# The factors affecting the psychometric function

# for speech intelligibility


**Alexandra MacPherson**


Submitted for the Degree of Doctor of Philosophy


Humanities and Social Sciences, School of Psychological Sciences and Health,
University of Strathclyde

MRC Institute of Hearing Research Scottish Section


2013

# Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.



Signed:
Date:

## Abstract

Older listeners often report difficulties understanding speech in noisy environments. Increasing the level of the speech relative to the background – e.g. by way of a hearing aid − usually leads to an increase in intelligibility. The amount of perceptual benefit that can be gained from a given improvement in signal-to-noise ratio (SNR), however, is not fixed: it instead depends entirely on the slope of the psychometric function. The shallower the slope, the less benefit the listener will receive. The aim of the research presented in this thesis was to better understand the factors which lead to shallow slopes.

A systematic survey of published psychometric functions considered the factors which affect slope. Speech maskers, modulated-noise maskers, and target/masker confusability were all found to contribute to shallow slopes. Experiment 1 examined the role of target/masker confusion by manipulating masker intelligibility. Intelligible maskers were found to give shallower slopes than unintelligible ones but subsequent acoustic analysis demonstrated that modulation differences between the maskers were responsible for this effect. This was supported by the fact that the effect was seen at low SNRs. Experiment 2 confirmed that the effects of modulation and target/masker confusion occur at different SNRs. Experiments 3 and 4 demonstrated that directing attention to the target speech could “undo” the effects of target/masker confusion. In Experiments 5 and 6 a new method was developed to study whether slope effects are relevant to “real-world” situations. The results suggested that using continuous speech targets gave shallower slopes than standard speech-in-noise tests. There was little evidence found to suggest that shallow slopes are exacerbated for older or hearing-impaired listeners.

It is concluded that in the complex demands of everyday listening environments the perceptual benefit received from a given gain in SNR may be considerably less than would be predicted by standard speech-in-noise paradigms.

## Acknowledgments

Firstly, I would like to say a big thank you to my supervisor Michael Akeroyd for the vast amounts of support, guidance and encouragement he has offered throughout the last few years. I would also like to thank the other members of my research panel, Steve Kelly and Kevin Durkin for their helpful comments and advice.

Thank you to all the staff, both past and present, at the MRC Institute of Hearing Research Scottish Section for all their help over the years. I am especially grateful to Alan Boyd, Chris Brennan-Jones, Kay Foreman, Fiona Guy, Neil Kirk, David McShefferty, and Sharon Suller for the many hours of listening that they have sat through between them and to Owen Brimijoin and Bill Whitmer for their invaluable words of wisdom on just about everything from Matlab programming to the perfect PhD writing tunes.

Last but not least, I want to say a massive thanks to Mum, Dad and Pete for all their love and support, it really has meant everything.

## Table of contents

# 1 Introduction

Many listeners find that as they get older they have increasing difficulty understanding speech in some situations. While following conversations in quiet environments pose few problems, the presence of background sounds can often have hugely detrimental effects on speech intelligibility (CHABA, 1988; Plomp & Mimpen, 1979). As most everyday listening situations contain at least some form of background noise these difficulties are far from trivial.

Several aspects of auditory functioning have been shown to deteriorate with age but a particularly common change is a loss of sensitivity to quiet sounds (CHABA, 1988; Davis, 1995). While simply amplifying quiet sounds cannot completely resolve speech-in-noise difficulties, listeners undoubtedly benefit from increases to the level of target speech in relation to that of an unwanted interferer (Plomp, 1986). Indeed, listeners often instinctively implement this simple tactic in everyday listening situations. When holding a conversation in a busy shop, for example, a listener can ask a talker to speak louder, or when watching television in a room where other people are talking the listener can increase the volume of the television. Much hearing aid technology essentially does a similar job. The aim of an aid is often to separate out and amplify a target signal while attenuating unwanted background noise – i.e. to improve the signal-to-noise ratio (SNR). A standard, modern hearing aid offers around a 2-3 dB improvement in SNR (Ricketts & Dittberner, 2002). A given SNR improvement will not, however, always translate to the same *intelligibility* improvement for the listener. Unlike level, which is an acoustical, physical variable, intelligibility is a perceptual variable and is, therefore, subjective. The relationship between the two is termed the psychometric function.

## 1.1 The psychometric function for speech intelligibility

The psychometric function describes the relationship between perceptual sensitivity and the acoustical level of a stimulus. Figure 1.1 is an illustrative example of a psychometric function for the intelligibility of masked speech, with proportion correct (a measure of perceptual sensitivity) plotted against SNR (a measure of relative stimulus level). Psychophysical theory suggests that as stimulus level increases sensitivity should increase monotonically (Macmillan & Creelman, 2005). Typically psychometric functions[1] are sigmoidal in shape and cumulative Gaussian or logistic distributions have traditionally been used as mathematical descriptions of these functions. Figure 1.1 also shows two fundamental parameters − threshold and slope – which can be calculated from the psychometric function. The threshold is the stimulus level required to meet some arbitrary perceptual criterion, e.g. 50% correct performance on a speech identification task, and is often referred to as the speech reception threshold. The slope is the *rate* at which the perceptual sensitivity grows with changes in the stimulus level. The slope of the psychometric function is important if we wish to understand the amount of perceptual benefit a listener is likely to gain from small changes in SNR; the kind that may be offered by a hearing aid, for example.

Many studies have demonstrated that speech reception thresholds (SRTs) vary greatly depending on the listening conditions or the listener (e.g. Carhart, Tillman, & Greetis, 1969; Duquesnoy, 1983; Festen & Plomp, 1990; Plomp, 1986; Plomp & Mimpen, 1979). The slope of the psychometric function for masked speech also varies, but the situations which result in these changes have been much less extensively studied. Figure 1.2 shows two psychometric functions with the same threshold but with different slopes. If the psychometric function has a steep slope, like the function in panel A, intelligibility will rapidly increase as SNR is increased. For such cases, therefore, large intelligibility benefits could be achieved with only small changes in stimulus SNR. On the other hand, if the psychometric function has a shallow slope, like the function in panel B, comparatively slower increases in intelligibility will occur with SNR changes. In these cases small changes in SNR may translate to little or no benefit in terms of intelligibility for the listener. If we wish to predict the magnitude of the intelligibility benefit that can be achieved from a gain in SNR then we need to know the slope of the psychometric function.

[1] The term psychometric function is used in this thesis to refer both to the *data* which relates SNR to perception and also to the mathematical *function* fitted to the data.

***Figure 1.1:*** *A diagrammatic illustration of a psychometric function for masked speech. The parameters of threshold and slope which can be derived from this function are also indicated.*

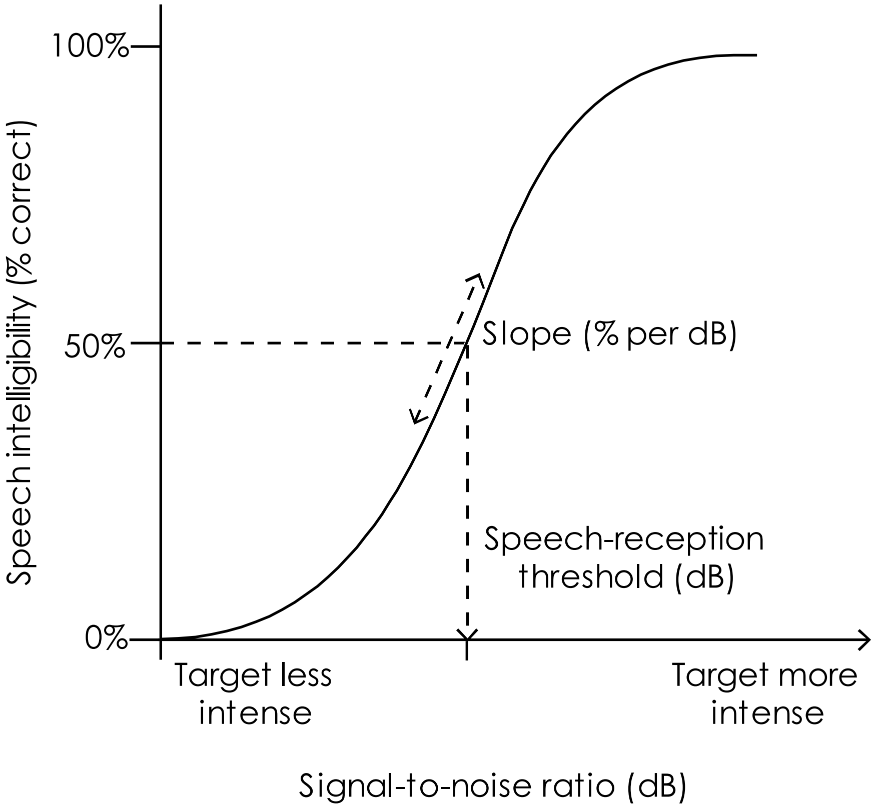
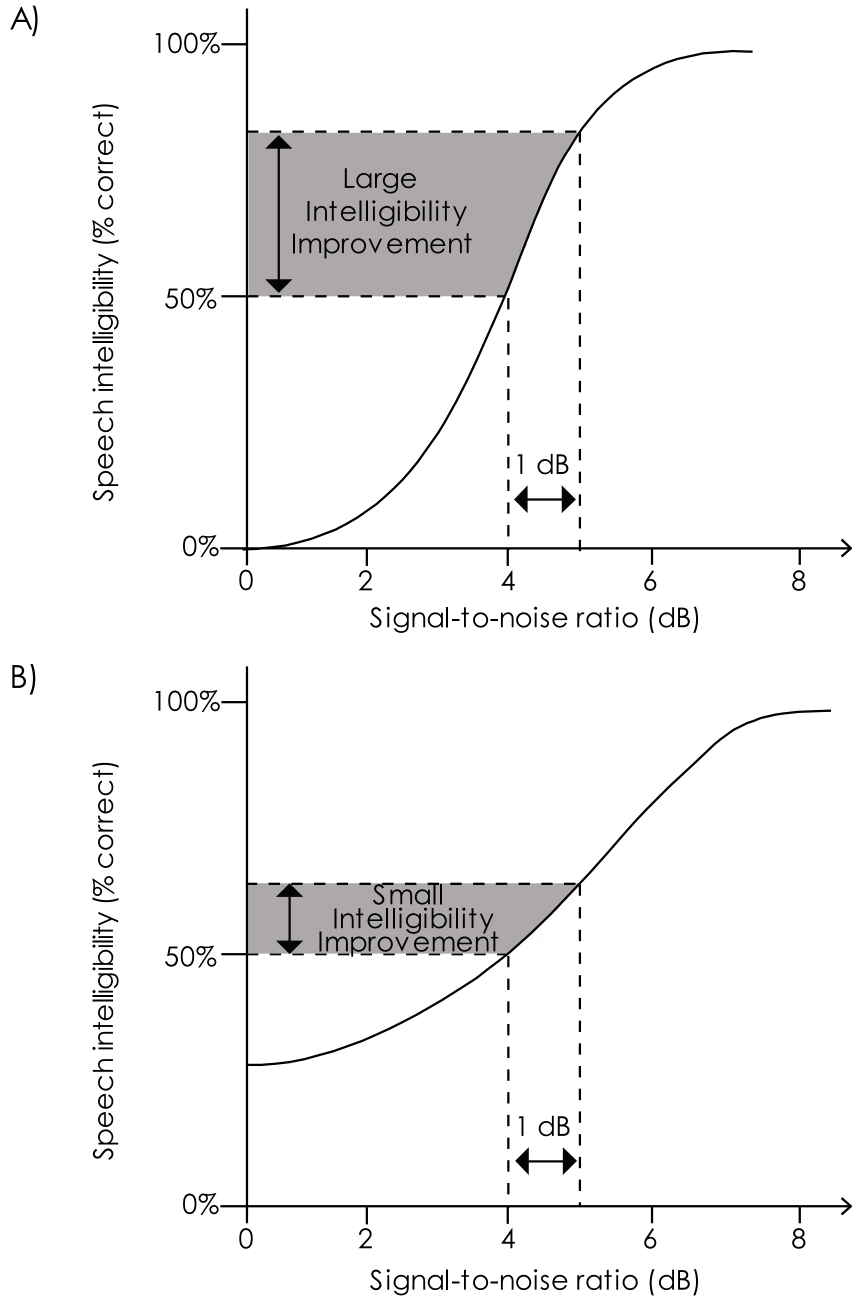


***Figure 1.2:*** *A diagrammatic illustration of two psychometric functions with the same 50% speech-reception threshold but with different slopes. Panel A illustrates the intelligibility improvement that can be gained if the slope is steep while panel B shows the intelligibility improvement gained if the slope is shallower.*



The research reported in this dissertation is concerned with (1) quantifying the variations in the slope of psychometric function for speech intelligibility, (2) understanding the conditions and mechanisms which may account for these variations in slope, (3) investigating whether realistic listening environments give shallower slopes (i.e. result in less sensitivity to changes in stimulus level) than conditions produced in standard audiological testing, and (4) considering whether shallow slopes are exacerbated in older or hearing-impaired listeners.

The remainder of this chapter and the following chapter is a review of the literature relating to psychometric functions for masked speech, focusing particularly on slope changes and summarising the methods used to measure psychometric functions, the proposed explanations and mechanisms for slope changes and the possible listening conditions which give shallow slopes.

## 1.2 Methods for measuring psychometric functions

All psychometric functions measured in this thesis, and most of those in the literature reviewed here, have used the method of constant stimuli. In this method, a range of SNRs are selected ranging from an SNR where the signal is clearly identifiable (ceiling, 100% performance) to an SNR where performance is at chance (floor, percent correct here depends on the number of response choices). Stimuli are then presented multiple times at each SNR in a quasi-random order. This prevents participants being able to predict the SNR of the next stimulus but ensures that all SNRs are heard an equal number of times. After each trial the listener is asked to make a judgment on what they heard: either discriminating between a set of possible responses or identifying the sound, word or sentence they just heard. Once these judgments have been made multiple times at each of the selected SNRs, the proportion of correct discriminations or identifications at each SNR is calculated. These data points can then be fitted with a sigmoidal function (e.g. a cumulative Gaussian or logistic function). The parameters of threshold and slope of the sigmoid are often calculated using maximum-likelihood procedures (Watson, 1979). The threshold can be computed by taking the inverse of the function for the required level of performance and the slope can be computed by calculating the gradient at this performance level (Wichmann & Hill, 2001b). In this thesis both threshold and slope are measured at 50% correct. For most psychometric functions, especially Gaussians, the slope will usually be at its maximum at 50% correct. While the method of constant stimuli provides reliable estimates of threshold and slope, a large number of trials are required to do so. This makes it a time-consuming method when several conditions are to be measured.

Estimates of slope can also be gained as a by-product of adaptive methods for measuring thresholds (Macmillan & Creelman, 2005). In adaptive procedures the presentation order of SNRs depends on the listeners' responses. In the down/up staircase procedure, for example, a starting SNR is chosen and this SNR is decreased with every correct response and increased after, usually, 2 or 3 successive incorrect responses (Levitt, 1971). When the direction of the SNR change is reversed after a correct or incorrect response this is termed a "reversal". The experiment usually runs for a fixed number of reversals and threshold is usually calculated by averaging the last few reversals (i.e. the first few reversals are usually not used). Adaptive methods are an efficient way to measure a single point of interest on a psychometric function, but with little information about performance at SNRs removed from the threshold they can give a less precise estimate of the shape of the psychometric function. Slopes can, however, be estimated based on performance at the different SNRs visited by the adaptive track. Several authors have suggested that the parameters of the adaptive task should be chosen carefully if the goal is to estimate slope. Levitt (1971) suggested, for example, that SNRs for trials should be placed one standard deviation away from the threshold whereas Leek, Hanna and Marshall (1992) proposed that the number of trials per track and the step size between presented SNRs must be considered if the estimated slope is to be reliable.

## 1.3 Factors affecting the slope of the psychometric function

If the intelligibility of speech depended wholly on its long-term SNR, it could be argued that the slope of the psychometric would be fixed regardless of other changes to either the target speech or the masking sound. That the slope of the psychometric does change, however, suggest that factors other than SNR also affect intelligibility. If we postulate that a complete dependence on SNR would give an "ideal" slope, any reduction in this dependence should result in a deviation in slope away from this ideal. A flattening of the psychometric function, therefore, represents instances where intelligibility becomes less dependent on SNR and more dependent on a second factor. Explanations of slope changes have been suggested in the literature and these can be grouped into four broad mechanisms. The following sections discuss the evidence for each of these slope-change mechanisms.

### 1.3.1 Slope changes as a consequence of fluctuating maskers

It has been proposed that changes to the shape of the psychometric function can be explained by temporal fluctuations in the masking sound. Festen and Plomp (1990), for example, presented short sentences in two types of masker; a steady noise and a noise whose amplitude had been modulated using the amplitude envelope extracted from a speech signal. They found that the slopes of psychometric functions were shallower when the amplitude modulated masker was used (11.9% per dB) than when the static noise maskers was used (21.0% per dB).

When target speech is presented in a fluctuating masker, there will be instances in which the speech sounds coincide with amplitude minima (or "dips") in the masking waveform. In these dips local SNR is improved allowing the listener to "glimpse" the target speech signal (Cooke, 2006; Miller & Licklider, 1950). Making use of these glimpses can greatly improve speech intelligibility and lower speech reception thresholds (Miller, 1947; Takahashi & Bacon, 1992; Wilson & Carhart, 1969). Glimpsing increases the SNR range over which target speech will remain audible (Rhebergen & Versfeld, 2005). Depending on the magnitude and duration of the fluctuations, glimpses of target speech may remain even as SNR is decreased. Small changes in SNR then would have less effect on intelligibility for a modulated masker than it would for a static masker. The result is a shallower psychometric function for modulated maskers (Speaks, Karmen, & Benitez, 1967). Models which are designed to predict speech intelligibility based on long-term spectra, such as the Speech Intelligibility Index (SII), have been extended to incorporate the effect that fluctuating maskers as opposed to static maskers have on the slope of the psychometric function (Rhebergen & Versfeld, 2005). The adjusted models have been found to be better predictors of actual speech identification data than previous models.

### 1.3.2 Slope changes as a consequence of target/masker confusion

Changes to the slope of the psychometric function have also been found to occur when a target and a masker become confused. The first notable piece of evidence to support this is the finding that speech maskers tend to give shallower slopes than noise maskers. Speech maskers are more similar to speech targets than static-noise maskers both in terms of their spectro-temporal and their linguistic features and it is argued that this similarity and the resultant confusion leads to changes in slope. Speaks et al., (1967) for example, reported that a static-noise masker gave a steep slope but that this slope was reduced if the masker was replaced with single-talker speech. Several other studies have repeated this finding with many different speech and noise stimuli (Brungart, 2001a; Dirks & Bower, 1969; Festen & Plomp, 1990; Freyman, Helfer, McCall, & Clifton, 1999; Wilson, Zizz, Shanks, & Causey, 1990; Wu et al., 2005; Yang et al., 2007).

To understand the effect on slope, it is useful to consider the general effects that similarity may have on speech-in-noise understanding. When speech is presented in a background of sounds, to be understood it first needs to be successfully segregated from the mixture. Several local spectro-temporal features are used to group together sounds that are likely to have originated from the same source. For example, sound elements with common onsets and offsets, those which are harmonically related, and those with common amplitude or frequency modulations are likely to be perceived as originating from the same source and grouped together into an auditory object (e.g. Darwin & Carlyon, 1995). Once individual auditory objects have been formed they must be linked together across time. The grouping of auditory objects, such as syllables, across time to form a speech stream has been referred to as "streaming"[2]

[2] The use of the term "streaming" here differs from its use in the majority of literature. Usually streaming refers to "denote the processes determining whether one stream or multiple streams are

(Shinn-Cunningham, 2008; Shinn-Cunningham & Best, 2008). Higher-order perceptual features such as pitch, timbre and location have been shown to be important for successful streaming of segregated auditory objects (Darwin, 1997; Darwin & Hukin, 2000). Next, the correct sound source (i.e. the target) must be selected and attended to (Shinn-Cunningham, 2008). A priori knowledge about the target stream, either lexical (i.e. the semantic content of the target) or indexical (i.e. features of the target voice such as location, pitch, intensity etc), can help to identify the target and to suppress the interfering stream (Best, Ozmeral, & Shinn-Cunningham, 2007; Helfer & Freyman, 2009; Kitterick, Bailey, & Summerfield, 2010).

A similarity between a target and a masker can cause interference at several levels of the speech understanding process. Similarity of spectro-temporal features can, for example, lead to difficulties segregating sound sources into individual auditory objects. This can result in a fusion of the different sound sources. Sounds which start and stop at the same time can often be perceived as arising from the same source (Bregman, 1990). If voices are similar, linking the syllables and words across time to form streams can also become difficult. The individual words from the mixture may be intelligible but a coherent meaning cannot be extracted due to interference from words from the competing message. Even if elements of target speech can be successfully grouped and streamed, any similarity between the target and the masking sounds can make it difficult to successfully select and attend to the target. Auditory attributes play an important role in allowing top-down attention to be directed correctly – for example features of the target voice, such as pitch and spatial location can be used to distinguish the target from a masker (Darwin, Brungart, & Simpson, 2003; Freyman et al., 1999) as can information about the linguistic content of the target or the masker (Helfer & Freyman, 2009). If target and masker are similar and these attributes are less well defined across the two sound sources (e.g. both voices have similar pitch, or semantic content is similar), there will be fewer cues available on which target selection can be based and attention can be focused.

---

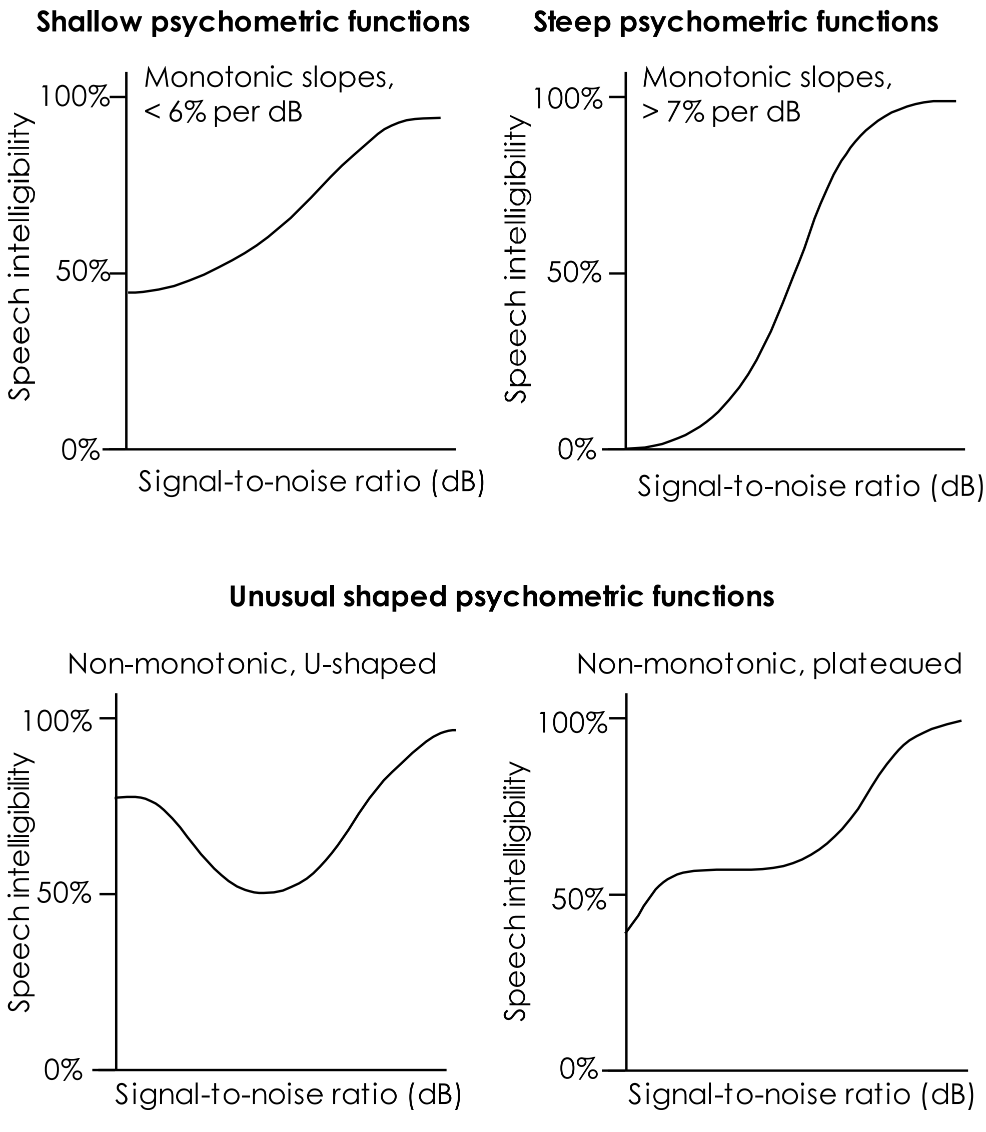
heard (Moore, 2012, page 300). Here it refers to the linking of temporally disjointed units or perceptual objects.

Further, it will be harder to inhibit or suppress the competing stream. The target will, therefore, become difficult to distinguish and the listener may become confused as to which stream is the target and which is the masker.

There is evidence to suggest that a high degree of similarity between targets and maskers results in a flattening of the psychometric function. The shallowest psychometric functions reported in the literature have been found to occur, for example, when targets and maskers are highly acoustically and/or linguistically similar. Egan, Carterette and Thwing (1954), for example, presented two competing speech sentences which were spoken by the same talker monaurally to listeners. They found that this condition gave unusual shaped psychometric functions: instead of the usual sigmoidal shape the function was "U" shaped. Figure 1.3 illustrates 4 different shaped psychometric functions and defines how these slopes are classified in this thesis. The figure shows that for the U-shaped function performance decreases as SNR is increased over the range of -10 to 0 dB, but begins to rise again as SNR is increased further.

Brungart (2001a) also reported non-monotonic psychometric functions when the target and masker speech were highly similar. In this study target and masker speech were taken from the Coordinate Response Measure (CRM) corpus (Bolia, Nelson, Ericson, & Simpson, 2000). CRM sentences are very similar with only three words differing from sentence to sentence; a call sign (e.g. Baron, Ringo) and two keywords (1 colour and 1 number), *"ready Ringo go to green three now"*. These sentences, therefore, give no semantic cues to aid in distinguishing the target from the masker. Brungart found that when target and masker sentences were spoken by the same person and presented monaurally, a plateau in performance occurred around 0 dB SNR. Figure 1.3 illustrates the psychometric function with a plateau and shows how performance remained constant at around 40% as SNR was increased from -12 to 0 dB but rapidly increased as SNR was increased above 0 dB. An analysis of the errors made by listeners showed that, when the SNR was around 0 dB, almost all of listeners' incorrect responses (roughly 90%) contained keywords from the masker. This result highlights the high degree of confusion between target and masker speech which occurred in these conditions.

***Figure 1.3***: *Diagrammatic illustrations of four different shaped psychometric functions and their classifications.*



Several other studies have also observed either U-shaped psychometric functions or functions with plateaus (e.g. Brungart, Chang, Simpson, & Wang, 2009; Brungart, Iyer, & Simpson, 2006; Brungart & Simpson, 2007; Cooke, Hershey, & Rennie, 2010; Darwin et al., 2003; Dirks & Bower, 1969; Freyman et al., 1999; Ihlefeld & Shinn-Cunningham, 2008a). This non-monotonicity can be reduced or even eradicated if the acoustic similarity between the target and the masker is reduced. Using different talkers of the same gender or talkers from the opposite gender to the target has been shown to reduce regions of non-monotonicity in the psychometric function (Brungart, 2001a; Brungart & Simpson, 2007), as has increasing the pitch differences between target and masker voices (Brungart & Simpson, 2007; Drullman & Bronkhorst, 2004). Introducing either a physical or perceived spatial separation between the target and the masker voices, so the signals no longer appear to come from the same location, has also been shown to eliminate the plateau or the U shape in the function (Freyman, Balakrishnan, & Helfer, 2001; Ihlefeld & Shinn-Cunningham, 2008b).

Reducing linguistic confusion between a target and a masker has also been shown to reduce non-monotonicity of the psychometric function. Many models for speech recognition suggest that the activation of word meanings is automatic on the presentation of spoken words (e.g. The Cohort Model, Marslen-Wilson, 1984). When target and maskers are both speech, then the masker will also activate linguistic and cognitive systems (Li, Daneman, Qi, & Schneider, 2004). It is possible that the semantic processing of the masker interferes with that of the target, making the target harder to select and attend to (Van Engen & Bradlow, 2007; Yost, 2006) and/or taxing cognitive resources which could otherwise be used to identifying the target speech (e.g. Francis, 2010). If the linguistic similarity between the target and the masker is removed, however, (i.e. by reducing the linguistic content of the masker) it follows that the interference which caused the confusion should also be reduced. Dirks and Bower (1969) noted U-shaped functions when target and masker speech were presented monaurally and spoken by the same talker. This dip in the function was found to be reduced if the masker speech was presented backwards. It was argued that while presenting the masker backwards reduced the temporal similarity between the target and the masker, it also eliminated semantic content in the masker, greatly reducing any linguistic interference with the target. Plateaus were also noted, however, when masker speech was played forward but spoken in a foreign language and it was concluded temporal similarity played a more important role as plateaus appeared regardless of intelligible semantic content in the masker. It has been subsequently suggested, however, that even maskers without meaningful lexical units, such as a foreign languages, may still engage speech understanding processes and interfere with target identification as they contain familiar phonological units. In other words, sources which are similar enough to speech may activate mechanisms which look for language-based information (Hawley, Litovsky, & Culling, 2004). This would mean that speech-like maskers may still induce linguistic confusion even if they are essentially unintelligible.

Freyman et al., (2001) also highlighted the role that reducing linguistic similarity could have on the slope of the psychometric function. It was demonstrated that a perceived spatial separation between target and masker speech had the greatest effect on slope if the masker speech was meaningful than if it was non-meaningful (either reversed or spoken in a foreign language). It was concluded that while intelligible semantic content played a role in difficulties attending to and identifying the target when stimuli were presented monaurally its role was negligible if speech was spatially separated. This result suggests then that linguistic similarity between a target and a masker may take a secondary role to acoustic similarity in terms of its effect on the slope of the psychometric function.

An increased dependence on the difference in level between a target and masker rather than on the overall SNR is the commonly suggested mechanism for *how* confusion — either linguistic or acoustic in basis — flattens the slope of the psychometric function. When few cues are available to distinguish the target from the masker, i.e. when they are both speech spoken by the same person, when they are presented monaurally, or when they are linguistically very similar, differences in level become a vital discrimination cue (e.g. Brungart, 2001a; Dirks & Bower, 1969; Egan et al., 1954). When the target is either less intense or more intense than the masker, this level difference can be used to identify the target speech. As the level difference between the two sources is reduced, however, as would be the case if SNR is increased from -10 dB to 0 dB, this cue begins to disappear. Despite an increase in the level of the target, the loss of a cue to aid discrimination means that performance remains either unchanged (i.e. a plateau) or decreases slightly (U-shaped) as SNR approaches 0 dB.

The additional masking that occurs when targets and maskers are either acoustically or linguistically similar is often referred to as "informational masking". A link with target/masker confusion has led shallow slopes to be associated with this type of masking. Informational masking (IM) is, however, an umbrella term usually used to describe any masking effects which cannot be adequately explained by another type of masking − "energetic masking" (e.g. Arbogast, Mason, & Kidd, 2002; Durlach et al., 2003). Energetic masking (EM) usually refers to masking that occurs at the periphery of the auditory system and that is the consequence of spectro-temporal overlap of the target with the masking sound. In essence, neurons responding to a particular frequency range will not be able to adequately represent elements of the target signal if more powerful masking elements are also competing for representation at the same neurons. The amount of EM experienced in different listening situations can, therefore, be largely estimated by filter-bank models of the auditory periphery (Moore & Glasberg, 1983). IM on the other hand usually refers to interference occurring "centrally" (e.g. Durlach et al., 2003) or at "higher levels" (e.g. Brungart, 2001a) in the auditory system and is often further classified as occurring due to stimulus uncertainty (Neff & Dethlefs, 1995; Oh & Lutfi, 2000) or stimulus similarity (Watson, 2005). Assessing the amount of IM that is experienced in a particular situation has proved to be harder to measure than EM. Shallow psychometric functions have, however, often been used as indicator of the occurrence of IM (e.g. Arbogast et al., 2002; Brungart, 2001a; Festen & Plomp, 1990; Freyman, Balakrishnan, & Helfer, 2004; Freyman et al., 1999; Kidd, Mason, Rohtla, & Deliwala, 1998).

There are, however, several issues with attributing the occurrences of shallow slopes *directly* to IM. First, IM is a very broad term encompassing the additional masking, interference and competition that can occur when a target is heard in the presence of a masker (Yost, 2006). It is not clear, therefore, whether all aspects of IM will result in a flattening of the slope of the psychometric function, nor whether all cases of shallow functions can indeed be attributed to IM per se. A second related issue is that, as there is still much debate as to what constitutes and causes IM, simply labelling shallow slopes as IM does not help explain the mechanisms underlying the slope changes or help give an accurate prediction about which listening situations result in shallow slopes. As such, while this thesis will consider much of the IM literature and touch on several factors related to IM, it will not be used by itself to explain shallow slopes. Instead, defined factors offering more specific mechanisms for changes to the shape of the psychometric function will be focused on.

### 1.3.3 Slope changes as consequence of the availability of top-down information

Qualities of the target speech, as well as those of the masker, have also been found to result in changes to the shape of the psychometric function. Miller, Heise and Lichten (1951) demonstrated, for example, that manipulating the size of the vocabulary set from which keywords could be selected affected slope with shallower slopes found as the set size was increased. It has also been demonstrated that manipulating sentence context can also affect slope. Kalikow, Stevens and Elliot (1977) developed a speech corpus, the Speech Perception in Noise (SPIN) test, to directly measure the effect of sentence context on speech identification. The corpus contained sentences where the last word could be strongly predicted by the previous context of the sentence ("probability high, "PH" – e.g. "I've got a cold and a sore *throat*") and also sentences where the last word could not be predicted from previous context ("probability low, PL" – e.g. "He is considering the *throat*"). It was found that PL test items gave psychometric functions that were considerably shallower than those given by PH test items. Several other studies have used the same speech corpus and have replicated this effect (Dirks, Bell, Rossman, & Kincaid, 1986; Dubno, Ahlstrom, & Horwitz, 2000; Elliott, 1979; Lewis, Benignus, Muller, Malott, & Barton, 1988; Pichora-Fuller, Schneider, & Daneman, 1995).

It has been suggested that the changes in slope which are seen when vocabulary size and sentence context are manipulated can be explained by the relative contributions of perceptual and cognitive factors (Pichora-Fuller et al., 1995). If no previous context is available, or if the vocabulary set is large, listeners must rely on perceptual, or bottom-up, information to identify target speech. Small increases in acoustic information, such as could be gained by an improvement in SNR, will lead to small increases in intelligibility. If, however, context is available, or if the vocabulary set is initially very small, speech identification does not need to rely *solely* on bottom-up information, as the top-down information can be employed to constrain the possible speech elements (phonemes/words) that are available as responses. Small increases in acoustic information may be sufficient to further constrain possible speech elements, thus increasing the probability that those elements will be guessed correctly (Bronkhorst, Bosman, & Smoorenburg, 1993).

### 1.3.4 Slope changes as a consequence of underlying variation.

It has been argued that changes in slope can reflect underlying variation or heterogenity either in listeners' decision making processes, the stimuli, or in the slope measurement method itself. Wichmann and Hill (2001b) suggested that stimulus independent mechanisms such as lapses in attention or guessing can affect estimates of both threshold and slope. They demonstrated through simulations that introducing errors at high stimulus levels (i.e. levels where performance might be expected to be at ceiling) led to a considerable decrease in the estimated slope value. Indeed, the literature concerned with the identification of tones-in-tones commonly attribute shallower slopes to listener inattention (e.g. Allen & Wightman, 1995).

Studies looking at speech-in-noise intelligibility have also queried whether shallow psychometric results reflect underlying sensory processes, or whether they instead represent variability in listeners' performance. Jonstone and Litovsky (2006) used an adaptive method to measure speech identification in several noise and spatial conditions for children and adults. They found that psychometric functions for children were steeper than those given by the adults. They also found that adaptive tracks where performance at individual stimulus levels was not consistent (tracks that "wandered") gave shallow psychometric functions. It was suggested that inattention could contribute to these shallow slopes and moreover that it could explain the individual differences in slopes values noted within listener groups. It was proposed that the slope differences noted across groups could be attributed to the likelihood of listeners making guesses at the correct answers. The extremely steep slopes given by the children could be taken as indication of a reluctance to guess when only a portion of the stimuli was heard, explaining why performance went from 0% to 100% with only a small increase in SNR. Slopes were shallower for adults as they were willing to make partial guesses meaning mid-range performance was also measured. This resulted in a seemingly shallower increase in intelligibility with SNR.

Conversely, factors which *reduce* variability in listeners' responses or in the method of measurement have been found to steepen the slope of the psychometric function. It has been suggested, for example, that increasing the number of trials used in adaptive procedures reduces the variability in the measurement of the underlying function (Leek et al., 1992) and indeed slopes have been found to be steeper when more trials, rather than fewer, are used (Saberi & Green, 1996).

### 1.3.5 Slope changes: Aging and hearing loss

It is not uncommon for older and hearing impaired listeners to struggle with speech understanding in noisy environments and many of them rely on hearing aids to provide an improvement in speech audibility. That the slope of the psychometric function is not fixed and instead depends on the listening situation is particularly pertinent for these listeners as any change in slope will relate directly to the amount of benefit they might expect to receive from their hearing aid.

The effects that aging and hearing impairment themselves have on the slope of the psychometric function are not, as yet, clear cut although there is evidence to suggest that slopes may differ between older and younger listeners. Wagener and Brand (2005) found, for example, that when looking at functions given by both static and modulated maskers at a range of different target presentation levels, slopes were on average shallower for older hearing-impaired listeners than they were for young-normal hearing listeners (median slope = 14.9% per dB and 17.3% per dB respectively). Bosman and Smoorenburg (1995) also reported shallower slopes for older hearing-impaired listeners, this time for static noise maskers, and Wilson, Carnell and Cleghorn (2007) reported a similar result for multi-talker babble maskers. Conversely, Wilson, McArdle et al., (2010) using interrupted-noise maskers reported little or no change in the slopes of the psychometric function given by young normal-hearing and older hearing-impaired listeners.

## 1.4 Summary

Slope changes have been noted in the literature and these can be grouped into four broad slope-change mechanisms. How these different mechanisms interact with one another to affect the slope of the psychometric function is, as yet, unclear. As a result of this, the specific listening conditions which may result in shallower psychometric functions, and therefore the situations where any applied gain to speech will be of less benefit to the listener, have not yet been identified. To quantify how much the slope of the psychometric function can vary across experimental designs and conditions and to gain a better insight into the conditions where shallow slopes are likely to occur, a systematic survey of the literature on psychometric functions for speech intelligibility was carried out.

# 2 A systematic survey of slopes

Many studies have looked at the factors which can affect the intelligibility of speech. A large proportion of these studies have done so by measuring psychometric functions and, as a result, there is a wealth of psychometric-function data available in the literature. A systematic corpus of this data would be an extremely useful resource for isolating and identifying factors associated with changes in slopes. Most of the published analyses of this data, however, focus on changes in *threshold*, with slope changes less commonly calculated and reported. Those studies which have only represented the psychometric function as a single value (i.e. the speech reception threshold) cannot directly tell us anything about the shape of the function, but if the full psychometric function has been published the slope of the function can still be calculated. Here the psychometric data presented in the literature is reanalysed in order to calculate psychometric function slopes, quantify the amount slopes vary across and within studies, and to identify the factors which lead to shallower slopes. Importantly the reanalysis uses a uniform method so enabling a direct comparison of data originating from different studies.

## 2.1 Method

### 2.1.1 Procedure

A computerized literature search was undertaken to find studies which had measured the intelligibility of speech as a function of SNR. Initially, Web of Science was searched for articles citing either Egan et al., (1954) or Brungart (2001a) − two key studies which found unusual shaped psychometric functions of masked speech. The reference list of Brungart's paper was also reviewed for possible studies to include in the survey. The first reports of common speech tests and the studies citing these speech tests were also reviewed as many of these studies include psychometric functions in different noise conditions. Other miscellaneous studies containing psychometric functions which were found over the course of the slope survey were also included. No studies after a cut off date of February 2012 are included.

Note that one branch of the psychometric function literature − that of tones-in-noise − was completely excluded from the slope survey. The slopes of psychometric functions for tone identification have been shown to vary (e.g. Allen & Wightman, 1994; Kidd, Mason, & Arbogast, 2002; Kidd et al., 1998; Lutfi, Kistler, Callahan, & Wightman, 2003) and as such this literature may provide further insights into the factors affecting slope. As the main aim of this thesis, however, is to consider how small changes in level affect *speech* identification, the tonal literature was not considered.

To be included in the slope survey, studies needed to include at least one psychometric function for speech identification which was 1) measured as function of signal-to-noise ratio (SNR) or some other unit of relative presentation level, 2) measured over at least three points, 3) presented clearly in graphical or tabular form and 4) averaged over several listeners. Individual data was excluded because this data tended to be practically hard to measure (multiple overlaying psychometric functions), harder to code (listener details often specified for listener groups rather than the individual listener) and subject to large amounts of individual difference.

The citation searches for Egan et al., (1954) and Brungart (2001a) produced 56 and 210 citations respectively. Brungart (2001a) cited a further 26 references. Thirty seven speech tests were reviewed and 2493 studies citing these speech tests were also found. After duplicates were accounted for, the remaining citations underwent a short initial review: titles and abstracts were checked and studies on unrelated topics excluded. The remaining studies were then assessed in more detail through manuscript review. Twelve studies which could not be accessed were excluded at this stage, nine studies were excluded because they presented modelled data, and ten studies were excluded because the published psychometric functions were taken from a previous study.

Of these studies 154 were included in the final survey. On average each of the included studies contained five or six psychometric functions which conformed to the inclusion criteria. A total of 1137 individual slopes were measured in all.

To gather as much detail as possible about the conditions which result in shallow slopes each psychometric function in the survey was subjected to detailed stimulus coding. The following information was coded for:

- Target speech corpus (see below for more information on sub-categories).
- Masker type (sub-categories = speech, modulated noise[3], or static noise).
- Number of maskers.
- Presentation of stimuli (sub-categories = monaural, diotic or dichotic).
- Spatial locations of target and masker.
- Target language.
- Target predictability (sub-categories = high predictability from context, low predictability from context).
- Whether target was primed before presentation or not.
- Processing of target or masker (sub-categories = vocoded, filtered or added reverberation).

If the masker was speech then further coding was carried out, including:

- Masker language.
- Masker corpus.
- Gender of the masker talker in relation to the target talker (sub-categories = same gender, different gender, same talker).
- Masker intelligibility (sub-categories = intelligible or unintelligible).

[3] Modulated maskers included all maskers with temporally fluctuating amplitude, regardless of the type and spectral shape of modulation.

- Masker uncertainty (sub-categories = Masker talker fixed from trial to trial, Masker content fixed from trial to trial).

- Pitch shift between target and masker voices (sub-categories = "small" if < 3 semitones, "medium" if 4-7, or "large" if > 8).

Finally, general information about the studies' participants was also coded:

- Age group (sub-categories = children, young adult or older adult).

- Hearing loss (sub-categories = Normal hearing, a reported hearing loss, cochlear implant user).

Some categories needed more extensive classification, for example, "speech corpus". Not all targets or maskers were taken from a specific standardised speech corpus. In these cases stimuli were categorised more generally as: *continuous speech*, *valid sentences*, *invalid sentences*, *words*[4], *digits*, or *short tokens*. *"Valid sentences"* described any stimuli consisting of syntactically and semantically correct sentences (e.g. sentences read from a history text book). *"Invalid sentences"* described any stimuli consisting of either syntactically incorrect sentences ("cat on sat the mat") or semantically incorrect sentences ("the thorn can wake the kettle"). "*Continuous speech"* described any speech stimuli longer than a single sentence, whereas *"short tokens"* described smaller speech units such as syllables and phonemes.

### 2.1.2 Analysis

For each psychometric function the individual data points were recorded. These values were either taken directly from the paper if the psychometric functions were reported in tabular form, or extracted using a custom written Matlab programme[5] if the psychometric functions were displayed graphically.

[4] Included VCV words.

[5] Uses the Matlab function "ginput" which gives X and Y coordinates for each mouse click.

The raw data points were then fitted with a logistic function using a non-linear least squares method (using Microsoft Excel Solver[6]):

$$P = 100\left(\frac{1}{1 + e^{m(x\text{-}c)}}\right) \tag{2.1}$$

where $x$ is the SNR level (decibels), $p$ is the percentage of correctly identified items at this level, $m$ and $c$ are constants: c is the SNR at which p = 50% correct and m is the slope of the function at $x$ = c. The derivative of the logistic function is:

$$\frac{dp}{dx} = \frac{\text{-}100m.e^{\,m(x\text{-}c)}}{(1 + e^{m(x\text{-}c)})^2} \tag{2.2}$$

which for $x = c$ is:

$$\begin{aligned} \frac{dp}{dx} &= \frac{\text{-}100m.e^{m.0}}{(1 + e^{m.0})^2} \\ &= \frac{\text{-}100\ .\ m}{(1+1)^2} \\ &= \text{-}25m \end{aligned} \tag{2.3}$$

For consistency, none of the logistic fits were corrected for either chance percent correct or maximum percent correct due to lack of information on this in many of the

[6] Solver uses the Generalized Reduced Gradient algorithm which, given certain constraints, iteratively optimises a target cell by adjusting selected parameters (Microsoft, Excel user's guide 2012).

studies found in the survey. The values of $m$[7] and $c$ were added to the database with the coding information. The reliability of the parameters $m$ and $c$ calculated by this method is covered via a bootstrap analysis in Chapter 3 (section 3.3.4).

To assess the fit of the logistic function to the original data points, a root-mean-square (RMS) error value was calculated for each function:

$$RMS\ error = \sqrt{\frac{\sum^{n}(p_1 - p_2)^2}{n}} \tag{2.4}$$

where $n$ is the number of data points in the psychometric function, $p_1$ is the percent correct value from original data at each data point and $p_2$ is the percent correct value predicted by the logistic function.

On the whole the fit to the data was regarded as good (mean RMS = 3.2%). There were, however, a small number of psychometric functions (n = 29) where the logistic curve was a poor fit to the data, with RMS values of 10% or greater[8]. These psychometric functions were excluded from the survey at this stage as slope values calculated from these logistic functions were unlikely to be good representations of the slope of the raw data. These exceptions will be considered separately in section 2.4.2.

As mentioned above, the slope parameter $m$ is calculated at $P$ = 50% correct. Several of the psychometric functions published in the literature reported performance where all data points were below 50% or above 50% correct. In these cases, $m$ defines an extrapolated logistic function. As such, slopes calculated in this way are also unlikely to be good representations of the true slope of the data, and were also excluded from further analysis (n = 195).

[7] m was converted to slope in % per dB for further analysis.

[8] An RMS cut off of 10% was chosen as it clearly represented the extreme cases where the logistic function was a poor fit to the data. Three quarters of the fitted functions in the slope survey had RMS values 5% or less, for example, and figure 3.9 shows that in the experimental work later in this thesis mean RMS values were around 6%.

To identify possible trends in the slope data, the distributions of slope values for different conditions were considered. A statistical analysis was also carried out to consider whether slope distributions for these conditions significantly differed from one another. As the data was not normally distributed non-parametric approaches were used. For conditions with more than two samples a Kruskal-Wallis one-way analysis of variance by ranks was carried out, followed by post-hoc Mann-Whitney tests (Field, 2009). For conditions with just two samples only the Mann-Whitney test was carried out. These methods test whether samples originate from the same distribution and while they do not assume samples are normally distributed they do assume that the distributions have the same shape (Fagerland & Sandvik, 2009). On the whole the slope distributions for different samples are shaped similar to that of the overall distribution (see section 2.2) and as such it was considered appropriate to use non-parametric methods here. Several of the specific distributions do not reflect the overall distribution in terms of shape but it is believed that this is due to the limited number of cases found for particular conditions. It is argued the specific distribution would resemble that of the overall distribution had a greater number of cases been found.

It was reasoned that as effect sizes for main effects are not that useful, reporting individual effect sizes for each pair of samples (i.e. for each post hoc analysis) would be more informative. The effect size r is, therefore, reported for each post-hoc Mann-Whitney test and interpreted in line with Cohen's estimates of effect sizes[9].

$$r = \frac{Z}{\sqrt{n}} \tag{2.5}$$

### 2.1.3 Comment on multiple comparisons

One of the main problems associated with using the slope survey to identify significant trends in slope is that multiple comparisons of the data must be made to do so. Multiple comparisons inflate the probability of incorrectly rejecting the null

[9] Small effect (r = 0.10), medium (r = 0.30) and large (r = 0.50) effect sizes.

hypothesis and so making a type 1 error. For a given α the probability of accepting at least one comparison as significantly different by chance is equal to:

$$1-(1-\alpha)^{n\ comparisons} \tag{2.6}$$

So for a single comparison using a significance level of 0.05 the probability that a non-significant difference between means would be accepted by chance is 0.05, but this probability increases to 0.14 with three comparisons and to 0.26 with six comparisons. A common method to deal with problems of multiple comparisons is to adjust the level of α so that the probability of making a type 1 error does not increase as the number of comparisons does. The Bonferroni correction adjusts α by dividing by the number of comparisons. So the chance of making a type 1 error using a Bonferroni correction deviates far less from 0.05 than it would had the correction not been made.

However, there are two problems with this method. Firstly, adjustments such as the Bonferroni correction are extremely conservative. Secondly, with the current data there is also the question of which tests – and therefore the divisor − to apply the correction to. There is an argument, for example, for just applying the correction within each set of comparisons (e.g. for the post-hoc comparisons made for the effect of masker type). There is equally an argument that the correction should be applied across all comparisons made in the slope survey, as the analysis of the survey is in essence multiple comparisons across the same data. If the Bonferroni correction was applied to all comparisons made in the survey, however, the significance level needed to be reached before accepting a slope difference would be $\alpha = 0.0019$.

Having considered the effects that making multiple comparisons can have on significance levels it was decided that applying conservative measures to avoid type 1 errors would be excessive in the current case. The primary purpose of the slope survey was to quantify trends, and so it was reasoned that allowing occasional type 1 errors would be acceptable. It was feared that applying a Bonferroni correction to all comparisons would limit the information that could be gained from the survey. While no adjustment was made for multiple comparisons, significant differences were treated warily and a greater emphasis was put on effect sizes to judge the relative importance of an effect (Nakagawa, 2004).

### 2.1.4 Comment on different sample sizes

Collecting data from such varied sources meant that there was large variation in the details of the stimuli used in different studies. It also meant that sample sizes for different categories and sub-categories (both in terms of the number of studies found and in individual psychometric functions) also varied greatly. More commonly used conditions had much larger sample sizes than the less commonly used conditions; while some sub-categories were based on 50 or more psychometric functions others were based on just three. The presence of unequal sample sizes had consequences for making comparisons across categories or sub-categories. While the Kruskal-Wallis test allows for unequal sample sizes, as the difference between samples sizes gets larger, $\chi^2$ values become more conservative (Meyer & Seaman, 2008). Comparisons across conditions with very unequal sample sizes were, therefore, treated tentatively.

## 2.2 Major trends in slope

After psychometric functions with high RMS values and those with slope values based on extrapolation were removed, 909 individual slope values remained in the survey, taken from 140 different studies. Table 2.1 (see end of chapter) summarises the general stimuli and participant information for each study. It can be seen that: 1) overall, 24 different speech corpuses were found plus several general classes of speech stimuli, 2) speech maskers tended to be taken from the same corpus category as the target, and 3) the participants in the majority of studies were young, normal hearing adults.

Figure 2.1 shows the overall distribution of slope values from the survey. The slope data has been fitted with a log-normal distribution as it gave an excellent fit to the data. There was a wide variation in the slope values found in the slope survey – the minimum and maximum values were 1% per dB and 29% per dB. The mean was

7.4% per dB and the median was 6.4% per dB. There was a clear positive skew, with the bulk of values, including the median, lying to the left side of the mean.

Table 2.2A reports the median slopes and interquartile ranges of psychometric functions measured for the six general classes of speech stimuli and the seven most-commonly reported speech tests when different types and numbers of maskers were used. Table 2.2B reports the number of studies and the number of individual psychometric functions that these values are based on. Note that as less than 20 psychometric functions each were found for the 17 remaining speech tests the data for these have been combined, and appear as "other" in tables 2.2A and 2.2B.

As mentioned in section 2.1.4, data could not be collected for all conditions and sample sizes do vary greatly, but where a given target was found in all three types of noise several overall trends in median slope values can be observed. Firstly, the type of masker affects slope, with speech maskers giving shallower slopes than either of the noise maskers. Secondly, the number of maskers affects the slope of the psychometric function, with one masker giving shallower slopes than when multiple maskers were used. These trends are discussed in detail in the following sections (sections 2.2.1 and 2.2.2).

With 909 individual psychometric functions it is not too surprising to find substantial variations across details of stimuli, maskers and other aspects of experimental design. The analysis here, therefore, concentrates on broad categories rather than on the highly specific individual combinations. Nevertheless, the coding schema means that many more specific questions could be asked. For example, we could consider the effect that CRM sentences in two-talker, different gender speech had on the slope of the psychometric function[10].

### 2.2.1 Major trend 1 - Type of masker

The first notable trend in the slope survey data is that speech maskers give shallower psychometric functions than either modulated-noise maskers or static-noise maskers. This trend is particularly evident in Table 2.2 when looking at the data for conditions

[10] This combination of conditions gave a median slope of 5.8% per dB.

***Figure 2.1:*** *A histogram displaying the distribution of the slope values measured in the systematic slope survey (n = 909). The data has been fitted with a log-normal distribution and mean and median slope values are indicated.*

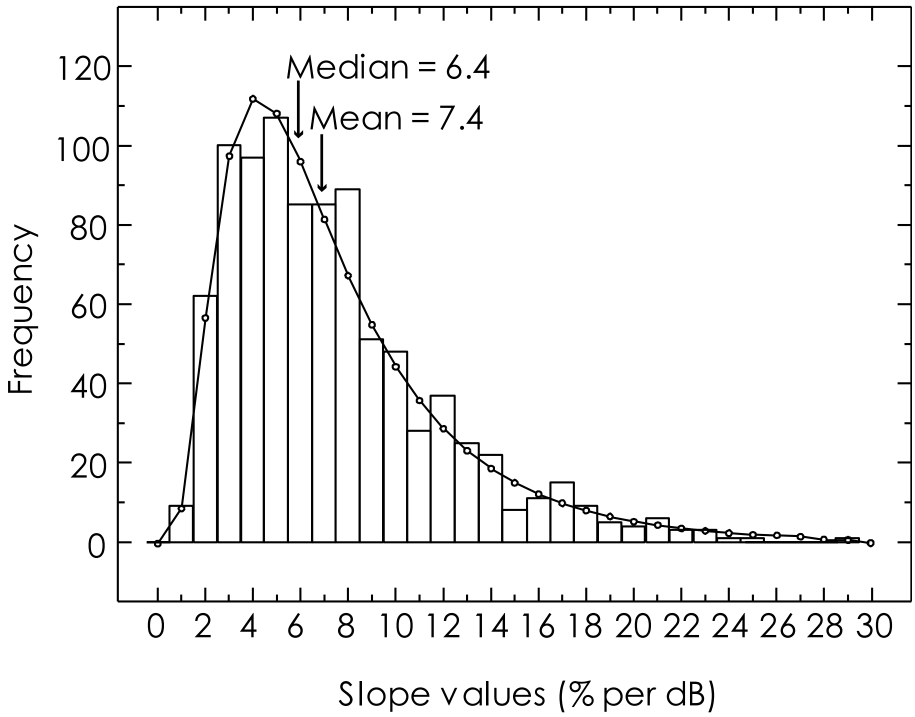


***Table 2.2A:*** *Median slope values (in bold) of psychometric data found for each of the general target/masker combinations identified in the systematic slope survey. Interquartile ranges for each condition are also included (in brackets and italics).*

| | Target | | | | | |
|---|---|---|---|---|---|---|
| **Masker** | Short tokens | Words | Digits | Valid sentences | Invalid sentences | Continuous speech |
| 1 speech masker | **5.1** *(3.3)* | **6.7** *(2.2)* | - | **2.6** *(1.2)* | **6.4** *(3.6)* | **5.7** *(-)* |
| 2 speech maskers | **7.9** *(2.2)* | **7.7** *(-)* | - | **7.7** *(5.5)* | **6.3** *(1.4)* | - |
| 3 speech maskers | - | - | - | - | **15.1** *(-)* | - |
| 4+ speech maskers | **2.3** *(-)* | **7.5** *(2.1)* | **9.9** *(3.6)* | **9.5** *(0.8)* | **8.6** *(4.5)* | - |
| 1 modulated masker | **4.4** *(1.2)* | **8.6** *(6.1)* | - | **13.5** *(-)* | **7.1** *(1.7)* | - |
| 2 modulated maskers | - | - | - | **5.0** *(-)* | - | - |
| 3 modulated maskers | - | - | - | - | - | - |
| 1 static noise masker | **4.0** *(3.4)* | **8.2** *(6.3)* | **13.7** *(6.3)* | **12.3** *(6.9)* | **7.9** *(1.6)* | **7.4** *(7.4)* |
| 2 static noise maskers | - | - | - | - | **8.8** *(1.0)* | - |
| 1 speech & 1 modultated maskers | - | - | - | **2.7** *(-)* | **1.9** *(-)* | - |
| 1 speech & 1 static noise maskers | - | **15.2** *(-)* | - | - | - | - |

*Table 2.2A continued.*

| | Target | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Masker** | CRM | HINT | IEEE | NU-6 | PB Lists | SPIN | SSI | Other |
| 1 speech masker | **3.7** *(1.5)* | **3.4** *(2.0)* | **4.5** *(1.3)* | - | - | **8.7** *(-)* | **4.6** *(3.2)* | **2.4** *(3.3)* |
| 2 speech maskers | **3.6** *(2.6)* | - | **4.3** *(-)* | - | - | - | - | **9.8** *(3.4)* |
| 3 speech maskers | **9.2** *(2.4)* | - | - | - | - | - | - | - |
| 4+ speech maskers | - | - | **15.9** *(-)* | **6.2** *(3.3)* | - | **8.2** *(7.2)* | **13.8** *(4.5)* | **9.2** *(3.8)* |
| 1 modulated masker | **5.8** *(2.2)* | **4.6** *(-)* | **4.3** *(1.1)* | **5.5** *(2.5)* | - | **3.1** *(5.5)* | **14.7** *(7.4)* | **11.2** *(-)* |
| 2 modulated maskers | **8.1** *(-)* | - | - | - | - | - | - | - |
| 3 modulated maskers | **10.2** *(-)* | - | - | - | - | - | - | - |
| 1 static noise masker | **10.7** *(3.5)* | **9.1** *(5.5)* | **5.0** *(4.0)* | **5.2** *(2.5)* | **6.6** *(3.4)* | **4.7** *(7.2)* | **17.1** *(4.2)* | **6.1** *(7.3)* |
| 2 static noise maskers | - | - | - | - | - | - | - | - |
| 1 speech & 1modulated maskers | **12.0** *(-)* | - | - | - | - | - | - | - |
| 1 speech & 1 static noise maskers | **4.5** *(1.8)* | - | - | - | - | - | **13.6** *(-)* | - |

***Table 2.2B:*** *Number of studies (in bold) to report psychometric data for each of the general target/masker combinations identified in the systematic slope survey. The number of Individual slopes measured in each condition is also included (in italics).*

| | Target | | | | | |
|---|---|---|---|---|---|---|
| **Masker** | Short tokens | Words | Digits | Valid sentences | Invalid sentences | Continuous speech |
| 1 speech masker | **2**/*10* | **1**/*4* | - | **5**/*43* | **4**/*13* | **1**/*2* |
| 2 speech maskers | **1**/*6* | **1**/*3* | - | **3**/*9* | **5**/*38* | - |
| 3 speech maskers | - | - | - | - | **1**/*2* | - |
| 4+ speech maskers | **2**/*3* | **5**/*13* | **2**/*13* | **2**/*7* | **3**/*7* | - |
| 1 modulated masker | **3**/*6* | **3**/*22* | - | **2**/*3* | **2**/*5* | - |
| 2 modulated maskers | - | - | - | **1**/*1* | - | - |
| 3 modulated maskers | - | - | - | - | - | - |
| 1 static noise masker | **15**/*59* | **22**/*67* | **4**/*20* | **12**/*30* | **9**/*40* | **2**/*5* |
| 2 static noise maskers | - | - | - | - | **1**/*4* | - |
| 1 speech & 1modulated maskers | - | - | - | **1**/*3* | **1**/*2* | - |
| 1 speech & 1 static noise maskers | - | **1**/*1* | - | - | - | - |

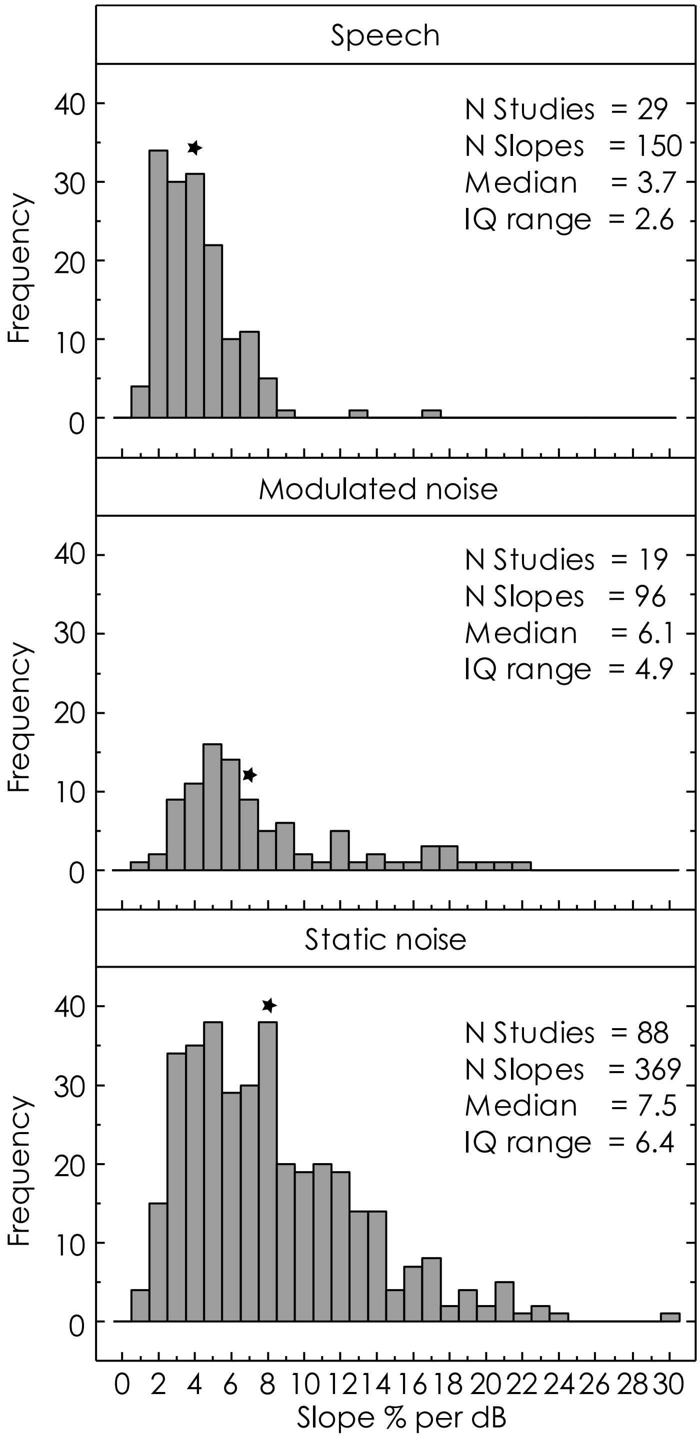
*Table 2.2B continued.*

| | Target | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Masker** | CRM | HINT | IEEE | NU-6 | PB Lists | SPIN | SSI | Other |
| 1 speech masker | **11**/*47* | **1**/*6* | **2**/*10* | - | - | **1**/*1* | **2**/*10* | **2**/*4* |
| 2 speech maskers | **9**/*50* | - | **1**/*3* | - | - | - | - | **1**/*10* |
| 3 speech maskers | **5**/*21* | - | - | - | - | - | - | - |
| 4+ speech maskers | - | - | **2**/*3* | **3**/*19* | - | **7**/*31* | **1**/*4* | **8**/*26* |
| 1 modulated masker | **3**/*21* | **1**/*3* | **1**/*4* | **1**/*12* | - | **2**/*6* | **1**/*12* | **2**/*2* |
| 2 modulated maskers | **1**/*1* | - | - | - | - | - | - | - |
| 3 modulated maskers | **1**/*1* | - | - | - | - | - | - | - |
| 1 static noise masker | **5**/*9* | **6**/*19* | **4**/*19* | **6**/*10* | **8**/*27* | **4**/*9* | **2**/*8* | **22**/*47* |
| 2 static noise maskers | - | - | - | - | - | - | - | - |
| 1 speech & 1modulated maskers | **1**/*1* | - | - | - | - | - | - | - |
| 1 speech & 1 static noise maskers | **2**/*15* | - | - | - | - | - | **1**/*1* | - |

where only one masker was used (i.e. 1 speech masker *vs*. 1 modulated masker *vs*. 1 static-noise masker). Of the 11 different target speech types that have been measured in both a speech masker and a noise masker (either modulated noise or static noise), 8 target types gave smaller median slope values for psychometric functions measured in a speech masker than they did in a noise masker [Words, Valid sentences, Invalid sentences, Continuous speech, CRM, HINT, SSI and Other]. There is also an indication that modulated-noise maskers may in turn give shallower slopes than static noise maskers. This trend, however, only holds strongly for 4 of the 11 speech types for which this comparison is available [CRM, HINT, SPIN, and SSI]. Only two speech types gave shallower slopes for the static-noise masker than for the modulated-noise masker [Valid sentences and Other], with the remaining speech types producing very similar slopes for the two masker conditions, within about 0.5% per dB of one another.

Figure 2.2 shows the overall distribution of slope values found for three different masker types: speech, modulated noise and static noise. The slopes of psychometric functions measured for mixed maskers (speech plus either static or modulated noise) are excluded from this particular analysis, and in an attempt to disentangle this effect from the slope effect seen when the number of maskers was increased, only cases where 1 masker was used are included. Unlike tables 2.2A and 2.2B, no distinction is made here between the psychometric functions of the different target types. Figure 2.2 shows there to be a substantial difference between the three slope distributions. While the slope distributions for all three maskers were skewed towards steeper slope values (as was the overall distribution of slopes in Figure 2.1), the measures of central tendency (i.e. median and mean slope values) follow a similar pattern, decreasing in value from static noise (median = 7.5% per dB) to modulated noise (median = 6.1 % per dB) to speech maskers (median = 3.7% per dB). The median slope value of the speech maskers was below that of the overall median slope (median = 6.4% per dB), but this was not the case for the other two masker conditions and suggests that the sample of shallower slopes in the survey were more densely populated by speech maskers.

***Figure 2.2:*** *Histograms showing the distributions of slope values given by three different categories of masker; speech, modulated noise and static noise.*



A Kruskal-Wallis test was carried out to ascertain whether the median slope values for the three masker types differed significantly from one another. These results plus subsequent statistical results for this chapter are summarised in Table 2.3. A significant main effect of masker type was found. Mann-Whitney follow-up tests indicated that speech maskers gave significantly smaller slope values (shallower slopes) than modulated noise maskers but no significant difference was found between the slope values given by modulated-noise maskers and static-noise maskers. These findings confirm that when only one masker was used the *type* of masker had an effect on the slope of the psychometric function, with speech maskers producing shallower slopes than either noise maskers (modulated or static).

### 2.2.2 Major trend 2 - Number of maskers

The second major trend notable in the slope survey data is that the slope of the psychometric function tends to increase as the number of maskers is increased. Table 2.2A shows that increasing the number of speech maskers from 1 to 2 increased the slope by, on average, 4% per dB. Further, three or more maskers produced psychometric functions whose slopes approach those produced by either a modulated noise or static noise masker.

Figure 2.3 shows the distribution of slope values as a function of the number of maskers used. To control for the affect of masker type on slope, only psychometric functions measured using speech maskers were included (few studies have looked at the use of more than one noise masker, so if all masker types were included the "1 masker" condition would be averaged over all masker types but the increased masker conditions would be biased towards speech maskers). It can be seen that the slope distributions shifted upwards to larger values as the number of maskers was increased from 1 to 3 or 4. The mean and median slope values also increased as the number of maskers were increased, but again only in the 1-masker condition was the median slope value (median = 3.6) below that of the overall median slope value. The bottom panel includes psychometric functions measured with 5 to 20 speech maskers. The distribution, mean, and median slope values for this condition were very similar to those produced when 3 or 4 maskers were used. This would suggest

***Figure 2.3**: Histograms showing the distributions of slope values given by 1, 2, 3 to 4, or greater than 5 speech maskers.*

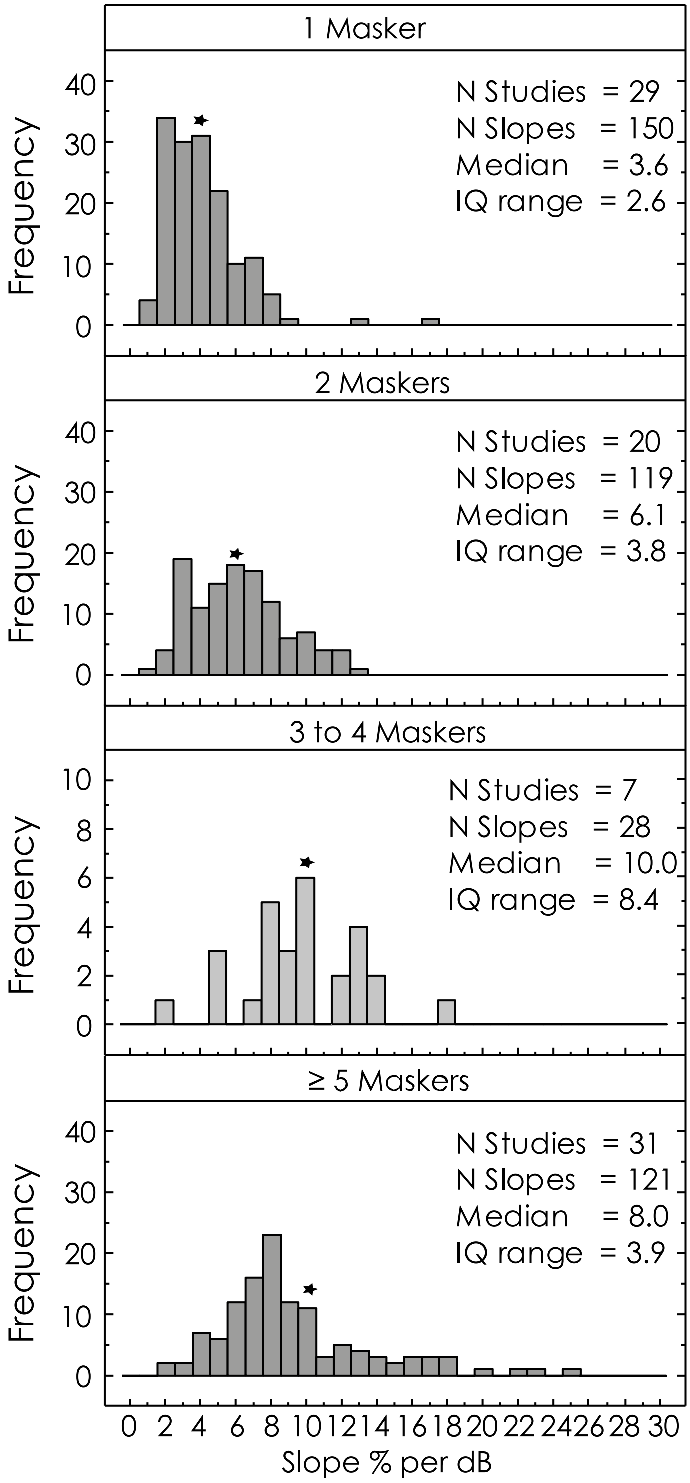


*Table 2.3: summarises the results of the statistical analyses made on the slope data collected in the systematic survey. Highlighted boxes indicate statistically significant results.*

| **Main Effect** | **Results of Kruskal_Wallis test** | **Post-hoc comparisons** | **Results of Mann-Whitney test** |
|---|---|---|---|
| *Type of masker* | H(2) = 120.48, p < 0.001 | Speech vs Modulated noise | U = 3117.0, Z = -7.50, p < 0.001, r = -0.50 |
| | | Modulated vs Static noise | U = 16161, Z = -1.32, p= 0.19, r = -0.06 |
| *Number of maskers* | °H(3) = 168.67, p < 0.001 | 1 vs 2 speech maskers | °U = 4597.0, Z = -6.82, p < 0.001, r = -0.42 |
| | | 2 vs 3/4 speech maskers | °U = 653.0, Z = -5.00, p < 0.001, r = -0.42 |
| | | 3/4 vs 5 speech maskers | °U = 1376.5, Z = -1.50, p = 0.12, r = -0.13 |
| *Target predictability* | | High vs low predictability | °U = 63.0, Z = -5.20, p < 0.001, r = -0.71 |
| | | High vs low predictability | †U = 16.0, Z = -1.86, p = 0.06, r = -0.45 |

°*Speech maskers*
†*Noise maskers*

*Table 2.3 continued.*

| Main Effect | Results of Kruskal_Wallis test | Post-hoc comparisons | Results of Mann-Whitney test |
|---|---|---|---|
| *Target corpus* | °H(3) = 6.8, p = 0.08 | | |
| | †H(3) = 28.30, p < 0.001 | CRM vs HINT target corpus | †U = 69.0, Z = -0.81, p = 0.44, r = -0.15 |
| | | CRM vs IEEE target corpus | †U = 18.0, Z = -3.32, p < 0.001, r = -0.63 |
| | | CRM vs SSI target corpus | †U = 3.0, Z = -3.17, p < 0.001, r = -0.77 |
| | | IEEE vs HINT target corpus | †U = 68.0, Z = -3.28, p < 0.001, r = -0.53 |
| | | SSI vs IEEE target corpus | †U = 1.0, Z = -4.04, p < 0.001, r = -0.78 |
| | | SSI vs HINT target corpus | †U = 19.0, Z = -3.03, p < 0.001, r = -0.58 |
| *Target length* | | Short token vs Sentences | °U = 434.0, Z = -1.86, p = 0.06, r = -0.16 |
| | †H(2) = 63.55, p < 0.001 | Short tokens vs Words | †U = 2534.5, Z = -4.99, p < 0.001, r = -0.34 |
| | | Words vs Sentences | †U = 8373.5, Z = -4.38, p < 0.001, r = -0.25 |
| *Prime of target or masker* | | Prime vs No prime | °U = 2880.5, Z = -3.22, p < 0.01, r = -0.16 |
| | | Prime vs No prime | †U = 2308.5, Z = -1.65, p = 0.10, r = -0.09 |

*Table 2.3 continued.*

| Main Effect | Results of Kruskal_Wallis test | Post-hoc comparisons | Results of Mann-Whitney test |
|---|---|---|---|
| *Meaningfulness of the Masker* | | Meaningful vs Non meaningful masker | °U = 111527, Z = -8.30, p < 0.001, r = -0.41 |
| *Similarity of target and masker voices* | °H(2) = 8.58, p < 0.05 | Same-talker vs Same-gender | °U = 344.5, Z = -2.40, p < 0.05, r = -0.30 |
| | | Same gender vs different gender | °U = 166.0, Z = -0.11, p = 0.92, r = -0.02 |
| *Similarity of target and masker content* | | Same vs Different speech corpus | °U = 1064.0, Z = -4.48, p < 0.001, r = -0.36 |
| *Listener age* | °H(2) = 13.72, p < 0.01 | Children vs Young adults | °U = 6768.5, Z = -3.68, p < 0.001, r = -0.19 |
| | | Young vs Older adults | °U = 5156.5, Z = -0.38, p = 0.70, r = -0.02 |
| | †H(2) = 0.33, p = 0.85 | | |
| *Spatial separation of target and masker* | | Spatially separated vs Colocated | °U = 174.5, Z = -0.97, p = 0.33, r = -0.15 |
| | | Spatially separated vs Colocated | †U = 935.0, Z = -0.54, p = 0.59, r = -0.06 |
| *Masker Intelligibility* | | Intelligible vs Unintelligible masker | °U = 1626.0, Z = -0.17, p = 0.87, r = -0.01 |

that once the number of maskers reached 3 any additional maskers had a negligible effect on the slope.

