CONTRASTING EFFECTS OF IRRELEVANT SPEECH AND NON-SPEECH SOUNDS ON SHORT-TERM MEMORY

MARIE ANN CAHILLANE

A thesis submitted in partial fulfilment of the requirements of the University of the West of England, Bristol for the degree of Doctor of Philosophy at Bath Spa University

School of Social Sciences, Bath Spa University

June 2008

Contents

APPENDICES

Abstract

The characteristics of speech that determine its greater disruption of serial recall relative to non-speech (the irrelevant sound effect) are investigated (c.f. Tremblay et al., 2000). Degraded non-words disrupted serial recall less than clear non-words. Tasks show that both vowels and consonants of degraded non-words were misperceived, with initial consonants misperceived to a greater degree. Measures that followed showed that clear sequences of non-words, with changing vowels were more disruptive than sequences with changing consonants. Degrading vowel only changing sequences reduced disruption of serial recall to a level observed with clear consonant only changing sequences, whereas degradation had no effect on disruption by consonant only changing sequences. In further experiments the acoustic complexity of speech was reduced while maintaining its intelligibility by removing fundamental frequency information. Whispered speech disrupted serial recall to the same degree as voiced speech. Alternating voiced and whispered speech sounds within a sequence did not reduce serial recall performance relative to a sequence of voiced-only speech sounds. Results indicate the formant structure of speech sounds and not fundamental frequency information is the important carrier of acoustic change. Reversing the fine structure of whispered speech damaged its intelligibility whilst preserving acoustic complexity and these sounds were as disruptive of serial recall as normal whispered speech. This indicates that vocal tract resonances (formants) of speech and not its intelligibility determine its disruptive power. The relative disruptiveness of speech and non-speech sounds was then examined. Sounds were matched for acoustic complexity, but their 'speech-likeness' was destroyed. Speech disrupted serial recall more than did non-speech. Results indicate that the biological nature of speech renders it more disruptive than non-speech. The findings refute the 'changing-state-hypothesis' which is derived from the object-oriented episodic record model. This hypothesis argues that it is the degree of acoustic variation within an irrelevant stream and not the nature of its component sounds which determines its disruption of serial memory. Biological sounds may disrupt serial memory to a greater degree since they are of behavioural relevance and provide information about the environment that may need to be attended to. The addition of an attentional mechanism to the object-oriented episodic record model that regulates the re-allocation of cognitive processing resources is proposed.

Acknowledgements

Thanks to Nigel Holt, my supervisor, for believing in me and making me realise my potential, whose special support and guidance made doing my PhD one of the most enjoyable experiences in my life so far. Also, for reading my thesis, providing many invaluable comments and being a profound source of inspiration and calm. Sincere thanks to Lance Workman, my director of studies, for mentoring me through my PhD and always being at hand. Also, for the many insightful discussions that have been so useful in my training as an academic. Thanks also to Sandie Taylor, for coaching me during my studies. Thanks to Clive Frankish, with whom it has been a privilege to work. For his enduring patience and support throughout my PhD, which went far beyond his role as external advisor. For the many enjoyable discussions and his guidance regarding the art of experimenting and whose astounding knowledge engaged my thinking and was a source of encouragement. Thanks to the many students who took part in my experiments, without whose voluntary participation this level of research would not have been possible. To the IT staff, for all their much appreciated assistance. Thanks extend to all in the psychology department, for introducing humour throughout that would brighten the most work loaded of days.

Thanks to Rachel, for making me see there is also life outside of experimental psychology and for all her support. To my parents and family, whose love, material and personal support was much valued and appreciated as always. To Beryl, whose support and wish for me to succeed I will always remember and cherish. Lastly, thanks to Ivor, whose love and desire for me to fore fill my dreams remained in my heart and helped me to perceive what seemed a closed door in my life to be one which would and did re-open into my receiving of the opportunity to do my PhD and work with some of the most amazing people I've ever known.

Glossary of Abbreviations

Abbreviation	Description
ANOVA	Analysis Of Variance
C-O-C	Consonant-Only-Changing
CSH	Changing-State-Hypothesis
CV	Consonant-Vowel
CVC	Consonant-Vowel-Consonant
fo	Fundamental Frequency
$f1$	First Formant
$f2$	Second Formant
$f3$	Third Formant
FM	Frequency Modulation
IP	Item Probe
ISE	Irrelevant Sound Effect
ISI	Inter-Stimulus Interval
ITI	Inter-Trial Interval
LED	Left Ear Disadvantage
LP	List Probe
LTM	Long-Term Memory
O-OER	Object-oriented-Record Model
OR	Oriented Response
OSPAN	Operation Span
PET	Positron Emission Tomography
PSH	Phonological Store Hypothesis
REA	Right Ear Advantage

RMS	Root Mean Square
RT	Reaction Time
SCN	Signal-Correlated Noise
SD	Standard Deviation
SNR	Signal-To-Noise-Ratio
STG	Superior Temporal Gyrus
STM	Short-Term Memory
TBR	To-Be-Remembered
TDT	Temporal Distinctiveness Theory
VDU	Visual Display Unit
V-O-C	Vowel-Only-Changing
VOT	Voice Onset Time
WMM	Working Memory Model

CHAPTER 1

1 LITERATURE REVIEW: EMPIRICAL FINDINGS

1.1 INTRODUCTION: THE IRRELEVANT SOUND EFFECT

The attenuation in performance on working memory tasks when performed in the presence of extraneous auditory information is known as the "irrelevant sound effect" (ISE) (Colle and Welsh, 1976; Jones and Macken, 1993, Salamé and Baddeley, 1982). The ISE was initially observed in experiments where the interfering sounds were sequences of spoken words or syllables (e.g. Colle & Welsh, 1976; Salamé & Baddeley, 1982). In a typical ISE experiment participants are presented visually with a series of between seven to nine digits or letters on a computer screen in sequence with an inter-stimulus interval (ISI), while irrelevant sounds are presented concurrently over headphones. Participants are instructed to ignore any sounds they may hear and are also notified that no test of the content of the auditory information will be made. At the recall stage, memory for the order of digits or letters is disrupted by the presence of the irrelevant auditory information, even though they are instructed to ignore the irrelevant sounds (e.g. Colle and Welsh, 1976; Jones and Macken, 1993; Salamé and Baddeley, 1982; Campbell and Dodd, 1984; Jones, 1994). In addition, irrelevant sound has been found to disrupt serial recall of auditory items (Campbell, Beaman and Berry, 2002) and lip-read items (Campbell and Dodd, 1984; Jones, 1994) in the same way as with graphically presented items.

The ISE has been widely and frequently replicated (e.g., Ellermeier and Zimmer, 1997; Tremblay and Jones, 1998) and has been interpreted in terms of attention, perception and their interaction with memory (Larsen and Baddeley, 2003). Empirical evidence shows that there is no habituation as within the serial recall task the effect remains stable over numerous trials and when trials take place days apart (Hellbrück, Kuwano and Namba, 1995; Jones, Macken and Mosdell, 1997; Tremblay and Jones, 1998). This chapter provides a review of the theoretical frameworks put forward to account for the effect of irrelevant sound on serial memory and the empirical findings relating to the irrelevant sound effect. The chapter aims to put across a body of knowledge regarding the characteristics of sounds and their organisation within the irrelevant stream that are known to influence the size of the ISE, and those which are shown to be unimportant. Also the types of task affected by irrelevant background sound and the practical implications of the ISE are surveyed.

1.2 EXPLANATIONS OF THE ISE

1.2.1 THE WORKING MEMORY MODEL AND THE PHONOLOGICAL STORE HYPOTHESIS

Baddeley and Hitch (1974) proposed a model of working memory. The working memory model (WMM) comprises an executive controlling attentional mechanism, known as the central executive, which directs and coordinates two subcomponent slave systems, the visuospatial sketch pad and the phonological loop. The visuospatial sketch pad is in charge of encoding, processing and manipulating visual information. Phonological information (speech stimuli) on the other hand is manipulated by the phonological or articulatory loop (see figure 1). The phonological loop is made of two components; a phonological store, which maintains speech-like stimuli and the articulatory control process, which manipulates inner speech. Representations in memory held within the phonological store are subject to decay after approximately one-and-a-half to two seconds and then become irretrievable. The articulatory control process acts to refresh items in the phonological store, to prevent them from decaying, by sub-vocally rehearsing the items using inner speech, therefore allowing items to re-enter the phonological store (Baddeley, 1990). Visual stimuli also gains access to the phonological store via the articulatory control process which converts visual items into phonological codes.

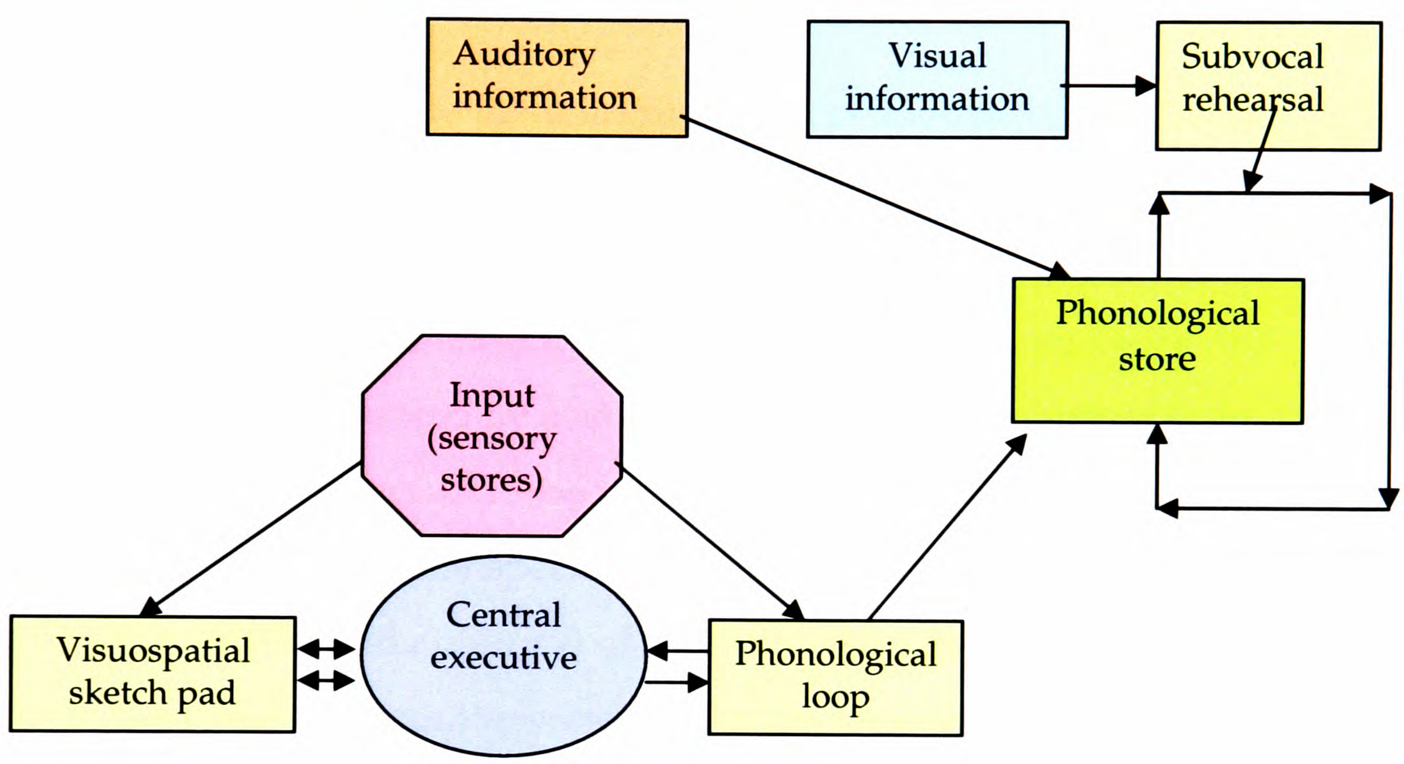


Figure 1. Simplified representation of Baddeley and Hitch's (1974) working memory model.

One explanation of the ISE argues that interference results from the similarity of items represented in memory. An early account of the ISE, known as the 'phonological store hypothesis' (PSH) distinguishes between an articulatory control process and a phonological store, which form the phonological loop (see figure 1) (Salamé and Baddeley, 1982). It is argued that irrelevant sounds and the to-be-recalled (TBR) items are both kept in the phonological store. The PSH proposes that the effect is limited to speech and that irrelevant speech gains automatic access to the phonological store, where it creates phonological codes that interfere with representations generated by rehearsal of the visual TBR items

(Salamé and Baddeley, 1982). In contrast to the irrelevant speech sounds, the visual TBR stimuli enter the phonological store through sub-vocal rehearsal, using the articulatory control process and are thus converted into a phonological code. The PSH assumes that all stimuli are represented as phonemes and that the confusion between auditory and visual phonemes results in the ISE (Salamé and Baddeley, 1982).

1.2.2 THE OBJECT-ORIENTED EPISODIC RECORD MODEL

An alternative view to the working memory model (WMM) is the object-oriented episodic record (O-OER). According to the O-OER model (Jones, 1993; Jones, Beaman and Macken, 1996) all auditory information gains access to short-term memory. It is assumed that both auditory and visual stimuli are represented by abstract representations, referred to in the O-OER model as *'objects'*. These representations are object-orientated because they are not modality specific. Rather they code all characteristics (visual and auditory) of incoming information. Objects from both the visual and auditory modalities are therefore represented in the same way in a single storage system. This feature of the O-OER model differentiates it from other models of Short-Term Memory (STM) such as the WMM (Baddeley and Hitch, 1974), which argue for the existence of two separate storage systems for auditory and visual items (Jones et al., 1996). Cues, referred to as episodic pointers, which are associated with each object in memory provide a code for their serial order. Streaming occurs in the different modalities, where items are assigned to either the same or a different source. The creation of episodic pointers is determined by the number of acoustic changes in state from item-to-item and once created, their strength decays over time. Episodic pointers are generated pre-attentively for auditory information. In contrast, episodic pointers are generated and preserved by a rehearsal process for visual TBR information (see figure 2). Serial recall constitutes moving from object to object using these episodic pointers (Jones et al., 1996).

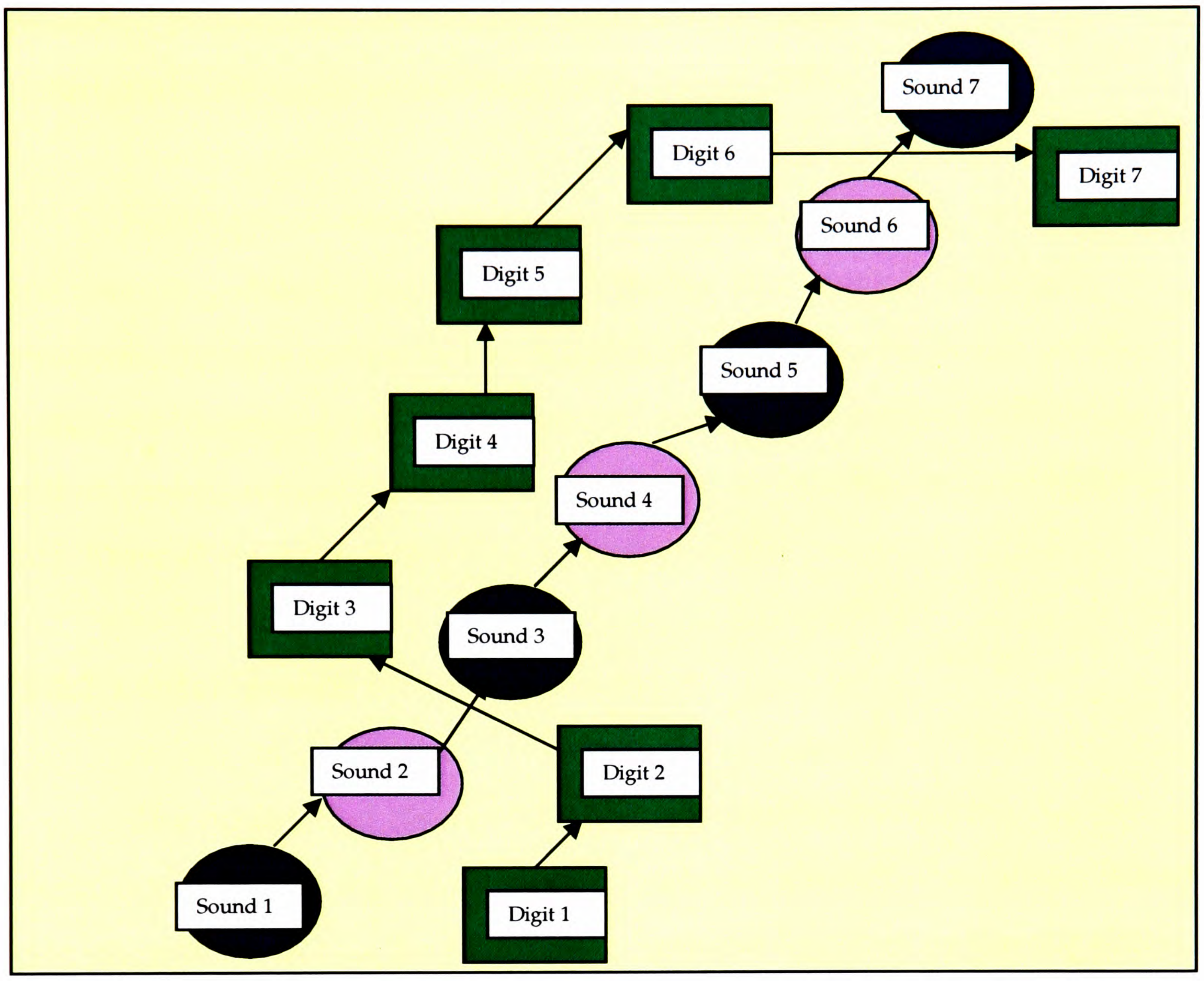


Figure 2. Schematic diagram of the interference of serial recall by task irrelevant sound upon the episodic record of the O-OER model.

In the ISE paradigm, serial recall errors occur when serial recall cues (episodic pointers) from an irrelevant sequence of distinct items interfere with a different collection of cues. Therefore, if the same auditory item is repeated, it forms only one object with a single episodic pointer for reference. In contrast, if a changing-state auditory sequence is presented, multiple objects are created, one representing each auditory item along with episodic pointers linking adjacent items. Serial order cues, automatically generated by acoustic change serve to maintain the serial (temporal) integrity of the representation of sound sequences in auditory STM (Jones, Madden and Miles, 1992; Jones, et al., 1996). Thus it has been suggested that changing-state stimuli disrupt the serial rehearsal mechanism not because they generate competing phonological codes, but because sequences of discrete sounds automatically create competing cues to serial order. These rival ordering cues disrupt the rehearsal of associations between objects in memory, causing deterioration of serial recall (Beaman & Jones, 1997).

Arising from the more general O-OER model (Jones, et al., 1996) is the changing-state-hypothesis (CSH) (Jones, et al., 1996). The CSH proposes that the critical factor in the irrelevant sound effect is not the nature of the sound, but the nature and extent of acoustic variation in either timbre, tempo or frequency within an unfolding auditory stream (c.f. Jones et al., 1992; Tremblay and Jones, 1999).

1.2.2.1 Interference by process account

The notion that disruption is determined by competing cues to serial order lead to the idea that interference is a result of a conflict based not on content, but on the similarity of process between irrelevant and relevant streams (Jones, 1999; Jones and Tremblay, 2000). This assumption was developed as a refinement of the CSH. Critically, it is thought that the breakdown of selectivity occurs because of interference between two synchronised processes of order maintenance (seriation). Remembering the order of items in the serial recall task provides one source of information. In addition, the second source of order maintenance (seriation) is produced when there is variation in the irrelevant auditory stream, as the processes related to the perceptual organisation of sound produce information about the order of the sounds. Two key predictions have been generated from these assumptions. The first concerns the level of seriation involved in the recall task, in that if the use of serial order is diminished; the level of disruption by irrelevant sound will also decrease (Jones, 1999). The second prediction of the 'interference-by-process' proposition is that if the level of seriation in the irrelevant stream is reduced, the level of disruption will be reduced.

1.2.2.2 PHONOLOGICAL DISSIMILARITY

The PSH predicts that as the phonological similarity between the irrelevant stimuli and the to-be-recalled items increases, the degree of interference should also increase (Salamé and Baddeley, 1982). However, empirical data demonstrates that the degree of phonological similarity between the relevant and irrelevant material is not a strong dictator of the degree of disruption (e.g. Bridges and Jones, 1996; Jones and Macken, 1995a; Larsen, Baddeley and Andrade, 2000; LeCompte and Shaibe, 1997). Only Salamé and Baddeley (1982) found evidence to support the hypothesis that between-stream similarity serves to increase the ISE. Rather, what determines disruption is not the similarity of the relevant and irrelevant streams of information, but the degree of phonological dissimilarity of auditory items within the irrelevant sound stream (Jones and Macken, 1995a). For instance, a stream consisting of non-rhyming words, where each successive item is distinct produces more disruption than a stream of rhyming words. For example, *hat, cow, nest* is more disruptive than *sea, flea, key* (Jones and Macken, 1995a). In addition, Larsen, Baddeley and Andrade (2000) replicated this phonological dissimilarity effect, but not in all conditions tested. The absence of a consistent phonological dissimilarity effect can be accounted for by the 'token-dose' – the number of auditory events presented per time unit in each trial. Memory interference increases as the irrelevant within-stream token-dose increases (Bridges and Jones, 1996). Jones and Macken (1995a) used 34 tokens, whereas Larsen et al (2000) synchronised only one speech token with each of the 6 to-be-remembered (TBR) items and the token set may not have been large enough to be sensitive to a phonological dissimilarity effect.

These findings refute the phonological interference account of the PSH but are consistent with the claim of the CSH, that a stream of sounds demonstrating a considerable degree of acoustical change will cause more interference (Jones et al., 1992). Furthermore, the phonological dissimilarity effect is considered within the CSH as a phonological example of a less specific 'acoustic' changing-state effect. Specifically, when speech tokens form an irrelevant auditory stream, disruption is not a function of the similarity between tokens at an abstract phonological, speech-based level of analysis but instead dissimilarity at an acoustic level.

1.2.2.3 THE EFFECT OF PRESENTING NON-SPEECH SOUNDS

Recall is also disrupted by many types of irrelevant non-speech sound, such as simple tones (Jones and Macken, 1993) or music (e.g. Klatte, Kilcher and Hellbrück, 1995; Salamé and Baddeley, 1989), presented either during presentation of to-be-recalled information or during the post-presentation retention interval (e.g. Jones & Macken, 1993; Beaman & Jones, 1998). Moreover, research has shown that irrelevant sounds which change physically produce memory interference, regardless of whether they are speech or pure tones. In contrast, stimuli, which change less, for example sequences of identical tones or speech do not (Jones and Macken, 1993; Neath, Surprenant, and LeCompte, 1998). The former changing sounds are described as 'changing-state' whereas less changeable sounds are described as 'steady-state'. Whereas there is research that does not demonstrate a 'steady-state effect' (e.g., LeCompte, 1996) some research has shown a steady-state sequence to produce more disruption than a quiet control, an effect which has been shown to be statistically reliable (e.g., LeCompte, 1995 and Hughes et al, 2005). Similarly, music varying greatly in pitch or tempo interferes with cognitive performance more than music comprised of numerous legato variations (Klatte et al., 1995). It is argued then, that the primary determinant of disruption is the degree of change within the irrelevant sound stream.

1.2.3 INTEGRATED MODEL OF ATTENTION AND MEMORY: DISRUPTION BY DIMINISHED 'ATTENTIONAL RESOURCES'

Cowan's (1995) integrated model of short-term memory and attention, which acts as a general approach to information processing, has been used to attempt to account for the ISE. Cowan (1995) assumes that immediate memory involves the activated area of what is deemed as a more long-term store and that only a portion of the activated memory is attended to at a given point in time. In terms of the irrelevant sound paradigm, visual TBR items are activated through rehearsal, and therefore are in the focus of attention. Incoming auditory stimuli, such as irrelevant sound can disrupt the attentional focus by automatically attracting attention and thus shifting processing resources away from the attended-to visual TBR items. The model can then, account for the ISE in terms of rehearsal disruption by auditory distraction. Cowan's (1995) model predicts that changing-state sequences produce an ISE, whereas steady-state (repetitive) sequences in general do not (Jones and Macken, 1995a; but see Hughes et al., (2005). Cowan (1995) proposed that a changing-state auditory sequence produces more interference than a steady-state sequence because the attentional mechanism will quickly habituate to a repetitive sequence but not to a sequence that changes. This is because sequences consisting of an identical repeated item would not elicit an involuntary attentional, Orientated Response (OR), an account which is in line with the critical assumption of the CSH (LeCompte, 1995). In contrast, changing-state sound sequences would be more distracting because the novelty of each item would bring about an involuntary attentional OR. The level of effect of the OR is determined by the goodness of fit between the features of the auditory and visual list information and the degree to which these features are represented in the cognitive representation. The mental representation of the TBR item sequence is progressively established.

However, consensus exists on the balance of experimental evidence that factors associated with the evocation of the attentional OR have a small role in the realm of selective attention (e.g. Allport, 1989). With regards to a repeated changing-state stimulus sequence, the disruption would be assumed to occur at the early stage in the development of the TBR item's mental representation, but there is nothing in the literature to support this possibility. Repeated presentation of an irrelevant changing-state sequence of sounds does not produce a reduction in the magnitude of disruption and there is no modulation of the magnitude of disruption over trials (Ellermeier and Zimmer, 1997). In addition, the changing-state effect does not diminish over experiments which normally consist of several trials (Hellbrück et al., 1995; Jones, Macken and Mosdell, 1997; Tremblay and Jones, 1998) or between experimental sessions separated by a period of time (Banbury and Berry, 1997; Morris and Jones, 1990).

The explanation of the ISE currently offered by Cowan's (1995) model is problematic due to its abstract description, which prevents the formulation of testable a priori hypotheses concerning the ISE. For example, it cannot predict in an a priori manner that although changing speech and tones produce an ISE, speech has been found to be more disruptive of memory (LeCompte, Neely and Wilson, 1997) Also, in its current state it does not comply with findings regarding acoustic-based manipulations of irrelevant sequences. The model argues an increase in disruption and thus distraction is the product of the 'novelty' of successive items, not by the amount of acoustic change between distinct items. It predicts that as the number of novel items within an irrelevant sound stream increases, so should the level of disruption. However, two novel items in an irrelevant sequence cause a level of disruption equivalent to that observed in the presence of five novel items (Tremblay and Jones, 1998). This finding is inconsistent with the predicted OR of Cowan's model and thus cannot be explained.

1.2.4 THE FEATURE MODEL

Nairne (1990) developed the feature model of primary memory, which has been used as a framework to account for the ISE (Neath, 2000). The feature model operates in a primary and secondary memory framework. Primary memory is a system that preserves and processes features relating to the items entering memory. Here a 'feature' of an item refers to an element of a sound. In primary memory items are represented as a set of features, and are recalled from secondary memory, which Neath (2000) refers to as 'memory proper'. It is assumed that primary memory representations are made of modality-dependent features and modality-independent features. Modality-dependent features encode an item's physical attributes that are specific to presentation modality. Modality-independent features encode the internal responses to an item (e.g. converting a graphical item into a phonological representation). Modality-independent features are not modality specific in that they represent attributes of an item that are identical regardless of from which modality they emanate from. This distinction between an abstract form of representation and a form that represents the physical aspects shares consensus with general models of dual coding (Surprenant and Neath, 1996).

During sequential presentation of items, items have access to the primary and secondary store. Items located in primary memory referred to as features are degraded as a consequence of retroactive interference. If a feature of an irrelevant sound item is similar to a feature of a visually presented TBR item, the feature of the TBR item will be overwritten. The model assumes no capacity limit and items are not subject to decay. At retrieval, the item in secondary memory that best matches a degraded feature in primary memory is located (Neath, 2000). Auditory information does not share any modality-dependent features with visual information. This is because memory representations of visually presented items are generated internally (e.g. by sub-vocal rehearsal) and thus convey only modality-independent features. Thus, it is argued the impairment caused by irrelevant speech is determined by the modality-independent features. It is assumed that the disruption of serial recall performance is due to irrelevant sound items adding modality-independent features to the cues of the TBR item representations. This would decrease the likelihood of a match between their primary memory representations and the correct secondary memory representations at recall.

The feature model is supported by the finding that the word-length effect is removed by irrelevant speech. The word length effect refers to the finding that shorter words are recalled more accurately than longer words. However, no difference in serial recall performance is found for short and longer words learnt in the presence of irrelevant speech (Neath, Surprenant and LeCompte, 1998). The observations that adding a suffix (an irrelevant item at the end of a list) to a sequence of repeated irrelevant speech produces a suffix effect above and beyond the ISE and that a suffix effect occurs even when irrelevant speech items are different (Surprenant, LeCompte and Neath, 2000) have provided additional support for the general theoretical stance of the model. The model is able to explain the finding of an ISE when the irrelevant sound is presented during a retention interval (between list presentation and recall) as well as when it is presented concurrently with visual TBR items. This is because the feature model proposes that a feature is only overwritten by another feature from an adjacent item (Neath, 2000).

During the retention interval the TBR items are rehearsed and this allows for the TBR items and irrelevant sounds to be analysed together.

The feature model predicts that disruption is due to interference by content (Neath, 2000) and thus items in the irrelevant sequence which are similar to those in the relevant sequence should produce more serial recall disruption. However, irrelevant sounds and TBR visual items which are phonologically dissimilar produce more serial recall interference than when they are phonologically similar. For example, irrelevant sounds which do not rhyme with the TBR items are more disruptive than sounds which do rhyme with the TBR items (e.g. Jones and Macken, 1995a).

The original model is also unable to simulate the CSH (Beaman, 2000) and cannot explain the effect of irrelevant non-speech stimuli on serial recall (e.g. Jones and Macken, 1993; Jones et al., 1993). The effect of irrelevant speech is explained in terms of feature adoption, which may not be applied to the effect of irrelevant non-speech sounds. The feature model can only provide an account of the differential effect of speech and non-speech stimuli (e.g. LeCompte et al., 1997) if it is argued that speech has more modality-independent features that are more similar to the features of the TBR items. In addition, variations in pitch of items in both speech and non-speech streams produces an ISE, an effect that the feature model cannot explain by the notion of feature adoption. The modality-independent features of the irrelevant sounds are argued to be adopted into the modality-independent features of the TBR items. Pitch is a modality-dependent (physical) feature and thus variations in pitch would not result in feature adoption (e.g. Jones et al., 1999a).

In order to accommodate these effects the model has adopted an attentional parameter, which mediates the overall level of attention or available cognitive processing resources (Neath, 2000). The presence of irrelevant information, whether speech or non-speech, results in a dual task paradigm where ignoring the sounds (secondary task) reduces the availability of processing resources for the memory task (primary task) (Elliot, 2002). This attentional element of the model allows it to account for the 'changing-state effect' – the greater impairment of serial recall by sequences of sounds which change more physically than sequences of identical repeated sounds. Sequences which vary acoustically are argued to re-direct more processing resources from the memory task at hand because sounds which change more physically are harder to ignore than repeated sounds (Neath, 2000). However, this adjustment is argued to be underspecified as it cannot predict in an a priori fashion the amount of attentional resources that will be recruited away from the memory task by a particular irrelevant sound manipulation (Hughes and Jones, 2005).

1.2.5 TEMPORAL DISTINCTIVENESS THEORY

The temporal distinctiveness theory (TDT) (Glenberg and Swanson, 1986) has also been suggested as possibly providing an explanation of the ISE (LeCompte, 1996). The theory assumes that when no other recall cues are present, temporal information is used to aid retrieval of information from primary memory. Therefore, the number of items in a search set determines recall performance. It follows that fewer items in the search set will result in improvement of recall performance (Macken, Mosdell and Jones, 1999). Temporal distinctiveness is therefore characterised by the degree to which stimuli share a search set with other stimuli. Fewer items in a search set are considered to be more temporally distinct (Macken et., 1999). At recall, there exist temporal search sets that incorporate both relevant and irrelevant information at a given point in time.

The ISE is accounted for in terms of the irrelevant sound being subsumed by the same search set as the TBR items. As a consequence memory interference is argued to be determined by the temporal proximity of irrelevant sounds and TBR items. The additional load on the search set by the presence of irrelevant sound results in cue overload and reduced recall (LeCompte, 1996). It follows that irrelevant and relevant items presented in the same temporal interval (e.g. presented at the same time) are more likely to interfere with each other. However, this cannot account for the disruptive effect caused by irrelevant sound presented after TBR items in a retention interval between TBR item presentation and recall (Miles, Jones and Madden, 1991). Interference observed after list presentation is as marked as at the time of list presentation, in spite of the fact that the amount of exposure to irrelevant auditory information is equivalent (Miles et al., 1991). LeCompte (1996) modified TDT in order to account for the post list presentation effect of irrelevant sound. LeCompte (1996) proposed that the delay between presentation of TBR items and their recall would act to diminish the distinctiveness of the TBR items by widening the temporal window which makes the search set. Furthermore, LeCompte (1996) argues the delay would increase the overload of the recency portion of the search set. As a consequence recall performance would be reduced, particularly for the last few items of the list, which has been observed (e.g. LeCompte, 1994). Thus, a reduction in TBR items distinctiveness heightens the susceptibility of the TBR items to interference by unattended sound.

The search set is argued to extend outside the TBR list and TDT wrongly assumes that irrelevant sound presented before list presentation will also impair recall. It is also incorrect in assuming the magnitude of disruption produced will be equal to that produced when irrelevant sound is presented after the start of the list. Macken, Mosdell and Jones (1999) tested this assumption empirically. Five-second intervals of irrelevant speech were presented during one of five intervals in a serial recall task. Participants were exposed to irrelevant speech just before list presentation, at the first half of the list, at the second half of the list, at the first half of a retention interval or during the later half of the retention interval. TDT assumes irrelevant sound will have a greater effect the closer it is to list presentation. Therefore, sound presented from the start of the list should disrupt recall to a degree which is equivalent to disruption observed when sound is presented after the list. Instead, irrelevant sound presented before list presentation had a small effect on recall of TBR items which was non-significant and the disruption was statistically less than that which occurred when irrelevant sound was presented directly after the list. Further, irrelevant sound was more disruptive when it occurred during the last five seconds of the retention interval than irrelevant sound played directly before the TBR list (Macken et al., 1999). Irrelevant sound is also significantly more detrimental just before recall, than sound presented at the first half of the TBR list. Equally problematic for TDT, is that irrelevant auditory information presented directly after the TBR list produced a more marked effect on recall of the first half of the TBR list than when irrelevant sound was presented at the first half of the list (Macken, et al., 1999). In light of the empirical data, it seems TDT needs much refinement in order to account for the ISE and the difference found with irrelevant speech and non-speech stimuli.

1.3 ISE – MEMORY BASED LOCUS

Primarily, most ISE studies have investigated the detrimental effect of irrelevant sound on visually presented graphic stimuli e.g. digits, letters or words presented and read on a visual display unit (VDU). These studies all have demonstrated a changing-state pattern of interference (c.f. Jones, Beaman and Macken, 1996) in that a sequence of sounds which vary physically is more disruptive of serial recall than a sequence of repeated identical sounds. This observed pattern of disruption occurs at an equivalent degree whether irrelevant sound is presented currently or during a retention interval after the encoding of visual information (Baddeley and Salamé, 1986; Beaman and Jones, 1998; LeCompte, 1996; Macken, Mosdell and Jones, 1999). It follows that studies involving the recall of visual, graphic stimuli have established that disruption occurs within memory rather than at perception or encoding (Baddeley and Salamé, 1986; Miles, Jones and Madden, 1991). This consensus is critical for the CSH, which assumes a changing list of items disrupts recall by way of the order cues connecting the irrelevant auditory events, which conflict with conscious efforts at seriation (memory for order) of the TBR stimuli in memory, and therefore has no mechanism to account for a perceptual effect (Jones et al., 1996).

Campbell and Dodd (1984) demonstrated approximately 14% disruption of immediate recall of lip-read sequences from the single, repeated utterance *'bah'*, presented concurrently with TBR material. However, no reliable effect of a single utterance with a graphic visually presented sequence was found. Further studies with lip-read lists have shown a significant changing-state effect with sequences of distinct successive sounds relative to a silent control condition, whereas a single repeated utterance produced only a small non significant effect with immediate recall disruption (Jones, 1994). Campbell, Beaman and Jones (2002) have suggested that differences in the interference patterns found by Campbell and Dodd (1984) and Jones (1994) may be explained by reference to individual differences in sensitivity of participants tested with irrelevant sound.

Campbell, Beaman and Berry (2002) conducted three experiments which investigated the disruption of lip-reading by irrelevant auditory utterances (*bah, dah, gah*, and *lah*) with the aim of distinguishing between any effects of irrelevant sound interference at perception and in memory. Using the stimuli of Jones (1994) they showed that a changing irrelevant auditory sequence interfered with the perceptual identification of lip-read items more than a steady-state auditory sequence. Campbell et al (2002) subsequently replicated the above result and implemented a tighter experimental design, which prevented the occurrence of a memory effect as a different randomisation algorithm was used. In this instance the digits 1-9 were quasi-randomly sampled with replacement as opposed to without replacement, in order to prevent the use of a memory strategy by participants such as 'checking off ' in memory the digits which had been presented. Further, the lip-read stimuli were screened for how identifiable they were to control for the possibility that perceptual irrelevant sound effects observed in their first experiment were produced by stimuli that were already perceptually ambiguous. The third experiment questioned whether or not the changing-state effect observed with lip-read items was indeed an effect located at encoding as suggested by the above experimental series, or in memory, as established by previous studies (e.g. Jones, 1994). Irrelevant sound was presented in the retention interval only. The findings demonstrated a disruptive effect in memory, which is in line with previous experiments of delayed recall of lists of visual, graphic items. This finding is also consistent with the primary assumption of the changing-state hypothesis, that disruption is a consequence of the conflict between two seriation processes within memory. The findings in addition support the observations of Jones (1994) on the assumption that a shared mechanism supports the changing-state interference within memory for lip-read and graphically presented information by streams of irrelevant auditory material.

The lack of a disruptive effect of steady-state sound challenges the previous observations of Campbell and Dodd (1984). The disruptive effect of changing sounds presented concurrently with lip-read TBR stimuli can be accounted for if it is assumed that features of the irrelevant sound overlapped with the representations of the lip-read stimuli, as suggested by the feature model (Neath, 2000). This would result in a reverse McGurk effect (audio-visual fusion), an 'auditory blend' illusion (Campbell and Dodd, 1984). The McGurk effect was first demonstrated by McGurk and MacDonald (1976). These authors showed that when video recordings of a speaker producing a syllable were combined with audio recordings of syllables, so that they occurred simultaneously, participants would perceive a third syllable. When the audio recording of the repeated syllable 'ba' had been dubbed onto the lip movements of the speaker producing the syllable 'ga', participants heard the syllable 'da'. A reverse McGurk effect may have been more detrimental with the numerous features of the changing auditory sequence than with the repeated features of the steady-state auditory information. This would account for the greater disruptive effect of changing sounds on memory for lip-read items (e.g. Campbell and Dodd, 1984; Jones, 1994). However, the feature model would need to be adjusted in order to explain the results of Campbell et al (2002) as it lacks a mechanism to account for the overlap of sound and TBR item features in memory (e.g. Beaman and Jones, 1998; Macken et al., 1999).

It can be argued that the possible role of early (perceptual based) and late (memory based) selection is not a primary issue. This is because the primary task with regard to the degree of interference has been viewed as a specific cognitive process as opposed to a technique of using multiple cognitive resources.

The presence of irrelevant sound after list presentation has been found to produce significant impairment in recall of the first half of the TBR item list, more so than when irrelevant sound is presented concurrently with the first half of the list (Macken et al., 1999). Therefore the proximity of irrelevant sound to the TBR items does not determine the disruption of recall. It is whether the TBR items are retained in memory during the presentation of the irrelevant sound that matters. TBR items at the beginning of the list will remain in short-term memory (STM) as they are rehearsed. It is these items that are most disrupted by irrelevant sound. This is contrary to the assumptions of the TDT, as unlike the CSH, this does not view rehearsal as a critical modulating factor, but as a stage of recollecting encoded events.

Perceptual based organisational factors are important in modulating the effect of irrelevant sound within memory as will be discussed later. Therefore, the separate functioning of perceptual and memory based processing is not clear. Moreover, the problem with attempting to separate perceptual and memorial roles is emphasised upon consideration of the fact that grouping factors involved in perceptual organisation that are seen as pre-categorical have the same effect as factors related to rehearsal in memory, rendering the difference between pre and post categorical representation less obvious.

1.4 THE INTENSITY OF IRRELEVANT SOUND

Generally, empirical research converges with respect to several aspects of the ISE. It is known that the intensity of the sound is not an important factor at least within the range of 40 to 76 dB(A) (Salamé and Baddeley, 1987; Ellermeier and Hellbrück, 1998; Tremblay and Jones, 1998) as it has been established that interference occurs with moderate intensity levels (e.g., Colle and Welsh, 1976; Salamé and Baddeley, 1989). This is true whether sound level is manipulated within or between trials (Tremblay and Jones, 1999). Colle (1980) showed that there is no difference in the amount of disruption produced by irrelevant speech and only at the lowest level of 20 dB[A] does the effect disappear. It has been suggested that the removal of the ISE at a level of 20 dB[A] was most likely due to the auditory signal being only 12 dB above the detection threshold of the listener (Ellermeier and Hellbrück, 1998), however, it can be construed that 12dB is much above the detection threshold and the removal of the ISE at 20 dB[A] may depend on the listener due to individual difference beyond the detection threshold. The observation that varying the intensity level of irrelevant sound has no effect on disruption offers support for a cognitive as opposed to a perceptual basis of the ISE, which further demonstrates the role of pre-attentive processing and acts as evidence against the role of arousal. The absence of an effect of varying the intensity level of irrelevant sound in the ISE paradigm also makes evident the difference between irrelevant sound research and previous research investigating the effect of broadband signals which demonstrated the objective effects of noise on cognitive performance. Broadband noise produced reliable attenuation in performance of vigilance and other mental tasks, only at high levels (more than 85-95 dB), close to levels causing hearing damage (Smith and Jones, 1992).

Ellermeier and Hellbrück (1998) replicated Colles's (1980) finding demonstrating that changing the intensity level of irrelevant sound has no effect on the degree of serial recall disruption. Further, this finding was extended to non-speech sounds (music). Although comparable effects were found with irrelevant music, these were smaller than that observed in the presence if irrelevant speech. Soft and loud conditions did not statistically differ in error rates for either irrelevant speech or musical backgrounds. Further, error rates produced under a control condition of uniform pink noise, comparable in loudness to the loud speech and music conditions did not differ statistically to error rates produced under the silent control. The absence of an ISE in the presence of pink noise is consistent with no effect of intense broadband uniform noise (e.g., Salamé and Baddeley, 1982, 1989). At what signal-to-noise ratio (SNR) the ISE disappears with respect to everyday listening environments, in which background noise is always present was also investigated. As an effect of music and speech is found, which is not the case with uniform pink noise (Salamé and Baddeley, 1989); this suggests that the time-varying acoustic changes in irrelevant sound are important in determining the relative level of disruption by irrelevant sound. This led Ellermeier and Zimmer (1998) to propose that adding additional uniform pink noise to the signal would decrease the ISE, even though loudness increased (speech kept at a constant level). The ISE near the masked threshold was examined by manipulating speech-to-noise level (Ellermeier and Hellbrück, 1998). The masked threshold is the level of degradation below which speech sounds are not detected and are thus masked by noise. Above this masked threshold the speech sounds are detected. Three SNRs of auditory items were generated and went from just beneath the threshold of detectability to perfect discrimination of auditory objects within the irrelevant auditory stream, by varying the level of added uniform pink noise, with the level of the speech signal at a constant level. Over a range no greater than 16dB, the ISE increased from no disruption to maximum disruption. As the level of pink noise mixed with the signal increased, so that the SNR was reduced, a decrease in error rate was observed.

Ellermeier and Hellbrück (1998) went on to reveal the shape of the function linking SNR to the magnitude of the ISE. More SNRs were generated to isolate the range that diminishes the ISE from the highest level of interference to no interference at all. The addition of uniform pink noise systematically to reduce the SNR at higher levels of noise addition removed the ISE. Thus, Ellermeier and Hellbrück's (1998) revealed that loudness plays no role in the ISE and when masking the speech signal the ISE disappears even though overall loudness increases. The removal of the ISE by decreasing the SNR may be explained by the assumptions of the CSH in that there are more non-interfering aspects of the auditory signal than interfering changing-state characteristics of the auditory signal which, when audible disruption cognitive performance. It follows, that masking time-varying acoustical structure of sounds will act to diminish its effect on immediate serial recall and that changing-state sounds are the basis of the generation of the ISE.

It is known that a binaural advantage in intelligibility and detectability is seen when the auditory system can utilise differences between auditory inputs at the two ears (Moore, 2004). Ellermeier and Hellbrück (1998) used the binaural unmasking method to investigate whether a binaural gain in processing the speech signal would serve to reduce recall performance. Conditions of binaural unmasking can be created by presenting the signal to one channel (ear) only and noise in phase to both channels. This situation creates a gain in detectability of the speech signal and is referred to as a dichotic stimulus (Ellermeier and Hellbrück, 1998). This is compared with the situation where both noise and signal occur at both ears in phase, an example of a diotic stimulus.

Ellermeier and Hellbrück (1998) reasoned if masking the time-varying properties of irrelevant speech reduces the negative effect irrelevant sound has on serial recall performance, then the removal of the effect by masking brought about by binaural listening would result in a reduction in serial recall performance. In Ellermeier and Hellbrück's (1998) study two diotic listening conditions were tested, in which speech and masking noise were presented in phase to both ears. In addition, two dichotic listening conditions were tested; where one auditory channel was fed speech and noise which was electronically mixed, and the other channel was fed only noise (which ear received the speech was counterbalanced across participants). Both dichotic and diotic conditions were created for two low SNRs, thus four conditions of speech in the presence of masking noise were constructed. In addition to the four signal-to-noise conditions, silence and speech alone were presented as controls to permit analysis by the level of the ISE. The error rates found at the four signal-to-noise ratio conditions and silence were not statistically different from each other. The main effect of sound input was a function of the increase in errors produced by clear speech alone. However, Ellermeier and Hellbrück (1998) argued that the pattern of results produced by the diotic and dichotic listening conditions was at least in line with a binaural unmasking explanation of the ISE in a qualitative meaner. The findings reveal that when the binaural mechanism is able to partially distinguish the speech signal from the background noise, the error rates increase by about 10 per cent. The binaural-monaural difference is removed at the higher signal-to-noise ratio of -4dB, which was expected due to the fact that the binaural advantage in loudness tends to vanish as the speech signal level is above threshold. It was proposed that the binaural mechanism plays a small role in relation to the magnitude of the ISE (Ellermeier and Hellbrück, 1998).

The finding that loudness is irrelevant to the ISE and that the effect can be removed, even when overall loudness increases provides strong evidence for the CSH in that the relative mixture of steady-state and changing state sound predicts the level of the ISE. To a degree, this result can also be explained by the phonological store hypothesis (PSH), which assumes disruption is determined by the confusion between phonemes in the phonological store (Salamé and Baddeley, 1982; 1989). This theory would assume that fewer phonemes may enter the store as the speech signal is increasingly masked by noise, which would result in a reduction in the interference between the phonologically transposed visual items and the automatically registered sound items. However, the monotonic linear relationship between SNR and memory interference seen (Ellermeier and Hellbrück's (1998) is evidence against a speech filter model that detects phonemes by way of categorical processing. If a threshold of SNR needed to be exceeded in order for the phonemes in the speech signal to be intelligible, a discontinuity in the SNR and memory performance function would have been observed as opposed to a monotonic function. In practical terms, this study suggests adding uniform noise masks the disruptive changing acoustic components of irrelevant sound when presented during a serial recall task. However, annoyance produced by the noise would need to be measured, though difficult to quantify, during different cognitive tasks in order for this to be considered as an effective noise abatement technique in the work place.

Ellermeier and Hellbrück's (1998) demonstrated that a rise in the SNR provided by the binaural hearing mechanism in the near-threshold range during a dichotic listening condition does not produce an ISE that is statistically different from that obtained by diotic listening, which is when both speech and noise is presented in phase to both ears.

1.5 THE ROLE OF NON-ACOUSTIC FACTORS IN THE ISE

Research evidence seems to clarify that the semantic similarity between the irrelevant sound and the to-be-recalled items is not a primary factor in producing interference. For instance, two experimental examples can be given here, both of which afford adequate experimental power. First, if streams of digits are presented concurrently with either two-digit numbers, non-words made of the phonemes of the digits, or words with phonemes that were similar to those of rehearsed digits, no difference in recall disruption is evident (Buchner et al., 1996). Second, if the type of the TBR items is the same as the irrelevant sound items, manipulating the degree of similarity between the modality streams does not increase the magnitude of performance disruption. Buchner et al (1996) found recall performance after presenting a visual stream of TBR digits, along with irrelevant spoken digits which were either in the same range as the TBR digits or a different range was statistically indistinguishable.

Empirical research concerning the meaning of sound (when speech is presented) has established that meaningfulness plays a small role, if any, in the effect (Buchner, et al., 1996). For instance, in terms of the meaningfulness of the irrelevant sound sequence itself, when comparing narrative speech in the participant's native language with a foreign language, no significant differences in disruption have been observed (Colle and Welsh, 1976; Jones, Miles and Page, 1990; Salamé and Baddeley, 1982). Likewise, reversed speech produces a level of interference relatively similar to forward speech (Jones et al., 1990). Only one study observed an effect of meaningfulness in a serial recall task (LeCompte et al., 1997), but the manipulation of meaning was poor and the size of the effect was small (Jones, 1999). The irrelevant meaningful speech in one experiment of LeCompte et al (1997) featured the words *chair*, *sky*, *box* and *egg*. The meaningless condition comprised reversed versions of the words in the meaningful condition. It follows that the meaningful words may have disrupted serial recall more than the meaningless reversed words because the speech-likeness of the reversed stimuli was reduced as opposed to the meaningfulness of the words having a greater effect. However, Jones et al (1990) found no difference between the disruptive effect of speech played forwards and backwards. Jones (1999) has argued that LeCompte et al's (1997) study does not carry much empirical weight alongside studies which have manipulated large features of the irrelevant sound stream (e.g. Jones, Miles and Page, 1990; Salamé and Baddeley, 1982). LeCompte et al's (1997) study is inconsistent with other studies that do not show an effect of meaning as it suggests that weak manipulations of meaning modulate the size of the ISE whereas more substantial manipulations of meaning do not. Another cause to reject the notion of an important effect of meaningfulness concerns the finding that changes at the supra-segmental level do not determine the degree of interference, rather the level of disruption is determined by variation at the item-to-item level, a notion clearly demonstrated by the finding that looping a short sequence many times does not attenuate the magnitude of the ISE (Jones, Madden and Miles, 1992; Tremblay and Jones, 1998).

In contrast to studies of the susceptibility of tasks incorporating a serial recall element, meaning has been found to have an effect in terms of increasing disruption patterns in a primary task in which semantic processing is important, such as in tasks requiring memory for prose, or in some proofreading tasks (Jones et al., 1990). Oswald, Tremblay and Jones (2000) found meaningful and meaningless irrelevant speech interfered with a reading comprehension task, but meaningful speech caused more disruption. Given that rehearsal and semantic processes required during reading comprehension are subject to the disruptive effects of meaningful speech, it can be assumed that semantic properties of irrelevant sound increases interference in cognitive tasks that require meaning to be processed. Therefore, tasks involving memory for prose are likely to be disrupted by the semantic properties of irrelevant sound, as opposed to serial recall tasks where meaning is not required to be processed. In other words, whether the magnitude of disruption is determined by either acoustic or semantic attributes of the sound is dictated by the type of cognitive processing elicited by the task, which is in the attentional focus.

Buchner and Eldfelder (2005) found irrelevant low frequency (infrequent) distracter words disrupted serial recall to a greater degree than did irrelevant high frequency (frequent) distracter words. This effect of word frequency refutes the assumptions of working memory models that do not include the operation of an attentional mechanism, such as the modular working memory model (Baddeley, 1996) and the O-OER model (Jones, 1993). These models argue linguistic, non-acoustic based features of sounds will not influence performance during serial recall and that the probability of an intact representation of a TBR item should not vary due to distracter word frequency. These notions are not supported. The modular working memory model (Baddeley and Hitch, 1974; Baddeley, 1986) proposes irrelevant speech gains automatic and privileged entry into a phonological store. Immediate serial recall of the visually TBR items requires them to be converted into a phonological representation (since the memory module features representations of phonologically based codes), which indirectly enter the store through a process of sub-vocal rehearsal using the limited capacity phonological loop component of working memory. Memory interference results from competing maintenance of phonological representations leading to confusion between irrelevant and TBR phonemic codes (Salamé and Baddeley, 1982). This model cannot account for the effect of word frequency since there is no indication of how low frequency words would be encoded more than high frequency words and thus cause more confusion (Buchner and Eldfelder, 2005).

The object-oriented episodic record (O-OER) model (Jones, 1993; Jones et al., 1996; Jones and Macken, 1993) assumes the important process during serial recall is that of maintenance of the order of the TBR items. The TBR items temporarily enter an episodic surface and are represented by abstract memory representations referred to as 'objects'. These objects are linked by a series of production rules. These act as cues to their temporal order and are referred to as episodic pointers. The objects representing in memory the TBR items and the links forming connections between them form via articulation. Importantly, the representational formats of objects on the episodic surface are of an amodal nature. This refers to the idea that no information with regards to the modality they emanated from is provided. Irrelevant speech items form auditory objects on this episodic surface through preattentive segmentation, separating the sounds into individual auditory objects. A reduction in serial recall performance is a consequence of the acoustic links generated automatically between auditory objects interfering with the links established among visual TBR objects (Jones et al., 1996). The O-OER model would assume that a memory representation in working memory is either accessible if a link points to it or not, which would be where an irrelevant sound successfully competes for that link. This model, however, does not predict that memory representations, or objects as it refers to, can be degraded and so it assumes irrelevant speech items will not effect the probability of successful redintegration (reconstruction) of a TBR item (Buchner and Eldfelder, 2005).