A Kruskal-Wallis test indicated a significant difference in median slope values for the four masker-number conditions (1 masker, 2 maskers, 3-4 maskers or $\geq$ 5 maskers). Mann-Whitney follow-up tests found significant differences in slope values as the number of speech maskers were increased from 1 to 2 and from 2 to 3/4. No difference was seen, however, as the number of speech maskers were increased further, from 3/4 to 5 plus maskers. These results indicate that the shallowest slopes were seen when one speech masker was used. Slopes became progressively steeper as the number of speech maskers were increased to 3 or 4 but no further steepening of the slope was seen if more speech maskers were added.

The mixed masker conditions found in the systematic survey (i.e. 1 speech plus 1 modulated masker, or 1 speech plus 1 static masker) were excluded from Figures 2.2 and 2.3 and the analysis of the major trends. Slope values for these conditions do, however, appear in Table 2.2A and are noteworthy because, given the two major trends discussed above, the median slopes for these conditions are surprisingly low. More than one masker is present in each case and at least one of those is a noise (modulated or static), factors which, as discussed above, tend to result in a steepening of the slope. It should be noted, however, that these values are based on small sample sizes; usually data has been taken from only 1 or 2 studies, and the median slope calculated from only 3 or 4 psychometric functions. There is a possibility that these values are outliers.

## 2.3 Minor trends in slope

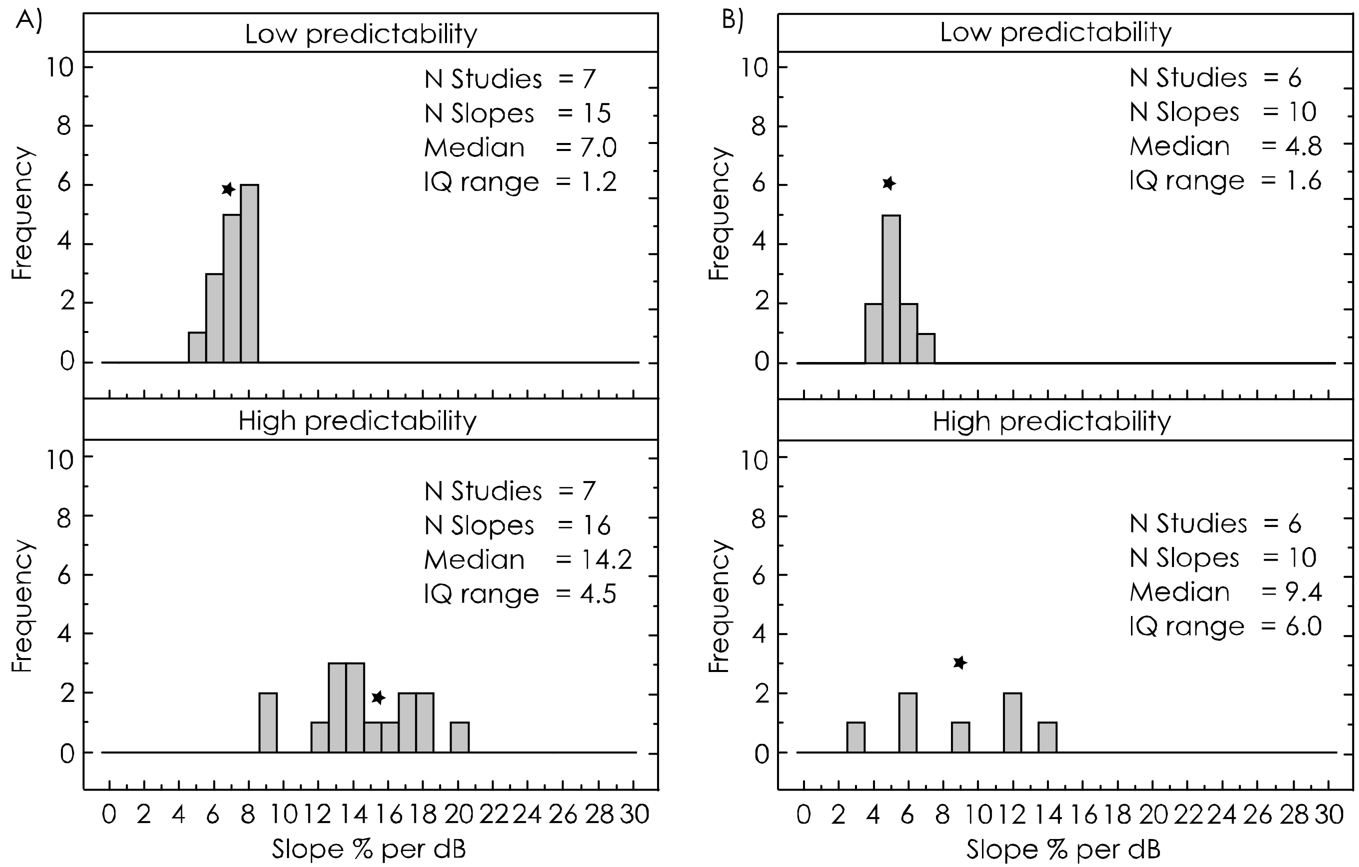
While the type and number of maskers used had a large effect on slope these factors cannot solely account for all the slope variation seen in the survey. For example, there is a range of 17% per dB between the lowest and highest slopes values for cases with one speech-masker. The following sections therefore go on to discuss further analysis of the slope-survey data with the aim of identifying several minor trends that occur within these two major trends.

### 2.3.1 Minor trend 1 – target predictability

One property which was commonly found to be manipulated in the studies included in systematic survey was that of target predictability. The SPIN test, for example, contains sentences which are classed as having high predictability (i.e. strong contextually cues and high frequency target words) and sentences which are classed as having low predictability (i.e. incongruent context cues and low frequency words) (Kalikow et al., 1977). Figure 2.4 compares the slopes of psychometric functions for highly predicable target speech to those for less predictable target speech. Only psychometric functions where the degree of target predictability was specifically stated (usually by the use of the SPIN test) were included in this comparison; i.e. no subjective classification stimulus predictability was made in this analysis. Panel A includes slope values for speech maskers only whereas panel B includes slope values for noise maskers only (due the low number of slopes found for 1 speech masker, the number of speech maskers was not restricted in this case and includes all speech masker cases). For the speech maskers[11] in panel A, a clear effect was found, with less predictable targets producing markedly shallower slopes (median = 7.0% per dB) than the highly predictable targets (median = 14.2% per dB), although both these are well above the overall median value for speech maskers (median = 3.7% per dB). This slope difference was reduced if the masker was noise, however (low predictability median slope = 4.8% per dB, high predictability median slope = 9.4% per dB). Mann-Whitney tests found a significant main effect of target predictability on slope if the masker was speech but not if the masker was noise.

[11] As no suitable non-parametric equivalent to a two-way anova was identified (alternatives such as the Friedman analysis of variance by ranks were for related rather than independent samples), separate one-way Kruskal-Wallis tests were carried out to consider the main effects of each category individually for speech maskers and for static-noise maskers (Siegal & Castellan, 1988). Possible interactions between particular categories (e.g. target predictability) and the type of masker used (speech or noise) could not, therefore, be considered.

***Figure 2.4:*** *Histograms showing the distributions of slope values given by target speech categorised as having either low or high predictability. Panel A compares slope distributions for speech maskers and panel B compares distributions for static-noise maskers.*



### 2.3.2 Minor trend 2 – Target corpus

That target predictability seems to affect slope suggests that the content of the target as well as the content of the maskers plays a role in the shape of the psychometric function. It is possible then that the corpus from which the target speech originates may also affect the slope of the psychometric function. Figure 2.5 compares the distribution of slope values for targets originating from different corpuses. The slopes measured using four standard speech tests (CRM, HINT, IEEE and SSI) are displayed separately for speech maskers (panel A) and static noise maskers (panel B). Only cases where one masker was used have been included; other standardised speech corpuses have been excluded from this analysis due to small sample sizes with just one masker. The figure indicates that when a speech masker is used target corpus seems to have little effect on slope (median slopes = 3.7, 3.4, 4.5 and 4.6% per dB for CRM, HINT, IEEE and SSI respectively). Steeper slopes and a greater difference in slope distributions were seen, however, when the masker was a static noise (median slopes = 10.1, 9.1, 4.8, and 17.1% per dB). Kruskal-Wallis tests found no significant difference in median slope values across target corpuses conditions for speech maskers but did find a significant difference in mean slope values if the masker was a static noise. Post-hoc tests revealed that while the distributions of slope values for the CRM and HINT corpuses did not significantly differ from one another, median slopes for all other comparison of target corpus were significantly different (see table 2.3).

### 2.3.3 Minor trend 3 – target length

It is possible that the length of the target utterance also has an effect on slope. Figure 2.6 compares the distribution of slope values for target tokens of differing lengths. To calculate these, all target speech tokens, regardless of originating corpus, were re-classed into four sub-categories; Short tokens, Words, Sentences, or Continuous speech. Panel A includes slope values produced when one speech masker was used whereas panel B includes the slope values produced when one static-noise masker was used. The median slope values for the Short tokens, Words, and Continuous target lengths were relatively similar although Sentences gave shallower slopes when a speech masker was used (median = 5.9, 6.7, 3.5 and 5.7% per dB for the Short

***Figure 2.5:*** *Histograms showing the distributions of slope values given by four different target corpuses: CRM, HINT, IEEE and SSI. Panel A compares slope distributions for speech maskers and panel B compares distributions for static-noise maskers.*

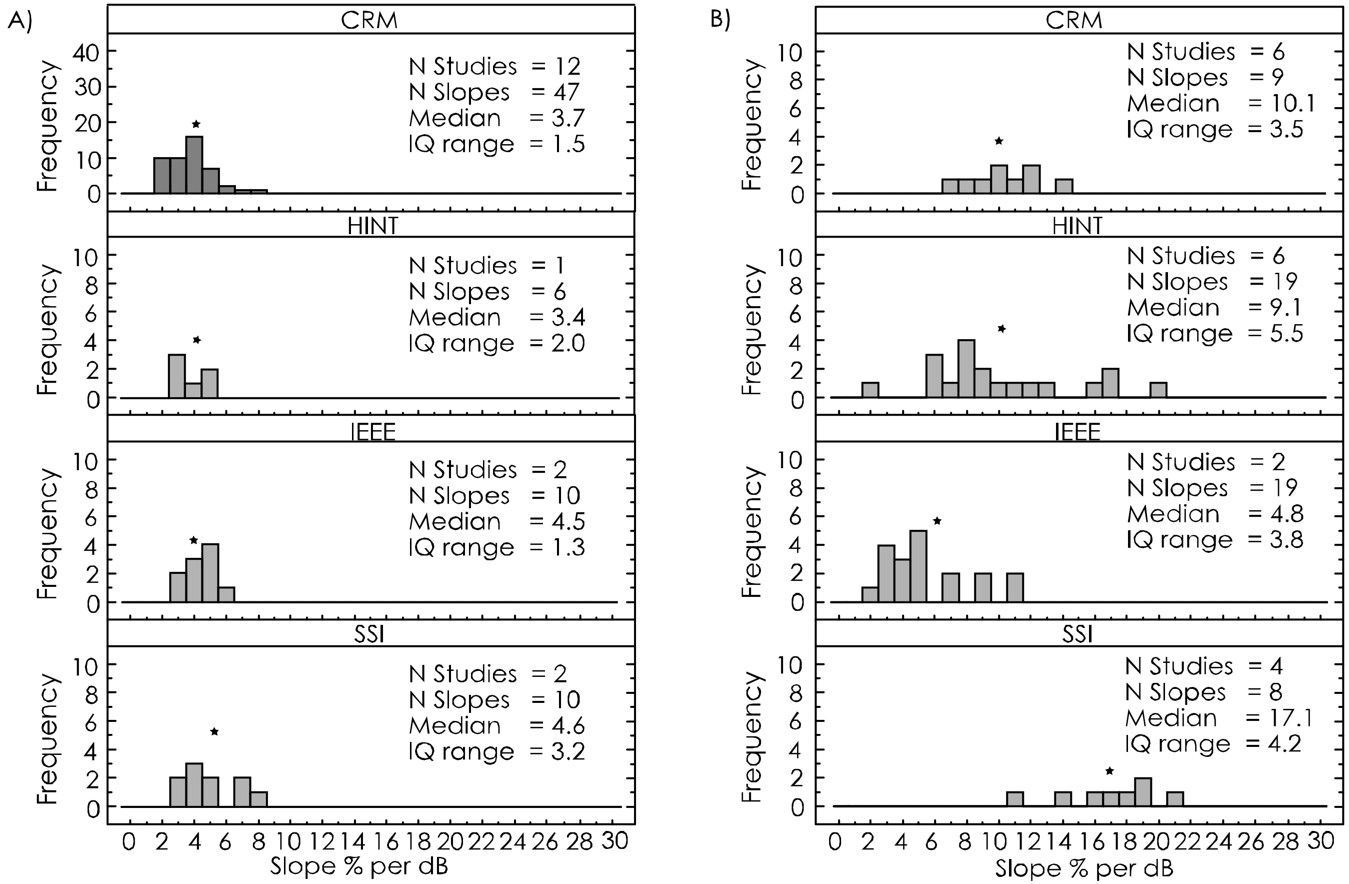
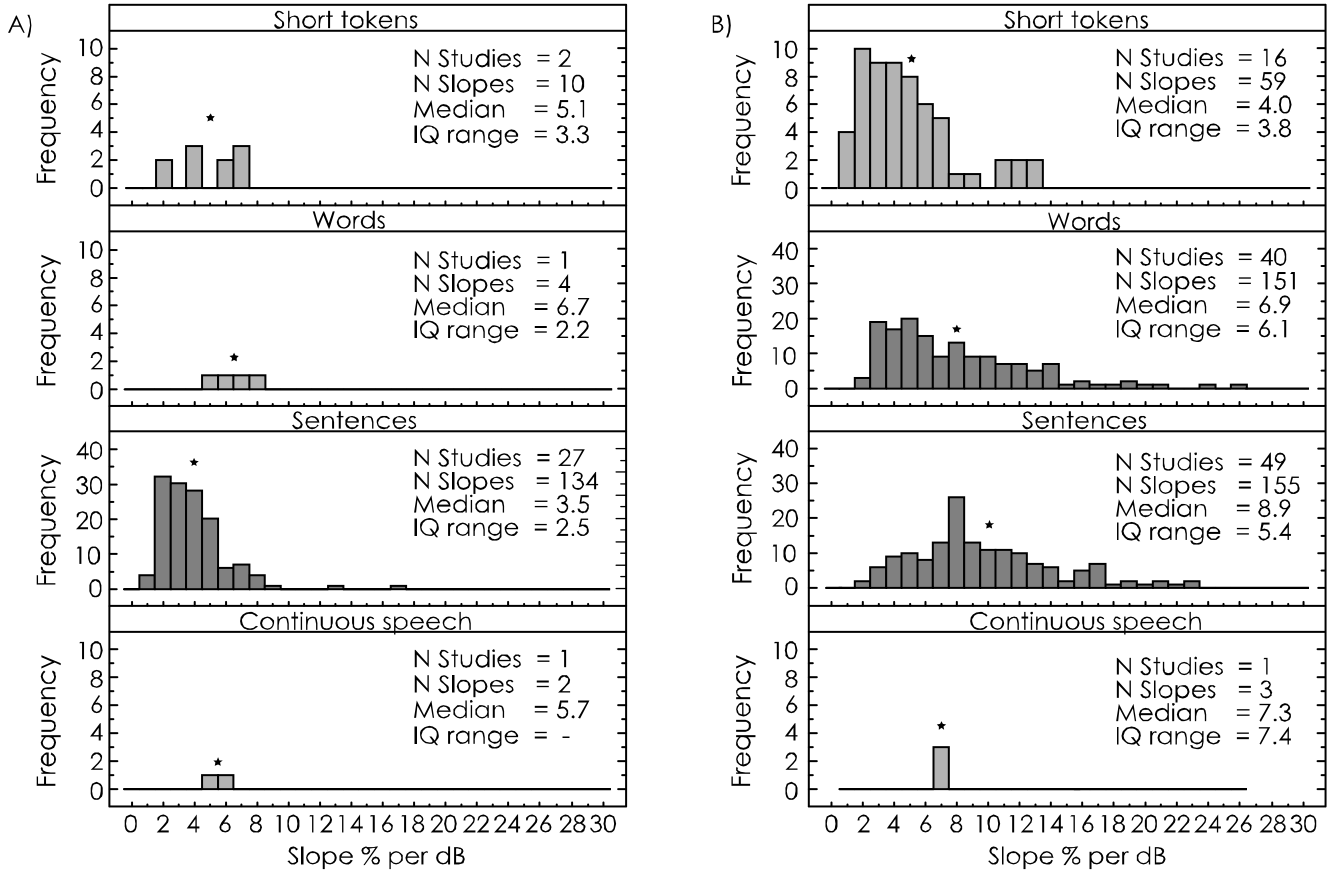


***Figure 2.6:*** *Histograms showing the distributions of slope values given by four different target lengths: Short tokens, Words, sentences and Continuous speech. Panel A compares slope distributions for speech maskers and panel B compares distributions for static-noise maskers.*



tokens, Words, Sentences and Continuous sub-categories respectively). There was an indication, however, that for noise maskers, Sentences resulted in *steeper* slopes than Words or Short tokens (median slope of 9.1% per dB for Sentences compared to 6.8 and 4.0% per dB for Words and Short tokens respectively). A Mann-Whitney test was carried out to establish whether slopes of the different target lengths differed significantly for the speech masker. Words and Continuous conditions were excluded from this analysis as very few psychometric functions were found for these sub-categories (for Words n = 4, for Continuous n = 2). No significant difference was found between the slopes of psychometric functions measured with Short tokens compared to those measured with Sentences when a speech masker was used.

A Kruskal-Wallis test was carried out to consider the effect of target length on slope when a static-noise masker was used (again the Continuous subcategory was excluded from this analysis, n = 3). A significant effect of target length was found. Mann-Whitney post hoc tests were carried out to see if slope values increased as the length of the target speech token was increased for the static noise masker. The results showed that slopes were shallower for short tokens than for word targets and slopes for word targets were in turn shallower than sentence targets, though inspection of the r values indicates that both of these effects were rather small. These results suggest that while target length had an effect on the slope of the psychometric function with a noise masker it did not have an effect on slope when the masker was speech.

### 2.3.4 Minor trend 4 – Prime of target or masker speech

In several studies identified in the survey the target or the masker sentence was primed in some way before presentation. Helfer and Freyman (2009), for example, primed either the target voice or content before presentation and Freyman et al., (2004) primed both target content and talker. To look at the effect of priming on the slope of the psychometric function, Figure 2.7 compares slope distributions given when a prime was present to those given when no prime was present. These distributions are displayed separately for speech (panel A) and static noise maskers (panel B). Primed cases included: 1) acoustic primes, where target or masker voices were primed, 2) linguistic primes, where the content of the target or masker were primed, and 3) dual primes, where both the acoustic and content of the target or masker were primed (e.g. the prime was the start of the test sentence). As the vast majority of the psychometric functions found in the slope survey were measured without using primes, the slope distributions given for the *no prime* conditions are more complete and as such more strongly resemble the overall distribution seen in section 2.2, i.e. skewed with a long tail at the higher slope values. The *prime* distribution is less complete but there does look to be a shift in central tendency to the right and to higher slope values compared to the *no prime* cases (medians of 7.5 compared to 5.9% per dB for speech maskers and 8.3 to 7.3% per dB for the static noise maskers). Mann-Whitney tests were carried out to assess the effect of priming on slope. A small but significant effect of a *prime* was found for speech maskers but not for noise maskers. These findings suggest that − at least when maskers are competing speech − priming speech stimuli can have the effect of steepening the slope of the psychometric function.

### 2.3.5 Minor trend 5 – Meaningfulness of the masker.

Figure 2.8 compares the distribution of slopes values produced for meaningful speech maskers and non-meaningful speech maskers. Non-meaningful speech maskers include unintelligible speech (which in turn includes time-reversed speech and foreign language speech) and speech void of meaningful linguistic information (e.g. invalid sentences, or babble maskers with 5 or more talkers). The non-meaningful maskers (median = 7.3% per dB) were found to give steeper slopes than the meaningful maskers (median = 4.0% per dB). A Mann-Whitney test was carried out and a significant difference between the two conditions was found. This result would suggest that speech maskers containing meaningful information produce shallower slopes than non-meaningful maskers. A caveat to this finding, however, is that a large number of the "non-meaningful" cases were babble maskers. The number of maskers used and the resultant changes to modulations in the masker have not, therefore, been accounted for here.

***Figure 2.7:*** *Histograms showing the distributions of slope values given with and without a prime. Panel A compares slope distributions for speech maskers and panel B compares distributions for static-noise maskers.*

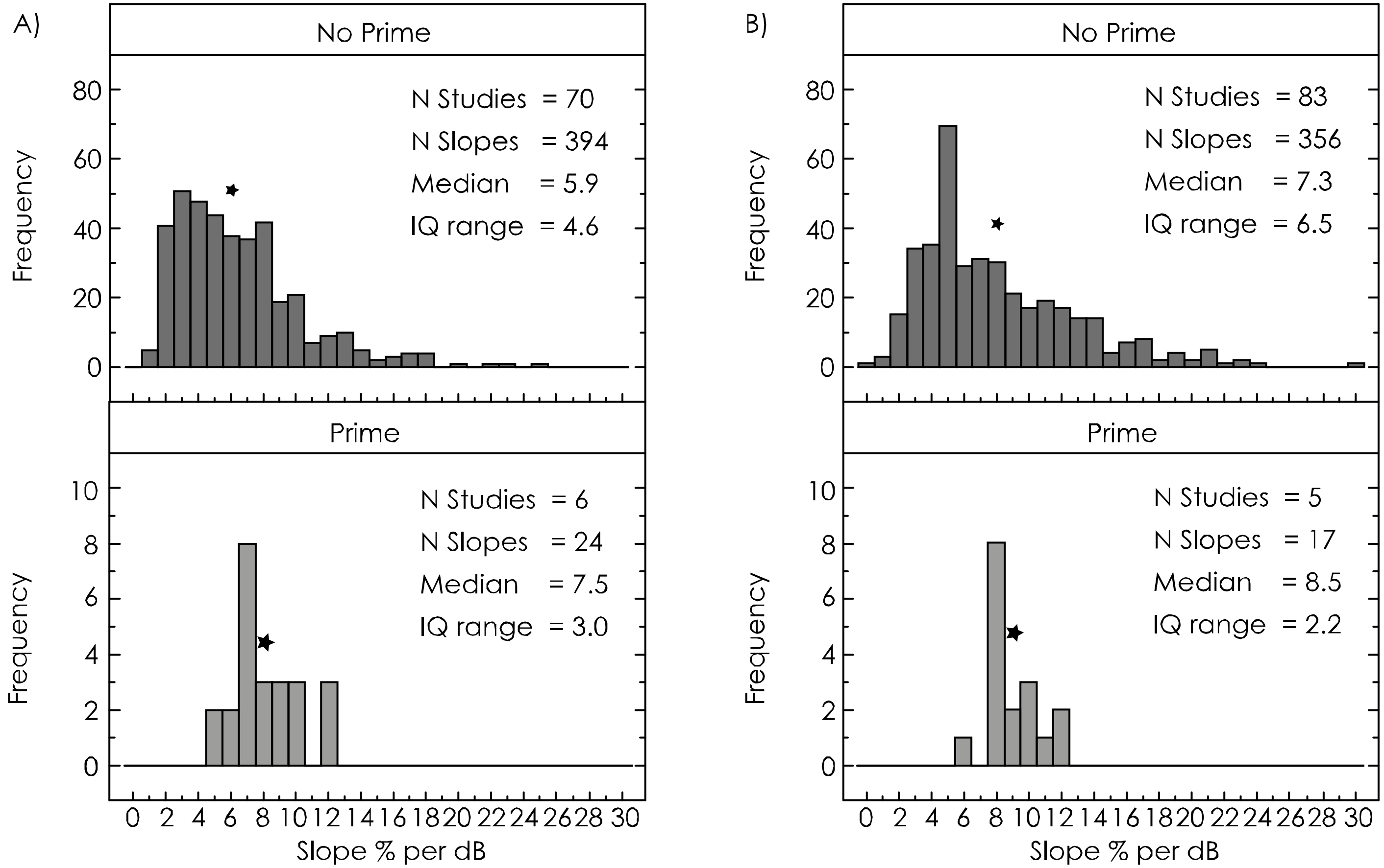
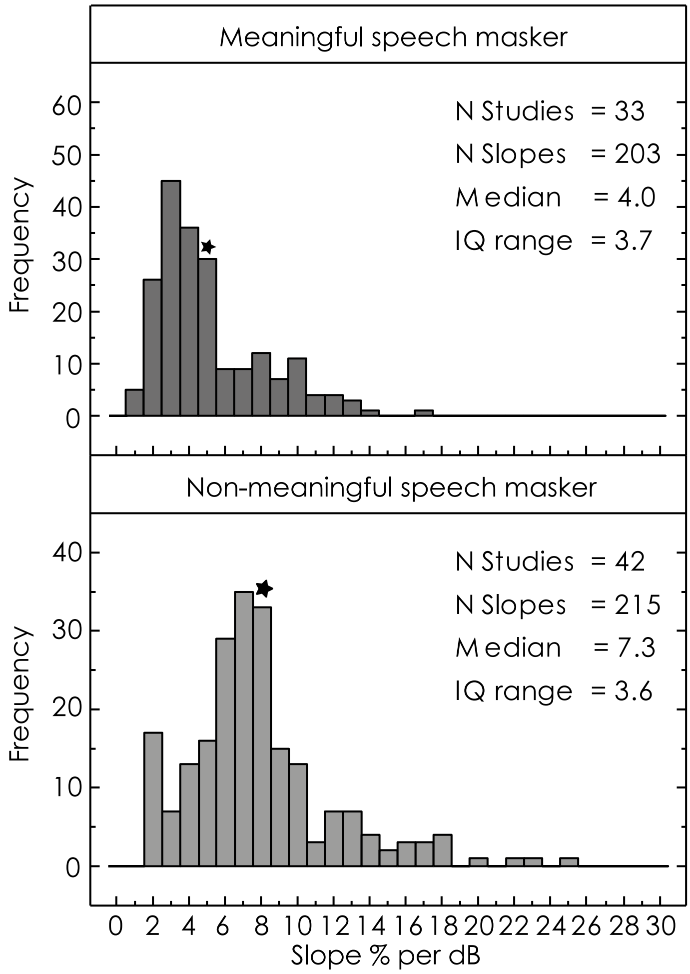


***Figure 2.8:*** *Histograms showing the distributions of slope values given by meaningful and non-meaningful speech maskers.*



### 2.3.6 Minor trend 6 – Similarity of target and masker voices

Figure 2.9 compares the distribution of slope values for varying degrees of target/masker voice similarity. The sub-categories of similarity include the target and masker spoken by: 1) the same person, 2) by a person of the same gender, 3) by a person of different gender. The sub-categories include all cases found in the survey where only one speech masker was used. All cases where the masker or the target speech had been manipulated or processed in some way, e.g. the fundamental frequency was shifted or the speech was vocoded, have been excluded as these would affect the similarity between the target and masker voices. Slope distributions for the same-talker maskers and different-gender maskers are similar with the different-gender maskers distribution shifted to the right, indicating steeper slopes (median = 2.6 and 4.5% per dB respectively). The same-gender maskers shows more variability in slope values although the median value (median = 4.7% per dB) is close to that of the different-gender maskers. Only the same-talker maskers give a median slope value below the overall median slope value for speech maskers, outline in section 2.2.

A Kruskal-Wallis test found a significant difference in slope distribution between the three sub-categories. Post-hoc tests revealed that same-talker maskers gave significantly shallower median slopes than same-gender maskers, however further distinctions in target and masker voices, i.e. moving from same gender to a different gender voice, did not have a significant effect on slope. These findings would suggest that only extreme similarity between target and masker voices, i.e. spoken by the same person, had a marked effect on slope over and above that seen due to the masker being speech rather than a static noise. This is discussed further in section 2.4.2, where the causes for unusual shaped psychometric functions are considered.

### 2.3.7 Minor trend 7 – similarity of target and masker content

It is possible that a high degree of similarity in the linguistic content − i.e. in the words used – of target and masker speech may have an effect on slope. Figure 2.10 shows the slope distribution when targets and maskers were taken from the same speech corpus and compares it to the distribution when the target and masker were

***Figure 2.9:*** *Histograms showing the distributions of psychometric function slope values given by speech masker with three different levels of talker similarity to the target speech; Same talker, Same gender talker and Different gender talker.*

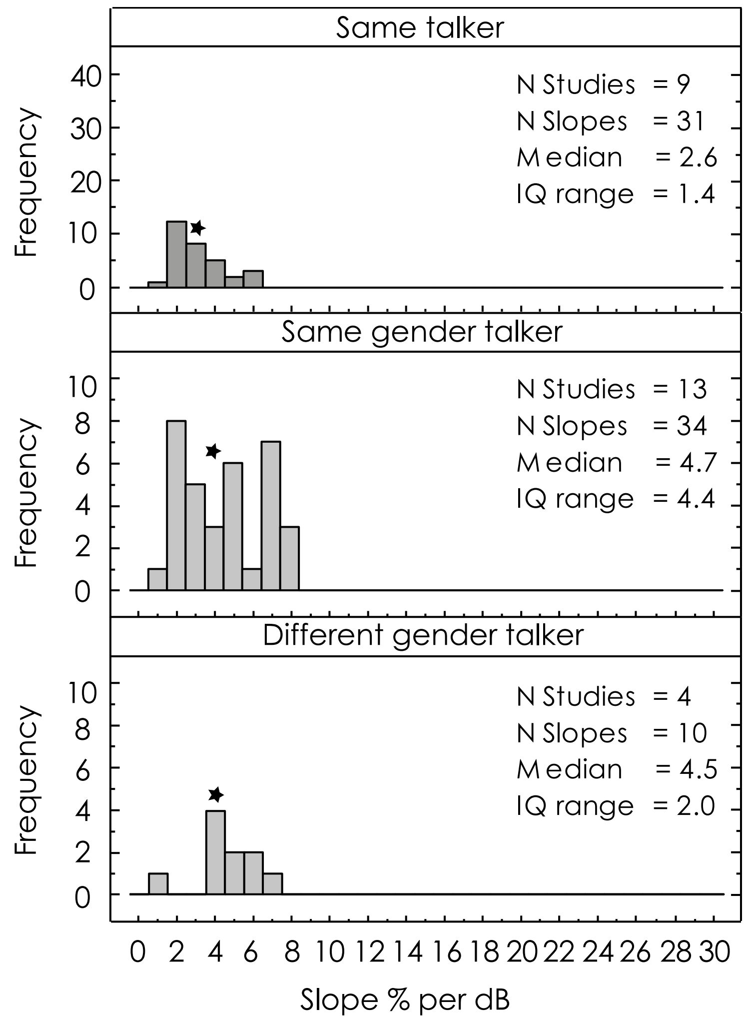
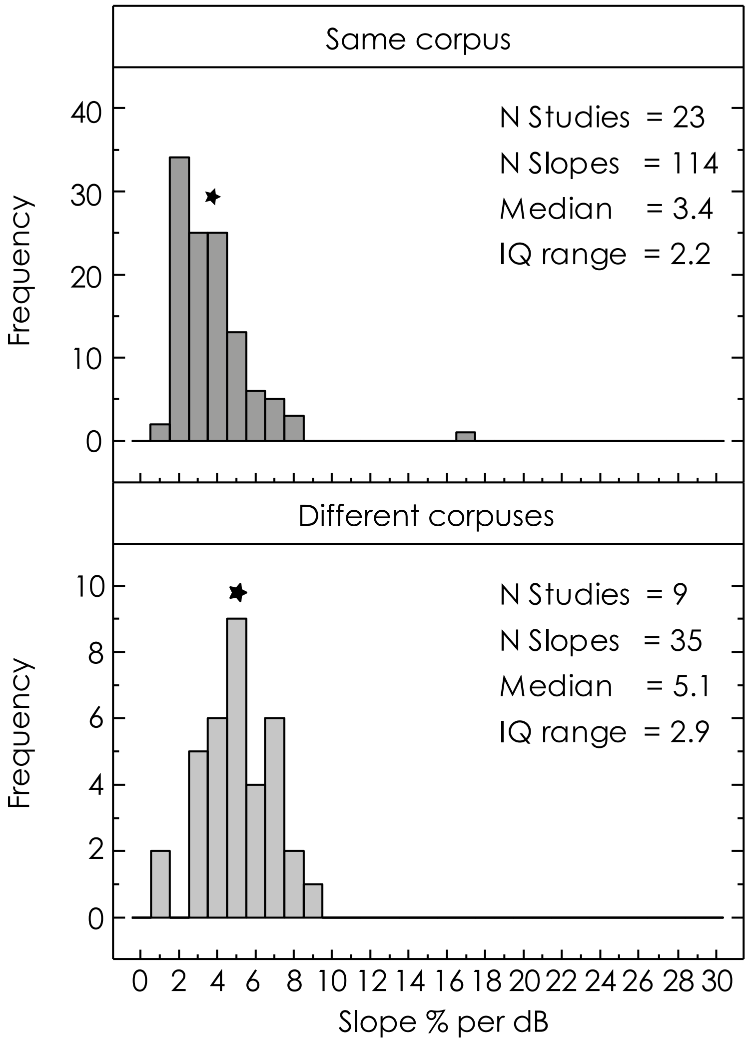


***Figure 2.10:*** *Histograms showing the distributions of slope values given when target and maskers speech stimuli were taken from the same (top panel) and different (bottom panel) corpuses.*

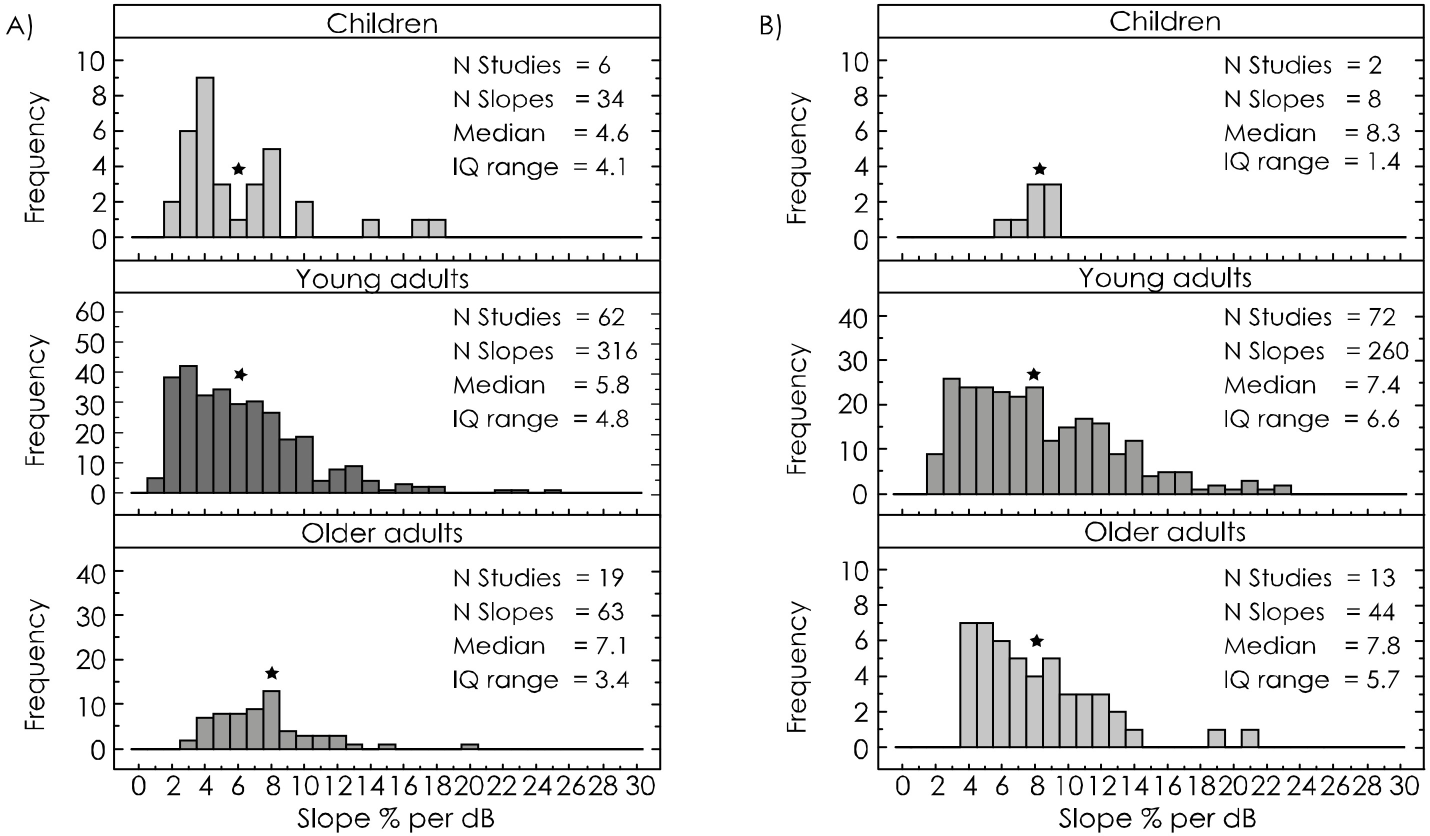


taken from different speech corpuses. The distribution is shifted to the right when different speech corpuses were used, and median slope values were higher for this condition (median = 5.0% per dB) than for the same corpus condition (median = 3.0% per dB). A Mann-Whitney test was carried out and a significant effect of corpus similarity was found. This result indicates that slopes were shallower when target and masker speech were taken from the same corpus (again see section 2.4.2 on unusual shaped psychometric functions).

### 2.3.8 Minor trend 8 – Listener Age

It is also of interest how factors such as listener age might affect the slope of the psychometric function. Figure 2.11 shows the distribution of slope values for three listener age groups: children (0 to 17 years), younger adults (between 18 to 49 years) and older adults (≥ 50 years). As in earlier figures, these have been further categorised by the type of masker used; speech (panel A) or static noise (panel B). Due to a small sample sizes for some of the age groups, the cases were not restricted to only one masker for this analysis. The figure indicates that age had little effect on slope values when a static-noise masker was used (median = 8.3, 7.4, and 7.8% per dB, for children, young adults, and older adults respectively). Conversely there is a trend of increasing slope values with age when a speech masker is used (median = 4.6, 5.8, and 7.1% per dB or children, young adults, and older adults respectively). Kruskal-Wallis tests support this, as a significant effect of age was found within the speech-maskers cases, but not within the noise-masker cases. Post-hoc tests suggested that this significant difference for the speech maskers lay between the slopes produced by young and older adults. The difference between slopes produced by children and young adults was not found to be significant. These findings suggest that the slope of the psychometric function may become steeper with age, although the effect only held for speech maskers and was rather small. The sample sizes for this comparison were also particularly unequal, as most studies used young normal-hearing listeners. Slope differences should, therefore, be treated tentatively here.

***Figure 2.11:*** *Histograms showing the distributions of psychometric function slope values given by listeners from three different age groups: Children, Young adults and Older adults. Panel A compares slope distributions for speech maskers and panel B compares distributions for static-noise maskers.*



## 2.4 Additional findings

### 2.4.1 Notable cases which did not affect slope

There were a few conditions which, given the mechanisms identified in chapter 1, were expected to affect slope but which were not found to do so in the slope survey. Figure 2.12 compares the slope distributions of psychometric functions measured when the target and the masker were spatially separated to those measured when the target and masker were co-located. These sub-categories include cases where targets and maskers were also only *perceived* to be in these spatial configurations, for instance via a perceptual shift induced by the precedence effect (Freyman et al., 1999). Again, only slopes measured using one masker were included, and the distributions for speech and static noise maskers are displayed separately. The figure indicates that the distributions of slopes are similar for the two spatial configurations in both masker conditions. Mann-Whitney follow-up tests support this as no significant differences in the slope distributions were found when either a speech or a static-noise masker was used.

In section 2.3.5 the effect of the meaningfulness of the masker on the slope of the psychometric function was considered. Non-meaningful maskers included maskers which were considered to be void of *meaningful* information, but they could still contain some intelligible linguistic information. For example, babble and invalid speech were classed as non-meaningful as semantic meaning could not be extracted from these maskers, however, individual words may still have been intelligible. To look at the effect that intelligibility had on slope, maskers which contained *any* intelligible speech[12], regardless of whether it carried semantic meaning or not were categorised as intelligible. Maskers where no intelligible linguistic units could be extracted, e.g. reversed speech, were conversely categorised as unintelligible. Figure 2.13 shows the distributions of slopes given for one unintelligible speech masker compared to those given for one intelligible masker. Fewer cases in the unintelligible subcategory mean the distribution is less complete for this condition but median

[12] Multi-talker babble was excluded from this analysis as it made sample sizes extremely uneven (386 intelligible cases to just 36 unintelligible cases).

***Figure 2.12:*** *Histograms showing the distributions of slope values given when the target and masker are co-located compared to when they are spatially separated. Panel A compares slope distributions for speech maskers and panel B compares distributions for static-noise maskers.*

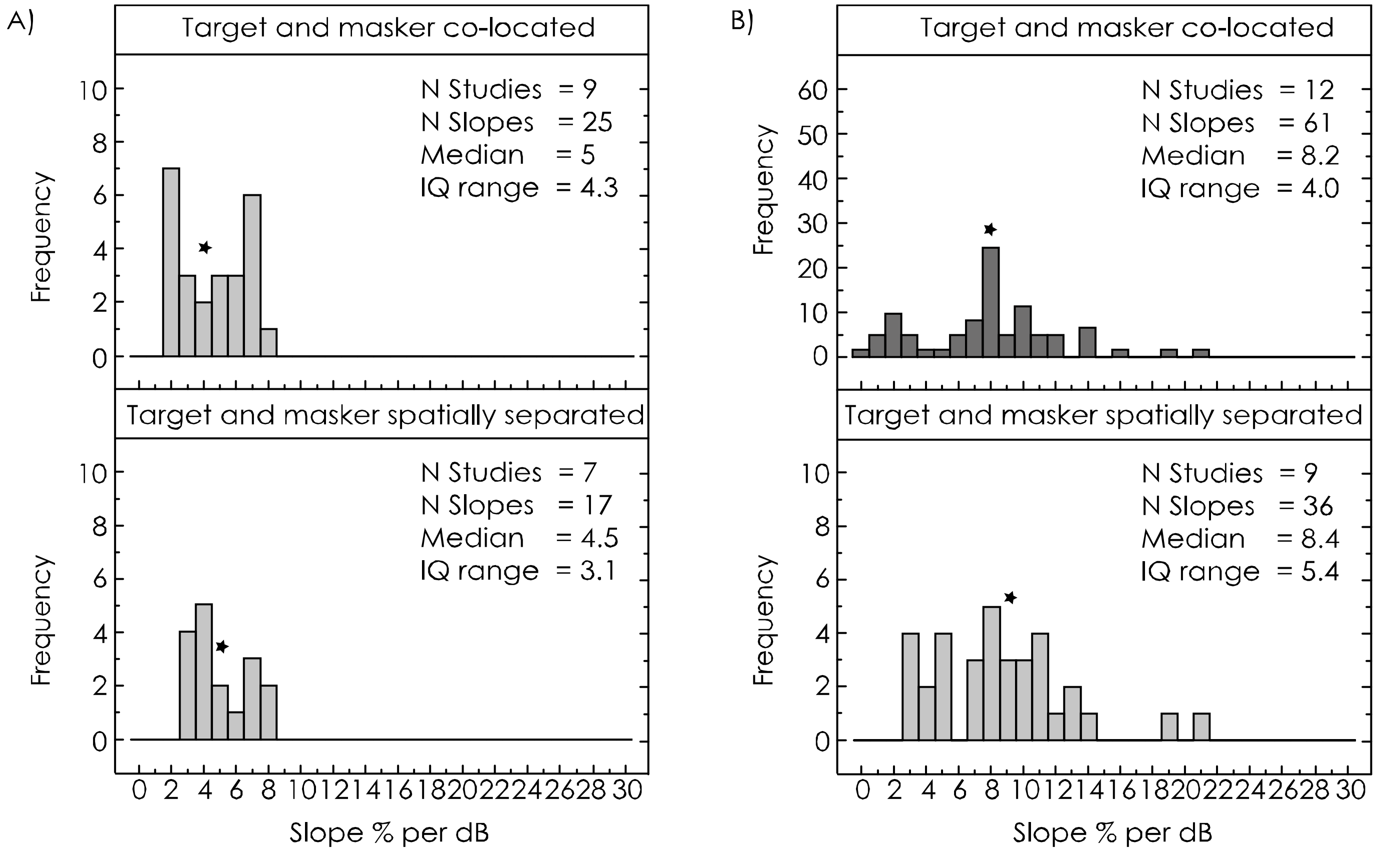
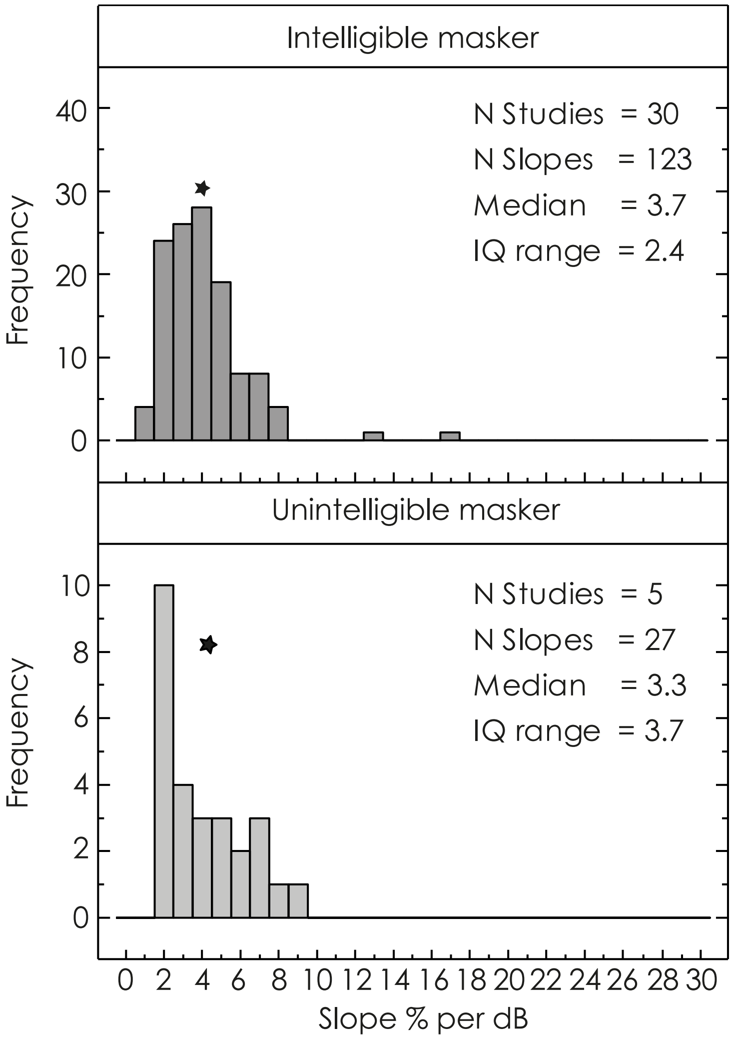


***Figure 2.13:*** *Histograms showing the distributions of slope values given when the masker consists of intelligible speech (top panel) and when it consists of unintelligible speech (bottom panel).*



values are similar across both sub-categories. A Mann-Whitney test confirmed that there was no significant difference between the two slope distributions. This result suggests the intelligibility of a masker did not have an effect on the slope of the psychometric function.

There were also several factors which may affect the slope of the psychometric function but which were unable to be explored due to a shortage of relevant cases identified in this survey. For instance, the effect of shifting the pitch of the target or the masker was one example of a factor that previous research suggests should have an effect on slope, but only 13 cases in all were found in the slope survey. A second example was the effect that the depth and rate of amplitude fluctuations might have on slope, but again too few studies provided this information so a comparison across variations in these conditions could not reliably be made. Target presentation level was also adjusted in a handful of studies but insufficient cases made it difficult to look for any trends in slope. Individual studies have, however, reported that adjusting presentation level did not have any material effect on slope (Wagener & Brand, 2005). Lastly, as outlined in section 1.3.5 in Chapter 1, the effects that age and hearing impairment may have on the slope of the psychometric function are not yet fully understood. Unfortunately, the effect of hearing impairment could not be accurately assessed in this slope survey as the majority of studies identified used young normal-hearing listeners and the hearing-impaired listeners were also usually older, so making any differences between the groups hard to discern.

### 2.4.2 Unusually shaped psychometric functions

As explained in section 2.1.2, any cases where the data had to be extrapolated to fit a logistic function or cases where the logistic functions were a poor fit to the data were excluded from the slope survey. As the data was always fitted with a monotonic curve, the fitted functions for the extremely shallow or unusual shaped psychometric functions outlined in section 1.3.2 are likely then to have been very poor fits and would have been excluded from the survey. As these cases were the most extreme examples of shallow slopes in the literature the listening conditions which give these functions should not be overlooked. Twenty four cases were excluded from the survey due to high RMS values but which also were non-monotonic in shape (e.g.

plateaus or U-shaped). Figure 2.14 shows a series of pie charts which display the proportion of this sub-set of psychometric functions which are given by different listening conditions. Panel A shows that the majority of functions in this sub-set are given by speech maskers rather than static noise maskers, and panel B shows that for the majority of these slopes only one masker was used. Panel C indicates that while these unusual shaped psychometric functions were measured using several different speech stimuli the largest portion of slopes resulted from using the CRM stimuli. Panel D demonstrates that 80% of slopes were given when the target and masker were taken from the same corpus. Finally panel E shows that the vast majority of the slopes in this group were given when the same talker was used in the target and the masker, be it natural, processed or mixed with other maskers. It seems then that the listening conditions giving the shallowest slopes fit with the trends reported above for shallow slopes identified in the main slope survey. The conditions sufficient and required for the occurrence of unusual shaped psychometric functions are returned to in greater detail in Chapters 4 and 5.

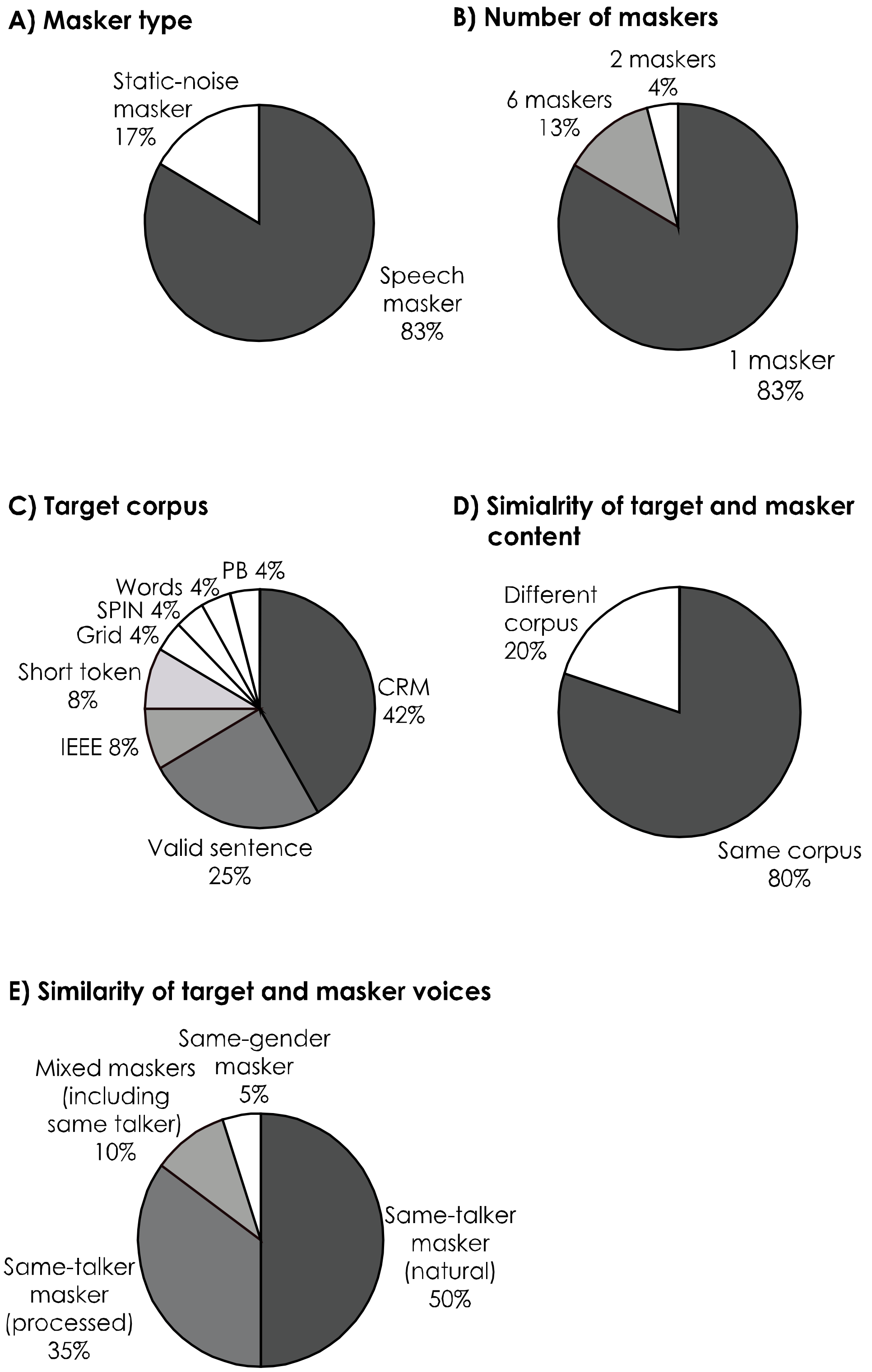
## 2.5 Discussion

The systematic slope survey used data already available in the literature to quantify the variation seen in slopes and identify slope-change trends. Large variations in slope were found with slopes ranging from as shallow as 1% per dB to as steep as 29% per dB. Differences in slope of 2-3% per dB or more between sub-categories suggested that several categories, such as masker type and number, were having an effect on slope. In the following sections these slope trends are discussed in terms of the four slope-change mechanisms outlined in Chapter 1.

### 2.5.1 Evidence for slope changes as a consequence of fluctuating maskers

One of the main trends that emerged from the slope survey was that masker type affected the slope of the psychometric function, with speech maskers found to give shallower slopes than noise maskers, be they modulated or static noise. As outlined

***Figure 2.14:*** *Pie charts showing the proportion of unusual shaped psychometric functions given by different listening situations. N = 24 psychometric functions.*



in Chapter 1, shallow slopes have in the past been accounted for by the presence of amplitude fluctuations in the masker. This trend provides partial support for this theory as amplitude modulations are vital for speech perception (e.g. Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995) and speech can be thought of as, in essence, the sum of multiple amplitude modulated frequency bands (Drullman, Festen, & Plomp, 1994). That amplitude-modulated noise maskers were not also found to give significantly shallower slopes than static-noise maskers was, however, surprising. It is plausible that not all modulated maskers are sufficient to flatten the slope of the psychometric function. Howard-Jones and Rosen (1993), for example, demonstrated that only fluctuations with relatively long durations (greater than 200ms) resulted in shallow psychometric functions. In the slope survey the modulated-masker category contained any noise masker whose amplitude was temporally varied. Variations in the depth, frequency and duration of these modulations were rarely reported in individual studies and were not recorded in the slope survey. As a result the modulated noise category was very broad; maskers ranged from interrupted white noise to modulated speech-shaped noise. It is possible then that the more subtle effects that amplitude modulation may have on slope were lost in such a varied category.

A second major trend that emerged from the slope survey was that the number of speech maskers used also affected the slope of the psychometric function: as the number of voices were increased from 1 to 4 the slope of the function increased. This finding was again in support of the suggestion that amplitude fluctuations in the masker result in slope changes. The quality of the amplitude variations present in the masker will change greatly as the number of speech maskers is increased. When a single competing talker is used temporal fluctuations are relatively slow and there are likely to be many opportunities where the target speech will coincide with a dip in the amplitude of the masker, i.e. there will be many opportunities for glimpsing the target speech (Miller & Licklider, 1950 etc.). As more maskers are added spectral and temporal dips begin to fill in (Cooke, 2006; Miller, 1947). The chance that the target will temporally overlap with at least one of the maskers becomes greater and overall amplitude modulations in the masking mixture effectively become shallower and more brief. The opportunities for glimpsing the target, therefore, become fewer.

In terms of slope, the reduction in dynamic range due to the filling-in of temporal dips would steepen the slope of the psychometric function. Theoretically, if enough voices were added to the masking signal it would approach that of a speech-shaped static noise. Cooke (2006) noted that when six or more masking voices were present, intelligibility was not significantly different from that of a speech-shaped static noise masker. The slope survey demonstrates that even fewer masking voices were needed before the slopes of psychometric functions became equivalent to those given by a static noise; median slopes given by only three or more speech maskers were of the order of those given by static noises (10% and 7.5% per dB respectively).

### 2.5.2 Evidence for slope changes as a consequence of target/masker confusion

The trend of shallower slopes for speech than for noise maskers also provides support for a second slope-change mechanism. As well as containing amplitude modulations, speech maskers are also more acoustically similar to speech targets than noise maskers are. The similarity, and the resultant confusion, between a target and a masker may, therefore, directly result in the occurrence of shallow slopes. Manipulating the acoustic similarity of the target to the speech masker also affected the slope of the psychometric function further supporting this theory: shallower slopes were found when target and masker voices were spoken by the same person than when they were either spoken by someone of the same gender or by someone of a different gender (Brungart, 2001a; Brungart & Simpson, 2007).

An analysis of the unusual shaped psychometric functions identified in the literature also demonstrated that a high degree of similarity between the target and the maskers gave extremely shallow slopes. Most of these functions (17/24) were produced when a speech masker was used and nearly all of those functions (23/24) were given when at least one of the speech maskers was spoken by the same voice as the target. This included cases where the same-speech maskers had been processed to give small shifts in fundamental frequency. This would give an indication then that slope was not affected by small changes in F0. There were, however, too few cases to consider the effect that further increasing F0 differences and, therefore, target/masker voice similarity, may have on slope.

Spatially separating a target from a masker has been found to improve speech intelligibility (e.g. Dirks & Wilson, 1969b; Freyman et al., 1999; Hirsh, 1950; Plomp & Mimpen, 1981). It has been argued that this improvement is the result of a decrease in confusion between a target and masker making it easier to select and attend to a desired speech source (Freyman et al., 1999). The slope survey did not, however, find a significant effect of spatial separation on the slope of the psychometric function. This finding was initially surprising as some studies have shown an effect of spatial separation on the slope of the psychometric function (Freyman et al., 2001; Helfer & Freyman, 2005; Ihlefeld & Shinn-Cunningham, 2008a). In each of these studies a plateau or U-shaped function was given when the target and masker speech were co-located but spatially separating the two signals returned the function to a monotonic curve. Several other studies, however, have noted very little difference in the slope of functions measured for co-located and spatially separated stimuli (Freyman, Helfer, & Balakrishnan, 2007; Helfer & Freyman, 2009; Li et al., 2004; Wu et al., 2005). In each of these cases functions in the co-located condition were already monotonic and spatially separating the target from the masker had no effect on slope. These results suggest that spatial separation only results in slope changes if the target and masker stimuli are already highly similar and that spatial separation alone has little effect on slope. This may explain why no overall effect of spatial separation was found in the slope survey.

As an aside, it should be mentioned that many of the studies that did not find a slope change when the target and maskers were spatially separated still demonstrated a release from masking (i.e. a decrease in threshold). This threshold improvement was attributed to a reduction of informational masking due to spatial separation. We argue that this provides evidence that not all "types" of informational masking flatten the slope of the psychometric function and again emphasises the need to separate these two concepts.

The trend for shallower slopes when a prime was provided in speech-in-speech cases further supports the theory that difficulties differentiating between targets and maskers, and the confusion that this leads to, can flatten the slope of the psychometric function. No change in slope was seen if a static noise rather than a speech masker was used, suggesting that the prime affected slope in speech-in-speech cases by helping to reduce similarity between the target and the masker. It has been proposed that for a desired source to be successfully selected from mixture of sounds, a prior knowledge of its distinguishing features is needed (Shinn-Cunningham & Best, 2008). Knowledge of these features can be used to direct top-down attention and reduce confusion between the target and the masker. A prime can, therefore, act as clue to where to direct selective attention. That providing a prime can steepen the slope of the psychometric function gives an indication that target/masker confusion flattens the slope due to failures selecting the target rather than failures segregating the target from the masker. This role of improving selective attention on slope will be further considered in the experimental work of Chapter 5.

Overall, the slope survey has shown that introducing any additional cue which can then be used to differentiate the target from the masker seems to change the slope of the psychometric function when the target and masker are highly similar. This lends support to the theory that confusion causes shallow slopes due to an increased reliance on the level difference between a target and a masker. Presumably, providing an additional cue aids in the selection of the target speech meaning a difference in level is no longer the sole cue. Once stronger cues are available the plateau in performance, which occurs as SNR reaches 0 dB, is eradicated. An additional finding of the slope survey, however, was that target voice and spatial separation per se have no *further* effect on slope once this basic differentiation is achieved. It could be argued, therefore, that acoustic similarity between targets and maskers has an "all or none" effect on slope.

It has been suggested that as well as acoustic similarity, linguistic similarity between a target and a masker might also result in shallow psychometric functions. The slope survey provided support for this theory, finding that slopes tended to be shallower when target and masker speech were taken from the same corpus. Speech stimuli from the same corpus are likely to be either syntactically or semantically similar. In the case of some corpuses, such as the CRM corpus, target and masker stimuli will be both syntactically and semantically very similar. That shallower slopes were found when targets and maskers originated from the same corpus supports the theory that similar linguistic features in the target and the masker affected the slope of the psychometric function.

Reducing linguistic similarity between the target and the masker by removing meaningful linguistic information from the masker was also found to affect psychometric functions in the slope survey. This finding again provides support for a confusion based, slope-change mechanism. It could be argued then that meaningful information in the masker diverts attention and/or cognitive resources away from the target making it harder to select and attend to and ultimately causing confusion between it and the masker. If, however, conditions were categorised as intelligible (including semantically meaningless but linguistically intelligible maskers such as invalid sentences) vs. unintelligible no significant effect of intelligibility on slope was found. This was surprising, as it could be reasoned that if removing meaning (but not intelligibility) could reduce target/masker confusion enough to steepen the slope, removing intelligibility completely should have the same, if not a greater, effect on slope. One possibility is that the slope difference seen when meaning was removed was confounded by the number of maskers used. Multi-talker babble was, for example, included as a meaningless masker as the increased number of talkers makes the probability of extracting meaningful linguistic information unlikely (Hoen et al., 2007). It was not, however, included in either subcategory of intelligibility. As noted above, increasing the number of talkers also increased the slope of the psychometric function. The greater number of maskers in the non-meaningful category may, therefore, have confounded the result. This confound was removed when intelligibility was considered and the slope effect disappeared. The slope survey therefore gives mixed findings as to the effect that linguistic similarity may have on slope. It should also not be forgotten that linguistic similarity is intrinsically related to acoustic similarity and the effects of the two could not be disentangled in the slope survey. Indeed, Freyman et al.'s (2001) finding that reducing meaningful information in the masker only affected slope when speech maskers were already highly acoustically similar seems to suggest that linguistic similarity may play a more secondary role to acoustic similarity. This will be considered further in the first experiment of this thesis.

### 2.5.3 Evidence for slope changes as consequence of the availability of top-down information

The third group of explanations for changes in slope proposed in the literature are related to the availability of top-down information. When top-down information can highly constrain word possibilities, less weight needs to be placed on perceptual bottom-up information and speech identification can increase rapidly with increases in level. If, however, there is little top-down information available to constrain word options, intelligibility will be based on bottom-up information alone and will thus increase more slowly as level is increased. Several slope trends highlighted in the slope survey provided support for such a mechanism. Target stimuli which contained keywords which were predictable from their content were found to give steeper slopes, for example, than those whose keywords were unpredictable. Although individual studies have demonstrated this slope effect with both speech and noise maskers (Dirks et al., 1986; Dubno et al., 2000; Elliott, 1979; Kalikow et al., 1977; Lewis et al., 1988; Pichora-Fuller et al., 1995) the effect only held in the slope survey when a speech masker was used. The effect of target predictability on slope was still moderate to large in size ($r = -0.45$) for the static-noise masker and it is plausible that it did not quite reach statistical significance. It has been suggested that congruent previous-context constrains possible word options, shifting the influence on word identification from bottom-up to top-down information (Pichora-Fuller et al., 1995). With a greater dependence on top-down information, word identification can increase more rapidly with changes in level, hence the steeper slopes for predictable targets. Two other minor trends identified in the slope survey can be explained by similar theories, namely the effect of target length and the effect of target corpus.

The length of the target stimuli was found to have an effect on the slope of the psychometric function when the masker was a static noise. It could be argued that this slope effect is closely related to that seen when the predictability of the target was manipulated. The shorter the target token, the less contextual cues there are available to aid in top-down identification. Identification will, therefore, be dependent on perceptual, bottom-up information. As the token becomes longer, however, preceding context will begin to constrain possible response options increasing the probability that a word will be identified from the same amount of perceptual information, the result of which is to steepen the slope of the psychometric function. Very few studies were found which measured psychometric functions for speech tokens longer than a sentence in length. As such the slope survey was unable to ascertain the effect on slope that longer, continuous speech targets may have on the slope of the psychometric function. The role of continuous stimuli on slope is considered further in Chapter 6.

Changing the target corpus resulted in changes to the slope of the psychometric function when a static-noise masker was used. While some corpuses gave shallower slopes (e.g. IEEE) others gave steep slopes (SSI). Again this slope difference can be explained as difference in the degree to which top-down information was able to constrain word responses. The SSI, for example, is usually presented as closed-set corpus (Speaks & Jerger, 1965). Before the experiment listeners are presented with a list of possible sentences numbered from 1 to 10. They are then presented with a sentence over headphones and asked to match it to one of the sentences on the list by reporting the corresponding number. Top-down information in this case can very effectively constrain identification; only part of the sentence needs to be audible for identification to be successful as the listener has full knowledge of all possible sentences. Small changes in audibility can have large effects on intelligibility resulting in a steep slope. The IEEE corpus on the other hand is open set (Rothauser et al., 1969). The corpus consists of 720 sentences on different topics. The listener is presented sentences one at a time and asked to repeat what they heard; their responses are not restricted by a set of options. Top-down information is far less constraining in this case. While the context of the sentence may allow some top-down influence, speech identification will be much more heavily dependent on bottom-up information for these speech stimuli compared to the SSI, thus giving shallower slopes for these stimuli. Steep slopes were also found for other closed-set corpuses like the CRM corpus. The CRM corpus gave shallower slopes than the SSI corpus. This may be explained by the fact that while they are closed-set sentences (keywords can be selected from a combination of 32 options), there is no contextual or semantic cues available in CRM sentences to identify the keywords. The CRM keywords are, therefore, less constrained by top-information than the SSI corpus.

The HINT sentences which, like the IEEE sentences, are also open set, gave relatively steep slopes in noise maskers. The HINT sentences are, however, more predictable than the IEEE sentences. The slope difference between these corpuses may reflect this additional top-down information.

The effects of target corpus and target length were only found to be significant when noise maskers were used. It is plausible that slope differences were not found with speech maskers for the corpus and length effects as the overall effect of target type was greater. All slopes were relatively shallow in these conditions for speech maskers, meaning these more minor trends in slope may have been lost. That these slope effects were significant for static noise maskers also gives an indication that this mechanism is distinct from slope changes which result from target/masker confusion.

### 2.5.4 Evidence for slope changes as consequence of underlying variation

The fourth slope-change mechanism proposed in Chapter 1 suggested that slope changes arise when variation in listeners' responses are averaged. The source of the variation has been suggested to be anything from lapses in attention to variability in the measurement of the underlying function. The slope survey could not provide any evidence for this mechanism, however, as these sources of variation could not be extracted from the studies and encoded in the slope survey. For example, no assumptions could be made about how well listeners maintained attention in a particular study or whether the stimuli used might result in variations in listeners' responses across trials. Evidence for this mechanism will be returned to later, however, in the experimental work presented in this thesis.

### 2.5.5 Evidence for slope changes due to aging and hearing impairment

There is an indication from the slope survey that slopes may also be dependent on some listener factors. Slopes were found, for example, to be steeper for older listeners than they were for children or young adults. Unfortunately there were not a sufficient number of suitable cases to look separately at the effect that degree of hearing loss may also have on slope. This result is contrary to that of individual studies which have specifically commented on the differences in slope between younger normal-hearing and older hearing-impaired listeners (e.g. Bosman & Smoorenburg, 1995; Wagener & Brand, 2005). Aging and hearing loss often go hand-in-hand and we were not able to disentangle these two factors in the slope survey so it is possible that hearing impairment may also play a role in the observed slope difference here.

The trend for an increase in slope with age noted in the slope survey can be explained by at least two of the slope-change mechanisms outlined in Chapter 1. It was suggested that amplitude modulations in a masking sound allowed glimpses of the target speech. These glimpses mean both that target speech is randomly masked and audible over a wider dynamic range. Several studies have demonstrated that older listeners are less able than younger listeners to make use of brief dips in the power of background noise to help identify target speech (Festen & Plomp, 1990). This reduced ability has been attributed to a reduced temporal resolution (Lutman, 1991; Schneider, 1997). Reduced glimpsing would, however, result in the masker acting more like a static noise masker and may explain the steeper slopes given by older adults.

A second explanation for why we might see steeper slopes for older listeners than for younger listeners is that older listeners tend to have a greater reliance on top-down information. As well as changes in the temporal domain, notable changes also occur in the frequency domain with age. Older listeners tend to have broader auditory filters and poorer frequency selectivity (Moore, 1995). Poorer spectral and temporal resolution means that older listeners are often dealing with a degraded speech signal, particularly in noisy environments. It has been suggested that as the speech signal becomes less audible a greater reliance is placed on top-down information, e.g. context and knowledge are employed to successfully identify speech. Indeed, older listeners have been found to more greatly benefit from contextual cues than younger listeners (Pichora-Fuller et al., 1995). As outlined in section 1.3.3, when more reliance is put on top-down information the slope of the psychometric function tends to become steeper. The steeper slopes found for older listeners may, therefore, be a reflection of this. The effects of aging and hearing impairment on slope are considered further in the experimental work reported in this thesis.

### 2.5.6 The listening conditions which gave the most extreme slopes

The analysis of the slope data collected in the survey concentrated on slope trends within only one or two categories (for example the trend of target predictability for speech maskers). This gave us general listening conditions where slopes are likely to be steep or shallow; one speech masker will give shallower slopes, for example, than multiple speech maskers or one static-noise masker. The slope survey can, however, also be used to identify the precise combination of listening conditions which give the shallowest slopes − or conversely the steepest slopes. The shallowest slope in the survey (0.4% per dB ) was given by cochlear implant users identifying short tokens (consonants and vowels) in static noise with both target and maskers co-located at the same loudspeakers (Friesen, Shannon, Baskent, & Wang, 2010). The shallowest slope given by normal-hearing listeners (1.2% per dB) was found when words taken from the NU-6 corpus were presented monaurally in an interrupted noise masker with a slow interruption rate (Dirks, Wilson, & Bower, 1969). The steepest slope in the survey (29.7% per dB) was found for young normal-hearing listeners when French digits[13] were presented in average speech spectrum static noise (HearCom, 2009).

## 2.6 Summary

The slope of the psychometric function for masked speech varies greatly both from study to study and within studies. Understanding the factors affecting the slope of the psychometric function and the mechanisms which underlie these slope changes is important as it could give us a means of gauging the amount of perceptual benefit that can be expected given a specific change in level in a specific listening condition.

Four broad underlying mechanisms have been proposed to explain slope changes including slope changes as the result of: amplitude modulations in the masker;

[13] French was listeners' native language in this study.

confusion between the target and the masker; the availability of top-down information and random variability in performance. The systematic slope survey demonstrated that these slope-change mechanisms can be translated into general listening/stimulus conditions. Speech maskers, for example, give particularly shallow slopes as they contain amplitude modulations, share acoustic features with the target speech, and convey linguistic and semantic meaning.

One of the major problems with making evaluations from the slope survey, however, is that while individual listening conditions can be artificially considered (for example the effect of using a single speech masker on slope) the influence of other factors which were also present in the stimuli when each of the functions were measured could not be excluded. The slope survey has identified both major and minor trends in slope suggesting many factors have an effect on slope. It is possible then that weaker trends may be overlooked or that the observed trends were really the result of an interaction between several factors. The experimental work in this thesis, therefore, aims to investigate the listening conditions which are believed to lead to changes in the shape of the psychometric function while tightly controlling for other factors which might also affect slope. It is hoped that as a result a better understanding of slope changes can be achieved, the knowledge of which could be of use, not only, in future experimental work but also in computing the benefits that can be gained from any device that aims to improve speech intelligibility by offering a gain in SNR.

## 2.7 Experimental work

Following on from the concepts and issues raised in the slope survey the experimental work in this thesis will consider the following:

- Chapter 3 (Experiment 1) considers the role that the linguistic similarity of a target and masker, as distinct from their acoustic similarity, plays in flattening the slope of the psychometric function.

- Chapter 4 (Experiment 2) considers the situations in which unusual shaped psychometric functions arise and the relative contributions of two mechanisms − masker modulations and target/masker confusion − in their occurrence.

- Chapter 5 (Experiments 3 + 4) further explores the occurrence of unusual shaped psychometric functions considering the role that difficulties selecting a target (as opposed to difficulties segregating or streaming a target) have on these functions.

- Chapter 6 (Experiments 5 + 6) considers the shape of psychometric functions given by more realistic listening conditions, namely continuous target speech.

- Chapter 7 summarises and discusses the main findings and issues raised in the thesis.

As there was an indication that age and/or hearing loss may in itself affect the slope of the psychometric function, much of the experimental work in this thesis is carried out with older listeners, for whom changes in slope may not only be different to those of young normal-hearing listeners but may also be most consequential.

***Table 2.1:*** *Displays all of the studies found in the systematic survey in alphabetical order. Key aspects of stimuli and design are reported for each study. Listener ages are coded as: children, young (refers to young adults) and older (refers to older adults). Speech corpuses are coded as:*

*AB Lists – Isophonemic Monosyllabic Word test (Boothroyd 1968).*

*ASL – Audio-visual Sentence Lists (MacLeod & Summerfield, 1990).*

*BKB – Bamford-Kowal-Bench sentence lists (Bench,Kowal & Bamford, 1979).*

*CAT – Call Sign Acquisition test (Rao & Letowski, 2006)*

*CCT – California Consonants Test (Owens & Schubert, 1977).*

*CID sentences - Central Institute for the deaf sentences (Silverman & Hirsh, 1955).*

*CRM – The Coordinate Response measure (Bolia et al., 2000).*

*DANTALE II –Danish sentence test (Wagener, Josvassen & Ardenkjoer, 2003).*

*FAAF – The Four Alternative Auditory Feature tests (Foster & Haggard, 1987).*

*Hagerman sentences – Swedish sentences (Hagerman, 1982).*

*HINT – Hearing In Noise Test (Nilsson, Soli & Sullivan, 1994).*

*IEEE – IEEE sentences (Rothauser et al., 1969).*

*MCDT - Multiple Choice Discrimination Test (Schultz & Schubert 1969).*

*MRT – Modified Rhyme Test (House, Willaims, Hecker & Kryter, 1965).*

*NU-6 – Northwestern University Auditory Tests No. 6.(Tillman & Carhart, 1966).*

*PB List – Phonetically balance words lists (Egan, 1948).*

*Picture Identification Task (Wilson and Antablin, 1980)*

*PSI Test - Pediatric Speech Intelligibility Test (Jerger, Jerger & Lewis 1981).*

*SPIN – Speech in noise test (Kalikow, Stevens & Elliott, 1977).*

*SSI – Synthetic Sentence Identification Test (Speaks & Jerger, 1965).*

*TMV – open-set sentences (Helfer & Freyman, 2009).*

*W-22 – Central Institute for the deaf word lists (Benson, Davis, Harrison, Hirsh, Reynolds & Silverman, 1951).*

| Study | N slopes | Target Corpus | Masker type | Masker Corpus | Presentation | Listener Age | Listener hearing levels |
|---|---|---|---|---|---|---|---|
| Acton (1970) | 2 | PB List | Static noise | - | Loudspeaker | Young | Normal |
| Arbogast, Mason & Kidd (2002) | 24 | CRM | Speech & modulated noise | CRM | Loudspeaker | Young | Normal |
| Barker and Cooke (2007) | 1 | Valid sentences | Modulated noise | - | Headphones | Young | Normal |
| Beattie (1989) | 1 | W-22 | Static noise | - | Headphones | - | Impaired |
| Beattie, Barr & Roup (1997) | 1 | W-22 | Speech | Valid sentences | Earphone | Young | Normal |
| Beattie & Clark (1982) | 4 | SSI | Speech | Valid sentences | Earphone | Young | Normal |
| Bernstein & Grant (2009) | 12 | IEEE | Speech, modulated & static noise | HINT | Headphones | Young & Older | Normal & impaired |
| Best, Gallun, Mason, Kidd & Shinn-Cunningham (2010) | 6 | CRM | Speech & static noise | CRM | Headphones | Young | Normal & impaired |
| Bhattacharya & Zeng (2007) | 17 | Phonemes & HINT | Static noise | - | Headphones | Young | Normal & CI User ● |
| Blue-Terry & Letowski (2011) | 1 | MRT | Static noise | - | Headphones | Young | Normal |
| Boothroyd (2008) | 6 | Phonemes | Static noise | - | - | Young | Normal |
| Boothroyd & Nittrouer (1988) | 5 | Phonemes and Invalid sentences | Static noise | - | Headphones | Young | Normal |
| Bosman & Smoorenburg (1995) | 17 | Syllables, valid & Invalid sentences | Static noise | - | Earphone | Young & Older | Normal & impaired |
| Bronkhorst, Bosman & Smoorenburg (1993) | 4 | Words | Static noise | - | Headphones | Young | Normal |

● *Cochlear implant user*

*Table 2.1 continued*

| Study | N slopes | Target Corpus | Masker type | Masker Corpus | Presentation | Listener Age | Listener hearing levels |
|---|---|---|---|---|---|---|---|
| Brungart (2001a) | 1 | CRM | Static noise | - | Headphones | Young | Normal |
| Brungart (2001b) | 2 | CRM | Speech, modulated & static noise | CRM | Headphones | Young | Normal |
| Brungart, Iyer & Simpson (2006) | 8 | CRM | Speech | CRM | Headphones | Young | Normal |
| Brungart, Chang, Simpson & Wang (2006) | 6 | CRM | Modulated & static noise | - | Headphones | Young | Normal |
| Brungart, Chang, Simpson & Wang (2009) | 15 | CRM | Speech | CRM | Headphones | Young | Normal |
| Brungart & Simpson (2002) | 6 | CRM | Speech, modulated & static noise | CRM | Headphones | Young | Normal |
| Brungart & Simpson (2004) | 9 | CRM | Speech | CRM | Headphones | Young | Normal |
| Brungart & Simpson (2007) | 15 | CRM | Speech | CRM | Headphones | Young | Normal |
| Brungart, Simpson, Darwin, Arbogast & Kidd (2001) | 3 | CRM | Speech | CRM | Headphones | Young | Normal |
| Brungart, Simpson, Ericson & Scott (2001) | 18 | CRM | Speech & modulated noise | CRM | Headphones | Young | Normal |
| Brungart, Simpson & Freyman (2005) | 3 | CRM | Speech & static noise | CRM | Headphones | Young | Normal |
| Cienkowski & Speaks (2000) | 1 | Words | Static noise | - | Earphone | Young & Older | Normal & impaired |
| Cooke, Hershey & Rennie (2010) | 2 | Grid | Speech | W-22 | Headphones | Young | Normal |
| Cooper & Cutts (1971) | 2 | NU-6 | Static noise | - | Earphone | Young | Normal |
| Craig (1988) | 9 | SPIN | Speech | SPIN | Earphone | Young | Normal |
| Crandell (1993) | 1 | BKB | Speech | SPIN | Headphones | Children | Impaired |

*Table 2.1 continued*

| Study | N slopes | Target Corpus | Masker type | Masker Corpus | Presentation | Listener Age | Listener hearing levels |
|---|---|---|---|---|---|---|---|
| Danhauer, Doyle & Lucks (1986) | 3 | Invalid sentences | Speech, modulated & static noise | Valid sentences | Headphones | Young | Normal |
| Danhauer & Leppler (1979) | 3 | CCT | Speech & modulated noise | Valid sentences | Headphones | Young | Normal |
| Darwin, Brungart & Simpson (2003) | 12 | CRM | Speech | CRM | Headphones | Young | Normal |
| Dirks, Bell, Rossman & Kincaid (1986) | 3 | SPIN | Speech & static noise | Valid sentences | Earphone | Young | Normal |
| Dirks & Bower (1969) | 37 | Valid sentences | Speech & modulated noise | Valid sentences | Earphone | Young | Normal |
| Dirks, Morgan & Dubno (1982) | 8 | NU-6 | Speech | Valid sentences | Earphone | Young | Normal |
| Dirks & Wilson (1969)a | 18 | SSI | Static noise | - | Headphones | Young & Older | Normal & impaired |
| Dirks & Wilson (1969)b | 20 | PB Lists | Static noise | - | Headphones & loudspeaker | Young | Normal |
| Dirks, Wilson & Bower (1969) | 43 | NU-6 | Modulated & static noise | - | Earphone | Young | Normal |
| Drullman (1995) | 2 | Valid sentences | Static noise | - | Headphones | Young | Normal |
| Drullman & Bronkhorst (2004) | 4 | Valid sentences | Speech | Valid sentences | Headphones | Young | Normal |
| Dubno, Horwitz & Ahlstrom (2005) | 3 | NU-6 | Static noise | - | Earphone | Young | Normal |
| Egan (1948) | 1 | Syllables | Static noise | - | Headphones | - | - |
| Egan, Carterette & Thwing (1954) | 1 | Valid sentences | Speech | Valid sentences | Earphone | Young | Normal |
| Eisenberg, Dirks & Bell (1995) | 4 | SPIN | Modulated & static noise | - | Earphone | Young | Normal |

*Table 2.1 continued*

| Study | N slopes | Target Corpus | Masker type | Masker Corpus | Presentation | Listener Age | Listener hearing levels |
|---|---|---|---|---|---|---|---|
| Elliot (1979) | 6 | SPIN | Speech | Valid sentences | Headphones | Children | Normal |
| Erber (1971) | 4 | Words | Static noise | - | Headphones | Young | Normal |
| Ezzatain, Li, Pichora-Fuller & Schneider (2010) | 27 | Invalid sentences | Speech & static noise | Invalid sentences | Loudspeaker | Young & older | Normal & impaired |
| Feeney & Franks (1982) | 4 | Syllables PB Lists & MRT | Static noise | - | Headphones | Young | Normal |
| Festen & Plomp (1990) | 6 | Valid sentences | Modulated noise | - | Headphones | Young | Normal |
| Foster & Haggard (1987) | 1 | FAAF | Static noise | - | Headphones | Young | Normal |
| Freyman, Balakrishnan & Helfer (2001) | 16 | Invalid Sentences | Speech & modulated noise | Invalid sentences | Loudspeaker | Young | Normal |
| Freyman, Balakrishnan & Helfer (2004) | 15 | Invalid Sentences | Modulated & static noise | Invalid sentences | Loudspeaker | Young | Normal |
| Freyman, Balakrishnan & Helfer (2007) | 6 | Invalid Sentences | Speech | Invalid sentences | Loudspeaker | Young | Normal |
| Freyman, Helfer, McCall & Clifton (1999) | 12 | Invalid Sentences | Speech & static noise | Invalid sentences | Loudspeaker | Young | Normal |
| Friesen, Shannon, Baskent & Wang (2001) | 19 | Vowels, consonants & HINT | Static noise | - | Loudspeaker | Older | Normal & CI users |
| Fu, Shannon & Wang (1998) | 9 | Vowels & consonants | Static noise | - | Headphones | Young | Normal |
| Gelfand (1998) | 2 | Phoneme and words | Static noise | - | Headphones | Young | Normal |
| Grant & Braida (1991) | 7 | IEEE sentences | Static noise | - | Headphones | Young | Normal |

*Table 2.1 continued*

| Study | N slopes | Target Corpus | Masker type | Masker Corpus | Presentation | Listener Age | Listener hearing levels |
|---|---|---|---|---|---|---|---|
| Griffiths (1967) | 4 | MRT | Static noise | - | Earphone | Young | Normal |
| Hagerman (1982) | 1 | Hagerman sentences | Modulated noise | - | Headphones | Young | Normal |
| Hallgren, Larsby & Arlinger (2006) | 1 | HINT | Static noise | - | Loudspeaker | Young | Normal |
| HearCom (2009) | 13 | Digits | Static noise | - | Headphones | - | - |
| Helfer (1997) | 2 | Invalid sentences | Speech | Invalid sentences | Headphones | Young | Normal |
| Helfer & Freyman (2005) | 8 | Invalid sentences | Speech & static noise | Invalid sentences | Loudspeaker | Young | Normal |
| Helfer & Freyman (2008) | 4 | Invalid sentences | Speech & modulated noise | Invalid sentences | Loudspeaker | Young & older | Normal & impaired |
| Helfer & Freyman (2009) | 13 | TMV | Speech | TMV | Loudspeaker | Young | Normal |
| Hirsh, Reynolds & Joseph (1954) | 2 | Words & Invalid sentences | Static noise | - | Headphones | Young | Normal |
| Horii, House and Hughes (1971) | 3 | Vowels & consonants | Modulated & static noise | - | Headphones | Young | Normal |
| House, Williams, Hecker & Kryter (1965) | 1 | Words | Static noise | - | Headphones | Young | Normal |
| Howard-Jones & Rosen (1993) | 5 | Words | Modulated & static noise | - | Headphones | Young | Normal |
| Ihlefeld & Shinn-Cunningham (2008) | 3 | CRM | Speech | CRM | Headphones | Young | Normal |
| Jerger & Jordan (1992) | 2 | Continuous speech | Speech | Continuous speech | Loudspeaker | Young | Normal |
| Jerger, Jerger & Lewis (1981) | 2 | PSI Test | Speech | PSI Test | Loudspeaker | Young | Normal |

*Table 2.1 continued*

| Study | N slopes | Target Corpus | Masker type | Masker Corpus | Presentation | Listener Age | Listener hearing levels |
|---|---|---|---|---|---|---|---|
| Johnstone & Litovsky (2006) | 12 | Spondees | Speech & modulated noise | IEEE | Loudspeaker | Children & Young | Normal |
| Kalikow, Stevens & Elliot (1977) | 4 | SPIN | Speech | Valid sentences | Headphones | Young & older | Normal |
| Kates & Arehart (2005) | 1 | HINT | Modulated noise | - | Earphone | Young | Normal |
| Keith & Talis (1970) | 3 | W-22 | Static noise | - | Earphone | Young & older | Normal & impaired |
| Kidd, Mason & Gallun (2005) | 3 | CRM | Speech & static noise | CRM | Headphones | Young | Normal |
| Krull, Choi, Kirk, Prusick & French (2010) | 1 | Words | Static noise | - | Headphones | Children | Normal |
| Kryter (1962) | 7 | Syllables, PB, MRT & Valid sentences | Static noise | - | Earphone | Young | Normal |
| Kryter & Whitman (1965) | 2 | PB & MRT | Static noise | - | Earphone | Young | Normal |
| Lewis, Benignus, Muller, Mallot & Barton (1988) | 3 | SPIN | Static noise | - | Earphone | Young | Normal |
| Li, Daneman, Qi & Schneider (2004) | 18 | Invalid sentences | Speech & static noise | Invalid sentences | Loudspeaker | Young & older | Normal |
| Li & Loizou (2009) | 7 | IEEE | Speech & static noise | IEEE | Headphones | Young | Normal |
| MacLeod & Summerfield (1990) | 1 | ASL | Static noise | - | Headphones | Young | Normal |
| Martin & Mussell (1979) | 2 | Words & SSI | Speech & static noise | Valid sentences | Headphones | Young | - |
| McArdle, Wilson & Burks (2005) | 5 | NU-6, digits & IEEE | Speech | Valid sentences | Headphones | Young & older | Normal & impaired |

*Table 2.1 continued*

| Study | N slopes | Target Corpus | Masker type | Masker Corpus | Presentation | Listener Age | Listener hearing levels |
|---|---|---|---|---|---|---|---|
| Miller, Heise & Litcten (1951) | 14 | Syllables, Words, digits & valid sentences | Static noise | - | Headphones | - | - |
| Ng, Meston, Scollie & Seewald (2011) | 6 | BKB | Speech | Valid sentences | Loudspeaker | Children | Normal & impaired |
| Niederjohn & Grotelueschen (1976) | 4 | PB Lists | Static noise | - | Earphone | Young | Normal |
| Nielsen & Dau (2009) | 2 | HINT | Static noise | - | Headphones | Young | Normal |
| Oxenham and Simonson (2009) | 12 | HINT | Speech & modulated noise | IEEE | Headphones | Young | Normal |
| Ozimek, Kutzner, Sek & Wicher (2009) | 1 | Valid sentences | Speech | Valid sentences | Headphones | Young | Normal |
| Ozimek, Warzybok & Kutzner (2010) | 1 | Matrix | Speech | Rhyme Test | Headphones | Young | Normal |
| Pederson & Studebaker (1972) | 2 | Words | Static noise | - | Headphones | Young | Normal |
| Pichora-Fuller, Schneider & Daneman (1995) | 6 | SPIN | Speech | SPIN | Headphones | Young & older | Normal & impaired |
| Pichora-Fuller, Schneider, MacDon - ald, Pass & Brown (2007) | 4 | SPIN | Speech | SPIN | Earphone | Young | Normal |
| Plomp and Mimpen (1979) | 2 | Valid sentences | Static noise | - | Headphones | Young | Normal |
| Rakerd, Aaronson & Hartmann (2006) | 4 | CRM | Speech | CRM | Loudspeaker | Young | Normal |
| Rao & Letowski (2004) | 6 | CAT Test | Static nose | - | Headphones | Young | Normal |
| Rogers, Lister, Dashielle & Besing (2006) | 3 | W-22 | Static nose | - | Headphones | Young | Normal |
| Schultz & Schubert (1969) | 2 | W-22 & MCDT | Static nose | - | Headphones | - | - |

*Table 2.1 continued*

| Study | N slopes | Target Corpus | Masker type | Masker Corpus | Presentation | Listener Age | Listener hearing levels |
|---|---|---|---|---|---|---|---|
| Scott, Rosen, Wickham & Wise (2004) | 1 | BKB | Static noise | - | Headphones | Young | Normal |
| Sergeant, Atkinson & Lacroix (1979) | 3 | MRT | Static noise | - | Earphone | Young | Normal |
| Sherbecoe & Studebaker (2002) | 2 | Continuous speech | Static noise | - | Earphone | Young | Normal |
| Speaks & Karmen (1967) | 1 | Valid sentences | Static noise | - | Earphone | Young | Normal |
| Speaks, Karmen & Benitez (1967) | 4 | SSI | Speech | SSI | Earphone | Young | Normal |
| Speaks, Parker, Kuhl, Harris, (1972) | 3 | Continuous speech | Static noise | - | Earphone | Young | Normal |
| Stickney, Zeng, Litovsky & Assmann (2004) | 10 | IEEE | Speech & static noise | IEEE | Headphones | Young | Normal & CI users |
| Studebaker, Taylor & Sherbecoe (1994) | 4 | Words | Static noise | - | Earphone | Young | Normal |
| Surprenant (2007) | 2 | Syllables | Static noise | - | Headphones | Young & older | Normal |
| Surr & Schwartz (1980) | 1 | CCT | Speech | Valid sentence | Headphones | Young | Impaired |
| Suter (1985) | 4 | MRT & CID sentences | Speech | Valid sentence | Loudspeaker | - | Impaired |
| Tabri, Abou Chacra & Pring (2011) | 8 | SPIN | Speech | Valid sentence | Headphones | Young | Normal |
| Takahashi & Bacon (1992) | 8 | SPIN | Modulated & static noise | - | Earphone | Young & older | Normal & impaired |
| Theodoridis & Schoeny (1988) | 3 | W-22 | Static noise | - | Headphones | Young | Normal |

*Table 2.1 continued*

| Study | N slopes | Target Corpus | Masker type | Masker Corpus | Presentation | Listener Age | Listener hearing levels |
|---|---|---|---|---|---|---|---|
| Theodoridis, Schoeny & Anne (1985) | 2 | W-22 | Static noise | - | Headphones | Young | Normal |
| Trammell & Speaks (1970) | 2 | Valid sentence | Speech | Valid sentence | Earphone | Young | Normal |
| Tun (1998) | 6 | Valid sentence | Speech | Valid sentence | Headphones | Young & older | Normal |
| Van Wieringen & Wouters (2008) | 2 | Digits | Static noise | - | Headphones | Young | Normal |
| Vestergaard, Fyson & Patterson (2009) | 3 | Syllables | Speech & static noise | Syllables | Headphones | Young | Normal |
| Wagener, Josvassen & Ardenkjoer (2003) | 1 | Dantale 2 | Static noise | - | Headphones | Young | Normal |
| Whitmal, Poissant, Freyman & Helfer (2007) | 5 | Syllables & valid sentences | Speech & static noise | Invalid sentence | Headphones | Young | Normal |
| Wightman & Kistler (2005) | 28 | CRM | Speech & static noise | CRM | Headphones | Children & young | Normal |
| Wilson & Antablin (1980) | 2 | Picture Identification task & NU-6 | Static noise | - | Earphone | Young | Normal |
| Wilson & Burks (2005) | 3 | Words | Speech | Words | Earphone | Older | Impaired |
| Wilson, Burks & Weakley (2006) | 12 | NU-6 & digits | Static noise | - | Headphones | Older | Impaired |
| Wilson, Carnell & Cleghorn (2007) | 4 | Words | Speech & static noise | Valid sentence | Earphone | Young & older | Normal & impaired |

*Table 2.1 continued*

| Study | N slopes | Target Corpus | Masker type | Masker Corpus | Presentation | Listener Age | Listener hearing levels |
|---|---|---|---|---|---|---|---|
| Wilson & Cates (2008) | 2 | Words | Speech | Valid sentences | Headphones | Younger & older | Normal & impaired |
| Wilson, Farmer, Gandhi, Shelburne & Weaver (2010) | 8 | Words | Static noise | - | Earphone | Children & young | Normal |
| Wilson & McArdle (2007) | 4 | Words | Speech | Valid sentences | Headphones | Older | Impaired |
| Wilson, McArdle, Betancourt, Herring, Lipton & Chisolm (2010) | 13 | Words | Speech, modulated & static noise | Valid sentences | Earphone | Younger & older | Normal & impaired |
| Wilson, McArdle & Roberts (2008) | 15 | PB Lists, W-22, NU-6 & digits | Static noise | - | Earphone | Young | Normal |
| Wilson, McArdle & Smith (2007) | 6 | Words, IEEE & BKB | Speech | Valid sentences | Earphone | Younger & older | Normal & impaired |
| Wilson & Oyler (1997) | 2 | W-22 & NU-6 | Static noise | - | Earphone | Young | Normal |
| Wilson & Strouse (2002) | 8 | NU-6 | Speech | Valid sentences | Headphones | Younger & older | Normal & impaired |
| Wu, Wang, Chen, Qu, Li, Wu, Schneider and Li (2005) | 6 | Invalid sentences | Speech & static noise | Invalid sentences | Loudspeaker | Young | Normal |
| Yang, Chen, Huang, Wu, Wu, Schneider and Li (2007) | 18 | Syllables & words | Speech & static noise | Invalid sentences | Loudspeaker | Young | Normal |
| Young, Goodman & Carhart (1979) | 3 | Words | Speech | Valid sentences | Headphones | Young | Normal |

# 3 Experiment 1: The effect of masker intelligibility on the slope of the psychometric function for masked speech.

The systematic survey of slopes detailed in chapter 2 indicated that speech maskers produce shallower psychometric functions than noise maskers. It has been suggested that this slope difference is related to the degree of confusion that exists between the target and the masker (Arbogast et al., 2002; Bronkhorst, 2000; Brungart, 2001a). The exact origin of this confusability is still, however, under debate. The confusion could arises from difficulties determining the sources of the each sound, separating the two messages or selecting one message while rejecting the other, i.e. the *acoustic* similarity of the masker could play a major role. It is also possible, however, that confusion could arise because the masking signal also contains words which, like the target, may evoke higher-order speech processing, i.e. the *linguistic* similarity of the masker plays a role. The slope survey provided evidence that changes in slope do occur if the acoustic similarity is reduced when targets and maskers are highly similar, the evidence for an affect of linguistic similarity on slope was, on the other hand, weaker. It was not possible, however, to separate the effects of acoustic and linguistic similarity from one another in the slope survey and it is possible that the stronger effect of the former overshadowed that of the later. Experiment 1, therefore, aims to manipulate a masker's linguistic similarity to a target while making minimal changes to its acoustic similarity, and observing the effect this manipulation has on the slope of the psychometric function.

The main challenge of designing such an experiment is the need to create maskers which vary only in their intelligibility, i.e. maskers that retain comparable acoustic features even when any intelligible linguistic content has been removed. In the past unintelligible or meaningless maskers have been created by time-reversing speech messages (e.g. Dirks & Bower, 1969; Rossi-Katz & Arehart, 2009) or by using maskers spoken in a foreign language (e.g. Dirks & Bower, 1969; Freyman et al., 2001). These maskers retain the long-term characteristics of speech but are essentially meaningless. Increasing the number of talkers in the masker has also been used to manipulate masker intelligibility (Hoen et al., 2007). While individual words are still intelligible in a masker with four talkers, the greater temporal overlap in speech signals that occurs as the number of talkers is increased to eight greatly reduces any available linguistic information.

There are, however, several drawbacks to using the aforementioned maskers to assess the effect that masker intelligibility alone may have on the slope of the psychometric function. Firstly, manipulations that increase the number of talkers in the masker would not be an appropriate way of manipulating intelligibility in the current experiment. It was demonstrated in the slope survey, for example, that the slope of the psychometric function increased as the number of talkers were also increased. It was argued that this was due to the reduction in the overall amplitude modulations as temporal overlap between masking voices increased. If such a manipulation was used in the current study, therefore, we would be unable to attribute any slope changes solely to the intelligibility manipulation as modulation changes are also likely to occur as the masker becomes unintelligible. Secondly, for several of these maskers acoustic and linguistic similarity cannot be separated, changes in the intelligibility of a masker ultimately result in changes in its acoustic similarity to a target. Playing a message backwards, for example, not only eliminates the intelligibility of the message, but also inadvertently alters its similarity to the target in other ways. Rhebergen, Versfeld and Dreschler (2005) noted, for example, that when speech is played backwards the envelope is also reversed. Instead of the temporal envelope being characterised by quick onsets and slow decays, it would instead be characterised by slow onsets and abrupt decays. The abrupt offsets result in greater amount of forward masking as well as introducing a difference between the temporal patterns of the target and the masker − a factor suggested to effect slope (Dirks & Bower, 1969). This suggests that again this masker manipulation would not allow slope changes to be attributed to intelligibility changes alone.

It may be preferable when looking at the effect that the linguistic similarity (as a separate factor to the acoustic similarity) of the masking speech has on slopes to use a masker which is *already* perceptually distinct from the target before the intelligibility manipulation. Brungart, Darwin, Arbogast & Kidd (2005), while looking at the effects of contralateral interference in dichotic listening tasks, used synthetic speech maskers such as noise vocoded and sine vocoded speech. The process of vocoding greatly alters the acoustic properties of speech, reducing spectral cues and eliminating temporal fine structure. Speech maskers which have been vocoded sound quite different to natural speech targets and are, therefore, likely to be extremely distinct. Brungart, Darwin, et al., (2005) used vocoded maskers to explore the acoustic characteristics needed to cause across-ear interference as vocoded speech can be changed systematically from “noise-like” to “speech-like” by varying the number of component bands. During the vocoding process the speech is filtered into logarithmically (or similar) spaced frequency bands, the amplitude envelope of each of these bands is then extracted, often by half or full-wave rectification followed by low-pass filtering. The extracted envelopes are then used to modulate a carrier noise which has been filtered into corresponding frequency bands to the original signal (Shannon et al., 1995). By varying the number of frequency bands or by varying the cut-off frequencies of the envelopes the output will sound more or less speech-like. For example, Dorman, Loizou and Rainey (1997) found speech recognition decreased as either the number of frequency bands or the cut-off frequency of these bands were reduced. Brungart, Darwin, et al., (2005) found that the contralateral interference increased systematically as the number of frequency bands in the vocoded speech increased, i.e. performance decreased as the masker became more speech like.

Intelligible and unintelligible maskers which were created by vocoding speech and manipulating the number of frequency bands (or the cut-off frequency of these bands) would still differ somewhat in their acoustic similarity to natural target speech despite being qualitatively distinct from it. For example, a vocoded masker with a greater number of frequency bands would not only be more intelligible than one with fewer frequency bands, but it would also be more speech-like. The current study, therefore, used vocoded speech but used a different method to manipulate its intelligibility. Instead of using just one sentence to create each masker, several different sentences were selected. Envelopes extracted from *different* sentences were then used to modulate the different bands of the carrier noise. For example, envelopes from bands 1 and 2 of sentence 1 were used to modulate bands 1 and 2 of the carrier noise, while bands 3 and 4 of sentence 2 were used to modulated bands 3 and 4, and so on. Combining the bands from disparate sentences resulted in a masker that was unintelligible but that still sounded equivalent to a masker that has been vocoded to remain intelligible. Both intelligible and unintelligible maskers had the same number of frequency bands and the cut-off frequencies and modulations within a channel were the same. It is argued, therefore, that their acoustic similarity to the target was not differentially affected. Nevertheless, it is possible that the overall modulations of the two signals were different – this is returned to in the discussion.

The two masking conditions used in this experiment were intelligible vocoded speech, termed "IVS", and unintelligible vocoded speech, termed "UVS". Since both maskers are perceptually distinct from the target speech any changes in slope can, therefore, be attributed to the intelligibility manipulation. If a speech masker gives rise to a shallower slope than a noise masker because of intelligibility per se then it would be expected that the slopes would be shallower for the IVS masker than for the UVS masker.

In addition to these two experimental conditions, psychometric functions were also measured in two control conditions; a natural speech masker, termed "S" (spoken by the same talker as the target) and a static noise masker, termed "N". The two control conditions were expected to give two extremes of slope, the S masker producing the shallowest slopes (being acoustically and linguistically similar to the target) and N maskers producing the steepest slopes (being neither acoustically and linguistically similar to the target). It was of interest then, where the slopes for the vocoded conditions would fall in relation to these two extremes: would the vocoded conditions produce slopes that are more similar to those produced by speech maskers or those produced by noise maskers? If, as indicated by the slope survey, acoustic similarity *primarily* affects slope we would expect the IVS masker, which is acoustically distinct from the target speech, to produce steeper slopes than the S masker.

A group of older listeners and a control group of younger listeners took part in the study. As discussed in the introduction of chapter 1, older listeners often report speech understanding difficulties in noisy environments (CHABA, 1988) but situations where more than one person is speaking at once are particularly challenging (Duquesnoy, 1983; Tun & Wingfield, 1999). It has been suggested that these problems are at least in part, associated with greater difficulties dealing with irrelevant linguistic information. It has been proposed that older listeners have difficulties inhibiting the processing of competing speech (Tun, Wingfield, & O'Kane, 2002). Several theories suggest that speech processing is an automatic, obligatory activity (e.g. Hawley et al., 2004) and as such suppressing an intelligible speech masker, as opposed to an unintelligible one, requires additional attentional resources. Age-related declines in cognitive resources (e.g. Schneider, Pichora-Fuller, & Daneman, 2010), or the redeployment of these resources to compensate for peripheral degradations (Pichora-Fuller et al., 1995) may result in a reduced ability to exert attentional control and inhibit processing of the masker. Several studies support this finding, demonstrating that removing the meaning from a masker had a greater effect on intelligibility for older adults than it did for younger adults (Rossi-Katz & Arehart, 2009; Tun et al., 2002). It could be argued that greater difficulties suppressing competing speech would result in greater linguistic confusion for older listeners than for younger listeners. The effect of manipulating the intelligibility of a masker in the current experiment was expected, therefore, to have a particular impact for older listeners.

## 3.1 Method

### 3.1.1 Listeners

Twenty-one older listeners and seven younger listeners took part in the experiment. The older listeners were all volunteers who had previously agreed to take part in research at the MRC Institute of Hearing Research. Their ages ranged from 56-73 (mean age = 68). Of the younger listeners, one listener was the author, another was a member of the public, and the remaining five listeners were staff at the MRC Institute of Hearing Research. Their ages ranged from 21-38 (mean age = 27). The hearing levels of all listeners were assessed using pure tone audiometry. The older group had four frequency average (4FA) hearing thresholds ranging from 12.5 dB HL (normal) – 42.5 dB HL (moderate hearing loss). All younger listeners had 4FAs of 10 dB HL or less. A small participation allowance was offered to listeners for their attendance and travelling expenses were refunded.

### 3.1.2 Stimuli

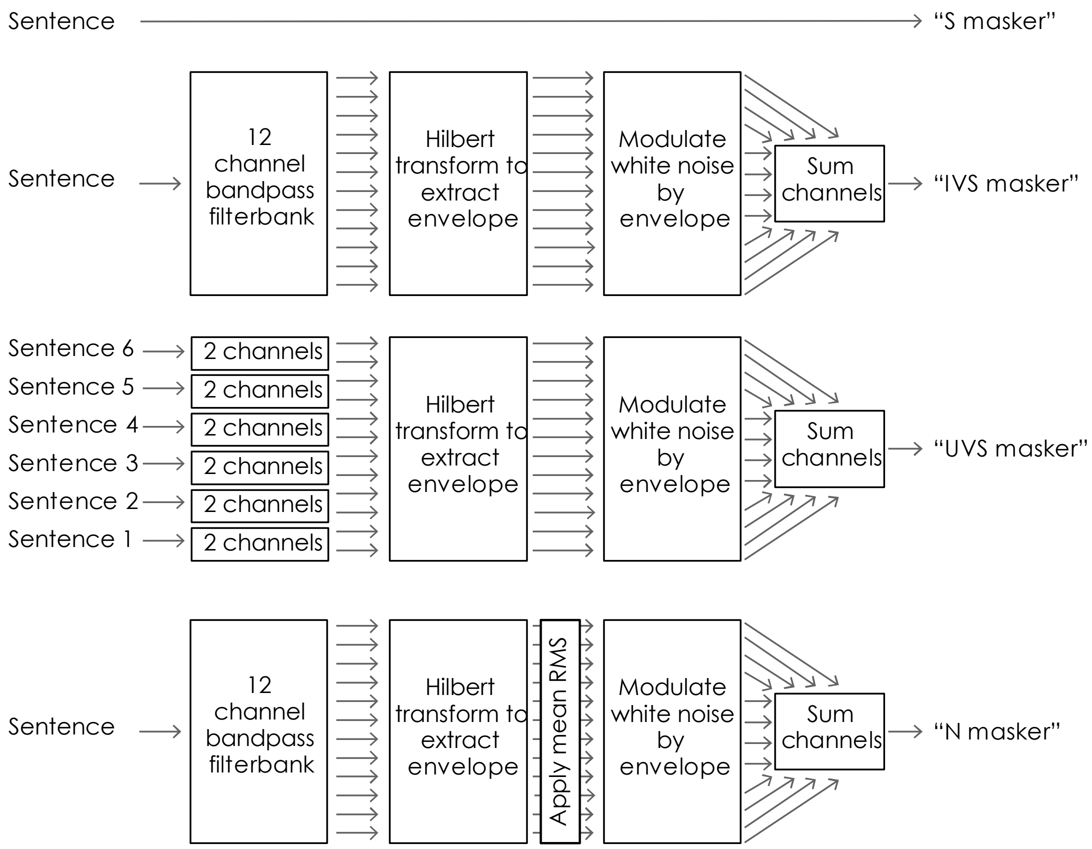
The stimuli consisted of target speech presented in a speech masker, two types of vocoded speech maskers, or a noise masker. The target sentences were taken from Kitterick et al.'s (2010) recordings of the Coordinate Response Measure (CRM) corpus (Bolia et al., 2000) and were spoken by a British-English male. The full corpus consists of 256 sentences with the form "*Ready [call sign] go to [colour] [number]* and includes eight possible call signs (Arrow, Baron, Charlie, Eagle, Hopper, Laker, Ringo, and Tiger), four colours (red, blue, green, and white), and eight numbers (1-8). In the current experiment listeners were not required to identify the call sign of target sentences and so for simplicity this was fixed and only the 32 sentences with the call sign Baron were used.

Four masker conditions were used in the current experiment: natural speech (S), intelligible vocoded speech (IVS), unintelligible vocoded speech (UVS) and a static noise (N). The S maskers were created by concatenating two different, randomly chosen IEEE sentences together without a gap between them (Rothauser et al., 1969). These were spoken by the same male talker used in the target sentences.

The IVS, UVS and N maskers were created by applying a vocoder (Shannon et al., 1995) to the S maskers (see figure 3.1). First each S masker was bandpass filtered. Bands spanned a range from 72.5 – 8000 Hz and this range was divided into 12 bands based on the Greenwood filter map[14]. The envelope of each band was extracted using the Hilbert transformation. These envelopes were then used to modulate white noises which had been filtered by the same bandpass filters used to filter the original stimuli. For the IVS maskers the 12 channels of modulated noise were then summed. For the UVS maskers 12 channels from six different randomly selected S stimuli were summed: channels 1 and 2 were modulated by envelopes extracted from channels 1 and 2 of the first randomly selected sentence; channels 3

[14] The Greenwood map divides bands based on equal basilar membrane distance (Greenwood 1990).

***Figure 3.1:*** *Illustration of the processes used to create the four maskers. For the natural speech masker (S masker) sentences underwent no processing. For the remaining three maskers sentences were vocoded to produce intelligible (IVS masker), unintelligible (UVS masker) or static noise (N masker) maskers.*



and 4 were modulated by envelopes extracted from channels 3 and 4 of the second randomly selected sentences, and so on[15]. For the N masker, the RMS of the envelope was calculated and applied to the entire envelope in each channel, thus removing any amplitude modulation, the 12 channels were then summed. 100 different versions of each of the four maskers were created in total and stored for use in the experiment. A further 25 IEEE sentences were also vocoded following the same method used to create the IVS maskers. These sentences were used as training stimuli.

The target sentences had an average duration of 2.5 seconds while the maskers had an average duration of 5.1 seconds. The target sentences were added to the start of the maskers with a randomly chosen onset delay of between 1500-2000 ms (the range was calculated so that the key words from the target sentences always occurred in the second sentence of the masker) The maskers continued for 50 ms after target offset. 25 ms raised cosine gates were then applied to target/masker stimuli.

### 3.1.3 Apparatus

The experiment was carried out in a sound-treated booth. The stimuli were digitally generated on a PC and presented using a RME DIG196/8 PAD soundcard, an Arcam A80 amplifier and Sennheiser HD580 precision headphones. The stimuli levels were measured with a B&K 2260 Observer sound-level meter, using a “slow” response, A-weighting filter, and a B&K artificial ear (type 4189) with a ½ inch microphone (type 4134). Maskers were presented at an average level of 70 dB SPL.

### 3.1.4 Procedure

On each trial listeners heard a target sentence presented in a masker. They were then shown an eight-by-four array on a computer screen in front of them, containing all 32 possible colour/number combinations available in the CRM target sentences. Listeners were asked to identify the colour and the number from the target sentence by pressing the appropriate button on the screen. Once the listener had made their selection a new trial began. The target sentences were presented at 7 different signal-

[15] No listener reported being able to hear any words in this masker after this process.

to-noise ratios (SNRs) which covered a 12 dB range with 2 dB steps. The 7 presentation levels were chosen individually for each listener on the basis of thresholds attained in a pre-test.

The experiment was conducted over two sessions. In each session listeners completed 4 blocks of trials. Each of the four blocks contained 105 trials (15 trials at each of the 7 levels) presented in a random order. Within each block the masker condition was fixed and over the two sessions listeners completed 2 blocks of each of the masker conditions. Block order was counterbalanced across the two sessions, with the first 11 listeners completing conditions IVS and UVS on their first visit and the remaining 10 listeners completing conditions S and N on their first visit.

On each visit listeners completed two short practice blocks before testing. During the first practice block target sentences were heard in the absence of any masking sound. During the second practice block a masker was introduced and the target sentence was heard at a SNR of 10 dB. Before testing was carried out in the IVS condition, and to ensure that these maskers were in fact intelligible, listeners also completed a short session of vocoded speech training. This training involved the participants listening to, and repeating back, vocoded IEEE sentences. As feedback has been shown to accelerate understanding of vocoded speech (Hervais-Adelman, Davis, Johnsrude, & Carlyon, 2008), after making an initial response listeners also heard the sentence in its natural, clear form and then again in its vocoded form. Listeners heard a total of 15 vocoded sentences with feedback. A further 10 vocoded IEEE sentences were then played without feedback and listeners were asked to identify them. All listeners were able to fully identify at least 7 out of 10 of these vocoded sentences successfully.

A short pre-test followed training and preceded the main experimental blocks. The aim of the pre-test was to find an appropriate range of presentation levels for each listener. The pre-test also served as additional practice at the main experimental task. Listeners heard a target sentence in a masker and were asked to identify the colour and number of the target sentence. The sentence was played at a SNR of 9 dB. If they responded correctly, the level was then decreased in 2 dB steps until they made 3 consecutive incorrect responses. The SNRs for the psychometric function were set so that the lowest SNR was 2 dB further down than the lowest point achieved on the pre-test. As unmodulated maskers have been found to give considerably higher thresholds than modulated maskers (Festen & Plomp, 1990) the pre-test was run twice, once with an the unmodulated N masker, to give the SNR range for this condition and once with a modulated masker, to give the SNR range for the S, IVS and UVS conditions.

### 3.1.5 Analysis

A logistic function (see equation 2.1) was fitted to the data. The speech reception threshold (SRT) was defined as that SNR which gave an identification score of 50%. The slope was defined as the value of the differential of the function at the SRT (see equation 2.3). The values were found using the Solver function in Excel with free parameters of slope and SRT, fitting all the data from one listener simultaneously.

## 3.2 Results

Of the 21 older listeners originally tested, four listeners were unable to complete both sessions and provided data for only two of the four masking conditions. These listeners were, therefore, excluded from further analysis. Figure 3.2 shows the individual psychometric functions measured for the remaining 17 older listeners and Figure 3.3 shows the psychometric functions for 7 younger listeners. The symbols show the data for the four different masking conditions: IVS (open circles), UVS (closed squares), S (asterisks), and N (open triangles). The solid lines show the logistic curves fitted to the data. To allow easier comparison across listeners, the data has been plotted as a function of “relative SNR”, given by the difference between the actual SNR and the SRT for the N condition (these SRTs are shown in Table 3.1). Overall the results showed that while there was some variation across listeners, the functions for the N conditions were steeper and located to the right of the functions for the UVS, IVS or S.

**_Figure 3.2:_** _Each panel contains the individual psychometric functions for each older listener measured in the four different maskers: the IVS masker (open circles), the UVS masker (closed squares), the S masker (asterisks), and the N masker (open triangles). The solid lines are the fitted logistic curves._

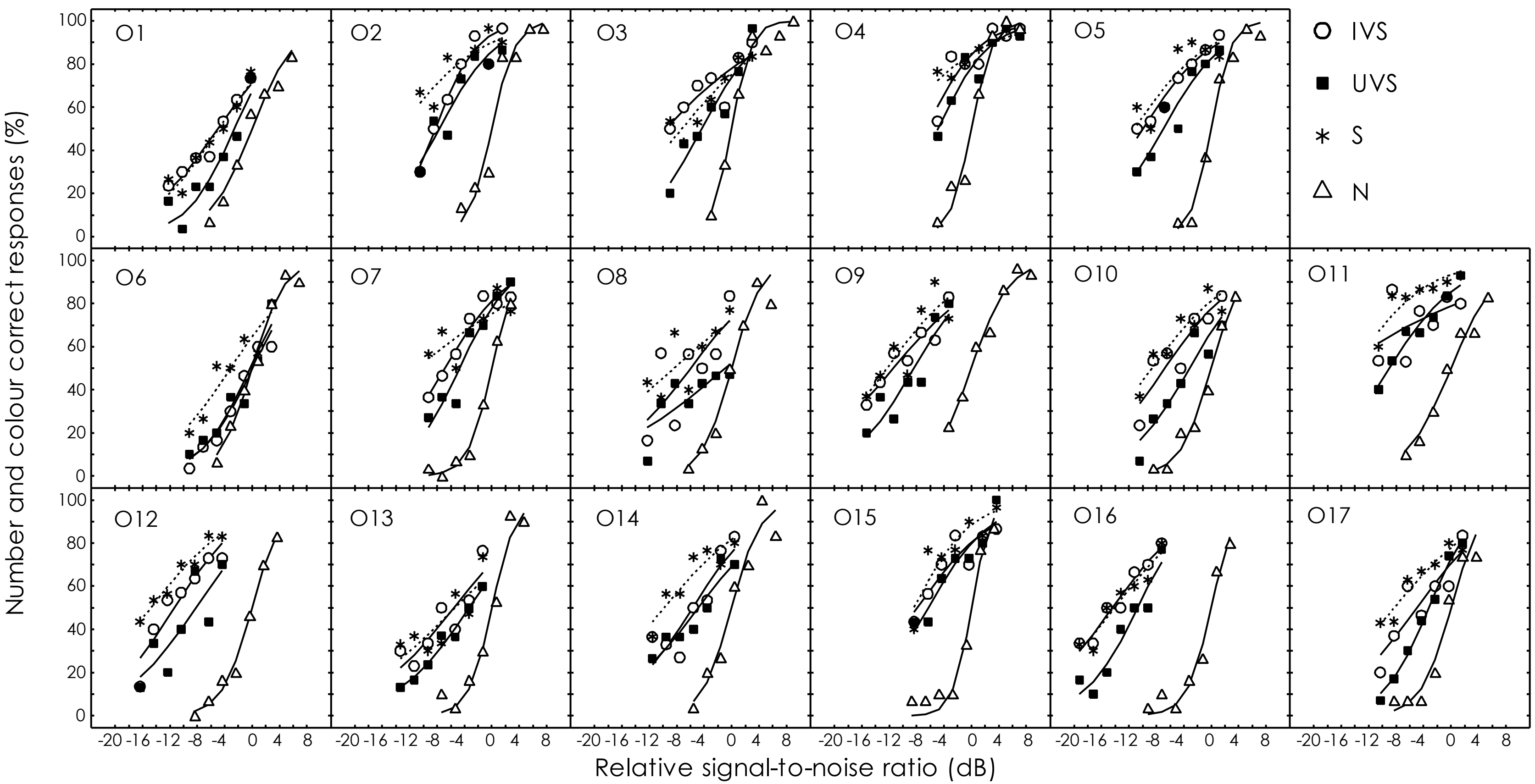


*__Figure 3.3:__ Each panel contains the individual psychometric functions for each younger listener measured in the four different maskers: the IVS masker (open circles), the UVS masker (closed squares), the S masker (asterisks), and the N masker (open triangles). The solid lines are the fitted logistic curves.*

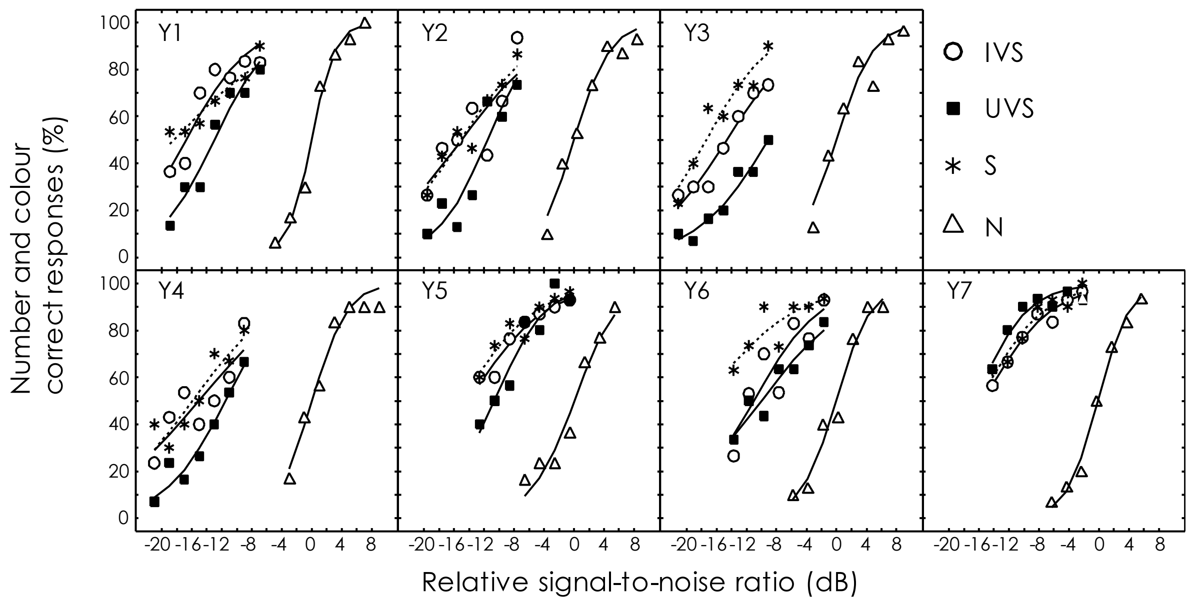


***Table 3.1****: Speech reception thresholds (SRT) measured in each of the four masking conditions for both older (O) and younger (Y) listeners. Listeners of each age group have been sorted by their SRT in the S masker condition. The table also contains listeners' ages and better ear average (BEA) hearing levels.*

| | | | SRT (dB) | | | |
|---|---|---|---|---|---|---|
| Listener number | Age | Better ear average | S | IVS | UVS | N |
| O1 | 73 | 38.8 | -12.0 | - 8.8 | - 9.1 | - 8.9 |
| O2 | 67 | 38.8 | -12.0 | -12.2 | -10.0 | - 6.7 |
| O3 | 69 | 42.5 | -12.7 | -12.8 | -10.4 | - 7.8 |
| O4 | 73 | 37.5 | -16.1 | -13.7 | - 8.6 | - 7.7 |
| O5 | 62 | 17.5 | -16.3 | -18.6 | -13.4 | - 9.0 |
| O6 | 59 | 32.5 | -18.0 | -14.6 | -12.9 | - 9.7 |
| O7 | 69 | 32.5 | -18.0 | -17.6 | -16.0 | - 9.3 |
| O8 | 56 | 35.0 | -18.2 | -16.3 | -12.7 | - 9.5 |
| O9 | 70 | 32.5 | -18.3 | -14.1 | -13.3 | - 8.4 |
| O10 | 71 | 36.3 | -19.7 | -15.6 | -13.5 | - 8.8 |
| O11 | 70 | 23.8 | -20.4 | -19.1 | -16.1 | - 9.2 |
| O12 | 72 | 28.8 | -20.4 | -15.5 | -13.9 | - 9.1 |
| O13 | 70 | 38.8 | -21.8 | -24.0 | -16.4 | - 7.4 |
| O14 | 72 | 22.5 | -23.0 | -17.8 | -17.2 | - 9.5 |
| O15 | 69 | 3.8 | -23.0 | -22.1 | -19.0 | -10.7 |
| O16 | 72 | 25.0 | -24.4 | -20.9 | -17.8 | - 9.6 |
| O17 | 63 | 18.8 | -26.9 | -25.2 | -21.5 | -10.8 |
| Mean | | | -18.9<br>SD = 4.3 | -17.0<br>SD = 4.4 | -14.2<br>SD = 3.6 | - 8.9<br>SD =1.1 |
| Y1 | 26 | 7.5 | -25.5 | -25.4 | -22.0 | -10.4 |
| Y2 | 28 | 7.5 | -26.9 | -26.1 | -22.5 | -11.0 |
| Y3 | 24 | 5.2 | -27.7 | -27.2 | -28.2 | -11.8 |
| Y4 | 38 | 6.5 | -28.3 | -25.5 | -20.0 | -10.9 |
| Y5 | 21 | 10 | -28.4 | -27.4 | -24.1 | -13.4 |
| Y6 | 25 | -5 | -30.0 | -23.3 | -22.3 | -12.3 |
| Y7 | 26 | -1.2 | -31.5 | -29.9 | -26.1 | -13.1 |
| Mean | | | -28.9<br>SD = 2.0 | -26.4<br>SD = 2.0 | -23.6<br>SD = 2.8 | - 11.8<br>SD =1.2 |

### 3.2.1 Speech reception thresholds

Speech reception thresholds (SRTs) are reported in Table 3.1. These values represent the SNR at which the listener identified *both* the colour and the number of the target sentence on 50% of presentations in each of the four masking conditions. Overall the results showed that for both older and younger listeners SRTs were highest when measured in the N condition, intermediate in the UVS and IVS conditions and lowest in the S condition.

A two-way mixed ANOVA was carried out to compare the mean SRTs measured in each of the four masking conditions by both older and younger listeners. Table 3.2 summarises the results from the ANOVA. A significant effect of age group was found demonstrating that, regardless of masker condition, the older listeners had higher SRTs than the younger listeners.  A significant effect of masker condition was also found. Planned simple contrasts were carried out to compare the mean SRTs measured in the IVS condition to those measured in each of the other three other conditions. SRTs measured in the IVS masker were found to be significantly lower than those measured in either the UVS masker or N masker and significantly higher than those measured in the S masker. A significant interaction between age group and masker condition was also found. Post hoc analysis revealed that older listeners SRTs were significantly higher than those of the younger listeners in all masker conditions except the N masker.

***Table 3.2:*** *Summarises the results of the two-way mixed ANOVA on SRTs.*

| Effect | Results |
|---|---|
| Main effect of age group on SRT | $F(1,22) = 38.7$, $P < 0.001$, Partial $\eta^2 = 0.64$ |
| Main effect of masker condition on SRT | $F(3,66) = 185.9$, $P < 0.001$, Partial $\eta^2 = 0.89$ |
| Planned contrasts - comparing SRTs for IVS maskers to those given by: | |
| UVS maskers | $F(1,22) = 39.0$, $P < 0.001$ |
| N maskers | $F(1,22) = 209.5$, $P < 0.001$ |
| S maskers | $F(1,22) = 14.1$, $P < 0.01$ |
| Age group X masker condition interaction | $F(3,66) = 14.5$, $P < 0.001$, Partial $\eta^2 = 0.40$ |

Figure 3.4 displays scatter plots of the SRT measured in each of the four masking conditions for older listeners (open circles) and younger listeners (open triangles). Correlations were carried out between the SRTs measured in the IVS and UVS maskers (top-left panel), the IVS and S maskers (top-right panel), the UVS and S masker (bottom-left panel) and the S and N maskers (bottom-right panel). It is clear that while younger listeners tended to have lower SRTs than the older listeners in each of the four maskers, their data points follow the same basic pattern and data from both groups were, therefore, included in each correlation. All correlations were found to be significant, with r values of at least 0.86 (individual r values for each correlation are displayed in figure 3.4 along with regression lines calculated using the Deming method[16]). These significant correlations indicate that the listeners with the lowest SRTs in one masking condition were also the listeners with the lowest SRTs in the other masking conditions. Each panel also contains a dotted line representing 1:1, i.e. where SRT for the two maskers is equal. A binomial analysis of the data was carried out to confirm that data points were not equally distributed about this line (i.e. to confirm that all listeners were consistently achieving higher thresholds in the same maskers). The results showed that a significant proportion of data points were above the line for the IVS vs. UVS condition, ($p < 0.001$), thus indicating listeners consistently achieved higher threshold in the UVS than the IVS maskers. The results also showed that data points were significantly below the 1:1 line for the IVS vs. S condition ($p < 0.01$), the UVS vs. S condition ($p < 0.001$) and the N vs. S condition ($p < 0.001$) showing that thresholds were higher in the IVS, UVS and N conditions respectively. The results of the binomial tests conform to the mean results presented in table 3.1, and indicate this pattern of results holds at the individual listener level as well as at the mean level.

Figure 3.5 shows scatter plots for SRT as a function of better ear average for each of the four masking conditions. Significant correlations were found between listeners'

[16] Regression usually assumes that only Y measurements are associated with random measurement errors. Deming methods take measurement errors for both axes into account. As measurements on both axes are slope values and, therefore, both subject to measurement errors the deming method was selected here for fitting regression lines (Cornbleet & Gochman, 1979). Deming regression minimizes the sum of squares of the perpendicular distances of the points from the line.

***Figure 3.4:*** *Individual scatter plots of SRT in each of the four masking conditions for both older listeners (open circles) and younger listeners (open triangles). The solid line represents the regression line calculated using the Deming regression method, R2 values are included (* indicates p < 0.05, ** indicates p < 0.01).*

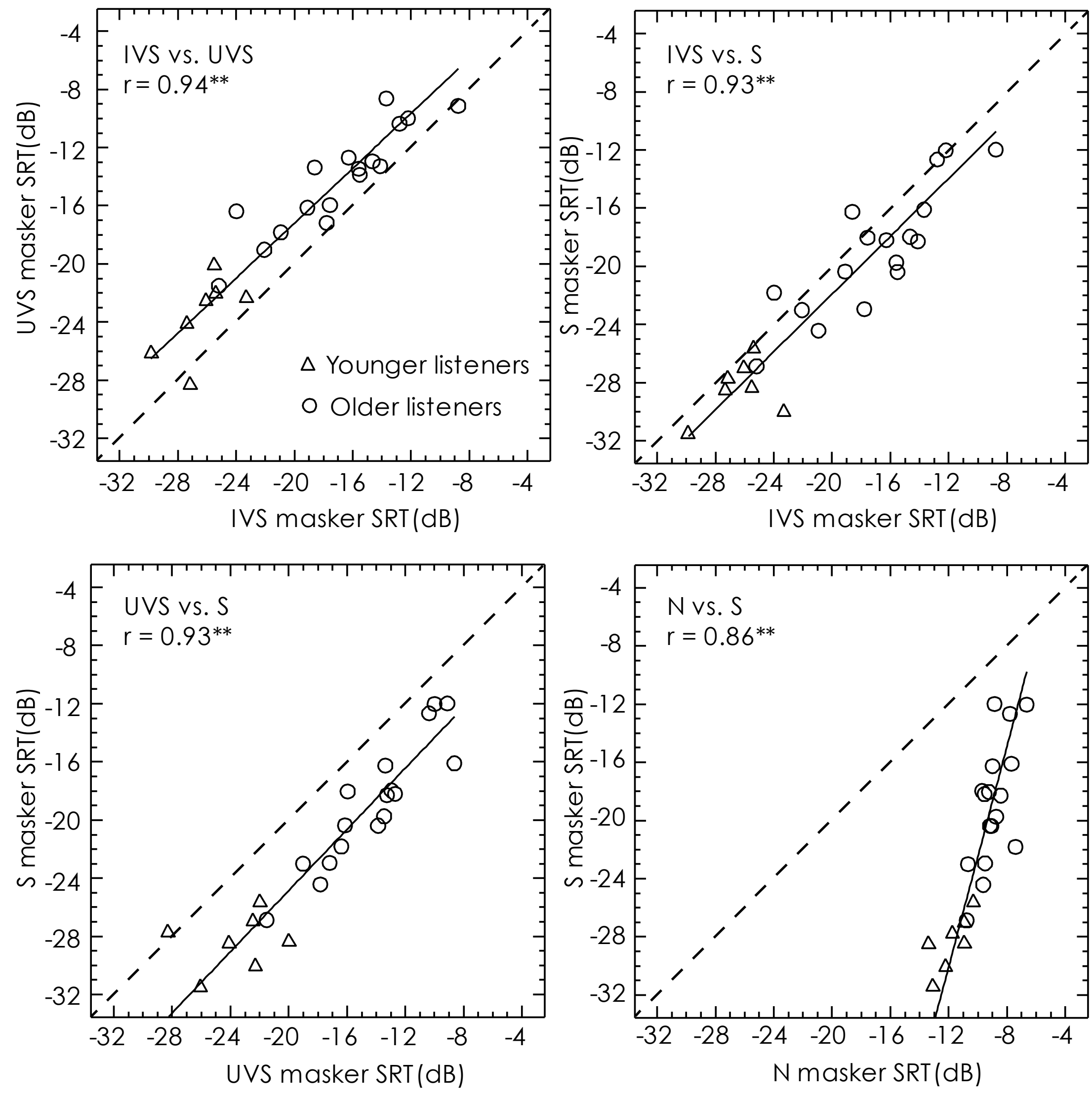
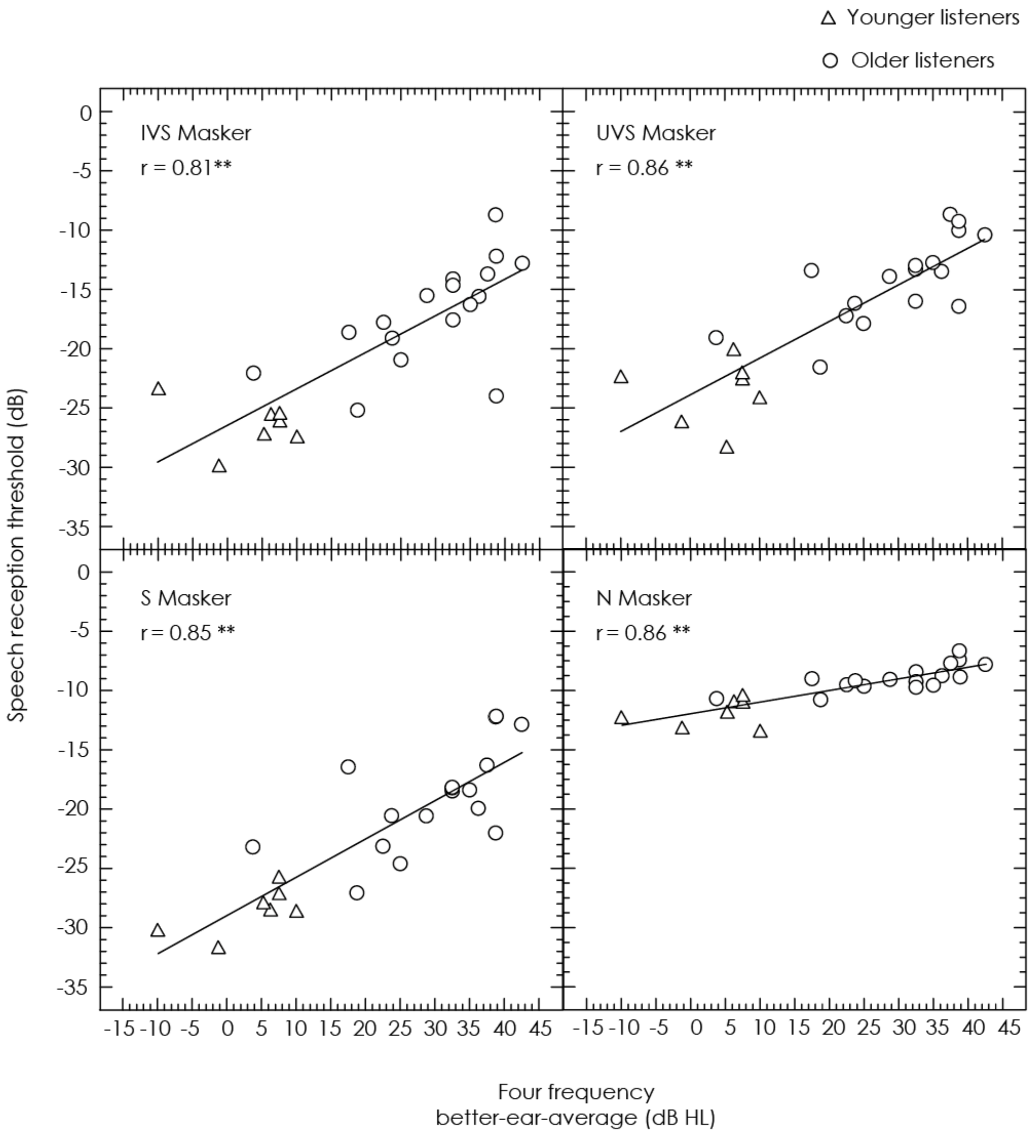


***Figure 3.5:*** *The four panels show older (circles) and younger (triangles) listeners speech reception thresholds (dB) in each of the four masking conditions, plotted against their better ear average (dB HL). Pearsons's correlation coefficients (r) are included (* indicates p < 0.05, ** indicates p < 0.01).*



SRTs in each of the four masking conditions and their degree of hearing loss, indicating that SRTs increased with increased hearing loss. The r values are reported on each panel: and all were at least 0.81.

### 3.2.2 Slopes

Figure 3.6 presents a summary of the slope data for older and younger listeners. For both listener groups mean slopes measured in the N condition were clearly steeper (older, M = 12.9%/dB, SD = 3.32 and younger, M = 11.3%/dB, SD = 2.3) than those measured in the IVS masker (older, M = 4.1%/dB, SD = 1.5 and younger M = 5.0%/dB, SD = 0.8), UVS masker (older, M = 5.7%/dB, SD = 1.3 and younger, M = 6.4%/dB, SD= 1.2), or S masker (older, M = 4.1% /dB, SD = 0.9 and younger, M = 4.8%/dB, SD = 1.0). Differences between the other conditions were, in comparison, relatively small.

A two-way mixed ANOVA was carried out on this slope data. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(5) = 25.4$, $p < 0.001$), and so the degrees of freedom were corrected using Greenhouse-Geisser estimates ($\varepsilon = 0.55$). The effect of masker type was statistically significant (results are summarised in Table 3.3). Planned simple contrasts were carried out to compare the mean slope measured in the IVS condition to each of the other three other conditions. Significant differences in mean slope were found between the two experimental conditions (the IVS and the UVS maskers) and the IVS and N maskers. However, no significant difference in mean slope was seen for the IVS and S maskers. The main effect of age group was not found to be significant nor was the interaction between masker condition and age group. The results demonstrate that psychometric function slopes were shallowest when measured in either the S or IVS maskers, slightly steeper in the UVS masker, and then steepest in the N masker.

***Figure 3.6:*** *Mean slopes of psychometric functions measured in the IVS, UVS, S, and N masking conditions for older (clear bars) and younger (striped bars). Error bars display 95% confidence intervals.*

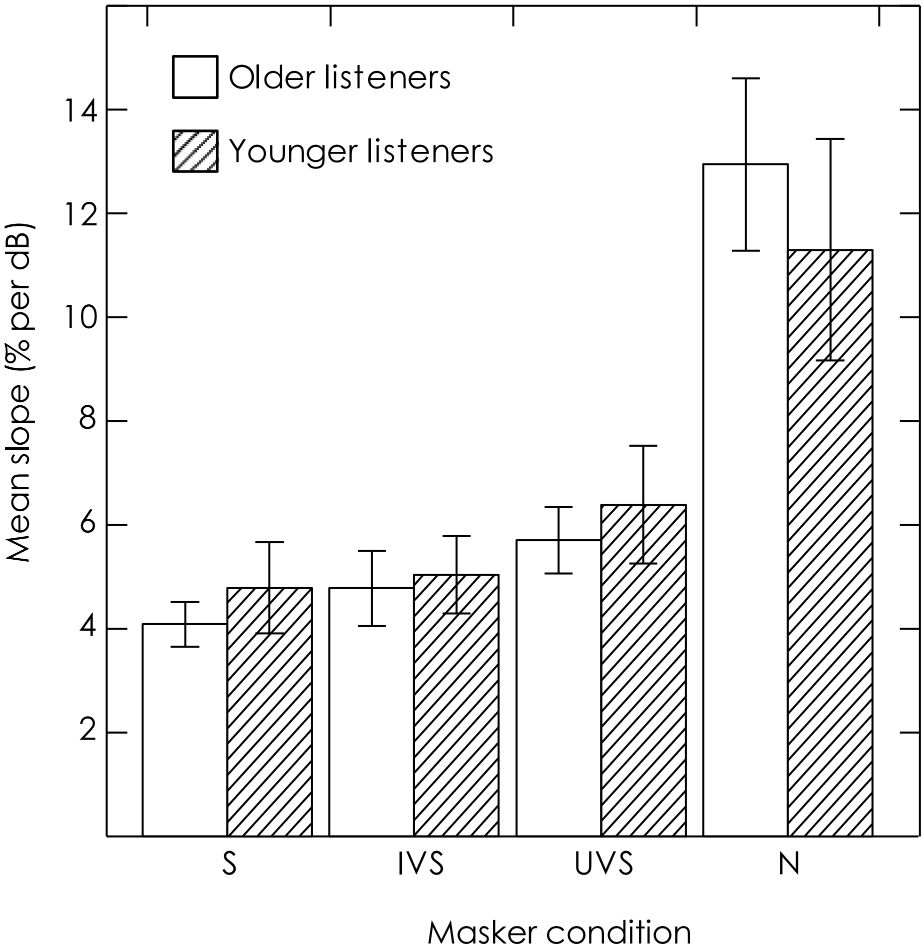


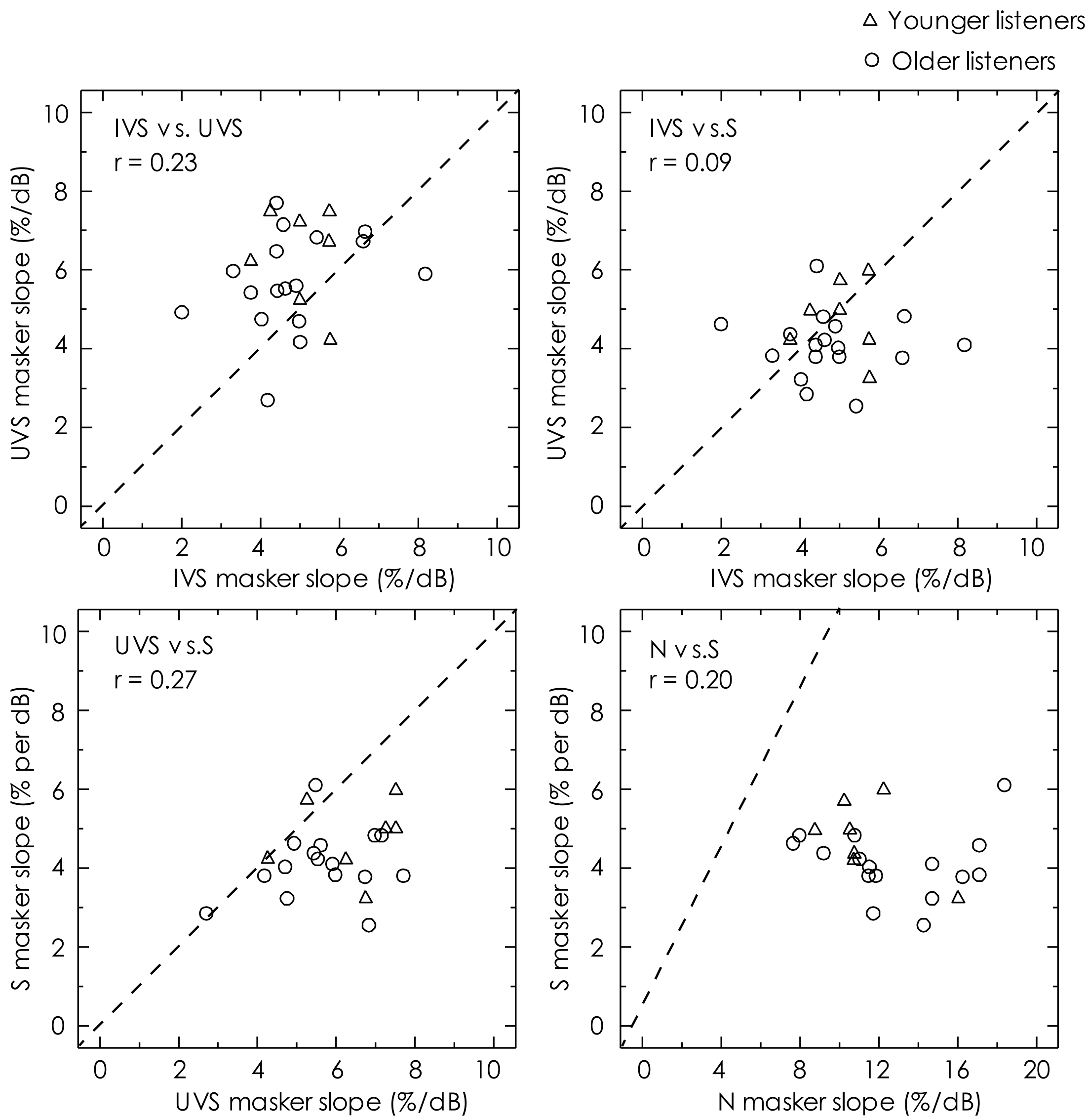
***Table 3.3:*** *Summarises the results of the two-way mixed ANOVA carried out on the slope data given by the four different masking conditions.*

| Effect | Results |
|---|---|
| Main effect of age group on slope | $F(1,22) = 0.01$, $P = 0.99$, Partial $\eta^2 = 0.01$ |
| Main effect of masker condition on slope | •$F(1.65, 36.5) = 85.0$, $P < 0.001$, Partial $\eta^2 = 0.79$ |
| Planned contrasts - comparing slope for IVS maskers to those given by: | |
| UVS maskers | $F(1,22) = 10.4$, $P < 0.01$ |
| N maskers | $F(1,22) = 124.1$, $P < 0.001$ |
| S maskers | $F(1,22) = 1.8$, $P = 0.19$ |
| Age group X masker condition interaction | •$F(1.65.36.4) = 2.02$, $P = 0.36$, Partial $\eta^2 = 0.08$ |

• Degrees of freedom corrected using Greenhouse-Geisser estimates ($\varepsilon = 0.55$).

Figure 3.7 shows scatter plots of the slopes measured in each masking condition for older and younger listeners. As was indicated by the results of the ANOVA, the figure shows no clear distinction between the slopes values produced by older or younger listeners. Correlations were carried out between the IVS and UVS maskers (top-left panel), the IVS and S masker (top-right panel), the UVS and S maskers (bottom-left panel) and the S and N maskers (bottom-right panel). R values were found to be 0.23 or less and no correlation reached significance (individual r values for each correlation are displayed in Figure 3.7). These results indicate that listeners with the steepest psychometric functions in one masking condition did not necessarily have the steepest functions in the other masking conditions. Each scatter plot contains a dotted line representing 1:1 and again binomial tests were carried out to examine whether statistically significant proportions of data points were above or below this line. The results showed that a significant proportion of data points were above the line for the IVS vs. UVS condition, ($p < 0.01$), indicating listeners consistently produced steeper slopes in the UVS than in the IVS maskers. The results

***Figure 3.7:*** *Individual scatter plots comparing slope measurements made in each of the four masker conditions for older (open circles) and younger (asterisks) listeners. The dotted line represents 1:1, i.e. where the slopes are equal. The r value for the data in each panel is also included.*



also showed that data points were significantly below the 1:1 line for the UVS vs. S condition ($p < 0.01$) and the N vs. S condition ($p < 0.001$) showing that most listeners produced steeper slopes in the UVS and N conditions than in the S condition. For the IVS vs. the S condition neither the proportions of data points above or below the line were found significant thus indicating no specific pattern across listeners. Again the results from the binomial tests conform with the mean results already presented (Figure 3.6) and indicate that despite the lack of an overall correlation in slope measurements across conditions, the individual listener data is still consistent with the mean results.

Figure 3.8 shows scatter plots for slope measurements plotted as a function of better ear average for each of the four masking conditions. The correlations between slope and the degree of hearing loss were found to be non-significant all four masking conditions, with r values of around 0.07 or less. The r values are reported on each panel.

### 3.2.3 Accuracy of logistic fits

An assessment of the "goodness of fit" of these functions was also carried out. The RMS value was calculated for each function. Figure 3.9 shows the distribution of the RMS values for psychometric functions measured in each of the four masker conditions. A value of 0% indicates a perfect fit between the raw data and the fitted logistic. The figure shows that while mean RMS values were relatively similar in the four conditions, the distributions of fits varied slightly. The distribution was the narrowest for the N masker and this masker also had the lowest mean RMS value, suggesting then that the model was a good fit of the data for this condition. The IVS masker gave the widest distribution of RMS values suggesting the logistic model was not a consistently good fit to the data in this condition. Greater variability of listeners' responses for this condition across SNR would account for this variability of fit and it is possible that an increased number of trials at each SNR are needed to achieve stable (smoother) psychometric functions.