The feature model (Nairne, 1990; Neath, 2000) can account for an effect of word frequency because it includes an attentional parameter that can be adjusted according to the amount of processing resources available for immediate memory. It can be argued that processing low frequency (rare) words in comparison to high frequency (common) words attracts more of a limited attentional resource and therefore results in less of an attentional focus on the visual TBR items. Although the proposed attentional parameter can be included to model effects of word frequency, Cowan's (1995) integrated model of attention and memory provides the best theoretical framework in which to make sense of the effect of non-acoustic characteristics of sound on serial recall. This model would predict that low frequency words would result in more of a reduction in memory for order because the likelihood of a TBR item representation being intact is further diminished if the TBR items are presented along side low frequency words, which automatically attract attention and re-direct processing resources away from the TBR items due to the irrelevant sounds being rare. These infrequent words would act to elicit more of an involuntary attentional oriented response (OR).

Buchner and Erdfelder (2005) proposed that Schweickert's (1993) process model of immediate recall can account for the disruptive difference observed between low and high frequency words because it argues that the identity of TBR items can be reconstructed by successfully matching degraded memory representations with long-term memory traces. During immediate recall it may be possible for an intact representation of a TBR item to remain and therefore it would be recalled successfully. This would happen on the basis of a probability estimate of successful recall at each serial position in Buchner and Erdfelder's (2005) explanation. This means that the probability of an item being successfully recalled can vary for items at different serial positions. A separate parameter is assumed for each serial position and this allows the likelihood of an intact memory representation to differ at different serial positions (Buchner and Erdfelder, 2005). A degraded memory representation could still occur depending on whether or not the item representation can be redintegrated (reconstructed). An additional parameter represents a general likelihood of accurate reconstruction, equivalent for all items at the different serial positions. This parameter would successful predict how distracter word frequency influenced the level of difficulty attached to reconstructing the TBR items from degraded short-term memory (STM) representations, which would be dependent on the level of degradation of the TBR item representations. The accessibility of the long-term memory (LTM) traces determines the likelihood a TBR item in its degraded form in STM will successfully be matched with a representation in LTM. Therefore, high frequency words have more accessible LTM traces than low frequency words and would in turn be more easily reconstructed in working memory.

Other studies have found evidence revealing how linguistic features can serve to moderate the influence of irrelevant sound on immediate memory, through supposedly modulating attentional based resources. Elliott (2002) demonstrated that as age increased memory interference caused by the presence of irrelevant sound was reduced. This was explained with reference to processes of selective attention improving with age in children. Buchner et al (2004) discovered that positive and negative emotionally valent irrelevant words produced more immediate serial recall interference than neutral words. In turn, negative emotionally valent words were more disruptive than positive emotionally valent words.

Buchner, et al (2006) provided a replication of Buchner et al.'s (2004) study in order to rule out a possible confound of differences that exist between acoustic profiles of naturally spoken words related to different emotional expressions. They reasoned negatively valent words may slightly differ in how they are articulated. This might map onto differences in fundamental frequency, the energy envelope or vocal energy between the sounds. Acoustic difference due to emotional expression may be construed as moderating the number of perceived acoustic changes-in-state, which does indeed mediate serial recall performance (e.g. Jones et al., 2000). Artificial associations of negative valence were made with meaningless non-words. The first part of the experiment involved valence induction. Non-words were also associated with neutral valence or were trained to be heard as irrelevant. The valence induction occurred as participants were required to classify non-words by their final consonant. For each trial the consonant judgement task required participants to respond as to whether the final consonant of the non-words presented was I or Z. Three different classes of non-words were formed, depending on the vowel they contained (which was either *'a', 'e' or 'o').* Participants learnt that negative non-words required a correct and fast response or their score for the 'game' would be markedly reduced. Participants learnt that they could take as long as required when responding to the neutral non-words as consequences of making a correct/wrong response would only result in small increments or reductions in game their score. For the irrelevant non-words, participants were not required to make a response. The three classes of non-words then featured in a serial recall task. Buchner et al (2006) found that negatively valent non-words interfered with serial recall of visual TBR items more than did neutral or irrelevant non-words. This data successfully replicated the findings of Buchner et al (2004). In addition, consistent with an effect of emotionally valent words on the modulation of attention stems from the emotional stroop task, which demonstrates that emotionally, loaded words delay the naming of the print colour of a word more than neutral words (Pratto and John, 1991; Wentura, Rothermund and Bak, 2000).

This finding is compatible with models of working memory that specify a role of attention in preserving the order of TBR items for serial recall (e.g. Cowan, 1995 see also Neath, 2000). Within Cowan's (1995) integrated model of attention and memory the visual TBR items are in the focus of attention as they are being rehearsed and are therefore maintained at the highest level of activation. It can be assumed that emotionally charged words attract more attentional resources than do neutral words and that negative words cause more of a shift in attention from simultaneous cognitive demands of the primary task than positive words. Emotionally valent distractor words thus attract and redirect processing resources to states of the environment that need to be attended to. The cognitive system is alerted to these states of the environment by the emotionally valent sounds and this acts to disrupt cognitive processes such as serial recall of TBR items. (Rothermund, Wentura and Bak, 2001). The finding of a significant difference in level of serial recall disruption when comparing the effect of negative and neutral non-words is evidence for them acting to signal a threat in the environment. It follows the emotional valence of the non-words is the important variable mediating the influence of irrelevant sounds as opposed to simple stimulus-response associations, as both negative and neutral non-words required a response in the consonant judgement task in the valence induction phase, whereas irrelevant distractor non-words did not and yet negative non-words were more disruptive.

The feature model (Nairne, 1990; Neath, 2000) can explain the effect of the emotional valence of words, with reference to its additional attentional parameter. This parameter is indicative of the processing resources available for serial memory. It suggests that re-directing attention from the maintenance of the order of the TBR items to the emotionally valent distractor sounds uses more attentional based resources than does re-directing attention to neutral and irrelevant distractor sounds.

The modular working memory model (Baddeley, 1986) and the O-OER model (Jones, 1993) cannot explain the observed differential effect between emotionally valent words because they argue only phonologically-based characteristic of sounds moderate immediate memory for order of TBR items. Furthermore, they do not include an attentional mechanism. The modular working memory model (Baddeley, 1986) does outline an attentional system that of the central executive, but this is argued not to be associated with STM storage of information (Baddeley, 1990). As valence of distractor sounds does indeed influence serial recall performance, Meiser and Klauer (1999) argue for the modular working memory model to be extended to incorporate the central executive, whose role is to coordinate and supervise information processing and storage. The foundation of this suggested extension was the observation that tasks with lower demands on the central executive disrupted serial recall less than secondary tasks demanding a lot of central executive function. An extension of the model assuming attentional based and STM storage processes are not managed by separate modular components would move away from the modules central premise, that being the modularity of working memory function.

The O-OER model (Jones, 1993) could be developed so that seriation of objects in memory requires attention and is subject to the detrimental effect of attentional redirection of processing resources to environmentally relevant irrelevant sounds. But seriation of TBR objects would also still be susceptible to competing auditory processes of seriation occurring on the amodal episodic surface use also the maintain the order of TBR items (Buchner et al., 2006).

1.6 ROLE OF WORKING MEMORY CAPACITY: INDIVIDUAL DIFFERENCES AND THE ISE

A general assumption regarding the basis of the ISE is that a link between working memory capacity and the magnitude of the ISE should exist. This assumption stems from several of the models attempting to account for the effects of irrelevant sound which assume a limited immediate memory capacity. Irrelevant sound either fills the restricted capacity of the phonological store (Baddeley, 1986) or an episodic space (Jones et al., 1996). From this notion the prediction that individuals with a low memory capacity would be impaired more than those with a high memory capacity was derived. Thus, an effect of irrelevant sound would stem from the restricted capacity of memory space left for the maintenance of TBR items. Supportive of this prediction is the observation that increasing the number of items within an irrelevant auditory sequence results in an increase in the size of the ISE (Bridges and Jones, 1996).

Ellermeier and Zimmer (1997) examined individual differences by asking participants to memorise lists of visually presented digits in the presence of foreign speech, pink noise and silence. These auditory conditions were randomly mixed from trial-to-trial. Individual ISEs were a function of the difference in recall errors obtained in the presence of speech and in the quiet control condition. The numbers of recall errors were normally distributed over a wide range and after four weeks the finding of individual differences in serial recall errors was replicated. Both large and reliable individual differences were documented in the participants' levels of susceptibility to disruption by irrelevant sound. In the presence of irrelevant speech the number of recall errors on the serial recall task increased by 50% and the effects were stable over time and the reported individual differences in levels of serial recall interference were independent of gender differences. That is, no difference was observed between serial recall errors for males and females. Ellermeier and Zimmer (1997) failed to find a correlation between working memory span and susceptibility to auditory distraction, as have several other studies adopting different measures of working memory capacity (Beaman, 2004; Elliott and Cowan, 2005; Neath, Farley and Surprenant, 2003). Therefore, the degree of serial recall interference is not influenced by working memory span. If this was the case, experiments manipulating the characteristics of irrelevant sound, such as the experimental series of this thesis, would have to match participants for working memory span.

Complex working memory capacity tasks have been shown to modulate the effect of unattended speech in dichotic listening tasks (Conway et al. 2001). Dichotic listening tasks have often adopted the 'shadowing' method. Here participants are asked to shadow the to-be-attended sequence presented in one ear whilst arguably ignoring irrelevant sound presented to the other ear. Studies of dichotic listening have shown that participants are unable to detect the task-irrelevant speech if it is foreign and if it is reversed (Cherry, 1953). In terms of shared phenomenon, foreign and reversed speech has been demonstrated to disrupt serial recall to the same extent as does speech played forwards (Jones et al, 1990).

Conway et al (2001) used the operation span (OSPAN) task as a measure of working memory and examined whether this could predict performance on a dichotic listening task. The OSPAN task was developed by Turner and Engle (1989). Here a mathematical problem is shown along with a word. The participants are then asked to read aloud the problem and the word and then indicate whether the operation is correct. After a succession of trials, the participant is prompted to recall the words in the correct serial order. The thinking behind using this more complex task is that it would measure the participant's ability to coordinate resources between processing and storage needs. The results demonstrated that participants with a low working memory span were more likely than those with a high working memory span to detect their own name presented to the ignored ear. This was argued to occur because the role of working memory capacity is to maintain activation of relevant information and suppressing irrelevant sounds. Therefore differences in the emergence of the cocktail party effect as referred to by Cherry (1953) can be accounted for in relation to how participants differ in their ability to control their attention; an ability to keep relevant items in the focus of attention, whilst stopping attentional resources from shifting to task-irrelevant sounds. Thus, OSPAN represents a measure of the modulation of an attentional control mechanism. This view is consistent with Cowan's (1995) attention based model of immediate memory where the ISE is the product of diversion of attentional resources from the processing of TBR items by irrelevant sound.

Beaman (2004) questioned whether the type of memory span measure was of importance and whether a more complex working memory task as opposed to number correct in a silent control condition (Ellermeier and Zimmer, 1997) may reveal a relationship between working memory capacity and disruption by irrelevant sound. Beaman (2004) took advantage of the fact that the irrelevant sound paradigm shares a characteristic of dichotic listening, that being they both involve examination of the extent of analysis carried out on task-irrelevant speech. It was assumed that if an affect of memory capacity was found in dichotic listening situations it might also explain individual differences in the affect of irrelevant sound on serial recall performance.

Inconsistent with this was the absence of an effect of OSPAN on the size of the ISE. If executive control of attention, reflected by working memory capacity, mediated the individual differences in the size of the ISE, it was too weak to be detected by using the OSPAN task (Beaman, Bridges and Scott, 2007). Beaman (2004) did, however, find OSPAN predicted the size of the ISE when the disruption was reflected in the number of intrusions from the irrelevant speech sequence mistakenly recalled in a free recall task as present in attended list of visually presented items. The attended list was constructed from low-exemplars of a specific category. The task irrelevant speech stream in the related condition consisted of high exemplars of the same category. In the unrelated condition, the irrelevant stream consisted of high exemplar members of a different category. High span participants were less likely than low span participants to recall irrelevant speech tokens when the irrelevant speech and TBR items were members of the same category, and thus semantically related. There was no confound of guessing, as the probability of a categorically-linked item being mistakenly recalled in the silent control condition would be reflected by the number of intrusion errors occurring in silence. Contrary to this, for the unrelated speech condition the probability of an item related to the same category being mistakenly recalled is represented by the number of intrusions occurring in the presence of categorically-unrelated speech items (Beaman, 2004). Therefore, the silent condition acts as a conservative measure of guessing, whereas the unrelated speech condition is a measure not as conservative, when a decrement in recall occurs and resistance to semantic intrusions is weakened in decision processes whilst recalling TBR items. It seems the finding that the ISE was affected by working memory capacity was due to the semantic analysis applied to the task-irrelevant speech and it is therefore semantic processing of irrelevant speech which is mediated by working memory capacity (Beaman, 2004).

The lack of a link between working memory capacity and susceptibility to distraction by irrelevant sound refutes the claim that dichotic listening and the ISE paradigm examine the same mechanism of interference. This can be explained in terms of disruption by supposedly unattended speech in the two paradigms stemming from two different mechanisms, at least when the standard ISE is considered. Previous research has shown that meaning does not modulate the disruption of serial recall by irrelevant speech. Further, non-speech sounds have been found to be sufficient to produce the standard ISE during serial recall and counting tasks (Buchner et al, 1996; Buchner et al, 1998) of which demonstrate no meaning. The only crucial factor in bringing about the ISE is that acoustically changes occur between each successive item (Jones and Macken, 1993; Jones et al., 1990). Beaman et al. (2007) points out that Conway et al. (2001) measured the occurrence of hearing one's own name in the unattended channel and Beaman's (2004) study involved recalling an intruding irrelevant speech item, both of which are examples of meaningful and indeed intelligible speech. Beaman et al. (2007) suggest that individual differences in distraction in the cocktail party effect and in the number of semantic intrusions during free recall (Beaman, 2004) can be construed as evidence of a shared interference or inhibitory control mechanism required to ignore irrelevant meaningful speech. It is suggested that a different mechanism modulates the ability to screen out the disruptive variable acoustic characteristics of task-irrelevant speech.

1.7 THE TOKEN-DOSE AND TOKEN-SET SIZE EFFECT

Evidence exists suggesting the degree of disruption is related to the number of tokens per time unit in the irrelevant stream, referred to in the literature as the token-dose effect. A token is another term for an irrelevant sound. If, within an irrelevant stream the number of tokens presented per time unit increases, whilst the length of the signal is at a constant, the degree of disruption will improve. However, no such effect is observed for a repeated sequence (c.f. Bridges and Jones, 1996). This finding supports a key assumption of the habituation hypothesis embedded within Cowan's (1995) integrated model of attention and memory, in that the involuntary attentional oriented response (OR) to irrelevant sound should take longer to habituate as more information is presented (see section 1.6 on Cowan's model). The greater speech effect observed by LeCompte et al (1997) is therefore explained in terms of speech having more inherent attention-recruiting properties than tones. However, the model cannot account for the equivalent disruptive effect of speech and tones observed by Jones and Macken (1993).

Empirical data suggests that habituation only plays a small role in the ISE. The habituation hypothesis states that the level of disruption should increase as the number of different tokens in the irrelevant sound stream increases. However, research on token-set size demonstrates that disruption increases substantially when a set of discrete tokens increases from one to two, and beyond this the magnitude of interference does not increase significantly. One study presented sequences with set-sizes of one, two, five and seven tokens and demonstrated a non-linear disruptive function between token-set size and disruption for speech and tones (Tremblay and Jones, 1998), which is evidence against the habituation hypothesis (Cowan, 1995) but supports the CSH (Jones, 1993). The concept of habituation also suggests that the ISE should decrease as trials are repeated, but empirical data demonstrates no evidence of reduction in disruption within (Jones, Macken and Mosdell, 1997) and between experimental sessions (Tremblay and Jones, 1998). Tremblay and Jones (1998) experimental series also showed that token-set size effects are statistically equivalent in nature over adjacent trials. That is, performance has not been demonstrated to improve at a faster rate in trials composed of few tokens relative to trials consisting of many tokens. Furthermore, the ISE cannot be due only to distraction of attention from the task, as the nature of the task accounts partially for the degree of disruption (e.g., Beaman and Jones, 1998).

Factorial combinations of token-set size and token-dose reveal no interaction between these two factors, which is problematic for the habituation based OR theory's mental model framework. Instead, a strong effect of dose and no viable effect of set size is observed (Tremblay and Jones, 1998). The changing-state hypothesis (CSH) can readily account for this lack of an interaction by assuming that the first contrast between two mismatched (distinct) tokens provides information concerning memory for order. Thus, increased token-dose as opposed to token-set size results in more information regarding memory for order that subsequently conflicts with cues to seriation in the serial recall task.

The effect of token-dose is problematic for the O-OER model, which predicts it is the number of changes between successive different items and not the nature of the sound changes is critical. In contrast, the PSH is not developed enough to account for the effect of token-dose or token-set size and the original feature model cannot account for the token-dose effect as it does not possess a mechanism for relating the probability of overwriting to the number of irrelevant items displayed (Bridges and Jones, 1996).

1.8 AN 'ORDER-INCONGRUENCE EFFECT'

As previously discussed, Buchner et al's (1996) research indicates that the semantic similarity between the irrelevant sound and the to-be-recalled items is not a primary factor in producing interference in a serial recall task (see page, 24-26). Hughes and Jones (2005) found that serial recall of visual digits was disrupted more by the irrelevant auditory presentation of the same lexical set of digits than by the presentation of consonants as irrelevant sound, but only when the order of the irrelevant digits was incongruent with that of the TBR digits, an effect referred to as an 'order-incongruence effect'. Thus the content of items in the irrelevant sound stream per se does not result in an increase in the size of the ISE. Rather, This novel effect was replicated and it was also demonstrated that interference was a function of the number of order-incongruent transitions (number of digits in a serial position that differed from the serial position of the same digits in the irrelevant stream, as opposed to the number of distinct items within the irrelevant auditory stream (Hughes and Jones, 2005). This is compatible with an effect of 'token-dose' (Bridges and Jones, 1996) and is consistent with the notion that acoustic variation between successive items is viewed as the critical factor in the modulation of the ISE's magnitude (Tremblay and Jones, 1998). Evidence of no token-set size effect was indicated by the finding that the presentation of eight different consonants and eight digits whose order was congruent with the digits in the TBR list produced no more memory interference than presenting two different irrelevant consonants. The absence of an effect of token-set size is consistent with the CSH's view that acoustic change between two successive items is key to determining the magnitude of the ISE and that the addition of tokens thereafter would produce no more interference. In contrast, there was a clear incongruent-transitions set-size effect. As the addition of more tokens into the irrelevant sequence (from the same set used to generate the TBR item list) invoked more transitions that were incongruent with transitions in the TBR item stream, memory for the order of the digits was reduced (Hughes and Jones, 2005). Therefore, the eight digits incongruent order condition disrupted serial recall performance more than the two digits incongruent order condition. For instance, when the eight digit relevant TBR sequence (e.g. *5,1,7,3,8,4,6,2*) was presented along with eight identical irrelevant digits, but whose order was incongruent with the TBR sequence (e.g. *7,4,1,5,3,2,8,6*), serial recall was disrupted reliably more than when an irrelevant sequence of only two digits, whose transition was incongruent with the transition of the digits in the TBR sequence was presented (e.g. 4,7,4,7,4,7,4,7). Interestingly, the presentation of an irrelevant stream that consisted of two transitions that were incongruent with those in the TBR stream caused more memory interference than two consonants, but also more interference than conditions with eight items, but where there was no order incongruence (8 digits congruent order and 8 consonants conditions) (Hughes and Jones, 2005). This demonstrates that the presence of an irrelevant stream whose tokens are identical to those of the TBR list causes more interference only if their order is mismatched with the order required by the memory task. Therefore, this adds additional weight to the consensus that the ISE is not driven by the number of items or the content of the irrelevant and relevant item lists.

The above findings are problematic for the phonological store hypothesis (PSH) (Salamé and Baddeley, 1982) which argues that disruption stems from the content of the irrelevant sound. Interference resulting from phonological confusions between items cannot account for the order-congruence effect just as it cannot explain the standard ISE. Likewise the feature model (Neath, 2000) cannot address Hughes and Jones's (2005) data. This is because the additional disruption from order-incongruence between the task-irrelevant and relevant sequences is not the result of the content of the irrelevant sounds overwriting the memory representations of the TBR items. Theories advocating that sounds divert attentional resources away from the memory task by an involuntary attentional orientating response (OR) (e.g. Cowan, 1995; see also Neath, 2000) do not predict an order-incongruence effect. These theories cannot account for why irrelevant auditory digits interfere with the serial recall of the TBR digits to a greater degree than do other unattended sounds (such as consonants) when the digits in both attended and unattended streams are incongruent. This suggests that any theory advocating distraction as the sole cause of interference is an inadequate theoretical framework within which the effect can be abstractly explained (Hughes and Jones, 2005).

Hughes and Jones's (2005) data can be linked with the interference-by-process account of the interference between irrelevant and relevant information (Jones and Tremblay, 2000). The task of serial recall requires the serial rehearsal of an episodic record containing the order cues linking the representations of the TBR list of items. Changing-state auditory sequences pre-attentively generate strong order cues and these must be inhibited to prevent them from suppressing the rehearsal of the episodic record of the TBR items. The larger ISE obtained under changing-state sounds is explained in terms of the automatic seriation of irrelevant order cues, a process that is congruent and thus conflicts with the general requirement of the primary task, that of maintaining the order of TBR items. However, at the same time, the irrelevant order cues are incongruent with the need to maintain serial order of the TBR items as seriation of the irrelevant cues uses articulatory transitions in a different order than those which are required for seriation of the TBR items.

The order-incongruence effect can be explained in terms of there being a mismatch between relevant and irrelevant episodic records containing seriation cues during the processing of the order of the TBR items. To illustrate, when a TBR stream consists of the digit list, *5, 7, 2, 4,* then during an order incongruent condition the irrelevant order cues (information in the unattended sound) would reflect transitions that are highly congruent with the type of articulatory transitions that are required to rehearse the order of the TBR information. For example, the serial articulatory rehearsal transitions between the digits *7,4,5,2* are highly congruent with those of 5,7,2,4, which make up the TBR digit list. The transitions within the irrelevant auditory stream of digits would in turn be incongruent with the transitions needed to recall the TBR items in their correct order. This is referred to as the 'action' required by the primary task (Hughes and Jones, 2005). Therefore, the order-incongruence effect is assumed to be the product of the irrelevant sounds being congruent as well as incongruent with the processes of order maintenance during serial recall, and hence require more inhibition. This is not what occurs when irrelevant sounds are unrelated to the TBR items, for which less inhibition is applied (e.g., Jones and Macken, 1993; LeCompte et al., 1997; Jones et al., 2000).

These observations are compatible with a 'selection-for-action' approach to selectivity in attention and interference between competing objects. The disruption from irrelevant sound does not stem from a limited-capacity module or phonological store (e.g., Baddeley and Salamé, 1986) becoming 'full' or the redirection of limited attentional resources (e.g., Cowan, 1995). Instead, disruption of memory for order is the consequence of an inhibitory system that acts to stop irrelevant sounds that are congruent with serial rehearsal processes required by the serial recall task, but incongruent with the maintenance of order of TBR items., from taking control of the seriation process (Hughes and Jones, 2005).

This approach can also provide a framework from within which the effects of unattended speech on performance during a dichotic listening task (requiring participants to shadow verbal information presented to the other ear whilst ignoring information in the unattended ear) can be explained. It is thus reasonable to assume that attentional selectivity plays a role in mediating the ISE. Disruption of serial rehearsal processes by irrelevant sound represents the activity of attentional mechanisms that afford the primary task of seriation of the TBR items to take control of task-directed cognitive processing (Hughes and Jones, 2005).

Bridges and Jones (1996) did not find more interference during the presence of irrelevant auditory tokens that were identical to the TBR sets of permutations of the digits 1-9 than when the irrelevant tokens were unrelated disyllabic words. Hughes and Jones (2005) suggest that this was because these words would elicit more order cues due to sharp acoustic mismatches between syllables within each word. These would therefore produce a larger changing-state effect than the auditory digits 1-9 and in turn may have prevented the emergence of an order-incongruence-effect. Hughes and Jones (2005) intuitively point out that the unrelated disyllabic words (such as *bed, sap, pick and stop*) presented by Bridges and Jones (1996) have a bilabial (/P/), velar (/k/) or palatoalveolar (/t/, /d/) offset and thus it is likely that they displayed sharper energy transitions at word boundaries and more reliable seriation cues than the digits 1-9, which all apart from 'eight' have an alveolar offset (/s/, /n/, /v/) or vowel offset.

1.9 TYPE OF MEMORY TASK AND THE ISE

Research has demonstrated that tasks that encourage or rely on a serial rehearsal strategy, such as serial recall are more susceptible to interference than tasks which do not (e.g., Beaman and Jones, 1997, 1998, Jones and Macken, 1993). One technique (Beaman and Jones, 1997; Jones and Macken, 1995b; 1995c) allows the same sequence of visually presented TBR items, such as a list of days of the week, to be presented in two different ways and thus requiring a different method of recall strategy. Participants may be presented with *Friday/ Tuesday/ Saturday/ Wednesday/ Sunday/ Thursday.* The missing item requires information regarding item identity, with no reliance on order information. This is tested by asking participants to recall the missing day from the list of days presented, which would be *Monday.* In contrast, memory for order information can be tested using a probe. When given the list of days as above and a probe features at the end, participants are required to specify the day following the probe in the list. If presented with the same sequence of days, but given the probe *Thursday*, the item following it would be *Wednesday.* When these tasks are performed in the presence of irrelevant sound the probe task is disrupted more by irrelevant sound than is the missing item task (Beaman and Jones, 1997; Jones and Macken, 1995b; 1995c). The probe version of the task involves seriation, which is susceptible to disruption by irrelevant sound due to the order of changeable sounds being automatically encoded and thus conflicting with seriation of the visual TBR items. This is in line with the prediction of the CSH in that irrelevant sound is more disruptive of tasks that call on memory for serial order (Beaman and Jones, 1997).

Beaman and Jones (1997) found a small but significant effect of irrelevant sound on a missing item task, where lists of digits from the sequence 1-9 were presented. Informal reports of participants revealed that a 'checking-off' strategy was the primary mnemonic strategy adopted. That is, mentally removing digits from the stimulus list as they were presented. Although this task involves memory for items and not their presentation order, rote rehearsal was used by some participants. In order to clarify whether the effect irrelevant sound was indeed due to some participants engaging in rote rehearsal, two variations of the missing item task were contrasted. Here stimulus lists were either learnt in a fixed order or a random order. Learning lists in a random as opposed to a fixed order meant that participants could not rely on a fixed order representation of the stimulus lists in memory which lead them to encode their order. An effect of irrelevant sound was found for the random order condition. In contrast, irrelevant sound had no effect on memory for items in the fixed order condition (Beaman and Jones, 1997).

Other tasks relying on memory other than strict serial recall, such as the missing-item task (Beaman and Jones, 1997, Jones and Macken, 1993) and memory for prose (Banbury and Berry, 1998) are adversely affected by irrelevant sound but to a lesser degree (e.g., Jones and Macken, 1993). Further, tasks that do not rely on seriation or memory, for example perceptual tasks (e.g., Baddeley and Salamé, 1986; Burani, Valker and Buttini, 1991) were found to be unaffected by irrelevant sound. The sensitivity of memory for order to disruption is further emphasised by the observation that other demanding tasks, interrupted in other ways by selectivity in attention, remain unaffected by the presence of irrelevant sound (Jones, 1993).