***Figure 3.8:*** *The four panels show older (circles) and younger (triangles) listeners' average slope scores (% per dB) in each of the four masking conditions, plotted against their better ear average (dB HL). Pearson's correlation coefficients (r) are included. All correlations were found to be insignificant.*

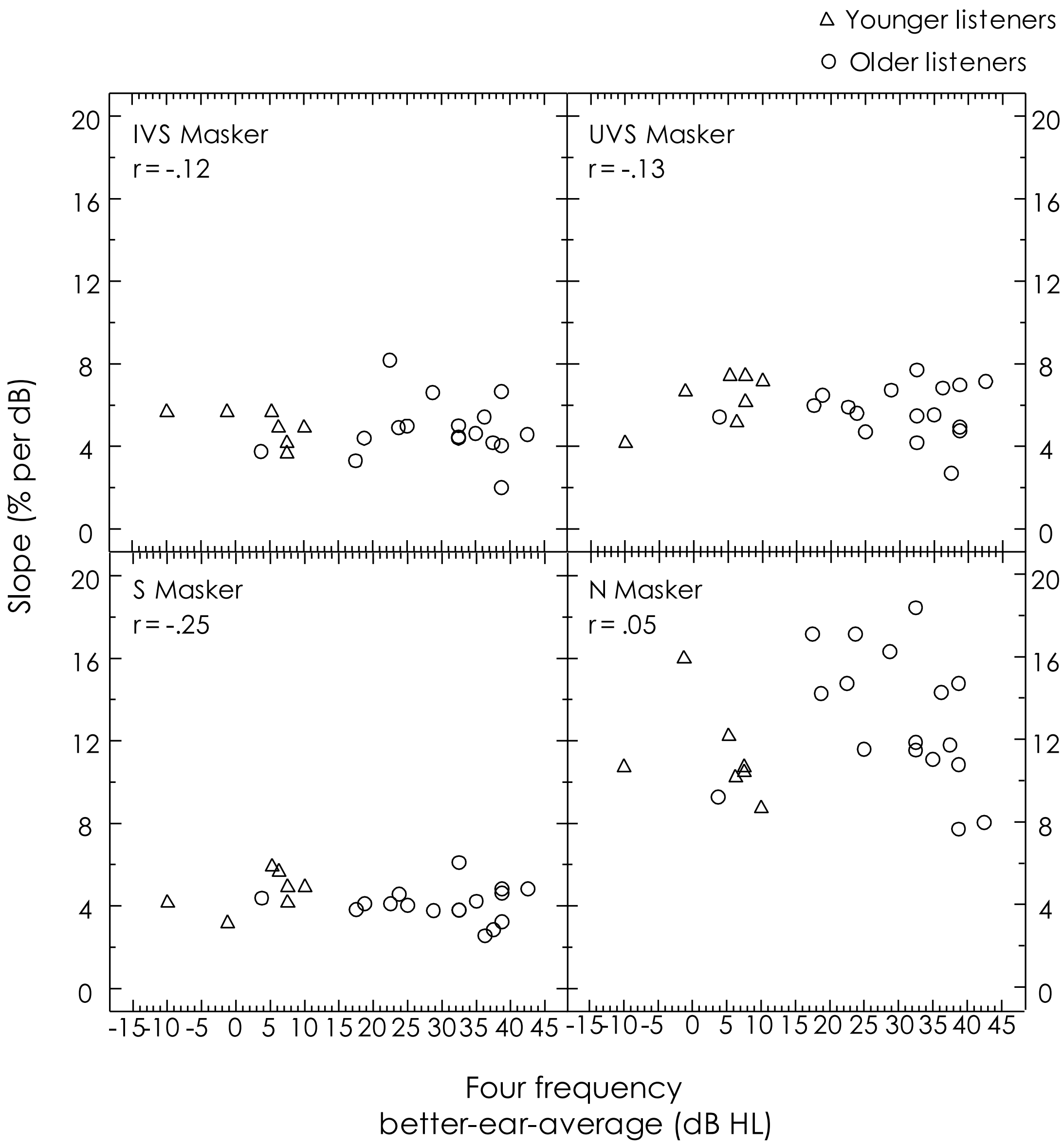
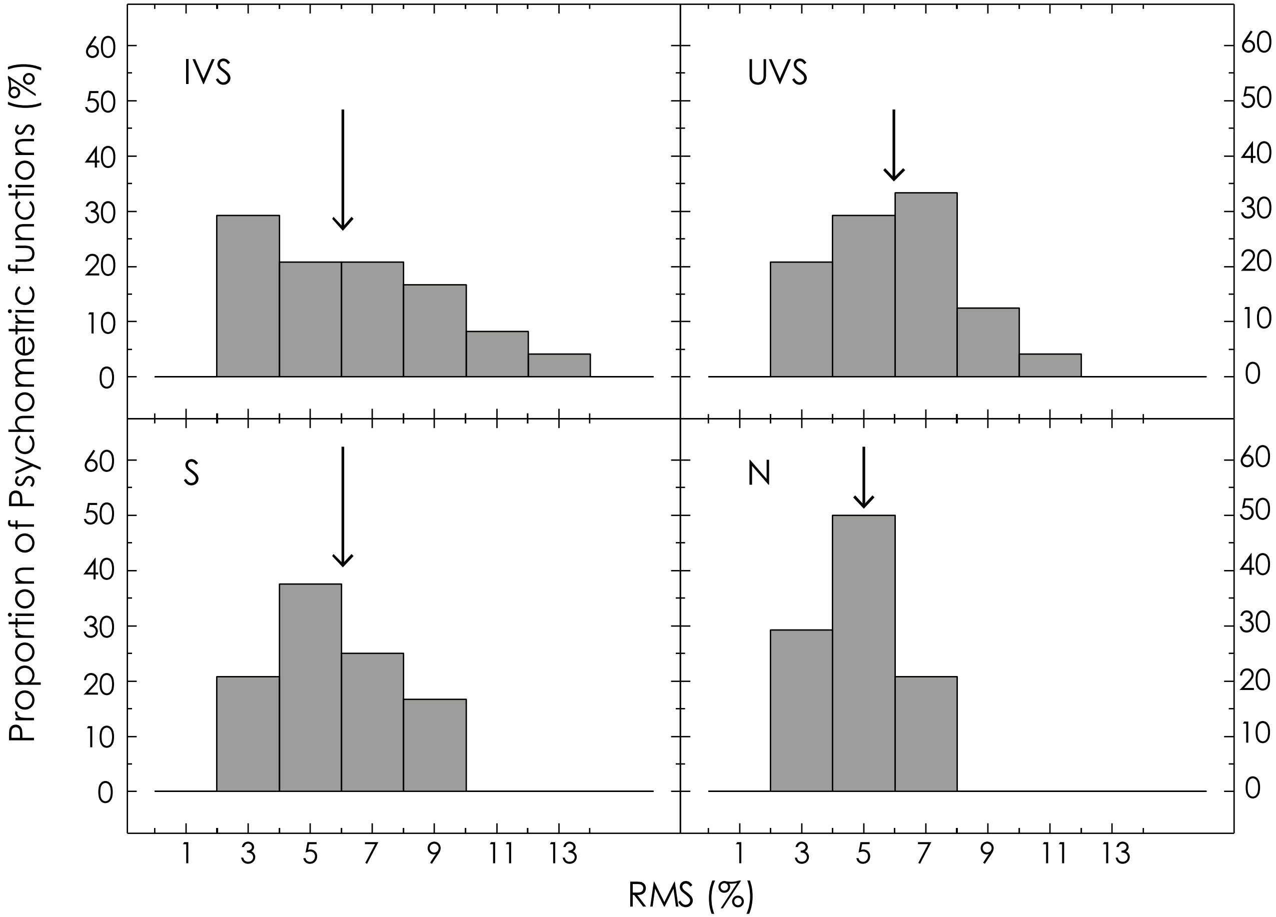


***Figure 3.9:*** *The four panels show the distribution of RMS values calculated to assess the degree of fit between logistic functions and the raw psychometric function data in each of the four masking conditions. The arrows indicate the mean RMS value for each condition.*



### 3.2.4 Summary of results

- The static-noise masker gave significantly steeper slopes than the remaining three masker types.
- A small but significant slope difference was found between the functions given by the intelligible and unintelligible maskers.
- There was no effect of age on the slope of the psychometric function.
- SRTs for the intelligible masker were lower than those given by either the static noise or the unintelligible masker but higher than those given by the speech masker.
- Older listeners SRTs were significantly higher than those of young listeners in all masker conditions except the static-noise.

## 3.3 Discussion

Psychometric functions were measured for a group of older listeners and for a group of younger listeners in four masking conditions: same talker speech (S), intelligible vocoded speech (IVS), unintelligible vocoded speech (UVS), and static noise (N). The S and N maskers were used to provide extremes of both slope and threshold. It was expected that if linguistic confusion gives rise to shallow slopes, psychometric functions would be shallower when measured in the IVS than the UVS masker. While an initial review of the data seem to support this hypothesis there is an indication that this interpretation may not be straightforward. Both speech reception threshold and slope effects are discussed in detail in the following sections.

### 3.3.1 Speech reception thresholds

For both older and younger listeners, significant differences were found between the SRTs measured in each of the four maskers. Thresholds were substantially higher in the N condition than they were in the UVS condition, which were in turn higher than those measured in the IVS masker. A small, but significant, difference was also found between SRTs measured in the IVS and S maskers. This pattern of results demonstrates that speech identification was harder in the N condition than it was in the UVS, IVS or S conditions.

Previous studies have indicated that thresholds tend to be higher in static or steady noises than they are in modulated noises (Bacon, Opie, & Montoya, 1998; Festen & Plomp, 1990). The current results support this and are consistent with the idea of dip listening (Miller, 1947). The interpretation is that thresholds were lower in the IVS, UVS and S conditions as these maskers contained modulations of both amplitude and frequency. Listeners were able to use brief dips in the masker to catch glimpses of the target speech. These glimpses were often enough to allow target speech to be reconstructed and so gave successful speech identification at lower SNRs than could be achieved with a non-modulated masker.

Older listeners were found to have higher speech reception thresholds than the younger listeners in three of the four masking conditions, highlighting the increased difficulty older listeners have understanding speech in background sounds. This difference in performance was found, however, to be greater for the modulated maskers (UVS, IVS and S maskers) than for the non modulated masker (the N masker). This result is also consistent with previous findings as it has been reported that older listeners are not as able to profit from dips in the masker to improve speech intelligibility as younger listeners are (Dubno, Horwitz, & Ahlstrom, 2002).

### 3.3.2 Slopes

The slopes of the psychometric function were also found to be affected by the type of masker used. Slopes measured in the N masker were considerably steeper than those measured in the other three maskers. Slopes were also found to be significantly steeper when the masker was unintelligible (in the UVS masker) than when it was intelligible (in the IVS masker). No difference in slope was found, however, between the IVS and S conditions. This pattern of results was found for both older and younger listeners, with no significant difference in the magnitudes of the slopes seen between the two age groups. Steeper slopes found for the static noise masker (N) than the speech maskers (S or IVS) supports the finding from the slope survey that noise maskers tend to give steeper slopes than speech maskers. The static-noise masker gave even steeper slopes (older, M = 12.9% per dB and younger, M = 11.3% per dB) than were found in the slope survey for equivalent maskers (8% per dB), and the speech masker (older, M = 4.1% per dB and younger, M = 4.8%/dB) gave slopes of very similar magnitude to those identified in the survey (4% per dB). A slope change that was expected but that did not reach significance in the slope survey − between static noise and modulated noise − was also found in the current study. The slope was significantly steeper in the N condition, for example, than the UVS conditions. This supports the role that amplitude modulations in the masker have on the slope of the psychometric function (Rhebergen & Versfeld, 2005). The noted slope difference between the IVS and UVS maskers supports the predicted outcome for an effect of masker intelligibility. It can be argued then, that despite the IVS masker sounding acoustically distinct from the target speech, the lexical and phonetic information it contained were available to interfere with the identification of the target (Hoen et al., 2007; Tun et al., 2002). The precise mechanism for this interference and, presumably the resultant target/masker confusion, however, is still unclear. For example, intelligible maskers may put extra burden on working memory and disrupt cognitive processes required for the successful identification of the target (e.g. Francis, 2010; Li et al., 2004). Alternatively, intelligible maskers may simply distract attention away from target, making it harder to select and attend to (Yost, 2006).

Despite findings supporting our initial hypotheses, caution must be taken before concluding that linguistic similarity has an effect on the slope of the psychometric function. There were several other indications which suggest that this interpretation may not be sound. Firstly, the slope survey in chapter 2, gave a strong indication that target/masker confusion due to the linguistic content of speech would produce a weaker slope effect than confusion due to acoustic similarity. The finding that slopes did not significantly differ in the natural intelligible speech (S) condition and the vocoded intelligible speech (IVS) condition was, therefore, surprising as it was presumed that when the masker could interfere both acoustically and linguistically with the target, slopes would be shallower than when they could only interfere on a linguistic level. This gives a first indication that the changes in slope seen in the current experiment may not necessarily be, as was intended, the result of changes in the degree of confusion between the target and the masker.

Secondly, many of the studies which have attributed slope changes to some form of confusion (either acoustic or linguistic) have reported unusual shaped psychometric functions (flat or U-shaped) and have noted these changes in performance at SNRs between -10 and 0 dB (e.g. Brungart, 2001a; Dirks & Bower, 1969). While slopes were shallow in the current study, they were still monotonic, even for the natural, intelligible speech-masker. Functions in the current study were also measured at considerably lower SNRs (mean SRTs were -14.8 dB and -22.5 dB for older and younger listeners respectively) as listeners performance was at ceiling for the presentation levels around 0 dB. It has been argued that the reason psychometric functions become shallow when the target and masker are highly confusable is that the listener has to rely on level differences to disentangle the competing signals (e.g. Arbogast et al., 2002; Drullman & Bronkhorst, 2004). Such an explanation can explain the occurrence of shallow slopes at SNRs around 0 dB, but at low SNRs the level difference between the target and masker offers a clear cue for disentanglement. Shallow psychometric functions at the levels measured in the current study are, therefore, unlikely to have arisen due to confusion of the target and masker.

Thirdly, slopes were not found to be affected by age. Several studies have shown that older listeners find it harder to inhibit the processing of competing stimuli than younger listeners do and are further hindered when the competing stimuli contains meaningful content (Rossi-Katz & Arehart, 2009; Tun et al., 2002). If the shallow slopes noted in the current experiment were down to the linguistic interference from a speech masker and older listeners are prone to greater interference, then it might be expected that the difference seen between the slopes measured in the IVS masker and those measured in the UVS masker would be greater for older listeners than for younger listeners. No significant age differences were found however, again suggesting that an alternative interpretation of the data may be needed.

Significant slope differences were measured in the current experiment between the intelligible and unintelligible maskers. If these were not the result of the intended manipulation (i.e. the result of confusion between the target and masker due to linguistic similarity) another, unforeseen factor, must have affected slope. It was assumed that there were no material differences in the amplitude modulations of the UVS and IVS maskers. However, if this is not the case and these stimuli do differ in terms of modulation rate or depth, it is possible that these differences were responsible for the differences in slope seen between the UVS and IVS maskers. As previously mentioned, dips in the amplitude of a masker offer opportunities to glimpse target speech. If these dips are deep enough many of these glimpses will continue even as the SNR is lowered. In these cases speech intelligibility will remain relatively stable over small changes in SNR. As modulations in the masker are reduced, however, glimpses of target speech quickly disappear with lowering SNRs. For these maskers small decreases in target level can have relatively large effects on speech intelligibility.

While UVS and IVS maskers were created to be as equivalent as possible there were slight changes in the vocoding process that might have resulted in modulation difference between the two stimuli. The UVS maskers were created by joining channels from six different sentences together. While the modulations *within* a channel are equivalent to those of a single sentence, the modulations *across* channels will not necessarily correspond. This may then have resulted in reduced modulations in the UVS masker. Accordingly, the modulations of the present stimuli were analysed.

### 3.3.3 Amplitude modulation differences in the stimuli

The basic process used to analyse amplitude modulations was as follows; first the sample for analysis was filtered using a gammatone filter and the amplitude envelope extracted using the Hilbert transformation. A 1000 msec section of this envelope was then selected, scaled to a mean level of 1 dB and a fast Fourier transform (FFT) applied to produce the modulation spectrum. Before analysing the stimuli from the current experiment, the process was tested using noises with modulations of known depth and rate.

The modulation analysis was carried out for all four maskers (S, IVS, UVS and N). It had been expected that, due to the vocoding process used to create the maskers, the

IVS and UVS maskers would have modulations equivalent to those of the S maskers *within* discrete channels; however it was plausible that the modulations in the UVS masker would be reduced *across* channels. To compare the modulations that were occurring within a channel to those occurring across channels, the modulation analysis was carried out at eight different centre frequencies; four corresponding to the geometric mean frequencies of channels 4, 6, 8, and 10 of the vocoder filterbank (479 Hz, 998 Hz, 1947 Hz, and 3681 Hz respectively) and four corresponding to the frequencies between channels 4/5, 6/7, 8/9, and 10/11 of the vocoder filterbank (583 Hz, 1187 Hz, 2289 Hz and 4306 Hz respectively). For each gammatone filter value, all 100 sentences for each of the four masking conditions were analysed separately and an average modulation spectrum for each masker condition calculated. In order to put any modulation differences between the masking stimuli into context, this whole process was repeated for 2, 4, 6 and 8-talker babble.

Figure 3.10 displays the mean modulation spectrums for the S, IVS, UVS and N maskers along with the mean modulation spectrums for 2, 4, 6 and 8-talker babble. Each panel shows the depth of modulations at the centre frequency of one channel. While modulation depth is greatly reduced for the N masker, modulation depths in the IVS, UVS and S conditions look to be relatively equivalent. As more talkers are added to the babble the dips in the masker are smoothed out and the modulation spectrum approaches that of the steady noise. The modulation spectrum of the UVS masker, however, more closely resembles those of the IVS and S maskers than it does the two-talker babble, suggesting that modulation depths for these stimuli within a channel are equivalent. Table 3.4 presents a summary of modulations depths averaged over modulations rates of 1- 10 Hz[17]. These averages are displayed for all 8 maskers at each of the four centre frequencies analysed. Again these results show modulation depths to be similar in the IVS, UVS and S maskers.

[17] This range of modulation rates was chosen for convenience and is within the envelope range defined by Rosen (1992). Figures 3.10 and 3.11, however, display amplitude modulation depths over a wider range modulation rates (up to 32 Hz).

***Figure 3.10:*** *The four panels show the average modulation spectrums for the IVS (closed circles), UVS (closed squares), S (closed diamonds), and N maskers (closed triangles), at 479 Hz, 998 Hz, 1947 Hz and 3681 Hz (centre frequencies of bands 4, 6, 8, 10, from vocoder filterbank). The modulation spectrums of 2-talker (open circle), 4-talker (open triangle), 6-talker (open diamond), and 8-talker (open square) are also included.*

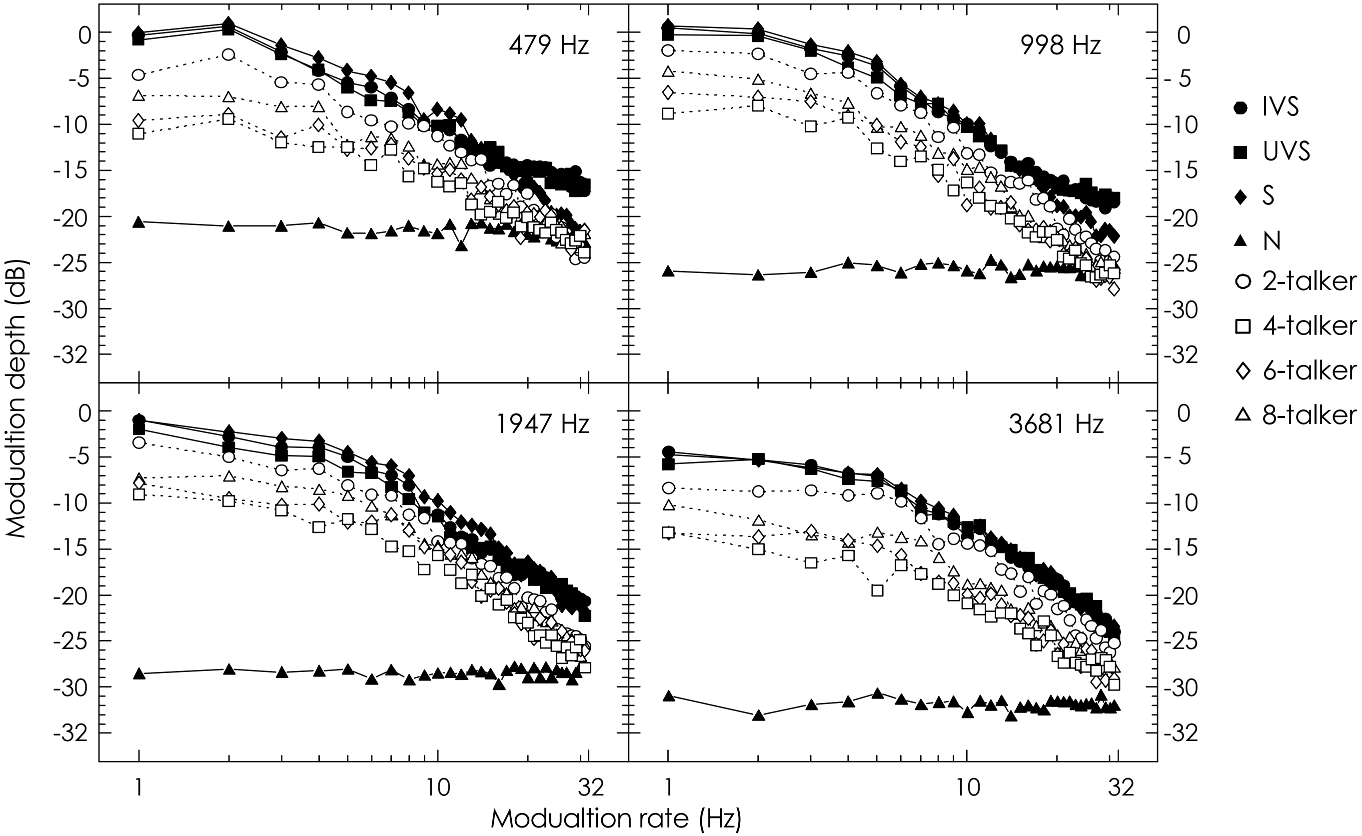


***Table 3.4:*** *The average modulation depth at modulations rates between 1 and 10 Hz, for the IVS, UVS, S and N maskers at 479 Hz, 998 Hz, 1947 Hz and 3681 Hz (centre frequencies of bands 4, 6, 8, and 10 from vocoder filterbank).*

| Masker condition | 479 Hz | 998 Hz | 1947 Hz | 3681 Hz | Mean Hz |
|---|---|---|---|---|---|
| S | -5.9 | -4.5 | - 5.6 | -7.4 | - 5.9 |
| IVS | -7.6 | -5.5 | -6.6 | -7.8 | - 6.9 |
| UVS | -7.9 | -5.9 | -7.5 | -7.9 | - 7.3 |
| 2-talker babble | -9.9 | -7.7 | -9.0 | - 9.8 | - 9.1 |
| 4-talker babble | -12.7 | -10.2 | -10.7 | -13.5 | - 11.8 |
| 6-talker babble | -14.1 | -11.8 | -12.0 | -15.0 | - 13.2 |
| 8-talker babble | -15.1 | -12.9 | -13.4 | -16.6 | - 14.5 |
| N | -23.1 | -25.6 | -28.5 | -30.5 | - 26.9 |

Figure 3.11 again displays the mean modulation spectrum for all eight analysed samples, but this time each panel shows the modulation depths at a frequency between two channels. Again modulation depth is clearly reduced in the N masker, but now a greater difference can also be seen between the IVS and S maskers, and the UVS masker. At these analysis frequencies the depth of modulations in the UVS masker are more similar to those seen in the two-talker babble at some rates than they are to those of a single talker (i.e. the S masker). This suggests that the modulations in the IVS and UVS maskers are not equivalent between channels. Table 3.5 displays the mean modulation depths for maskers at modulation rates between 1-10 Hz. The results further demonstrate the reduction in the amplitude of modulations across channels for the UVS maskers.

***Figure 3.11:*** *The four panels show the average modulation spectrums for the IVS (closed circles), UVS (closed squares), S (closed diamonds), and N maskers (closed triangles), at 583 Hz, 1187 Hz, 2289 Hz and 4306 Hz (frequencies between bands 4/5, 6/7, 8/9, 10/11, from vocoder filterbank). The modulation spectrums of 2-talker (open circle), 4-talker (open triangle), 6-talker (open diamond), and 8-talker (open square) are also included.*

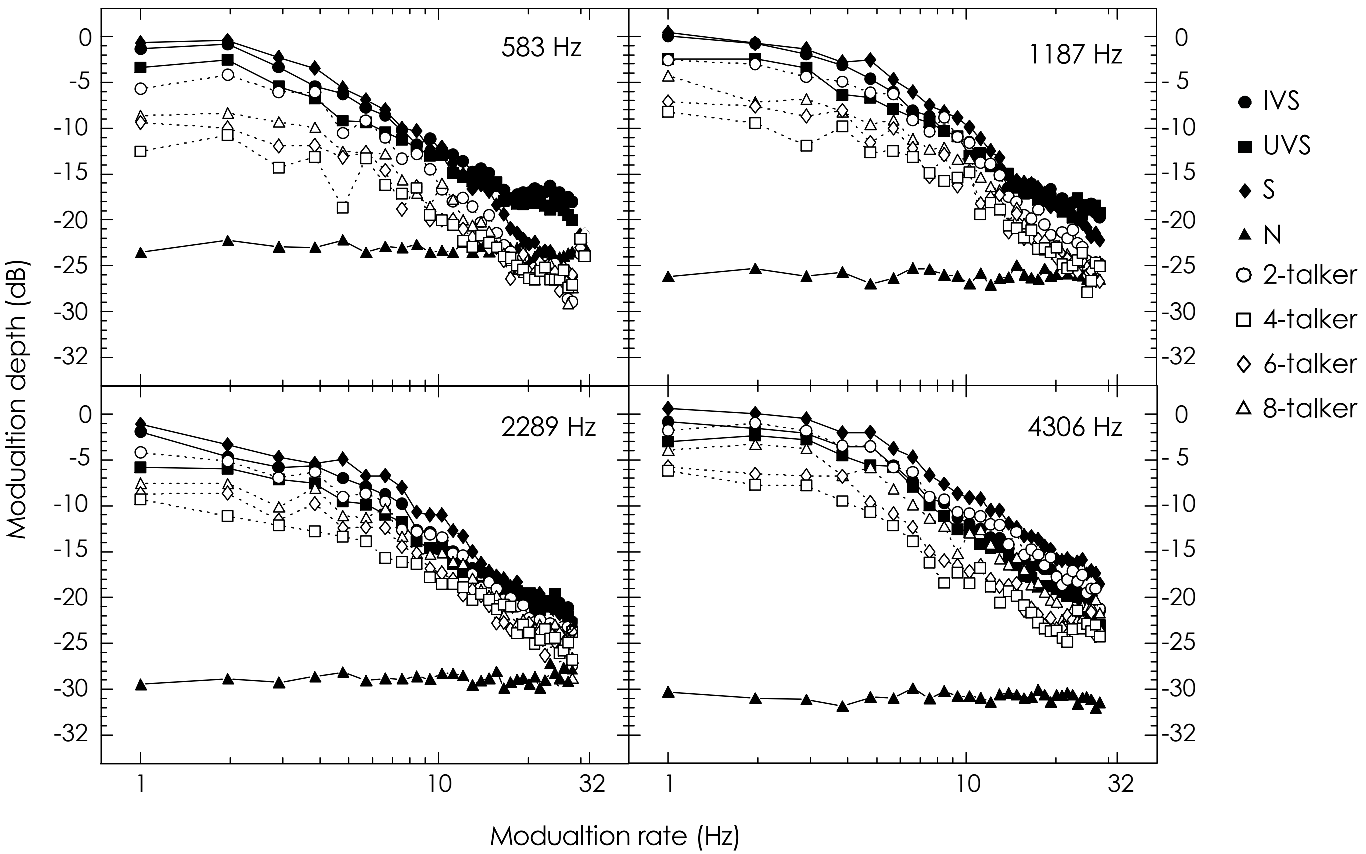


***Table 3.5:*** *The average modulation depth at modulation rates between 1 and 10 Hz, for the IVS, UVS, S and N maskers at 583 Hz, 1187 Hz, 2289 Hz and 4306 Hz (frequencies between bands 4/5, 6/7, 8/9, 10/11, from vocoder filterbank).*

| Masker condition | 583 Hz | 1187 Hz | 2289 Hz | 4306 Hz | Mean Hz |
|---|---|---|---|---|---|
| S | -6.6 | -4.7 | - 6.7 | -3.9 | - 5.5 |
| IVS | -7.3 | -6.0 | -8.2 | -8.9 | - 7.6 |
| UVS | -8.9 | -7.3 | -10.0 | -6.8 | - 8.3 |
| 2-talker babble | -9.9 | -7.1 | -9.2 | - 5.5 | - 7.9 |
| 4-talker babble | -13.0 | -10.0 | -10.8 | -8.4 | -10.6 |
| 6-talker babble | -14.4 | -11.4 | -12.4 | -11.0 | - 12.3 |
| 8-talker babble | -15.5 | -12.8 | -14.2 | -12.5 | - 13.8 |
| N | -22.9 | -25.9 | -28.6 | -30.7 | - 27.0 |

It can be concluded, therefore, that while within individual channels the modulations were equivalent in the UVS and IVS maskers, the modulations across channels were reduced in the UVS maskers. Comodulated glimpses across frequency bands, such as those available in the IVS masker, have been found to be more important than uncomodulated ones, such as those in the UVS masker (Howard-Jones and Rosen 1993). The uncomodulation across channels in the UVS masker more closely resembled modulations of a two-talker masker. It was demonstrated in the slope survey that increasing the number of talkers from 1 to 2 also increased the slope of the psychometric function by, on average 4%. It is plausible, therefore, that the increased slope in the UVS masker compared to the IVS was due to this modulation difference. Overall, slope changes between the IVS and UVS maskers explained as the result of modulation differences, rather than the result of changes in target/masker confusion, provides a better interpretation of the current data.

### 3.3.4 A bootstrap analysis of the psychometric functions.

The rationale of the current experiment (and the other experiments reported in this thesis) depends on being able to compare slope (and occasionally SRT) differences across conditions. This relies on being able to accurately fit the experimental data with a mathematical function. All psychometric functions measured in the current study were fitted with a logistic function in which two parameters – slope and SRT – were set to minimize the RMS difference between the model and the experimental data. To be confident in the values of each parameter, it is important to assess the variability of the fitted parameters (Wichmann & Hill, 2001a).

One method for calculating error estimates for the parameters of slope and SRT is the "bootstrap" (Efron, 1979). The bootstrap is a Monte Carlo resampling technique that generates a large number of trials to attain accuracy. The essence of the bootstrap is that a large number of synthetic data sets ("bootstrap samples") are generated. Each bootstrap sample has *n* elements created by resampling *with replacement* from the original data set *n* times. The parameter of interest is then calculated for each bootstrap sample and the process is then repeated a large number of times. By creating a histogram of the distributions of these bootstrap estimates, the $5^{th}$ and $95^{th}$ percentiles can be derived and used to generate 95% confidence intervals (CI) for the parameter.

A purpose-built programme was created to bootstrap the psychometric data collected in the current experiment. First, the experimental data was fitted with a logistic function and estimates of slope (m) and SRT (c) were derived, as explained in section 2.1.2. Each of the seven points on the psychometric function were then individually bootstrapped by sampling with replacement from the number of correct/incorrect responses which made up each data point. For example, if the target colour and number were identified 6 times out of 30 at one SNR then the bootstrap sample would be drawn from a sample of 6 correct responses and 24 incorrect responses. Possible bootstrap samples could consist, therefore, of 8 correct responses and 22 incorrect responses, or 5 correct responses and 25 incorrect responses, or 10 correct responses and 20 incorrect responses.  This re-sampling was repeated for each point on the psychometric function. The bootstrapped function was then fitted with a logistic curve from which values of slope ($m_{*1}$) and SRT ($c_{*1}$) were derived. The whole process was repeated a large number of times to produce multiple bootstrapped values of slope and SRT (e.g. $m_{*1}$, $m_{*2}$, $m_{*3}$, $m_{*4}$…… $m_{*1000}$). These values were then averaged to find the mean value for slope and SRT and their distribution used to calculate 95% confidence intervals (CI). A validation of the bootstrap method was also carried out and appears in Appendix A.

The bootstrap process outlined above was carried out individually on the data of all 24 listeners in each of the four masking conditions tested. Figure 3.12 shows the distribution of the resultant 95% CI calculated for the slope and threshold parameters in each masking condition. For the three modulated, speech-like maskers (S, IVS and UVS) the mean 95% CI calculated for the slope parameter was approx +/- 2.5% per dB. This rose to +/- 4% per dB for the N condition. There was slightly more variation across masking conditions for the SRT parameter. This time CIs were very small for the N masker (+/- 0.9 dB) and largest for the S masker (+/- 3.9 dB). The confidence intervals displayed in Figure 3.12 give an indication of the underlying accuracy of the slope and threshold parameters measured for individuals. While the confidence intervals are not of an unreasonable magnitude, they do highlight the inter-listener variability of these measures and compound the need to collect data on more listeners to look at population means. Figure 3.13 shows the mean 95% confidence intervals calculated for the slope parameter by the bootstrap imposed on the population slope means for Experiment 1 (mean slopes averaged over younger and older listeners). While the relatively small slope difference between the IVS and UVS conditions fall within these 95% CI, a clear distinction can be still be made between the slopes measured in the static noise masker (N) and the modulated speech-like maskers (IVS, UVS, and S).

An interesting observation was also made from Figure 3.12; while SRT CIs are extremely small for psychometric functions measured in the N condition, slope CIs are rather large. It is likely that there is an inherent relationship between slope magnitude and the size of the error for this measurement; meaning CIs would get larger as slope values increase. Figure 3.14 examines this possibility by plotting slope data collected in Experiment 1 as a function of the calculated 95% CIs. It is

***Figure 3.12:*** *Shows distributions of confidence interval ranges generated by the bootstrap method. Column a, shows confidence interval ranges for the slope parameter and column b shows ranges for the threshold parameter. These distributions are shown for each of the four conditions used in Experiment 1.*

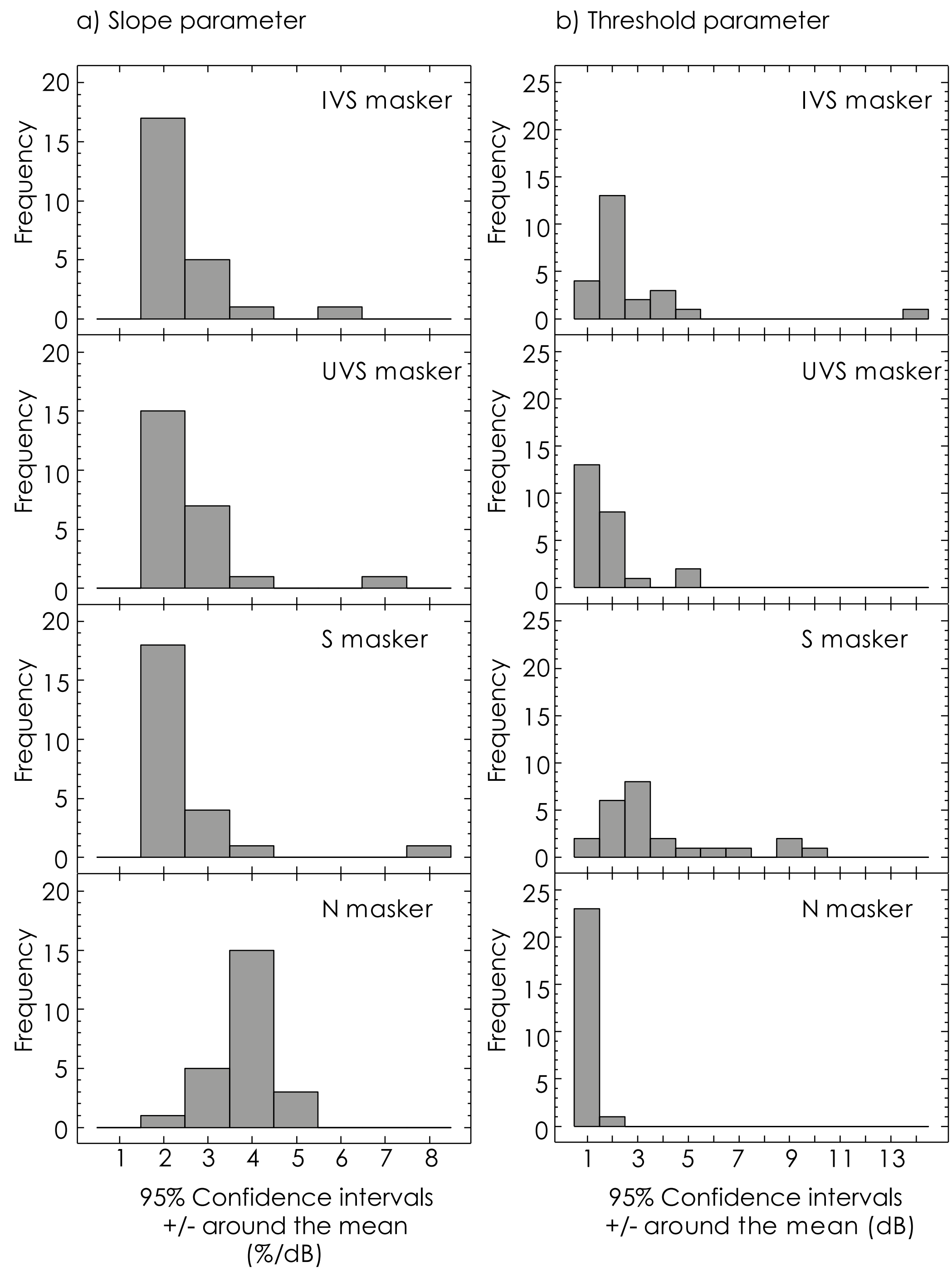


***Figure 3.13:*** *Mean slope is shown here for each of the four masking conditions. This graph is equivalent to figure 3.6, however mean slope is now averaged over young and older listeners. Error bars show the mean 95% confidence interval calculated by the bootstrap method for each condition.*

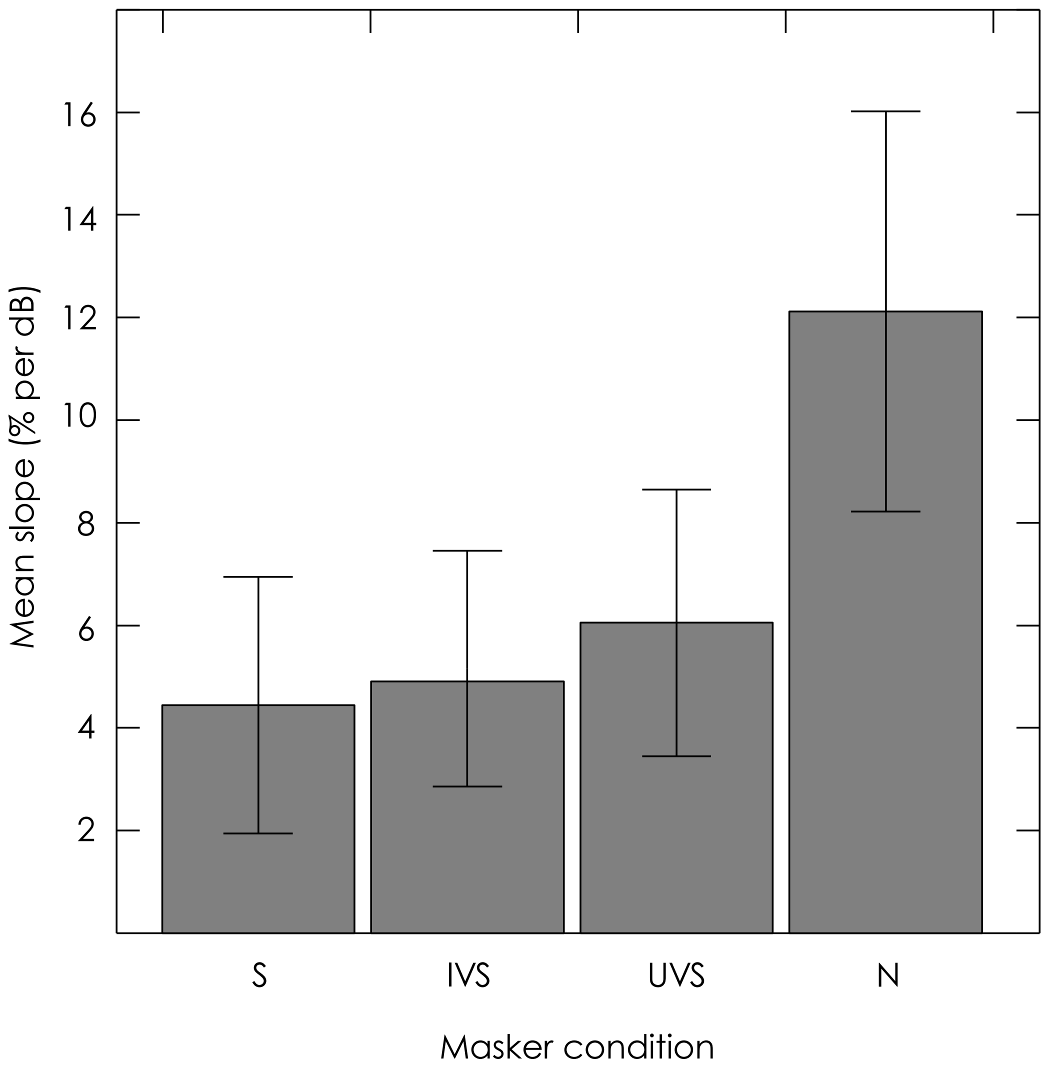
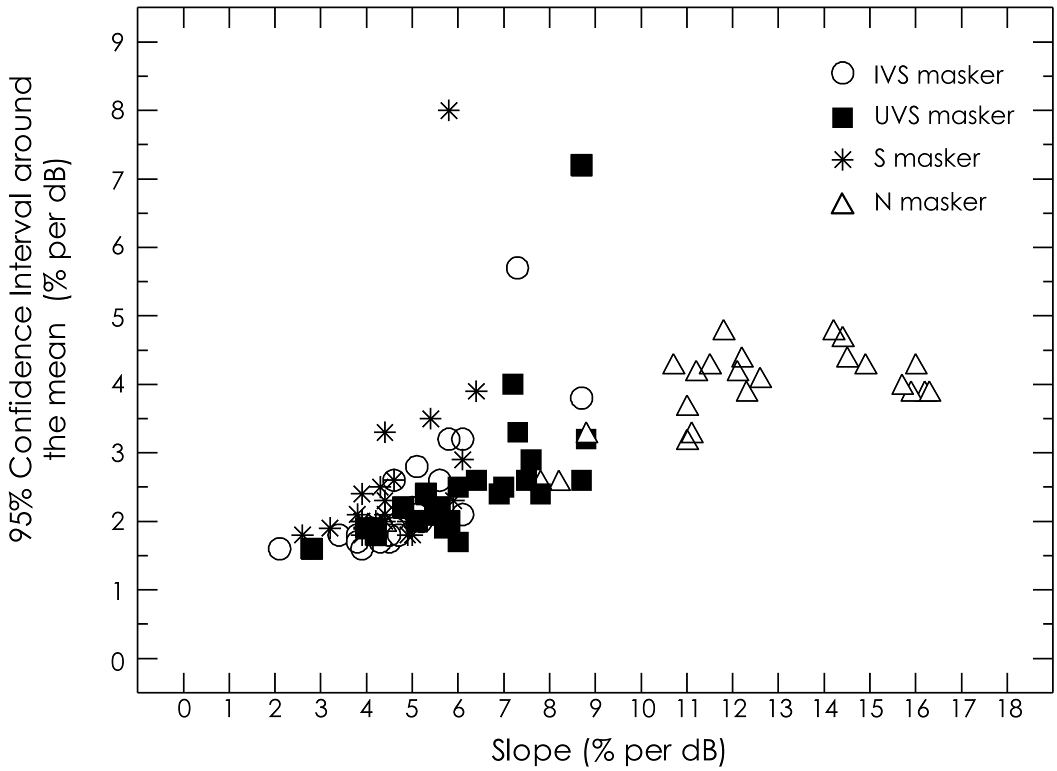


***Figure 3.14:*** *The slope parameters generated by the bootstrap for each listener in each of the four masker conditions (IVS – open circles, UVS - Closed squares, S - asterisks, N – open triangles) are plotted against their 95% confidence intervals.*



clear from this figure that CIs do indeed increase as the slope value increases. This is unsurprising, however, if we consider that as a slope becomes steeper it approaches infinity (i.e. a vertical line). A small error in such a case, while having little effect on SRT, and so giving a small CI for the threshold parameter, would result in a large absolute error value for slope.

Carrying out a bootstrap analysis highlights another limitation of the current method, namely that there is a certain degree of inter-listener variability in slope and threshold measurements suggesting that, where possible, slope comparisons should be made with population rather than individual means. This inter-listener variability along with the inherent relationship between slopes and CIs means that small differences in slopes between conditions (in the order of <2.5% per dB) are unlikely to be reliable. This range of inaccuracy is also likely to get larger as the slope gets steeper (i.e. increases to 4% per dB for a slope of 12% per dB).

## 3.4 Summary

Experiment 1 found that slopes for speech identification were shallower for modulated noises than they were for static noises. While initial results also suggested that maskers containing intelligible words (IVS and S) produced shallower slopes than maskers that did not contain intelligible words (UVS), further analysis questioned whether this result really supports the role of linguistic confusion in slope change. An analysis of the amplitude modulations in the stimuli suggested this slope difference was likely instead to be the result of reduced modulations in the UVS condition. It was concluded, therefore that the shallow slopes seen in the current experiment were not the result of confusion between the target and the masker.

Experiment 2 aims to measure slope changes that do arise due confusion between the target and the masker and attempts to establish whether these slope changes can be completely distinguished from those that arise due to variations in the amplitude modulations of the target.

# 4 Experiment 2: Manipulating linguistic and acoustic similarity to identify different *types* of shallow psychometric functions.

The aim of Experiment 1 was to consider the role that linguistic similarity between a target and a competing speech masker played in flattening the slope of the psychometric function. To separate *linguistic* similarity from the role that *acoustic* similarity between the target and masker might also play, masker stimuli were vocoded. While the slopes measured in Experiment 1 were relatively shallow (average slope ≈ 4.5% per dB for speech maskers) there was a strong indication that this flattening was not produced due to either form of confusion. Slope differences between conditions were instead attributed to amplitude modulation differences. The aim of Experiment 2 is to consider what listening conditions are necessary to induce target/masker confusion and considers the relative contributions that masker modulations and target/masker confusion have on slope as SNR changes.

The psychometric functions measured in Experiment 1 showed two inconsistencies to the slopes which have been highlighted in the literature as having occurred due to target/masker confusion (Brungart, 2001a; Egan et al., 1954):

1) The slopes, while shallower than static noise, did not show any regions of non-monotonicity; plateaus and U-shaped functions were common in previous studies where confusion occurred between the target and the masker.
2) Psychometric functions were measured at considerably lower SNRs; in the region of -15 dB SNR or lower rather than around 0 dB SNR.

It is likely that the comparatively steeper slopes and lower thresholds found in Experiment 1 can be explained by methodological variations. It is not uncommon in speech-in-speech studies, for example, to use target and masker sentences from the *same* speech corpus. For many corpuses, all of its sentences will have the same syntactic structure so if a competing sentence is taken from the same corpus as the target they are likely to be very linguistically similar. A number of recent studies have used target and masker stimuli from a particular speech corpus; the Coordinate Response Measure (CRM; Bolia et al., 2000). The closed nature of the CRM-in-

CRM task makes it a useful tool for measuring speech understanding in multi-talker environments but the content of target and masker are extremely similar: four out of seven words in each sentence are exactly the same (Ready, go, to, now). In section 2.3.7, the slopes of studies whose stimuli were taken from the same corpus were compared to those which were taken from different corpuses. The former gave shallower slopes than the latter. Further, of the unusually shaped psychometric functions noted in the slope survey (section 2.4.2), 42% were given when the target and the masker were taken from the CRM corpus. This gives an indication then that *high* degrees of similarity are required to give extremely shallow psychometric functions.

In Experiment 1 we tried to emulate more realistic listening conditions. While the target was taken from the CRM corpus, the masker was taken from a different corpus whose context was not related to that of the target (IEEE corpus; Rothauser et al., 1969). It is possible then that targets and maskers were not similar enough to produce psychometric functions of the class seen in previous literature (Brungart, 2001a; Egan et al., 1954). It is thus hypothesised that targets and maskers must be both acoustically very similar (i.e. spoken by the same or similar voice) and linguistically very similar (similar words or word order) to result in confusion and plateaus in the psychometric function.

To test this hypothesis, Experiment 2 measured psychometric functions where both the target and the masker were taken from the same corpus (CRM in CRM), and thus linguistically very similar, and compared them to functions given when targets and maskers were taken from different corpuses (CRM in IEEE) and, therefore, less linguistically similar. The acoustic similarity was likewise manipulated, the masker being either same-talker natural speech and, thus, acoustically similar to the target or same-talker vocoded speech and, thus, acoustically distinct from the target.

Psychometric functions were measured over a very wide range of SNRs (-32 to +8 dB) to encompass presentation levels used in both Experiment 1 and those of previous research. It was reasoned that looking over a wider SNR range could provided evidence for different classes of shallow slope, resulting from different mechanisms. It was expected that if different mechanisms were responsible for the shallow slopes seen at different SNR ranges they would be differentially affected by the acoustic and linguistic manipulations made in the experiment. If the shallow slopes seen in Experiment 1 for the speech and speech-like maskers were, as concluded, the result of amplitude modulations in the masker, they should be unaffected by the similarity manipulations. Arguably these manipulations should only affect shallow slopes which occur at higher SNRs and that are the result of confusion. Only two listeners participated in the current experiment as the results were found to be particularly clear.

## 4.1 Method

### 4.1.1 Listeners

Two listeners, both aged 26, took part in Experiment 2. One listener was the author and the other listener was a staff member at the MRC Institute of Hearing Research who had also taken part in Experiment 1. Both listeners had audiometric four frequency averages within the normal range (i.e. below 25 dB HL).

### 4.1.2 Stimuli

As in Experiment 1, the stimuli consisted of target speech presented in a masker. The target speech was again taken from the CRM corpus (Recordings: Kitterick et al., 2010) and, as before, only sentences with the call sign Baron were used. All selected target sentences were spoken by the same British male talker.

The maskers were either taken from the same corpus as the target (CRM-in-CRM condition) or from a different speech corpus (CRM-in-IEEE corpus). For maskers taken from the CRM corpus, sentences were selected from the remaining eight call signs (Arrow, Charlie, Eagle, Hopper, Laker, Ringo, or Tiger) ensuring that neither the colour nor number matched that of the selected target sentence. For maskers taken from the IEEE corpus, two different randomly chosen IEEE sentences were selected and concatenated together without a gap, thus ensuring that the masker was always longer than the target sentence. Both sets of maskers were spoken by the same British male used in the target sentence. Vocoded versions of both maskers were then produced using the same vocoding method used to create the intelligible (IVS) maskers used in Experiment 1. Four masker conditions were therefore used in the current experiment: natural CRM ("nCRM"), vocoded CRM ("vCRM"), natural IEEE ("nIEEE"), and vocoded IEEE ("vIEEE"). On each trial, the target and masker onsets were aligned and the two sentences added together. Target sentences had an average duration of 2.5 seconds. For the nCRM and vCRM maskers, which had the same average duration as the target, no further editing was required. For the nIEEE and vIEEE maskers, which had an average duration of 5.1 seconds, the maskers were terminated 25 ms after the offset of the target. Raised cosine gates of 25 ms were then applied to all target/masker stimuli.

The experiment was carried out in the same sound-treated booth and with the same equipment used in Experiment 1.

### 4.1.3 Procedure

The procedure followed closely that of Experiment 1, though with a few minor adjustments.

To ensure that psychometric functions were measured across the full range of SNRs where an effect of slope could occur, speech identification was measured at three SNR ranges instead of just one: an extremely unfavourable SNR range (-30 to - 20 dB), an unfavourable SNR range (-18 to -8) and a favourable SNR range (-6 to +8 dB). Data points in each SNR range were separately fitted with logistic curves and the slope and SRT calculated.

The number of trials completed at each individual SNR was also increased from 30 to 60 trials. It was reasoned that this would increase the accuracy of the psychometric functions measured, a particular issue considering the reduced number of listeners used in this study.

### 4.1.4 Analysis

The analysis of the psychometric functions measured in Experiment 2 differed slightly from the analysis used in the slope survey and Experiment 1. There, the data points were fitted with a standardised logistic function (Equation 2.1) and the slope and threshold at 50% correct derived. Experiment 2, however, measured psychometric functions over three SNR ranges. If performance in a particular SNR range was always greater than 50% correct then the logistic function would be extrapolated to 50% to calculate the slope and threshold values. Extrapolation can lead to inaccurate slope and threshold values particularly in SNR ranges where performance is always well above 50%. As a result, the parameters from the logistic function are not reported. The functions were, however, still fitted to the data, the primary reason for this being to give best fit lines which took into account all data points (i.e. the overall shape of the function) in each SNR range. The overall change in percent correct (i.e. the difference in performance from the lowest SNR in each range to the highest SNR in each range) was calculated from the logistic functions and used to calculate slope values as per equation 4.1:

$$\text{Slope} = \frac{(Y_2 - Y_1)}{(X_2 - X_1)} \tag{4.1}$$

where 1 is the X and Y values at the lowest SNR and 2 is the X and Y values at the highest SNR.

## 4.2 Results

Figure 4.1 shows the psychometric functions for the four masking conditions for both listeners. Table 4.1 summarises the slope values for each condition and listener. In the extremely unfavourable SNR range (-32 to -20), relatively little difference was seen between the slopes of the four masking conditions. All the slopes in this SNR range were found to be between 1.4 and 3.8% per dB. Two of the maskers (nIEEE and vIEEE) were also measured at this range in Experiment 1 (though there they were termed S and IVS). The present slopes for those maskers were shallower than before; on average 2.6 and 3.6% per dB for the nIEEE and vIEEE respectively in the current experiment compared to an average of 5.0 and 6.4% per dB for the same listeners in Experiment 1[18]. Nevertheless, in both experiments 1 and 2 it is clear that, at an unfavourable SNR range, speech and speech-like maskers give shallower slopes than the 12% or so that would be expected from a steady noise (see Experiment 1). That the vIEEE and vCRM slope values and the nIEEE and nCRM slope values were very similar to one another indicates that, at extremely unfavourable SNRs, slope was not affected by the type of corpus used, i.e. by the similarity of the content in the target and masker, or by the acoustic similarity of the target and masker voices.

At the unfavourable SNR range (-18 to -8 dB), greater variation in performance was seen across listeners. While performance reached ceiling in all four conditions for listener 1, for listener 2, performance only reached ceiling at this level for the IEEE maskers. The CRM maskers gave poorer performance at this level for this listener which may point to a possible effect of content similarity.

At the favourable SNR range (-6 to +8 dB), an interaction between similarity of linguistic content and acoustic similarity occurred. While performance reached 100% for the nIEEE, vIEEE and vCRM maskers, performance dropped significantly in the nCRM masker giving a "U-shaped" function for both listeners for this masker. The U shape was broader and slightly shallower for listener 2 than for listener 1, with the decline in performance starting earlier (-8 dB), levelling out for longer (-6 to -2 dB), and increasing again more slowly (0 to +6 dB).

### 4.2.1 Summary of results

- At extremely unfavourable SNRs little difference in slope was seen across the four masker types.
- At favourable SNRs speech maskers which were both acoustically and linguistically similar to the target gave U-shaped psychometric functions.

[18] The bootstrap analysis of Experiment 1 demonstrated that slopes of this magnitude were likely to be only measurable to within +/- 2.5% dB of accuracy. This measurement error may account for the slope difference seen for these conditions across experiments.

***Figure 4.1:*** *shows the psychometric functions measured for four masker conditions; nCRM, vCRM, nIEEE, vIEEE. Each condition is measured over three different SNR ranges; extremely unfavourable (-32 to -20), unfavourable (-18 to -8), and favourable (-6 to +6). The right-hand panels show the psychometric functions for listener 1 and the left-hand panels show those for listener 2.*

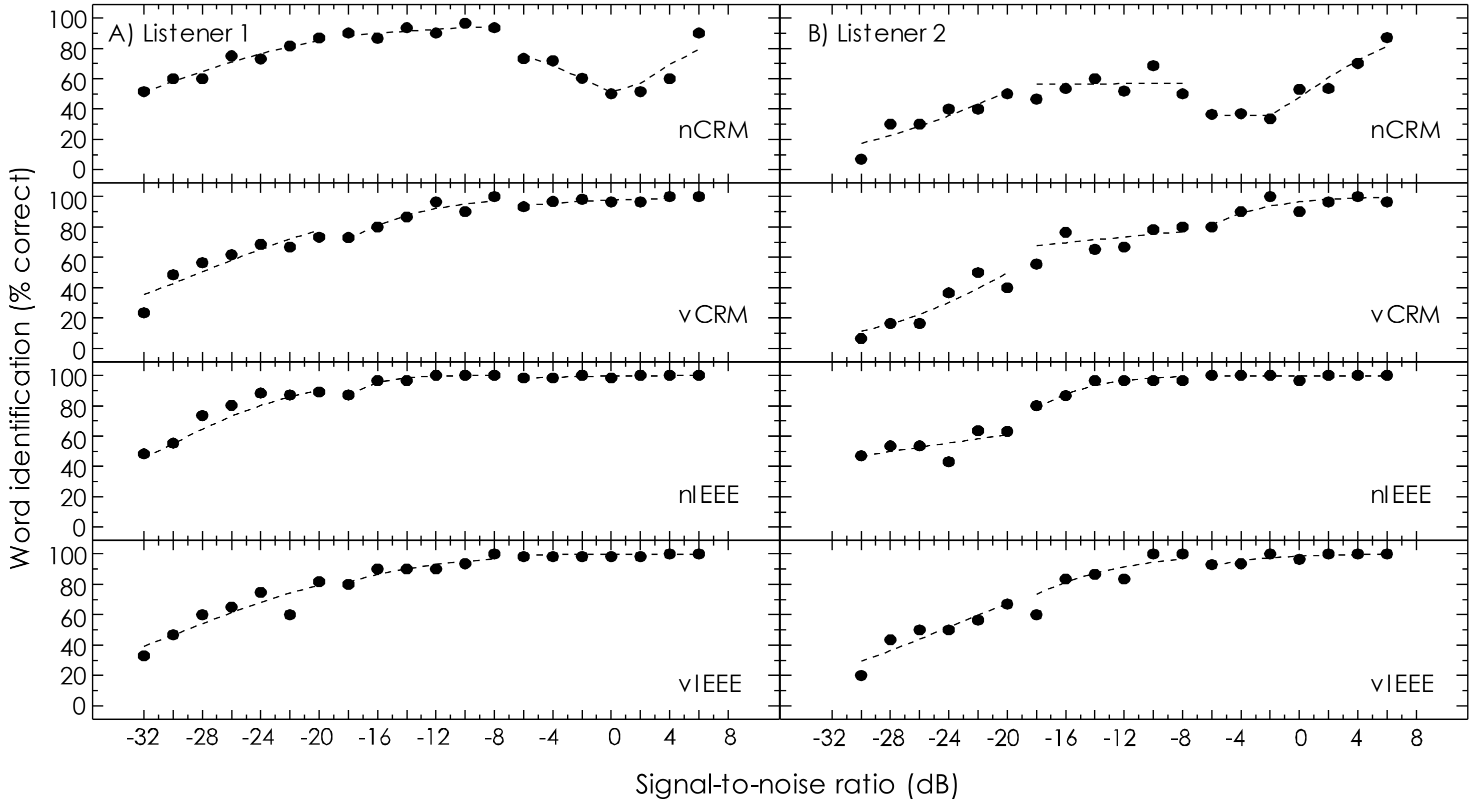


***Table 4.1:*** *Summarises the slope values (% per db) for all four masker conditions at three different SNR ranges for both listeners. Two slope values are given in the nCRM condition at the favourable SNR for both listeners; this is to fully represent the U-shaped function noted here.*

| **Masker** | **Type** | **Condition name** | | **SNR range** | | |
|---|---|---|---|---|---|---|
| | | | | Extremely unfavourable | Unfavourable | Favourable |
| CRM | Normal | nCRM | L1 | 2.8 | 0.6 | -2.5 5.9 |
| | | | L2 | 3.4 | 0.1 | 0.0 5.6 |
| CRM | Vocoded | vCRM | L1 | 3.5 | 2.4 | 0.4 |
| | | | L2 | 3.8 | 1.5 | 0.9 |
| IEEE | Normal | nIEEE | L1 | 3.8 | 1.3 | 0.1 |
| | | | L2 | 1.4 | 2.1 | 0.0 |
| IEEE | Vocoded | vIEEE | L1 | 3.4 | 1.5 | 0.4 |
| | | | L2 | 3.8 | 2.3 | 0.6 |

## 4.3 Discussion

In the current experiment the acoustic similarity of target and masker speech and the similarity of their linguistic content were parametrically varied. The effect these manipulations had on the slope of the psychometric function at three different SNR ranges were also observed. Of the three SNR ranges tested, only one, the favourable SNR range, showed a clear difference in slope between any of the four masking conditions. While relatively shallow slopes were noted at lower SNRs, there was very little difference in slope across the similarity manipulations. These results, when combined with those from Experiment 1, support the contention that there are at least two types of shallow slopes: those that occur at extremely low SNRs and those which occur at more favourable levels. The results also suggest that different mechanisms are responsible for these different types of shallow slope.

As was seen in Experiment 1, the shallow slopes given by speech maskers at low SNRs seem to be minimally affected by the presence of linguistic factors in the masker or by acoustic similarity between the target and masker (natural speech maskers produced marginally shallower slopes than the vocoded maskers in Experiment 1 and this difference was not seen at all in Experiment 2). Instead, they are more likely to result from temporal and spectral modulations in the maskers as opposed to any linguistic or higher level features. This was further supported in the current experiment with the added observation that even when targets and maskers were linguistically very similar, i.e. when both speech tokens were taken from the same highly confusable corpus, this had very little effect on slope at extremely unfavourable SNRs.

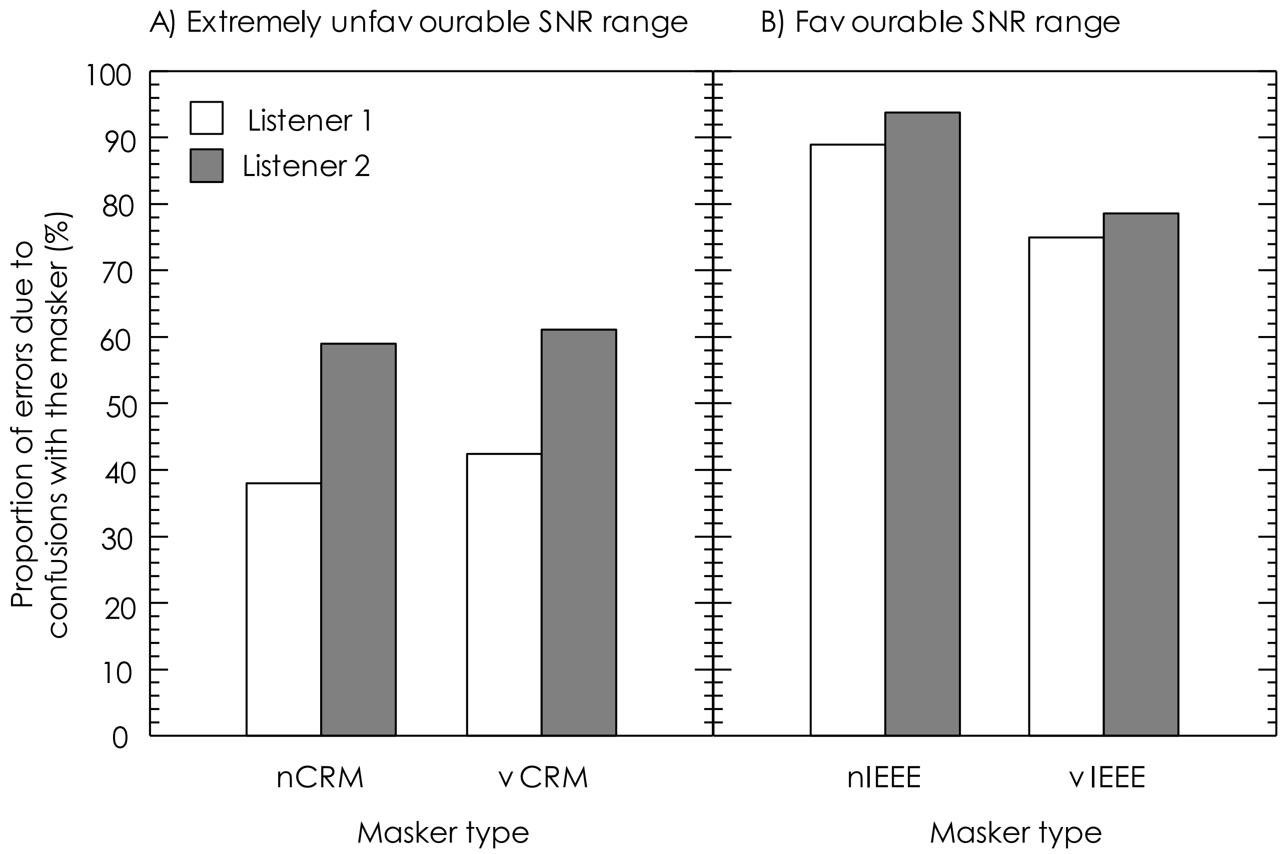
In contrast, the shallow slopes that occurred at more favourable SNRs were clearly highly dependent on both the acoustic and the linguistic similarity between the target and the masker. "U-shaped" functions were seen when both the target and the masker were taken from the CRM corpus but not when the masker was taken from a different corpus. While some previous research has noted plateaus in performance for similar conditions, rather than the U shapes seen here, it is highly likely that these slope changes were driven by the same mechanism; U-shaped functions will average out to a flat function if data is averaged across listeners (e.g. Brungart, 2001a). Other consequences of averaging data have also been discussed by Estes (1956). As mentioned previously (see Chapters 1 and 2), it has been suggested that an increased reliance on the level differences between target and masker speech causes the dissociation between intelligibility and relative target level that results in the plateau or dip in performance. The results of the current study provide further evidence that these functions result from confusion between the target and the masker. The precise source of this confusion is considered further in Experiment 3. Differences in the functions measured at the favourable SNR range for the two listeners gives an indication that individual differences may play a role in the general shape of these functions. A wider U-shape would suggest, for example, that the listener needs a larger SNR difference to distinguish the target from the masker and may indicate a greater susceptibly to confusion for that listener.

These two different mechanisms for shallow slopes can be further examined by looking at the types of errors listeners made during the experiment. For the nCRM and vCRM conditions it is possible to determine whether errors were the result of confusions with the masker or not. The errors for these conditions were examined at both the extremely unfavourable SNR range and the favourable SNR range for both listeners. A response was defined as a confusion error if the responded colour, number, or both were taken from the masker instead of the target sentence.

Figure 4.2 displays the proportion of confusion errors made by each listener in each condition. Both listeners showed less confusion errors in the extremely unfavourable SNR range conditions than they did in the favourable SNR range conditions. The figure also shows that while the proportions of confusion errors were relatively similar between the nCRM and vCRM conditions at the extremely unfavourable SNR range, they differed somewhat at the favourable SNR range. In the favourable SNR range confusion errors were higher for both listeners in the nCRM condition than the vCRM condition, i.e. they were higher in the condition where there is high acoustic similarity between the target and the masker. These patterns of errors further support the suggestion of differences between the shallow slopes seen at extremely unfavourable SNRs and the shallow slopes seen at favourable SNRs. While shallow slopes seen at higher SNRs (i.e. those around 0 dB) seem very dependent on the degree of target/masker confusion, those seen at lower SNRs are much less dependent on this confusion. A proportion of the errors at lower SNRs do still seem to be the result of confusion (approximately 40% for listener 1 and 60% for listener 2). Ihlefeld and Shinn-Cunningham (2008b), using a same-talker CRM-in-CRM task also found that confusion errors occurred at low SNRs when the more intense sentence should have been easily identified as the masker and ignored. It was argued that these errors may have occurred because either the listener was unsure if the sentence they just heard was the masker or whether it was simply easier to report the audible keywords when the target was inaudible.

As mentioned above a U-shaped or plateaued function similar to those seen in the literature (Brungart, 2001a; Dirks & Bower, 1969; Egan et al., 1954) was reproduced in the current experiment in the nCRM condition at the favourable SNR range. This

***Figure 4.2:*** *The proportion of confusion errors made by each listener in two masking conditions (natural CRM and vocoded CRM) at two SNR ranges (the extremely unfavourable "negative range" SNRs and the favourable "zero range" SNRs).*



implies that three conditions must hold for these unusual shaped functions to occur: 1) the target and masker must be presented at roughly equivalent levels (i.e. the SNR was near 0 dB), 2) the target and masker must be acoustically similar (i.e. spoken by the same talker and not vocoded), and 3) the target and masker content must be highly similar (i.e. taken from the same speech corpus). If any of these three conditions were removed in the current experiment the effect was eradicated and performance reached a ceiling level. This supports several of the trends identified in the slope survey, in which it was argued that confusion only affects slope in extreme conditions, i.e. when no cues are available other than level to distinguish the target from the masker. Once another cue is introduced, e.g. voice difference, corpus difference or spatial separation, any further effect on slope resulting from an exaggeration of this additional cue is minimal. But, while robust experimentally, the effect on slope due to extreme confusion may be unlikely to occur in everyday listening situations. Conditions (1) and (2) in themselves are often fulfilled in everyday listening environments, for example listening to a female talker while a similar-sounding female is talking nearby[19]. It seems unlikely, however, that condition (3) would also be fulfilled. For this, both conversations would have to contain sections which were semantically similar and further that these sections be temporally aligned as they are with CRM-in-CRM stimuli[20]. That being said, same-talker CRM-in-CRM presumably provides a situation of "maximum confusion" and therefore provides a good experimental tool for understanding the exact nature of the confusion.

[19] Plomp (1977) suggested that SNR is around 0 dB in typical "cocktail-party conditions" when the target speaker is about 0.7m away. Thus condition (1) would plausibly be fulfilled in everyday listening environments.

[20] One listening situation where all three of these conditions could possibly arise is at the self-checkout machines in supermarkets. The machines are located close together, the same female voice is used for all machines and there is a limited vocabulary so the speech from different machines is often remarkably similar. With enough machines there is a reasonable chance of a temporal alignment of the sentences.

## 4.4 Summary

Experiment 2 demonstrated that different types of shallow slopes can occur, which are both quantitatively and qualitatively different from one another. Amplitude modulations in the masker result in shallow slopes when speech is at unfavourable SNRs, while high degrees of confusion between targets and maskers results in unusual-shaped psychometric functions but only at SNRs around 0 dB. Experiment 3 goes on to consider further the unusually shaped, confusion-based psychometric functions by looking at the role of attention in the CRM-in-CRM paradigm.

# 5 Experiment 3: The effect of improving selective attention on confusion-based shallow slopes.

It was demonstrated in Experiment 2 that unusually-shaped psychometric functions were observed when a large amount of confusion existed between target and masker speech. These confusion-based shallow slopes occurred around 0 dB SNR presentation levels when the competing speech stimuli were both acoustically and linguistically very similar to the target. The purpose of Experiment 3 was to address the role that selective attention can play in relieving this confusion and therefore in steepening the slope of the psychometric function.

As was outlined in Chapter 1, the interference that arises due to confusion between speech stimuli is often termed "informational masking" (Arbogast et al., 2002; Brungart, 2001a; Brungart & Simpson, 2002; Freyman et al., 2001, 2004). Yost (2006) has argued that the term informational masking is unnecessary and suggests that this interference can be more simply explained as a failure of selective attention. Early research on the perception of speech in competing speech put great emphasis on the role of attention. Broadbent (1952), for example, ascribed mistakes made by listeners when asked to report one of two simultaneous messages as a failure to *select* the correct message. More recently, focus has returned to this approach with general theories of selective attention being proposed to explain the interference that arises when speech stimuli are presented in competing speech (Francis, 2010; Shinn-Cunningham, 2008).

Shinn-Cunningham (2008) argued that a failure at any one of three stages may result in a failure to selectively attend to a target in the presence of a similar competing sound. Firstly, similarity between the target and masker may lead to difficulties grouping together the sounds of individual components from the two sentences into separate auditory objects (Bregman, 1990). Problems at this stage are sometimes referred to as failures of "object formation" (Shinn-Cunningham & Best, 2008). Secondly, even if individual objects are successfully formed into coherent units, difficulties might arise in successfully linking these units together. Difficulties at this stage have been referred to as failures of "automatic streaming". Thirdly, even if objects and streams are correctly formed, the similarity of the two sentences may make it hard for the listener to focus and maintain attention on the desired sentence. This third type of confusion has been referred to as a failure of "object selection". It is unclear, however, at which of these three stages difficulties arise causing confusion and a flattening of the psychometric function in the CRM-in-CRM task.

We found in Experiment 2 that, when asked, listeners could report *both* call signs from the two competing CRM sentences. Putting this into the attentional framework described above, this would suggest that separate auditory objects had been successfully formed, at least near the beginning of the sentences. This result corroborates two earlier reports. Best, Gallun, Ihlefeld and Shinn-Cunningham (2006), also using a CRM-in-CRM task, demonstrated that when asked to report *both* sets of colour and number keywords from competing sentences (i.e. to divide attention between the two sentences) listeners were able to do so with approximately 40% accuracy. This result indicates that object formation could likewise be accomplished, at least part of the time, later on in the sentence. Brungart and Simpson (2004) reported that a large proportion of incorrect responses during CRM-in-CRM tasks were taken directly from the masker phrase (as did the results of Experiment 2). This also provides evidence that listeners were able to understand syllables and words from both messages, i.e. that auditory objects were properly formed, but that listeners were unable to determine which phrase belonged to which speaker. Indeed, listeners have been found to be able to make use of very small spectro-temporal differences between targets and maskers to successfully group sounds locally (Darwin & Carlyon, 1995). This would suggest that even when spectral and temporal features are very similar, as they are when speech is presented in competing speech, short-term grouping is usually robust and syllables and words are likely to be correctly formed (Shinn-Cunningham & Best, 2008).

There is also evidence to suggest that difficulties linking objects over time to form proper streams are unlikely to be responsible for the high degree of confusion seen in the CRM-in-CRM task. Ihlefeld and Shinn-Cunningham (2008b) carried out a selective attention task using a same-talker, CRM-in-CRM task. It was noted that when target and masker sentences were co-located and presented at SNRs between -10 and 10 dB, the number of response errors where both colour and number keywords were taken from the masker (confusion errors) peaked around 0 dB while the number of response errors where one keyword was taken from the masker and one for the target (mixed errors) remained constant over the TMR range. Confusion errors would suggest listeners had difficulties selecting the correct segments or steam, while mixed errors would suggest that listeners had difficulties linking segments together into streams. That confusion errors were directly affected by a change in level but mixed errors were not suggests that the drop in performance around 0 dB, and therefore the resultant U-shaped function, was the result of increased difficulties selecting the target keywords rather than increased difficulties streaming them over time.

It seems likely then that a failure to select auditory objects, rather than a failure to segment and stream auditory objects results in the confusion that arises in CRM-in-CRM task around 0 dB SNR. It can be hypothesised that the effect, a U-shaped function, is seen when there are few cues other than level available to aid object selection. When two CRM sentences are played, at the start top-down cues derived from the task instructions (e.g. "listen for the Baron sentence") may help listeners identify which sentence to attend to. As the sentences progress, however, fewer cues are available which listeners can use to *maintain* attentional focus. If the two CRM sentences are acoustically very similar (i.e. spoken by the same person), the listener may lose track of which sentence they are listening to, and so be forced to guess.