An important finding with respect to theoretical considerations is that robust irrelevant sound effects have only been demonstrated in experimental tasks requiring memory for the serial order in which the to-be-recalled items are presented (Beaman and Jones, 1997). Tasks requiring memory for the order of items (seriation) are described as having a 'serial component'. Irrelevant sound has been found to effect free recall, recognition and paired-associate tasks (LeCompte, 1994). However, participants undertaking tasks that do not directly require seriation may still use serial rehearsal as a memory strategy (Beaman and Jones, 1997; 1998). It has been suggested that in a free recall task, where participants are instructed to recall the list items in any order, serial rehearsal may be the dominant strategy adopted in free recall based tasks (LeCompte, 1994). Thus, if the primary strategy used in such a task involves serial recall, this would account for the observed disruption by background irrelevant sound (c.f. Beaman and Jones, 1997, 1998; LeCompte, 1994). It follows that the disruptive effect of irrelevant sound in free recall is on order as opposed to item errors (Beaman and Jones, 1997).

In a recognition task, participants were presented with a list of words (Beaman and Jones, 1997). After an eight second delay participants were given two words, one which was from the list and one which was not. Participants had to report which word had featured in the stimulus list. Half the participants had to engage in articulatory suppression by repeating aloud the alphabetical sequence of A-G during list presentation and at recall. Articulatory suppression was used to attenuate the reliance on serial rehearsal at recall. It is argued that the articulatory loop, a component of the working memory model (WMM) is used by participants to subvocally rehearse the presentation order of items (Salamé and Baddeley, 1982). The other half of the participants did not perform articulatory suppression. The performance of participants who engaged in articulatory suppression was improved in the presence of irrelevant sound, whereas the performance of those who did not was disrupted, though not reliably (Beaman and Jones, 1997).

In a paired associate task lists of pairs of words are presented. After each list, the left word from one of the pairs is presented and participants are asked to report the word that had been presented alongside this cue word. As in the recognition task, half of the participants engaged in articulatory suppression during list presentation and the other half did not. Suppressing sub-vocal rehearsal using articulatory suppression resulted in a non-significant effect of irrelevant sound on the paired-associates task (Beaman and Jones, 1997). On the basis of the above findings the effect of irrelevant sound on a recognition and paired-associate task found by LeCompte et al (1994) can be attributed to the predominance of rote rehearsal as a recall strategy (Beaman and Jones, 1997).

Henson et al. (2003) distinguished between the effects of irrelevant sound on a list probe (LP) task and an item probe (IP) task, with reference to several computational models of verbal short-term memory (e.g. Brown et al., 2000; Burgess and Hitch, 1999; Page and Norris, 1998). These models suggest the existence of a 'timing signal' that reflects serial order information over time and consider the coding of items and their order to entail separate processes. The timing signal emanates from a group of internal temporal oscillators and allows for the serial position of items to be encoded. Patterns of errors in memory tasks requiring phonological output, for example the recall of verbal items, can be accounted for by the action of the oscillators during encoding and retrieval of verbal information.

During the LP task a list of items is presented. A probe list is subsequently presented in sequence and participants are required to state whether it is the same or different from the original list. The probe list is always constructed of the same items as the original list and when it differs this is due to a difference between the transpositions of two contiguous items only (Henson et al., 2003). It is postulated this would require a serial processing strategy in that participants would contrast contiguous items in the probe against their memory of the original list. In contrast, the IP task tests for item information as participants are presented with a list of items, followed by the presentation of a single probe, of which they respond by stating whether or not the probe featured in the list (Henson et al., 2003). In both LP and IP tasks the participant responds with a yes or no answer. Performance under the IP task has been accounted for by direct access and how much the item representations are subject to decay. So as to avoid the adoption of serial rehearsal strategies and only require item information, retention intervals were short and items were presented rapidly. It was argued that the LP task rather than the IP task would involve a timing signal, and thus any characteristic of irrelevant sound (which has a temporal component) that moderated the timing signal would affect the LP task more than the IP task. It was observed that performance was reduced under both LP and IP tasks, but the affect on performance was greater for the LP than the IP task (Henson et al., 2003). This is congruent with previous studies indicating tasks that require the adoption of serial rehearsal and thus maintain serial order to be particularly susceptible to disruption by irrelevant sound (Beaman and Jones, 1997; Salamé and Baddeley, 1990).

For the LP task, half of the probe lists were positive and the other half were negative. A positive probe was a probe list that matched the experimental list in order. Negative probes were probe lists that did not match the order of the original list. The errors induced in the presence of irrelevant speech were mainly associated with positive probes. It was assumed processing the irrelevant sound increased transpositions in the order of items in STM, leading to participants making incorrect no responses to list probes that actually were identical to the original list.

For this reason, fewer errors, inconsistent with item transposition, were made for responses to negative probe lists (Henson et al., 2003). Reaction Time (RT) as a function of the serial position of probes measured the degree of serial rehearsal, if adopted, in each task. The reaction time function observed for the IP task was indicative of varying amounts of decay occurring for each item, as opposed to serial scanning (see McElree and Dosher, 1989). The shape of the function for the LP task corresponded to an increase in reaction time over serial position of items in the probe list. This resembled the rate of sub-vocal rehearsal observed for familiar monosyllables (e.g. Baddeley, 1986) and is evidence that LP makes use of a serial rehearsal strategy.

Phonological similarity was manipulated in the probe lists. For the LP tasks a phonologically confusable probe was one where the adjacent transposed letters were phonologically similar. For the IP task, however, a phonologically confusable probe was where at least one of the items in the original list was phonologically similar to the probe item. Performance under both tasks was impaired when probes were phonologically confusable, which is evidence that both tasks had accessed phonological STM (Henson et al., 2003).

Interestingly there was an effect of irrelevant speech on the IP task. Henson et al (2003) suggest that this may be due to the unattended speech having different interfering effects. As the IP task was affected as well as the LP tasks, this may be because of a general distraction of attention by the irrelevant sounds as well as unattended sound having a more specific effect on seriation processes that only affects the LP task. This assumption would be compatible with Cowan's (1995) integrated model of attention and memory and Neath's (2000) use of the feature model, which both propose that all tasks requiring attention and memory are susceptible to the detrimental effect of irrelevant sounds.

1.10 THE ORGANISATION OF AUDITORY OBJECTS: STREAMING EFFECTS

The role of seriation in the perceptual organisation of sound becomes clearer when the findings of studies into the organisation of sound are considered. Auditory stream segregation results in the perceptual organisation of sounds in the environment. That is sounds emanating from different sources are streamed apart, and form separate streams (Bregman, 1990). The auditory scene is therefore partitioned into relatively stable and temporally extended perceptual objects. As an extension to partitioning the sound into separate perceptual objects, the process of perceptual organisation of objects also incorporates the maintenance of order in which acoustic events occur in each object. An example of an auditory 'object' would be words produced by a single speaker. It seems that perceptual streaming incorporates two concurrent processes, one partitioning the acoustic objects (e.g. voices), and the other maintaining the order of events within the streams produced by those objects. Crucially, it is thought that this process of order maintenance is performed automatically and is thus rendered obligatory (Bregman, 1990).

The link between the perceptual organisation of irrelevant sound and the maintenance of the order of TBR items has been demonstrated experimentally in two ways. First, the finding that when items within an unattended auditory stream are similar and hence less distinct, the level of interference is attenuated in a monotonic linear manner (Jones et al., 1999) is empirically robust. However, it is misleading in that one may assume that a sequence whose object members are highly dissimilar would be more detrimental than a sequence which comprises indistinct objects. Empirical data however does not support this assumption. Jones et al (1999a) states that in contrasting the disruptive effect of a stream of indistinct vowels, spoken in the same voice with neutral intonation with a stream of different musical instruments (e.g. horn, guitar, violin and trumpet), it is the vowel sequence that is more disruptive. This could be regarded as problematic for the 'distinctiveness' assumption, that argues as auditory objects become more distinct, so does memory impairment increase.

It is well established that changing sounds produce more disruption than repeated sounds (e.g. Jones and Macken, 1993; Neath, Surprenant and LeCompte, 1998). This effect can be changed, by the perceptual organisation of sound with regard to spatial location. The sequence of syllables 'x, j, w', when presented to both auditory channels simultaneously is perceived as a fixed coherent stream, and in keeping with the changing-state effect sequences such as this are very disruptive of immediate serial memory. However, if each of the three syllables is presented from a separate auditory location in space, so that the 'x' is presented in the left auditory channel, 'j' in the centre of the head and the 'w' in the right auditory channel, then three streams are perceived, each consisting of a repeated auditory object. Therefore, when these streams of extended auditory objects feature as irrelevant sound, serial recall interference is reduced (Jones and Macken, 1995b; Jones, Saint-Aubin and Tremblay, 1999b). This reveals how auditory streaming can mediate the disruptive potency of irrelevant sound. Thus, the link between memory for order and dissimilarity is non-monotonic, such that as sounds elicited by the same source, such as a voice become increasingly different, temporal order information is improved. However, beyond a threshold of change when the auditory items become separate objects (such as different musical instruments) and thus form separate streams of repeated items, order information is diminished.

Pitch is another characteristic of sound that can be manipulated so that either one coherent changing stream or multiple steady-state streams are perceived. If the pitch difference between successive tones or vowels in an irrelevant stream is increased, initially serial recall interference increases. However, as the pitch difference exceeds the threshold of change, which might be referred to as the binding threshold, the level of serial recall disruption decreases (Jones et al., 1999a; Macken et al., 2003).

These findings show that small variations on an attribute shared by sounds provide more order information than a sequence which comprises sounds from very different sources. Order information is produced automatically when change exists in an irrelevant auditory stream, and this therefore conflicts with order information generated from the sub-vocal rehearsal of TBR item lists in the serial recall task. This is consistent with the changing-state hypothesis (CSH) which incorporates the role of the perceptual organisation of sound in explaining the ISE (Jones et al., 1996).

Early research has shown that individuals have great difficulty in judging the order of attended to auditory objects if the sounds within the auditory stream come from a variety of sources (Broadbent and Ladefoged, 1959; Warren and Obusek, 1972; Warren, et al., 1969). For example, participants have difficulty judging the order of a looped sequence of four unrelated sounds (burst of white noise, a tone, a vowel sound and a buzz), despite the fact that the sounds are unrelated and distinct. Interestingly, memory for a sequence reveals a 50% increment in accuracy of order judgements if the sequence features two objects that are variations of one another and if they feature adjacent to each other within the irrelevant sequence (e.g. white noise, a high pitched tone, a low pitched tone, a vowel sound and a buzz). In this instance, memory for order is good, because two of the auditory items making up the sequence differ on an attribute shared by the tones, that being their pitch (Warren and Obusek, 1972; Warren, et al., 1969). Although these examples of sequential ordering are based on the retrieval of attended sounds, they demonstrate the same non-monotonic relationship between dissimilarity of auditory items and seriation observed with unattended sound in the ISE paradigm.

1.11 PRACTICAL IMPLICATIONS OF THE ISE

The rationale and generation of interest surrounding the study of the ISE has centred on the finding that the processing of sound is obligatory, requiring no conscious control or effort. The obligatory processing found with unattended sound has several implications not just theoretically, but of a practical nature. Irrelevant sound research has many important implications for the understanding of noise interference in a range of settings. Banbury et al (2001) point out that the number of manual jobs has decreased as the number of jobs involving cognitive tasks has increased, where the accuracy of short-term memory, particularly memory for order is important. Irrelevant background sound is an inconspicuous cost to both industry and individuals. Irrelevant sound research can aid the development of methods of acoustic alteration, which aim to diminish the variability of extraneous sound (Banbury et al., 2001).

As irrelevant speech has been shown to affect a range of cognitive tasks such as serial recall, reading comprehension and reasoning to name a few (Beaman and Jones 1997; Oswald et al. 2000), it is reasonable to assume that its presence in an office environment will reduce work performance. This research has had the particular aim of looking at how to reduce the effect of irrelevant speech on cognitive performance.

Reverberation is the product of multiple sound reflections produced by the signal bouncing off of the surfaces of objects within a room. It is defined as is the time taken in seconds for a sound to drop 60db below its original level before decaying (Beaman and Holt, 2007; Perham, Banbury and Jones, in press). Single sounds come into contact with various surfaces which either absorb (e.g. soft ceilings) or reflect (e.g. hard ceilings) depending on the absorption rate of its physical properties. Auditory reflections occur on large surfaces, such as ceilings, walls and windows. A room which causes reverberation to a higher degree will in essence prevent the sounds from decaying and thus it would take longer for the sounds to attenuate by 60dB. In contrast, in a highly absorptive room, sounds drop more quickly. Soft acoustic ceilings that are applied to offices act to attenuate reverberation and also reduce the intensity of sound, but leaving speech comprehensible (Beaman and Holt, 2007).

Reverberation has often been thought to be detrimental to the working environment. This has lead to the instalment of acoustically treated ceilings and wall panels which act to attenuate the level of reverberation. For example, engineers have installed soft ceilings which absorb rather than reflect sound (Beaman and Holt, 2007). However, recent research has clearly shown that reverberation, at extremely high levels is less disruptive than low level reverberation. Beaman and Holt (2007) ran a serial recall experiment using a high and low level of reverberation and a silent control as irrelevant auditory conditions. Highly reverberated speech interfered with immediate serial memory no more than the level of error found during the quiet control condition. The CSH is able to account for this finding if one considers how reverberation acts to smooth the profile of the waveform envelope, effectively attenuating the abrupt multidimensional acoustic changes in the speech signal. The effect of longer reverberation times on speech is similar to that found with 'babble', which is where multiple speakers are present within the irrelevant auditory stream. Jones and Macken (1995c) varied the number of voices making up the babble from five to eight and observed error rates that fell between those found for performance in silence and performance in the presence of a single speaker. The addition of more voices made the irrelevant sound more noise-like, arguably because this too acts to smooth the temporal envelope of the sound. Babble both reduces and attenuates the peaks and troughs in the speech signal, and therefore this reduces the number of acoustic based 'changes-in-state' (Jones and Macken, 1995c). The increase in number of voices masks the onset and offsets of individual sound elements, which reduces acoustic change.

However, Perham, et al (in press) argue that the high reverberation level of 5 seconds used by Beaman and Holt (2007) is not one which typically acts on background sound within open plan offices, but are instead those experienced in large auditoriums and concert halls, which is good for musical performances as it creates a richer sound, but degrades intelligibility. Perham et al (in press) ran an experiment with reverberation times that were more representative of an office environment. The experiment consisted of a high, low and no reverberation condition and a quiet control. The primary memory task was the typical immediate serial recall task. The high reverberation condition consisted of the multiple reflections of a speech signal mimicking that produced by bouncing off a hard ceiling. The reverberation time for the high level condition was at 0.95 and 0.75 for the low reverberation condition, which as Perham et al (in press) argue, are more representative of the typical office reverberation time of between 0.45-1 seconds. In contrast, multiple echoes of a signal produced in the presence of a soft ceiling were used to reflect low reverberation levels for the low reverberation condition. It was found that there was no difference in performance between low (soft ceiling) and high (hard ceiling) reverberation conditions (Perham et al., in press). This demonstrates that soft ceilings which contain sound absorptive materials that reduce reverberation time do not act to reduce the deleterious effects of irrelevant sound to a reliable degree in comparison to hard reflective ceilings. Instead, both reverberation times disrupted memory relative to a silent control, revealing reverberation times were not long enough to attenuate disruption.

The use of automated systems in aviation has also increased the amount of cognitive tasks the pilot has to engage in. These automated systems have increased the amount of irrelevant sound in the cockpit by these systems outputting auditory messages. Research evidence indicates the importance of seriation in a pilot's ability to sustain adequate situation awareness of the aircraft systems, with regards to both the prediction of future systems states and immediate comprehension of situation (Endsley, 1995).

CHAPTER 2

2 LITERATURE REVIEW: THE EFFECT OF SPEECH AND NON-SPEECH SOUNDS

2.1 INTRODUCTION

Empirical findings relating to the processing of unattended sound is reviewed in this chapter. As this thesis looks at which characteristic(s) of the speech signal can explain its greater disruption of serial recall in comparison to non-speech (e.g. LeCompte et al., 1997 and Tremblay et al., 2000), the literature that looks at the effect of speech and non-speech on serial recall will be discussed. Sounds within the irrelevant stream have been manipulated in various ways. For example, both speech and non-speech has been degraded systematically and the effect of degradation on the level of serial recall interference will be examined. The characteristics of irrelevant sounds that have been manipulated, such as the phonemes that change within a sequence of speech sounds to determine their relative disruptive power will also be discussed. The chapter closes with a statement of the aims of the present research and predictions generated by existing hypotheses.

2.2 HEMISPHERIC PROCESSING OF UNATTENDED AND ATTENDED SOUND

Long-standing accounts regarding the processing of attended auditory information converge on making a distinction between speech and non-speech and the role of the left and right cerebral hemispheres The majority of the literature concerning the hemispheric location of mental functions (e.g., Kimura, 1961a; 1961b) refers to the speech dominant left hemisphere and the non-speech dominant right hemisphere.

Kimura's research has been used as a theoretical framework to investigate the existence of a distinct processing mechanism for unattended sound. Kimura (1961a; 1961b) conducted dichotic listening tasks and showed that the majority of right-handed participants were able to shadow or identify verbal material which was presented to the right ear more accurately than when the same stimuli were presented to the left ear. Kimura referred to this observation as the right ear advantage (REA). The finding that performance is more accurate or faster for verbal information presented to the right auditory channel than it is when presented to the left auditory channel is argued to indicate functional asymmetry, in this case demonstrating specialisation of the left hemisphere for language processing (Voyer and Flight, 2001). The REA indicates stronger transfer of auditory information to the contralateral hemisphere. This shows that contralateral as opposed to ipsilateral connections are stronger (Voyer and Flight, 2001). Evidence demonstrating this functional distinction comes from other sources including behavioural studies using brain damaged patients which have identified a clear functional dichotomy between the two cerebral hemispheres (e.g., Baum, Pell, Leonard, and Gordon, 1997).

Other studies focusing on the processing of sound have also provided evidence for contralateral activation during monaural auditory presentation. This has been observed when Consonant-Vowel-Consonant (CVC) syllables and tones have been presented monoaurally (to one ear only, e.g. the right ear) (Jancke, Wustenberg, Schulze, and Heinze, 2002). This research also supports the idea that sound presented monaurally activates the contralateral hemisphere faster and more efficiently than when auditory information is presented binaurally (e.g. Hirano et al., 1997; Jancke et al., 2002). The contralateral direction of processing for auditory information is further made stronger because ipsilateral routes are suppressed, or access to them is prevented by the presence of contralateral auditory stimuli. Verbal information played to the left ear, which is processed by the right hemisphere is routed to the left hemisphere to be sufficiently processed. However, this route is suppressed by concurrent verbal stimuli received at the right ear (Beaman et al., 2007)

In contrast, Voyer and Flight (2001) put forward an alternative account of the processing of sound, which argued that attentional factors, due to individual predispositions, act to bias or attenuate the occurrence of the REA in dichotic listening tasks. It is also known that when the right ear is stimulated by verbal information, the processing regions of the left hemisphere are activated, which then prime this speech-dominant hemisphere to receive more information from auditory space. In turn, the increased activation of the left hemisphere increases the participant's awareness of the right side of both visual and auditory space, and thus results in more accurate reports of items presented to this side.

Data from imaging studies indicate verbal STM to be predominately localised in the left hemisphere (Baddeley, 2003; Henson et al., 2000; Logie et al., 2003; Paulesu et al, 1993). For example, Paulesu et al (1993) found neural correlates of the verbal mechanisms of Baddeley's working memory model in the left inferior parietal cortex, the left premotor cortex and the right cerebellum regions. In addition, Hickok and Buchsbaum (2001) have argued for the involvement of temporal lobe speech perception systems in verbal working memory. On the basis of this evidence it was originally assumed that if irrelevant speech was presented to the right ear only this would be afforded direct and obligatory processing by verbal STM. Further to this, irrelevant speech presented to the left ear only may follow a weaker transfer route to gain access to the left hemisphere and would disrupt memory less. This is said to be due to suppressed or blocked ipsilateral connections between the left hemisphere and left ear (Haddlington, Bridges and Darby, 2004).

The CSH (Jones et al, 1996) argues that it is the time-varying acoustic makeup of the irrelevant sound, as opposed to its nature, which is the basis of the interference seen in the ISE. In view of the literature on the behavioural affects of unattended sound in dichotic listening tasks, if speech holds special status in the ISE paradigm, irrelevant sound presented to the right ear and thus processed in the left hemisphere should produce a larger ISE. Hadlington, et al. (2004) and Hadlington, Bridges and Beaman (2006) investigated whether the sound's physical composition as opposed to its nature is an essential element, as assumed by the CSH in determining obligatory processing. If this were the case sound would be processed more efficiently in the right as opposed to the left cerebral hemisphere. This assumption is based on the right hemisphere specialisation for the processing of the structural characteristics of sounds (e.g. Kimura, 1961a, 1961b). Hadlington et al (2004) and Hadlington et al (2006) observed a left ear disadvantage (LED) for the processing of irrelevant sound. A speech sequence made from the letters *B/I/J/N/Z* and a sequence of tones differing in pitch played to the left ear produced more immediate memory interference than when these sequences were presented to the right ear. This effect was demonstrated using a mental arithmetic task and a serial recall task. An LED was not found, however with steady-state unattended sounds (Haddlington et al. 2006). These results provide further support for the main assumption of the CSH, that it is not the nature of the sound but the changeable acoustic structure of the sound that is the critical determinant of the interference seen in the irrelevant sound effect (Jones and Macken, 1993, and Jones et al, 2000). Also, an LED contradicts the hypotheses derived from the dichotic listening and imaging literature, detecting a left hemisphere specialisation for both speech processing and verbal working memory. Beaman et al (2007) infer that this is further support for the notion that unattended sound in dichotic listening situations is processed differently from unattended sound in the ISE paradigm. In addition, all sounds presented to the left ear only produced more interference than sounds presented to both ears. One inference was that processing of sound in the right hemisphere is modulated when both hemispheres receive the same input (Beaman et al., 2007).

Hadlington et al's (2004; 2006) demonstration of a LED for the processing of irrelevant sound is in line with the original notion that irrelevant sound may have a fundamental area of disruption located in the right hemisphere, or certainly may have some of its direct pre-attentive processing occurring within the right hemisphere. According to the assumptions of the CSH, this would be perceptual attributes related to the prosodic, spectral and temporal form of the sounds presented (e.g., Jones et al., 1999a; Jones et al., 2000). The greater disruption of memory for order that is observed when irrelevant sound is presented to the left ear can be accounted for by previous theoretical ideas concerning the ISE and the functional specialisation with respect to different processing preferences of each cerebral hemisphere. The findings of Hadlington et al (2004; 2006), provide support for the suggestion that the right hemisphere is specialised for processing stimuli in a non-relational or holistic way. Taking this evidence along with the assumption that the content of irrelevant sound is not processed (e.g. Buchner, Irmen, and Erdfelder, 1996; Jones and Macken, 1995a) and the finding of a LED for the presentation of both speech and non-speech sounds is particularly damaging for the PSH (Salamé and Baddeley, 1982; 1989) which describes interference as resulting from phonological confusions within STM. Also, these observations do not fit with the original prediction that because the left hemisphere is specialised for processing speech and maintains the neural correlates of working memory function, a right ear disadvantage would be observed.

Imaging studies have used positron emission tomography (PET) to investigate the location of the ISE (Gisselgård et al., 2003; 2004). Gisselgård et al.'s (2003) study featured visual TBR items presented concurrently with irrelevant speech. The baseline for a comparison of performance with serial recall was a digit repetition task. When neural activity in the speech conditions was contrasted with that occurring under a silent control, high activation was observed bilaterally in the superior temporal region. Changing versus steady-state (sequences of identical sounds) speech comparisons demonstrated a significant decrement in activation in the left superior temporal cortex and a weaker but still significant decrement in the left inferior parietal cortex, bilateral secondary auditory and inferior/middle fontal areas. These observed reductions in activation during the serial recall task were related to the effect of 'changing-state' irrelevant speech and are consistent with Hadlington et al's. (2004; 2006) data indicating a right hemisphere preference for processing unattended speech in the ISE. Gisselgård et al. (2003) observed the decrement in levels of activation was greater in the left than in the right superior temporal area. They also assumed that the little activation apparent in the left parietal cortex in serial recall compared with that seen during the digit repetition may be linked to an overall inhibitory effect of varying irrelevant speech (Gisselgård, 2003).

The neural processing of unattended speech has been easier to decipher using PET studies regarding speech perception in the presence of irrelevant competing speech, because only the processing of two types of sound is examined. The advantage of this is neural activity is not confounded by memory based tasks (e.g. Gisselgård et al., 2003; 2004). These studies provide data that convey no evidence upon which to conclude that the neural processing mechanisms devoted to attended speech would differ from those analysing and processing unattended speech (e.g. Narain et al., 2003; Scott et al., 2000). In contrast, the right superior temporal lobe is activated in the presence of sounds demonstrating dynamic pitch changes, independent of intelligibility (Patterson et al., 2002; Scott et al., 2000; Zatorre et al., 1992). If it is the dynamic pitch changes processed in acoustically changeable irrelevant sounds then this would provide a framework within which the critical assumption of the CSH can be explained.

2.3 THE IMPORTANCE OF DYNAMIC PITCH VARIATION

Speech demonstrating sufficient change between adjacent items results in the ISE whether or not speech is intelligible (Tremblay et al., 2000). Thus the appearance of the ISE is independent of an effect of intelligibility. Changing-state sounds convey abrupt and variable pitch changes which steady-state (repeated) sounds do not. The largest ISE is produced through the presentation of sound at the left ear. In terms of speech perception, analysis of speech varying in pitch is observed through the activation of the right superior temporal gyrus (STG). Scott et al. (2004) examined whether or not the above finding for speech processed in the focus of attention matched that of unattended speech. Listeners had to shadow a female speaker in the presence of either a male speaker or continuous noise. When neural activity whilst listening to speech in the presence of unattended speech was examined much bilateral activity was observed, compared to neural activity elicited when listening to speech in the presence of unattended continuous noise. The bilateral activation observed by Scott et al.'s (2004) supports the findings of Gisselgård et al. (2003; 2004) who compared neural activity in changing-state and steady-state (repeated) speech conditions with a silent condition, even though the cerebral activity observed by Gisselgård et al. (2003; 2004) was confounded by additional neuronal excitation by a digit serial recall task. The bilateral activity found by Scott et al.'s (2004) could have been the product of the semantic analysis of the unattended speech or its acoustic analysis.

Scott et al (submitted, cited in Beaman et al., 2007) examined the processing of lexical-semantic characteristics relative to the acoustic processing of unattended speech masking the to-be-attended speech. This was conducted to investigate whether or not the findings above were in some way dependent on the masking stimulus chosen. Here the to-be-attended speech (female speaker) was presented simultaneously with three different types of unattended masking stimuli. These were speech (male speaker), spectrally rotated speech (male speaker) generated using a spectral inversion technique (Blesser, 1972) or unattended signal correlated noise (SCN) versions of the speech spoken by a male speaker. All versions of the unattended sounds were amplitude modulated, so that periods of silence appeared for all unattended auditory sequences. Therefore, in all conditions, silent gaps would afford glances at the unattended sounds, which was not possible for the SCN condition in Scott et al's (submitted, cited in Beaman et al., 2007) study. Spectrally rotated speech acted as the baseline for a more precise investigation of how attended and unattended sound is differentially processed by the auditory system. The signal inversion (rotation) transformation maintains the spectral and temporal structure of the speech signal whilst rendering it unintelligible (Narain et al., 2003; Scott et al., 2000). Figures 3a and 3b display the spectrograms for the untransformed and spectrally rotated versions of the non-word /lowch/ (l6J) as an example (see appendix 5 for examples of disc phonetic symbols).

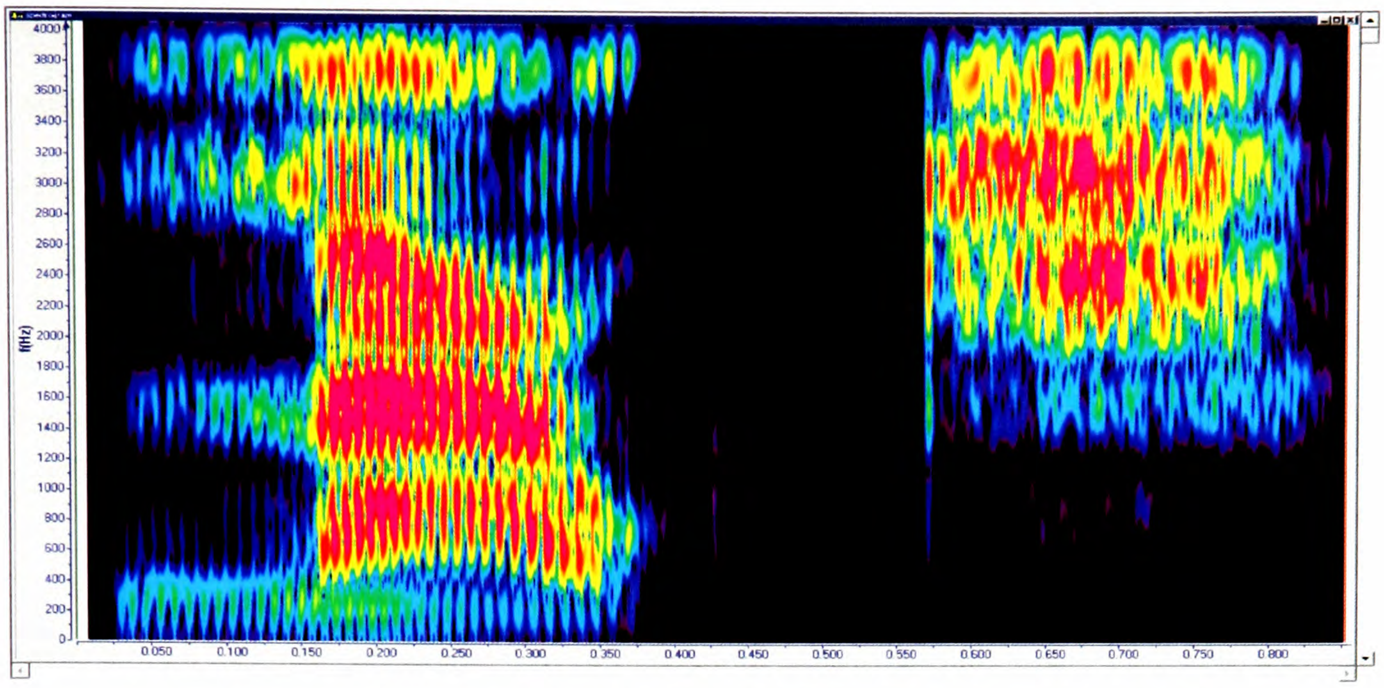


Figure 3a. Untransformed version of the non-word 'lowch' spoken by a male speaker.

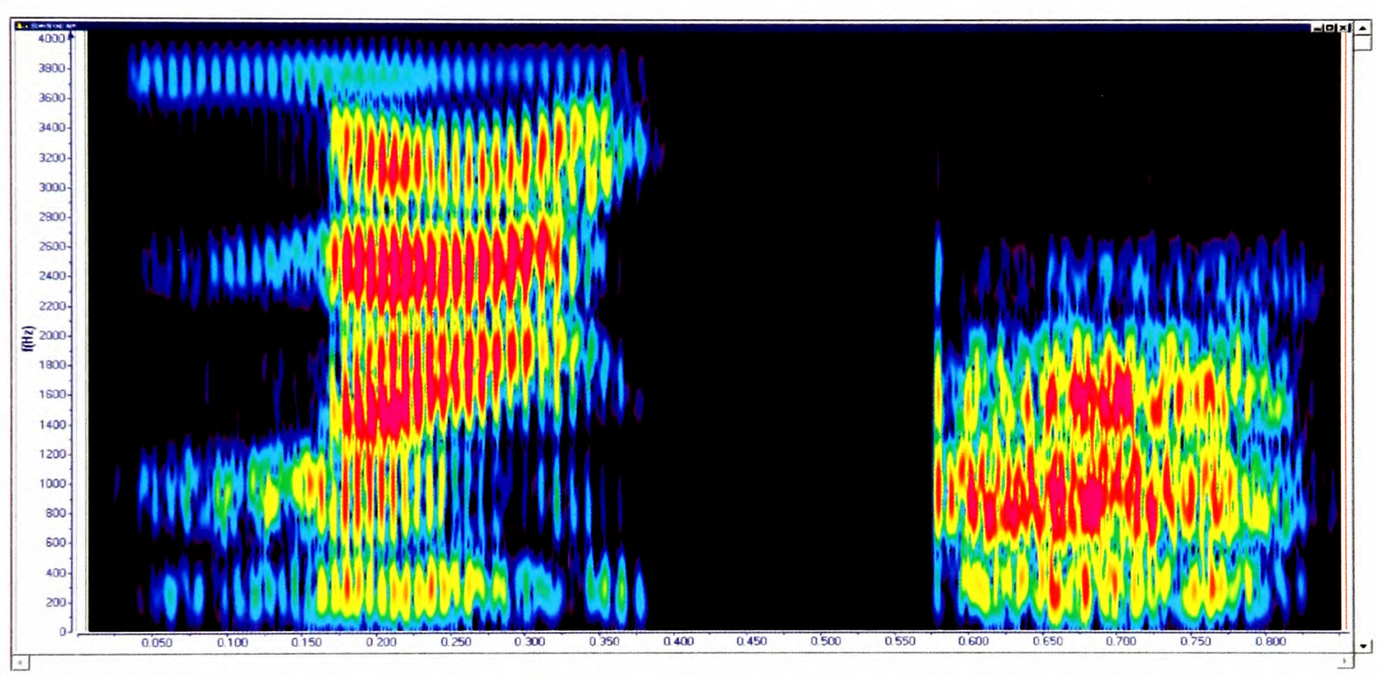


Figure 3b. Spectrally rotated version of the non-word 'lowch' spoken by a male speaker.

It follows that spectral rotation of the speech signal preserves acoustic characteristics associated with changes in pitch which represent important acoustic based changes-in-state. However, it does not preserve meaning of the signal and so lexical activations cannot confound the data (Beaman et al., 2007). By comparing spectrally rotated speech and SCN, Scott et al. (submitted, cited in Beaman et al., 2007) were able to look at neural activation by pitch variation in unattended transformed speech and activation elicited by the lexical-semantic identity of unattended untransformed speech in isolation from each other. Listening to speech in the presence of unattended speech relative to speech in the presence of unattended SCN led to bilateral activity in the STG. This shows that unattended speech was subject to more analysis than unattended SCN. When neural activation in the presence of unattended spectrally-rotated speech relative to unattended SCN was examined most activation was in the right STG. This supports Hadlington et al.'s (2004; 2006) data, where an LED was documented, indicating greater processing of unattended sound in the right hemisphere. This comparison also directly demonstrates the neural processing applied to irrelevant spectrally-rotated speech with pitch changes is equivalent to irrelevant untransformed speech (Beaman et al., 2007). Therefore, right hemisphere processing of pitch variation must be related to the ISE. Irrelevant spectrally-rotated speech which is unintelligible, but conveys the same pitch variation as untransformed speech is processed in the right hemisphere. This is evidence that processing of unattended speech occurs solely at an acoustic level and is based on the analysis of pitch variation, as opposed to analysis being at a lexical-semantic level (Scott et al., submitted, cited in Beaman et al., 2007). Beaman et al (2007) argue that changes in pitch analysed by the right hemisphere represent the changing-state characteristics of unattended sound that result in the LED in the ISE (Hadlington et al., 2004; 2006).

2.4 IRRELEVANT SPEECH AND NON-SPEECH IN THE ISE

Jones and Macken (1993) directly compared the disruptive effect of speech and non-speech stimuli. In their second experiment they presented quiet, irrelevant pitch varied tones or irrelevant speech (i.e., random sequences of the four syllables (C, H, J, and U) while participants saw and then recalled a series of letters in serial order. Relative to the quiet condition, the presentation of both the irrelevant tones and irrelevant speech substantially impaired recall performance. Furthermore, there was no significant difference between the effect of speech and tones. As an extension to this test, Jones and Macken (1993) compared the same four tones with the utterance "ah" spoken at four different pitches. The results were consistent with their previous experiment. It was concluded that speech is not critical to producing the irrelevant sound effect and that speech and non-speech equally impair recall of visually presented stimuli. This supports the CSH notion that tones and speech are coded in the same way and are therefore equally capable of disrupting serial recall. Even when a syllable is repeated, changes in pitch are sufficient to cause changing-state effects. Jones and Macken (1993) challenge the differentiation between speech and non-speech (tones) suggested by the PSH. Furthermore, the evidence that a sequence of tones impairs recall is evidence against a speech-based store (Jones and Macken, 1993).

The suggestion that speech and non-speech stimuli have equivalent affects on primary memory is not compatible with data on the suffix effect. The suffix effect refers to the finding that when a redundant item, such as the word 'go' (stimulus suffix) is presented at the end of an auditory list, recall of the final list item is significantly attenuated (Surprenant, LeCompte and Neath, 2000). A speech suffix impairs recall more than does a non-speech suffix (e.g., LeCompte and Watkins, 1995; Neath, Surprenant and Crowder, 1993). For example, Neath et al (1993) manipulated whether or not a speech suffix was heard as speech or non-speech. The sound "ba" could be perceived as spoken by the speaker who read the auditory list preceding the suffix or as a sound produced by a sheep. The "ba" sound was presented along with other lists that did feature real animal sounds. When "ba" was interpreted as speech, recall of the final item was dramatically reduced relative to when "ba" was interpreted as an animal sound. The proposition that speech and non-speech stimuli affect memory to the same degree is also inconsistent with primary memory findings outside the suffix effect. For example, LeCompte and Watkins (1993) observed that speech tokens (e.g., spoken words *whistle, bugle,* and *horn*) were recalled at nearly twice the rate of non-speech tokens (e.g., sounds of the above words).

LeCompte et al (1997) questioned the findings of Jones and Macken (1993). First, in Jones and Macken's (1993) manipulation, the irrelevant speech sequence consisted only of the utterance *"ah"* at four different pitches. It has been shown that when a sequence of background speech consists of phonologically similar items, the degree of the disruption produced by irrelevant speech is greatly diminished relative to phonologically dissimilar irrelevant speech. Thus, the repetition of a single syllable, even at different pitches, may have resulted in a weak irrelevant speech effect. Furthermore, the participants may not have interpreted the sound *'ah'* at four different pitches as speech (LeCompte et al., 1997). LeCompte et al. (1997) also suggest that Jones and Macken's (1993) use of a small participant sample may account for their failure to observe a difference between the effect of speech and tones due to low statistical power and so they used a larger sample of participants.

The results of LeCompte et al. (1997) had higher statistical power. Contrary to the results of Jones and Macken (1993), LeCompte et al. (1997) found that an irrelevant sound background consisting of a series of four frequency-changing tones caused less impairment than a random arrangement of the words, 'hey', 'you', 'me', and 'no'. Likewise, it was demonstrated that meaningful speech impaired recall more than did tones or nonsense syllables. They went on to show that meaningless speech (reversed speech) disrupted recall more than did tones, which further emphasises that the critical factor in this speech/non-speech distinction cannot be the semantic content of the words. These findings led these authors to suggest a special role for speech in the ISE.

Investigation of the auditory stimuli used by Jones and Macken (1993) and LeCompte et al. (1997) suggests an alternative account of the differential findings obtained is an account that is in line with the CSH. The words that LeCompte et al. (1997) presented vary in timbre, frequency, envelope characteristics, such as attack (rise time) and decay, and spectral complexity. This contrasts with the tones used by Jones and Macken (1993) that changed in frequency only. In particular the words were dissimilar (non-rhyming), began with different consonant sounds and within the theoretical framework of the CSH demonstrate a higher degree of change than the stream of tones. In contrast, the nonsense syllables were relatively acoustically similar to each other. This assumption can account for the equivalent impairment produced by nonsense syllables and tones (LeCompte et al., 1997). Therefore, the finding that a series of frequency-changing tones impaired serial recall as much as did a series made of four phonologically similar vowels is consistent with the CSH (c.f. Jones and Macken, 1993).

Jones and Macken (1993) and LeCompte et al. (1997) used small stimulus samples (4 items per condition) in their experimental series. Clark (1973) have argued that treating linguistic variables such as words as having fixed effects in analysis of variance is fallacious. Clark (1973) suggest that any effect may be due to attributes of the words used (such as age of acquisition (AOA), frequency, etc) and this must be taken into account, as well as the variability of participant's responses. This means one cannot assume the difference between speech and tones observed by LeCompte et al. (1997) generalises to other stimuli, since the words chosen for the speech conditions were selected from a wider population. Thus Clark (1973) would argue the finding that words disrupt serial recall more than frequency-changing tones (c.f. LeCompte et al., 1997) is not a fixed effect but a random effect. It follows that if another sample of words were chosen, it could be the case that a difference between speech and tones would no longer be observed. However, contrary to Clark's (1973) suggestions, robust irrelevant sound effects have been shown with various sounds in the irrelevant stream, including different words (c.f.

Jones et al., 2000 and Tremblay et al., 2000) and non-speech (c.f. Jones and Macken, 1993; Jones et al., 2000 and Tremblay et al., 2000). The literature shows that the irrelevant sound effect does not depend on the words in the irrelevant stream at all; rather, it is the identification of the items in the irrelevant stream as speech that determines the size of the ISE.

The comparison of two different types of background sound, which change along different physical characteristics emphasises a critical weakness in the CSH. It is not clear whether a prediction of distinct variation along one physical attribute between the repeated components of one stream will result in more or less robust connections, than a stream that varies across a number of dimensions (Divin et al., 2001). Furthermore, LeCompte et al's (1997) research was a direct replication of Jones and Macken (1993). However, the findings revealed that a sequence of changing consonants caused more impairment than four frequency varying tones. It could be argued that the mismatch between a series of consonants occurs along more physical dimensions than do the tones, which change only in frequency. Therefore, a sequence of changing consonants may have been more disruptive than a sequence of tones changing in frequency only because they vary more acoustically.

The disruption of serial recall of graphically presented items by tones of different frequencies that contain no phonological information has been extended to lip-reading (Divin et al., 2001). This finding provides further evidence against encoding occurring on a phonological basis, but supports the CSH. First, the CSH predicts that tones will impair recall performance. Second the two devices suggested by the PSH do not fit with current data. The filter hypothesis of the PSH could explain why tones enter the store since they may be viewed as not noise, but that disruption within the store would occur due to phonological similarity is not consistent with empirical evidence (Jones and Macken, 1995a). Likewise, a speech detector may allow tones entry on the basis of some physical attribute that renders them speech-like. However, an additional non-speech store would be required (Salamé and Baddeley, 1989). Therefore, only the CSH predicts the pattern of impairment, as the only way in which the PSH could account for the data is if tones are recoded phonologically, which is highly implausible or that the underlying representations are not phonological. Divin et al. (2001) found a significant interaction between the two background sound conditions of speech and tones at the last two serial positions. Although small, this difference is not consistent with Jones and Macken's (1993) assumption that speech and tones produce equivalent impairment. However, a critical difference between this study and previous research is that lip-read digits were presented, thus suffix interference needs to be considered (Divin et al., 2001).

It has been suggested that the enhanced recency effect of lip-read and auditory lists is due to phonetic or speech processes occurring in a module peripheral to working memory (c.f. Frankish, 1996). Therefore, even if speech and tones have equivalent memorial effects, differences may occur because the irrelevant speech items, not the tones, enter this speech module. The finding that tones impair recall is incompatible with the assumption that the short-term storage system is speech-based, as can be inferred by the access given to tones.

Tremblay, Macken and Jones (2001) investigated whether the disruptive effect seen at low levels with periodic sounds (speech/tones) can be demonstrated with aperiodic sounds. When irrelevant sound consisted of broadband noise, in which the centre frequency changed with each noise burst, serial recall was substantially impaired. In contrast, a stream of irrelevant sound in which the same band-pass noise burst was repeated did not cause significant disruption. Furthermore, serial recall for both visual-verbal and visual-spatial items was susceptible to the increase in interference caused by changing irrelevant noise. These findings demonstrate a changing-state effect, showing that sounds that are mainly aperiodic can cause significant impairment of serial recall similar to that caused by the presentation of periodic sounds. These results challenge the idea of a memory system limited to the storage of periodic sound information. This system proposes that a filter allows access of tonal items rendering aperiodic sounds, for example noise to be excluded (Salamé and Baddeley, 1989). This study provides further evidence against the phonological store explanation and the feature model that are derived from the assumption that similarity of identity is a primary factor, as no content is shared between the irrelevant noise bursts and either the visual spatial or visual-verbal to-be-recalled stimuli (Tremblay et al., 2001).

2.5 EFFECT OF DEGRADATION ON IRRELEVANT SOUND

Several studies have degraded stimuli by gradually reducing the amount of variation within the irrelevant stream. For example, Ellermeier and Hellbrück (1998) investigated changes in the signal-to-noise ratio (SNR) of the irrelevant sequence, where the degree of change was manipulated by adding more or less uniform pink noise to a speech sequence. The speech sequence consisted of a 15-min recording of a Japanese male speaker reading a text. The participants did not understand Japanese. The findings were consistent with the CSH as when the SNR became smaller, the degree of interference was reduced and thus a monotonic relationship was found between degradation and memory disruption. Further analyses revealed that speech mixed with a low level of noise produced an ISE that was statistically equivalent to that found when only speech was present. Both speech alone and speech with low level noise produced more interference relative to speech mixed with noise at a lower SNR and the control conditions of pink noise alone and silence. When speech was mixed with noise at a lower SNR the detectability of speech was at the absolute hearing threshold and the degree of acoustic change within the speech signal is reduced dramatically compared with speech mixed with low level noise.

The linear decrement in serial recall performance found with speech degraded by the addition of different levels of pink noise has also been observed with non-speech stimuli. In an experiment conducted by Ellermeier and Wolski (1998, cited in Ellermeier and Hellbrück, 1998) Japanese speech sounds used by Ellermeier and Hellbrück (1998) were replaced by sinusoidally frequency-modulated (FM) tones. The same linear decrement in level of disruption was found as different levels of pink uniform noise were mixed with the tonal signal.

It can be assumed that sequences of noise-masked stimuli have less prominent acoustic features, and therefore the degree of variation is reduced gradually as the signal-to-noise ratio is reduced. Therefore, some of the items would be more susceptible to masking than others, resulting in the signal-to-noise ratio being reduced. These tokens would be reduced to a level below audibility within the noise, which would result in fewer tokens being perceived. Therefore, adding noise reduces the *number of tokens* rather than the *relative degree of change*.

Jones et al. (2000) demonstrate a clear relationship between the degree of change within the irrelevant sequence and interference. This was found by degrading a stream of words spoken by the same voice by low-pass filtering. The rate of roll-off of the filter was manipulated, which acted to progressively attenuate the frequencies above the fundamental. The sound was never totally obliterated and therefore, the token dose remained constant at all degrees of roll-off. To-be-recalled lists consisted of the random arrangement of nine letters. As the degree of filtering was increased more acoustic attributes of the sound stream were removed. That is the difference between irrelevant sounds would be reduced. Eventually the stream heard was not intelligible. As the degree of filtering was increased, there was a monotonic improvement in performance. This reveals the continuous nature of changing-state stimuli, as there was no evidence of a threshold above which disruption was more pronounced. These results are consistent with the notion that the ISE is not determined by the phonological identity of the sounds. If the occurrence of the ISE was dependent on the phonological identity of the sounds, a discontinuous function would have been observed in the relationship between the level of stimulus degradation and the level of memory interference. In other words, the observed linear pattern of interference indicates that a particular level of signal degradation did not have to be reached to obtain an ISE.

Jones et al (2000) point out that the low-pass filtering technique is problematic. First, it is difficult to apply when contrasting stimuli that differ in complexity. For example, when contrasting the effects of degradation of a speech sequence with that of a sequence of instrumental tones, substantially different degrees of roll-off for each auditory sequence would have to be applied. This is because the distribution of energy across the auditory spectrum is narrower in a typical musical instrument. Auditory stimuli would need to have the same fundamental frequency, which would not be the case when contrasting speech and non-speech (e.g. tones). Different ranges of levels of roll-off may not produce the same effects with each stimulus (Jones et al., 2000). Jones et al. (2000) responded to these disadvantages by using a common metric. Digital signal processing was used to digitally sample speech and cello notes. The polarity of each of the sample points making up the stimulus was reversed with a certain probability. By systematically changing the likelihood of reversal a range of degraded speech and cello notes were generated and at one point the stimuli consisted of amplitude-modulated noise. A linear effect was observed; as the stimulus degradation was reduced, disruption of memory by background sound increased. This finding replicates the linear relationship observed between degradation and serial recall disruption when irrelevant speech sounds were progressively low-pass filtered. It also provides further evidence supporting the notion that categorical processing of lexical, sub-lexical and phonological items cannot explain the ISE due to there being no discontinuity in the effect of degradation on disruption when changing speech sounds form the irrelevant stream. These findings are consistent with the notion that the ISE is not a function of the semantic, lexical or phonological identity of the sounds.

Further evidence in support of the claim that irrelevant speech and non-speech differ in their effect on serial memory is provided by the finding that fully degraded speech was more disruptive than fully degraded cello notes. One explanation is that fully degraded speech tokens still have some spectral changes from item-to-item because of different amplitude envelopes associated with the different speech tokens. Although levels of performance differed for speech and cello notes, a linear relationship between degradation and disruption was observed for both classes of sound. That is, as speech sounds and cello notes are disrupted, memory disruption is reduced. This is evidence that although speech is found to be more disruptive of serial recall (e.g. LeCompte et al., 1997; Jones et al., 2000 and Tremblay et al., 2000) speech and non-speech are functionally similar in producing disruption of serial recall. Other evidence for the functional equivalence of speech and non-speech is the fact that the same relationship between token-set size and disruption is found for speech and non-speech (Tremblay and Jones, 1998).

The evidence that speech holds no special status within the ISE paradigm matches the predictions of the CSH. Rather, perceptual variations between distinct and segmentable auditory items are crucial for the appearance of the ISE. It can be assumed that the effect of degrading the sounds within a stream, either by lowering the SNR or reversing the polarity of sample points at varying degrees, acts to remove the time-varying features of the sound. This is argued to be the basis for the removal of the ISE as opposed to a loss in segmentation of the auditory items. A loss of segmentation account is refuted by the finding of an ISE in the presence of a continuous sequence of vowels, linked by smooth formant transitions, where no pauses featured in the signal.

One generalisation that has arisen from these studies is that the degree of change seems to determine the manner in which the brain automatically processes information about the order of events. The relationship between distinctiveness and memory for seriation has been found to be non-monotonic. When sounds produced by the same objects (e.g., a voice) become increasingly different; seriation information is enhanced, whereas consistent with an explanation that when acoustic change is very great, the events are streamed into separate objects, (e.g., different voices), seriation information is diminished (Jones et al., 1999b). The assumption is that incoherent streams consist of unconnected 'objects'. Therefore, although the objects are intelligible, information about their order is relatively reduced, and so their effect on serial recall is relatively small. For example, in contrasting the disruptive effects of indistinct vowels produced by the same voice in a monotone with that of a stream of different musical instruments, the vowel series is more disruptive. Thus, it could be argued that speech may indeed be special in some way because it produces disruption even when there are relatively small variations within the irrelevant stream. Furthermore, modest variations on a common fundamental produce more information about order than do sound sequences from very different sources (Jones, 1999).

A slightly more objective approach was adopted by Tremblay et al (2000). The disruptive effect of speech was compared with that obtained by two sine-wave speech conditions. Sine-wave speech is an ambiguous stimulus consisting of a number of sinusoids that resemble speech formant features. Sine-wave speech has been used to provide a more objective test of the speech/non-speech distinction and the CSH. Sine-wave stimuli exclude some of the spectro-temporal acoustic attributes of natural speech, but maintain the global pattern of the first three formants over time (Remez, Rubin, Pisoni and Carell, 1981; Remez and Rubin, 1990). Untrained listeners typically indicate that the sine-wave stimuli sound like electronically generated sounds, or music, but when made aware of the 'speech-likeness' of sine-wave stimuli, listeners find intelligibility to be good even though sine-wave speech is like a sketch of speech, with significantly less information than natural speech (Tremblay et al., 2000). This shows how top-down knowledge processed by experience-driven mechanisms guides and changes the perceptual experience. One condition had participants that were trained to perceive the sine-wave speech as speech, whereas in the other condition, listeners were unaware of its 'speech-likeness' (Tremblay et al., 2000). The findings revealed that there was no significant difference between the two sine-wave speech conditions and that natural speech was significantly more disruptive than either of the two sine-wave speech conditions. This finding refutes the finding of no difference between the magnitude of disruption produced by changing speech and frequency-changing tones (c.f. Jones and Macken, 1993), but is consistent with LeCompte et al's. (1997) observation of a significant difference between the effect of speech and tones.

Tremblay et al (2000) suggest the greater disruption of serial recall by speech relative to sine-wave speech is due to the greater acoustic complexity of the natural speech signal, an explanation in favour with the CSH. More elements change between items in natural speech than in the relatively acoustically simpler sine-wave speech. Natural speech is still more complex than sine-wave speech and because sine-wave speech perceived as speech is not as disruptive as natural speech, this suggests a distinctive perceptual attribute which is more prominent in natural speech, such as timbre, or acoustic energy.

In addition, sine-wave stimuli heard as speech was numerically more disruptive than sine-wave stimuli perceived as non-speech. Tremblay et al (2000) argue that this highlights the potential for a small role of top-down knowledge, which in this case is the awareness of the 'speech-likeness' of the irrelevant sound and therefore, over-learned acoustic properties inherent within the signal. The familiarity with the speech signal may act to mediate the power of the irrelevant sound to disrupt immediate memory for order. Tremblay et al (2000) argue that if top-down processing from familiar variables within speech accounted for its greater disruptive power an observed statistical difference between sine-wave stimuli perceived as speech and non-speech would be expected. Several studies have reflected the importance of bottom-up knowledge, such as primitive streaming and acoustic changes as opposed to semantic attributes in producing memory interference (e.g. Jones and Macken, 1995b, Jones et al., 1999a; Jones et al 1999b; Jones et al., 2000). More specifically, top-down or context dependent processing may have a larger role in the disruption of cognitive tasks of a higher order than memory for order (Tremblay et al., 2000). This is evident when semantic processes feature in a memory task, where meaning is manipulated and affects the magnitude of the ISE. For instance, irrelevant speech presented in a participant's native language produces no more serial recall disruption than irrelevant speech presented in a language that is unfamiliar to participants (Jones et al., 1990). However, a semantic effect is observed during a free recall task for words grouped in semantic categories (Neely and LeCompte, 1999).