Several studies have shown that priming some aspects of target speech before presentation can create a release from masking (Freyman et al., 2004; Helfer & Freyman, 2009; Yang et al., 2007). Freyman et al., (2004), for example, found that playing the start of a target sentence improved performance on a speech identification task with co-located talkers even if the fragment of the target sentence – the prime − was spoken by a different talker from the target. Further, the effect was found to be maintained if the prime was printed on paper and read silently. The prime added no extra improvement in intelligibility, however, if the two talkers were perceptually spatially separated using the precedence effect or if the target was presented in a static noise. Freyman et al., argued, therefore, that the prime acts as a top-down cue to direct attention when the target and masker are highly confusable.

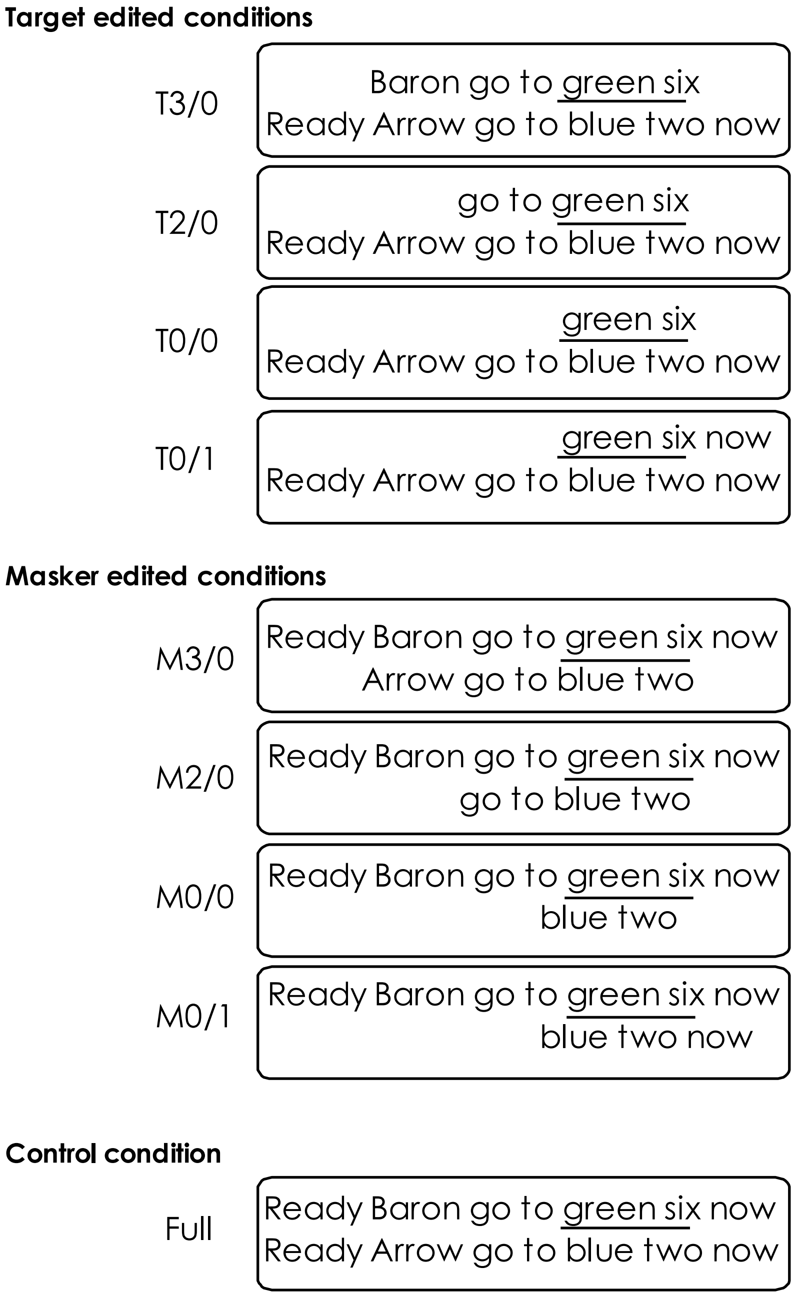
The systematic slope survey in Chapter 2 demonstrated that priming the target sentence also had an effect on the slope of the psychometric function. The median value of the slopes for speech-in-speech cases where no prime was provided was 5.9% per dB, but was 7.5% per dB if either the target voice, content or both were primed (see figure 2.7). These results suggest that in highly confusable situations, improving the listeners' ability to select the target speech can increase the slope of the psychometric function.

This hypothesis can be tested with the CRM paradigm by providing listeners with some top-down cues on to which attention can be "latched". This would require giving the target some distinguishing feature with which it can be distinguished from the masker and indicating this feature, before presentation, to the listener. Previous studies have, for example, varied pitch or spatial location and have found that these differences can indeed eradicate the U-shape/plateau in the psychometric function (Brungart & Simpson, 2007; Freyman et al., 1999; Ihlefeld & Shinn-Cunningham, 2008b). Experiments 1 and 2 have demonstrated that even small changes in the acoustical properties of the target can greatly reduce confusion and eradicate the U-shape/plateau in the function. Whether this reduced confusion is solely down to an improved ability to select the target is unclear, however, as these cues will also improve segmentation and object formation and the streaming of segments across time (Ihlefeld & Shinn-Cunningham, 2008b). By introducing such cues then we cannot establish the effect that simply directing listeners' attention towards the target sentence can have on the degree of confusion and on the U-shaped function which arose in the CRM-in-CRM task. Thus a different approach to manipulating attention was used in the present experiment: the relative salience of the sentence as induced by onset differences. The focus of attention is affected by the inherent salience of a sound (Conway, Cowan, & Bunting, 2001). It has also been shown that the onset of an abrupt sound can involuntarily draw attention (e.g. Treisman & Gelade, 1980). We reasoned, therefore, that if the target sentence started *after* the masker sentence, its abrupt onset would have the effect of directing listeners' attention towards the target thus making it easier to distinguish from the masker.

In terms of a CRM-in-CRM task, an abruptly starting target which consists of just the colour and number keywords, yet still aligned with the keywords of a full-sentence CRM masker, should be more easily identified than a full sentence target. Aside from giving an improvement in performance, if failures of object selection are involved in flattening the slope of the psychometric function, then the U-shaped psychometric function should also be eradicated by directing listeners' attention to the target in this way. It was further reasoned that any improvement seen when using an abruptly starting CRM target compared to a full CRM target should decrease as more of the words preceding the keywords were added back into the target sentence. More words after the attention-grabbing onset would mean more opportunities for the listener to become distracted by the masker or lose track of the target sentence before the keywords were reached. This increase in confusion should cause a flattening of the slope and an incremental return to a U-shaped function.

Experiment 3 tested these predictions using several conditions of abruptly starting edited CRM targets ("target-edited" conditions), Figure 5.1 illustrates these conditions: including 1) just the keywords, 2) the keywords plus two preceding words ("go to") and 3) the keywords plus three preceding words ("Baron go to"). A fourth condition where the keywords and the following word ("now") were included was also tested. The reason for including this last condition was to rule out contributions of a suffix effect in the CRM-in-CRM paradigm, as several studies of memory have shown that an irrelevant word placed at the start or end of a list of to-be-remembered words can have a detrimental effect on recall performance (Crowder, 1967; Crowder & Morton, 1969). Nicholls and Jones (2002) argued that a suffix causes most interference when it is strongly perceptually grouped with the to-be-remembered list. As it can be argued that the CRM task has similarities with memory tasks (it requires the memory and recall of two keywords), it is possible that for some of the difficulties recalling keywords are the result of the suffix "now". If this is the case, we would expect to see reduced performance when the word "now" is included compared to when the target consists of the keywords alone.

***Figure 5.1:*** *Nine different stimulus conditions were used in Experiment 3. Four different target length configurations (T0/0, T0/1, T2/0, T3/0), Four different masker length configurations (M0/0, M0/1, M2/0, M3/0), and a full sentence control.*



As in the previous experiments using the CRM-in-CRM paradigm, psychometric functions were constructed using the results of the identification of just the keywords from the target sentences. Importantly, in all conditions the timing of the keywords was the same as they would have been had no editing been done, i.e. the keywords were masked as they would have been had a full sentence been presented. Any threshold or slope changes seen for the different conditions cannot, therefore, be attributed to acoustic changes to the masking of the keywords.

A further four conditions were created by applying the same editing procedure to the masker sentence (see the bottom of figure 5.1). In these "masker-edited" conditions the target was the longer of the two sentences and thus it was the abrupt onset of the masker that should draw attention. In all other respects, however, the masker-edited conditions were exactly the same as the target-edited conditions. It was reasoned that directing listeners' attention away from the target and towards the masker instead should have differential effects on the resultant psychometric function if object selection is playing a role.

## 5.1 Method

### 5.1.1 Listeners

Eight listeners took part in Experiment 3. Their ages ranged from 18 to 38 (mean age = 26). One listener was the author, 4 listeners were staff members at the MRC Institute of Hearing Research and 3 listeners were volunteers recruited from the University of Strathclyde student population. All listeners had audiometric four frequency averages below 25 dB HL and so within the normal range.

### 5.1.2 Stimuli

In Experiment 3, both target and masker sentences were taken from the CRM corpus (Kitterick et al., 2010). All sentences were spoken by the same British-English male, with sentences containing the call sign Baron used as target sentences and those containing the call sign Arrow used as masker sentences.

The experimental conditions are illustrated in Figure 5.1. Either the target ("T") or the masker ("M") sentence could be edited. The four types of edit, denoted as the number of words which appeared before/after the keywords, were:

- "0/0" (just the keywords, e.g. "red four").
- "0/1" (the keywords plus the following word now, e.g. "red four now").
- "2/0" (the keywords plus the preceding words "go to", e.g. "go to red four").
- "3/0" (the keywords plus the preceding call sign and the words "go to", e.g. "Baron go to red four now").
- A control condition "4/1" or "full", where neither the target nor the masker was edited, was also included. This condition is equivalent to the "nCRM" condition in Experiment 2.

The onset and offset times corresponding to the four edit types were measured for each of the target sentences (those with the call sign Baron) and masker sentences (those with the call sign Arrow) using the speech editing software Praat (Appendix B reports these values). On each stimulus presentation, the target and masker sentences were aligned at the beginning and, depending on the edit type, the corresponding onset and offset durations of the to-be-edited sentence were replaced with silence. This method ensured that target/masker stimuli would be aligned in each edit type as they would have been had both sentences been played in full, i.e. with the same degree of keyword overlap. 20ms raised cosine gates were then applied to the complete stimulus (target plus masker) waveform.

The experiment was carried out in the same sound-treated booth and with the same equipment used in Experiments 1 and 2. As before stimuli were presented diotically. Masker sentences were always presented at 70 dB and the level of the target sentence varied to create 7 different SNRs.

### 5.1.3 Procedure

Psychometric functions were measured in the nine different conditions following much the same procedure as followed in Experiments 1 and 2. Listeners were asked to identify keywords from the target sentences, making their selection by pressing the appropriate button on the eight-by-four array on the computer screen in front of them. As Experiment 3 was designed to follow up on the U-shaped psychometric functions observed in experiment 2, seven SNRs were used to cover the expected range of the "U" shaped function − -6, -4, -2, 0, +2, +4, and +6 dB.

Each listening block consisted of 105 trials (= 15 trials at each SNR). The edited sentence (target or masker) and the type of edit were fixed within a block. Listeners completed two blocks of each of the target-edited conditions (T0/0, T0/1, T2/0 and T3/0), and two blocks of each of the masker-edited conditions (M0/0, M0/1, M2/0 and M3/0). Listeners also completed 2 blocks of the full sentence control condition (4/1). Listeners therefore completed 18 blocks in total, each block taking 15- 20 minutes. Trial order was counter-balanced across listeners.

### 5.1.4 Analysis

In both the slope survey and in Experiment 1 psychometric functions were fit with a logistic function using the least squares method. This allowed a direct comparison of slopes to be made across studies and conditions. The logistic function is unable, however, to capture the U-shaped psychometric functions produced in the CRM-in-CRM paradigm. As the key objective of experiment 3 was to establish whether editing CRM sentences can eradicate the U-shaped dip and increase the slope of the function, an alternative fitting method was required for this experiment.

The method is illustrated in Figure 5.2. It was reasoned that 1) averaging speech identification scores at the two lowest SNRs (-6 and -4 dB) and at the two highest SNRs (+4 and +6 dB), 2) fitting a straight line between these values, and then 3) calculating the slope of this line would give a good indication of the overall gradient of the function. Also, the difference between the speech identification score predicted by this regression line at 0 dB (the usual centre of the U-shaped portion of the function) and the average speech identification score actually recorded around 0 dB (taken as the mean performance for -2, 0, and 2 dB SNRs) provided a second parameter and quantification of any dip in performance. This is termed the "confusion difference" (CD). A low slope value and a high CD value would represent a classic U-shaped function, while a high slope and zero CD value would represent a normal monotonic S-shaped function.

Algebraically, if X denotes points on the x axis in dB, and Y denotes points on the y axis in % correct, then the lowest points are

$$X_a = \frac{-6 + -4}{2} = -5$$

$$Y_a = \frac{P_{-6} + P_{-4}}{2}$$

the highest points are

$$X_b = \frac{4 + 6}{2} = 5$$

$$Y_b = \frac{P_4 + P_6}{2}$$

the slope* of the line between the two is [21]

$$\text{Slope*} = \frac{Y_b - Y_a}{X_b - X_a}$$

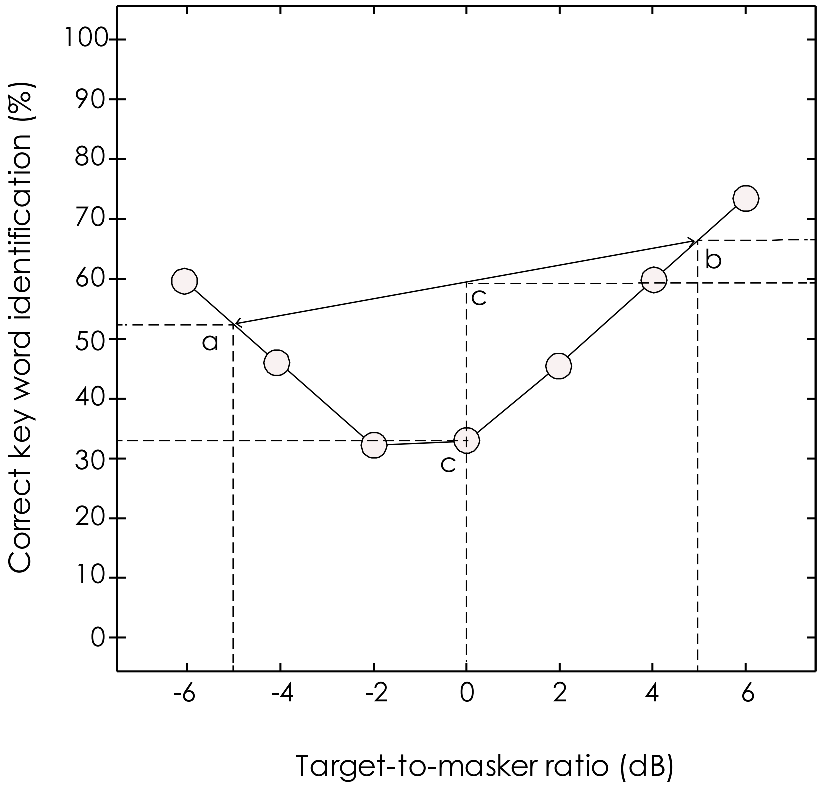
The predicted value at X = 0 is

$$Y_c = Y_b - \text{Slope*} \times X_b$$

and so, $CD = Y_c - Y_0$

[21] The asterisk is used here to denote that the method for calculating "slope" differs from the method outlined in section 2.1.2 which was used to calculate slope in the slope survey and in Experiment 1.

***Figure 5.2:*** *Schematic illustration of the fitting process used in Experiment 3. The method for measuring both slope and confusion difference are also indicated.*



## 5.2 Results

Figure 5.3 shows the individual psychometric functions for each listener in each of the target-edited conditions (open symbols) and the full-sentence control condition (asterisks). For 6 of the 8 listeners (excepting listeners 3 and 4), the full sentence condition gave a U-shaped function. Generally, editing the target sentence gave an improvement in performance at all SNRs. The only exception to this was for listeners 1 and 5, for whom an increase in performance was only seen at higher SNRs > 0 dB. As well as the positions of the functions, the shapes of the target-edited functions also differed from those of the full sentence control. For most listeners the U shape disappeared as the target was progressively edited. Indeed, only one listener (listener 7) still exhibited a dip in performance in the T0/0 condition (i.e. the shortest edited length).

Figure 5.4 displays the individual masker-edited functions for each listener. As was seen with the target-edited conditions, the psychometric functions for the masker-edited conditions again showed increased performance compared to the full-sentence control. Again this improvement was not seen at lower SNRs for listeners 1 & 5. Despite an overall improvement for the masker-edited condition, the psychometric functions were relatively flat. For most listeners, for example, performance in the T0/0 condition was between 80 – 90% across the whole SNR range.

Figure 5.5 is a scatter plot of slope* and CD values averaged across listeners for the four target-edited conditions, the four-masker-edited conditions, and the full sentence control. The plot gives an indication of both the shallowness of the slope and the degree of non-monotonicity of the psychometric function. There was a clear grouping of the data; A) while slopes for the T3/0 and full conditions were relatively shallow; the CD values for these conditions were high, suggesting dips in the function. B) the other target-edited conditions (T0/0, T0/1 and T2/0) had slightly steeper slopes and much smaller CD values, suggesting more monotonic functions. C) For the masker-edited conditions, again when only one word was edited (M3/0) the slope was shallow and the CD value was relatively high. Unlike their equivalent target-edited conditions, however, the remaining masker-edited conditions (M0/0, M0/1, and M2/0) gave even shallower slopes than the full sentence control.

***Figure 5.3:*** *Individual psychometric functions for each listener in each of the four different target edited conditions: T3/0 (circles), T2/0 (triangles), T0/1 (diamond), T0/0 (squares) and also in the full sentence control (asterisks).*

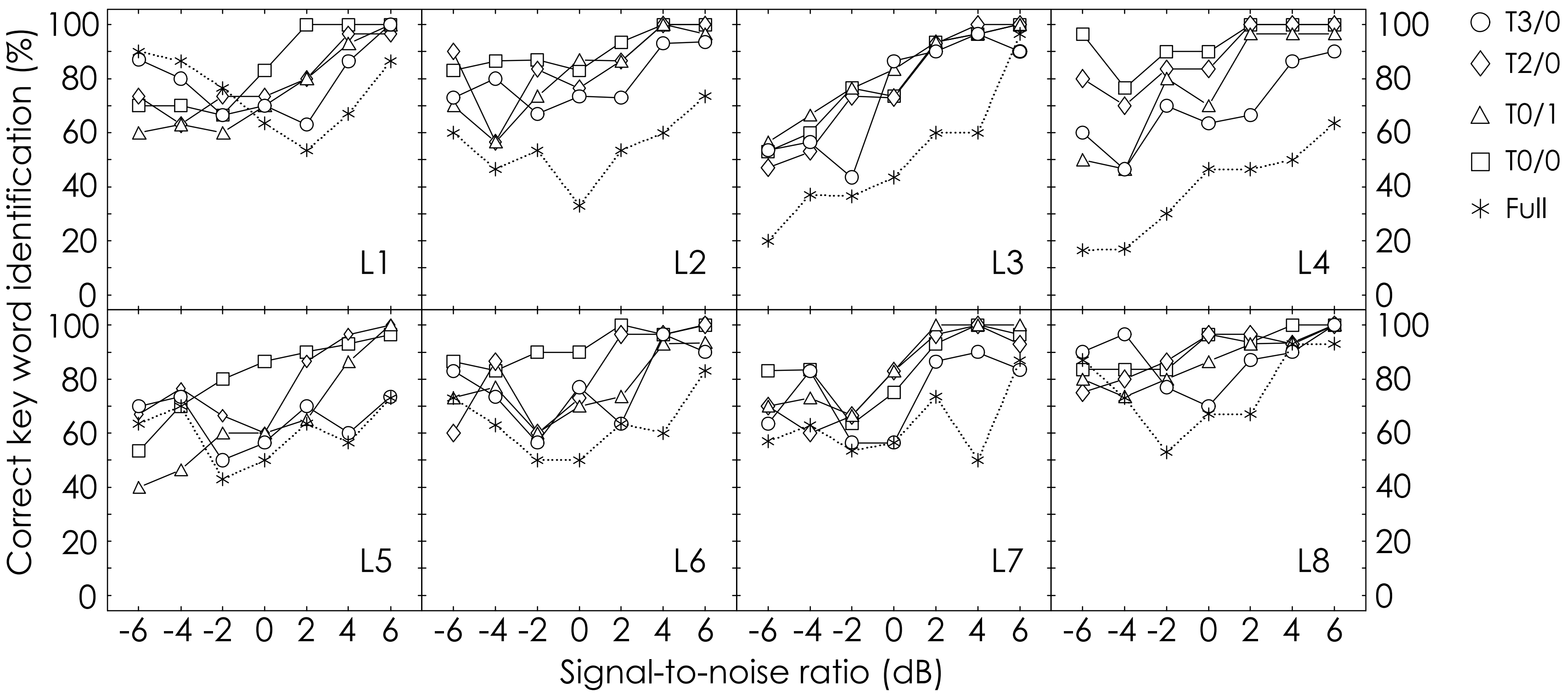


**Figure 5.4:** *Individual psychometric functions for each listener in each of the four different masker edited conditions: M3/0 (circles), M2/0 (triangles), M0/1 (diamond), M0/0 (squares) and also in the full sentence control (asterisks).*

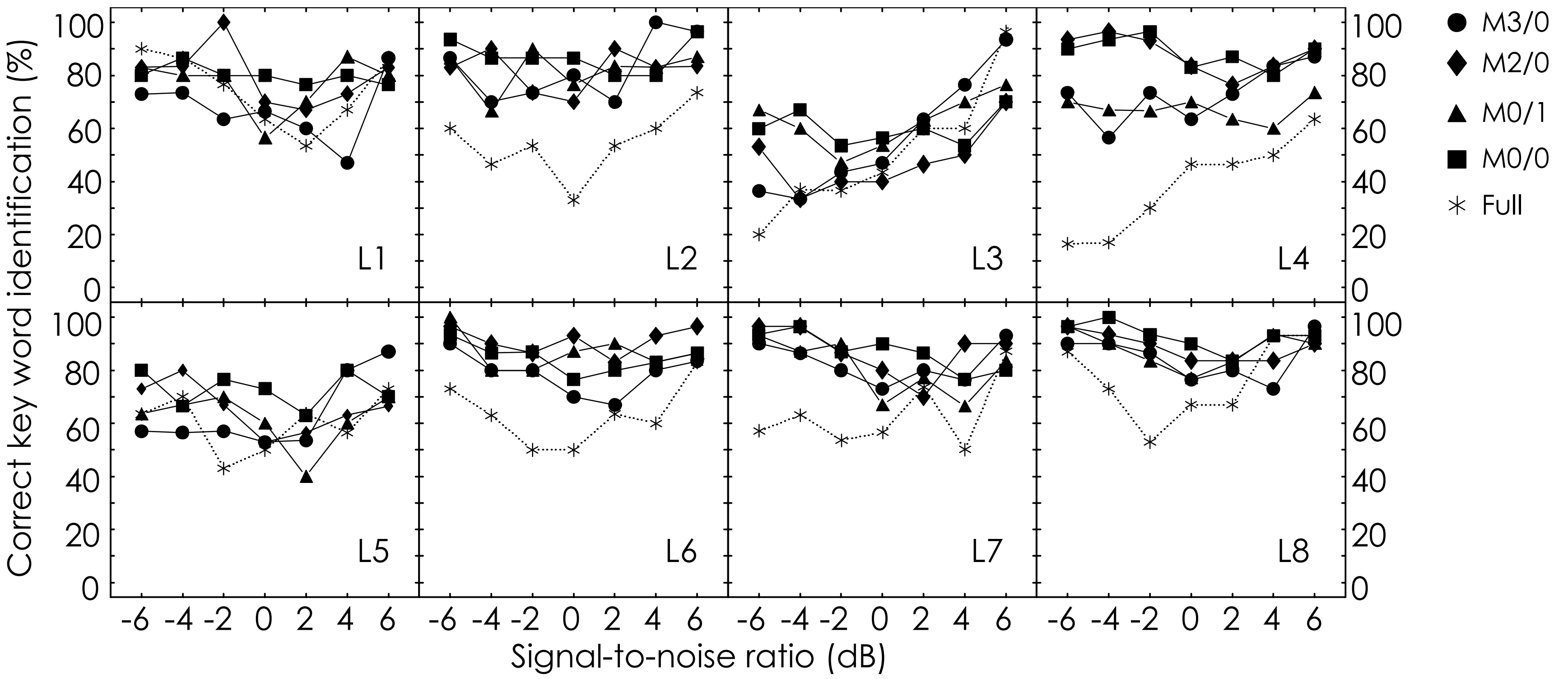
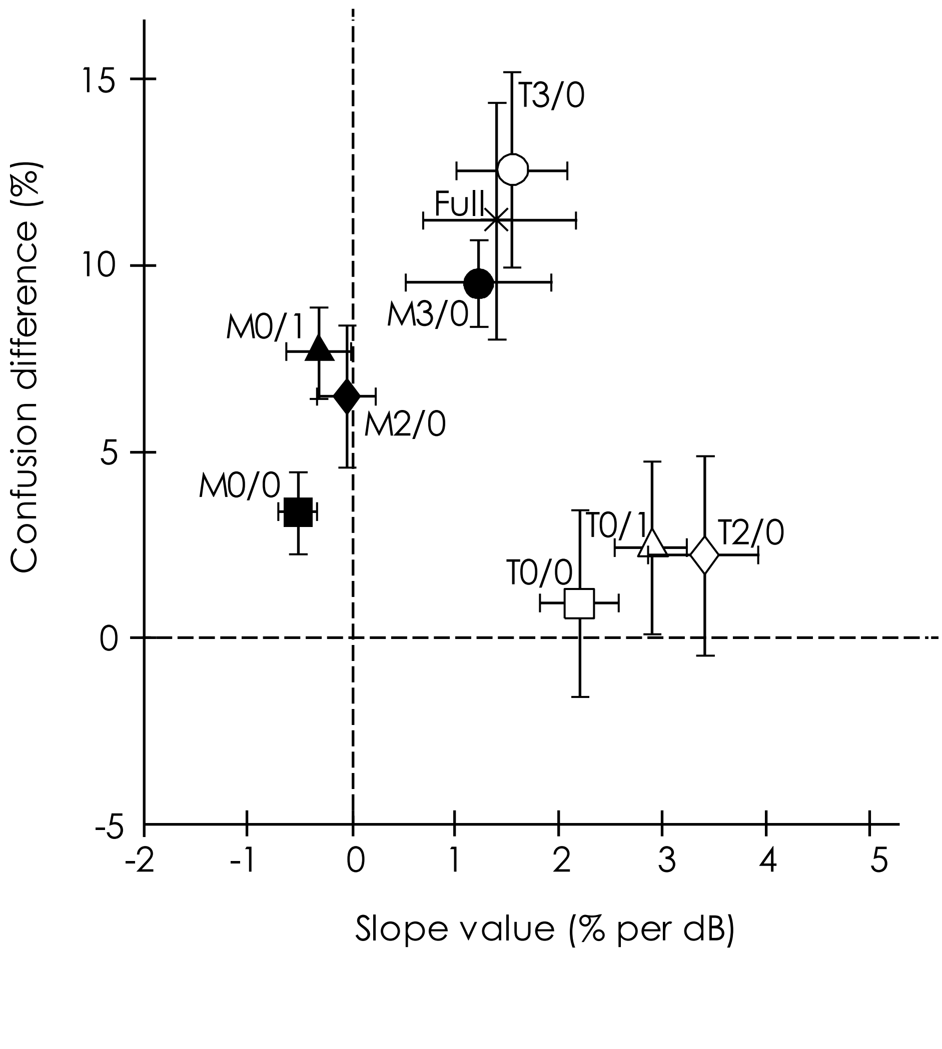


***Figure 5.5:*** *Mean slope* plotted against mean CD values for the four target-edited conditions (open symbols), the four masker-edited conditions (closed symbols), and the full-sentence control condition (asterisks). The error bars are the 95% confidence intervals for these values.*



### 5.2.1 Effect of editing the target

A one-way repeated measures ANOVA was carried out to establish the effect of type of target edit (T0/0, T0/1, T2/0, T3/0, or full) on the slope* of the psychometric function. Table 5.1 displays mean slope* values for these conditions and Table 5.2 displays the results of the ANOVA. A significant main effect of type of target edit was found and so planned contrasts compared slopes* for each of the four edit conditions with those of the full sentence control. Slopes* for the T0/1 and the T2/0 edits were both found to be significantly steeper than the full-sentence control but no significant difference was found between the slopes* of the T0/0 or the T3/0 conditions and the full condition.

***Table 5.1:*** *Mean slope values for the four different target edit types.*

| Condition | Mean (% per dB) | Standard deviation |
|---|---|---|
| T0/0 | 2.3 | 1.1 |
| T0/1 | 3.0 | 0.9 |
| T2/0 | 3.4 | 1.2 |
| T3/0 | 1.5 | 1.6 |
| T4/0 (full) | 1.4 | 2.1 |

That the slope* for the T0/0 edit length did not differ significantly from the full-sentence control was surprising, especially as less extreme edits (T0/1 and T2/0) did lead to the expected increases in the slope. It can be seen from looking at the individual psychometric functions in Figure 5.3, however, that performance in this condition for all but 3 listeners (L1, L3 and L5) was already above 80% at the lowest SNR. It is likely, therefore, that slopes* values were low because performance had reached a ceiling level rather than being due to difficulties distinguishing between the target and the masker as was the case for the full sentence control condition. A very low CD value (i.e. an eradication of the U-shaped dip) for this condition, which is discussed below, further supports this view.

A one-way repeated measures ANOVA was also carried out with CD values as the dependent variable, to examine the effect of type of target edit on the monotonicity of the psychometric function. Again a significant main effect was found and planned comparisons indicated that CD values were significantly smaller for the T0/0, T0/1, and T2/0 edits than they were for the full-sentence control. Smaller CD values for these edited conditions suggest that the dip in the function was significantly reduced or eradicated in these conditions. CD values were not found to be significantly different from the full condition for the T3/0 condition, however, suggesting that the dip in the function endured in this condition.

***Table 5.2:*** *Summarises the effect that editing the target sentence had on both slope* and CD Values.*

| Effect | | Results |
|---|---|---|
| Main effect of *target* edit type on slope*. | | $F(4,28) = 4.53$, $P < 0.01$, Partial $\eta^2 = 0.02$ |
| Planned contrasts - comparison to slope* in full condition (T4/0): | | |
| | T0/0 | $F(1,7) = 1.10$, $P = 0.33$, Partial $\eta^2 = 0.14$ |
| | T0/1 | $F(1,7) = 5.81$, $P < 0.05$, Partial $\eta^2 = 0.45$ |
| | T2/0 | $F(1,7) = 6.85$, $P < 0.05$, Partial $\eta^2 = 0.50$ |
| | T3/0 | $F(1,7) = 0.31$ $P = 0.86$, Partial $\eta^2 = 0.00$ |
| Main effect of *target* edit type on CD values. | | $F(4,28) = 5.39$, $P < 0.01$, Partial $\eta^2 = 0.47$ |
| Planned contrasts - comparison to CD values for full condition (T4/0): | | |
| | T0/0 | $F(1,7) = 5.80$, $P < 0.05$, Partial $\eta^2 = 0.45$ |
| | T0/1 | $F(1,7) = 5.66$, $P < 0.05$, Partial $\eta^2 = 0.45$ |
| | T2/0 | $F(1,7) = 10.12$, $P < 0.05$, Partial $\eta^2 = 0.59$ |
| | T3/0 | $F(1,7) = 0.31$ $P = 0.59$, Partial $\eta^2 = 0.04$ |

### 5.2.2 Effect of editing the target compared to editing the masker

A two-way repeated measures ANOVA was carried out to compare slope values measured in each of the four edit types for both target-edited and masker-edited stimuli. As the aim of this analysis was to establish whether editing had different effects depending on whether the target or the masker was edited, the full-sentence control was excluded from this analysis. The results demonstrated a significant effect of sentence edited with target-edited conditions giving significantly steeper slopes than masker-edited conditions. No significant effect of type of edit was found but the interaction between edit type and sentence edited was found to be significant.

Figure 5.6 shows this interaction. It indicates that editing the target instead of the masker resulted in a significantly different slope* values for all edit types except 0/3. An analysis of the simple main effects supports this with edited sentence found to have had a significant effect on slope* for the 0/0, 0/1, and 2/0 edits but not for the 3/0 edit.

### 5.2.3 Response errors

Looking at the keyword responses made by listeners gives an insight into the degree of confusion between the target and the masker that occurred in each condition and how this changed as SNR was increased. Figure 5.7 shows the mean proportion of colour and number keyword responses which matched the target (i.e. correct), or which matched the masker (i.e. a confusion error), or which matched neither the target nor the masker (i.e. a random error). The figure shows these responses at three SNR ranges (-6 to -4, -2 to 2, 4 to 6), for the full-sentence control condition, the target-edited conditions (averaged over all the four target-edited conditions), and the masker-edited conditions (averaged over all the four masker-edited conditions). It can be seen that listeners very rarely responded with a random error. This again supports the concept that dips in performance seen in the CRM-in-CRM paradigm are not down to audibility but to difficulties telling the target and the masker apart. This corresponds to the findings of Brungart (2001a), where it was also noted that very few random error were made around 0 dB in a CRM-in-CRM paradigm. The current

***Figure 5.6****: The interaction between edit type and edit length.*

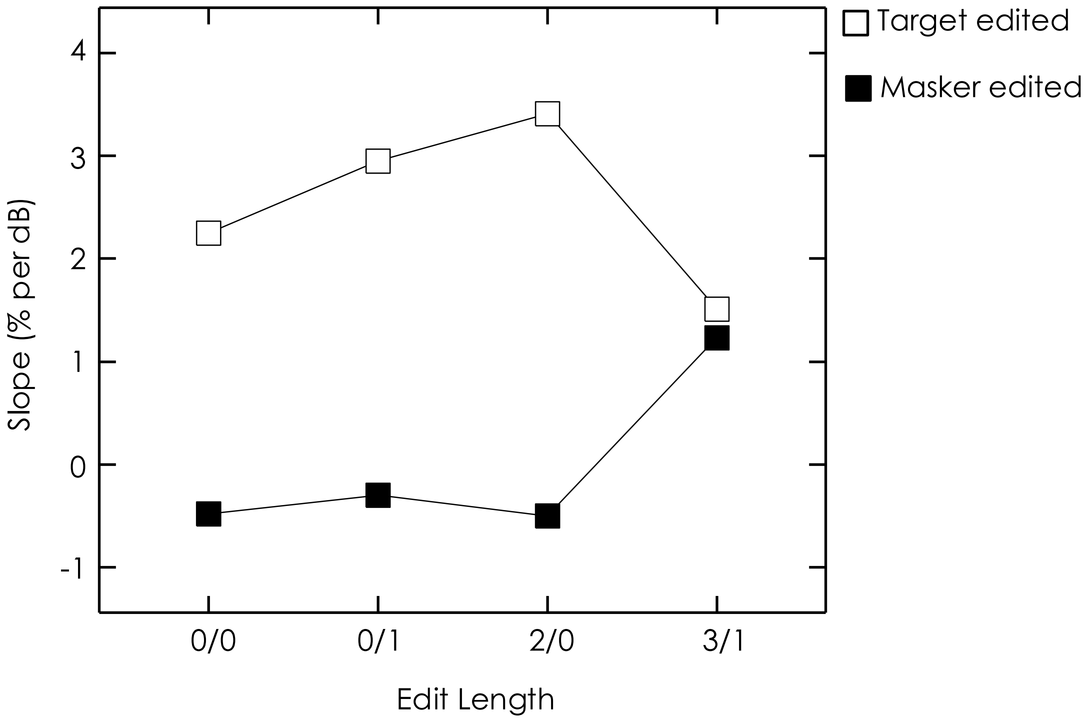
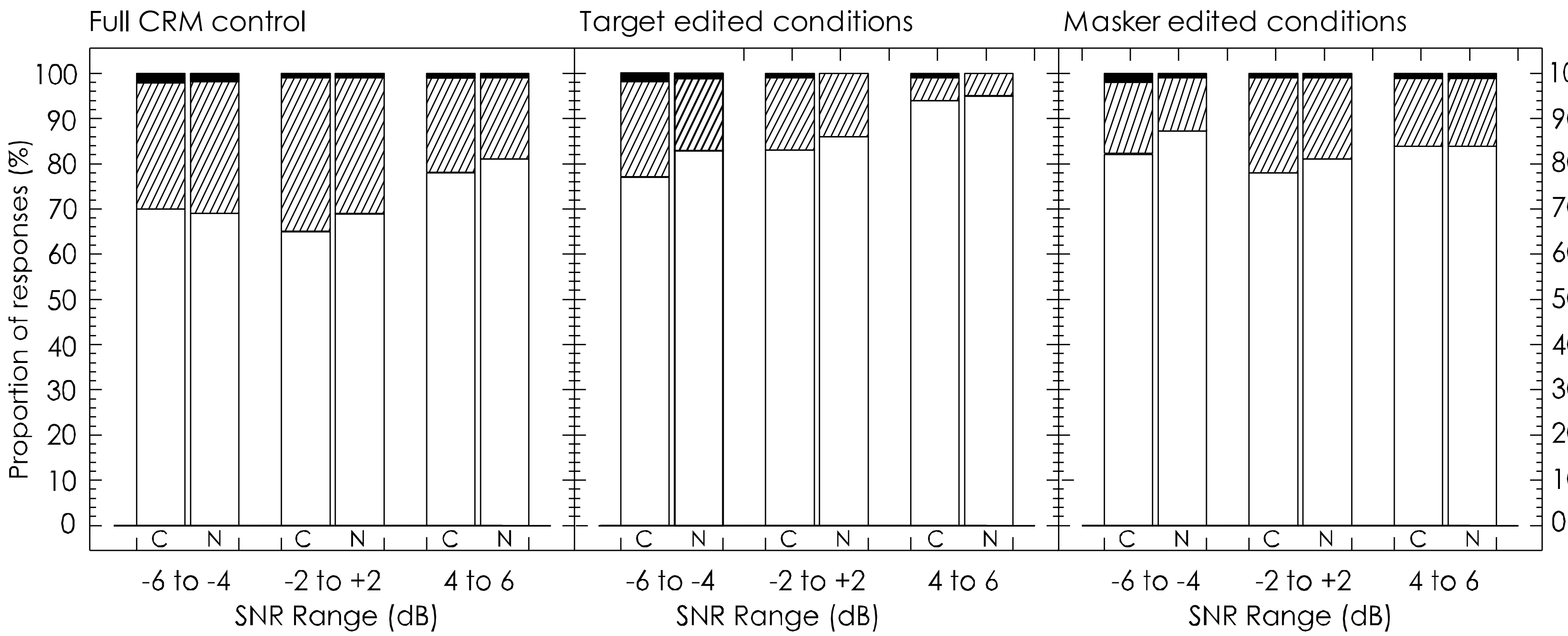


***Figure 5.7:*** *Listener responses made in each condition at three different signal-to-noise ranges. Responses are shown as the proportion of colour (C) and number (N) keywords taken from the target sentence (white bars), the masker sentence (hatched bar), or from neither the target or the masker sentence (black bars).*



results indicate that the proportion of confusion errors decreased (i.e. the number of correct responses increased) if either the target or the masker were edited.

The manner in which the confusion errors changed as SNR was increased was also different between the three conditions shown in Figure 5.7. For the control condition, the proportion of confusion errors stayed relatively constant across the lowest to mid SNR ranges, with slightly fewer confusion errors – on average, 11% − being made at the highest SNR range. This suggests that while increasing SNR from -6 to 2 dB had little effect on reducing target/masker confusions in this condition, increases at more favourable SNRs (4 to 6 dB) did lead to a reduction in the number of confusions being made. This partial independence of performance from SNR was not seen for the target-edited conditions, where increasing SNR led to increases in performance and decreases in the number of confusion errors being made (an average decrease of 5% from low to mid range SNR and a further decrease of 11% from mid to high range SNR). The masker-edited condition gave another pattern of confusion errors again. Confusion errors remained relatively constant but increased slightly – by on average 6% − at the mid-range SNRs and levelled out again at higher SNRs.

### 5.2.4 Summary of results

- For almost all listeners, editing either the target or the masker improved performance on the task.
- When compared to the slopes of psychometric functions measured in a full CRM control, editing the target to either the T0/1 or T2/0 edit types significantly increased the slope, but longer (T3/0) and shorter (T0/0) edit types did not.
- Confusion difference values for all target-edited conditions but the longest (T3/0) were found to be significantly smaller than those calculated for the full sentence control, suggesting the dip in the function was greatly reduced in these conditions.

- A significant interaction was found between edit type and the sentence edited, with significant differences in slope* seen between target and masker-edited conditions for all edit types except T3/0.
- Keyword responses showed a different pattern of target/masker confusions as a function of SNR for the full-sentence control, the target-edited and the masker-edited conditions.

## 5.3 Discussion

The current experiment was designed to test whether improving a listener's ability to select the target sentence could reduce target/masker confusion and eradicate the unusual shaped psychometric functions that this confusion can produce. The CRM-in-CRM paradigm has been found to reliably produce U-shaped/plateaued psychometric functions. It was hypothesised that when using this paradigm, increasing the salience of the target by editing it to just the keywords would steepen the slope of the function, eradicating the U-shape. The results partly supported this hypothesis; CD values for the T0/0 condition were much reduced compared to the full sentence control, suggesting that the dip in the function was eradicated. Nevertheless, the mean slope for this condition did not differ significantly from that of the control, but this shallowness was attributed to a ceiling effect as target/masker confusion had been greatly reduced.

It was also hypothesised that the slope of the psychometric function would progressively decrease and the U-shape would return as the words preceding the keywords were added back into the target sentence. Shorter edit lengths (T0/1 and T2/0) were indeed found to give steeper slopes than the full sentence control, while the longer edit length (T3/0) did not, supporting the hypothesised pattern for results.

Further it was hypothesised that editing the masker would have differential affects to editing the target sentence. In other words, if selective attention does play a role in reducing target/masker confusion then directing attention away from the target towards the masker should have a different effect on slope. This was again supported by the results; slopes were found to be significantly shallower in the masker-edited condition than in the target-edited condition at three out of four of the experimental edit lengths.

### 5.3.1 Directing attention to the target

The same-talker CRM-in-CRM paradigm offers few cues by which to tell the target and masker messages apart; targets and maskers are acoustically, semantically and linguistically similar. There are several opportunities for confusion to arise; even if individual auditory objects are grouped and successfully formed, these still need to be correctly streamed and the correct stream to be identified before the listener can successfully report both the colour and number keywords (Shinn-Cunningham & Best, 2008). By editing the target to just the keywords, we increased the salience of the target sentence. The abrupt onset of the target involuntarily drew listeners' attention and so provided a cue to which sentence was the target. It has been suggested that selective attention can enhance desired signals and suppress unwanted or interfering ones, thus improving listeners' ability to selectively attend to the target CRM utterance in turn reduced confusion between it and the masker. This attention cue was sufficient to allow the target to be identified and selected even when it was presented at roughly the same level as the masker (i.e. at -2, 0 or 2 dB) and as a result no dip in performance was seen at this level and the psychometric function returned to a normal monotonic shape. The underlying psychometric function for the listening conditions, i.e. the one that would be given if confusion was removed, was revealed.

For longer edit lengths (e.g., T3/0), psychometric functions were on the whole still non-monotonic, suggesting that improving the salience of the target was no longer a useful cue for distinguishing between the target and the masker and listeners were once again relying on a level difference to make this distinction. We hypothesise that this was because after the abrupt onset had drawn listeners' attention to the target, the subsequent words before the keywords provided an opportunity for the listener to lose track of which sentence was the target sentence. This suggests that the abrupt onset was only successful in improving selective attention over short durations.

### 5.3.2 Directing attention away from the target

Performance on the CRM-in-CRM task was also found to improve if the masker, not the target sentence, was edited. We presume this improvement was the result of an improved ability to distinguish the target from the masker. Listeners may have, for example, focused attention on the target and suppressed the abruptly starting masker sentence. Alternatively, attention could have been involuntarily drawn to the masker and the masker sentence identified. To identify the target, attention could then have been switched to the other message, replaying and recalling information from short-term memory to retrieve the target keywords if they had been missed during the switching of attention (Pashler, 1998). Listeners seemed to be extremely adept at using this method; mean performance across the masker-edited conditions was 77%. Nevertheless, switching of attention from the masker *back* to the target is likely to take time (Moray, 1969). There is further support for such a mechanism in the response errors: while colour and number errors do not materially differ, there is a tendency for slightly more confusion errors to be made on the colour keyword, i.e. the first keyword, than on the number keyword (3% on average).

None of the four masker-edited conditions showed a return to a monotonic psychometric function; the functions for these conditions stayed relatively constant over the full range of SNRs. The M0/0 condition gave the flattest slope (M = -0.48% per dB) and while, as mentioned above, performance was relatively high in all masker-edited conditions, performance hovered around this level and for most listeners never reached full identification (i.e. 100%). One explanation for this constancy of performance is that while listeners were relatively good at using the prominent masker in this condition to identify the target, the masker did still occasionally distract them. Crucially, however, unlike the target-edited and full conditions where this distraction was dependent on the level of the signal distraction in the masker-edited conditions were essentially independent of SNR: performance could be affected at any SNR, and averaging this variation in performance gave a flat function (Wichmann & Hill, 2001b). The proportion of confusion errors made at each SNR supports this theory. For the target-edited conditions the proportion of confusion errors decreased as the SNR increased suggesting a reduction in confusion.

For the masker-edited conditions, however, at the lower and higher SNRs the proportions of confusion errors were the same. The fact that confusion errors were just as likely to occur at high SNRs as they were at lower ones in the masker-edited condition suggests that they may be accounted for by lapses in concentration or inappropriately focussed attention on random trials. The unusual, plateau shaped functions given in the masker-edited cases can be thought of, therefore, as distinct from those given in the full CRM condition as they result from a different slope-change mechanism. The plateaued slopes in the full CRM condition are the result of target/masker confusion while those in the shorter masker-edited conditions are more likely to be the result of underlying variation in the decision making process (i.e. due lapses in concentration or attention).

### 5.3.3 Additional findings

The 0/1 condition was included to establish whether some of the difficulties in the CRM-in-CRM paradigm were the result of a suffix effect (Crowder, 1967). It was hypothesised that the 0/0 and 0/1 conditions would be equivalent in terms of their ability to draw attention (both start at the keywords), but that only the 0/1 contained the suffix word "now". Therefore, if the suffix did interfere with performance then scores in the 0/1 condition should be less than in the 0/0 condition. Figures 5.3 and 5.4 show that for several listeners their performance was indeed poorer in the T0/1 and M0/1 conditions than in the T0/0 and M0/0 conditions, although the detriment between the two edit types was largest when the target was edited (on average 4.8% at each SNR) compared to when the masker was edited (on average 2.8% at each SNR). The results were, therefore, consistent with the suffix effect.

Nevertheless, it is not entirely clear cut. The difference in performance between the 0/0 and 0/1 conditions was strongest at 0 dB; a reduction of 7.8% in performance was seen for both the M0/1 and T0/1 conditions at 0 dB. This hints that the benefit offered by the 0/0 over the 0/1 condition is due to a reduction in confusion rather than a removal of the suffix effect. The addition of a suffix should, arguably, have had the same effect on performance at each SNR, however, it had the greatest effect at 0 dB where confusion was at its highest. Another explanation for the difference in performance between the 0/0 and 0/1 conditions could be, therefore, that the offset of the keywords, as well as the onset, can be used as a cue to direct attention. If on one trial the onset cue was missed and the listener was still unsure which sentence was the target the offset cue could still be used in the 0/0 condition to identify it, giving the listener the chance to switch attention if need be and recall from memory the correct message. Better performance in the T0/0 and M0/0 conditions than in the T0/1 and M0/1 conditions can, therefore, be attributed to the additional selection cue.

### 5.3.4 Summary

Experiment 3 demonstrated that directing listeners' attention to either the target or the masker sentence in the CRM-in-CRM paradigm could improve performance and reduce target/masker confusion. The slope of the psychometric function was dependent on which sentence had been edited. Progressively editing the target led to an eradication of the U-shaped dip and steepened the slope, showing that improving listeners' ability to selectively attend to the target reduced target/masker confusion. Editing the masker, however, led to extremely flat psychometric functions. This flattening was attributed to lapses in attention rather than an inability to decide where to direct selective attention, as is the case in the standard CRM-in-CRM paradigm. It can be concluded, therefore, that due to a lack of top-down cues, difficulties selecting the target sentence, as opposed to difficulties segmenting or streaming segments of the sentence, resulted in the confusion-based, U-shaped/plateaued psychometric functions characteristic in the same-talker CRM-in-CRM task at SNRs around 0 dB.

As this experiment was carried out on young normal hearing listeners, it is possible that the results seen rely on the listeners being able to successfully suppress distracting information, rapidly switch attention and recall degraded information from short-term stores. It is plausible then that different results may be seen for older and hearing impaired listeners. This is considered in the next experiment.

## 5.4 Experiment 4: Attentional effects on confusion-based shallow slopes for older and hearing-impaired listeners

It was demonstrated in Experiment 3 that improving a listener's ability to selectively attend to a target − either directly by increasing the salience of the target or indirectly by making the masker more distinguishable – reduced target/masker confusion and had a large effect on the slope of the resultant psychometric function. It is possible that these "attentional" effects on slope may, however, be altered somewhat for older listeners. For a listener to selectively attend to a particular sound in a mixture, for example, they must be able to enhance the representation of the sound of interest (Shinn-Cunningham, 2008). There is evidence to suggest, however, that age related changes may lead to difficulties selecting and enhancing a particular sound source in a mixture.

It is well documented that older listeners tend to experience peripheral hearing degradations such as a reduction in spectral and temporal resolution (Lutman, 1991; Moore, 1995; Schneider, 1997). In terms of understanding speech-in-noise, these changes mean a reduced ability to make use of brief dips in the power of masking speech to identify the target and a reduced ability to segregate target and masker voices based on frequency. The cues usually used to group individual sounds together − for example, onsets, offsets and harmonic structures − may also become less distinct for these listeners (Leek & Summers, 2001). These factors tend to lead to a more ill-defined perception of the auditory scene; target and masker utterances become more perceptually similar and the features on which top-down information can be focused become less distinct. If this is the case then providing a cue or a prime will be less effective and as a result, older listeners may have greater difficulties listening selectively in noisy environments. If a degraded peripheral input due to aging leads to less effective selective attention then it is likely that older listeners with a hearing-impairment will have even greater difficulties than older listeners with normal hearing.

As well as being able to enhance a target source, efficient selective attention also relies on being able to suppress the sound sources which are not of immediate interest. It was demonstrated in Experiment 3 that young, normal-hearing listeners were able to use an increased salience of the masker sentence to identify the target sentence. It was argued that listeners were able to do this by either focusing attention on the target and suppressing the abruptly starting masker sentence, or by allowing attention to be drawn to the masker but then rapidly swapping attention back to the target sentence, and recalling from memory any information that had been missed. There is evidence to suggest that these aspects of selective listening may also be challenging for older listeners.

Older listeners have been found to have greater difficulties ignoring irrelevant interfering speech (Schneider, Daneman, & Murphy, 2005; Tun et al., 2002). This would suggest that they may find it harder to suppress sound sources which are not of primary focus. The greater difficulty that older listeners experience selecting a target sound source is likely to have concomitant effect on their ability to rapidly swap attention. It has been suggested that each time attention is shifted, grouping and selection processes are reset (Macken, Tremblay, Houghton, Nicholls, & Jones, 2003). Swapping attention could, therefore, be much more cognitively costly for older adults, and it is likely that they will miss parts of the target message as they do so. Any parts of the message that are missed could be recalled from memory, analogous to mentally replaying, but at this stage again, older listeners may be at a disadvantage. Memory traces decay over time and it is possible that the slower processing experienced by older listeners may make these memory traces less useful; the target message may be unintelligible, for example, by the time it comes to being recalled (e.g. Brown, 1958). These difficulties may be further exacerbated for older listeners with hearing impairment. If sensory input is already degraded when it enters temporary memory stores, as is it is likely to be for hearing-impaired listeners, then further time spent in a volatile memory store may quickly result in it being unintelligible (Mackersie, Boothroyd, & Prida, 2000; Pichora-Fuller, Schneider, MacDonald, Pass, & Brown, 2007).

In summary, there is evidence to suggest that older listeners may find it hard in complex listening situations to enhance sound sources of interest, suppress interfering sound sources and rapidly swap attention between two or more sound sources. If this is the case we might expect that manipulations to improve selective attention may not always be successful and so in turn will not be as effective at reducing target/masker confusion. In other words, improving selective attention may not be an effective way of eradicating the unusual shaped psychometric functions given in the CRM paradigm for older listeners as it was in the younger listeners.

To test this hypothesis, the current experiment used a similar method to that used in Experiment 3. To reduce the experiment time, a reduced subset of conditions was tested; the shortest edit type (T0/0) for both sentences (target and masker) were again compared to a full-sentence CRM control condition. As in Experiment 3, psychometric functions were measured in each of these edited conditions. It was expected that if improving listeners' ability to selectively attend to the target did reduce target/masker confusion, a reduction in the confusion difference (CD) value and an increase in the slope* of the psychometric function would be seen. Two groups of older listeners were tested; a group of audometrically normal-hearing listeners and a group of hearing-impaired listeners. It was expected that the experimental manipulations of selective attention would be less successful at reducing confusion between the target and masker sentences for the hearing-impaired listeners.

## 5.5 Method

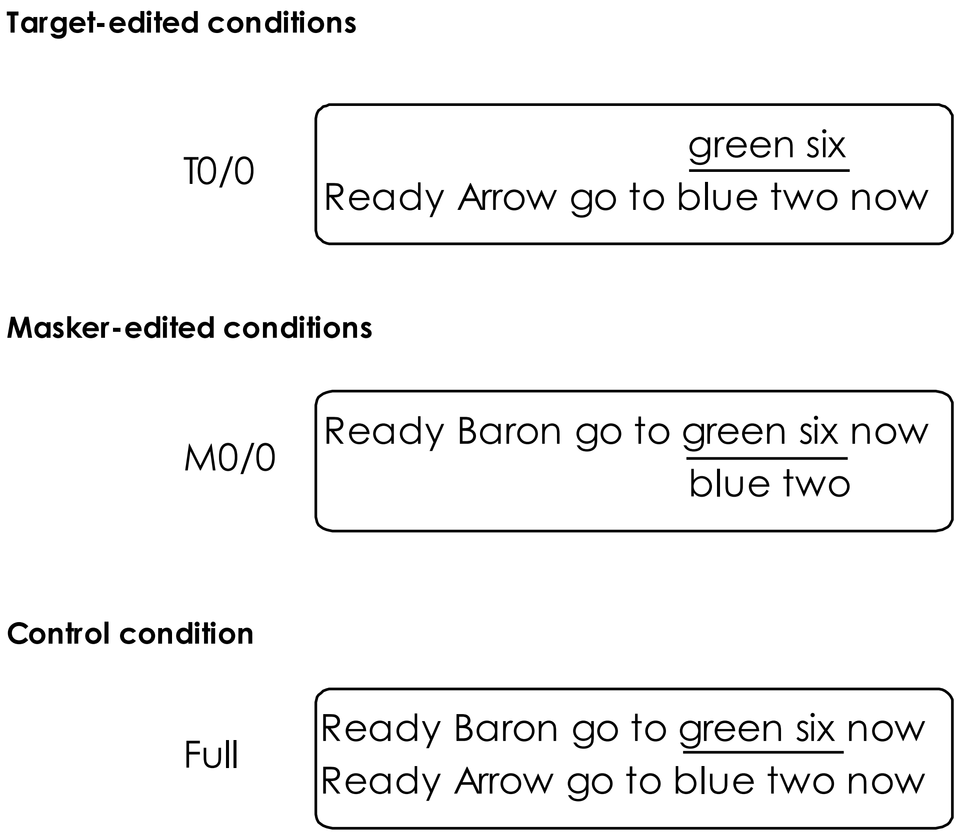
### 5.5.1 Listeners

23 listeners took part in Experiment 4. Their ages ranged from 61 to 78 (mean age = 71) and were recruited from the MRC Institute of Hearing Research's volunteer panel. Listeners' hearing levels were assessed using pure-tone audiometry. The group had better ear, four-frequency average (BEA) thresholds ranging from 7.5 dB HL to 60 dB HL (mean hearing loss = 35 dB HL). Eight of the listeners were classed as having normal hearing (BEA < 25 dB HL), while the remaining 15 listeners were classed as having mild to moderate hearing loss (25< BEA >60 dB HL).

### 5.5.2 Stimuli

Three of the nine conditions used in Experiment 3 were selected and used in Experiment 4: T0/0, M0/0 and the full sentence control (see figure 5.8).

***Figure 5.8:*** *The three stimulus conditions used in Experiment 4. A target-edited condition (T0/0), a masker-edited condition (M0/0), and a full sentence control condition (Full).*



The masker was again presented at 70 dB SPL and the target sentence varied parametrically within blocks to give 8 different SNRs: -8, -6, -4, -2, 0, +2, +4 and +6 dB. An extra SNR (-8 dB) was tested in this experiment to accommodate the expectation that if the older listeners were to exhibit the same U-shaped functions that were displayed by the younger listeners in some conditions in Experiments 2 & 3, then the dip in the function would be wider.

As before, stimuli were presented diotically over headphones. The same sound-treated booth and equipment used in the previous experiments were again employed.

### 5.5.3 Procedure

The listeners' task was much the same as in the previous experiments; i.e. to identify the colour and number which corresponded to the target sentence. Each listening block consisted of 120 trials (= 15 trials at each SNR). Listeners completed two blocks of each of the three conditions over two, 1 hour sessions. One block of the full sentence condition was always completed at the start of each session, followed by two blocks of either the T0/0 condition or the M0/0 condition. The T0/0 and M0/0 blocks were fixed within a visit so that only one set of target instructions needed to be given per session, thus minimising confusion for the listener. The order of these blocks were, however, counter-balanced across listeners so that half of the listeners completed the T0/0 blocks on their first visit and the M0/0 blocks on their second visit and visa versa for the remaining listeners.

## 5.6 Results

Figure 5.9 shows the eight normal-hearing older listeners' individual psychometric functions for the three stimulus conditions (the full condition, the target-edited condition and the masker-edited condition). Listeners 1, 2, 3, and 6 all exhibited either U-shaped or plateaued functions, centred around 0 dB SNR in the full-sentence condition. Listeners 4, 5, 7, and 8, however, showed more normal functions: performance in this condition was very low below SNRs of -2 to 0 dB but rose steeply as the signal was increased past this level for these listeners. For all listeners, editing the target gave an improvement in performance, with performance for many listeners approaching ceiling level for most SNRs. For the listeners who exhibited a U-shaped function in the full-sentence condition, editing the target led to either a reduction (listeners 1 and 6) or a complete eradication (listeners 2 and 3) of this dip in performance. Editing the masker also had a large effect on performance. Again performance for all listeners improved in this condition compared with that of the full sentence condition (the only exception being at the highest SNR, 6 dB). U-shaped functions were also greatly reduced, although for several listeners psychometric functions stayed flatter in this edit condition than in the target-edited condition (e.g. listeners 2, 4 and 8).

Figure 5.10 shows the corresponding results for the hearing-impaired older listeners. Only four of these listeners had clear U-shaped functions in the full sentence condition (listeners 10, 11, 12, and 13), the remaining listeners showed more monotonic functions. Like the normal-hearing older listeners, editing the target led to improved performance for all listeners. For those listeners who produced U-shaped functions in the full sentence condition, the dip was greatly reduced in the target-edited condition. For most listeners editing the masker did give an improvement in performance, at lower SNRs at least, but the shape of the psychometric function was much more variable across listeners than was seen for the full and target-edited conditions. On the whole, however, the functions for the masker-edited condition were generally flatter than in the other conditions.

Figure 5.11 shows the across-listener mean psychometric function for each of the three conditions for the normal hearing (panel A) and the hearing-impaired (panel B) listeners. Note that averaging the individual psychometric functions leads to the portion of the slope between -8 and 0 dB being artificially flattened (i.e. rather than a clear U-shape). For both listening groups editing the target removed this plateau at lower SNRs, although performance in this condition reached a higher level for the normal-hearing older adults. The function for the masker-edited condition was flat for both listening groups but overall performance in this condition was higher in the normal hearing group (around 80% at all SNRs) than in the hearing-impaired group (between 40 and 55% at all SNRs).

***Figure 5.9:*** *Individual psychometric functions for each of the normal-hearing older listeners in each of the three conditions; the full control (asterisks), the target-edited condition (open squares) and the masker-edited condition (closed squares).*

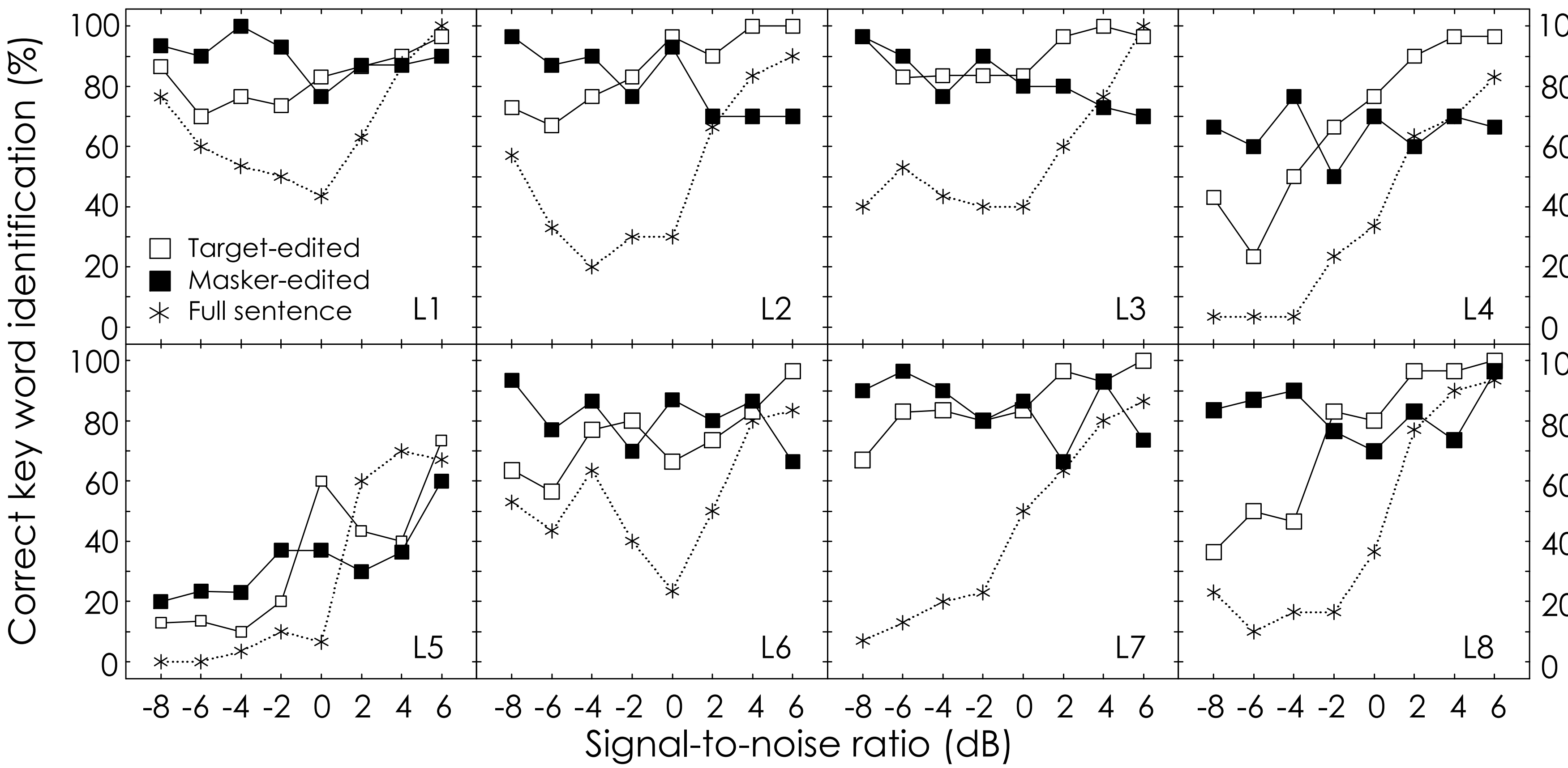


***Figure 5.10:*** *Individual psychometric functions for each of the hearing-impaired older listeners in each of the three conditions; the full control (asterisks), the target-edited condition (open squares) and the masker-edited condition (closed squares).*

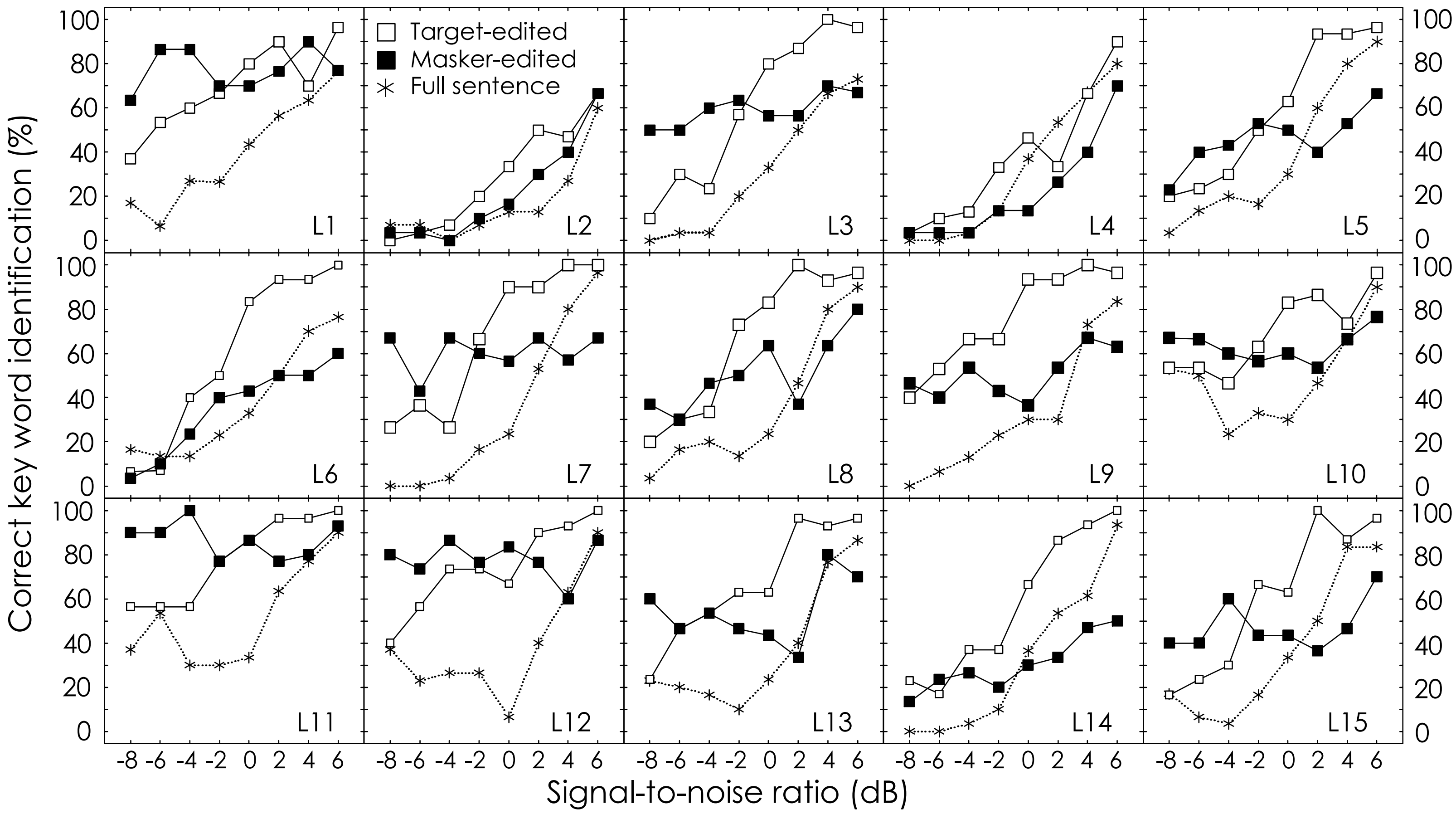
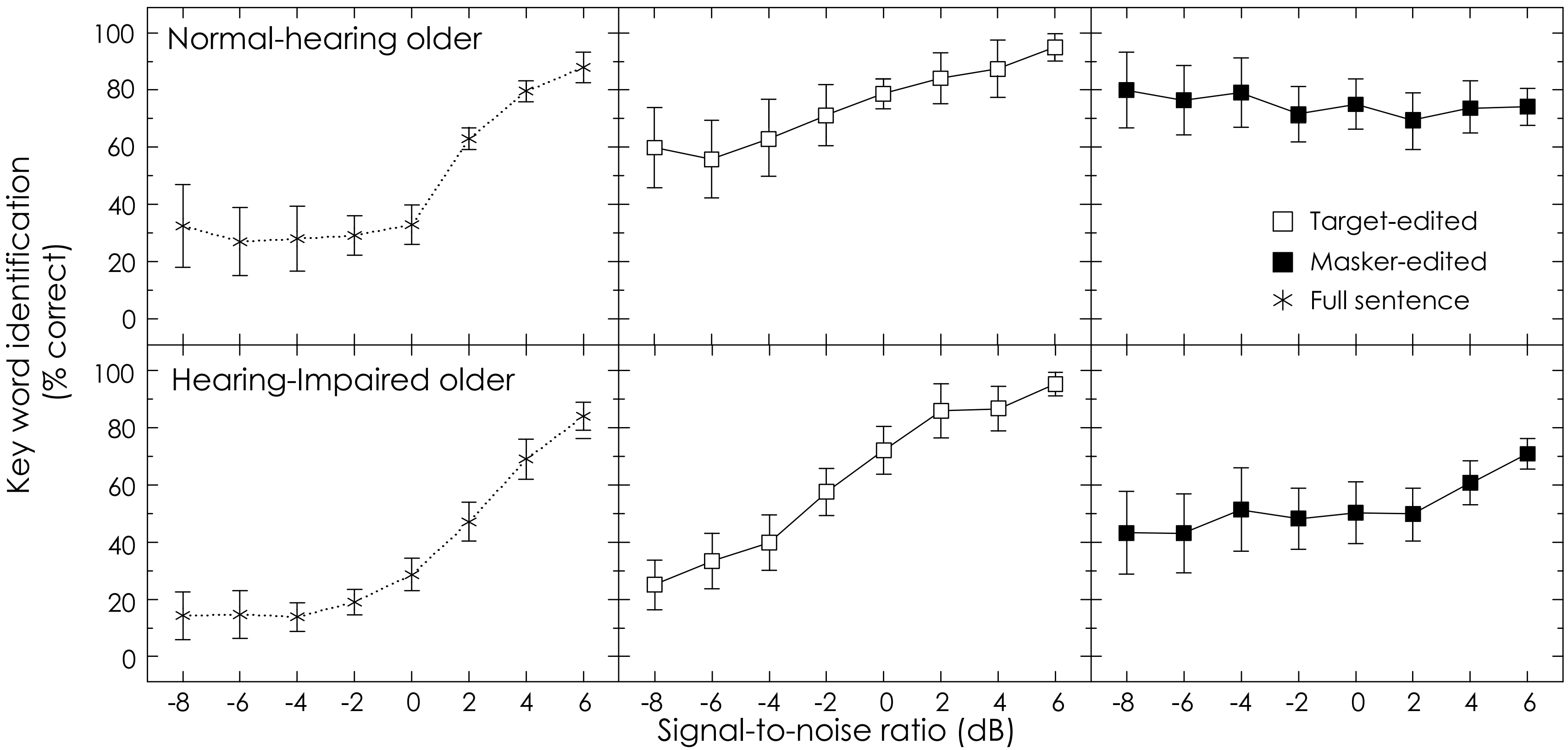


***Figure 5.11:*** *Mean psychometric functions for the target-edited (open squares), masker-edited (closed squares) and full sentence control conditions (asterisks), averaged across normal-hearing older listeners (top panels) and hearing-impaired older listeners (bottom panels).*



As in Experiment 3, slope* and CD values were calculated for each psychometric function measured. Figure 5.12 plots mean slope* and CD values for the target and masker-edited conditions averaged separately across listener groups. The figure shows a similar trend for both listener groups. The full-sentence condition gave moderate slopes* (around 5% per dB) but very large CD values (around 20%); i.e. large dips in performance were seen around 0 dB for this condition. Editing the target led to little change in slope* values (a decrease of 2% per dB for normal-hearing listeners and no change at all for hearing-impaired listeners) but to a large reduction in CD value (20% for normal-hearing listeners and 26% for hearing-impaired listeners), suggesting the dip in performance had been largely eradicated. The negative CD values seen in this condition for both sets of listeners suggests that at 0 dB, where the CD values were calculated, performance was actually greater than would have been estimated from the regression line fitted to calculate slope*. It is likely that this was due to performance approaching ceiling before or around 0 dB and then reaching an asymptote. Editing the masker led to a reduction in slope* and CD values for both normal (16% per dB and 12%) and hearing-impaired listeners (12% per dB and 16%). Slopes were on average shallower for the normal-hearing listeners in this condition; the negative slope value suggests that performance actually slightly decreased with increased level of the target.

### 5.6.1 Effect of editing the target compared to editing the masker

A 2x3 ANOVA was carried out to look at the effect of listener group (normal or hearing impaired) and the sentence edited (target, masker, or full) on the slope* of the psychometric function. Table 5.5 displays mean slope* values and Table 5.6 displays the results of the ANOVA. Significant main effects of both listener group and the sentence edited were found. A significant interaction between the two variables was also found.

The interaction is shown in figure 5.13. In order to interpret this interaction the effect of the sentence edited was considered for the two listener groups separately. A significant effect of sentence edited was found for both the normal-hearing and the hearing-impaired older adults. Post-hoc pairwise comparisons (with a Bonferroni correction) indicated that for the normal-hearing listeners, editing either the target or the masker resulted in a significant reduction in slope* compared to the full sentence control (p < .05 and p < .001 respectively). Further, the masker-edited condition was found to give significantly shallower slopes* than the target-edited condition (p < .01). For the hearing-impaired listeners, the masker-edited condition was still found to give significantly shallower slopes* than either the full sentence (< .001) or target-edited conditions (p < .001). No significant difference in slope* was found for these listeners however, between the target-edited and the full sentence condition (p = 1.00). These results suggest that for the normal-hearing listeners, editing either the target or the masker sentence led to shallower psychometric functions. For the hearing-impaired listeners, only editing the masker significantly affected the slope*.