The emotional valence of distractor words has been found to increase the effect of irrelevant speech on serial recall (Buchner et al, 2004 and Buchner et al., 2006). However, experiments that have manipulated meaningfulness by, for example, comparing the disruptive effect of irrelevant speech in the participant's native language with a foreign language have found no differences in serial recall disruption (e.g. Jones et al., 1990 and Salamé and Baddeley, 1982). Experiments that have manipulated the effect of the meaningfulness of sounds in the irrelevant stream differ from those that have manipulated the emotional valence of the words in the irrelevant stream. Valent words signal behavioural demands and thus give information about the environment that needs to be attended to (Buchner et al., 2006), whereas words that are meaningful in the sense that they are in a participant's native language as opposed to a foreign language, for example, do not give such environmental information.

If, when speech and non-speech stimuli are matched for spectral complexity, speech disrupts serial recall more than non-speech then this may indicate that some form of top-down processing is involved during pre-attentive processing of irrelevant speech. The presence of supposedly unattended sounds that are speech-like may have the additive effect of acting to signal possible behavioural demands in the environment, and consequently may attract attention. This seems plausible because the effect of speech is moderated by how emotionally valent it is, as negative words or non-words are more disruptive than positive words, which again are more disruptive than neutral words (Buchner et al, 2004; Buchner et al., 2006). Moreover, linguistic knowledge of speech sounds held in long-term memory may interact with pre-attentively processed speech sounds in STM in order to form more stable and integral mental representations (Buchner and Erdfelder, 2005; Schweickert, 1993). If this is the case, the interaction between STM and LTM may provide more cues to serial order, rendering speech more disruptive than non-speech stimuli.

Advocates of the CSH hypothesis hold that speech is more complex than non-speech sounds such as tones due to the rapid spectral variation in speech (e.g., Jones et al., 2000). The critical issue here is the identity of the perceptual attribute that needs to be changed in speech, and whether this attribute is more detailed and variable within the speech signal. Spectral variation over time is an acoustic feature required for speech comprehension, however, acoustically transformed speech can be understood (e.g. Shannon, et al. 1995). This emphasises the redundancy inherent in an untransformed clear speech signal.

2.6 THE EFFECT OF VOWEL AND CONSONANT CHANGES IN THE IRRELEVANT STREAM

Hughes, Tremblay and Jones (2005) using Consonant-Vowel-Consonant (CVC) syllables as irrelevant speech items investigated the disruptive power of changes only in the final consonant, initial consonant, or in the vowel of each item (Hughes et al., 2005). The findings indicate that vowel changes, as opposed to consonant changes are the dominant source of disruption, an effect referred to as the 'vowel-changing-state effect'. Furthermore, it was found that consonant variations do not cause more impairment than a steady-state irrelevant sequence and that vowels must change-in-state from token-to-token. The observation that vowels must change from one token to the next and are the important source of disruption is problematic for the CSH, which claims any acoustical change between mismatching tokens should cause more disruption than a repeating token (e.g., Jones et al., 1992).

Jones et al. (1999a) and Jones et al. (1999b) argue that the primary variable is not any form of acoustical change, but acoustical change carried on an attribute common to the sounds in an irrelevant stream. The suggestion that only variations on a shared quality cause substantial interference is plausible if it is thought of along with the phenomena of auditory perceptual streaming (Jones et al., 1999a; 1999b). When all the distinct sounds in an irrelevant stream are presented to both ears a changing-state effect is observed, as the sounds are perceived as forming a coherent changing stream. However, when an unattended auditory stream of changing-state items is presented, so that the different items traverse different spatial locations, the changing-state effect is diminished (Jones and Macken, 1995b; Jones et al., 1999b).

Research demonstrates that great difficulty arises when attempting to identify the order of auditory objects if those objects are perceived as originating from different auditory sources (Bregman, 1990). Furthermore, variations between temporally successive sounds that traverse separate perceptual streams, such as sounds presented from widely spaced frequency bands (Jones et al., 1999a) will produce little if any detail about serial order causing serial recall performance to remain relatively immune to the irrelevant sounds. This is because the irrelevant sounds would be perceived as separate streams of steady-state (repeated) sounds. The idea that it is acoustical change occurring on a shared fundamental that is important for the changing-state effect to occur provides an explanation for the difference between the effects of consonant and vowel variations.

Research on the perceptual organisation of speech has led to the suggestion that the perceptual integration of different speech utterances spoken by the same voice over time is based on a continuity in some percept (e.g., timbre) shared by the periodic (voiced) vowel sounds. However, unvoiced consonants make up the aperiodic, noisy onsets and offsets of these periodic utterances (c.f. Bregman, 1990). As a consequence, it is argued that changes between mismatching vowels produced by different vocal tract shapes which are carried on this common attribute (e.g. fundamental frequency ($f0$) and formant structure) may elicit serial order cues and are therefore more likely to interrupt recall than changes between consonants (Hughes et al., 2005).

A difference in spectral complexity between vowels and consonants may account for the vowel-changing-state effect (Hughes et al., 2005). The assumption that vowel changes are an important factor in producing a changing-state effect because vowel changes are more likely to provide serial order cues is further supported by research investigating the serial recall of items which feature changes in consonants only. A variety of experiments have demonstrated that vowel-only-changing items are better recalled in order than consonant-only-changing tokens (e.g., Surprenant and Neath, 1996). One indirect inference that has been drawn from these findings is that attended-to items that can be easily recalled in serial order are more likely to be more disruptive when implemented as irrelevant sound (Hughes et al., 2005). This indirect assumption is in line with the disruption by seriation processes account of the ISE (e.g., Jones et al., 2000).

In the light of the contradictory evidence in regard to the speech/non-speech distinction it is argued that in acoustic terms even simple speech sounds are typically more complex than non-speech sounds. Jones et al (2000) demonstrated a linear relationship between degree of change within the sound sequence and its disruptive potency for speech and non-speech. However, the exact mechanism of this effect is not clearly defined and few studies make a direct comparison of the interference caused by speech versus non-speech. Moreover, there has been almost no systematic investigation of the possible significance of acoustic cues distinguishing speech and non-speech sounds. What studies have done is systematically degraded natural speech and non-speech (e.g. cello notes) stimuli by various degrees, and compared the relationship between degradation and serial recall interference for both classes of stimuli (e.g. Jones et al., 2000). However, research has not degraded speech in order to examine what feature or features within its signal render it more disruptive than non-speech sounds.

2.7 AIMS OF PRESENT RESEARCH AND PREDICTIONS BY EXISTING HYPOTHESES

The present experimental series looks at which characteristic or characteristics of speech accounts for its ability to disrupt serial recall more than non-speech. Irrelevant speech will be degraded and distorted in order to manipulate its 'speech-likeness' in ways that allow systematic control of phonemic content and changing-state acoustic cues in speech. This will enable the investigation into what attribute(s) inherent within the speech signal is responsible for its ability to disrupt memory more than non-speech stimuli (c.f. Jones et al., 2000; LeCompte et al., 1997 and Tremblay et al., 2000). By investigating the attribute(s) in speech which is responsible for its disruptive power and then contrasting the effect of speech and non-speech on memory, the research aims to evaluate the explanatory power of the changing-state hypothesis (CSH) (Jones et al., 1992) as well as examining explanations put forward by other models of the ISE.

The research will explore the effects of degrading the intelligibility and acoustic complexity of speech, as well as reducing intelligibility whilst preserving its acoustic complexity on the degree of memory disruption from irrelevant speech. It will also examine the effect maintaining the intelligibility of speech whilst reducing its acoustic detail has on the size of the ISE. The phonological store hypothesis (PSH) would predict that speech, even when degraded or distorted would have privileged access to the phonological store and thus create codes that would interfere with the representations of the visual to-be–remembered (TBR) items, so long as its phonemic content is intelligible. In contrast, the CSH would predict that speech would produce more disruption than degraded speech, because degrading speech reduces the number and extent of acoustic changes in the signal. As non-speech stimuli (e.g. tones) are argued to be acoustically less complex than speech stimuli (e.g. Tremblay et al., 2000), when non-speech and speech are matched for acoustic complexity the CSH would predict equivalent levels of disruption by both classes of stimuli. If, when the acoustic complexity of speech and non-speech is matched speech is found to be more disruptive, this would point to a perceptual mechanism specialised for speech. This would be in support of the PSH, in so far as it suggests a greater disruptive effect of speech.

CHAPTER 3

3 GENERAL METHODOLOGICAL APPROACH

3.1 BACKGROUND

3.1.1 Memory tasks

The characteristics of a memory task determine its sensitivity to disruption by irrelevant sound. Research converges on the important role seriation in a memory task has in determining the degree of disruption (LeCompte, 1996; Beaman and Jones, 1997; 1998). As discussed in chapter 1, tasks for which the predominant mnemonic strategy drawn on by participants is rote rehearsal are sensitive to the effects of irrelevant sound, whereas tasks which do not draw on serial memory are markedly less susceptible. For example, during the missing-item task, participants are presented with a list of items, such as months of the year or a series of digits (Beaman and Jones, 1997; Jones and Macken, 1993). After the stimulus list is presented, participants are required to recall the month that was missing from the list. A small effect of irrelevant sound has been found with a missing-item task with speech (Beaman and Jones, 1997) and irrelevant tones (Jones and Macken, 1993).

The strategy adopted by participants can be constrained, influencing whether or not rote rehearsal is used as a mnemonic strategy. Beaman and Jones (1997) familiarised participants with the stimulus set from which names of religious buildings were taken to form the experimental lists in a missing item task. Half the participants were presented with the stimulus set in a fixed-order, with the items presented in alphabetical order. This meant that participants could draw on a fixed-order representation of the items in memory from which to retrieve the name missing from the stimulus lists by mentally excluding each item in a list as the list was presented. The other half saw the set in random order and so at recall they could not rely on a fixed-order memory representation. When the strategy drawn on during a missing-item task was constrained to that of 'checking-off', that is mentally checking-off each item in the stimulus list as the list is presented, no effect of irrelevant sound on memory performance was found. In contrast, when no fixed-ordered representation of the items was available in memory, irrelevant sound disrupted recall performance reliably. Beaman and Jones (1997) argue that participants may have had to draw on a temporary representation of the order of items by rote rehearsing items.

Other tasks, such as free recall (recalling items in any order) and recognition tasks (participants have to report which of two words presented featured in the experimental list for each trial) to name two are susceptible to the deleterious effects of irrelevant sound (LeCompte et al., 1996; Beaman and Jones, 1997; 1998). Although it is item information that is required by these tasks and not recall of their serial order, their susceptibility to irrelevant sound has been shown to be determined by the degree to which serial rehearsal is the dominant strategy relied on by participants (Beaman and Jones, 1997). Further, irrelevant sound interferes with item and order information in free recall as in serial recall. The finding that an equivalent degree of order information is retained for both free and serial recall tasks indicates that order information is important in recalling item information (Beaman and Jones, 1998). The importance of the serial recall component of a task is consistent with the primary assumption of the CSH. The CSH assumes the effect of irrelevant sound is on serial order cues as the automatic obligatory serial processing of irrelevant auditory items is argued to conflict with the serial rehearsal of the TBR items in short-term memory (STM).

3.1.2 Stimuli

3.1.2.1 Auditory stimuli

The present experimental series investigates the effect of different manipulations of the speech signal in order to examine which characteristic(s) afford speech to disrupt serial recall more than non-speech. As discussed in chapter 1, research indicates no effect of meaning on the size of the ISE. For example, no difference in the degree of disruption between words and nonsense-words, such as reversed words has been found and speech in a language foreign to participants is as disruptive of memory as speech in participant's native language (Colle and Welsh, 1976; Jones et al., 1990; Salamé and Baddeley, 1982). Although the semantic content of items in the irrelevant stream has been found not influence the size of the ISE, non-words will be used as auditory stimuli in the current experiments in order to isolate their phonemic content. This ensured any difference between speech and distorted speech would result from the experimental manipulation and not be confounded by any possible small effect of meaning. LeCompte et al. (1997) found a small but significant difference between the degree of disruption produced by words and nonsense syllables. However, in a subsequent experiment this difference did not reach significance, although the small effect size for the difference between words and nonsense (reversed) words was comparable to that found in the initial experiment when speech and nonsense syllables were compared (LeCompte et al., 1997).

3.1.2.2 Visual stimuli

In previous research, digits or letters have featured as TBR items in the visual stream (e.g. Jones et al., 2000; Jones and Macken 1995c; Beaman and Holt, 2007). The visual TBR items of the present experiment will consist of lists of random permutations of the digits from a digit set. There is no evidence in the literature to suggest that having TBR items and irrelevant items that are of the same class makes a difference to the size of the ISE. It follows that presenting TBR digits rather than letters should make no difference in terms of the sensitivity of item rehearsal to the disruptive effect of irrelevant sound during seriation.

3.2 AIMS AND OBJECTIVES

Evidence surrounding the controversy of whether or not speech is special shows that the ISE is not only constrained to speech (Jones et al., 2000; Jones and Macken, 1993; LeCompte et al., 1997). However, what these different findings cannot account for is what it is about the speech signal that renders it more disruptive (Jones et al., 2000; LeCompte et al., 1997). Few studies directly compare the interference caused by speech with that caused by non-speech sounds, and there have been few systematic investigations of the possible significance of acoustic cues that distinguish speech from non-speech sounds. This thesis investigates the explanatory power of the changing-state-hypothesis (CSH) in that it examines whether it is the nature of the irrelevant sounds or how acoustically changeable they are that determines the size of the ISE.

The experimental series reported in this thesis involves the manipulation of irrelevant sound with regard to the acoustic complexity of the signal and the intelligibility of phonemic information to determine how 'speech-like' the sound has to be to produce the magnitude of memory interference observed with clear untransformed speech. By manipulating speech in ways that allow systematic control of phonemic content and changing state acoustic cues, the research will examine whether in terms of the ISE 'speech is special', and if so, why? Natural speech is degraded to various degrees using various signal processing techniques in order to investigate what aspect(s) of the complex speech signal need to be reduced or removed in order to diminish the disruptive potency of the speech in the irrelevant stream. Moreover, the research investigates whether phonemic content, in terms of perceiving speech as speech, is the key characteristic of speech that causes the higher level of disruption in serial recall observed with speech relative to non-speech.

3.3 STIMULI

3.3.1 Auditory stimuli

Monosyllabic non-words were recorded digitally with 16 bit resolution at a sampling rate of 22050Hz, in a female voice with neutral intonation. The non-words were recorded directly onto a re-writeable compact disc and then moved digitally into a Pentium class PC for editing with Cool Edit Pro 1.2 (Syntrillium Software Corporation). The recording was then high pass filtered at 50 Hz, in order to minimise low frequency distortion. The sounds were broadly similar in loudness and free from background noise. Research has demonstrated that variations in both the degree of change in intensity and the overall intensity of irrelevant sounds have no effect on the level of immediate memory disruption (Salamé and Baddeley, 1989; Tremblay and Jones, 1999). The non-words were separated into individual sound files. The non-words in each sound condition were made of different phonemes so that maximum acoustic variation was demonstrated between stimuli. Each non-word in the irrelevant speech conditions was presented for 1000ms and was separated by an inter-stimulus interval (ISI) of 1000ms. During the silent condition no sound was played through the headphones. The relevant non-words recorded were subjected to different methods of signal manipulation, degrading or distorting different aspects of the signal. Table 1 lists the experimental manipulations performed in each experiment. More details regarding the experimental manipulations and the non-words used will be given in the appropriate chapters.

	Experimental Manipulation
Experiment 1	Degrading speech (signal correlated noise (SCN)).
Experiment 2	Degrading (SCN) versions of vowel-only-changing (V-O-C) and consonant-only-changing speech (C-O-C).
Experiment 3	Whispers.
Experiment 4	Alternating voiced and whispered speech.
Experiment 5	Reversal of the fine structure of whispers.
Experiment 6	Spectral rotation of speech.
Experiment 7	Spectral rotation of speech.

Table 1. Experimental manipulations performed in each experiment.

3.3.2 Visual Stimuli

Lists of digits to be recalled were presented serially on a visual display unit (VDU). Lists were constructed from the random arrangement of the 7 digits from the digit set 1-7, using a Latin square design, so that for each condition, each digit appeared equally often in each serial position and no runs of more than 2 digits in ascending or descending order were present (appendix 1). Each digit in a trial was presented in succession. Digits were displayed in the centre of a VDU in new courier font at size 20. Each digit was presented for 1000ms and was separated by an ISI also of 1000ms.

3.4 DESIGN

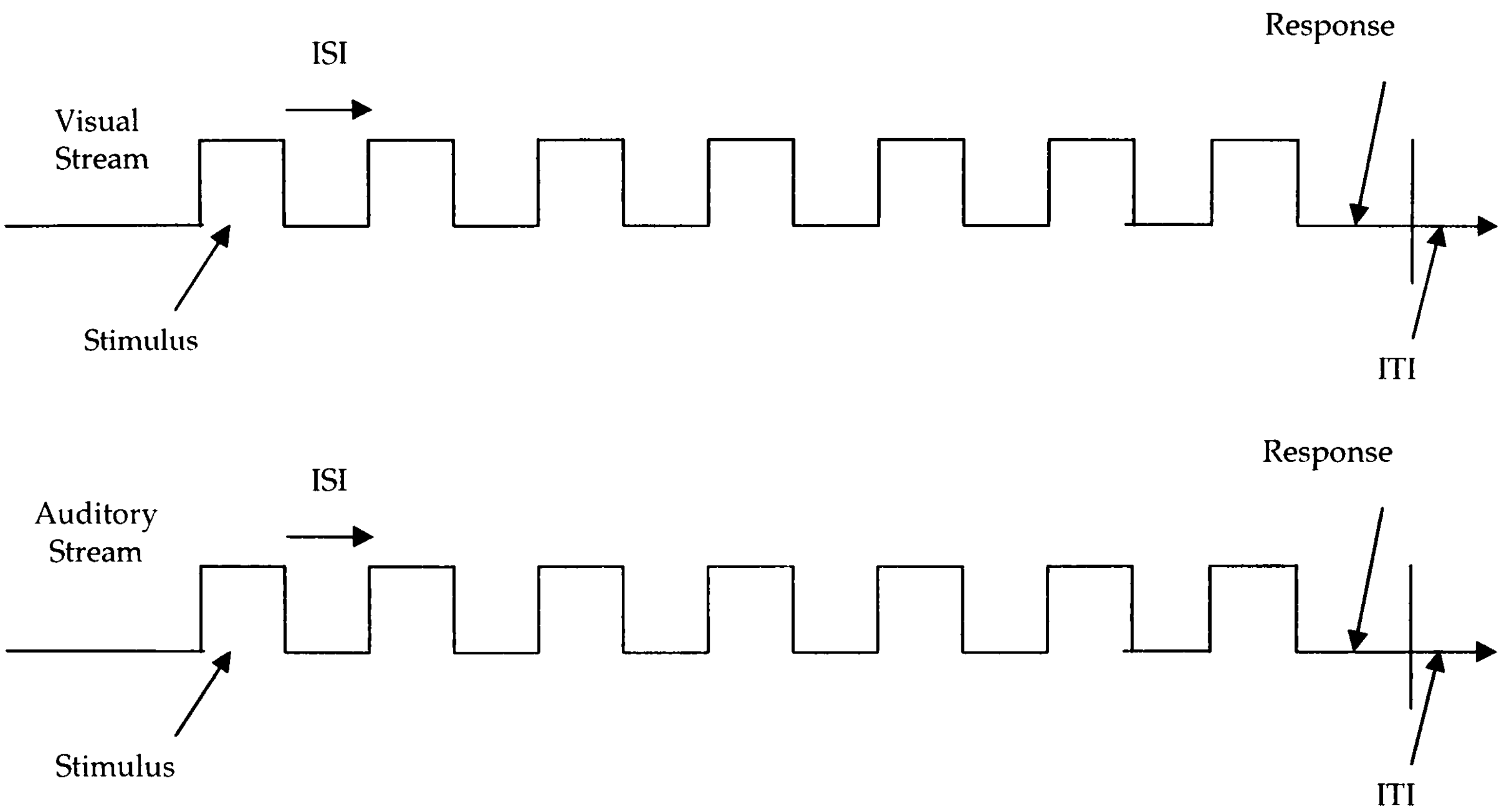
A repeated measures design was used with all participants undertaking the serial recall task under all auditory conditions. Participants performed the serial recall tasks for 84 trials in total, 28 for each of the three auditory conditions. The presentation order of conditions was fully counterbalanced such that the condition order for each participant was selected from the set of six possible orders without replacement. The same number of participants performed the serial recall task in each of the six possible condition orders. As 30 participants were tested, this was achieved by randomly assigning participants to one of five groups of six participants.

It could be argued that whilst a fully counterbalanced block design eliminates order effects, it may have implications for strategy use. For example, participants may concentrate on the first few digits and the final TBR digit presented, ignoring the digits presented at intermediate positions. If adopted more often in the irrelevant sound conditions, an irrelevant sound-specific strategy shift would result in any difference between the quiet and irrelevant sound conditions being confounded by such a strategy shift. Serial recall performance as a measure of memory disruption would, in this case, be less sensitive to differences between the effects of the irrelevant sound conditions. It could be argued that randomising the order of conditions so that each TBR digit sequence is assigned to one of the three conditions at random would eliminate the possibility of participants developing a strategy when learning TBR item lists, as participants would not know what sound condition would be presented next. However, in a study by Buchner et al. (2004) a succession of sequences of TBR words from 6 different sound conditions was presented at random and yet the use of a memory strategy by participants was reported. Buchner et al. (2004) in their study on the affect of distractor valency on serial recall disruption found that participant's often concentrated on the first, second, third and final TBR words presented. This strategy was adopted more often in the irrelevant sound conditions so that serial recall performance was found to be poorer compared to the quiet control condition. Six item lists were presented comprising TBR three syllable words. Buchner et al. (2006) suggest that presenting long lists of visual TBR items leads participants to adopt such a memory strategy when learning TBR sequences is difficult in the presence of irrelevant sound. Buchner et al. (2006) state that strategy shifts are not expected for short TBR sequences. Lists of seven TBR digits are much shorter than the six item lists comprising three syllable TBR words employed by Buchner et al. (2004). In addition, experiments examining the ISE have used a fully counterbalanced block design and no evidence of strategy use has been reported (e.g. Beaman and Holt, 2007 and Perham, Banbury, and Jones, in press).

3.5 MEMORY TASK: GENERAL PROCEDURAL OUTLINE

Participants were tested individually, seated in a quiet room approximately 0.5m from the computer screen. Standard instructions were given to participants and presented on VDU (appendix 2). Standard instructions were given verbally to participants before the experiment began. Instructions were also displayed on a VDU at the outset of each experimental condition. These informed them of the nature of the recall task, asked them to ignore any sounds they might hear, and reassured them that the content of the auditory items would not be tested and thus were irrelevant to the memory task. During each trial, the seven digits were presented in random order as described above.

The experiment generating software E-Prime (Psychology Software Tools INC, 2002) was used to present each digit for 1000ms with an inter-stimulus interval (ISI) also of 1000ms. The irrelevant sounds were synchronised with the onset of each digit. See schematic diagram of a trial in figure 4.



* ISI = inter-stimulus interval = 1000ms. Stimulus duration = 1000ms.

* ITI = inter-trial interval. During response period the instruction to "hit space bar for next trial" was presented. Transition through trials was under participant control.

Figure 4. Schematic diagram of a trial. 7 digits were presented per trial. Sounds were presented with the onset of each sound synchronised with the onset of a visual digit.

Sound was presented digitally over Sennheiser HD570 headphones through a Creative Extigy digital sound card connected to the PC. During the silent condition no sound was played through the headphones. Immediately after each digit list had been displayed, participants were required to recall the list in strict serial order upon the appearance of the prompt 'recall'. Responses were written on a blank grid comprising rows of seven boxes corresponding in order to the To-Be-

Remembered (TBR) digit sequences. Participants were required to recall each digit list from left to right. If participants could not remember a digit they were asked to guess before moving onto the next digit, so as not to omit any responses. They were instructed not to check responses and correct them even if they believed a digit to be incorrectly recalled in a serial position. Experimental trials were preceded by a short practice session of 3 trials that were not used in any analyses. The experiment lasted approximately 40 minutes. The irrelevant sounds and digit lists were presented during trials using the experiment generating software E-Prime (Psychology Software Tools, INC, 2002). At the end of each trial participants were instructed to move onto the next trial on seeing 'push space bar for next trial' on the VDU and therefore they controlled the speed at which each trial was presented. Participants completed a consent form before the experiment (appendix 3) and were fully debriefed at the end. Some of the experiments undertook involved variations on the general design and procedure outlined. These will be detailed in the methodological considerations section for the experiments in the following chapters as will the various signal manipulation techniques.

CHAPTER 4

4 PRELIMINARY INVESTIGATION INTO THE EFFECT OF PHONOLOGICAL DEGRADATION ON THE DEGREE OF SERIAL RECALL DISRUPTION FROM IRRELEVANT SPEECH

4.1 BACKGROUND

Past research (e.g. Jones et al., 2000) shows that speech and non-speech sounds produce a linear relationship between degradation and serial recall performance. That is, as the degree of degradation of speech and non-speech sounds is increased, the number of serial recall errors decreases. It is therefore clear that no threshold needs to be exceeded for digital degradation of the signal to have a noticeable effect on memory for order information, since a discontinuity in the degradation by disruption function would needed to have been observed. However, fully degraded speech is still more disruptive of memory than fully degraded non-speech sounds (Jones et al., 2000). Thus, it can be argued that degradation acts to remove or distort a phonetic-acoustic feature of speech which affords the observed increase in memory interference upon its presentation.

The experiment reported here explores the effect of phonological degradation on the degree of disruption from irrelevant speech. If an acoustic-phonetic feature is the key characteristic of speech that causes the marked level of disruption in serial recall, then acoustic degradation of speech, which will reduce the degree of acoustic variation, should reduce the intelligibility of the speech along with the degree of disruption. That is, clear speech will produce more serial recall interference than degraded speech because phonological degradation of the signal not only acts to attenuate the nature and extent of acoustic changeability within an irrelevant auditory stream but diminishes the intelligibility of the phonemes within the non-word samples. A reliable difference is predicted between serial recall performance in the presence of degraded speech and silence as degraded speech items will still be perceived as speech and adjacent items will still vary acoustically.

The present experiment aims to generate a reliable difference between the number of serial recall errors produced under clear and degraded speech. The degraded speech will be examined to see whether there are differences in the intelligibility of vowels and consonants. Hughes et al. (2005) found evidence that varying vowels within an unattended irrelevant sequence determine memory interference more than do changes in consonants between successive auditory items. vowel-only-changing (V-O-C) speech items disrupted memory significantly more than consonant-only-changing (C-O-C) speech items. Whether or not vowel identification within the degraded speech stream is affected by digital manipulation will be examined.

4.2 PILOT A (FOR EXPERIMENT 1A). PERCEPTUAL IDENTIFICATION TASK: INTELLIGIBILITY SCREENING OF NON-WORD SAMPLES

4.2.1 Participants

A group of 23 undergraduate students volunteered to take part in the experiment, each reporting normal or corrected-to-normal vision and normal hearing. All participants had English as their first language.