***Table 5.5:*** *Mean slopes* and standard deviations (% per dB) for hearing impaired and normal hearing listeners in each of the three sentence edited conditions (full-sentence, target edited and masker edited).*

| | Full-sentence | Target edited | Masker edited |
|---|---|---|---|
| Normal hearing | 4.5 (1.7) | 2.8 (1.6) | -0.4 (1.3) |
| Hearing Impaired | 5.2 (1.5) | 5.1 (1.4) | 1.9 (1.5) |

A 2x3 ANOVA was also carried out to look at the effect of listener group and the sentence edited on the CD values. A significant effect of sentence edited was found but CD values were not found to differ significantly between the two listener groups and no significant interaction between the two variables was found. Pairwise comparisons with a Bonferroni correction were carried out to further look at the effect of edit type. The results indicated that CD values were significantly reduced for both the target-edited and masker-edited conditions compared to the full sentence condition (significant p < 0.001). CD values were also significantly lower for the target-edited condition than they were for the masker-edited condition (p < 0.01). These results suggest that the dip in performance seen when two full CRM sentences were used could be reduced by editing either the target or the masker sentence. The results also suggest, however, that editing the target was the most effective way to improving performance around the 0 dB SNR range.

***Figure 5.12:*** *Mean slope* plotted against mean CD values for the three conditions; the full-sentence control (full), the target-edited condition (T0/0) and the masker-edited condition (M0/0). Panel A shows mean data for normal-hearing listeners and panel B shows mean data for hearing-impaired listeners. The error bars are the 95% confidence intervals for these values.*

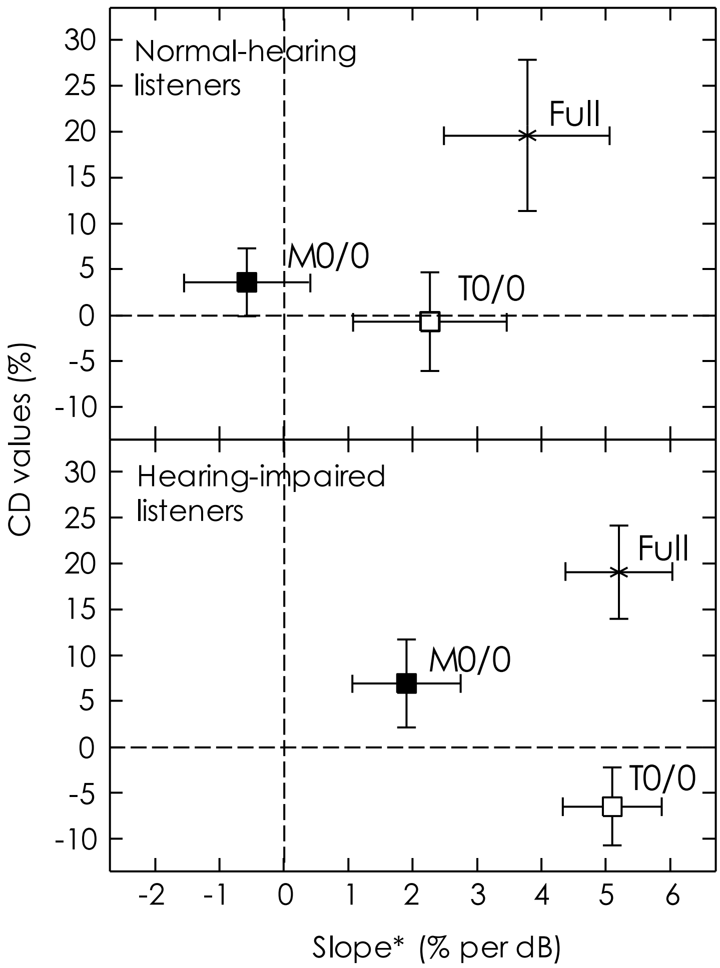


***Figure 5.13:*** *The interaction between listener group and edit type.*

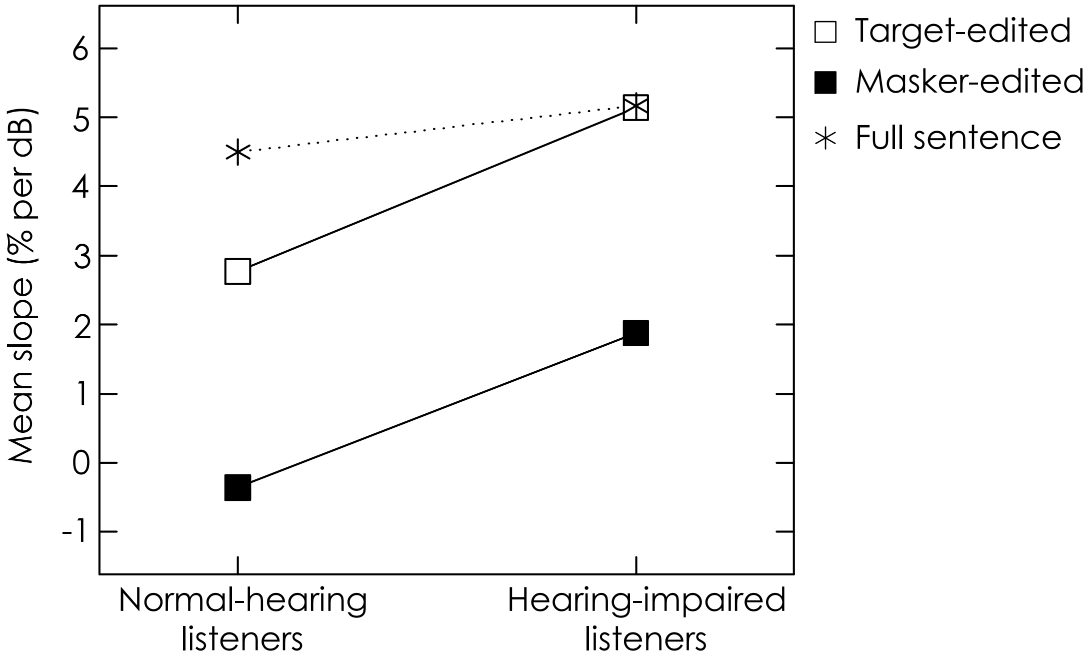


***Table 5.6:*** *Summarises the effect that editing the target sentence had on both slope* and CD Values.*

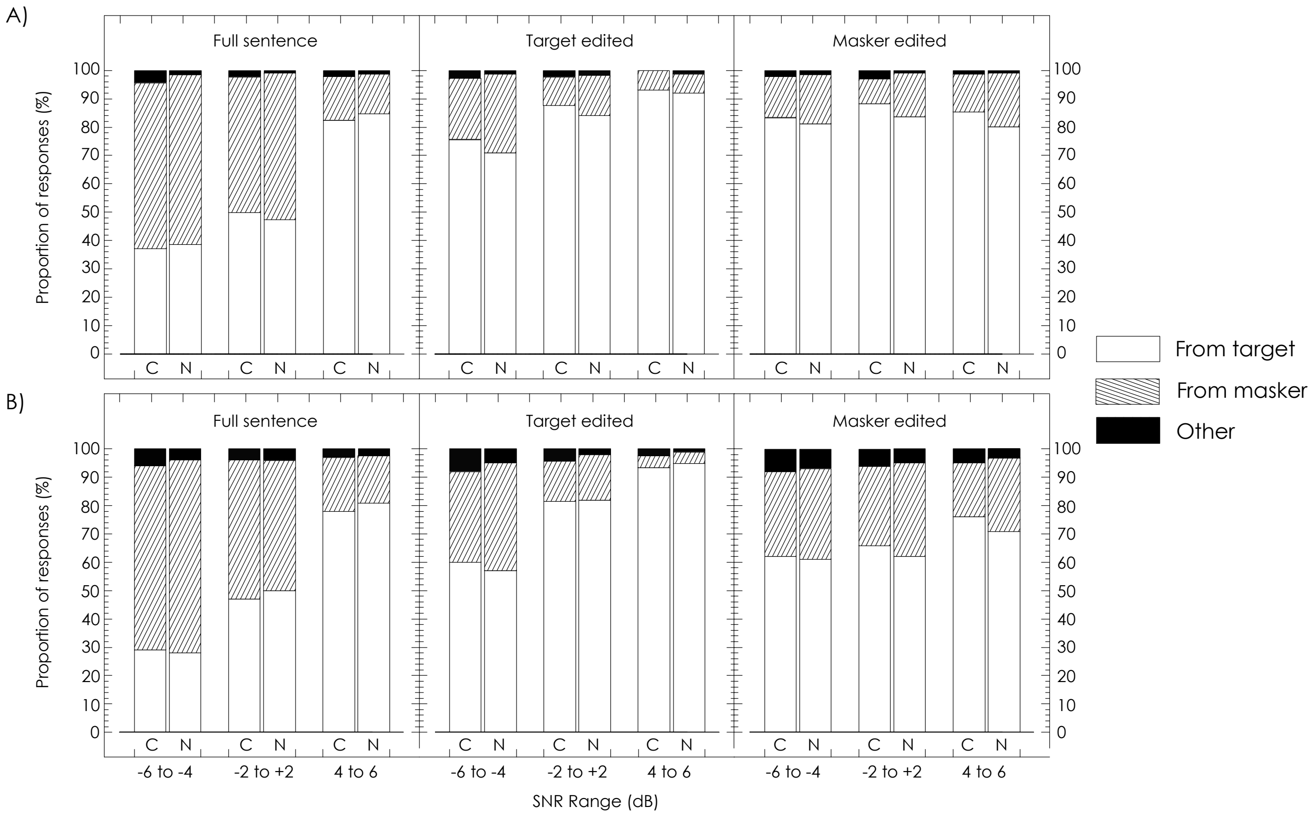
| Effect | | Results |
|---|---|---|
| Main effect of listenergroup on slope*. | | F(1,21) = 90.7, $P < 0.001$, Partial $\eta^2 = 0.81$ |
| Main effect of the sentence edited on slope*. | | F(1,42) = 10.5, $P < 0.01$, Partial $\eta^2 = 0.33$ |
| Listener group x sentence edited interaction. | | F(2,42) = 4.39, $P < 0.05$, Partial $\eta^2 = 0.02$ |
| Simple main effects of sentence edited at each level of listener group: | Normal Hearing | F(2,28) = 49.47, $P < 0.01$, Partial $\eta^2 = 0.87$ |
| | Hearing impaired | F(2,14) = 48.32, $P < 0.001$, Partial $\eta^2 = 0.78$ |
| Main effect of the sentence edited on CD. | | F(2,42) = 46.98, $P < 0.001$, Partial $\eta^2 = 0.69$ |
| Main effect of listenergroup on CD. | | F(1,21) = 0.21, $P = 0.65$, Partial $\eta^2 = 0.02$ |
| Listener group x sentence edited interaction. | | F(2,42) = 1.86, $P = 0.17$, Partial $\eta^2 = 0.08$ |

### 5.6.2 Response errors

Figure 5.14 plots the proportion of listener keyword responses where either the colour or number was present in the target sentence (i.e. a correct response), present in the masker sentence (i.e. a confusion error) or not present in either the target or the masker sentence (i.e. a random error). These proportions are shown at three different SNR ranges (-6 to -4, -2 to +2 and +4 to +6) for each of the three edit conditions (full sentence, target-edited and masker-edited). Responses are also shown separately for the normal-hearing (panel A) and hearing-impaired listeners (panel B).

For the full sentence condition, both listener groups made a high proportion of confusion errors, particularly at the low and mid SNR ranges. At these ranges, over 50% of listeners' responses were made up of the colour or the number from the masker keyword, suggesting listeners had difficulty ignoring the masker (which was the more intense sentence). The hearing-impaired listeners made slightly more confusion errors than the normal-hearing listeners did at each SNR range. For both listener groups, as SNR was increased the number of confusion errors steadily decreased. A similar pattern was seen in the target-edited condition as the SNR was

***Figure 5.14:*** *Listener responses made in each condition and by each listener group at three different target-to-masker ranges. Responses are shown as the proportion of colour (C) and number (N) keywords taken from the target sentence (white bars), the masker sentence (hatched bar), or from neither the target or the masker sentence (black bars).*



improved. The proportions of confusion errors at each SNR range were, however, substantially smaller for the target-edited condition than they were for the full sentence control, suggesting that editing the target made it easier for the listeners to ignore the more intense sentence. Smaller proportions of confusion errors were also seen when the masker was edited, however unlike the target-edited condition these errors did not seem to decrease as SNR increased for either listening groups. Instead, the proportion of confusion errors remained relatively constant over all three SNR ranges. As was seen in the other conditions, regardless of SNR range, the hearing-impaired listeners made a higher proportion of confusion errors than the normal-hearing listeners did.

### 5.6.3 Summary of results

- For both groups of listeners, editing the target or the masker improved performance on the task. Editing the target sentence was, however, most effective at improving performance around 0 dB.

- When compared to the full sentence control condition, editing either the target or masker sentence generally decreased the slope* of the psychometric function, except in the target-edited condition for the hearing-impaired listeners where there was no difference in slope* found between it and the full-sentence control.

- CD values for both groups of listeners were significantly smaller for the target-edited and masker-edited conditions than they were for the full sentence control. This suggests that editing greatly reduced any dips or plateaus in performance around 0 dB SNR.

- Response errors showed a different pattern of confusion errors for the full sentence, target-edited and masker-edited conditions. Hearing-impaired older listeners were also more susceptible to confusion errors than their normal-hearing contemporaries.

## 5.7 Discussion

The aim of Experiment 4 was to test the effect that selectively attending to either the target or the masker sentence in the CRM paradigm had on the degree of target/masker confusion experienced by normal-hearing and hearing-impaired older adults. As in Experiment 3, selective attention was manipulated by editing either the target or the masker sentence to just its keywords. It was proposed that a reduction in target/masker confusion would be evident by an increase in the slope of the psychometric function, and/or decrease in the CD value (degree of non monotonicity), and a reduction in the number of confusion errors made.

### 5.7.1 Effect of directing attention to the target

It was found that improving listeners' ability to attend to the target sentence did indeed reduce confusion between it and the masker sentence. When compared to the full sentence condition, CD values were found to be substantially reduced for the target-edited condition for both hearing-impaired and normal-hearing listeners, suggesting any plateaus or dips in performance had been eradicated. The number of confusion errors produced by listeners supported this view, with less confusions being made as the SNR was increased for the target-edited condition. These results suggest that the abrupt onset of the target was sufficiently salient to direct older listeners' attention to the target sentence and further, that improved selective attention significantly reduced the amount of target/masker confusion experienced.

It was expected that if directing listeners' attention to the target did reduce target/masker confusion then this would be reflected in a steepening of the psychometric function as well as a reduction in CD values and confusion errors. This was not found to be the case, however: no significant change in slope* was seen for older hearing-impaired listeners and slopes* were actually found to be shallower for the normal-hearing listeners when the target sentence was edited from full. As was suggested for the equivalent condition (T0/0) in Experiment 3, this decrease in slope for the older normal-hearing listeners was likely to be due to performance reaching a ceiling level when the target sentence was edited (performance for 5 of the 8 normal-hearing listeners in Experiment 4 was above 70% for at least 6 of the 8 measured

SNRs). The non-significant change in slopes* between the full and target-edited conditions for the older hearing-impaired listeners, however, may be the result of floor rather than ceiling effects. Hearing impaired listeners tended to perform poorly at low SNRs but well at high SNRs. Whilst editing the target eradicated shallow areas in the middle of the function (hence the significant reduction in CD value), the overall slope* of the function remained largely unchanged. So while not directly increasing the slope of the function, we can argue that editing the target sentence does uncover the underlying acoustic function for the situation once confusion has been removed.

### 5.7.2 Effect of directing attention to the masker

It was also found that improving the salience of the masker sentence improved performance on the CRM task. It has been argued that older listeners would find it particularly hard to ignore a salient distracter, such as the abrupt onset of the masker (Tun et al., 2002) and would also be slow at swapping attention back to the target once distracted (Shinn-Cunningham & Best, 2008). It was also suggested that, for the hearing-impaired listeners in particular, a degraded input would lead to a degraded memory trace making it harder to recall unattended or missed keywords (Mackersie et al., 2000). It was reasoned that these factors combined would make it difficult for older listeners to use an improved salience of the masker sentence in order to identify the target sentence. This was not, however, found to be completely the case. As with the target-edited condition, CD values were found to be significantly smaller, and fewer confusion errors were made when the masker sentence was edited than when two full CRM sentences were presented. There was an indication, which will be discussed in more detail below, that hearing-impaired listeners may still be partially affected by these difficulties, but in general the results suggest that older listeners were, in fact, able to make use of improved masker salience to reduce confusion with the target.

As with the target-edited condition, the reduction in confusion did not result in a steepening of the psychometric function as slopes* for the masker-edited condition were again found to be significantly shallower than those for the full sentence condition. Figure 5.11 shows that for the masker-edited condition, the psychometric functions were flat over the whole SNR range. As discussed in Experiment 3, this suggests that distraction for this condition was independent of SNR and instead the result of occasional diversions of attention. Interestingly, while the shape of the functions produced in this condition were relatively similar for both groups of older listeners, the threshold for the function is shifted to the right in the older normal-hearing listeners compared to the older hearing-impaired listeners. This suggests that while hearing-impaired listeners were able to use the masker to identify the target − this is evident by the improved performance on the masker-edited condition compared to the full condition for these listeners, particularly at lower SNRs – they found the condition more distracting than normal-hearing listeners did. Greater difficulties in ignoring an abruptly starting sound may explain this difference between listener groups, but it is also possible that a hearing-impairment reduces a listener's ability to recall the unattended stream from auditory memory. If input is already degraded due to hearing loss it may quickly become too degraded to recall from temporary memory stores (Shinn-Cunningham & Best, 2008). If this is the case lapses in attention would be less easily recovered from by hearing-impaired listeners and this may explain the lower threshold for these listeners in this condition.

### 5.7.3 Summary of Experiment 4

Experiment 4 demonstrated that manipulating selective attention did have an effect on the degree of target/masker confusion experienced by older and hearing-impaired listeners. As was seen in Experiment 3, the *shape* of the function depended on the manipulation, i.e. which sentence had been edited. While editing the target saw a return to a monotonic function for both groups of listeners, editing the masker sentence resulted in flat psychometric functions. The reduction in confusion seen from editing the target sentence was equivalent for both listening groups, suggesting that this was effective at improving selective attention for both listening groups. The overall differences in the thresholds for functions produced in the masker-edited condition, however, gave an indication that this manipulation of selective attention was less effective for hearing-impaired older listeners than it was for normal hearing older-listeners.

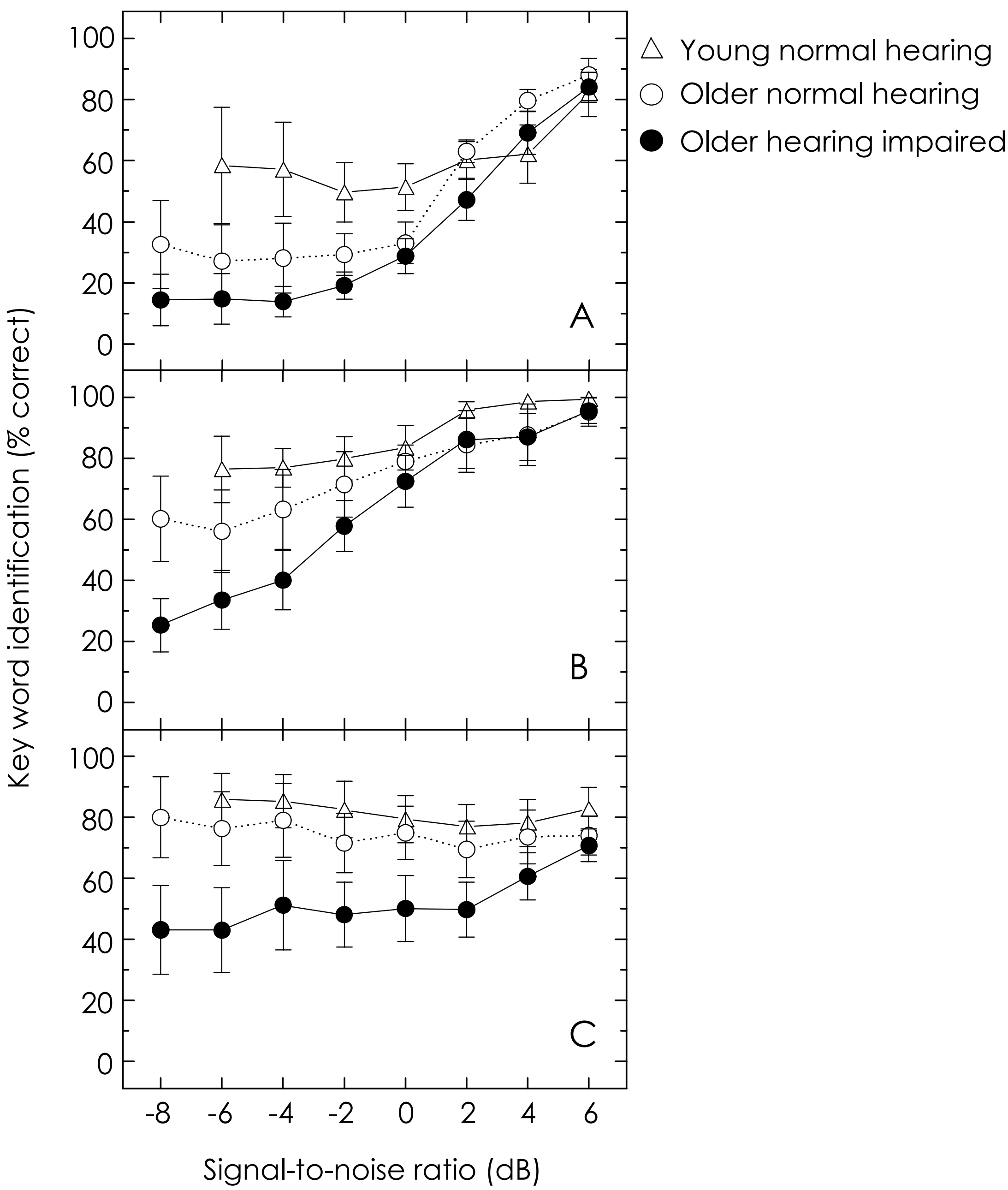
## 5.8 Experiment 3 vs. Experiment 4 – Comparing attentional effects on slope for three listener groups.

Experiments 3 and 4 were both designed to consider whether selective attention could play a role in steepening the slope of confusion-based psychometric functions. This section aims to briefly compare the results for the three groups of listeners tested and look at trends in the data common to all three listener groups.

Figure 5.15 shows the mean psychometric functions averaged across listeners for the three groups: young normal-hearing, older normal-hearing, and older hearing-impaired. Each panel shows the average psychometric function for the target-edited, masker-edited and full sentence control conditions. For the young normal-hearing listeners who were tested on several different edit types in Experiment 3, only the 0/0 edit type is considered here. It can be seen that there were three different shapes of functions for the three conditions which hold across the three listener groups – unusual (full condition), shallow (target-edited) and flat (masker-edit). There was also little difference between the threshold of these functions for the young and older normal-hearing listeners for the two edited conditions. The biggest difference for these groups was between the functions given by the full sentence control. This may suggest that older listeners, while able to make use of the attentional manipulations to reduce target/masker confusion, were initially more susceptible to this confusion than the younger listeners were. The psychometric functions for the full condition were, conversely, the most similar in terms of threshold for the older normal-hearing and older hearing-impaired listeners, suggesting similar degrees of confusion for these listeners in this condition. It was the masker-edited condition in particular where the biggest threshold difference occurred between the hearing-impaired and normal-hearing (older and younger) listeners. Performance dropped by about 20% for the hearing-impaired compared to the normal hearing listeners for this condition, suggesting that these listeners were less able to use the increased salience of the masker as a cue to identifying the target.

For all three listener groups the psychometric functions for the target and masker-edited conditions crossover at around -2 dB SNR. At low SNRs performance was at a higher level for the masker-edited condition than for the target-edited condition, but

***Figure 5.15:*** *Mean psychometric functions for the young normal-hearing (open triangles), older normal-hearing (open circle) and older hearing-impaired listeners (closed circle) in the full-sentence control (panel A), target-edited (Panel B) and maker-edited conditions (Panel C).*



at high SNRs this was reversed and performance was higher in the target-edited condition. Further this effect was not simply an artefact of averaging over listeners as the same pattern can be seen in many of the listeners' individual psychometric functions (see figures 5.5, 5.9, and 5.10). This is an interesting result as it seems at odds with the reasoning behind the experimental design. The masker-edited condition should, in theory, distract listeners' attention away from the target sentence while the target-edited condition should direct listeners' attention towards the target sentence. Even if, as was postulated earlier in the chapter, listeners were able to deduce the target from improved salience of the masker we would not expect this to lead to better performance at any SNR than directing attention to the target directly would give. This unexpected difference between the two conditions is yet more surprising as it occurs at low SNRs. At this level the masker is at its most favourable level and so, one would expect, at its most distracting.

To explain this, it is argued that there must be some other cue present in the masker-edited condition salient enough to override distraction from a more intense, abruptly-starting masker and further that this cue is not present in the target-edited condition. In the masker-edited condition the listener is able to gauge at what level the sentence is being played and, therefore, at which level to expect the target keywords before the masker starts. Freyman, Balakrishnan and Helfer (2004) demonstrated that priming the target had the ability to greatly improve performance by providing cues on which selective attention can be focused. It is possible that in the masker-edited condition the acoustic characteristics of the target are primed as it is heard on its own before the start of the masker (see figure 5.8). No such cue is available in the target-edited condition where the target is not heard on its own before the masker starts. Further, the SNR differences were created by adjusting the level of the *target* relative to the masker, i.e. the masker was always presented at the same level. Assuming the listener learns to make an internal reference level for the masker then there is an expectation that in the masker-edited condition, not only can the listener gauge the level of the target, but they can also decide whether the distracting sound is likely to be louder or quieter than the target. This top-down information may help to maintain focus on the target and aid suppression of the distracting masker. In the target-edited condition where the masker starts first, this extra information is not available;

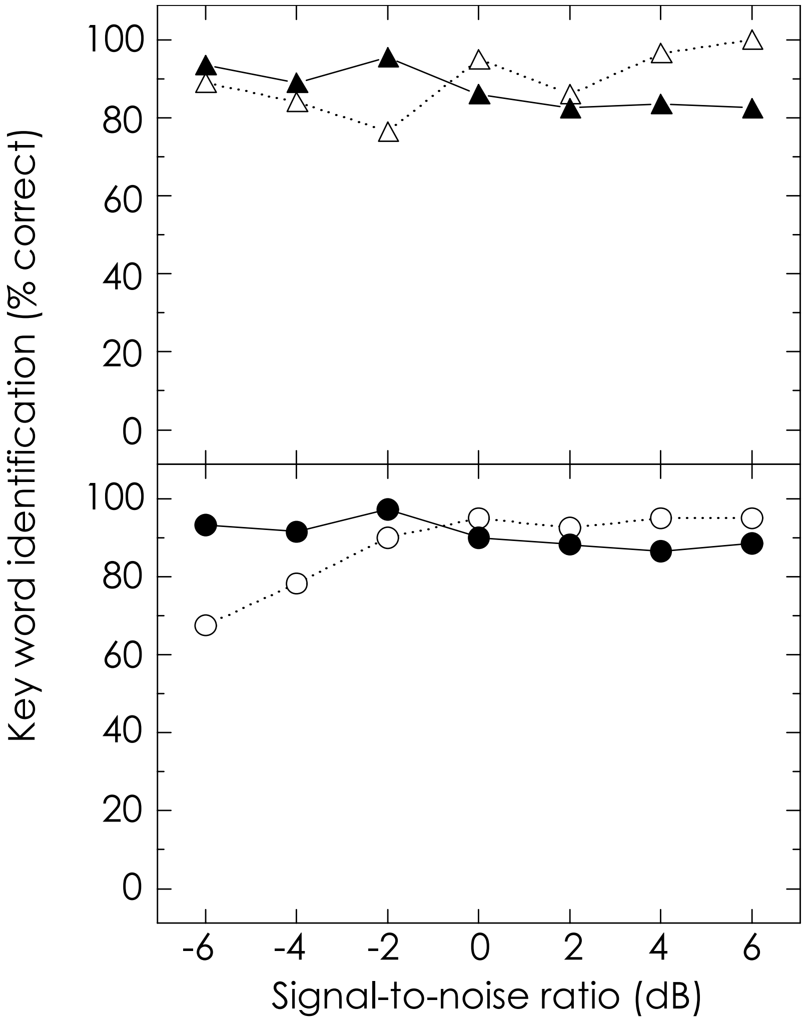
gauging the level of the initial sentence does not tell the listener anything about the level of later-starting sentence. In short, the masker-edited condition may provide additional cues to allow both the enhancement of the target sentence and the suppression of the masker sentence.

One experimental test of this concept would be to rove the level over which the masker sentence is presented. This would reduce at least one of the extra cues afforded in the masker-edited condition. The masker would not now always be heard at the same level and the listener would not be able to gauge the level of the target in relation to their internal representation of the masker. Hearing a quiet target would not now necessarily indicate a louder masker as it would have done previously. A short follow-up test was run to test this. It was expected that when the masking level is roved the advantage offered by the M0/0 condition at lower SNRs would be reduced and performance would more closely match that of the T0/0 condition.

Seven-point psychometric functions (SNRs of -6, -4, -2, 0, 2, 4, and 6) were measured using the target-edited (T0/0) and masker-edited conditions (M0/0). The procedure used in Experiments 3 and 4 was followed, but instead of the masker presentation level being fixed at 70 dB, six different presentation levels, roved over a 10 dB range, were used (from 66 to 76 dB in 2 dB intervals). Listeners completed two blocks; one for each edited condition. Listeners completed 18 trials at each SNR (3 trials at each of the 6 masker presentation levels), thus each block consisted of 128 trials in total. Edit type was fixed within a block but masker presentation level and SNR were randomised within a block. Two young normal-hearing listeners and two older normal-hearing listeners took part.

Results are shown in Figure 5.16. The top panel show the results for the younger normal-hearing listeners. For these listeners at lower SNRs there was little difference in performance between the two edit types but performance is very close to ceiling meaning there is little headroom to make this judgment. The lower panel shows the results for the older normal-hearing listeners. For these listeners performance only reached ceiling for the masker-edited condition, performance in the target-edited condition was still below that seen for the masker-edited condition at lower SNRs. Roving the level of the masker did not, therefore, seem to drop performance in the

***Figure 5.16:*** *Mean psychometric functions for two young normal-hearing (top panel) and two older normal-hearing (bottom panel) given for the masker-edited (closed symbols) and target-edited conditions (open symbols) when presentation level was roved over a 10 dB range.*



masker-edit case to match that of the target-edit condition. These results suggest that the benefit offered by the masker-edit condition compared to the target-edited condition at lower SNRs was not based on an ability to prejudge the level of the interrupting message.

## 5.9 Summary

Experiments 3 and 4 demonstrated that improving listeners' ability to selectively attend to the target sentence in the CRM-in-CRM paradigm could reduce target/masker confusion and reveal the underlying psychometric function for the situation. These experiments provided further evidence for confusion based slope-change mechanisms and highlighted the role that difficulties selecting a target − rather than segregating or streaming a target − played in their occurrence. The experiments showed similar slope patterns for all three listener groups; younger normal-hearing, older normal-hearing and older hearing-impaired all providing evidence to suggest that listeners were able to use the improved salience of either a target or masker to enhance the selection of a target when presented with an acoustically and linguistically similar masker. Nevertheless, there was evidence to suggest that older hearing-impaired listeners tended to be more susceptible to distraction than normal-hearing listeners, be they younger or older, supporting previous findings that this group do find it harder to ignore irrelevant interfering information.

While older listeners did not show a marked deficit in rapidly swapping attention or recalling portions of missed target speech from memory in the current experiment it is possible that the burdens of everyday speech might better highlight these difficulties. Experiments 5 and 6, therefore, considered the role that more complex, continuous speech has on the slope of the psychometric function for older listeners.

# 6 Experiment 5: Effect of a continuous speech target on the slope of the psychometric function.

The speech materials used to measure speech intelligibility in research and clinical practice are usually short utterances such as monosyllabic words (e.g. the NU No.6), digits (e.g. the triple digit test), or sentences of four to six words in length (e.g. the ASL sentences). These speech materials are commonly presented on a trial-by-trial basis, with one test utterance presented followed by a pause for the listener to respond before the next utterance is presented. These speech materials are useful tools for measuring intelligibility as many aspects of the test speech can be controlled for. Standardised lists can be, for example, phonetically balanced, equated for length, grammar and word frequency, and matched for linguistic content. Such corpuses are, therefore, easy to administer and score, and give good test-retest reliability. But one criticism of standardised speech materials and procedures, however, is that they lack the "naturalness" of the speech we encounter in everyday life (Bench, Kowal, & Bamford, 1979).

In everyday communication speech will often be continuous with only short pauses between words or sentences. When listening to the radio, the T.V. or to announcements, for example, speech is often rapid and the listener does not have the opportunity to stop the speaker and ask for clarification. Continuous discourse is more representative then of the speech encountered in everyday listening situations (Hirsh, 1954). With this in mind, the aim of Experiment 5 was to ascertain how the slopes of psychometric functions for continuous speech might differ from those given by standardised speech materials and procedures. Measuring psychometric functions with continuous targets will hopefully give a better indication of the magnitude of the intelligibility benefit that a listener will experience in normal listening environments from small increases in level.

Of the 909 psychometric functions measured in the systematic slope survey (see chapter 2), only seven functions were found that used continuous speech targets. These psychometric functions originated from just two studies. The first, carried out by Speaks, Parker, Kuhl, and Harris (1972), used short stories recoded for the purposes of the experiment. The target speech was presented monaurally in a static white noise. Psychometric functions with slopes of, on average, 7% per dB were reported. The second study, carried out by Jerger and Jordan (1992), used a 20 minute segment of a story. The continuous target was presented over a loudspeaker placed 1.6m away from one of the listener's ears and a delayed copy of the same speech was presented from a loudspeaker at the opposite ear while a multitalker babble was presented from a loudspeaker directly behind the listeners' head. A slope of 5.7% per dB was reported. These broadly match the median slope values of studies identified in the slope survey using similar maskers with short target stimuli (slope survey median = 7.4% per dB for 1 static-noise masker). The slope given for continuous speech in babble was markedly shallower, however, than the median of the survey for equivalent masker conditions (slope survey median = 9.1% per dB for multitalker babble presented in the free field). These results give an indication that, for speech maskers at least, the continuousness of the target speech may have an effect on slope. Neither of these studies reported a direct comparison of continuous speech to shorter speech stimuli, however, and as will be discussed in greater detail below, both studies used very different methods for measuring the intelligibility of continuous discourse which may have affected the results.

There are several possible ways in which it might be theorised that a continuous target might affect the slope of the psychometric function. An increase in the availability of top-down information is one mechanism by which slope changes are believed to occur. Miller, Heise and Lichten (1951) demonstrated, for example, that slopes were shallower for isolated words than they were for words presented in sentences. It is likely that when speech is presented in context, i.e. when words can be predicted from previous content, speech intelligibility becomes less influenced by bottom-up information alone and more influenced by top-down information. This shift means that small changes in SNR can be more effective, steepening the slope of the psychometric function (Bronkhorst et al., 1993; Pichora-Fuller et al., 1995). If short tokens give shallower slopes than longer tokens like sentences, it follows that an even longer target token, such as a few minutes of continuous discourse, would give steeper slopes still. Continuous speech is arguably more context rich than a single sentence. In usual trial-by-trial tasks sentences are presented one at a time and the preceding sentence will not be semantically related to the following sentence[22]. In continuous discourse, however, sentences are linked semantically, suggesting the content of previous sentences can aid the identification of subsequent sentences.

Other factors found to affect the slope of the psychometric function suggest that continuous speech targets might have the contrary effect on slope, with continuous speech targets giving shallower slopes than sentence targets. Lapses in attention have, for example been shown to flatten the slope of the psychometric function (Wichmann & Hill, 2001b). Similarly, it was found in Experiments 3 and 4 that a distracting masker could result in a completely flat psychometric function, while improving listeners' ability to direct their attention to a target increased the slope. For speech to be continuous it naturally has to be lengthier than sentence targets. As the stimuli span a longer period of time the listener will have to maintain their attention on the task for longer. The opportunity for lapses in attention will, therefore, become greater and so continuous tasks may be more susceptible to attentional factors than trial-by-trial tasks.

Continuous speech is also likely to be more cognitively taxing than single sentences presented one at a time. For listeners to successfully identify target speech, the individual speech sounds must first be formed into objects, streamed and then the correct stream selected and attended to. Cognitive processes can be employed to compensate for minor failures at each of these stages. Unsuccessfully segregated or masked speech sounds can be perceptually filled via "phonemic restoration" (Shinn-Cunningham & Wang, 2008; Warren, 1970) and speech missed due to lapses in, or misdirected, attention can be retrieved from temporary memory stores (e.g. Broadbent, 1958). These processes, however, put greater demand on cognitive resources and can slow down speech understanding. When speech is presented on a trial-by-trial basis listeners are given pauses in which to make their response. It has been argued that these pauses also give speech understanding processes an opportunity to "catch up" (Shinn-Cunningham & Best, 2008 page 9). In continuous

[22] One exception to this is the connected speech test (CST), where 10 related sentences form a paragraph. But in this test speech is still presented one sentence at a time with pauses for listener to repeat back the sentence they just heard (Cox, Alexander, & Gilmore, 1987)

speech where pauses are often unavailable, short, or far apart, listeners must simultaneously access and integrate stored material while processing ongoing speech. This may result in occasional difficulties keeping up with the rate of speech. Difficulties keeping up may result in "processing lapses" where words currently being spoken are missed because previously spoken words are still being streamed, selected, identified and integrated. Processing lapses, like lapses in attention, are likely to flatten the slope of the psychometric function; speech identification will become less dependent on level and more dependent on whether processing is able to keep up with the rate of speech. This is, in essence, an increase in the underlying variability on the task. Unlike attentional lapses, processing lapses are not entirely stimulus independent: more challenging listening situations would result in greater cognitive demands and slower speech processing, hence more processing lapses.

If continuous speech does have an effect on slope due to its more taxing nature, it would be expected that the greatest difference in slope between a continuous task and a standard trial-by-trial task would be seen in older or hearing-impaired listeners. As outlined in Chapter 5, due to degraded peripheral input, speech understanding processes are already more taxing for these listeners compared to normal-hearing younger listeners. A loss of spectro-temporal resolution means the acoustic features usually used to segregate and group sounds from the same source are less distinct (Leek & Summers, 2001). With auditory objects poorly defined selecting a desired stream from a mixture also becomes more difficult (Shinn-Cunningham, 2008). These factors mean that older and hearing impaired listeners will more frequently need to use top-down information and compensatory processes to understand speech, thus making speech understanding slower and more effortful (Pichora-Fuller et al., 1995). Age-related cognitive changes may further exacerbate older listeners' difficulties understanding continuous speech. Deteriorations in working memory, selective attention and a general slowing of processing speed, may result in greater difficulties dealing simultaneously with stored information and ongoing processing (Cohen, 1987). A greater detriment in speech understanding is likely to be seen in this group, therefore, in more challenging situations. It has been argued that standard speech-in-noise tests do not stress the listener and pauses allow some of the difficulties older and hearing-impaired listeners experience to be compensated for

(Shinn-Cunningham & Best, 2008). Support for this may be found in the finding that older listeners show a greater detriment in performance than younger listeners as the rate of speech is increased (Gordon-Salant & Fitzgibbons, 1997), which suggests a slowing of speech processing resulting in difficulties keeping up with the rate of speech (Wingfield, Poon, Lombardi, & Lowe, 1985).

The intelligibility of continuous speech is harder to quantify than shorter speech tokens. Standard word or sentence tests usually express intelligibility as the percentage of test items repeated correctly. Long passages of connected speech cannot, however, be recalled and stored in the same way. Several alternative methods for measuring the intelligibility of continuous speech have, therefore, been proposed.

One method is to present listeners with a continuous speech passage and then ask them a series of questions based on information presented in the speech (Giolas & Epstein, 1963). The intelligibility score in this case would be the number of questions correctly answered. It could be argued that, as such, this method does not directly assess speech "intelligibility". This method also does not take into account variables other than intelligibility that may equally affect a listener's performance on the task; variables such as intelligence or comprehension skills (Speaks et al., 1972). Also off-line methods, where responses are made after all the speech message has been heard, are likely to be subject to post-perceptual processes, with factors such as working memory span having a large effect on performance (Marslen-Wilson, 1985). Hafter, Xia and Kalluri (2012) have introduced a variation on this task where questions are asked throughout the speech, approximately 1 sec after the information has been heard. This reduces the working memory issue but does not address the role that comprehension skills may also play.

Another method which could be used to assess the intelligibility of continuous speech is "shadowing". Shadowing requires the listener to follow a message, repeating back each word as they hear it. In the past the method has been used to study the time course of the processes involved in "mapping sound to meaning" by looking at response delays (Marslen-Wilson, 1985 page 56) and to study selective attention by looking at listeners' ability to shadow messages while ignoring messages presented to the contralateral ear (Cherry, 1953; Moray, 1959; Treisman, 1964). In both cases the measure of performance was usually the percentage of words correctly reported. Using shadowing as a method for assessing the intelligibility of continuous speech has several advantages. Firstly, the duration of speech can be varied from a few sentences to a few minutes. This would make it possible, therefore, to assess the effect that the *duration* of continuous speech has on intelligibility. Secondly, shadowing allows an on-line method[23] of assessing the intelligibility of continuous speech. One major disadvantage of shadowing for the current purposes, however, is that it requires a verbal response while the stimulus continues. It has been suggested that the listener's own voice competes with that of the incoming message. This has been termed "competitive feedback" and refers to the interaction that occurs between listeners' attempts to listen to the speech and their attempts to repeat what has just been heard (Hopkinson, 1967). Competition from the listener's own voice is a particular issue if we wish to assess intelligibility of degraded speech. For example, if speech is presented in noise at low SNRs it is highly likely that the listeners own voice will significantly mask target speech making this method unsuitable for the current purposes.

A third approach is to use a "tracking" procedure. This method was proposed by Hawkins and Stevens (1950) and required listeners to adjust the level of ongoing target speech so that it "tracked" to a specific threshold. The threshold could be, for example, the level at which almost all sentences could just be understood ("threshold of intelligibility") or the level at which speech could just be detected ("threshold of detectability"). The intelligibility criterion was set by the experimenter but the decision of when this criterion was reached was made by the listener. Several other studies used similar approaches placing the emphasis on listeners' representation of "intelligibility" rather than on a statistical measure of word reception such as the percentage of words correctly repeated (Cox & McDaniel, 1989; Falconer & Davis, 1947; Lezak, Siegenthaler, & Davis, 1964). In each of these studies, speech intelligibility was given as a single threshold in dB; this method cannot be directly

[23] An on-line method is described here as one which can measure speech intelligibility as the speech is heard on a moment to moment basis. This method is in contrast to off-line methods where large segments or the entirety of the speech sample is heard before the intelligibility measurement is made.

used, therefore, to construct a psychometric function for continuous speech. Speaks et al., (1972) addressed this drawback by modifying the "tracking" procedure slightly so that intelligibility could be expressed as a magnitude in percent correct. Listeners in their experiment were asked to track to a "constant percentage criterion"; for example, they would be asked to find the level at which 75%, 50%, or 25% of sentence could be understood. As mentioned above, tracking procedures put a large emphasis on listener interpretations of the set criterions and are, as such, subjective measures of intelligibility. While the modifications made by Speaks et al., go some way to better defining each criterion, the psychometric function produced by such a method is still likely to be highly dependent on the listener's interpretations.

Jerger and Jorden (1992) also measured a psychometric function for continuous speech, although using a different method again. In this study a 20 minute segment of a first person story was used as the continuous stimuli and the listener's task was to press a response button every time they heard the word "I". SNR was adjusted and the percentage of targets correctly identified used as measure of intelligibility. The advantages of this method are that long sections of speech can be presented to listeners and an online measure can be obtained. It could be argued, however, that as listeners are simply detecting the presence or absence of the single word "I" it may overlook some of the more complex processes involved in understanding speech in everyday listening situations.

Other studies, while not testing the intelligibility of continuous speech per se, have used ongoing stimuli and therefore may use methods which, if adjusted, could be suitable for the purposes of the current experiment. Huckvale, Hilkhuysen and Frasi (2010), for example, used continuous speech stimuli in their "audio proof-reading task", a task designed to give a performance-based measure of speech quality. The audio proof-reading task involved participants listening to four minute extracts of a conversation between two talkers. Participants were asked to compare the speech they heard to typed speech which was displayed on the screen in front of them. Deliberate changes were made to the typed speech (word deletions, substitutions or insertions) and the participants' task was to identify these changes. It was argued that the task was challenging enough to preserve the cognitive demands experienced in everyday speech communication and would be an appropriate task for measuring changes in the quality of highly intelligible speech stimuli. Speech in this study was always presented at suprathreshold levels (levels that would give an SII value of 0.9 or greater), although it seems plausible that the intelligibility of speech (instead of the quality of speech) could be measured using a similar method if speech was presented at more unfavourable levels.

A similar substitution detection paradigm using continuous speech has also been used to assess the development of reading skills in children (e.g. Neville, 1975; Schneeberg, 1977). Participants were asked to "read along" with an audio stream, identifying any substitutions. In these studies audio was presented at different speaking rates and the number of substitutions identified was used as a measure of the child's ability to coordinate reading and listening. Results demonstrated that children as young as 7 were able to identify substitutions successfully at a natural speaking rate in quiet conditions (McMahon, 1983). These results indicate that this kind of audio/visual monitoring task involves processes that can be acquired at a young age and should, therefore, be suitable to be carried out with the older listeners.

The aim of the current experiment was to measure a psychometric function for continuous speech. An on-line method was desired, as it would minimise the effects of other factors which may affect performance (such as working memory) but also because it would eliminate the need for the artificial response pauses of the off-line, trial-by-trial type tasks. To avoid some of the problems associated with the current on-line methods (such as competitive feedback in shadowing tasks and subjective responses in tracking procedures), however, a new continuous task is proposed which is influenced by the audio/visual monitoring methods previously used to measure speech quality (Huckvale et al., 2010) and reading skills (McMahon, 1983). The continuous task in the current experiment involves participants listening to segments of continuous speech while simultaneously monitoring a written transcript of the same speech. The written transcript contains deliberate word changes and the listener task is to mark these substitutions as they occur. The rationale for the task is that if speech is intelligible then listeners should be able to easily spot word substitutions. As the speech becomes more degraded, however, speech will become less intelligible and the substitutions should be harder to identify. The number of substitutions identified then can be used as a measure of speech intelligibility. By measuring intelligibility at a range of different SNRs this method can be used to measure a full psychometric function.

Psychometric functions measured using the *continuous* task will be compared to those measured in a standard, *trial-by-trial* task. It is hypothesised that the slopes of the psychometric functions given by the *continuous* task will be significantly different to those given by the *trial-by- trial* task.

## 6.1 Method

### 6.1.1 Listeners

17 listeners aged between 60 and 73 (mean age = 68) took part in the experiment. All listeners were recruited from the MRC Institute of Hearing Research's (IHR) volunteer panel. Up-to-date audiometric data (measured at the IHR within the last six months) was available for all listeners. Better ear, pure-tone hearing thresholds (based on four-frequency averages) ranged from 25 dB HL (normal) to 50 dB HL (moderate hearing loss) (mean hearing loss = 39 dB HL). A small participation allowance was offered to listeners for their attendance and travelling expenses were refunded.

### 6.1.2 Stimuli

Both *trial-by-trial* and *continuous* targets were used in the current experiment. The *trial-by-trial* targets were taken from the Audiovisual Sentence Lists (ASL) and were spoken by a British English male (MacLeod & Summerfield, 1990). Each sentence had a simple syntactic structure and contained three designated keywords (e.g. *"The bag was very heavy*"). Their mean length was 1.5 seconds. The corpus contains 270 sentences in total which are split into 18 equally-identifiable lists. In the current experiment Lists 1 to 15 were used in the main experiment, and Lists 16 to 18 were used in a pre-test.

The *Continuous* targets were extracts taken from a commercial audiobook. The audiobook was an unabridged version of "The Memoirs of Sherlock Holmes" by Arthur Conan Doyle. Two stories ("Sliver Blaze" and "The Gloria Scott") were selected for the continuous extracts. The extracts were spoken by a British English actor and were edited to durations of 240-seconds (or the nearest word) using the speech editing programme Praat (Boersma & Weenink, 2012). Each 240-second extract was, on average, 647 words long (162 words per min). Eight, 240-second segments were used in total.

Both *trial-by-trial* and *continuous* targets were presented in a speech-shaped unmodulated ICRA noise (Dreschler, Verschuure, Ludvigsen, & Westermann, 2001). Short 3-second bursts were used as maskers for the *trial-by-trial* targets, while longer 250-second bursts, constructed by concatenating random 5-second segments of noise together with no gaps, were used as maskers for the *continuous* targets. On each trial targets were added into the middle of the noise masker and 25-ms raised-cosine gates were applied. Target speech was presented at a long-term average A-weighted level of 70 dB SPL. The level of the noise masker was adjusted relative to the target to achieve 7 different signal-to-noise ratios (SNRs). SNRs were chosen individually for each listener on the basis of thresholds attained in a pre-test (explained below). SNR was fixed over each 240-sec segment. Figure 6.1 illustrates the stimuli used for each task.

For the *continuous* condition, typed transcripts were taken from the published text of the book. In each transcript 50 words were changed so that they no longer matched those of the audio[24]. The words to be changed were selected in a pseudo-random fashion, adhering to several rules; no changes were made in the first line of each transcript, no function words were changed (e.g. "the", "his")[25], changed words

[24] The substitutions were only made to the typed transcript; no audio editing was done to the speech.

[25] Changing these almost always resulted in changing of the meaning of the sentence or made the sentence nonsensical. Take the sentence "he took a piece of white paper out of his waistcoat pocket": changing the adjective "white" to "blank", for example, does not notably change the sentence meaning. However, if the word "his" was changes to "her" it would be incongruous with the previous content and easily spotted as a substitution.

needed to be at least 4 words apart, and no word was changed if it would alter the context of the text. This last rule was crucial as we wanted to ensure that substitutions could not be spotted simply by reading the text alone. To test this, transcripts were administered to 2 people who were asked to read them and mark any words they believed had been changed. Only 4% of the errors were correctly identified from simply reading the text. Wherever possible, substituted words retained the same length and approximate word-frequency as the original word, as it was assumed that large changes in word lengths and/or frequency would make the substitution easier to spot.

### 6.1.3 Apparatus

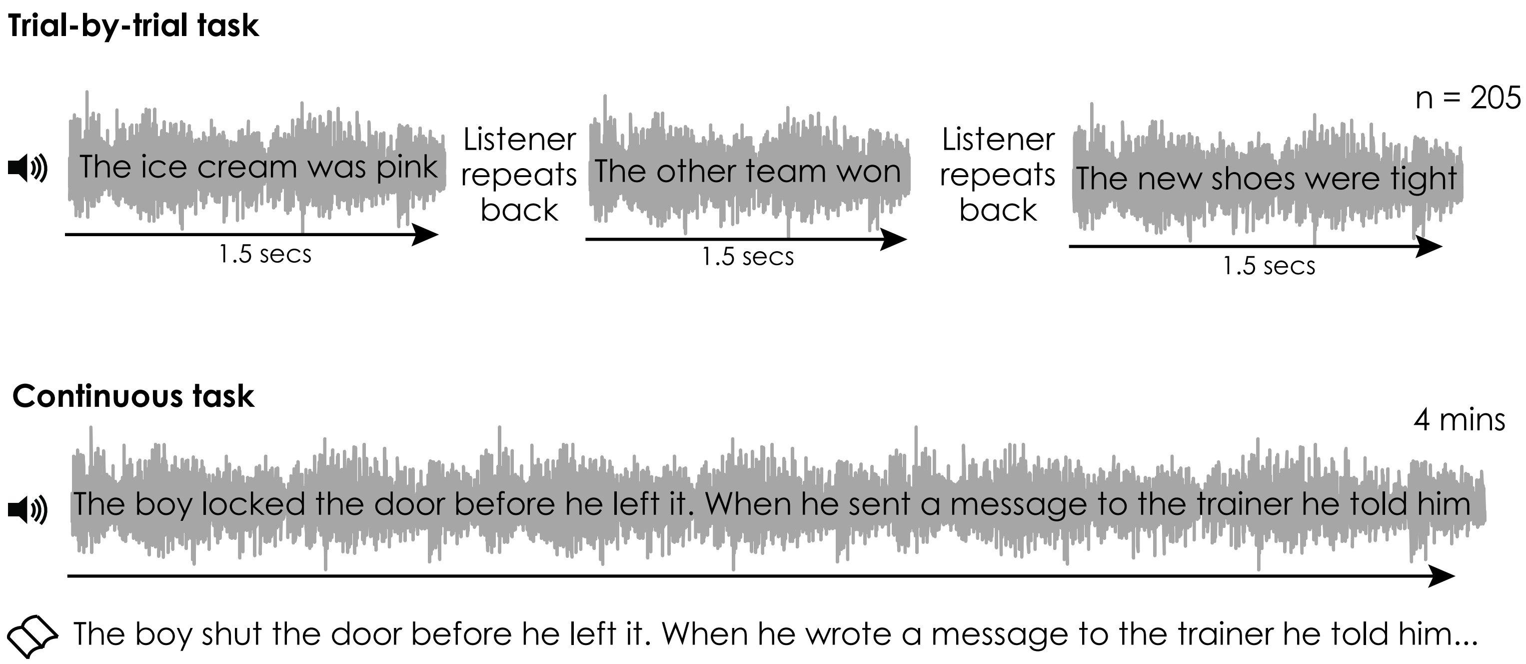
The experiment was carried out in a sound-treated booth. The same apparatus as previous experiments was used, with the addition of a wireless microphone to allow the experimenter to monitor listeners' responses in the *trial-by-trial* condition. The transcripts for the *continuous* condition were typed (double spaced, font size 14) on one side of A4 paper and broken into paragraphs of 4 to 5 lines long. An example transcript can be found in Appendix C.

### 6.1.4 Procedure

The experiment consisted of three parts; a pre-test, and two experimental tasks (the *trial-by-trial* task and the *continuous* task).

The pre-test was used to ascertain appropriate SNR ranges for each listener in the two experimental tasks. On each trial a random ASL sentence was presented which the listener was then asked to repeat back. Sentences were initially presented at a SNR of 6 dB. This level was reduced by 2 dB if the listener correctly identified at least two of the keywords from the target sentence. The SNR was reduced in 2-dB steps until the listener was unable to report more than one keyword on three consecutive trials. To ensure that a full psychometric function was measured in each task (i.e. from 20% to 90% word identification), the SNR ranges were set so that the lowest presentation level was 2 dB further down than the lowest point achieved on the pre-test. The remaining 6 presentation levels were then set at 2 dB steps above

***Figure 6.1:*** *A schematic illustration of the stimuli used in the trial-by-trial task and in the continuous task.*



these levels. For example, if the lowest point a listener achieved on the pre-test was -4 dB, speech would be presented at -6, -4, -2, 0, 2, 4, and 6 dB SNR. For the *continuous* task higher SNR ranges than those indicated on the pre-test were used for two listeners who struggled with the task when played in quiet.

Each trial of *trial-by-trial* task consisted of a random ASL sentence presented in noise. After each presentation the listener was asked to repeat back as much of the sentence as they could. Listeners were given as much time as they needed to respond and the next trial was not initiated until a response had been made. On each trial the number of correctly identified keywords (out of a possible three) was recorded. Homonyms and declensions of the target words were accepted as correct (MacLeod & Summerfield, 1990). The SNR at which the target speech was presented was randomised across trials. Listeners completed 2 blocks of the *trial-by-trial* task in total, each block consisting of 105 trials (i.e. 15 trials at each of the 7 SNRs). Psychometric functions were constructed by calculating the percentage of keywords correctly identified at each of the 7 SNRs.

On each trial of the *continuous* task participants listened to a 240-sec segment of continuous speech. Their task was to mark on the transcript any written words that did not exactly match those they had heard in the speech. Listeners were *not* able to pause the speech within a trial. Before completing the full 240-sec experimental trials, listeners were given practice on shorter, 60-sec segments. In these practice trials the first few substitutions were highlighted to further illustrate the listeners' task. The first experimental trial was always completed in quiet in order to measure listeners' baseline performance on the task. Subsequent trials were completed in noise. The SNR was fixed during a trial, but trial order was randomised. Listeners completed 2 blocks of the *continuous* speech task in total, each block consisting of 4 trials. Eight trials were, therefore, completed in total, one at each of the 7 SNRs and one in quiet. Psychometric functions were constructed by calculating the percentage of substitutions correctly identified at each of the 7 SNRs.

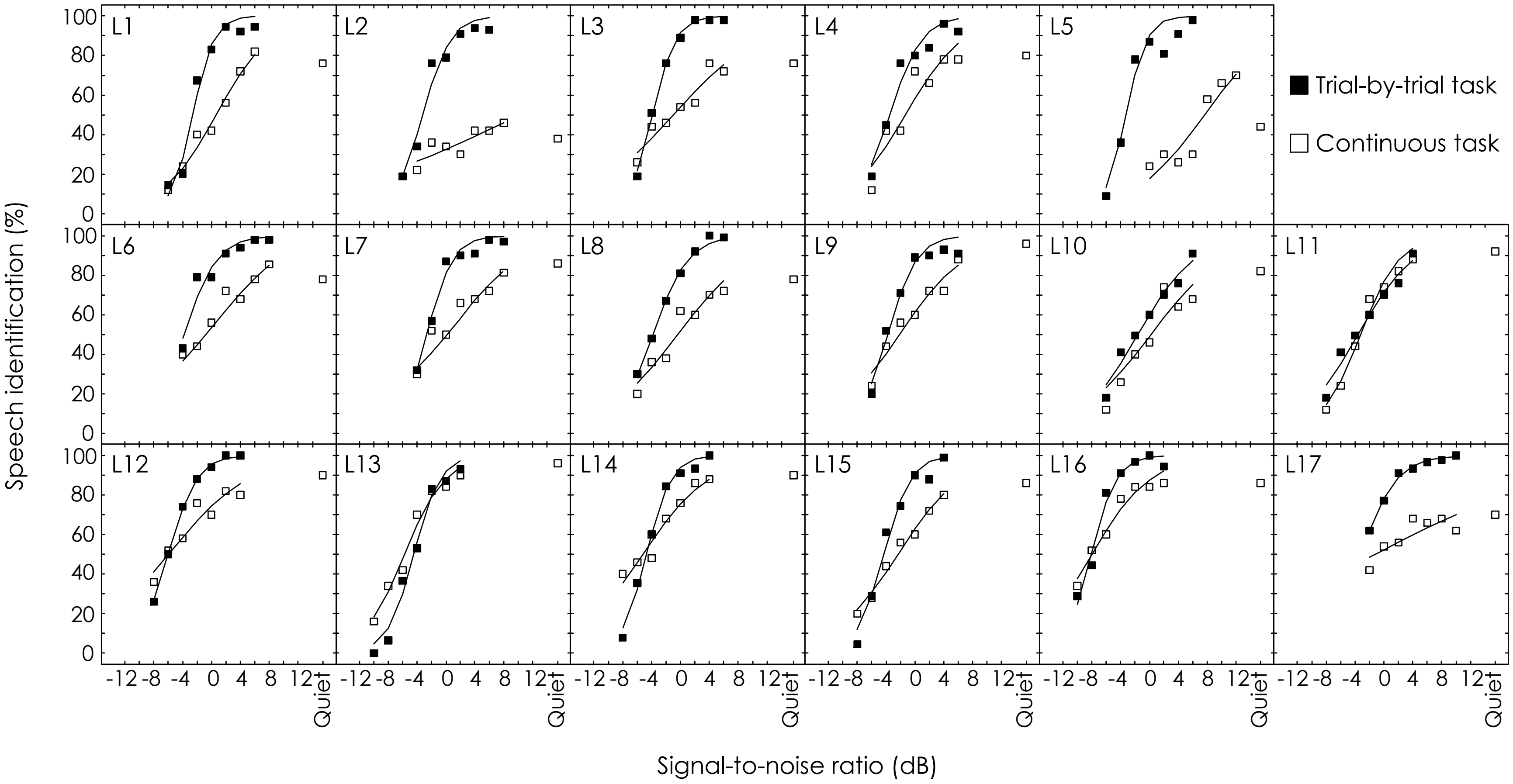
Each listener completed the experiment in one 1-hr 15-min session. In each session the pre-test was always completed first. The blocks of the experimental tasks were interleaved so that the listener did one block of one task followed by one block of the other task. This was counterbalanced across listeners so that half of the listeners started with a block of the *continuous* speech and the other half started with a block of the *trial-by-trial* task.

## 6.2 Results

Figure 6.2 shows individual psychometric functions for the *continuous* and *trial-by-trial* tasks. All psychometric functions have been fitted with logistic curves. For most listeners, the *trial-by-trial* task gave steep sigmoidal, psychometric functions. The one exception to this is listener 17, where there seems to be a ceiling effect, suggesting the full psychometric function was not successfully captured for this listener. In comparison to these steep functions, the functions for the *continuous* task were generally shallower for most listeners. Listeners 10, 11, and 13 are the only listeners for whom both tasks gave similar-shaped psychometric functions. For several listeners, performance on the two tasks was very similar at low SNRs and again at higher SNRs but differed in the mid SNR range (e.g. Listeners 1, 4, 9, 15 and 16). This gives an indication that the rate of intelligibility improvement with level was less for the *continuous* task than it was for the *trial-by-trial* task. The figure also shows that most listeners could do the *continuous* task; performance was generally good at baseline and at favourable SNRs in noise. Nevertheless, there was an indication that a couple of listeners did struggle with the task. Listeners 2 and 5, for example, performed poorly on the task in quiet (38% and 44% respectively). In noise, listener 2's performance on the task remained around this baseline level, suggesting the difficulty of the task, rather than audibility, was limiting performance. Similarly, listener 5 needed a considerably higher SNR range than was required in the *trial-by-trial* task to be able to perform the task, again suggesting this was not a problem of audibility.

Anecdotally, listeners reported that they enjoyed the *continuous* task. Many expressed a wish to listen to more segments so they could find out what would happen in the story. This suggests that listeners were engaged in the task and that attention was being maintained throughout each 4 minute segment. Several listeners also reported that they felt the rate of speech was increasing in the *continuous* task as

***Figure 6.2:*** *Individual psychometric functions are shown for all 17 listeners for the trial-by-trial task (closed squares) and the continuous task (open square). Each function is also fitted with a logistic function (solid black line).*



the level of the noise was increased. This gives a hint that a greater effort was required to keep up with speech as it became more degraded.

Figure 6.3 shows mean psychometric functions averaged across listeners for the two types of task. As the SNR ranges were individually selected for each listener the extreme data points represent fewer listeners than the more central points. SNRs where data for fewer than 5 listeners were collected have been excluded (i.e. -10, 8, 10, and 12 dB SNR). Both psychometric functions have been fitted with logistic curves. The graph shows the overall trend across listeners and suggests that shallower psychometric functions were given by the *continuous* task (M = 5.1% per dB, SD = 1.9% per dB) than by the *trial-by-trial* task (M = 12.3% per dB, SD = 3.1% per dB).

### 6.2.1 Slopes

Figure 6.4 displays slope values for each individual listener measured in the *trial-by-trial* task as a function of their slope values measured in the *continuous* task. The dotted line represents 1:1, i.e. where the two tasks give equivalent slopes. Data points for all but one listener sit above this line suggesting that slopes measured in the continuous task were consistently shallower in the *continuous* task than they were in the *trial-by-trial* task.

A one-way repeated measures ANOVA was carried out on this slope data. Slopes measured in the *continuous* task were indeed found to be shallower than those measured in the *trial-by-trial* task: $F(1,16) = 63.8$, $p < 0.001$, partial $\eta^2 = 0.80$. Figure 6.5 shows scatter plots of slope values on the two tasks as a function of the listener's four-frequency, better-ear average. The correlations between slope and the degree of hearing loss were found to be non-significant for both the *trial-by-trial* ($r = 0.22$, $p = 0.39$) and *continuous* task ($r = 0.25$, $p = 0.33$).

### 6.2.2 Speech reception thresholds

Figure 6.6 is a scatter plot of the speech reception thresholds measured for each listener in the two tasks. Again the dotted line represents 1:1. While data points are

***Figure 6.3:*** *Psychometric functions for the trial-by-trial task (closed squares) and the continuous task (open squares) averaged across listeners. The dotted line show the logistic fit and the error bars show 95% confidence intervals. As the SNR ranges for each listener varied slightly, the extreme data points here represent fewer listeners than more central points.*

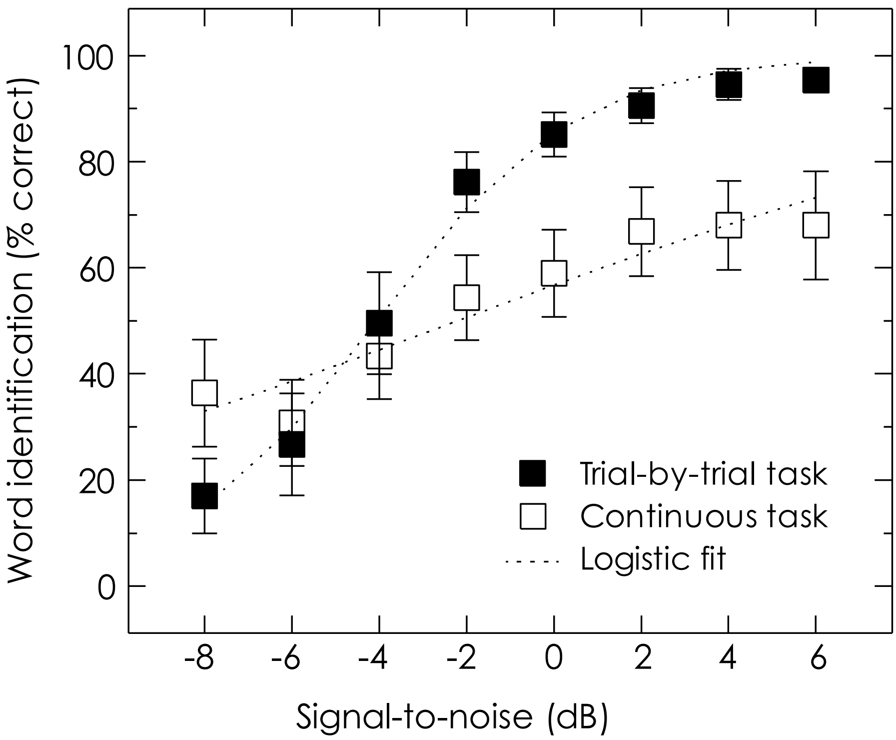


***Figure 6.4:*** *Individual listeners' slope data for the continuous and trial-by-trial tasks. The dotted line represents 1:1 where the slopes on the two tasks are equivalent.*

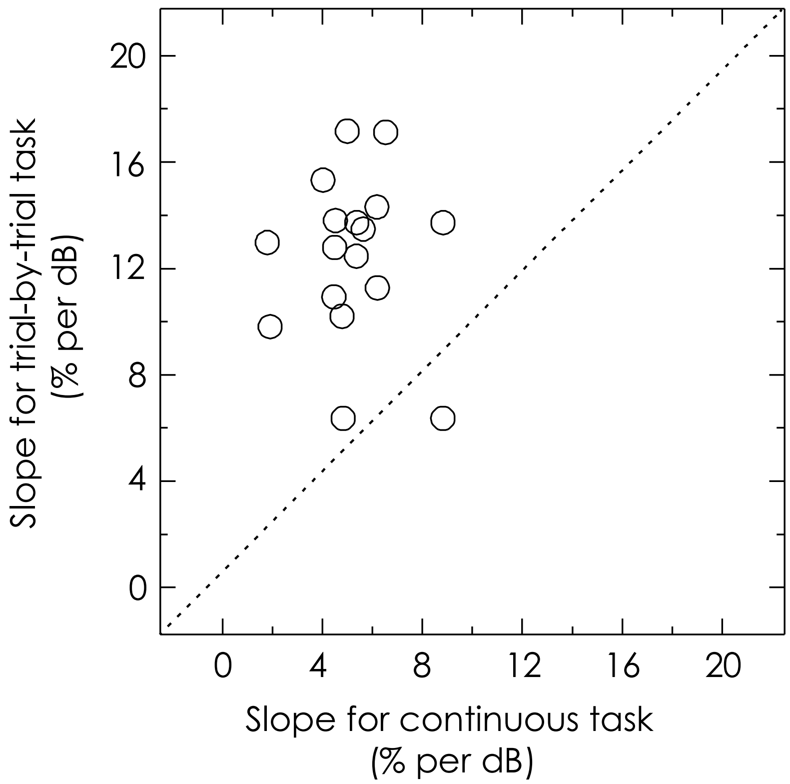


***Figure 6.5:*** *Individual listeners' slope data plotted as a function of better-ear average hearing loss. The left panel shows slope values given on the trial-by-trial task and the right panel shows slope values given on the continuous task. r values are also included.*

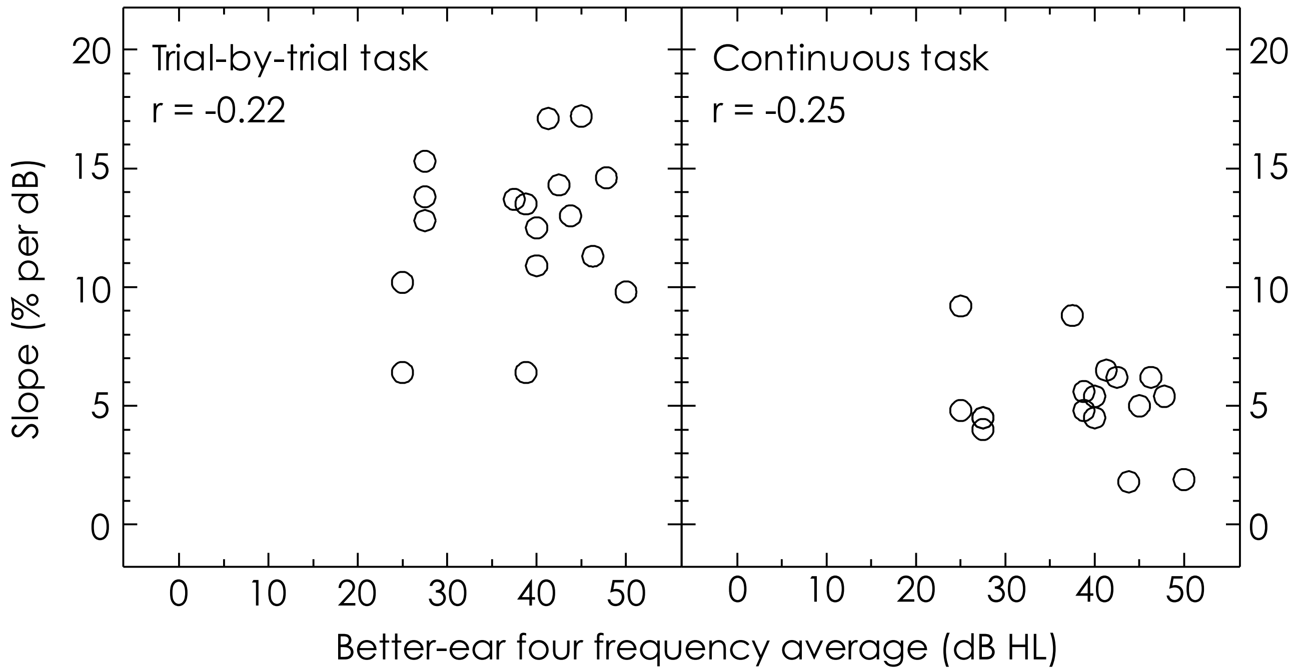
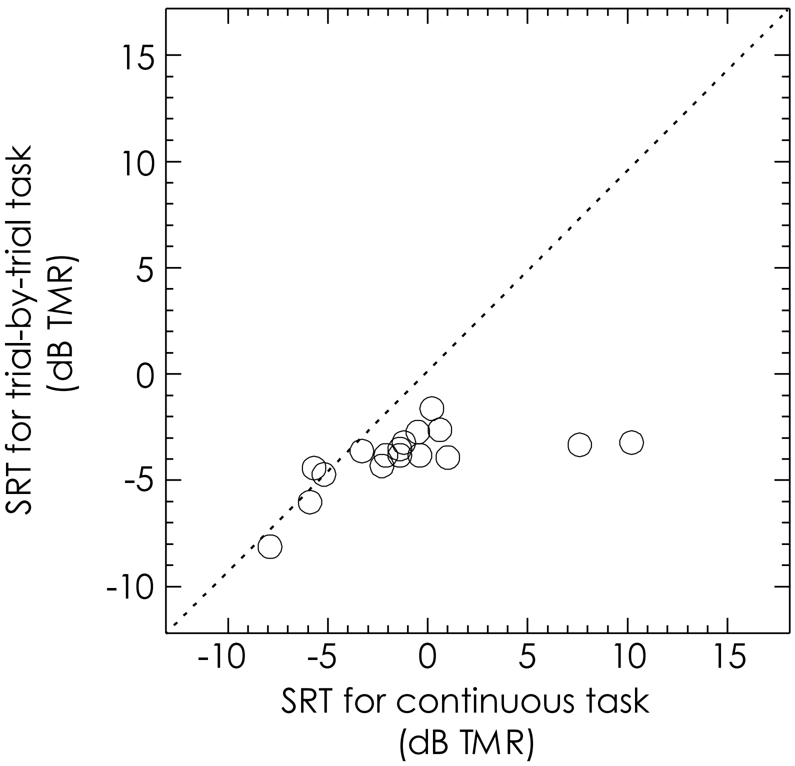


***Figure 6.6:*** *Individual listeners' speech reception threshold data for the continuous and trial-by-trial tasks. The dotted line represents 1:1 where the SRTs on the two tasks are equivalent.*



clustered relatively close to the line, the majority are below the line suggesting that SRTs were higher in the *continuous* task than they were in the *trial-by-trial* task.

A one-way repeated measures ANOVA found a significant effect of task on SRT with significantly higher SRTs given in the *continuous* task ($M = -1.0$, $SD = 4.5$) than in the *trial-by-trial* task ($M = -3.9$, $SD = 1.4$): $F(1,16) = 9.5$, $P < 0.01$, partial $\eta^2 = 0.37$. The effect size suggests that the effect of task on SRT is smaller than it was on the slope of the psychometric function. Figure 6.7, shows scatter plots of SRT values measured on each task plotted against listeners' better ear averages. No significant correlation was found between SRT and degree of hearing loss for either the *trial-by-trial* ($r = 0.02$, $p = 0.94$) or the *continuous* task ($r = 0.17$, $p = 0.53$). There was very little variation across listener in terms of thresholds on the sentence task (SD = 1.4 dB SNR), which would explain the lack of correlation with hearing loss for this task.

### 6.2.3 False alarm rate on the continuous task

Words which were incorrectly identified as possible substitutions in the *continuous* task were termed false alarms. In an attempt to identify where listeners were making these errors, the false alarms in each transcript were analysed. For each false alarm the number of words between it and the closest genuine substitution were counted; the closest substitution could be either before or after the false alarm. Whether the nearest genuine substitution had been identified (a hit) or not (a miss) was also recorded. The classification of false alarms is explained in greater detail in Appendix D. Listeners made in total 49 false alarms which were nearest to a hit (hit/false - alarms) and 138 false alarms which were nearest to a miss (miss/false-alarms).

Figure 6.8 shows how many words occurred before or after a genuine substitution for each of the hit/false-alarms. The figure shows that most hit/false-alarms were either made just 1 or 2 words after a genuine substitution or several words (6 or more) after the substitution. The errors made close to a substitution suggest that some confusion over the precise location of the substitution had occurred. That the substitution had also been correctly identified suggests, however, that listeners may have, after making the false alarm, subsequently corrected for their error (listeners

***Figure 6.7:*** *Individual listeners' speech-reception threshold data plotted as a function of better-ear average hearing loss. The left panel shows SRTs given on the trial-by-trial task and the right panel shows SRTs given on the continuous task. r values are also included.*

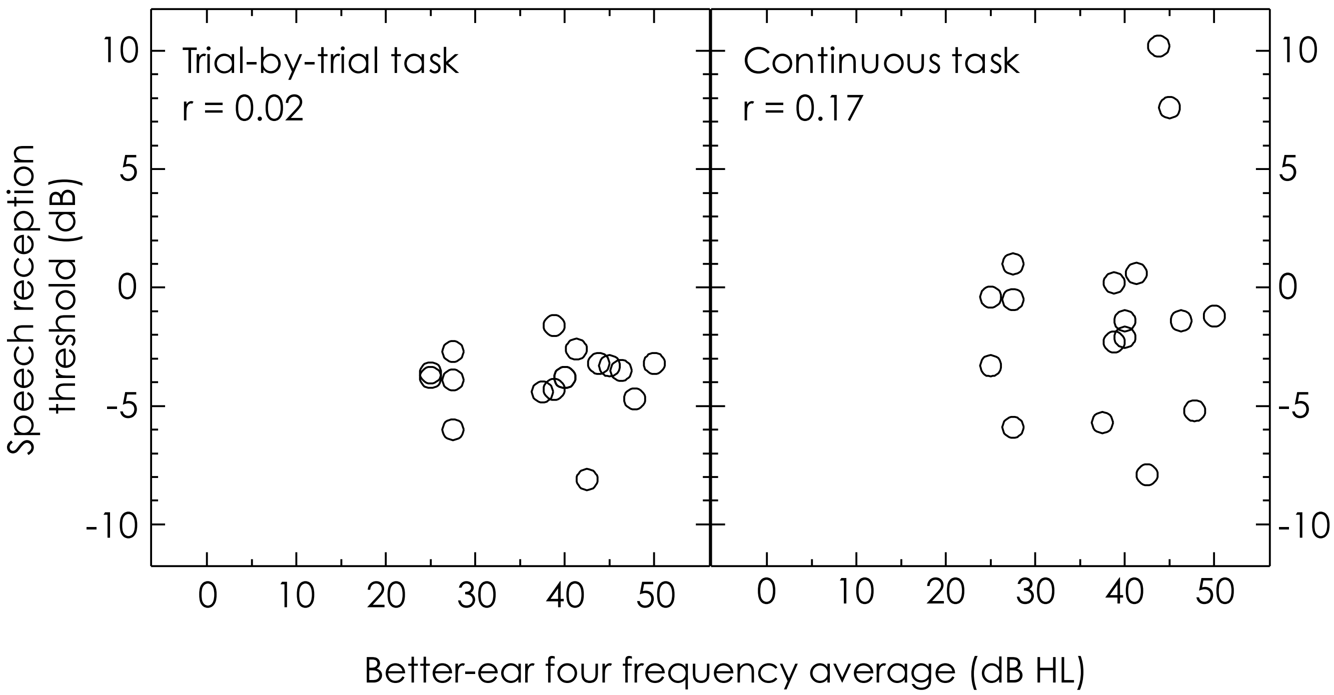


***Figure 6.8:*** *Distributions of the number of words away from a genuine substitution that hit/false alarms occurred.*

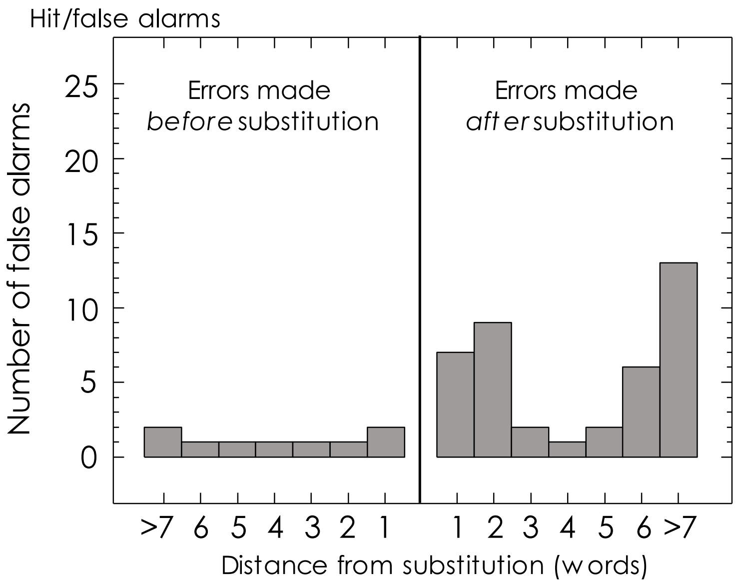
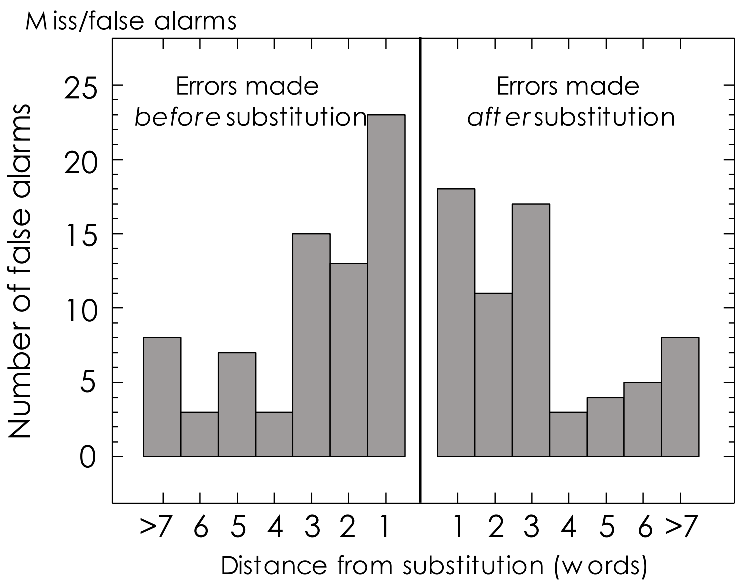



***Figure 6.9:*** *Distributions of the number of words away from a genuine substitution that miss/false alarms occurred.*



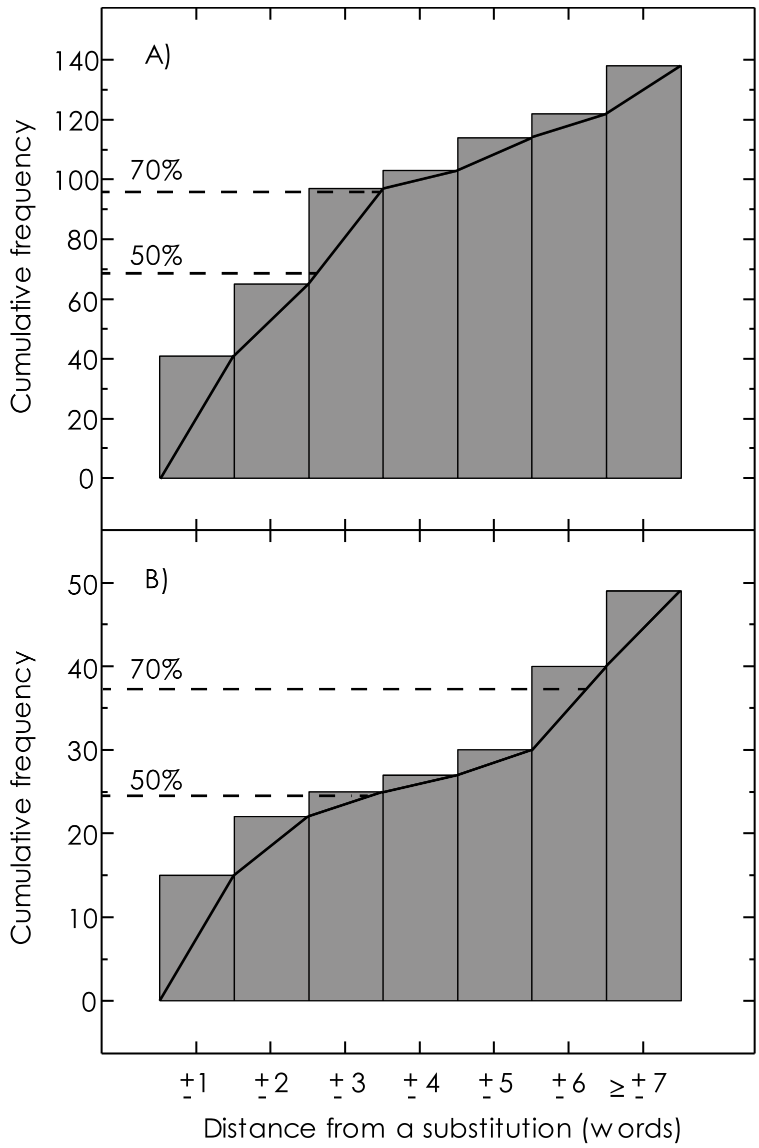
did not have the means, or time, to indicate to the experimenter when they were aware that they had made an error). The false alarms made further away from a substitution may reflect a different decision making process. As the substitutions were on average 13 words apart, the higher instance of false alarms more than 6 words after a substitution suggests listeners might start anticipating where the next substitution will be. In other words these false alarms, unlike those made close to a substitution, are likely to be additional guesses which are unrelated to the processing of a genuine substitution.

Figure 6.9 shows how many words either before or after a genuine substitution each of the miss/false-alarms occurred. A large number of these types of false alarms were made either just before or just after a genuine substitution. This again suggests that confusion as to the precise location of the substitution occasionally occurred which resulted in it being misassigned to a surrounding word. Figure 6.10 shows cumulative frequency graphs for the number of miss/false alarms (panel A) and hit/false alarms (panel B) and their relative distance to a genuine substitution. This graph highlights the large proportion (over 50%) of both miss/false alarms and hit/false alarms that occurred within three words of a genuine substitution.

### 6.2.4 Summary of results

- Slopes given by the *continuous* task were found to be significantly shallower than those given by the *trial-by-trial* task.
- SRTs were found to be significantly higher for the *continuous* task than for the *trial-by-trial* task.
- A large proportion of false alarms made in the *continuous* task were made within a few words of a genuine substitution.

***Figure 6.10:*** *shows the cumulative increase in the number of false alarm made with increasing distance from a substitution. Panel A shows miss/false alarms and panel B shows hit/false alarms. The dashed lines indicate 50 and 70% of the total number of each type of false alarms.*



## 6.3 Discussion

A large proportion of speech encountered in everyday situations is continuous. The aim of the current experiment was to consider the shape of the psychometric functions given by more realistic listening conditions. Psychometric functions given by a new *continuous* speech task were, therefore, compared to those given by a standard *trial-by-trial* task. It was hypothesised that the slopes for the two tasks would differ. This hypothesis was supported with the slopes given in the continuous task found to be significantly shallower than those given in the trial-by-trial task. These results suggest that the rate of intelligibility improvement with level was less when speech was presented continuously with no pauses than when it was presented one sentence at a time with pauses.

Very few studies have measured the psychometric function for continuous speech and, to our knowledge, no studies have compared these functions to the functions given by non-continuous speech. Jergen and Jorden (1992) did report a psychometric function for continuous speech in speech babble and when compared to the average slopes given by similar masking conditions identified in the systematic slope survey, their slopes were notably shallower. The results from Experiment 5 provide further evidence for this trend. Speaks et al., (1972) also measured a psychometric function for continuous speech. As was the case in the current study, a static noise was also used as a masker, however, no difference in slope was seen when this study was compared to average slope values from the slope survey. Differences to the current study in terms of stimuli and task may explain why slopes in Speaks et al.'s study were not shallower. The effect that choice of stimuli may have on slope will be discussed in greater detail below.

As well as a slope difference, there was also a small but significant threshold difference found between the two tasks; thresholds measured in the *trial-by-trial* task were, on average, 2.9 dB lower than those measured in the *continuous* task. It could be reasoned that, as listeners were provided with a written transcript of the speech in the continuous task, this task would give a lower threshold. Several studies have demonstrated that combining audio and visual cues, such as being able to see the speaker as they are talking, gives a significant benefit in terms of threshold – with improvement reported anywhere in the region of 1 – 11 dB (Grant & Seitz, 2000; Macleod & Summerfield, 1987). It was surprising therefore that thresholds were slightly *higher* in the *continuous* task, where there was a visual aid, compared to the *trial-by-trial* task where only audio information was available. It is possible, however, that the greater complexities involved with understanding continuous speech made the task slightly more difficult, hence the higher threshold. In the *continuous* task, for example, cognitive resources are likely to be more greatly taxed as processing works to keep pace with the speech rate.

As the masking noise used in the current experiment was static noise, the slope-change mechanisms (i.e. amplitude modulations and target/masker confusion) which have been identified as underlie the slope changes seen in the previous experiments cannot explain the slope change seen here. This suggests that a different mechanism is responsible for the slope change which arises when target speech is continuous. It was postulated that there were several aspects of continuous speech which could flatten the slope of the psychometric function. The longer duration of the *continuous* task would mean, for example, that there was a greater likelihood of errors occurring due to lapses in attention than there was in the *trial-by-trial* task. The substitution detection rate (i.e. speech intelligibility) would not be based solely on the audibility of the target speech but also on lapses in attention. As lapses in attention are likely to be stimulus independent and can occur regardless of the SNR they would have a tendency to flatten the slope of the psychometric function (Wichmann & Hill, 2001b). Though the *continuous* task was undoubtedly subject to occasional lapses in attention it is unlikely that these lapses alone could have resulted in the large, and consistent, slope difference seen between the two tasks.

It was also suggested that if the continuous task taxed cognitive resources and slowed down speech understanding processes this might also result in a flattening of the psychometric function. With no pauses to allow processing to catch-up (Shinn-Cunningham & Best, 2008), occasional lapses in processing are likely to occur and parts of the speech will be missed. These occasional lapses are partially independent of target level[26] and as such tend to flatten the slope of the psychometric function. It is highly likely, however, that processing lapses are dependent on other features of the target stimuli. For example, the complexity of the target speech, speech rate and the processing load imparted by the task are all likely to effect the amount to which cognitive resources are taxed and in turn the number of processing lapses which will occur. It is plausible then that changes in any of these features would change the slope of the psychometric function.

There were several indications that listeners were finding the *continuous* task more taxing than the *trial-by-trial* task. Anecdotally several listeners reported that the speech rate for the *continuous* task increased as the SNR was decreased. No change was made to the rate of speech of any of the continuous segments but listeners' perception that speech was speeding up as target speech became degraded by the noise gives an indication that processing was becoming more effortful and listeners were finding it more difficult to keep pace with the speech rate. No listeners made a similar observation for the *trial-by-trial* task. The type of false alarms made in the *continuous* task also gave an indication that this task taxed speech understanding. A large proportion of the false alarms were instances where listeners had marked words which were within a few words of an actual substitution (45% of hit/false-alarms and 70% of miss/false-alarms were within 3 words of a substitution). If substitution detection was down to audibility alone, it would be expected that false alarms would be distributed relatively evenly throughout the text, they would not necessarily be placed close to actual substitutions. One explanation for this is that the continuous nature of the speech meant that cognitive resources were devoted to speech understanding processing and fewer resources were available to devote to the task of matching the audio information as it arose to the visual information. The matching task became less fine-grained as a result and, rather than being able to indentify substitutions to the precise words, listeners were only able to identify substitutions to

[26] So called "processing lapses" are not completely independent of level. The occurrence of processing lapses are proposed to be related to a slowing down of speech processing. If speech processes become slower as the listening situations become more challenging (e.g. Shinn-Cunningham & Best, 2008), a greater number of processing lapses could be expected when speech is degraded e.g. when presented at an unfavourable SNR.

within a small cluster of words. Presumably, if the continuous task became less taxing, the number of these types of false alarms would fall.

Two very different sets of stimuli were used in the *continuous* and *trial-by-trial* tasks. Before it can be concluded that the continuousness of the stimuli resulted in the observed slope difference, other differences in the stimuli must be addressed. The stimuli used in the *continuous* task were, for example, not recorded specifically for the purposes of auditory testing; instead a commercially available audiobook was used. An audiobook was chosen as it provided convenient access to a large amount of continuous speech which could be quickly adapted. While the chosen stimulus does have ecological validity, i.e. it is analogous to listening to the radio or to an audiobook in a noisy environment such as a car, several drawbacks should also be noted. Firstly, the speech from audiobooks is often quite rapid. Secondly, the audiobook is spoken by an actor and is, as such, much more expressive than the ASL sentences; this may have resulted in greater level variations in the audiobook stimuli. Thirdly, the lengths of sentences range from 3 words to 50 words, the vocabulary is wider, word frequency fluctuates and the structures of the sentences also vary greatly. With both level and complexity differences between the *stimuli* used in the two tasks it is possible that the slope effect measured in the current experiment could be the result of these differences rather than the continuousness of speech. This will be addressed in detail in Experiment 6.

Several different methods for measuring the intelligibility of continuous speech have been used in the past (e.g. Giolas & Epstein, 1963; Speaks et al., 1972). But each of these methods have their drawbacks, such as subjective intelligibility measurements or intelligibility measurements influenced by other factors such as working memory. The continuous task used in the current experiment uses a new method for measuring speech intelligibility and while it avoids some of the problems of previous methods it has its own drawbacks. It could be argued, for example, that the *detection* of word substitutions is not a direct measure of speech *identification*. In a standard speech-in-noise test such as the ASL test intelligibility is based on the percentage of words the listener was able to fully identify. In contrast the continuous task could be thought of more as a discrimination task. For each word heard the listener is faced with a yes/no discrimination task "does the word I just heard match that of the word on the page". The listener does not need to identify the word to correctly spot the substitution, they only need to establish that it doesn't match the word on the page. A phoneme or a syllable may be sufficient for this discrimination. "Intelligibility", therefore, may not be equivalently measured across the two tasks. This being said, the *continuous* task may be more similar to other experimentally used speech-in-noise tests. The CRM (Bolia et al., 2000), Matrix (Hagerman, 1982), or any other closed-set task where the listener matches speech utterance to a list of keywords, also measures intelligibility based on an *n*-choice discrimination.

By asking listeners to match information extracted from an audio source to a visual one, the current *continuous* task avoids introducing some additional factors, such as working memory and competitive feedback, which in other methods may have affected speech intelligibility. It could be argued, however, that in doing so the *continuous* task involves two separate processes. Listeners are not just identifying speech, as they would in a standard speech in noise test, they are also reading the text and then comparing information from both sources (McMahon, 1983). This may affect the measurement of "intelligibility" in two ways. Firstly, a listener's performance on the *continuous* task might not, as is assumed, reflect their ability to keep up with and identify continuous speech but instead their ability to *read* the written transcript quickly enough to keep up with the rate of speech. A listener with a slower reading rate may, therefore, perform more poorly on the current task than they would if a different method for measuring the intelligibility of continuous speech was selected. Secondly, if the *continuous* task is thought of as a two separate processes − both reading *and* listening − it could be presumed that greater cognitive resources would be required to complete the current *continuous* task than say a shadowing task. It was expected that the continuousness of speech would increase cognitive load but if the task itself significantly increases cognitive load it would introduce a large confound to the results. It could be argued, for example, that the general taxing nature of the *continuous* task, rather than the continuousness of the speech per se, resulted in a flattening of the psychometric function in the current experiment.

There is a question, however, surrounding whether reading while listening can really be thought of as two separate processes. Daly, Neville and Pugh (1979, as cited by McMahon, 1983) suggested that reading and listening simultaneously involves two distinct processes between which attention and cognitive resources must be divided. Others have argued, however, that as the goal of both modalities is to comprehend a stream of semantically invariant information, and are, therefore, complementary not competing, reading while listening should be thought of as a single process (McMahon, 1983). Further, reading while listening can be thought of like any other situation where we integrate simultaneous information from different sources into a single perceptual event; like watching television for example (Gibson, 1969). Integrating text and audio is also something that many people are familiar with. Several listeners in the current experiment said, for example, that they often use subtitles while watching television to help understand what is being said. If the reading and listening aspects of the *continuous* task are considered as complementary processes for which resources do not need to be split, it seems unlikely that the task itself greatly increased cognitive load to the point where any increased difficulties due to the continuousness of speech might have been masked.

In summary, Experiment 5 has demonstrated that the psychometric functions for continuous speech were shallower than those measured for standard trial-by-trial tasks. While this slope difference has been attributed to the greater tax on cognitive resources that continuous speech produces, it is possible that other differences between the stimuli selected for the two tasks — such as differences in level or linguistic complexity — may also have played a role. This will be considered further in Experiment 6.

## 6.4 Experiment 6: The relative roles of continuousness and complexity on the slope of the psychometric function.

In Experiment 5, shallower slopes were found when target speech was continuous than when short, simple utterances were presented one at a time. The main aim of Experiment 6 was to establish whether this effect could be attributed to the “continuousness” of the stimuli or whether some other difference between the stimuli resulted in the change in slope.

The speech stimuli for the continuous task were taken from a commercial audiobook while the speech stimuli for the trial-by-trial task were taken from a speech corpus designed specifically for audiological testing. The speech in the two tasks, therefore, differed in several ways besides in their continuousness. Firstly, the auditory quality of the two types of stimuli was different. The ASL sentences were normalised, spoken at a steady rate and with a limited variation of intonation (MacLeod & Summerfield, 1990). With the main goal of the audiobook being to tell a story, however, there was greater expression in the talker’s voice, meaning there was also likely to be greater variations in speaking rate, overall level and in the underlying fundamental frequency track. Secondly, the content of the speech for the two stimuli types also differed. In the audiobook, the speech was quite varied: sentences were of different lengths and syntactic structure, semantic information spanned several sentences, narration and dialogue were mixed and both high and low frequency words were used. The ASL sentences were much less complex: each sentence was roughly the same length (4 – 6 words with 3 keywords) and had the same syntactic structure, sentences across lists were equated to be equally intelligible and contained a limited vocabulary with medium to high frequency words often used in multiple sentences (e.g. lady, man, Father, Mother).

Given these large auditory and qualitative differences between the stimuli, it is possible that at least part of the slope difference attributed to the continuousness of target speech in Experiment 5 was the result of these other differences. The greater variation in level and intonation in the audiobook stimuli means that it would have a larger overall dynamic range than the ASL stimuli. Its intelligibility may, therefore, be much more variable. If this is the case intelligibility scores would be more dependent on whether substitutions happened to be assigned to the more intense portions or the less intense portions of target speech than on the overall level of the speech (the SNR). If more substitutions were, for example, assigned to words which happened to be to be at a local minima for an overall presentation level of 4 dB SNR than for an overall presentation level of 2 dB SNR, intelligibility scores could be similar, in effect flattening the slope of the psychometric function.

The average speaking rate for the audiobook stimuli was also relatively rapid at 164 words per minute. While this is around the rate of an average conversation (Wingfield, McCoy, Peelle, Tun, & Cox, 2006) it is likely to be faster than the more steady speaking rate of the ASL sentences used in the trial-by-trial task. There was also likely to be greater variations in speaking rate in the audiobook stimuli. Several studies have demonstrated that even for short sentences, speech understanding for older listeners decreased as speech rate was increased (Gordon-Salant & Fitzgibbons, 1997, 1999; Tun, 1998). It has been argued that this is due, at least partly, to a decline in processing speed (Wingfield et al., 1985). It is possible then that the faster, or variable speech rate of the audiobook stimuli, especially when considered in conjunction with the continuousness of the stimuli, led to the occasional difficulties "keeping up" which were proposed in Experiment 5 as the reason for shallower slopes.

The greater variation in sentence length and structure in the audiobook stimuli means that these stimuli are more complex than the ASL sentences. It was proposed in Experiment 5 that the shallow slopes seen in that experiment may have been related to a slowing down of processing due to the greater tax on cognitive resources that continuous speech imposed. It is possible, however, that the increased complexity of the continuous stimuli also required increased cognitive processing. Syntactically complex sentences, for example, have been found to put a greater strain on working memory than simpler sentence structures, as materials must be held in memory as it unfolds over time before it can be integrated and resolved and the relationships between words and concepts understood (Cohen, 1987; Wingfield et al., 2006). Variations in sentence length and structure, and variations in word frequencies also result in less predictable speech. This greater unpredictability could result in increased processing times as unfamiliar words are retrieved (Greenberg, 2006) and means contextual cues and/or knowledge about the subject area will need to be accessed (Pichora-Fuller et al., 1995). Changes in the style of speech means the listeners also had to keep track of which character was talking to fully understand the audiobook. It is possible that complex sentences might be sufficient to tax speech understanding processes resulting in some parts of speech being missed even if speech is not continuous. The complexity of the audiobook stimuli then, rather than its continuousness per se, may be the factor that resulted in shallower slopes for these stimuli.

In an attempt to determine whether it was the continuousness of the target speech or the choice of stimuli which resulted in shallow slopes in Experiment 5, the current experiment switched the stimuli used in the two experimental tasks. The short ASL sentences used for the trial-by-trial task were concatenated to form four minute segments of speech which could be used in a new *continuous* task. Conversely, the four minute segments of continuous speech taken from the audiobook were edited into shorter sentences and presented one at a time to create the new *trial-by-trial* task. If slopes were shallow due to speech being continuous, then it was predicted that the same slope pattern noted in Experiment 5 should be seen. The *continuous* task would still give shallow slopes even though the ASL stimuli were used and the *trial-by-trial* task would still give steep slopes even though the audiobook stimuli were used. If, however, some other factor present in the audiobook stimuli but not in the ASL stimuli was causing the psychometric functions to become shallower an alternate slope prediction would be that the slope pattern noted in Experiment 5 would be reversed: shallower slopes would be seen in the *trial-by-trial* task because this task used the audiobook stimuli and steeper slopes would be seen in the *continuous* task because this task used the ASL sentences.

## 6.5 Method

### 6.5.1 Listeners

17 listeners took part in the experiment. These were a different set of listeners from those who took part in Experiment 5. All listeners were recruited from the MRC Institute of Hearing Research's volunteer panel and were aged between 62 and 75 (mean age = 68). Listeners' BEAs ranged from 24 dB (normal hearing) to 50 dB (moderate hearing loss). Mean hearing loss for the group was 36 dB HL. A small

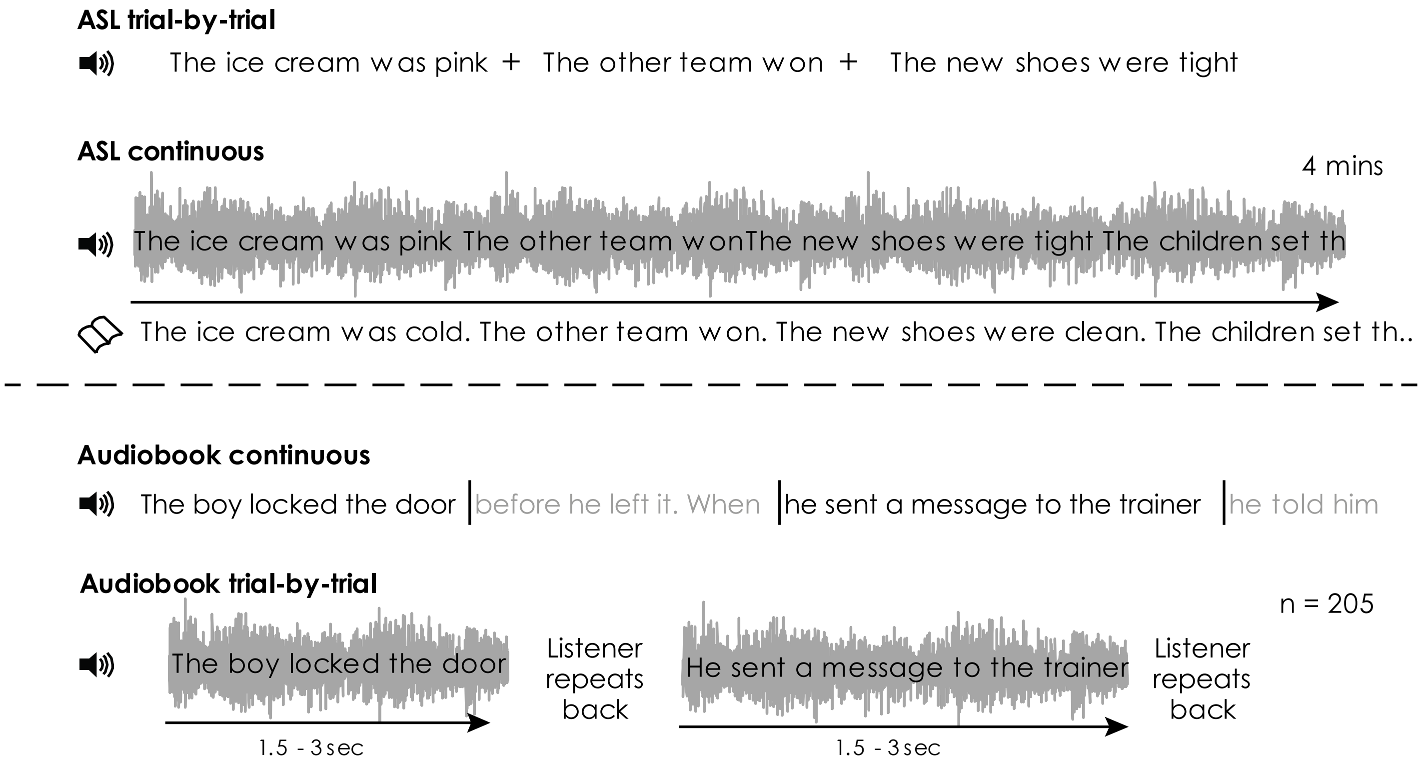
participation allowance was offered to listeners for their attendance and travelling expenses were refunded.

### 6.5.2 Stimuli

The stimuli used for the *trial-by-trial* and *continuous* tasks in Experiment 5 were swapped in the current experiment (see Figure 6.11). The *audiobook/trial-by-trial* targets were 210 sentences edited from the commercial audiobook used in the *continuous* task in Experiment 5 (an unabridged version of "The Memoirs of Sherlock Holmes" by Arthur Conan Doyle). The speech editing programme Praat was used to edit the continuous speech into short phrases of 4 – 12 words in length (average duration = 2.1secs). Phrases were often edited from longer sentences (e.g. the original sentence *"He took a piece of white paper out of his waistcoat pocket"* was edited to *"He took a piece of white paper"*). 25-ms raised-cosine gates were applied to each sentence. For scoring purposes 3 – 6 keywords were selected for each sentence depending on its length. Assignment of keywords followed Macleod & Summerfield (1990) (e.g. *"He took a piece of white paper"*, *"The boy locked the door"* the keywords are underlined). Thirty of the sentences were reserved for the pre-test and were not presented in the audiobook trial-by-trial task testing phase. All sentences are listed in Appendix E.

The *ASL/continuous* targets were six, 240-second segments of speech created by concatenating together randomly ordered ASL sentences (MacLeod & Summerfield, 1990). Due to the limited number of ASL sentences, additional sentences were also taken from the BKB corpus and randomly interspersed with ASL sentences. BKB sentences were selected as they were very similar, in terms of length and syntactic structure, to the ASL sentences (e.g. *"The clever girls are reading"* is an example of an ASL sentence and *"The dirty boy is washing"* is an example of a BKB sentence). Sentences from both corpuses were spoken by the same British English male. It was calculated that an average interval of 300 ms between sentences would give an average rate of 163 words per min (an equivalent rate to the *audiobook/continuous* task in Experiment 5). As the intervals between sentences in Experiment 5's *audiobook/continuous* task naturally varied, sentences in the current experiment were concatenated with a random interval of 100 to 500 ms (i.e. average overall interval of

***Figure 6.11**: A schematic illustration of the stimuli used in the audiobook/trial-by-trial task and in the ASL/continuous task.*



300 ms for the segment). Despite using both the ASL and BKB corpuses, there were only sufficient sentences to make six, 240-second segments (as opposed to the eight created in Experiment 5). Also, 30% of sentences in each 240-second segment were repeated in a second segment. It should be noted that segments which shared sentences were always presented with at least two intervening blocks separating them.

As in Experiment 5, both the *audiobook/trial-by-trial* and the *ASL/continuous* targets were presented in the middle of a speech-shaped unmodulated ICRA noise of the appropriate length (3 seconds or 250 seconds respectively). The target speech was presented at a long-term average A-weighted level of 70 dB SPL. To achieve different SNRs, the level of the masking noise was then adjusted relative to that of the target. As before the SNR range was selected individually for each listener by means of a pre-test. The SNR range was slightly reduced in the current experiment (from seven SNRs to five SNRs) due to the reduced number of 240-second segments available.

As in Experiment 5, typed transcripts of the six *ASL/continuous* targets were created with deliberate substitutions made to a set proportion of words. The rules regarding word changes set out in Experiment 5 were followed in the current experiment. One noteworthy change, however, was that the number of substitutions per segment was reduced from 50 to 40. The reason for this was that the ALS and BKB sentences were relatively short and reducing the number of substitutions ensured that substitutions could be spaced at least four words apart without falling into a predictable pattern (e.g. one substitution every other sentence) or push the overall duration over 240 seconds.

### 6.5.3 Apparatus and Procedure

The experiment was carried out in the same sound-treated booth and with the same equipment used in Experiment 5. The experimental procedure also followed that of Experiment 5 very closely. The method for selecting individual SNR ranges for the *ASL/continuous* task was adjusted slightly. In Experiment 5 SNR ranges for both tasks were selected by setting the lowest presentation level to be 2 dB less intense than the lowest point achieved on the pre-test. Pilot data for the current experiment indicated, however, that at this level performance on the *ASL/continuous* task was likely to be at ceiling. The *highest* presentation level was set to be 2 dB less intense than the lowest point achieved on the pre-test. Ad-hoc adjustments were made to the SNR range selected by the pre-test for a few listeners who were either performing close to ceiling or floor on selected SNRs.

## 6.6 Results

Figure 6.12 shows individual psychometric functions for each listener for the *audiobook/trial-by-trial* task and the *ASL/continuous* task. Each psychometric function was fitted with a logistic function. For most listeners, while the functions for the *ASL/continuous* task were shifted to the right, the slopes given by both were relatively similar. The slopes for both tasks were relatively steep. Two listeners (L4 and L13) showed noticeably different patterns to the other listeners. For these listeners there was no threshold shift and slopes were shallower in the *continuous* task than in the *trial-by-trial* task; i.e. their psychometric data was reminiscent of the functions measured in Experiment 5.

### 6.6.1 Slopes

Figure 6.13 is a scatter plot comparing slope values measured in the *ASL/continuous* and *audiobook/trial-by-trial* tasks. The dotted line represents 1:1. Data points are clustered around this line, indicating there is no consistent slope trend; the *audiobook/trial-by-trial* task gave steeper slopes for some listeners but for others the *ASL/continuous task* gave the steeper slopes ($r = 0.14$, $p = 0.59$).

A one-way repeated ANOVA was carried out. Slopes for the *audiobook/trial-by-trial* ($M = 9.3$, $SD = 2.8$) and *ASL/continuous tasks* ($M = 8.7$, $SD = 2.5$) were not found to significantly differ ($F(1,16) = 0.57$, $p = 0.46$, partial $\eta^2 = 0.03$). Figure 6.14 displays the correlation between listeners' BEAs and slope values measured on the two tasks. A significant correlation was found for the *audiobook/trial-by-trial* task ($r = -0.6$, $p < 0.05$) indicating that shallower slopes were found for listeners with the largest

***Figure 6.12:*** *Individual psychometric functions are shown for all 17 listeners for the audiobook trial-by-trial task (closed squares) and the ASL continuous task (open square). Each function is also fitted with a logistic function (solid black line).*

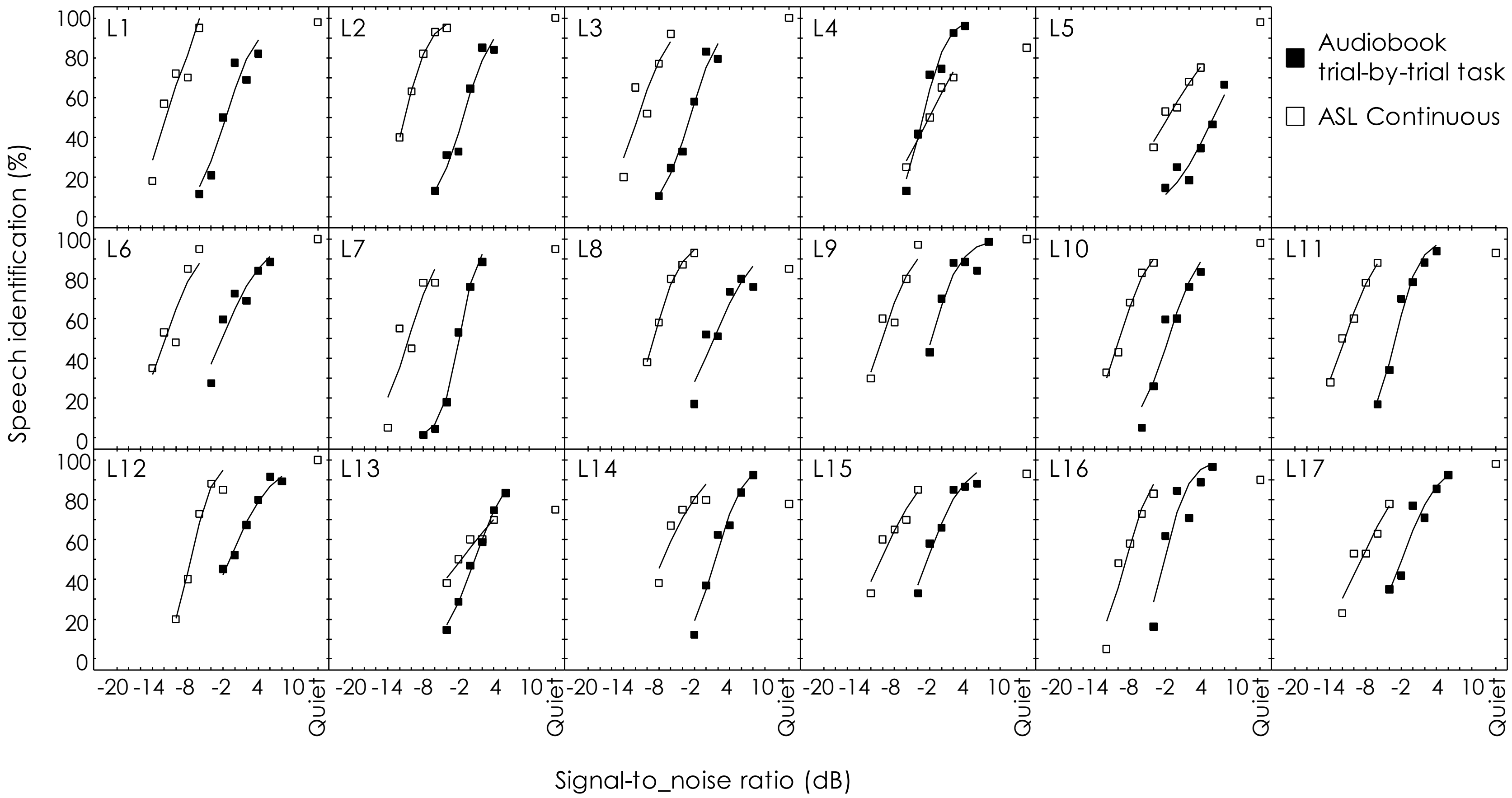


***Figure 6.13:*** *Individual listeners' slope data for the ASL continuous and audiobook trial-by-trial tasks. The dotted line represents 1:1 where the slopes on the two tasks are equivalent.*

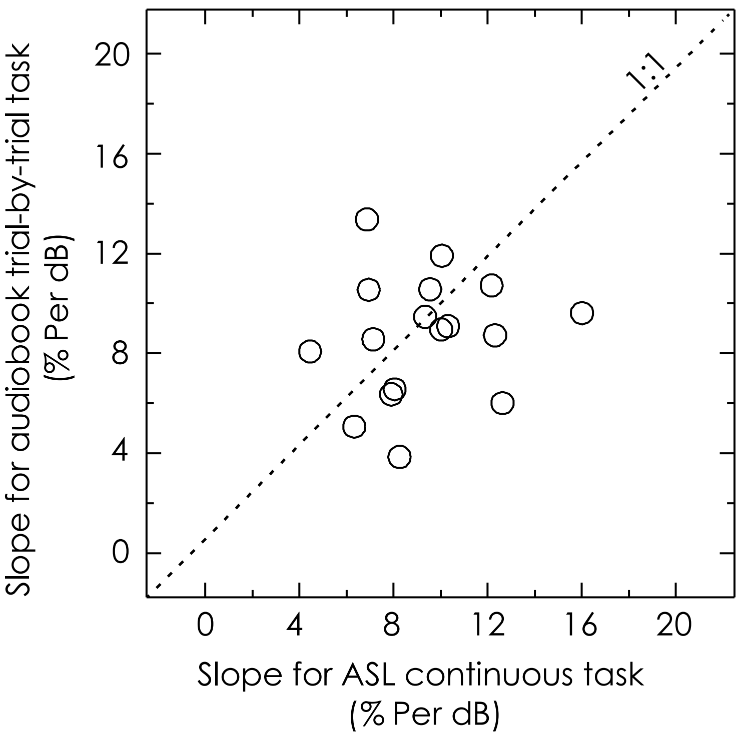
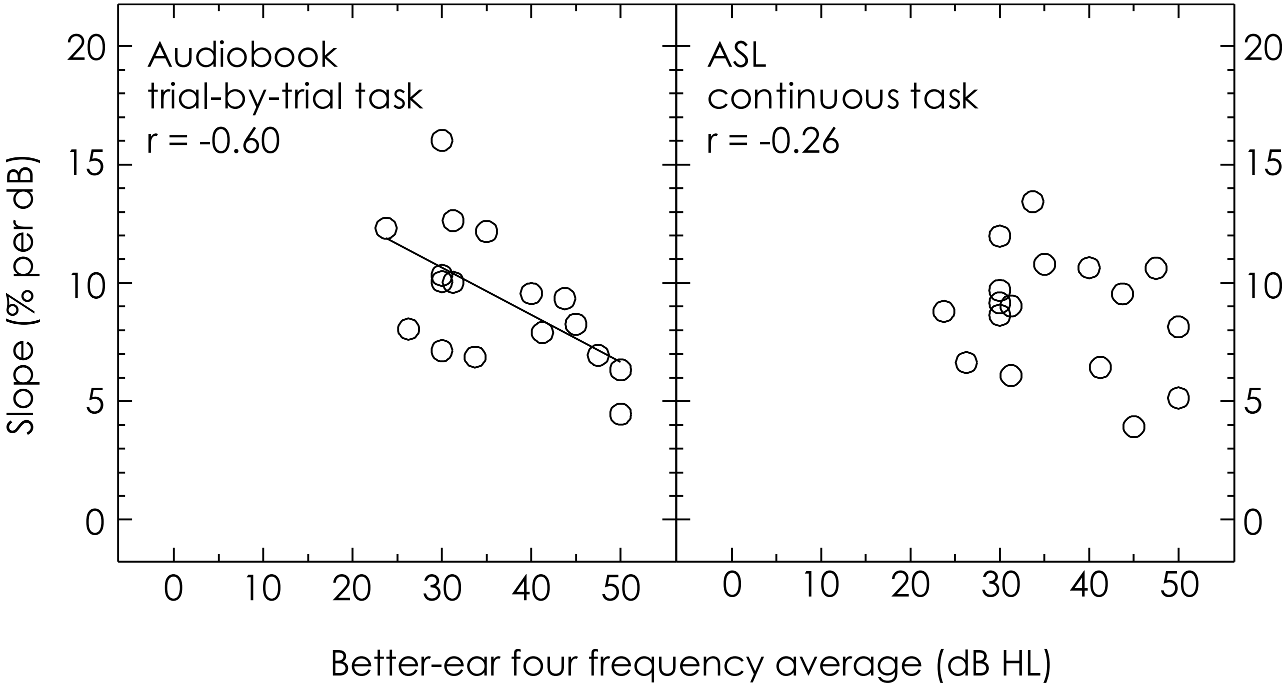


***Figure 6.14:*** *Individual listeners' slope data plotted as a function of better-ear average hearing loss. The left panel shows slope values given on the audiobook trial-by-trial task and the right panel shows slope values given on the ASL continuous task. r values are also included.*



hearing losses. No correlation was found between BEA and slope for the *ASL/continuous* condition ($r = -0.26$, $p = 0.31$).

### 6.6.2 Speech reception thresholds

Figure 6.15 is a scatter plot comparing speech reception values for each listener for each of the tasks. All data points are clearly above the line marking 1:1, indicating, therefore, that SRTs for all listeners were higher in the *audiobook/trial-by-trial* task than they were in the *ASL/continuous* task ($r = 0.55$, $p < 0.05$).

A one-way repeated ANOVA found SRTs to be significantly higher for the *audiobook/trial-by-trial* task ($M = -0.7$, $SD = 2.8$) than for the *ASL/continuous* task ($M = -8.4$, $SD = 3.6$; $F(1,16) = 105.5$, $p < 0.001$, partial $\eta^2 = 0.87$). Figure 6.16 shows SRT for the *audiobook/trial-by-trial* task (panel A) and *ASL/continuous* task (panel B) plotted as a function of BEA. A significant correlation with BEA was found for both tasks ($r = 0.80$, $p < 0.001$ and $r = 0.53$, $p < 0.05$ respectively). This result suggests that for both tasks as the degree of hearing loss increased so did SRTs.

### 6.6.3 False alarm rate on the continuous task

As in Experiment 5, false alarms made by listeners while identifying substitutions in the *ASL/continuous* task have been considered. Again false alarms have been split into two types, those closest to an identified substitution (hit/false-alarms) and those closest to an unidentified substitution (miss/false-alarms). Listeners made a total 32 hit/false-alarms and 43 miss/false-alarms. Figure 6.17 shows the number of intervening words between each hit/false-alarm and a genuine substitution. The graph demonstrates that the majority of hit/false-alarms were made seven or more words away from a substitution, suggesting these errors were unlikely to be related to the processing of a genuine substitution. Figure 6.18 shows the number of intervening words between each miss/false-alarm and a genuine substitution. The relative distance from a substitution was more variable for these types of false alarms, but most errors were made either close to (within two words) or relatively far away (more than seven words) from a substitution. Figure 6.19 shows a cumulative count of miss/false-alarms and hit/false alarms as the distance from a genuine

***Figure 6.15:*** *Individual listeners' speech reception threshold data for the ASL continuous and audiobook trial-by-trial tasks. The dotted line represents 1:1 where the SRTs on the two tasks are equivalent.*

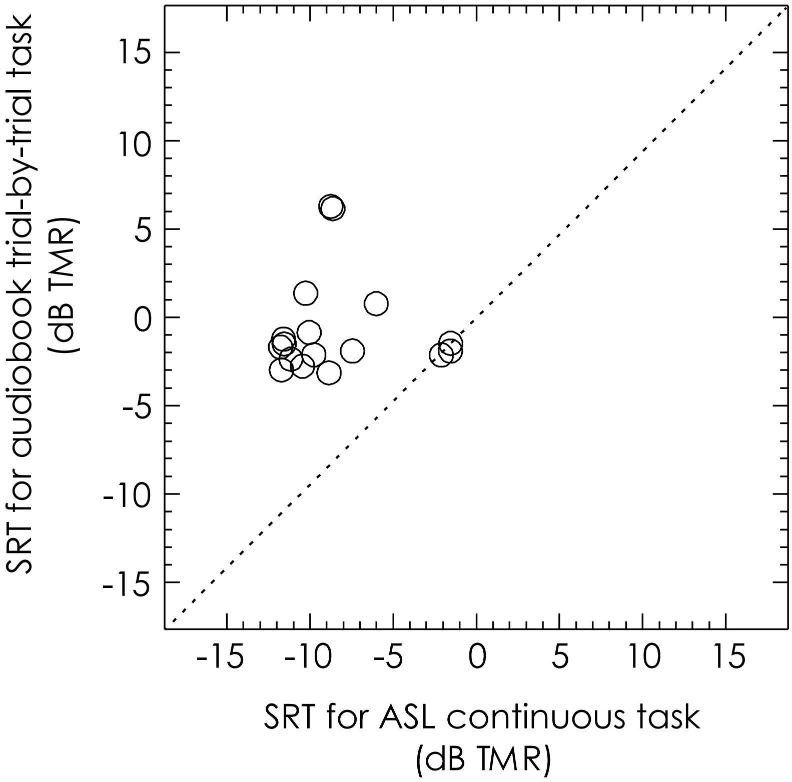


***Figure 6.16:*** *Individual listeners' speech-reception threshold data plotted as a function of better-ear average hearing loss. The Left panel shows SRTs given on the audiobook trial-by-trial task and the right panel shows SRTs given on the ASL continuous task. r values are also included.*

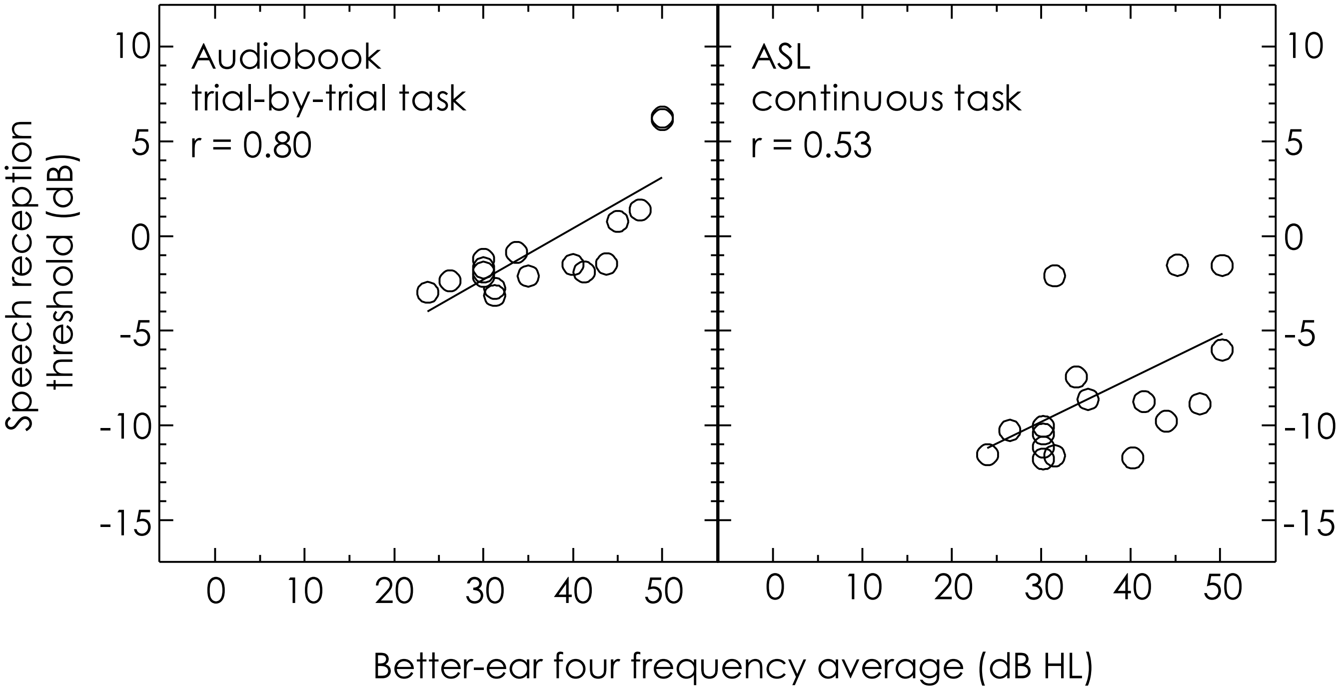


***Figure 6.17:*** *Distributions of the number of words away from a genuine substitution that hit/false alarms occurred.*

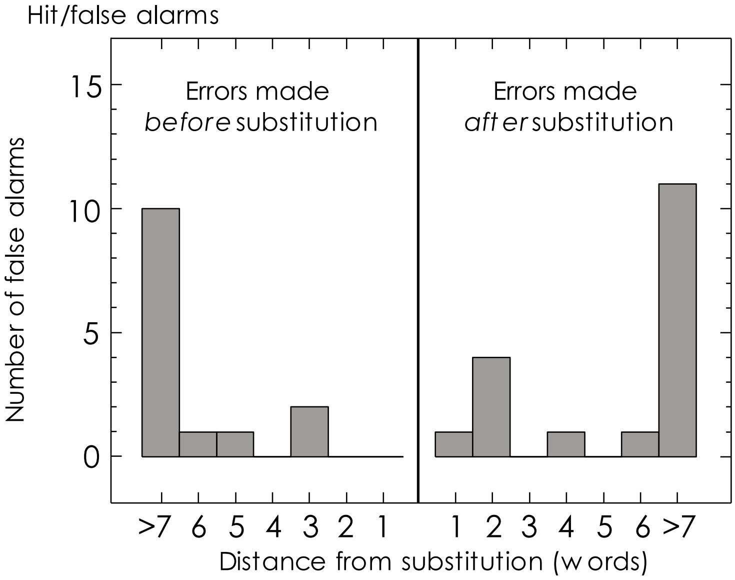



***Figure 6.18:*** *Distributions of the number of words away from a genuine substitution that miss/false alarms occurred.*

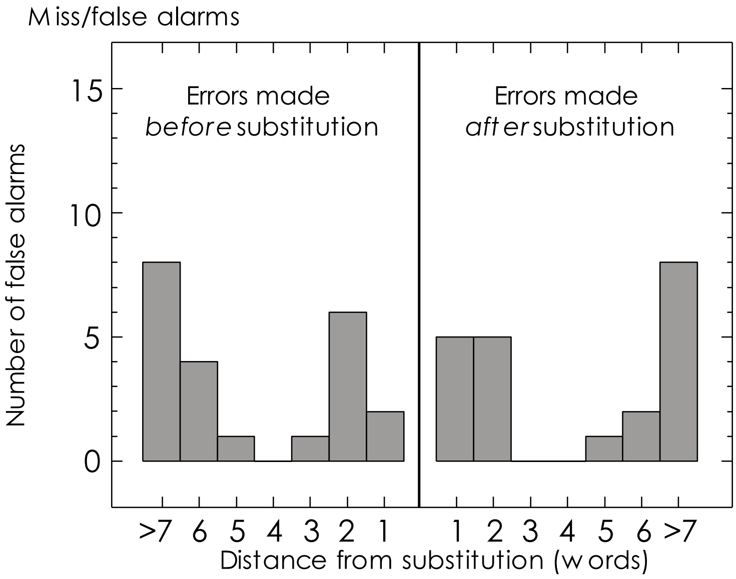
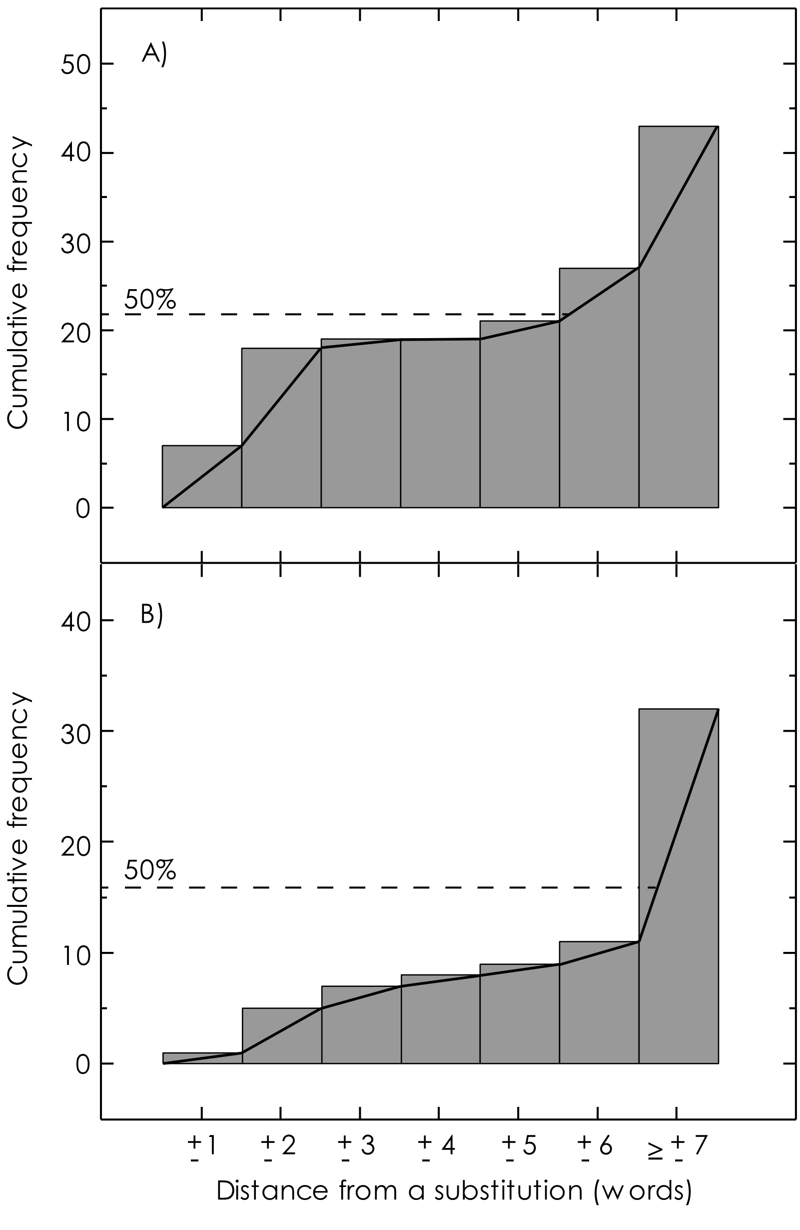


***Figure 6.19:*** *shows a cumulative count of the false-alarm made with increasing distance from a substitution. Panel A shows miss/false alarms and panel B shows hit/false alarms. The dashed lines indicate 50% of the total number of false alarms.*



substitution is increased. The graph shows that false alarms made within 3 words of a substitution accounted for 44% of the total miss/false-alarms made and only 22% of the total hit/false alarms.

### 6.6.4 Summary of results

- Slopes given by the *audiobook/trial-by-trial* and the *ASL/continuous* tasks were not found to significantly differ.
- SRTs were found to be significantly lower for the *ASL/continuous* task than for the *audiobook/trial-by-trial* task.
- Less than half of the false alarms made in the *ASL/continuous* task were made within a few words of a genuine substitution.

## 6.7 Discussion

The aim of this experiment was to establish whether the slope effect measured in Experiment 5 was the result of the continuousness of target speech or whether the stimuli selected for the tasks were playing a role. Two alternate slope predictions were made: 1) if continuousness did play a role, shallower slopes should still be seen in the continuous task even if the original stimuli were switched, and 2) if differences between the ASL and audiobook stimuli were causing the difference in slope then switching the stimuli should result in switching the slope pattern found in Experiment 5; shallower psychometric functions would be found for the *trial-by-trial* task and steeper slopes would be found for the *continuous* task. The results were in fact a combination of these two predictions. Table 6.1 summarises the mean slope values found in Experiments 5 & 6.

The psychometric functions were found to be relatively steep in the *audiobook/trial-by-trial* task, however, the average slope was still slightly shallower than that seen for the *ASL/trial-by-trial* task of Experiment 5. The steep slope found for the audiobook stimuli in Experiment 6 fits with prediction (1) and would suggest that the task, and therefore the continuousness of the speech, did play a role. For the audiobook stimuli, presenting sentences on a trial-by-trial basis gave steeper slopes than presenting it continuously; the slope increased by 4.2% per dB from the *audiobook/continuous* task to the *audiobook/trial-by-trial task*.

Psychometric functions were found to also be relatively steep for the *ASL/continuous* task. They were, for example, steeper than the slopes given by the *audiobook/continuous* task but less than those given by the *ASL/trial-by-trial* task from Experiment 5. The steeper slope found in this experiment for the ASL stimuli fits best with prediction (2) and suggests that the stimuli and not the task were playing a role here; presenting ASL stimuli in a continuousness form still gave relatively steep psychometric functions.

***Table 6.1:*** *A comparison of mean slope values (and standard deviations) for the different combinations of task type and stimuli used in Experiments 5 & 6.*

| **Stimuli type** | **Task type** | |
|---|---|---|
| | continuous | trial-by-trial |
| ASL | 8.7 (2.5) | 12.5 (3.1) |
| Audiobook | 5.1 (1.9) | 9.3 (2.8) |

Slope values in % per dB

There remains a possibility that shallower slopes may have been measured in Experiment 5 for the continuous task because the audiobook fluctuated in level from word to word. An analysis of the level variations was carried out to test this hypothesis. As the measures of intelligibility were based on the identification of only a subset of words from the continuous stimuli, i.e. the substitutions, it was reasoned that level variations across these words was important. The substituted words in all continuous segments were edited out using Praat. The words were then filtered using an A-weighted filter. This was done so that measured levels would approximately correspond with listener-perceived loudness (Moore, 2012). The words were then loaded into Matlab and the power for each word measured. The variations in power were found to be equivalent for both the ASL (SD = 3.2 dB) and audiobook stimuli (SD = 3.9 dB). The shallower slopes given by the *audiobook/continuous* task than by the *ASL/continuous* task cannot, therefore, be attributed to greater level fluctuations.

There was also the possibility that the shallow slopes for the *audiobook/continuous* task in Experiment 5 could have been the result of the complexity and unpredictable nature of the stimuli. Sentences from the audiobook were a mixture of lengths, syntactic structure, and style. This was not the case for the ASL sentences. Table 6.2 shows the results of a reading difficulty analysis on each of the continuous stimuli. The Flesch Reading ease test uses sentence and word length to judge the readability of a passage of writing (Flesch, 1948). The difficulty of each continuous segment from the *audiobook/continuous* and *ASL/continuou*s tasks were measured using a freely available web version of the Flesch Reading Ease Test (readabilityformulas.com). In terms of reading difficulty, scores of 60 to 69 are classified as standard, 70 to 79 fairly easy, 80 to 89 easy and 90 to 100 very easy. Readability scores on all *ASL/continuous* segments were found to be higher than those for the *audiobook/continuous* segments – i.e. ASL segments were easier to read. This analysis suggests that ASL segments were indeed less complex than the audiobook segments. It is possible then that simple sentences, even when presented continuously, were not sufficient to tax speech understanding processes enough to flatten the slope of the psychometric function.

Table 6.2 also shows that there was less variation in difficulty across segments for the *ASL/continuous* stimuli than there was for the *audiobook/continuous* segments. It is possible that this variability across segments may have affected results. Some segments would have been less complex than others. If complexity does affect the number of substitutions identified, listeners are likely to perform more poorly on some segments than others. This may also have had an effect on the slope of the psychometric function. If the less complex segments happened to be presented at high SNRs while the more complex segments were presented at low SNRs this would give a steeper slope than if the less complex segments were presented at low SNRs and the more complex segments were presented at high SNRs. Complexity across segments would have to be controlled for in future experiments to avoid this confound.

***Table 6.2:*** *Flesch Reading Ease scores for all continuous segments used in the ASL continuous and the audiobook continuous tasks. Scores between 60-69 are classified as standard, 70-79 fairly easy, 80-89 easy and 90-100 very easy (Flesch 1948).*

| **Task** | **Continuous segment number** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| ASL continuous | 99 | 99 | 99 | 99 | 97 | 97 | - | - |
| Audiobook continuous | 70 | 80 | 66 | 66 | 82 | 84 | 80 | 89 |

Comparing the false alarms made on the *ASL/continuous* task to those made on the *audiobook/continuous* task also gives an indication that the *ASL/continuous* task was the less taxing of the two. Firstly, fewer false alarms in general were made in the *ASL/continuous* task than in the *audiobook/continuous* task; a total of 75 errors were made by listeners in the *ASL/continuous* compared to 187 in the *audiobook/continuous* task. Listeners did complete two fewer continuous trials in the *ASL/continuous* task but with each listener making, on average, 1.4 errors per trial this would only account for a reduction of 48 errors rather than the 112 observed. Secondly, a smaller proportion of the false alarms made in the *ASL/continuous* task were within 3 words of a genuine substitution; this was both the case for hit/false-alarms and miss/false-alarms. It was argued in Experiment 5 that misassigning substitutions to surrounding words was an indication that cognitive resources were being taxed. A reduction in the number of errors made close to substitutions suggests then that the *ASL/continuous* stimuli were less challenging than the *audiobook/continuous*.

In summary it seems likely that shallow slopes were not seen in the *ASL/continuous* task because the stimuli consisted of sentences which lacked complexity; concatenating disjointed, simple sentences did not sufficiently tax speech understanding processes. That a slope change was seen, however, for the audiobook stimuli when it was presented continuously compared to when it was presented on a trial-by-trial basis suggests then that both complexity *and* continuousness are required to give shallow psychometric functions.

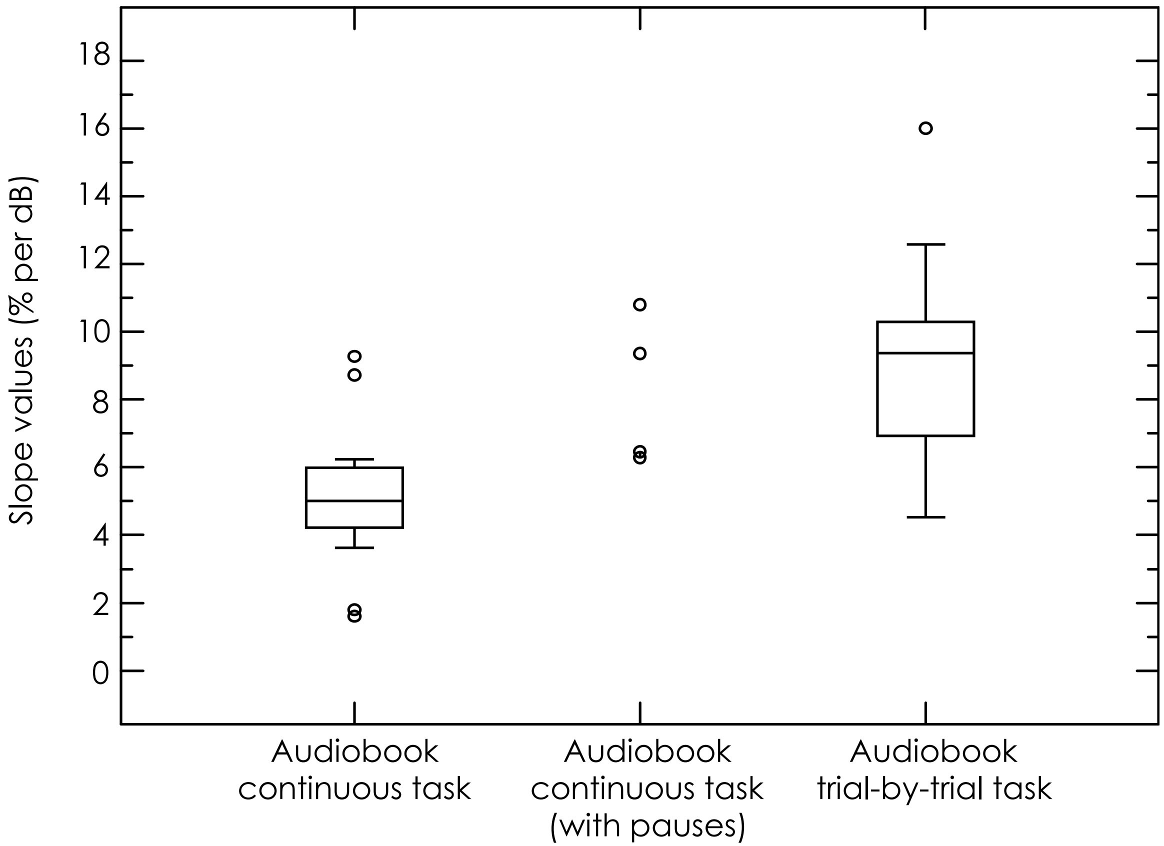
## 6.8 A follow-up experiment to further consider the role of continuousness on the slope of the psychometric function

To further consider the roles of continuousness and complexity a short follow up experiment to Experiments 5 and 6 was carried out. It was reasoned that if these factors play a role in producing the shallow slopes seen for the *audiobook/continuous* task then removing one of these factors should steepen the slope of the psychometric function. In Experiment 6, the complexity of target speech was reduced by using ASL stimuli in the continuous task and slopes for this condition were indeed found to be steeper than for the *audiobook/continuous* task. In the follow-up experiment the target speech remained complex and connected but the continuousness of the stimuli was reduced. It has been argued that the increased effort required to keep up with continuous speech can result in a flattening of the psychometric functions. It was hypothesised, therefore, that introducing pauses to connected, complex speech would result in steeper psychometric functions than if the pauses were not present (i.e. than those given by the *audiobook/continuous* task of Experiment 5).

The task and procedure was identical to that of the audiobook continuous task carried out in Experiment 5, the only difference was that a two second pause was inserted into the audio after each sentence. After each pause the static speech-shaped noise would start and the next sentence of the segment would be presented. Asterisks were included in the typed transcript to clearly mark where pauses would occur. Sentences from each audiobook segment were played in their original order, i.e. the sentences were thematically connected. To maintain the complex nature of the stimuli, pauses were put at the end of each natural sentence. This meant that sentence length varied from 2 to 50 words before pauses. Four older listeners completed the follow-up experiment. They had a mean age of 71 years and a mean hearing loss of 43 dB HL.

Figure 6.20 is a box plot showing the distributions of slope values measured in three conditions: the *audiobook/continuous* (from Experiment 5), the *audiobook/continuous-with-pauses* (Experiment 6 follow up) and the *audiobook/trial-by-trial* (Experiment 6). Box plots are, therefore, from different sets of listeners. Although based on just four listeners, the slopes for the *audiobook/continuous-with-pauses* task were steeper than those given by the *audiobook/continuous* task. All of the slopes measured in the follow-up experiment were steeper than those measured in the *audiobook/continuous* task (apart from two outliers). It is plausible that inserting pauses provided an opportunity for listeners to compensate for any slowing down of processing meaning fewer substitutions were missed due to processing lapses. As a result, the detection of substitutions would have been more influenced by the overall audibility of the speech and as such would increase more rapidly with increases in SNR, hence the steeper psychometric function. Figure 6.20 also indicates that the slopes were slightly shallower in the follow-up experiment than those given in the *audiobook/trial-by-trial* task. This would suggest that reducing the continuousness of the task was not entirely sufficient for altering the balance between the effects of processing lapses and audibility on intelligibility. The more complex nature of the stimuli in the *audiobook/continuous-with-pauses* task compared to the *audiobook/trial-by-trial* task was still enough to result in the occasional processing lapse. Further data is needed, however, to confirm these effects. The results of this follow-up experiment support the suggestion that complexity and continuousness interact to flatten the slope of the psychometric function. As both of these factors act to make the speech understanding more taxing on cognitive resources it is possible that other factors which increase processing load may also flatten the slope of the psychometric function.

***Figure 6.20:*** *Box plots to show the distribution of slope values given by three tasks: the audiobook continuous task, the audiobook continuous task with pauses and the audiobook trial-by-trial task.*



## 6.9 Summary

Experiment 5 used a new continuous speech task to ascertain whether slopes were shallower when target speech was presented continuously than when it was presented one sentence at a time with intervening pauses. Shallower slopes were indeed found for the continuous condition. Experiment 6 considered the relative contributions of the continuousness of speech and the choice of stimuli on this effect. The findings from this experiment suggested that while continuousness did play a role in flattening the slope of the psychometric function, it was only when it was in combination with the complexity of the target speech. A brief follow-up experiment further supported this, finding target speech stimuli which were complex but not continuous gave steeper slopes than target speech that was both continuous and complex.

The main aim of this series of experiments was to gain a better understanding of the shapes of psychometric functions given by realistic listening situations. Most of the speech encountered in everyday situations is likely to be both complex and continuous and, as such, it can be concluded that psychometric functions for everyday situations are likely to be shallower than those predicted by standard audiological tests which used syntactically simple speech on a trial-by-trial basis.

# 7 General Discussion

The research reported in this dissertation was concerned with gaining a better understanding of the changes that occur to the shape of psychometric function for speech intelligibility. The main aims were to 1) quantify the variations in the slope across studies and experimental conditions 2) identify the conditions and mechanisms which may account for these variations in slope, 3) investigate whether realistic listening environments give shallower slopes than standard speech-in-noise tests, and 4) consider how slopes are affected by aging and hearing-impairment. The results are discussed below in terms of each of these four initial aims.

## 7.1 Aim 1: The amount of variation in the slopes of psychometric functions.

Chapter 2 reported a systematic survey of psychometric functions that was carried out to collate the slope data already available in the literature. One aim of the slope survey was to establish how much the slope of the psychometric function varies across experimental designs and conditions. If the range of slope changes were relatively small then identifying factors which affected the slope would be of little importance as their differential effect would be negligible. A median slope of 6.4% per dB was found but the variation in slope was very large indeed; slopes of functions identified in the survey ranged from 1% per dB (or shallower[27]) to 29% per dB. This variation in slope highlighted the need to explore the situations which might result in shallow psychometric functions.

The bootstrap analysis carried out in Chapter 3 gave an indication of the degree of accuracy to which the slopes of the psychometric functions could be measured using current methods. The analysis demonstrated that there was an inherent relationship between the slope of the psychometric function and the size of the confidence interval; while relatively shallow slopes (around 4% per dB) could be measured to

[27] "Flat" and "u-shaped" psychometric functions were also identified in the slope survey which gave slopes of 1% per dB or less. These were, however, poor fits to the logistic curves fitted to the data and so were excluded from the main analysis.

within 2.5% per dB, the range of inaccuracy was inflated as slopes became steeper (e.g. slopes around 12% per dB could only be measured to within 4% per dB). This had consequences for interpreting any differences seen between slopes. A greater slope difference, for example, was required to establish an effect if the psychometric functions were steep (e.g. in Experiment 5 where static noise maskers were used) compared to when functions were shallow (e.g. in Experiments 3 and 4 where slopes were extremely shallow).

## 7.2 Aim 2: The listening situations and mechanisms which result in changes to the slope of the psychometric function.

In Chapter 1 four broad underlying mechanisms to slope changes were identified. Amplitude modulated maskers, high degrees of target/masker confusion, restricted top-down information, and stimuli or performance variability were all proposed to lead to the flattening of an "ideal" psychometric function. The main aim of the systematic slope survey reported in Chapter 2 was to translate these slope-change mechanisms into specific listening conditions which would give shallow psychometric functions. Two major trends emerged from the survey; (1) the general type of masker could affect slope, with speech maskers giving shallower slopes (median = 3.7% per dB) than noise maskers (median = 7.5 % per dB) and (2) the number of maskers could also affect slope, with slopes becoming shallower as the number of maskers were reduced (median = 10% per dB for 3 maskers and 3.6% per dB for 1 masker). Within the two main slope trends other, more minor, factors were also found to affect slope. Several of these indicated that slopes became shallower when target and masker speech become more similar to one another.