4.2.2. Signal processing

The 50 non-words (appendix 4, see also appendix 5 for examples of disc phonetic symbols) were recorded and edited as detailed in chapter 3 (p 90). The non-words were then digitally degraded using custom software by randomising a percentage of the samples in each sound, therefore turning a proportion of the signal into signal correlated noise (SCN). First, the polarity of samples within the waveforms of each non-word was reversed at a signal-to-noise ratio (SNR) of 0.65. Therefore, 65% of the speech signal was unchanged, leaving 35% of the signal replaced with random samples. The same 50 non-words were also manipulated at a SNR of 0.7. Here, 70% of the speech signal was left unmanipulated, with 30% of the signal replaced with random samples. Thus, two versions of the 50 non-words were generated. As the percentage of sample points whose polarity is reversed increases, so does the proportion of noise relative to the signal increase. The amplitude envelope of the sample remained the same. These two signal-to-noise ratios were chosen so that the intelligibility of the non-words would be reduced, but not to the extent that the non-words would be heard as noise. If there was more noise relative to the signal, this might have destroyed the intelligibility of some of the phonemes within the degraded non-words so that they were no longer perceived, rather than their identification being distorted. The aim of this experiment is to examine whether there are differences in the identification of vowels and consonants within the degraded non-words. Two SNRs were employed as seven distinct low intelligible non-words needed to be selected for the memory experiment (pilot B, for experiment 1). The SNR needed to afford the isolation of seven distinct non-words where the consonants and vowels between adjacent sounds differed. The SNR that produced a range of high to low intelligible non-words that was more distributed was selected.

4.2.3 Design

A repeated measures experimental design was used for this pilot study. All participants completed a perceptual identification task.

4.2.4 Procedure

A group of participants were tested in a free-field situation in a seminar room. Stimuli were projected using Microsoft PowerPoint software. Before the experimental trials were presented, standard instructions were given to the participants and provided on a screen informing them of what the experiment involved (appendix 6). The degraded non-words were presented over a single speaker and each non-word was played once to participants. A speaker icon was presented on screen symbolising each non-word to be played. Each speaker icon was clicked by the experimenter to present each degraded non-word. Immediately after the presentation of each non-word the response cue 'respond' was displayed and participants were instructed to write down phonetically what they thought they heard on a response sheet, used for scoring in a later session. That is, participants wrote down the word/non-word they thought they heard. The inter-stimulus interval (ISI) was controlled by participants, in that when all participants had finished writing down the sound they thought they heard the next sound was presented. Six practice sounds were given before participants proceeded to the experimental sounds, after which any questions were answered. These practice trials were not used in any analysis. After the first 50 non-words were presented, participants were given a five minute break. The intelligibility screening sessions lasted approximately 20 minutes. Participants completed a consent form (appendix 7) before the experiment and were fully debriefed at the end.

4.2.5 Results

The responses were scored correct if they corresponded phonetically to the non-word presented. For instance, when the non-word /muj/ (mV_) was presented (see appendix 5 for examples of phonetic symbols), if participants wrote "mudge" or "muj" the response was scored as correct. The non-words were then ranked in terms of the number of correct identifications. Non-words whose written response was questioned were judged by an independent committee, blind to the experimental predictions, and there was 100% agreement. This acted as a measure of intelligibility, as a range of low to highly intelligible non-words was generated (appendix 8). Intelligibility here referred to the identifiability of the degraded non-words. For the 50 non-words degraded at a SNR of 0.65 the intelligibility range went from 0-22 correct identifications. The total number of correct identifications was 461. This figure was calculated by adding the number of correct identifications for each of the 50 non-words. In contrast the intelligibility range for the same non-words at a SNR of 0.7 ranged from 0-21 correct identifications and the total number of correct identifications including all 50 non-words was 494. It was from the intelligibility range of the non-words at 0.7 SNR that seven non-words of low intelligibility were selected for the irrelevant degraded speech condition of pilot B (for experiment 1b).

Non-words that were low in intelligibility, with recognition scores in the range of 0-8 correct responses made, were isolated from the bottom quartile of the intelligibility range. Seven non-words for the clear (un-degraded) speech condition that differed from those making up the degraded speech condition where selected from the recorded non-words (appendix 9; see also appendix 5 for examples of disc phonetic symbols). Different non-words were used in the two sound conditions because the literature shows that phonological content does not determine the size of the ISE. For example, Jones et al. (1990) demonstrated that irrelevant speech in a language native to participants produces no more disruption than a foreign language. Also, words played forwards and reversed words were found to disrupt serial recall to the same extent (Jones et al., 1990). In addition, if the non-words in the two sound conditions were the same, the presentation of clear speech versions first may improve the intelligibility of the same non-words in their degraded format. This would present an example of perceptual learning. Noise-vocoded speech sounds are reported as being readily intelligible after a short training session (Narain et al., 2003). Noise-vocoded speech is generated by replacing a synthesised speech signal with several bands of noise, whilst preserving the amplitude envelope. Therefore, band passed filtered noise as opposed to the quasi-periodic vibrations of the vocal chords represent the spectral variation in the signal (Shannon et al., 1995). The seven non-words degraded at a SNR of 0.7 are not degraded to the same extent as noise-vocoded speech. Quasi-periodic vibrations of the vocal chords still convey spectral variation in the signal. Therefore, although participants are instructed to ignore the sounds, perceptual learning may occur. As phonological content does not determine the size of the ISE (Jones et al., 1990), using different non-words in the speech and degraded speech conditions would control for any affect of perceptual learning. The consonant and vowel sounds of the non-words all differed within the irrelevant stream for both conditions, so that each successive non-word in the irrelevant stream was distinct. The changing-state-hypothesis (CSH) argues serial recall disruption is a function of the degree of change and therefore distinctiveness between adjacent items in an irrelevant auditory stream, so long as the sounds are perceived as forming a coherent sequence (Jones et al., 1996).

4.3 THE EFFECT OF DEGRADATION ON THE IDENTIFICATION OF VOWELS AND CONSONANTS

Whether the digital manipulation of the sounds forming the degraded speech condition had a different effect on the consonants and vowels was examined. Hughes et al. (2005) demonstrated that irrelevant auditory sequences of CVC syllables in which the vowels of the syllables changed from item-to-item were more disruptive of serial recall than sequences of syllables in which the initial or final consonants changed between successive syllables. The intelligibility data for the seven non-words in the degraded irrelevant speech stream was subjected to a one-factor repeated measures ANOVA on the mean number of perceptual identification errors made for the initial consonants, vowels and final consonants (appendix 10). This was carried out in order to determine if the effects of degradation had in general degraded the vowels, but relatively preserved the initial and final consonants of the degraded non-words, which would provide a possible account of the marginal difference between the disruptive effect of clear and distorted speech.

	Mean Errors	*SD*
Initial	3.87	0.869
Vowel	2.52	0.994
Final	4.17	1.193

Table 2. Descriptive statistics indicating the mean errors in the perceptual identification of initial and final consonants and vowels in the seven degraded monosyllabic non-words. N = 23.

The descriptive statistics in table 2 displaying the mean perceptual identification errors indicates the initial and final consonants were more likely to be misperceived than the vowels. The data are summarised in figure 5.

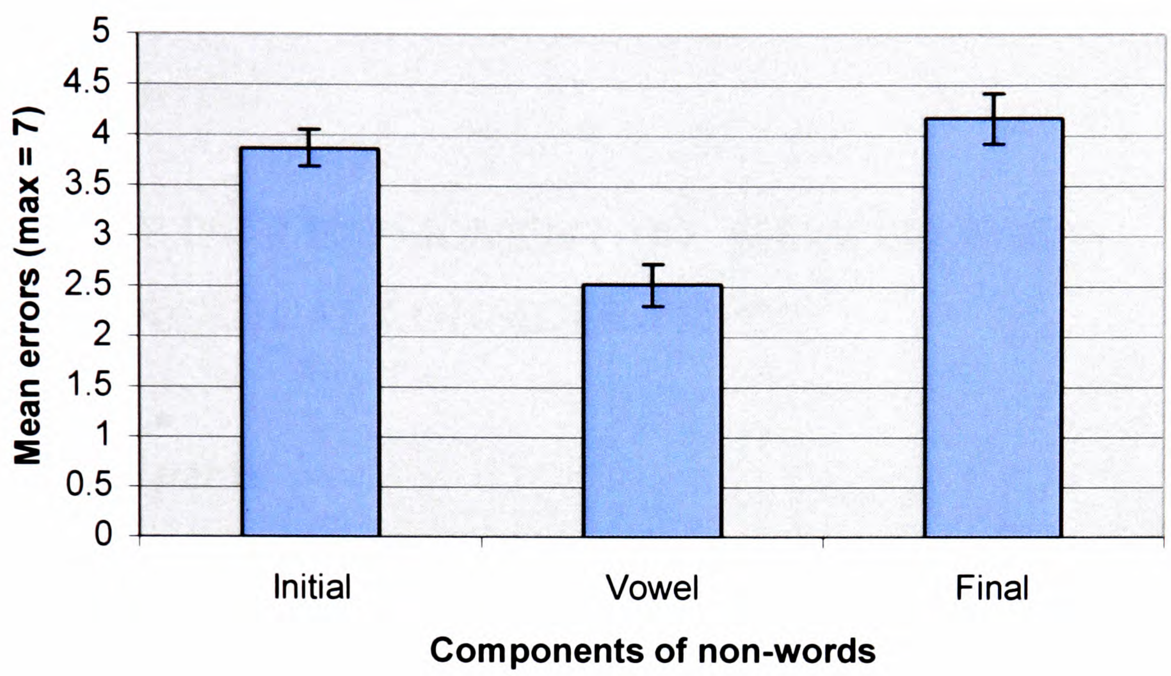


Figure 5. Mean error perceptual identification scores for the consonants (initial and final) and vowels of the non-words of the degraded speech condition. Error bars represent standard error above and below the mean. Initial = initial consonants. Final = final consonants.

	Vowel	*Final consonant*
Initial consonant	✓ $p < 0.001$	Non-sig $p \leq 0.887$
Final Consonant	✓ $p < 0.001$	xx

Table 3. Bonferroni corrected pairwise comparisons for the three non-word components of the seven non-words in the degraded speech condition.

The ANOVA showed that degradation had a different effect on the intelligibility of the initial consonants, vowels and final consonants of the seven monosyllabic non-words used as irrelevant speech [F (2,44) = 18.153, MSE = 17.783, $p < 0.001$]. Pairwise comparisons with Bonferroni correction displayed in table 3 (appendix 10) revealed that the initial and final consonants were misperceived more than the vowels ($p < 0.001$). No difference was found between the misperception of initial and final consonants ($p \leq 0.887$). Therefore, the results indicate that the vowel portion of the non-words making up the degraded speech stream was relatively preserved.

4.4 PILOT B (FOR EXPERIMENT 1B). MEMORY TASK: METHODOLOGICAL CONSIDERATIONS

4.4.1 Participants

20 undergraduate student volunteers took part in the study. The participants had English as their first language and each reported normal hearing and normal or corrected-to-normal vision. Participants were assigned to all three experimental background sound conditions:

1. Quiet control
2. Clear speech
3. Degraded speech (low in intelligibility)

4.4.2 Stimuli

4.4.2.1. Visual stimuli

Digit lists were presented serially on a VDU and were sampled quasi-randomly and without replacement from the digit set 1-7. In each trial each digit was displayed in succession. No list contained a run of more than 2 digits in ascending or descending order (appendix 11).

4.4.2.2. Auditory stimuli

Sounds were presented over a single speaker, free-field, in a seminar room. During the silent control condition no sound was played. The non-words were presented using Microsoft PowerPoint software.

4.4.3 Design

A repeated measures design was used, all participants undertaking the recall task under all three auditory conditions. There were 75 trials in all, 25 for each of the three auditory conditions.

4.4.4 Procedure

Standard instructions were given verbally to participants before the experiment began. Instructions were also displayed on a screen at the outset of each experimental condition (appendix 12). In each trial, the seven digits were displayed in random order as described above. Immediately after the presentation of each digit list, participants were required to recall the list in strict serial order. The irrelevant sound sequences were presented alongside the visual display of digits and were presented at a comfortable sound level (see schematic diagram of a trial in chapter 3, p94). The duration of irrelevant sounds and digits as well as the ISI between stimuli in the visual and auditory streams were as described in chapter 3. Participants were required to recall each digit list from left to right. The experimental trials were preceded by a short practice session of 3 trials that were not used in any analyses. The experiment lasted approximately 30 minutes. Participants completed a consent form (appendix 13) before the experiment and were fully debriefed at the end.

4.5 RESULTS

The results were scored to a strict serial recall criterion. Each digit in the recalled sequence was scored as correct only if it corresponded to the digit at that position in the to-be-remembered (TBR) sequence. Two participants were omitted from the analysis due to their poor performance in the silent condition as indicated in figure 6.

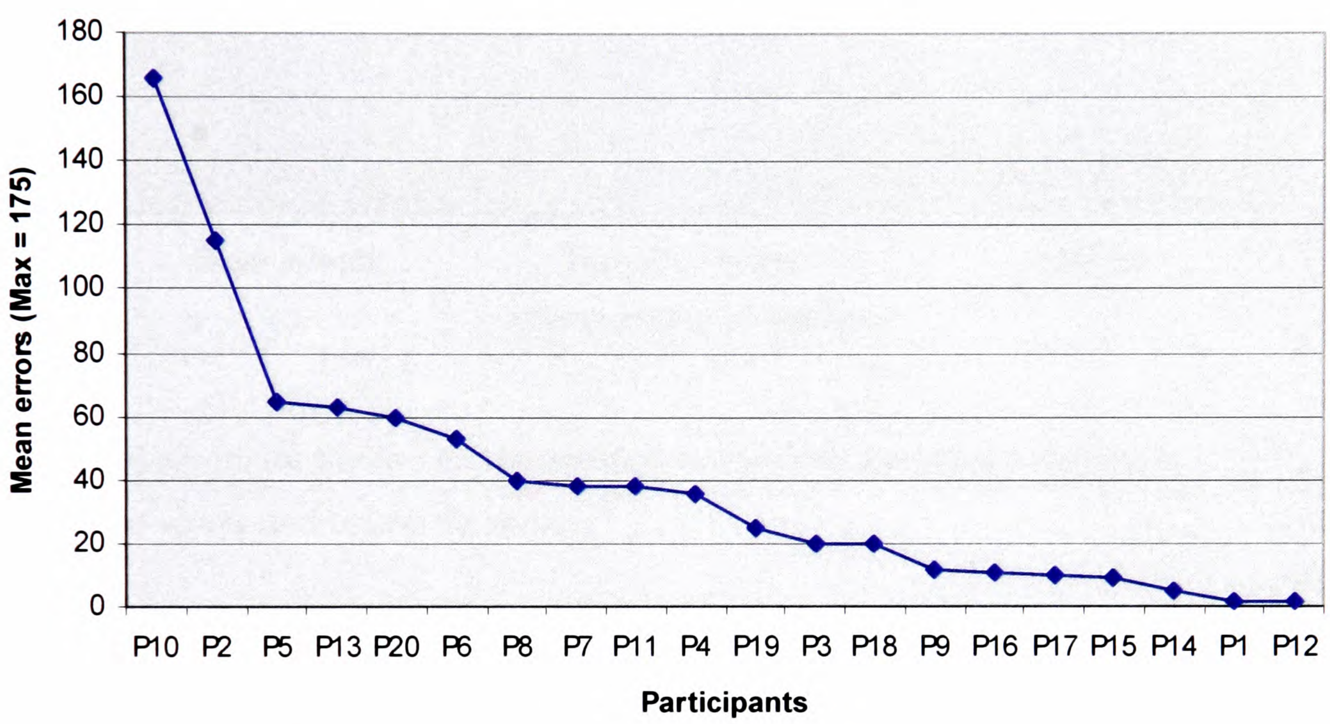


Figure 6. Scree plot of performance in the silent condition.

	Mean Errors	*SD*
Clear speech	75.89	40.202
Degraded speech	44.17	28.857
Silence	28.28	21.554

Table 4. Descriptive statistics for three experimental conditions; mean serial recall errors per condition. N = 18.

The mean recall errors made per auditory condition as shown in table 4 above show that clear speech impaired recall the most relative to degraded speech and silence. A small difference is evident between the affect of degraded speech and silence on serial recall performance. Figure 7 summarises the mean error scores for the three experimental conditions.

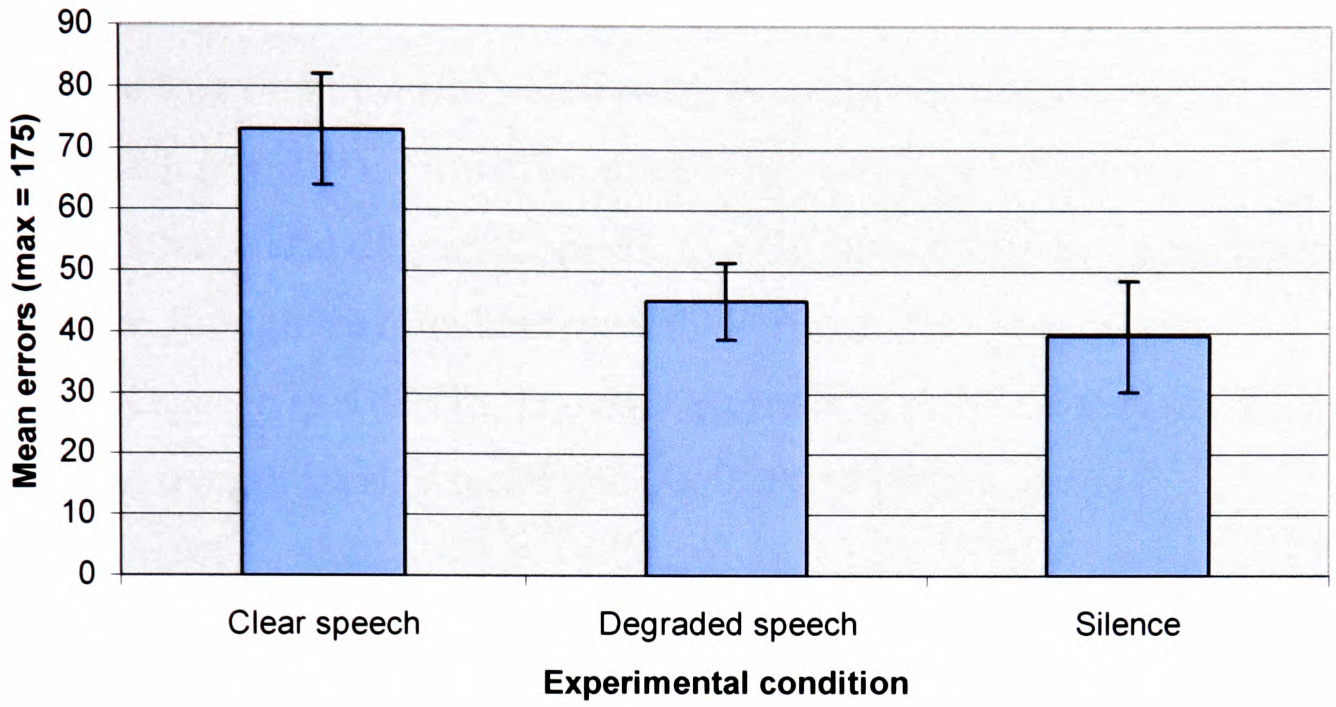


Figure 7. Mean errors for the three experimental conditions. Error bars represent standard error above and below the mean.

	Silence	*Speech*
Degraded speech	Non-sig $p \leq 0.073$	✓ $p \leq 0.049$
Speech	✓ $p < 0.01$	xx

Table 5. Bonferroni corrected pairwise comparisons for the three experimental conditions.

A one-factor repeated measures analysis of variance (ANOVA) with three levels (Clear speech, distorted speech and silence) was carried out on the mean number of serial recall errors for the three experimental conditions. Mauchy's test indicated that the assumption of sphericity had been violated ($\chi^2(2) = 7.482$, $p < 0.05$); therefore the degrees of freedom were corrected using Huynh-Feldt estimates of sphericity. The

ANOVA revealed that when compared to a silent control, serial recall is significantly disrupted by irrelevant sound [F (1.558, 26.478) = 12.133, MSE = 13581.485, $p < 0.001$] (appendix 14). Pairwise comparisons with Bonferroni correction as detailed in table 5 (appendix 14) show that when compared to a silent control, serial recall is significantly disrupted by clear speech ($p < 0.01$). However, there is no significant difference between silence and degraded speech ($p \leq 0.073$) but there is a significant difference, though marginal between the effect of clear speech and degraded speech ($p \leq 0.049$). The data are summarised in figure 8, which shows the overall level of recall collapsed across serial position.

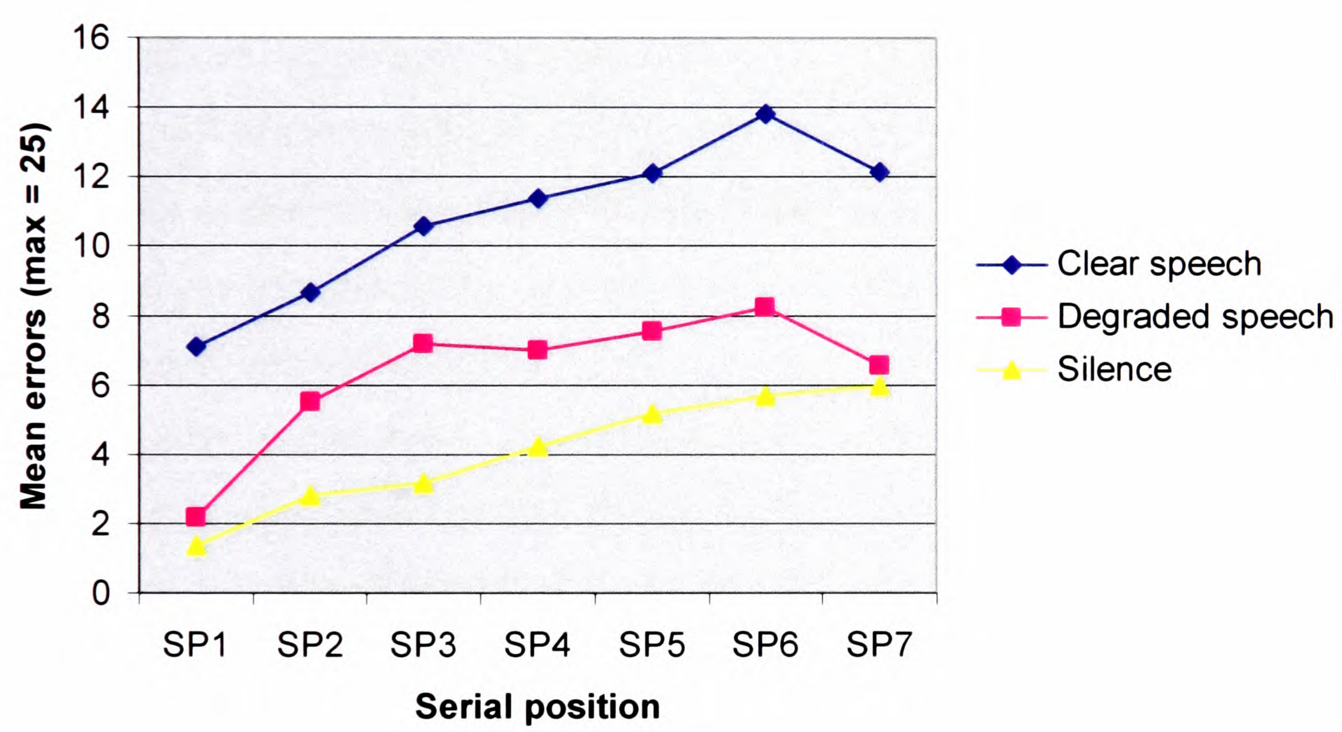


Figure 8. Mean errors for the three experimental conditions collapsed across serial position.

4.6 DISCUSSION

The pilot study provides evidence in agreement with the general principle of the CSH (Jones et al., 1996) that digitally degrading an irrelevant stream of speech attenuates the size of the irrelevant sound effect when compared with the effect of clear speech. The pilot data can therefore be explained by the CSH with reference to the proposition that the presence of noise attenuates the ISE by reducing the number of acoustic changes within the speech signal.

That there was no difference between the disruptive affect of distorted speech when compared to the silent control was not predicted as the degraded speech items still varied acoustically, though the digital manipulation reduced the degree of acoustic variation. Previous studies comparing the effect of steady-state sound with that of silence demonstrate equivalence in memory performance (e.g. LeCompte, 1996). Sequences of steady state sounds do not demonstrate physical change between adjacent tokens and this offers an account for the lack of an ISE being attained in its presence. In contrast, the degraded irrelevant speech samples still convey abrupt changes across several perceptual attributes, so the samples are not reduced to signals resembling steady-state sounds. The analysis of the effect of degradation on the identification of the vowels and consonants in the degraded non-words indicate that the vowels were relatively preserved. Therefore, a significant difference between the degraded speech and silent control condition would be expected with regards to the remaining presence of changing-state information, in particular the changing vowels.

One key objective of this study was to establish how much speech has to be degraded in order for it to disrupt serial recall at a level that significantly differs from recall in the clear speech and silent control condition. The finding of no difference between degraded speech and silence may be due to the level of degradation being too high at a SNR of 0.7. An increase in the SNR of the degraded speech may increase the variance of serial recall errors between those occurring under silence and in the presence of irrelevant degraded speech sounds. However, Jones et al (2000) degraded speech digitally by degrees. Speech reduced to amplitude modulated noise was still more disruptive than non-speech degraded to the same degree. As a silent control condition did not feature in the study of Jones et al. (2000) but did in the present experiment it does not form a fair comparison and conclusions cannot be drawn about whether or not a reliable ISE would have been observed with speech reduced to amplitude modulated noise. In addition, Tremblay et al (2001) found that bursts of broadband noise that changed in frequency disrupted serial recall relative to a quiet control. Therefore, as speech reduced to amplitude modulated noise disrupts memory more than non-speech degraded to the same degree and bursts of broadband noise arguably contain less changes in state than do degraded non-words, a difference between memory performance in the degraded speech and silent control condition would have been expected. This is because acoustic variation was still conveyed between degraded items, a characteristic in an irrelevant stream necessary for the observation of a changing-state effect (Jones et al., 1996). Therefore the lack of a reliable difference in serial recall disruption by degraded speech and silence in the present findings must be attributable to its weak experimental design. A weak experimental design was employed for this pilot experiment as an opportunity sample was available. This pilot was conducted in order to screen which SNR would afford the isolation of seven distinct non-words where the consonants and vowels differed between adjacent sounds.

There are several limitations inherent in this pilot's experimental design, which may explain why no difference in serial recall performance was observed between the degraded irrelevant speech condition and the silent control condition. First, the free-field presentation of the sounds over a single speaker to a group would result in additional distortion because of the room transfer function (Moore, 2004). Large rooms and lots of furniture in a room cause standing waves, which affect the transfer function of the sound. Thus, the sound reaching the ears (the receiver) is very different from that which came from the speaker (the transmitter). Second, the design was not fully counterbalanced as only one group of participants were tested and therefore received only one of the six possible orders. Therefore, the design has not controlled for practice and fatigue effects. Third, the sounds in clear and degraded speech conditions were not randomised between trials. Although there is no evidence for habituation to a repeated sequence of distinct irrelevant sounds (Tremblay and Jones, 1998), the randomisation of sounds for each list in each condition would be a more objective and systematic approach.