Experiment 1 (Chapter 3) was designed, therefore, to further investigate the role that confusion between targets and maskers could have on the slope of the psychometric function, focusing particularly on confusion caused by linguistic as opposed to acoustic similarity of the stimuli. It was reasoned that removing intelligible linguistic information from the masker should reduce its linguistic similarity to the target. This was indeed found, but further investigation indicated that an alternative mechanism was likely to underlie this slope change: differences in the modulation spectrums of the two maskers rather than differences in linguistic similarity to the target were proposed to have resulted in the slope change.

Experiment 2 (Chapter 4) further established the distinction between shallow slopes which occur due to amplitude modulations in the masker and those that occur due to confusion between the target and the masker. It was found that at low SNRs modulated maskers would give shallower slopes than static-noise maskers, but no further difference in slope was seen at this level if the modulated masker's similarity to the target, either acoustic or linguistic, was manipulated. It was concluded that the main factor flattening the slope of the psychometric function at this SNR range was the amplitude modulations in the masker. At higher SNRs (around 0 dB) target/masker similarity played a crucial role with only maskers which were both acoustically and linguistically similar to the target, giving shallow psychometric functions. It was concluded that the main factor flattening the slope of the psychometric function at this level was confusion between the target and the masker. Maskers containing amplitude fluctuations and maskers which are highly confusable with the target are, therefore, both factors which result in shallow psychometric functions. The psychometric functions that each of these conditions gives, however, are both quantitatively and qualitatively different and can thus be classed as different *types* of shallow slope which result from different mechanisms.

Experiments 3 and 4 (Chapter 5) focused on the confusion-based shallow slopes defined in Experiment 2. These unusual-shaped psychometric functions have been measured elsewhere in the literature often using the CRM-in-CRM paradigm or a similar highly confusable speech corpus (e.g. Brungart, 2001a). An attentional framework was applied in these experiments in an attempt to explain how confusion arises when using such stimuli and whether this confusion can be reduced. It was reasoned that it would be easier to distinguish between the target and masker and selectively attend to one if the two sentences were edited so that their onsets no longer aligned with one another. It was proposed that this release from confusion would be observed as a change in the shape of the psychometric function. Slope changes were indeed seen, both when attention was directed towards the target and when directed to the masker. It was concluded that improving selective attention can reduce target/masker confusion and alter the shape of confusion-based psychometric function.

Experiments 5 and 6 (Chapter 6) investigated a third type of shallow slope. The psychometric function for continuous target speech was compared to those given by more standard, trial-by-trial approaches. The functions for continuous speech were found to be shallower. As the masker in this experiment was a static noise, the change in slope could not be attributed to either of the mechanisms (amplitude modulations or target/masker confusion) used to explain the occurrence of shallow slopes in the previous experiments. Instead there was evidence to suggest that occasional difficulties keeping up with speech processing resulted in lapses, meaning some words were missed. Essentially, the change in slope is due to the variability in the underlying identification process.

The fourth slope-change mechanism proposed in the literature − the role of top-down information – was not considered in the experimental work of this thesis. Several trends did emerge from the slope survey, however, which provided evidence in support of this mechanism. Unpredictable sentences, short target utterances, and open-set target corpuses all gave shallow slopes. It was argued that in these cases slopes became shallower as less top-down information was available to help identify speech and a greater reliance was placed on the bottom-up information from the signal. It may be of interest in the future to consider a possible interaction between the availability of top-down information and the quality of the bottom-up signal. It seems plausible that a reduction in top-down information would have a greater detriment on speech intelligibility, and result in shallower slopes, if the target speech were degraded in some way (e.g. speech in reverberant conditions) than if it were not. An experiment such as this would help to further understand the effect that the balance between top-down and bottom-up information has on slope.

The effect of masker modulations on slope changes was also only partially considered in this thesis and could be an interesting area for further research. Howard-Jones and Rosen (1993) have previously reported the effects that different modulation durations have on the slope of the psychometric function and a comprehensive analysis of the effects that different rates, depths and durations of modulations have on the slope would further add to this information. It would help understand exactly how the acoustics of a particular masker (or mixture of maskers) might affect the psychometric function.

## 7.3 Aim 3: Shallow slopes and everyday listening situations.

The importance of studying situations which result in shallow psychometric functions has been emphasised throughout this thesis, as they directly reflect situations where the perceptual benefit received from a given increase in level is likely to be small. It is equally important, though, to establish whether shallow slopes actually occur in real life situations and can thus have consequences for actual listeners or whether they are simply a laboratory phenomenon.

Experiments 2-4 demonstrated that high degrees of confusion between a target and a masker can lead to extremely shallow psychometric functions. The CRM paradigm, for example, has been shown to produce flat or U-shaped psychometric functions (Brungart, 2001a). This effect has been shown to be robust with several studies replicating similar shaped psychometric functions (Brungart & Simpson, 2002; Ihlefeld & Shinn-Cunningham, 2008b). As demonstrated in Chapter 4, however, the situations required to produce these functions are extremely specific. They are only found to occur with a competing speech masker spoken by the same or a very similar sounding talker to the target. Both messages must also be semantically and linguistically similar and presented at equivalent levels. These "confusion-based" shallow psychometric functions are very unlikely, therefore, to occur in everyday listening conditions.

The slope survey in Chapter 2 demonstrated that manipulating the general type of background sound used could affect the slope of the psychometric function. Competing sounds often encountered in every day listening situations, such as speech, were found to give shallow slopes. This finding gave an initial indication that slope changes might then have consequences for speech understanding beyond the lab. Experiment 5 further considered the occurrence of shallow slopes in everyday listening conditions, this time looking at the effect that continuous target speech could have on the slope of the psychometric function. When compared to a trial-by-trial task, the continuous speech task gave shallower psychometric functions. It was postulated that part of this slope effect could be due to the complexity of the speech used in the continuous task: Experiment 6 and a follow up experiment confirmed that target speech needed to be *both* continuous and complex to result in a flattening of the psychometric function. Speech encountered on a daily basis *is* usually both continuous and complex. While most speech tests use short structured utterances, everyday speech is much more variable. Depending on who is speaking, the components (words), structure and length of sentences can vary greatly. Speech is also often ongoing with unpredictable pauses between words or sentences. The results reported in Chapter 6, therefore, provide strong evidence to suggest that slopes given in realistic conditions are likely to be shallower than those given in laboratory conditions. Again, this has consequences for predicting the benefit received from an increase in SNR provided, say, by a hearing aid, as standard speech-in-noise tests are likely to overestimate the benefit a listener is likely to receive in everyday listening situations.

The investigations reported in this thesis into the effects that more realistic situations can have on the slope of the psychometric function point to several interesting areas for future research. The results of Experiments 5 & 6 suggested, for example, that continuousness and complexity result in a flattening of the psychometric function, and while the stimuli used in these experiments were a better representation of everyday speech than trial-by-trial speech-in-noise tests, they still avoided some of the more complex situations faced in everyday listening situations. The continuous task stimuli were, for example, highly coherent stories on a single topic spoken by a single talker. One listening situation where older listeners often report particularly difficulty however, are those where there are several people talking at once (CHABA, 1988). Lively conversations often consist of partially formed sentences, are punctuated by interruptions, and topics and talkers change rapidly. If variations in processing (i.e. occasional lapses) result in shallow slopes for the relatively simple continuous task reported here, it seems likely that even shallower slopes still will be seen under the greater variations and more complex demands of mutlitalker conversations. Indeed, our data gives an indication that the more complex a listening situation becomes the shallower the psychometric function will be. This may go some way to explaining why some listeners find their hearing aids to be of little benefit in very complex listening situations. It is also a problem for which simply improving the gain on a hearing aid would do little to solve.

A longer version of the continuous speech task could also be a useful tool for examining listening effort. It has been suggested that even when audible older listeners find understanding speech much more effortful than younger listeners. Listener effort has been measured in the past using, among other things, response times (Larsby, Hallgren, Lyxell, & Arlinger, 2005), word recall (Sarampalis, Kalluri, Edwards, & Hafter, 2009) and EMG activity (Strauss, Corona-Strauss, & Froehlich, 2008). The advantage of a continuous speech task, such as the one introduced in this thesis, would be that listening effort could be measured over an extended period of time. Monitoring the number of substitutions that are identified as the task progressed would, for example, provide a way of quantifying the declines in speech intelligibility which build-up over time due to increased effort.

## 7.4 Aim 4: Slope changes and different listener groups

The slope survey demonstrated that a number of factors can affect the slope of the psychometric function. It is possible that, along with stimulus qualities, listener qualities might also affect slope; age and hearing impairment may, for example, also have an effect on the shape of the function. Such a listener effect would be of particular interest as slope changes are likely to be most consequential for these listeners. They often report great difficulties understanding speech in noisy environments and would, as such, be the groups to benefit most from any gain in SNR.

While there was an indication from the slope survey that slope changes might differ for listeners of different age groups (the distribution of slopes given by older listeners was significantly steeper than that given by younger listeners) little experimental evidence was found to support this. In Experiment 1 the psychometric function given by both older and younger listeners for four different types of maskers were measured. While threshold differences were seen between older and younger listeners across the different masking conditions, there was no significant difference between the age groups in terms of slope. The slope effects seen in Experiment 1 were attributed to modulation differences between the maskers and so it can be argued that, at least for this slope-change mechanism, there was no additional effect of age on slope.

Experiments 3 and 4 also allowed a comparison of psychometric functions given by older and younger listeners to be made. In this experiment shallow functions were the result of a different slope-change mechanism; target/masker confusion. While threshold differences indicated that when target and masker speech was highly confusable (two full aligned CRM sentences spoken by the same person) older listeners were marginally more susceptible to confusion than younger listeners, the same general slope patterns were still seen across conditions for both older and younger listeners.

In Experiment 5 the shallow slopes given by the continuous task were attributed to the underlying variation in listeners' performance due to difficulties keeping up with the continuous and complex nature of the target speech. Age-related effects on slope cannot be completely ruled out here as only older listeners with a normal to moderate hearing loss took part in this experiment. It could be speculated, however, that age may affect the slopes of these functions. Younger listeners may have fewer difficulties keeping up with the continuous speech and if this is the case they would make fewer processing lapses and this reduction in performance variability would lead to steeper slopes. The same slope difference between continuous, complex speech and speech presented on a trial-by-trial basis may not have been seen with these listeners. This would be an interesting question to address in the future because if this was indeed the case then a continuous task may be a better predictor of a listener's handicap in normal discourse than standard speech-in-noise tests. Improvements could be made to the current continuous task to test this hypothesis directly. Creating and recording new stimuli would, for example, remove some of the unwanted aspects of the audiobook stimuli used in Chapter 6, such as the variations in reading difficulty, intonation, and level. Automating the test so that the text scrolled across the screen at the same rate as the audio would also reduce issues with individual difference in reading rate and participants losing track of their place in the transcript. Comparing the psychometric functions given by older and younger listeners on this new continuous task and on a standard trial-by-trial task may then give us an insight into age-related effects of listening to continuous speech.

As with age, there was also little experimental evidence to suggest that hearing-impairment had any effect on slope. In several of the experiments carried out in this thesis, correlations were computed between listener's better ear averages and slope values found in particular conditions. Of the eight correlations made only one was found to be significant: when audiobook trial-by-trial sentences were presented in static noise slopes were found to become shallower as listeners' hearing losses increased (section 6.6.1). In all other conditions tested in this thesis no relation was found between the magnitude of the slope and the listeners' degree of hearing loss. Furthermore, in section 5.8 the slopes given by older hearing-impaired listeners were directly compared to those given by younger and older normal-hearing listeners. The slopes were not found to substantially differ across listener groups suggesting hearing impairment did not differentially affect slope.

## 7.5 Summary

This thesis has shown that the slope of the psychometric function for speech intelligibility varies greatly and that many factors can contribute to these slope changes. Understanding these slope changes not only has relevance for future experimental work, suggesting that care needs to be taken in selecting both target and masker stimuli, but is also crucial if we wish to quantify the amount of perceptual benefit a listener is likely to gain from any change in SNR offered by a hearing aid (for example, the perceptual benefit that can be gained from a given SNR increase is halved if a speech masker rather than a static-noise masker is used). Continuous listening conditions were found to reduce the perceptual benefit given by an improvement in SNR. The evidence also indicates that in the complex demands of everyday listening situations this benefit may be yet further diminished.

## Appendix A - Validation of the bootstrap method

To be sure that the bootstrap method used in Experiment 1 was itself valid. A simulation was carried out to derive confidence intervals for a data set in which the variation was known in advance. First, values of m and c were randomly taken from two predefined normal distributions with a mean and standard deviation of 5 and 2% per dB or -15 and 3 dB respectively. Second these values were then substituted into the logistic equation (equation 1.1) to generate $y$ values (speech identification scores) at seven selected SNRs, so simulating the experiment. Third, the synthetic data set was then bootstrapped using the process outlined above to generate 100 bootstrapped values of slope and SRT. Further, the process was repeated for 100 synthetic data sets, all of which had slopes and SRTs with the specified variance. The result of this simulation was that the mean and standard deviations of the bootstrapped values of slopes and SRT equalled those put in of the predefined distributions. This simulation suggest, therefore that the bootstrap method provides a realistic representation of the possible variance in the data.

## Appendix B - CRM target and masker edit lengths

Below is a table displaying the onset and offset times corresponding to the four CRM edit lengths used in Experiments 3 & 4. Times were measured using Praat. Onset and offset times (ms) are given for all CRM sentences spoken by a male talker with the callsign Baron or with the callsign Arrow (Kitterick, Bailey and Summerfield, 2010). Baron sentences were used as targets in the experiments and Arrow sentences were used as maskers.

| Keywords | Ready | Callsign | go | to | colour | number | now | offset |
|---|---|---|---|---|---|---|---|---|
| **Baron** | | | | | | | | |
| Blue one | 0 | 248 | 943 | 1089 | 1256 | 1485 | 1807 | 2375 |
| Blue two | 0 | 256 | 1060 | 1207 | 1340 | 1683 | 2073 | 2582 |
| Blue three | 0 | 289 | 1041 | 1209 | 1370 | 1588 | 2063 | 2576 |
| Blue five | 0 | 273 | 1030 | 1150 | 1351 | 1604 | 1977 | 2470 |
| Blue four | 0 | 275 | 1000 | 1100 | 1266 | 1549 | 2037 | 2484 |
| Blue six | 0 | 279 | 1081 | 1215 | 1340 | 1640 | 2127 | 2573 |
| Blue seven | 0 | 272 | 960 | 1080 | 1182 | 1470 | 1897 | 2485 |
| Blue eight | 0 | 300 | 1053 | 1184 | 1300 | 1633 | 2026 | 2500 |
| Green one | 0 | 254 | 1012 | 1240 | 1450 | 1763 | 2183 | 2554 |
| Green two | 0 | 250 | 895 | 1040 | 1185 | 1613 | 1950 | 2450 |
| Green three | 0 | 295 | 1185 | 1350 | 1475 | 1810 | 2242 | 2714 |
| Green four | 0 | 291 | 920 | 1085 | 1138 | 1513 | 1888 | 2369 |
| Green five | 0 | 252 | 845 | 991 | 1064 | 1420 | 1963 | 2436 |
| Green six | 0 | 270 | 1086 | 1228 | 1404 | 1637 | 2198 | 2620 |
| Green seven | 0 | 263 | 1002 | 1151 | 1236 | 1638 | 2101 | 2499 |
| Green eight | 0 | 290 | 1094 | 1228 | 1336 | 1760 | 2206 | 2656 |
| Red one | 0 | 277 | 974 | 1095 | 1210 | 1574 | 1975 | 2370 |
| Red two | 0 | 305 | 973 | 1129 | 1305 | 1582 | 1967 | 2393 |
| Red three | 0 | 257 | 960 | 1109 | 1212 | 1456 | 1933 | 2379 |
| Red four | 0 | 249 | 1066 | 1192 | 1305 | 1572 | 1964 | 2415 |
| Red five | 0 | 296 | 1040 | 1198 | 1313 | 1603 | 2121 | 2554 |
| Red six | 0 | 268 | 1094 | 1247 | 1417 | 1548 | 2055 | 2509 |
| Red seven | 0 | 286 | 958 | 1137 | 1283 | 1457 | 1991 | 2361 |
| Red eight | 0 | 295 | 1029 | 1181 | 1290 | 1617 | 2048 | 2536 |
| White one | 0 | 277 | 912 | 1045 | 1123 | 1516 | 1835 | 2310 |
| White two | 0 | 268 | 973 | 1121 | 1241 | 1682 | 2010 | 2457 |
| White three | 0 | 268 | 938 | 1112 | 1282 | 1500 | 2000 | 2587 |
| White four | 0 | 288 | 910 | 1056 | 1223 | 1457 | 1877 | 2367 |
| White five | 0 | 296 | 818 | 1335 | 1451 | 1686 | 2225 | 2719 |
| White six | 0 | 287 | 1227 | 1360 | 1473 | 1775 | 2255 | 2700 |
| White seven | 0 | 298 | 891 | 1042 | 1164 | 1432 | 1908 | 2496 |
| White eight | 0 | 276 | 1130 | 1287 | 1424 | 1704 | 2141 | 2622 |

## Appendix B continued

| Keywords | Ready | Callsign | go | to | colour | number | now | offset |
|---|---|---|---|---|---|---|---|---|
| **Arrow** | | | | | | | | |
| Blue one | 0 | 364 | 1000 | 1131 | 1247 | 1583 | 1879 | 2417 |
| Blue two | 0 | 324 | 1044 | 1194 | 1339 | 1671 | 1980 | 2508 |
| Blue three | 0 | 394 | 1210 | 1363 | 1500 | 1788 | 2184 | 2641 |
| Blue five | 0 | 346 | 975 | 1119 | 1207 | 1526 | 1893 | 2357 |
| Blue four | 0 | 394 | 1139 | 1273 | 1382 | 1705 | 2182 | 2669 |
| Blue six | 0 | 389 | 1129 | 1273 | 1467 | 1715 | 2143 | 2561 |
| Blue seven | 0 | 401 | 1219 | 1369 | 1474 | 1798 | 2218 | 2677 |
| Blue eight | 0 | 399 | 947 | 1233 | 1257 | 1519 | 1955 | 2517 |
| Green one | 0 | 355 | 1268 | 1422 | 1514 | 1971 | 2299 | 2755 |
| Green two | 0 | 368 | 970 | 1121 | 1316 | 1699 | 2085 | 2623 |
| Green three | 0 | 336 | 1114 | 1269 | 1379 | 1729 | 2188 | 2677 |
| Green four | 0 | 403 | 1252 | 1402 | 1548 | 1870 | 2292 | 2694 |
| Green five | 0 | 311 | 1042 | 1176 | 1300 | 1601 | 2060 | 2561 |
| Green six | 0 | 347 | 932 | 1080 | 1201 | 1525 | 1952 | 2406 |
| Green seven | 0 | 334 | 1127 | 1250 | 1421 | 1769 | 2161 | 2631 |
| Green eight | 0 | 349 | 1184 | 1329 | 1500 | 1884 | 2315 | 2787 |
| Red one | 0 | 358 | 941 | 1080 | 1267 | 1559 | 1890 | 2338 |
| Red two | 0 | 362 | 1025 | 1154 | 1321 | 1668 | 2011 | 2484 |
| Red three | 0 | 338 | 1069 | 1187 | 1320 | 1533 | 2059 | 2549 |
| Red four | 0 | 324 | 1135 | 1285 | 1442 | 1660 | 2012 | 2513 |
| Red five | 0 | 382 | 1084 | 1196 | 1330 | 1521 | 2046 | 2497 |
| Red six | 0 | 343 | 1114 | 1251 | 1427 | 1632 | 2093 | 2549 |
| Red seven | 0 | 369 | 1218 | 1358 | 1502 | 1666 | 2182 | 2669 |
| Red eight | 0 | 348 | 1036 | 1195 | 1364 | 1598 | 1963 | 2507 |
| White one | 0 | 353 | 1239 | 1373 | 1546 | 1834 | 2175 | 2661 |
| White two | 0 | 362 | 1204 | 1324 | 1474 | 1905 | 2251 | 2687 |
| White three | 0 | 322 | 807 | 942 | 1094 | 1340 | 1855 | 2411 |
| White four | 0 | 382 | 1080 | 1248 | 1407 | 1730 | 2105 | 2582 |
| White five | 0 | 315 | 823 | 967 | 1124 | 1322 | 1847 | 2413 |
| White six | 0 | 320 | 1013 | 1143 | 1295 | 1568 | 2021 | 2479 |
| White seven | 0 | 329 | 870 | 1016 | 1149 | 1420 | 1881 | 2382 |
| White eight | 0 | 385 | 1124 | 1260 | 1430 | 1751 | 2052 | 2612 |

## Appendix C - Example transcript from the continuous task

Below is a to scale copy of one of the transcripts from the audiobook/continuous task used in Experiment 5. In this annotated version substitutions are underlined and the original words (i.e. the words that appeared in the audio) are written above - these annotations did not appear in the copies given to the participants.

Page | 1

IHR number =      Date =      Time =

Noise condition =      SNR =

Under him were three lads; for the establishment was a small one, containing only
*Sat*
four horses in all. One of these lads waited up each night in the stable, while the
*loft*
others slept in the barn. All three bore excellent characters. John Straker, who is a
*small*
married man, lived in a little villa about two hundred yards from the stables. He
*comfortably*
has no children, keeps one maid-servant, and is adequately off. The country round
*north*
is very lonely, but about half a mile to the east there is a small cluster of villas
*contractor*
which have been built by a Tavistock investor for the use of invalids and others
*pure* *lies*
who may wish to enjoy the fresh Dartmoor air. Tavistock itself sits two miles to
*moor*
the west, while across the hills, about two miles distant, is the larger training
*establishment* *managed*
stables of Capleton, which belongs to Lord Backwater, and is run by Silas Brown.
*complete*
In every other direction the moor is a vast wilderness, inhabited only by a few
*roaming*
wandering gypsies. Such was the general situation last Monday night when the
*catastrophe*
trouble occurred.

*exercised*
"On that evening the horses had been fed and watered as usual, and the stables
*lads*
were locked up at nine o'clock. Two of the boys walked up to the trainer's house,
*supper* *third*
where they had dinner in the kitchen, while the other, Ned Hunter, remained on

## Appendix C continued.



IHR number =　　　　Date =　　　　Time =

Noise condition =　　　　SNR =

guard. At a few minutes *after* before nine the maid, Edith Baxter, carried down to the *stables* yard his supper, which consisted of a *dish* bowl of curried mutton. She took no liquid, as there was a water-tap in the stables, and it was the rule that the lad on *duty* guard should drink nothing else. The maid *carried* took a lantern with her, as it was very dark and the path ran across the open moor.

"Edith Baxter was within thirty yards of the stables, when a man *appeared* sprung out of the darkness and called to her to *stop* wait. As he stepped into the *circle* glow of yellow light thrown by the lantern she saw that he was a *person* man of gentlemanly bearing, dressed in a *grey* brown suit of tweeds, with a *cloth* flat cap. He wore gaiters, and carried a *heavy* walking stick with a knob to it. She was most *impressed* surprised, however, by the extreme pallor of his face and by the *nervousness* uneasiness of his manner. His age, she *thought* guessed, would be rather over thirty than under it.

" 'Can you tell me where I am?' he asked. 'I had *almost* nearly made up my mind to sleep on the moor, when I saw the light of your lantern.' " 'You are close to the King's Pyland training stables,' said she.

## Appendix C continued.

Page | 3

IHR number = Date = Time =

Noise condition = SNR =

“ ‘Oh, indeed! What a stroke of luck!’ he cried. ‘I understand that a stable-boy sleeps there alone every night. Perhaps that is his *supper* dinner which you are carrying to him. Now I am sure that you would not be too proud to earn the price of a *new* nice dress, would you?’ He took a piece of *white* grey paper folded up out of his waistcoat pocket. ‘See that the boy has this to-night, and you shall have the *prettiest* loveliest frock that money can buy.’

“She was frightened by the earnestness of his *manner* speech, and ran past him to the window through which she was accustomed to *hand* pass the meals. It was *already* quickly opened, and Hunter was seated at the *small* low table inside. She had begun to tell him of what had *happened* occurred, when the stranger came up again.

“ ‘Good-evening,’ said he, *looking* peering through the window. ‘I wanted to have a word with you.’ The *girl* maid has sworn that as he spoke she noticed the *corner* edge of the little paper packet protruding from his closed *hand* fist.

“ ‘What business have you here?’ asked the *lad* boy. “ ‘It’s business that may put something into your pocket,’ said the other. ‘You’ve two horses in for the Wessex

## Appendix C continued.



IHR number =　　　Date =　　　Time =

Noise condition =　　　SNR =

Cup—Silver Blaze and Bayard. Let me have the straight tip and you won't be a

*weights*

loser. Is it a fact that at the fence Bayard could give the other a hundred yards in

*money*

five furlongs, and that the stable have put their bets on him?'

## Appendix D - Classification of false alarms

Words which were erroneously identified as substitutions in the continuous tasks of Experiments 5 & 6 were termed false alarms. The example transcript below illustrates how false alarms were classified. Genuine substitutions have been underlined and the shaded rectangles illustrate listener responses. The example transcript contains 5 numbered errors. Errors were classified as either hit/false-alarm or miss/false-alarm depending on whether they were closest to an identified substitution or not. Error (1), for example would have been classified as a minus-4, hit/false-alarm while errors (2) and (4) would be classed as a minus-5, miss/false-alarm and a plus-1, miss/false-alarm respectively. On occasion listeners would mark more than one word at once as in errors (3) & (5). If, like error (3), the substituted word was clearly marked in the same stroke as a non-substituted word, this was assumed to be “a slip of the pen” and the non-substituted word was not classed as a false-alarm.  If however, like error (5) neither marked word was a genuine substitution the mark is considered as just one word and the number of intervening words to a genuine substitution counted, i.e. error (5) would be classed as a plus-2, miss/false-alarm.

They still had hopes that the trainer had for some purpose taken out the horse for early exercise, but on reaching the knoll near the house, from which all the neighbouring lands were visible, they not only could see no ① signs of the missing pair, but they perceived something which alerted them that they were in the presence of a tragedy.

“Around a quarter of a mile from the ② stables John Straker’s overcoat was dangling from a furzebush. Immediately beyond there was a basin shaped depression in the moor, and at the bottom of this was found the dead body of ③ the ill-fated trainer. His head had been dashed by a savage blow from some blunt ④ weapon, and he was wounded on the leg, where there was a long, clean cut, inflicted evidently by some very sharp instrument. It was evident, however, that Straker had defended himself vigorously against his attackers, for in his right hand he held a small knife, which was clotted with blood up to the handle, while in his left he gripped a red and black silk cravat, which was identified by the maid as having been worn on the preceding evening by the gentleman who ⑤ had visited the stables. Hunter, on recovering from his stupor, was also quite positive as to the ownership of the cravat.

## Appendix E - Audiobook trial-by-trial sentence corpus

Below is a list of the 210 sentences used in Experiment 6 in the *audiobook/trial-by-trial* task.

He is now in his fifth year
We should look for his tracks
The trainer was a retired jockey
He took a piece of white paper
He would surely have been seen
The girl fled away to the house
As she ran she looked back
The reds had all faded to greys
The boy locked the door
They are booth sound sleepers
The two women ran out
He stood pointing
He was wounded in the thigh
He held a small knife
Like a dog with its master
The sun was beginning to sink
He turned very pale
On Tuesday they were gone
A silver watch and a gold chain
Like a branch in the wind
Her face was haggard and thin
He followed the inspector outside
A short walk across the moor
There was no wind that night
He turned upon his heel
He took the boots from the bag
I follow my own methods
We walked slowly across the moor
A tall young man entered the room
He tried to bluster out of it
He turned back with the inspector
With his finger and thumb in his waistcoat pocket
What has become of the horse
Supposing he broke away
Where could he have gone to
Every precaution was taken to guard the favourite
I walked down the bank
I heard him give a shout
She noticed the corner of the little paper packet
Again the ground sloped
About a quarter of a mile from the stables
We lost them for half a mile
His face was ashy pale
The stranger was leaning through the window
They do not wish to be pestered by the police
He was a man of excellent birth and education
He grasped a red and black silk cravat
Your instructions will be done
There must be no mistake
He started violently and flushed to the temples
I shall write to you about it
You shall hear from me tomorrow
He had picked it up as he had left the room
His bullying overbearing manner was all gone
He will guard it as the apple of his eye
The ground has been trampled up a good deal
Say nothing to him about the horse
The inspector opened his eyes
You have a few sheep in the paddock
His attention had been keenly aroused
I glanced at the card to see the entries
One mile and five furlongs

## Appendix E continued

Yellow and black stripes
He handed me a short note
He lit his pipe
He was the only friend I made
I heard a whisper close to my ear
He was a young man
At first it was only a minutes chat
But soon his visits lengthened
The house seemed to be at his mercy
At first I thought it meant nothing
The stateroom was next to the cabin
A small but select library
My friend was his only son
There had been a daughter I'd heard
The father interested me extremely
He was a man of little culture
He was a thick set burly man
We heard a cry for help
He stared at me with great surprise
You have been in New Zealand
He gave a gasp or two and sat up
He wore an open jacket
His crinkled hands were half closed
He was back in a moment
He had grown thin
He explained the situation
The lapse of two years
She is away upon a visit
But at last he made a trip
That is to say on Monday next
He sat down opposite to me

I can write with all truth and honesty
Jumping out of his chair he ran into the house
He carried his head very jauntily in the air
I was glad to find that he was my neighbour
The house was an old fashioned building
I remembered hearing of his case
Well where do you suppose the balance is
Right between my finger and thumb
He held up a little crumpled piece of paper
The man knew what was up in an instant
He had unlocked the door that lead to the deck
We found we had some subjects in common
In five minutes it was all over
There was no one left of our enemies
It was a long hour before we reached it
There was no sign of life
The old man thought his son was exaggerating
Next day we were picked up
During the two years I was at college
There had been three accounts in the press
The only light in the room came from the lamp
Might I trouble you for a match
He held out his hand
I saw in the light of the lamp
We were dashing along a country road
There was something very strange in all this
He saw the question in my eyes
Glimmering in the red light of the sun
I couldn't sit quiet in my chair
I could already see the high chimneys
Night after night I heard him pacing his room

## Appendix E continued

The clearing up of over 40 mysteries
They may slip out of our hands
All would have been well
He saw every step which I took
He strove to break away
He cut deep
My heart turned to lead
Few details came out
He is extremely tall and thin
He is clean shaven
He peered at me with great curiosity
As I left the dinning room
The situation is an impossible one
Danger is part of my trade
I know every move of your game
You can do nothing before Monday
A brick came down from the roof
You must stop work now
The girl was taken to her room
His bed had not been slept in
She fell back against the wall
His boots were left behind
His clothes were in his room
A chill of fear had come over me
The doors had all been shut
Every precaution is still necessary
I saw a tall man
No great harm was done
He will catch us there
The lake there is eight feet deep
Close to the gravel path

Lined by glistening coal black rock
We stood near the edge
Walk slowly over the hill
A look of surprise past over his face
This morning the last steps were taken
Pulling the letter from my pocket
If I may make a full confession
His appearance was quite familiar to me
The best and the wisest man
You crossed my path on the fourth of January
She had only recently recovered from an illness
Come back when you are stronger
She looked at me with so strange an expression
We will see what the doctor says
I called the police and had the place examined
You might find me a dangerous guest
With shriek after shriek of laughter
Horrified at this sudden hysterical attack
There were slates and bricks piled upon the roof
I had often admired my friends courage
Both windows and doors were fastened
The second first class carriage from the front
He turned the carriage and dashed away
We ransacked every room and the attic
Rain had fallen on the night before
A nurse had been employed to sit up with her
When she woke in the early morning
It would be to ruin the work of three months
Let us have him arrested on his arrival
Old rusted and discoloured metal
About two hundred yards from the building

## Appendix E continued

For five days at a stretch
He lives at the lodge with his mother
The truth was still discovered
It came through the post
One day in early spring
Whoever wrote that note
The villain got clean away
To some land beyond the seas
It was nine feet in length
I tapped upon the floor
The back door was open
One of them fired a shot

His smile showed that it had pleased him
The twilight had closed in
The lock has been forced
She is very old and deaf
He sank his head upon his knees
They belong to men who are blood relatives
The top most branches of the old oak
We could get no information from her
Bent and twisted out of its original shape
Where is the rest of that sheet of paper
I examined the ground carefully
It was torn out of the dead mans hand

## References

•Acton, W. I. (1970). Speech intelligibility in a background noise and noise-induced hearing loss. *Ergonomics, 13*(5), 546-554.

Allen, P., & Wightman, F. (1994). Psychometric functions for children's detection of tones in noise. *Journal of Speech and Hearing Research, 37*(1), 205-215.

Allen, P., & Wightman, F. (1995). Effects of signal and masker uncertainty on children's detection. *Journal of Speech and Hearing Research, 38*(2), 503-511.

•Arbogast, T. L., Mason, C. R., & Kidd, G. (2002). The effect of spatial separation on informational and energetic masking of speech. *Journal of the Acoustical Society of America, 112*(5), 2086-2098.

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

•Barker, J., & Cooke, M. (2007). Modelling speaker intelligibility in noise. *Speech Communication, 49*(5), 402-417.

•Beattie, R. C. (1989). Word recognition functions for CID W-22 test in multitalker noise for normally hearing and hearing impaired subjects. *Journal of Speech and Hearing Disorders, 54*, 20-32.

•Beattie, R. C., Barr, T., & Roup, C. (1997). Normal and hearing-impaired word recognition scores for monosyllabic words in quiet and noise. *British Journal of Audiology, 31*, 153-164.

•Beattie, R. C., & Clark, N. (1982). Practice effects of a four-talker babble on the synthetic sentence identification test. *Ear & Hearing, 3*(4), 202-206.

Bench, J., Kowal, A., & Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *British Journal of Audiology, 13*(3), 108-112.

•Bernstein, J. G. W., & Grant, K. W. (2009). Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America, 125*(5), 3358-3372.

Best, V., Gallun, F. J., Ihlefeld, A., & Shinn-Cunningham, B. G. (2006). The influence of spatial separation on divided listening. *Journal of the Acoustical Society of America, 120*(3), 1506-1516.

•Best, V., Gallun, F. J., Mason, C. R., Kidd, G., Jr., & Shinn-Cunningham, B. G. (2010). The impact of noise and hearing loss on the processing of simultaneous sentences. *Ear & Hearing, 31*(2), 213-220.

Best, V., Ozmeral, E. J., & Shinn-Cunningham, B. G. (2007). Visually-guided attention enhances target identification in a complex auditory scene. *Journal of the Association for Research in Otolaryngology, 8*(2), 294-304.

•Bhattacharya, A., & Zeng, F. (2007). Companding to improve cochlear-implant speech recognition in speech-shaped noise. *Journal of the Acoustical Society of America, 122*(2), 1079-1089.

•Blue-Terry, M., & Letowski, T. (2011). Effects of white-noise on Callsign Acquisition Test and Modified Rhyme Test. *Ergonomics, 54*(2), 139-145.

Boersma, P., & Weenink, D. (2012). Praat: Doing phonetics by computer (Version 5.3.23) [computer program].

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

Boothroyd, A. (1968). Developments in speech audiometry. *British Journal of Audiology, 2(1)*, 3-10.

•Boothroyd, A. (2008). The performance/intensity function: An underused resource. *Ear & Hearing, 29*, 479-491.

•Boothroyd, A., & Nittrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *Journal of the Acoustical Society of America, 84*(1), 101-114.

•Bosman, A. J., & Smoorenburg, G. F. (1995). Intelligibility of Dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment. *Audiology, 34*(5), 260-284.

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

Broadbent, D. E. (1952). Failures of attention in selective listening. *Journal of Experimental Psychology, 44*(6), 428-433.

Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon.

Bronkhorst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility *Acustica - Acta Acustica, 86*, 117-128.

•Bronkhorst, A. W., Bosman, A. J., & Smoorenburg, G. F. (1993). A model for context effects in speech recognition. *Journal of the Acoustical Society of America, 93*(1), 499-509.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology, 10*(1), 12-21.

•Brungart, D. (2001a). 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. (2001b). Evaluation of speech intelligibility with the coordinate response measure. *Journal of the Acoustical Society of America, 109*(5), 2276-2279.

•Brungart, D. S., Chang, P. S., Simpson, B. D., & Wang, D. L. (2006). Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. *Journal of the Acoustical Society of America, 120*(6), 4007-4018.

•Brungart, D. S., Chang, P. S., Simpson, B. D., & Wang, D. L. (2009). Multitalker speech perception with ideal time-frequency segregation: Effects of voice characteristics and number of talkers. *Journal of the Acoustical Society of America, 125*(6), 4006-4022.

•Brungart, D. S., Iyer, N., & Simpson, B. D. (2006). Monaural speech segregation using synthetic speech signals. *Journal of the Acoustical Society of America, 119*(4), 2327-2333.

•Brungart, D. S., & Simpson, B. (2002). Within-ear and across-ear interference in a cocktail-party listening task. *Journal of the Acoustical Society of America, 112*, 2958-2995.

•Brungart, D. S., & Simpson, B. D. (2004). Within-ear and across-ear interference in a dichotic cocktail party listening task: Effects of masker uncertainty. *Journal of the Acoustical Society of America, 115*(1), 301-310.

•Brungart, D. S., & Simpson, B. D. (2007). Effect of target-masker similarity on across-ear interference in a dichotic cocktail-party listening task. *Journal of the Acoustical Society of America, 122*(3), 1724-1734.

•Brungart, D. S., Simpson, B. D., Darwin, C. J., Arbogast, T. L., & Kidd, G. (2005). Across-ear interference from parametrically degraded synthetic speech signals in dichotic cocktail-party listening task. *Journal of the Acoustical Society of America, 117*(1), 292- 304.

•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*(5), 2527-2538.

•Brungart, D. S., Simpson, B. D., & Freyman, R. L. (2005). Precedence-based speech segregation in a virtual auditory environment. *Journal of the Acoustical Society of America, 118*(5), 3241-3251.

Carhart, R., Tillman, T. W., & Greetis, E. S. (1969). Perceptual masking in multiple sound backgrounds. *Journal of the Acoustical Society of America, 45*(3), 694-703.

CHABA (1988). Speech understanding and aging. *Journal of the Acoustical Society of America, 83*(3), 869-895.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America, 25*(5), 975-979.

•Cienkowski, K. M., & Speaks, C. (2000). Subjective vs. Objective intelligibility of sentences in listeners with hearing loss. *Journal of Speech, Language, and Hearing Research, 43*, 1205-1210.

Cohen, G. (1987). Speech comprehension in the elderly the effects of cognitive changes. *British Journal of Audiology, 21*(3), 221-226.

Conway, A. R. A., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review, 8*(2), 331-335.

Cooke, M. (2006). A glimpsing model of speech perception in noise. *Journal of the Acoustical Society of America, 119*(3), 1562-1573.

•Cooke, M., Hershey, J. R., & Rennie, S. J. (2010). Monaural speech separation and recognition challenge. *Computer Speech and Language, 24*(1), 1-15.

•Cooper, J. R., & Cutts, B. P. (1971). *Journal of Speech and Hearing Research, 14*, 332-337.

Cornbleet, P. J., & Gochman, N. (1979). Incorrect least-squares regression coefficients in method-comparison analysis. *Clinical Chemistry, 25*(3), 432-438.

Cox, R. M., Alexander, G. C., & Gilmore, C. (1987). Development of the Connected Speech Test (CST). *Ear & Hearing, 8*(5), S119-S126.

Cox, R. M., & McDaniel, D. M. (1989). Development of the Speech-Intelligibility Rating (SIR) Test for hearing-aid comparisons. *Journal of Speech and Hearing Research, 32*(2), 347-352.

•Craig, C. H. (1988). Effect of three conditions of predictability on word-recognition performance. *Journal of Speech & Hearing Research, 31*, 588-592.

•Crandell. C. (1993). Speech recognition in noise by children with minimal degrees of sensorineural hearing loss. *Ear & Hearing, 14*(3), 210-216.

Crowder, R. G. (1967). Prefix effects in immediate memory. *Canadian Journal of Psychology, 21*(5), 450-450.

Crowder, R. G., & Morton, J. (1969). Precategorical Acoustic Storage (PAS). *Perception & Psychophysics, 5*(6), 365-373.

•Danhauer, J. L., Doyle, P. C., & Lucks, L. E. (1986). Effects of signal-to-noise ratio on the nonsense syllable test. *Ear and Hearing, 7*(5), 323-324.

•Danhauer, J. L., & Leppler, J. G. (1979). Effects of four noise competitors on the California Consonant Test. *Journal of Speech and Hearing Disorders, 44*(3), 354-362.

Darwin, C. J. (1997). Auditory grouping. *Trends in Cognitive Sciences, 1*, 327-333.

•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*(5), 2913-2922.

Darwin, C. J., & Carlyon, R. P. (1995). Auditory grouping. In B. C. J. Moore (Ed.), *Hearing* (pp.387-420). San Diego: Academic press.

Darwin, C. J., & Hukin, R. W. (2000). Effectiveness of spatial cues, prosody, and talker characteristics in selective attention. *Journal of the Acoustical Society of America, 107*(2), 970-977.

Davis, A. (1995). *Hearing in adults*. London Whurr Publishers Ltd.

•Dirks, D. D., Bell, T. S., Rossman, R. N., & Kincaid, G. E. (1986). Articulation index predictions of contextually dependent words. *Journal of the Acoustical Society of America, 80*(1), 82-92.

•Dirks, D. D., & Bower, D. (1969). Masking effects of speech competing messages. *Journal of Speech and Hearing Research, 12*, 229-245.

•Dirks, D. D., Morgan, D. E., & Dubno, J. R. (1982). A procedure for quantifying the effects of noise on speech recognition. *Journal of Speech and Hearing Disorders, 47*(2), 114-123.

•Dirks, D. D., & Wilson, R. H. (1969a). Binaural hearing of speech for aided and unaided conditions. *Journal of Speech and Hearing Research, 12*(3), 650-664.

•Dirks, D. D., & Wilson, R. H. (1969b). Effect of spatially separated sound sources on speech intelligibility. *Journal of Speech and Hearing Research, 12*(1), 5-38.

•Dirks, D. D., Wilson, R. H., & Bower, D. R. (1969). Effect of pulsed masking on selected speech materials. *Journal of the Acoustical Society of America, 46*(4), 898-906.

Dorman, M. F., Loizou, P. C., & Rainey, D. (1997). Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *Journal of the Acoustical Society of America, 102*, 2403-2411.

Dreschler, W. A., Verschuure, H., Ludvigsen, C., & Westermann, S. (2001). ICRA noises: Artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment. *Audiology, 40*(3), 148-157.

•Drullman, R. (1995). Speech-intelligibility in noise: Relative contribution of speech elements above and below the noise-level. *Journal of the Acoustical Society of America, 98*(3), 1796-1798.

•Drullman, R., & Bronkhorst, A. W. (2004). Speech perception and talker segregation: Effects of level, pitch and tactile support with multiple simultaneous talkers. *Journal of the Acoustical Society of America, 116*(5), 3090-3098.

Drullman, R., Festen, J. M., & Plomp, R. (1994). Effect of reducing slow temporal modulations on speech reception. *Journal of the Acoustical Society of America, 95*(5), 2670-2680.

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

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

•Dubno, J. R., Horwitz, A. R., & Ahlstrom, J. B. (2005). Word recognition in noise at higher-than-normal levels: Decreases in scores and increases in masking. *Journal of the Acoustical Society of America, 118*(2), 914-922.

Duquesnoy, A. J. (1983). Effect of single interfering noise or speech source upon the binaural sentence intelligibility of aged persons. *Journal of the Acoustical Society of America, 74*(3), 739-743.

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*(1), 368-379.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics, 7*, 1-26.

•Egan, J. P. (1948). Articulation testing methods. *Laryngoscope, 58*, 955-991.

•Egan, J. P., Carterette, E., & Thwing, E. (1954). Factors affecting multichannel listening. *Journal of the Acoustical Society of America, 26*, 774-782.

•Eisenberg, L. S., Dirks, D. D., & Bell, T. S. (1995). Speech recognition in amplitude-modulated noise of listeners with normal and listeners with impaired hearing. *Journal of Speech and Hearing Research, 38*, 222-233.

•Elliott, L. L. (1979). Performance of children aged 9 to 17 years on a test of speech-intelligibility in noise using sentence material with controlled word predictability. *Journal of the Acoustical Society of America, 66*(3), 651-653.

•Erber, N. P. (1971). Auditory detection of spondaic words in wide-band noise by adults with normal hearing and by children with profound hearing losses *Journal of Speech and Hearing Research, 14*, 372-381.

Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin, 53*(2), 134-140.

•Ezzatain, P., Li, L., Pichora-Fuller, K., & Schneider, B. A. (2010). The effect of priming on release from informational masking is equivalent for younger and older adults. *Ear & Hearing, 32*(1), 84-96.

Fagerland, M. W., & Sandvik, L. (2009). The Wilcoxon-Mann-Whitney test under scrutiny. *Statistics in Medicine, 28*(10), 1487-1497.

Falconer, G. A., & Davis, H. (1947). The intelligibility of connected discourse as a test for the threshold for speech. *Laryngoscope, 57*(9), 581-595.

•Feeney, M. P., & Franks, J. R. (1982). Test-retest reliability of a distinctive feature difference test for hearing aid evaluation. *Ear & Hearing, 3*(2), 59-65.

•Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *Journal of the Acoustical Society of America, 106*(6), 3578-3588.

Field, A. (2009). *Discovering statistics using SPSS*. London: Sage Publications Ltd.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32*(3), 221-233.

•Foster, J. R., & Haggard, M. P. (1987). The Four Alternative Auditory Feature Test (FAAF) - linguistic and psychometric properties of the material with normative data in noise. *British Journal of Audiology, 21*, 165-174.

Francis, A. L. (2010). Improved segregation of simultaneous talkers differentially affects perceptual cognitive capacity demands for recognizing speech in competing speech. *Attention, Perception & Psychophysics, 72*(2), 501-516.

•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*(5), 2112-2122.

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

•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*(2), 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*(6), 3578-3588.

•Friesen, L. M., Shannon, R. V., Baskent, D., & Wang, X. (2010). 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*(2) 1150-1163.

•Fu, Q. J., Shannon, R. V., & Wang, X. S. (1998). Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing. *Journal of the Acoustical Society of America, 104*(6), 3586-3596.

•Gelfand, S. A. (1998). Optimizing the reliability of speech recognition scores. *Journal of Speech, Language, and Hearing Research, 41*, 1088-1102.

Gibson, E. J. (1969). *Principles of perceptual learing and development*. New York: Meredith Corporation.

Giolas, T. G., & Epstein, A. (1963). Comparative intelligibility of word lists and continuous discourse. *Journal of Speech and Hearing Research, 6*(4), 349-358.

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

Gordon-Salant, S., & Fitzgibbons, P. J. (1999). Profile of auditory temporal processing in older listeners. *Journal of Speech Language and Hearing Research, 42*(2), 300-311.

•Grant, K. W., & Braida, L. D. (1991). Evaluating the articulation index for auditory-visual input. *Journal of the Acoustical Society of America, 89*(6), 2952-2960.

Grant, K. W., & Seitz, P. F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *Journal of the Acoustical Society of America, 108*(3), 1197-1208.

Greenberg, S. (2006). *Understanding speech understanding: Towards a unified theory of speech perception.* Paper presented at the ESCA workshop on the "Auditory Basis of Speech Perception", Keele University, Staffordshire.

•Griffiths, J. D. (1967). Rhyming minimal contrasts: A simplified diagnostic articulation test. *Journal of the Acoustical Society of America, 42*(1), 236-241.

Hafter, E., Xia, J., & Kalluri, S. (2012). *A naturalistic approach to the cocktail party problem*. Paper presented at the International symposium on Hearing, St John's College, Cambridge.

•Hagerman, B. (1982). Sentences for testing speech-intelligibility in noise. *Scandinavian Audiology, 11*(2), 79-87.

•Hallgren, M., Larsby, B., & Arlinger, S. (2006). A Swedish version of the Hearing in Noise Test (HINT) for measurement of speech recognition. *International Journal of Audiology, 45*, 227-237.

Hawkins, J. E., & Stevens, S. S. (1950). The masking of pure tones and of speech by white noise. *Journal of the Acoustical Society of America, 22*(1), 6-13.

Hawley, M. L., Litovsky, R. Y., & Culling, J. F. (2004). The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer. *Journal of the Acoustical Society of America, 115*(2), 833-843.

•HearCom. (2009). *D-7-1b: Speech recognition tests for two additional languages: Polish and French.* Public deliverable, retrieved January 10, 2012 from, www.hearcom.eu.

•Helfer, K. S. (1997). Auditory and auditory-visual perception of clear and conversational speech. *Journal of Speech Language and Hearing Research, 40*(2), 432-443.

•Helfer, K. S., & Freyman, R. L. (2005). The role of visual speech cues in reducing energetic and informational masking. *Journal of the Acoustical Society of America, 117*(2), 842-849.

•Helfer, K. S., & Freyman, R. L. (2008). Aging and Speech-on-Speech Masking. *Ear & Hearing, 29*(1), 87-98.

•Helfer, K. S., & Freyman, R. L. (2009). Lexical and indexical cues in masking by competing speech. *Journal of the Acoustical Society of America, 125*(1), 447-456.

Hervais-Adelman, A., Davis, M. H., Johnsrude, I. S., & Carlyon, R. P. (2008). Perceptual learning of noise vocoded words: Effects of feedback and lexicality. *Journal of Experimental Psychology: Human Perception and Performance, 34*(2), 460 - 474.

Hirsh, I. J. (1950). The relation between localization and intelligibility. *Journal of the Acoustical Society of America, 22*(2), 196-200.

Hirsh, I. J. (1954). *The measurement of hearing.* New York: McGraw-Hill.

Hirsh, I. J., Davis, H., Silverman, S. R., Reynolds, E. G., Eldert, E. & Benson, R. W. (1950). Development of materials for speech audiometry. *Journal of Speech and Hearing Disorders, 17*, 321-337.

•Hirsh, I. J., Reynolds, E. G., & Joseph, M. (1954). Intelligibility of different speech materials. *Journal of the Acoustical Society of America, 26*(4), 530-538.

Hoen, M., Meunier, F., Grataloup, C., Pellegrino, F., Grimault, N., Perrin, F., Perrot, X., & Collet, L. (2007). Phonetic and lexical interferences in informational masking during speech-in-speech comprehension. *Speech Communication, 49*, 905-916.

Hopkinson, N. T. (1967). Combined effects of interruption and interaural alternation on speech intelligibility. *Language and Speech, 10*, 234-243.

•Horii, Y., House, A. S. & Hughes, G. W. (1971). A masking noise with speech-envelope characteristics. *Journal of the Acoustical Society of America, 49*(6), 1849-1856.

•House, A. S., Williams, C. E., Hecker, M. H. L., & Kryter, K. D. (1965). Articulation-testing methods: Consonantal differentiation with a closed-response set. *Journal of the Acoustical Society of America, 37*(1), 158-166.

•Howard-Jones, P. A., & Rosen, S. (1993). The perception of speech in fluctuating noise. *Acustica, 78*(5), 258-272.

Huckvale, M., Hilkhuysen, G., & Frasi, D. (2010). *Performance-based measurement of speech quality with an audio proof-reading task*. Paper presented at the 3rd ISCA workshop on perceptual quality systems. Dresden, Germany.

Ihlefeld, A., & Shinn-Cunningham, B. (2008a). Spatial release from energetic and informational masking in a divided speech identification task. *Journal of the Acoustical Society of America, 123*(6), 4380-4392.

•Ihlefeld, A., & Shinn-Cunningham, B. (2008b). Spatial release from energetic and informational masking in a selective speech identification task. *The Journal of the Acoustical Society of America, 123*(6), 4369-4379.

•Jerger, J., & Jordan, C. (1992). Age-related asymmetry on a cued-listening task. *Ear and Hearing, 13*(4), 272-277.

•Jerger, S., Jerger, J., & Lewis, S. (1981). Pediatric speech-intelligibility test II. Effect of receptive language age and chronological age. *International Journal of Pediatric Otorhinolaryngology, 3*(2), 101-118.

•Johnstone, P. M., & Litovsky, R. Y. (2006). Effect of masker type and age on speech intelligibility and spatial release from masking in children and adults. *Journal of the Acoustical Society of America, 120*(4), 2177-2189.

•Kalikow, D. N., Stevens, K. N., & Elliot, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America, 61*, 1337-1351.

•Kates, J. M., & Arehart, K. H. (2005). Coherence and the speech intelligibility index. *Journal of the Acoustical Society of America, 117*(4), 2224-2237.

•Keith, R. W., & Talis, H. P. (1970). Use of speech in noise in diagnostic audiometry. *Journal of Auditory Research, 10*(3), 201-204.

Kidd, G., Mason, C. R., & Arbogast, T. L. (2002). Similarity, uncertainty, and masking in the identification of nonspeech auditory patterns. *Journal of the Acoustical Society of America, 111*(3), 1367-1376.

•Kidd, G., Mason, C. R., & Gallun, F. J. (2005). Combining energetic and informational masking for speech identification. *Journal of the Acoustical Society of America, 118*(2), 982-992.

Kidd, G., Mason, C. R., Rohtla, T. L., & Deliwala, P. S. (1998). Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns. *Journal of the Acoustical Society of America, 104*(1), 422-431.

Kitterick, P. T., Bailey, P. J., & Summerfield, A. Q. (2010). Benefits of knowing who, where, and when in multi-talker listening. *Journal of the Acoustical Society of America, 127*(4), 2498-2508.

•Krull, V., Choi, S., Kirk, K. I., Prusick, L., & French, B. (2010). Lexical effects on spoken-word recognition in children with normal hearing. *Ear & Hearing, 31*(1), 102-114.

•Kryter, K. D. (1962). Methods for the calculation and use of the Articulation Index. *Journal of the Acoustical Society of America, 34*(11), 1689-1697.

•Kryter, K. D., & Whitman, E. C. (1965). Some comparisons between Rhyme and PB-word intelligibility tests. *Journal of the Acoustical Society of America, 37*(6), 1146-1146.

Larsby, B., Hallgren, M., Lyxell, B., & Arlinger, S. (2005). Cognitive performance and perceived effort in speech processing tasks: Effects of different noise backgrounds in normal-hearing and hearing-impaired subjects. *International Journal of Audiology, 44*(3), 131-143.

Leek, M. R., Hanna, T. E., & Marshall, L. (1992). Estimation of psychometric functions from adaptive tracking procedures. *Perception & Psychophysics, 51*(3), 247-256.

Leek, M. R., & Summers, V. (2001). Pitch strength and pitch dominance of iterated rippled noises in hearing-impaired listeners. *Journal of the Acoustical Society of America, 109*(6), 2944-2954.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America, 49*(2), 467-477.

•Lewis, H. D., Benignus, V. A., Muller, K. E., Malott, C. M., & Barton, C. N. (1988). Babble and random-noise masking of speech in high and low context cue conditions. *Journal of Speech and Hearing Research, 31*(1), 108-114.

Lezak, R. J., Siegenthaler, B. M., & Davis, A. J. (1964). Bekesy-type audiometry for speech reception threshold. *Journal of Auditory Research, 4*(3), 181-189.

•Li, A., Daneman, M., Qi, J. G., & Schneider, B. A. (2004). Does the information content of an irrelevant source differentially affect spoken word recognition in younger and older adults? *Journal of Experimental Psychology Human Perception and Performance, 30*(6), 1077-1091.

•Li, N., & Loizou, P. C. (2009). Factors affecting masking release in cochlear-implant vocoded speech. *Journal of the Acoustical Society of America* (126), 338.

Lutfi, R. A., Kistler, D. J., Callahan, M. R., & Wightman, F. L. (2003). Psychometric functions for informational masking. *Journal of the Acoustical Society of America, 114*(6), 3273-3282.

Lutman, M. E. (1991). Degradations in frequency and temporal resolution with age and their impact on speech identification. *Acta Oto-Laryngologica, 111*(s476), 120-126.

Macken, W. J., Tremblay, S., Houghton, R. J., Nicholls, A. P., & Jones, D. M. (2003). Does auditory streaming require attention? Evidence from attentional selectivity in short-term memory. *Journal of Experimental Psychology-Human Perception and Performance, 29*(1), 43-51.

Mackersie, C. L., Boothroyd, A., & Prida, T. (2000). Use of a simultaneous sentence perception test to enhance sensitivity to ease of listening. *Journal of Speech Language and Hearing Research, 43*(3), 675-682.

Macleod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology, 21*(2), 131-142.

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

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Marslen-Wilson, W. D. (1984). *Function and process in spoken word recognition.* Hillsdale, N.J: Erlbaum.

Marslen-Wilson, W. D. (1985). Speech shadowing and speech comprehension. *Speech Communication, 4*(1-3), 55-73.

•Martin, F. N., & Mussell, S. A. (1979). The influence of pauses in the competing signal on synthetic sentence identification scores. *Journal of Speech and Hearing Disorders, 44*(3), 282-292.

•McArdle, R. A., Wilson, R. H., & Burks, C. A. (2005). Speech recognition in multitalker babble using digits, words, and sentences. *Journal of the American Academy of Audiology, 16*(9), 726-739.

McMahon, M. L. (1983). Development of reading-while-listening skills in the primary grades. *Reading Research Quarterly, 19*(1), 38-52.

Meyer, J. P., & Seaman, M. A. (2008). *A comparison of the exact kruskal-wallis distribution to asymptotic approximations for n <= 105.* Paper presented at the annual meeting of the American Educational Research Association, New York.

Microsoft (2011). *Excel solver user's guide*. Retrieved September 12, 2012 from http://support.microsoft.com/kb/82890

Miller, G. A. (1947). The masking of speech. *Psychological Bulletin, 44*(2), 105-129.

•Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology, 41*(5), 329-335.

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

Moore, B. C. J. (1995). *Perceptual consequences of cochlear damage.* Oxford: Oxford University Press.

Moore, B. C. J. (2012). *An introduction to the psychology of hearing* (Sixth edition). London: Academic Press.

Moore, B. C. J., & Glasberg, B. R. (1983). Suggested formulas for calculating auditory-filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America, 74*(3), 750-753.

Moray, N. (1959). Attention in dichotic-listening - affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology, 11*(1), 56-60.

Moray, N. (1969). *Listening and attention*. Middlesex, England: Penguin science of behaviour.

Nakagawa, S. (2004). A farewell to Bonferroni: The problems of low statistical power and publication bias. *Behavioral Ecology, 15*(6), 1044-1045.

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*(1), 125-134.

Neville, M. H. (1975). Effectiveness of rate of aural message on reading and listening. *Educational Research, 18*(1), 37-43.

•Ng, S. L., Meston, C. N., Scollie, S. D., & Seewald, R. C. (2011). Adaptation of the BKB-SIN test for use as a pediatric aided outcome measure. *Journal of the American Academy of Audiology, 22*(6), 375-386.

Nicholls, A. P., & Jones, D. M. (2002). Capturing the suffix: Cognitive streaming in immediate serial recall. *Journal of Experimental Psychology-Learning Memory and Cognition, 28*(1), 12-28.

•Niederjohn, R. J., & Grotelueschen, J. H. (1976). Enhancement of speech-intelligibility in high noise-levels by high-pass filtering followed by rapid amplitude compression. *IEEE Transactions on Acoustics Speech and Signal Processing, 24*(4), 277-282.

•Nielsen, J. B., & Dau, T. (2009). Development of a Danish speech intelligibility test. *International Journal of Audiology, 48*(10), 729-741.

Nilsson, M., Soli, S. D. & Sullivan, J. A. (1994). Development of the Hearing In Noise Test for measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America, 95*(2), 1085-1099.

Oh, E. L., & Lutfi, R. A. (2000). Effect of masker harmonicity on informational masking. *Journal of the Acoustical Society of America, 108*(2), 706-709.

Owens, E. & Schubert, E. D. (1977). Development of the California Consonant Test. *Journal of Speech and Hearing Research, 20*, 463-474.

•Oxenham, A. J., & Simonson, A. M. (2009). Masking release for low- and high-pass-filtered speech in the presence of noise and single-talker interference. *Journal of the Acoustical Society of America, 125*(1), 457-468.

•Ozimek, E., Kutzner, D., Sek, A., & Wicher, A. (2009). Polish sentence tests for measuring the intelligibility of speech in interfering noise. *International Journal of Audiology, 48*, 433-443.

•Ozimek, E., Warzybok, A., & Kutzner, D. (2010). Polish sentence matrix test for speech intelligibility measurement in noise. *International Journal of Audiology, 49*, 444-454.

Pashler, H. E. (1998). *The psychology of attention.* Cambridge: MIT Press.

•Pederson, O. T., & Studebaker, G. A. (1972). A new minimal-contrasts closed-response-set speech test. *Journal of Auditory Research, 12*, 187-195.

•Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America, 97*(1), 593-608.

•Pichora-Fuller, M. K., Schneider, B. A., MacDonald, E., Pass, H. E., & Brown, S. (2007). Temporal jitter disrupts speech intelligibility: A simulation of auditory aging. *Hearing Research, 223*(1-2), 114-121.

Plomp, R. (1977). Acoustical aspects of cocktail parties. *Acustica, 38*(3), 186-191.

Plomp, R. (1986). A signal-to-noise ratio model for speech-reception threshold of the hearing impaired. *Journal of Speech & Hearing Research, 29*, 146-154.

•Plomp, R., & Mimpen, A. M. (1979). Speech-reception threshold for sentences as a function of age and noise-level. *Journal of the Acoustical Society of America, 66*(5), 1333-1342.

Plomp, R., & Mimpen, A. M. (1981). Effect of the orientation of the speakers head and the azimuth of a noise source on the speech-reception threshold for sentences. *Acustica, 48*(5), 325-328.

•Rakerd, B., Aaronson, N. L., & Hartmann, W. M. (2006). Release from speech-on-speech masking by adding a delayed masker at a different location. *Journal of the Acoustical Society of America, 119*(3), 1597-1605.

•Rao, M. D., & Letowski, T. (2006). Callsign Acquisition Test (CAT): Speech intelligibility in noise. *Ear & Hearing, 27*(2), 120-128.

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*(4), 2181-2192.

Rhebergen, K. S., Versfeld, N. J., & Dreschler, W. A. (2005). Release from informational masking by time reversal of native and non-native interfering speech. *Journal of the Acoustical Society of America, 118*(3), 1274-1277.

Ricketts, T. A., & Dittberner, A. B. (2002). Directional amplification for improved signal-to-noise ratio: Strategies, measurements, and limitations. In Valente M (Eds.): *Hearing Aids: Standards, options and limitations,* 2nd edition. New York: Thieme, 274-346.

•Rogers, C. L., Lister, J. J., Dashielle, M. F., & Besing, J. M. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Applied Psycholinguistics, 27*(3), 465-485.

Rosen, S. (1992). Temporal information in speech - acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London. Series B-Biological Sciences, 336*(1278), 367-373.

Rossi-Katz, J., & Arehart, K. H. (2009). Message and talker identification in older adults: Effects of task, distinctiveness of the talkers' voices, and meaningfulness of the competing message. *Journal of Speech, Language, and Hearing Research, 52*, 435-453.

Rothauser, E. H., Chapman, W. D., Guttman, N., Nordby, K. S., Silbiger, H. R., Urbanek, G. E., and Weinstock, M. (1969). IEEE recommended practice for speech quality measurements. *IEEE Trans Audio Electroacoustics, 17*, 227-246.

Saberi, K., & Green, D. M. (1996). Adaptive psychophysical procedures and imbalance in the psychometric function. *Journal of the Acoustical Society of America, 100*(1), 528-536.

Sarampalis, A., Kalluri, S., Edwards, B., & Hafter, E. (2009). Objective measures of listening effort: Effects of background noise and noise reduction. *Journal of Speech Language and Hearing Research, 52*(5), 1230-1240.

Schneeberg, H. (1977). Listening while reading – A four year study. *Reading Teacher, 30*(6), 629-635.

Schneider, B. A. (1997). Psychoacoustics and aging: Implications for everyday listening. *Journal of Speech-Language Pathology* and *Audiology, 21*, 111-124.

Schneider, B. A., Daneman, M., & Murphy, D. R. (2005). Speech comprehension difficulties in older adults: Cognitive slowing or age-related changes in hearing? *Psychology and Aging, 20*(2), 261-271.

Schneider, B. A., Pichora-Fuller, M. K., & Daneman, M. (2010). The effects of senescent changes in audition and cognition on spoken language comprehension. In S. Gordon-Salant, R. D. Frisina, R. R. Fay & A. N. Popper (Eds.), *The aging auditory system*. New York: Springer.

•Schultz, M. C., & Schubert, E. D. (1969). A multiple choice discrimination test (MCDT). *Laryngoscope, 79*(3), 382-399.

•Scott, S. K., Rosen, S., Wickham, L., & Wise, R. J. S. (2004). A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. *Journal of the Acoustical Society of America, 115*(2), 813-821.

•Sergeant, L., Atkinson, J. E., & Lacroix, P. G. (1979). The NSMRL tri-word test of intelligibility (TTI). *Journal of the Acoustical Society of America, 65*(1), 218-222.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primary temporal cues. *Science, 270*, 303-304.

•Sherbecoe, R. L., & Studebaker, G. A. (2002). Audibility-index functions for the connected speech test. *Ear & Hearing, 23*(5), 385-398.

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences, 12*(5), 182-186.

Shinn-Cunningham, B. G., & Best, V. (2008). Selective attention in normal and impaired hearing. *Trends in amplification, 12*(4), 283-299.

Shinn-Cunningham, B. G., & Wang, D. (2008). Influences of auditory object formation on phonemic restoration. *Journal of the Acoustical Society of America, 123*(1), 295-301.

Siegal, S., & Castellan, N. J. (1988). *Nonparametric statistics for behavioral sciences*. New York: McGraw-Hill International Editions.

Silverman, S. R. & Hirsh, I. J. (1955). Problems related to the use of speech on clinical audiometry. *Annals of Otology, Rhinology, laryngology, 64*(4), 1234-1244.

Speaks, C., & Jerger, J. (1965). Method for measurement of speech identification. *Journal of Speech and Hearing Research, 8*(2), 185-194.

•Speaks, C., & Karmen, J. L. (1967). The effect of noise on synthetic sentence identification. *Journal of Speech and Hearing Research, 10*, 859-864.

•Speaks, C., Karmen, J. L., & Benitez, L. (1967). Effect of a competing message on synthetic sentence identification. *Journal of Speech and Hearing Research, 10*(2), 390-396.

•Speaks, C., Parker, B., Kuhl, P., & Harris, C. (1972). Intelligibility of connected discourse. *Journal of Speech and Hearing Research, 15*(3), 590-602.

•Stickney, G. S., Zeng, F. G., Litovsky, R., & Assmann, P. (2004). Cochlear implant speech recognition with speech maskers. *Journal of the Acoustical Society of America, 116*(2), 1081-1091.

Strauss, D. J., Corona-Strauss, F. I., & Froehlich, M. (2008). Objective estimation of the listening effort: Towards a neuropsychological and neurophysical model. *Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society Conference, 2008*, 1777-1780.

•Studebaker, G. A., Taylor, R., & Sherbecoe, R. L. (1994). The effect of noise spectrum on speech recognition performance-intensity functions. *Journal of Speech and Hearing Research, 37*(2), 439-448.

•Surprenant, A. M. (2007). Effects of noise on identification and serial recall of nonsense syllables in older and younger adults. *Aging Neuropsychology and Cognition, 14*(2), 126-143.

•Surr, R. K., & Schwartz, D. M. (1980). Effects of multi-talker competing speech on the variability of the California Consonant Test. *Ear and Hearing, 1*(6), 319-323.

•Suter, A. H. (1985). Speech recognition in noise by individuals with mild hearing impairments. *Journal of the Acoustical Society of America, 78*(3), 887-900.

•Tabri, D., Abou Chacra, K. M. S., & Pring, T. (2011). Speech perception in noise by monolingual, bilingual and trilingual listeners. *International Journal of Language & Communication Disorders, 46*(4), 411-422.

•Takahashi, G. A., & Bacon, S. P. (1992). Modulation detection, modulation masking, and speech understanding in noise in the elderly. *Journal of Speech and Hearing Research, 35*(6), 1410-1421.

•Theodoridis, G. C., & Schoeny, Z. G. (1988). Comparison of various modes of presenting sentence materials in tests of speech-perception in noise. *Journal of the Acoustical Society of America, 84*(6), 2270-2272.

•Theodoridis, G. C., Schoeny, Z. G., & Anné, A. (1985). Measuring the contribution of printed context information to acoustical word recognition by normal subjects. *Audiology, 24*(2), 104-116.

Tillman, T. W. & Carhart, R. (1966). An expanded test for speech discrimination utilizing CNC monosyllabic words. Northwestern University Auditory test No. 6. SAM-TR-66-55 (Technical report). Brooks Air Force Base (TX): USAF School of Aerospace Medicine; 1966.

•Trammell, J. L., & Speaks, C. (1970). Distracting properties of competing speech. *Journal of Speech and Hearing Research, 13*(2), 442-445.

Treisman, A. M. (1964). The effect of irrelevant material on the efficiency of selective listening. *American Journal of Psychology, 77*(4), 533-546.

Treisman, A. M., & Gelade, G. (1980). Feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97-136.

•Tun, P. A. (1998). Fast noisy speech: Age differences in processing rapid speech with background noise. *Psychology and Aging, 13*(3), 424-434.

Tun, P. A., & Wingfield, A. (1999). One voice too many: Adult age differences in language processing with different types of distracting sounds. *Journals of Gerontology Series B-Psychological Sciences and Social Sciences, 54*(5), 317-327.

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

Van Engen, K. J., & Bradlow, A. R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *Journal of the Acoustical Society of America, 121*(1), 519-526.

•Van Wieringen, A., & Wouters, J. (2008). LIST and LINT: Sentences and numbers for quantifying speech understanding in severely impaired listeners for Flanders and the Netherlands. *International Journal of Audiology, 47*, 348-355.

•Vestergaard, M. D., Fyson, N. R. C., & Patterson, R. D. (2009). The interaction of vocal characteristics and audibility in the recognition of concurrent syllables. *Journal of the Acoustical Society of America, 125*(2), 1114-1124.

Wagener, K. C., & Brand, T. (2005). Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: Influence of measurement procedure and masking parameters. *International Journal of Audiology, 44*(3), 144-156.

•Wagener, K. C., Josvassen, J. L., & Ardenkjoer, R. (2003). Design, optimization and evaluation of a Danish sentence test in noise. *International Journal of Audiology, 42*, 10-17.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science, 167*(3917), 392-393.

Watson, A. B. (1979). Probability summation over time. *Vision Research, 19*(5), 515-522.

Watson, C. S. (2005). Some comments on informational masking. *Acta Acustica, 91*(3), 502-512.

•Whitmal, N. A., Poissant, S. F., Freyman, R. L., & Helfer, K. S. (2007). Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience. *Journal of the Acoustical Society of America, 122*(4), 2376-2388.

Wichmann, F. A., & Hill, N., J. (2001a). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics, 63*(8), 1314-1329.

Wichmann, F. A., & Hill, N. J. (2001b). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics, 63*(8), 1293-1313.

•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*(5), 3164-3176.

•Wilson, R. A., & Antablin, J. K. (1980). A picture identification task as an estimate of the word-recognition performance of nonverbal adults. *Journal of Speech and Hearing Disorders, 45*, 223-238.

•Wilson, R. H., & Burks, C. A. (2005). Use of 35 words for evaluation of hearing loss in signal-to-babble ratio: A clinic protocol. *Journal of Rehabilitation Research and Development, 42*(6), 839-851.

•Wilson, R. H., Burks, C. A., & Weakley, D. G. (2006). Word recognition of digit triplets and monosyllabic words in multitalker babble by listeners with sensorineural hearing loss. *Journal of the American Academy of Audiology, 17*(6), 385-397.

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

•Wilson, R. H., Carnell, C. S., & Cleghorn, A. L. (2007). The Words-in-Noise (WIN) test with multitalker babble and speech-spectrum noise maskers. *Journal of the American Academy of Audiology, 18*(6), 522-529.

•Wilson, R. H., & Cates, W. B. (2008). A comparison of two word-recognition tasks in multitalker babble: Speech Recognition in Noise Test (SPRINT) and Words-in-Noise test (WIN). *Journal of the American Academy of Audiology, 19*(7), 548-556.

•Wilson, R. H., Farmer, N. M., Gandhi, A., Shelburne, E., & Weaver, J. (2010). Normative data for the Words-in-Noise Test for 6-to 12-year-old children. *Journal of Speech Language and Hearing Research, 53*(5), 1111-1121.

•Wilson, R. H., & McArdle, R. (2007). Intra-and inter-session test, retest reliability of Words-in-Noise (WIN) Test. *Journal of the American Academy of Audiology, 18*, 813-825.

•Wilson, R. H., McArdle, R., Betancourt, M. B., Herring, K., Lipton, T., & Chisolm, T. H. (2010). Word-recognition performance in interrupted noise by young listeners with normal hearing and older listeners with hearing loss. *Journal of the American Academy of Audiology, 21*(2), 90-109.

•Wilson, R. H., McArdle, R., & Roberts, H. (2008). A comparison of recognition performances in speech-spectrum noise by listeners with normal hearing on PB-50, CID W-22, NU-6, W-1 spondaic words, and monosyllabic digits spoken by the same speaker. *Journal of the American Academy of Audiology, 19*(6), 496-506.

•Wilson, R. H., McArdle, R. A., & Smith, S. L. (2007). An evaluation of the BKB-SIN, HINT, QuickSIN, and WIN materials on listeners with normal hearing and listeners with hearing loss. *Journal of Speech Language and Hearing Research, 50*(4), 844-856.

•Wilson, R. A., & Oyler, A. (1997). Psychometric functions for the CID W-22 and NU Auditory Test No. 6. Materials spoken by the same speaker. *Ear & Hearing, 18*(5), 430-433.

•Wilson, R. H., & Strouse, A. (2002). Northwestern University Auditory Test No. 6 in multi-talker babble: A preliminary report. *Journal of Rehabilitation Research and Development, 39*(1), 105-113.

Wilson, R. H., Zizz, C. A., Shanks, J. E., & Causey, G. D. (1990). Normative data in quiet, broad-band noise, and competing message for Northwestern University Auditory Test No. 6 by a female speaker. *Journal of Speech and Hearing Disorders, 55*(4), 771-778.

Wingfield, A., McCoy, S. L., Peelle, J. E., Tun, P. A., & Cox, L. C. (2006). Effects of adult aging and hearing loss on comprehension of rapid speech varying in syntactic complexity. *Journal of the American Academy of Audiology, 17*(7), 487-497.

Wingfield, A., Poon, L. W., Lombardi, L., & Lowe, D. (1985). Speed of processing in normal aging - effects of speech rate, linguistic structure, and processing time. *Journals of Gerontology, 40*(5), 579-585.

•Wu, X. H., Wang, C., Chen, J., Qu, H. W., Li, W. R., Wu, Y. H., Schneider, B. A. & Li, L. (2005). The effect of perceived spatial separation on informational masking of Chinese speech. *Hearing Research, 199*(1-2), 1-10.

•Yang, Z., Chen, J., Huang, Q., Wu, X., Wu, Y., Schneider, B. A., & Li, L. (2007). The effect of voice cuing on releasing Chinese speech from informational masking. *Speech Communication, 49*(12), 892-904.

Yost, B. (2006). *Informational masking: What is it?* Retrieved February 28, 2010 from www.isr.umd.edu/Labs/NSL/Cosyne/Yost.htm.

•Young, L. L., Goodman, J. T., & Carhart, R. (1979). Effects of whitening and peak-clipping on speech-intelligibility in the presence of a competing message. *Audiology, 18*(1), 72-79.