CHAPTER 5

5 THE EFFECT OF PHONOLOGICAL DEGRADATION ON THE DEGREE OF SERIAL RECALL DISRUPTION BY IRRELEVANT SPEECH

5.1 AIMS AND OBJECTIVES

Experiment 1b aimed to control for noise distortion that featured in pilot A and B (chapter 4) by presenting auditory stimuli over quality headphones and testing participants individually in a quiet listening room. The non-words used were digitally degraded using the same technique as pilot A and B (chapter 4). During the perceptual identification task (experiment 1a) the non-words were also presented over quality headphones rather than being presented over a single speaker used to deliver the sounds in the pilot perceptual identification task (pilot A). Partially degrading a signal at 0.7 SNR as in pilot A and B does not remove or attenuate all of the acoustic changes within the spectrally complex speech signal. The presentation order of the sound conditions during the memory task was fully counterbalanced so that the condition order for each participant was taken from the six possible condition orders without replacement. In addition a larger sample size was used to allow for full counterbalancing of the six possible sound conditions. A tighter experimental design was predicted to bring out a reliable difference between degraded speech and silence and remove any effect of practice or fatigue that may have confounded the findings of pilot B.

The prediction that a reliable difference in serial recall interference will be observed between degraded speech and silence is derived from the literature, which shows any physical change between perceptually distinct auditory items will produce an ISE (e.g., Jones et al., 2000; Jones and Macken, 1995a, Ellermeier and Hellbrück, 1998), so long as it does not breach the temporal coherence boundary so that the auditory objects are no longer perceived as one coherent stream (e.g., Jones and Macken, 1995b; Jones et al., 1999a; Jones et al., 1999b). A reliable difference between the disruptive effect of clear and degraded speech as found in pilot B (for experiment 1b) is also predicted on the grounds that distortion of the speech signal reduces the number of acoustic changes from item-to-item which leads to an attenuation in serial recall performance (e.g. Jones et al., 2000).

5.2 EXPERIMENT 1A. PERCEPTUAL IDENTIFICATION TASK: ADDITIONAL METHODOLOGICAL CONSIDERATIONS

5.2.1 Participants

23 undergraduate students were used. The participants who did the pilot perceptual identification task (pilot A, chapter 4) were not used for this experiment as 50 of the non-words that were previously screened for intelligibility would seem more common to them. All participants had English as their first language and reported normal or corrected-to-normal vision and normal hearing.

5.2.2 Preparation of auditory stimuli

100 monosyllabic non-words were screened for intelligibility, 50 of which featured in the previous pilot perceptual identification task (appendix 15, see also appendix 5 for examples of disc phonetic symbols). The non-words were digitally edited and prepared as described in the pilot methodology. The level of degradation for all 100 non-word

samples was set at a SNR of 0.7 and therefore 30% of the signal was replaced with random samples. This level of degradation was chosen on the basis of it generating a better distributed range of intelligibility for the non-words in the pilot study. The experiment generating software E-prime (Psychology Software Tools, INC, 2002) was used to present each sound for 1000ms and the Inter-Trial Interval (ITI) was controlled by the participant's according to the speed with which they gave their written response.

5.2.3 Design

A repeated measures design was used as all participants undertook the perceptual identification task. E-prime (Psychology Tools INC, 2002) was used to present the non-words in random order for each participant.

5.2.4 Procedure

Participants were tested individually, seated in a quiet room approximately 0.5m from the computer screen. At the outset, standard instructions were given to the participants and presented on VDU (appendix 16). During each experimental session, each non-word was presented through Sennheiser HD570 headphones in random order as described above. The random order of the 100 non-words was recorded and each non-word was presented for 1000ms. After the presentation of each non-word, participants were instructed to write down phonetically what they thought they heard on a response sheet. The intelligibility screening sessions lasted approximately 20 minutes. Participants completed a consent form (appendix 7) before the experiment and were debriefed at the end.

5.3 RESULTS

The responses were scored as detailed in pilot A (see chapter 4). From the intelligibility range (appendix 17), seven non-words of low intelligibility were selected for the irrelevant degraded speech condition of experiment 1a. Non-words that were low in intelligibility (with recognition scores in the range 0-4 correct identifications) were isolated from the bottom quartile of the intelligibility range. Seven different non-words were selected for the clear (un-degraded) speech condition. Different non-words were used in the speech and degraded speech conditions as the literature shows that phonological content does not determine the size of the ISE (e.g. Jones et al., 1990). Having different non-words in the two conditions also eliminates any possibility of perceptual learning leading to the identification of the degraded non-words. This may be the case when the clear speech condition occurs before the degraded speech condition (as discussed on p100). The consonant and vowel sounds of the non-words all differed within the irrelevant stream for both conditions (appendix 18; see also appendix 5 for examples of disc phonetic symbols).

5.4 THE EFFECT OF DEGRADATION ON THE IDENTIFICATION OF VOWELS AND CONSONANTS

	Mean Errors	*SD*
Initial	4.78	0.951
Vowel	2.78	1.085
Final	2.35	1.027

Table 6. Descriptive statistics indicating the mean errors in the perceptual identification of initial and final consonants and vowels of the seven digitally degraded non-words. N = 23.

As in pilot A (for experiment 1a), a follow up analysis was performed, to examine the effect of phonological degradation on the components of the non-words (initial and final consonants and vowels). The descriptive statistics in table 6 indicate the initial consonants of the seven non-words in the degraded speech condition were misperceived more than the vowels or final consonants. However, there is only a small numerical difference between the mean number of identification errors for the vowels and final consonants. The data are summarised in figure 9.

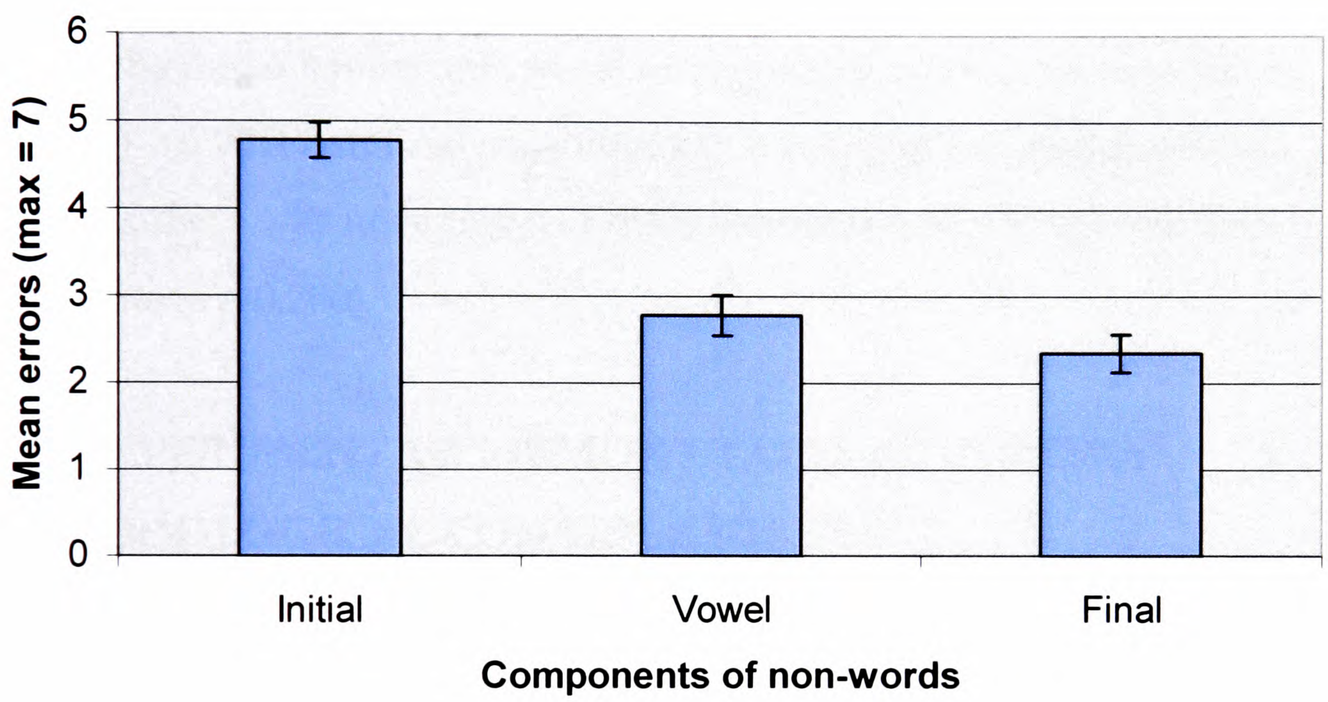


Figure 9. Mean perceptual identification errors for the components of the seven non-words in the degraded speech condition. Error bars represent standard error above and below the mean.

	Vowel	*Final consonant*
Initial consonant	✓ $p < 0.001$	✓ $p < 0.001$
Final Consonant	Non-sig $p \leq 0.700$	XX

Table 7. Bonferroni corrected pairwise comparisons for the three non-word components of the seven non-words in the degraded speech condition.

A one-way repeated measures ANOVA (appendix 19) was carried out on the number of incorrect identifications for the initial and final consonants and the vowels of the seven non-words in the degraded speech condition. Mauchly's test indicated that the assumption of sphericity had been violated (χ^2 (2) = 7.170, $p < 0.05$); therefore the degrees of freedom were corrected using Huynh-Feldt estimates of sphericity. The results show there was an effect of phonological degradation [F (1.647, 36.235) = 33.393, MSE = 47.093, $p < 0.001$]. Pairwise comparisons with Bonferroni correction displayed in table 7 (appendix 19) showed the initial consonants were misperceived more relative to the vowels ($p < 0.001$) and final consonants ($p < 0.001$) of the speech sounds. However, there was no significant difference between vowels and final consonants ($p \leq 0.700$).

5.5 EXPERIMENT 1B. MEMORY TASK: ADDITIONAL METHODOLOGICAL CONSIDERATIONS

5.5.1 Participants

30 undergraduate student volunteers took part in the study. The participants used had English as their first language and each reported normal hearing and normal or corrected-to-normal vision. They were not paid for their time.

5.5.2 Stimuli

5.5.2.1 <u>Visual stimuli</u>

As in pilot B (for experiment 1b) the digit lists were presented serially on a VDU and were sampled quasi-randomly and without replacement from the digit set 1-7. In each trial each digit was displayed in succession. No list contained a run of more than 2 digits in ascending or descending order (appendix 20).

5.5.2.2 Auditory stimuli

The sounds for the clear and degraded speech conditions were presented over Sennheiser HD570 headphones (see general procedural outline in chapter 3). The non-words for the clear speech condition and degraded speech condition can be seen in appendix 18. The non-words were recorded and edited as described in chapter 3 (p90). The non-words for the degraded speech condition were degraded as detailed in chapter 4 (p98). For each digit list, the appropriate set of irrelevant sounds was randomised. The sounds were presented concurrently with the digits as detailed in chapter 3 (p94).

5.5.3 Design and procedure

The presentation order of conditions was fully counterbalanced and the general procedure was as detailed in the general procedural outline section of chapter 3 (p93).

5.6 RESULTS

	Mean Errors	*SD*
Clear speech	55.63	24.480
Degraded speech	47.33	23.227
Silence	35.63	21.384

Table 8. Descriptive statistics for three auditory conditions; mean number of serial recall errors per condition. N = 30.

More recall errors occurred under clear speech relative to degraded speech and the silent control. There is a difference between clear and degraded speech and degraded speech and silence (table 8). The data is summarised in figure 10.

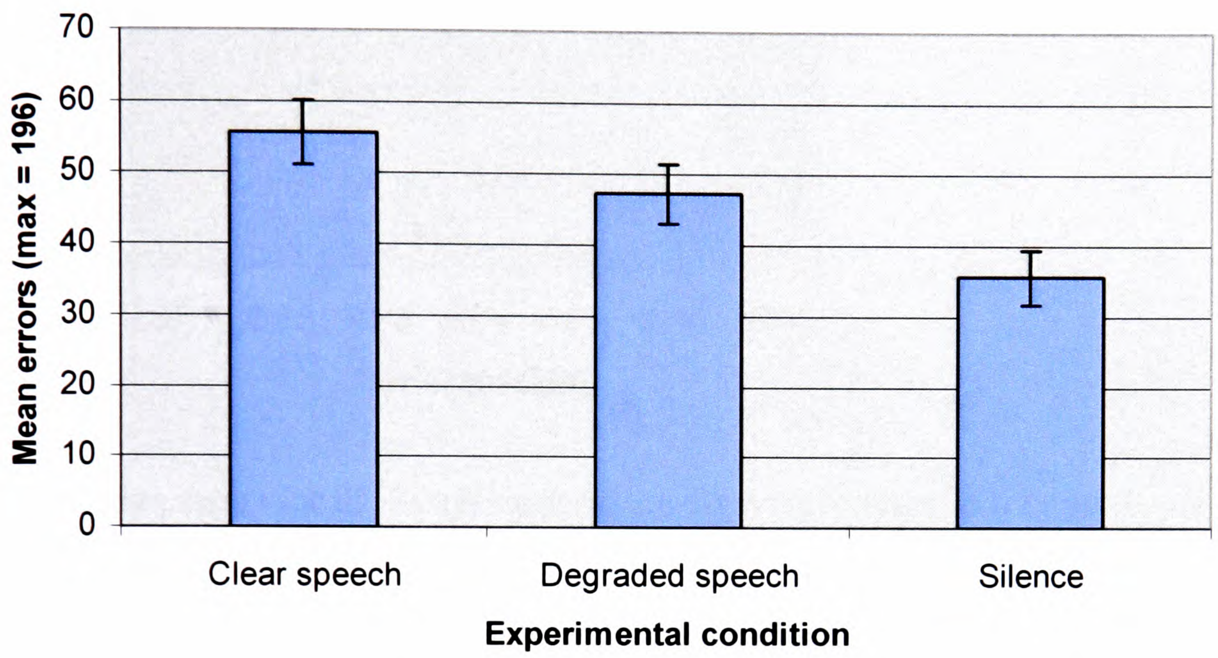


Figure 10. Mean errors per experimental condition. Error bars represent standard error above and below the mean.

The mean numbers of digits incorrectly recalled on three levels (clear speech, degraded speech and silence) were subjected to a one-factor repeated measures ANOVA (appendix 21). Again, when compared to a silent control, serial recall is significantly disrupted by irrelevant sound [F (2, 58) = 26.518, MSE = 3028.900, p < 0. 001]. The data are summarised in figure 11, which shows the overall level of recall collapsed across serial position.

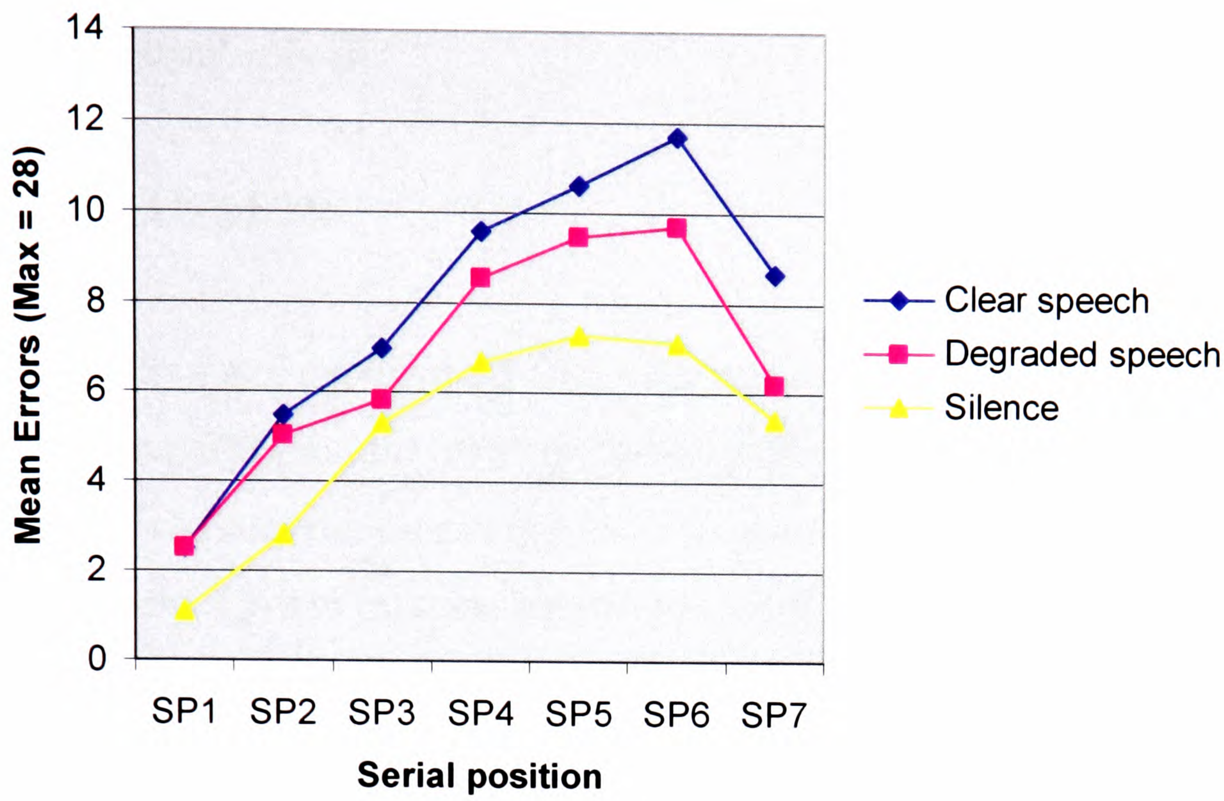


Figure 11. Mean errors for the 3 experimental conditions collapsed across serial position.

	Silence	***Speech***
Degraded speech	✓ $p < 0.01$	✓ $p < 0.01$
Speech	✓ $p < 0.001$	XX

Table 9. Bonferroni corrected pairwise comparisons for the three irrelevant sound conditions.

Pairwise comparisons with Bonferroni correction as displayed in table 9 were carried out to see which conditions differed significantly (appendix 21). Clear speech disrupted recall the most relative to a silent control ($p < 0.001$) and disrupted recall significantly more than degraded speech ($p < 0.01$). Contrary to the previous results, degraded speech not only differed from clear speech in its ability to disrupt recall, but also differed significantly from silence ($p < 0.01$). This demonstrates that the digital degradation of the non-word samples has removed some important perceptual attribute of the speech signal which allows clear (natural) speech to be more disruptive of immediate memory than degraded natural speech.

5.7 DISCUSSION

The results of experiment 1b replicate the general finding observed in the pilot experiment and confirm that digitally degrading speech significantly reduces the size of the ISE when its effect is compared to that of clear speech. However, once conditions are fully counterbalanced and presented in a more appropriate testing environment, distorted speech impaired serial recall at a level that significantly differed from when the TBR items were presented in silence.

The changing-state-hypothesis (CSH) (Jones et al., 1992) can account for the higher disruptive effect of clear speech relative to degraded speech, by assuming that the reduction in the ISE is due to the fact that the number and extent of acoustic change is greater for clear speech than for degraded speech. Acoustic variation in this paradigm is defined as the variation in terms of all acoustic characteristics between distinct items, except for overall amplitude (Tremblay and Jones 1999). However, the acoustic characteristic(s) that need to change-in-state from item-to-item in an irrelevant sound stream is not proposed by the CSH, and it is clear that several acoustic changes within the multi-dimensional speech signal have been attenuated along with phonemic intelligibility due to digital degradation.

Although much evidence to contradict the phonological store hypothesis (PSH) exists, such as the observation of an ISE with irrelevant non-speech stimuli as well as speech stimuli (e.g., Jones and Macken, 1993 and Jones et al., 2000), the present findings suggest that phonological degradation removes or distorts an important characteristic of the phonetic content of speech which serves to significantly attenuate its disruptive effect. The ability to correctly identify the phonemic content of speech sounds may account for the significantly higher level of memory interference of irrelevant clear speech. It can be argued that the CSH's assumption that more changes in state between successive irrelevant items accounts for the greater power of clear speech to disrupt serial recall does not quantify the effect of clear speech relative to degraded speech. Degrading speech not only acts to reduce acoustic variation from item-to-item in the irrelevant stream, it also reduces the intelligibility of spoken items, in terms of their identification.

Recent research using consonant-vowel-consonant (CVC) syllables as irrelevant speech items investigated the disruptive power of changes only in the final consonant, the initial consonant, or in the vowel of each item (Hughes et al., 2005). The findings indicate that *vowel changes*, as opposed to consonant changes are the dominant source of disruption, an effect referred to as the 'vowel-changing-state effect'. Surprenant and Neath (1996) found that attended to vowel-only-changing (V-O-C) items are better recalled in their serial order than consonant-only-changing (C-O-C) items. Hughes et al. (2005) argue that items that are easily recalled in their correct serial order when in the attentional focus will in turn be equally as disruptive to serial recall as irrelevant sound. This suggests, for verbal serial memory, vowels provide more serial order cues than do consonants. When sound is unattended, vowel changes would generate more conflicting order cues in memory based on the seriation process being similar for irrelevant sound and relevant visual TBR item streams.

Hughes et al. (2005) also point out how research into the perceptual organization of sound provides further support for the suggested dominant role of changing vowels in the ISE. It is argued that the attribute shared by speech sounds used by the perceptual system to integrate different speech utterances spoken by the same voice over time is afforded by the continuity in some percept, for example a common fundamental and formant pattern shared by the phonated-periodic vowels of these utterances. In contrast, consonants give the periodic sounds their characteristic noisy onsets and offsets (Hughes et al., 2005). Further, the spectral complexity of vowels is greater than consonants (Bregman, 1990). Acoustic differences between discrete vowels produced by varying the shape of the vocal tract occur on the common fundamental and formant structure of spoken utterances and as a consequence create strong serial order cues, which would cause more serial recall interference than would consonant variations.

In light of the findings of Hughes et al (2005) and Surprenant and Neath (1996) a subsequent analysis was performed on the intelligibility data for the seven low intelligible non-words making up the degraded irrelevant sequence. The analysis indicated that the initial consonants were misperceived more than the vowels, which were relatively well preserved. However, the level of degradation resulted in the vowels and final consonants of the degraded non-words being statistically equivalent in their misperception. For example, for the non-word /thet/ (TEt) the initial consonant sound /T/ was identified incorrectly significantly more than the vowel sound /E/ and the final consonant sound /t/ (see appendix 5 for examples of disc phonetic symbols). One explanation is that in these samples enough information corresponding to the vowels may have been removed from the speech signal in order to make the disruptive potency of degraded speech lower than that of clear speech. This may have been the case even though consonants were misperceived more than vowels since Hughes et al (2005) demonstrated vowel changes as opposed to consonant changes to be important in serial recall disruption by irrelevant speech. Also, there are fewer consonants that can legally follow a vowel. Therefore, it may have been easier for participants to guess the final consonant of the non-word samples.

In considering the finding of the equivalent misperception of the vowels and final consonants of the non-words in the degraded speech condition, there is insufficient evidence to generalise to the whole syllable population, since only seven monosyllabic non-words acted as the degraded stimuli across all the trials within this auditory condition. It may be that this result was only observed because these non-words were employed as irrelevant stimuli.

It follows that the result may be different if a larger set of CVC syllables acted as the irrelevant stimuli in regards to examining how consonants and vowels are affected by digital signal degradation. Accordingly, a more objective and systematic methodological approach to exploring the affect of degradation on speech sounds would be to use a larger sample of non-words, so that each irrelevant changing-state sequence contains different distinct non-words.

CHAPTER 6

6 THE RELATIONSHIP BETWEEN DEGRADATION AND SERIAL RECALL PERFORMANCE FOR VOWEL-ONLY AND CONSONANT-ONLY-CHANGING NON-WORDS

6.1 BACKGROUND

Experiment 1b (chapter 5) showed that digitally degrading speech attenuated the intelligibility of a selection of non-words. The degraded non-words low in intelligibility went on to reduce serial recall performance when presented as irrelevant sound (experiment 1b, chapter 5). As changing vowels are argued to generate more competing cues to serial recall because when attended, vowel-only-changing (V-O-C) sequences are better recalled in serial order than consonant-only-changing (C-O-C) sequences (Surprenant and Neath, 1996), it is useful to consider the effect of noise on verbal to-be-remembered (TBR) item sequences.

Several studies have demonstrated that distortion of the signal of verbal TBR items has a detrimental effect on serial recall of these sequences. Luce, Feustel and Pisoni (1983) showed that naturally spoken words were better recalled in serial order than synthetically spoken words. One characteristic shared by the research of Luce et al (1983) and the findings of experiment 1b (chapter 5) is that the sounds making up both speech conditions differed in their intelligibility. Therefore, the reduction in serial recall performance in the presence of degraded speech (experiment 1b, chapter 5) and the poorer serial recall of synthetically spoken words relative to naturally spoken words (Luce et al., 1983) may be due to the reduced intelligibility of the phonemes making up the speech sounds, in particular the vowels. However, the sounds in the clear and degraded speech conditions (experiment 1b, chapter 5) and in the naturally and synthetically spoken words of Luce et al (1983) differed also in the degree of acoustic variation conveyed from-item-to-item. The intelligibility of items in the clear and degraded speech condition would need to be matched whilst reducing the acoustic variation between successive items. This would be necessary in order to make conclusions as to whether it is the reduction in intelligibility of items or the acoustic variation conveyed from item-to-item that is responsible for the improvement in performance in the presence of degraded speech.

The phonological store hypothesis (PSH) argues that memory interference is the product of the confusion between irrelevant and relevant phonological codes within the hypothetical phonological store. The PSH as described earlier cannot account for the effect of non-speech stimuli on serial recall performance (e.g. LeCompte et al., 1997 and Jones et al., 2000). Non-speech sounds have no phonetic content and it is not possible for them to be re-coded into phonemes and so they should not disrupt immediate memory. Therefore, the confusion between irrelevant and relevant phonological codes cannot account for the ISE. However, speech is found to be more disruptive of serial recall than non-speech (e.g. LeCompte et al., 1997 and Jones et al., 2000) and degraded speech (experiment 1b, chapter 5), and so the intelligibility of the phonetic content of speech may explain the greater disruptive power of speech.

There is research evidence which suggests that the correct identification of words may not predict their ability to disrupt memory. Rabbitt (1991) demonstrated that participants with mild peripheral hearing loss recalled fewer words than did an age-matched control group with normal hearing, even though, the words for both groups were identified equivalently. Identification performance was measured by asking participants to overtly shadow the words as each one was presented. Rabbitt (1991) argued the decrement in memory performance may be due to the extra difficulty experienced by the participants with peripheral hearing loss in interpreting the words. That the participants had difficulty in interpreting the TBR words may have lead to a reduction in the encoding and rehearsal of the words. This demonstrated that even though both groups of participants demonstrated equivalent identification performance for the TBR words, memory performance was still adversely affected more for the group with peripheral hearing loss.

The reduction in intelligibility and the physical complexity of the consonant-vowel-consonant (CVC) syllables through degradation of the signal in experiment 1b (chapter 5) may also have resulted in more difficulty in interpreting the CVC syllables and may have required the use of more cognitive resources. Considering these CVC syllables are irrelevant to the task and are ignored, this account seems unlikely. It may be that the degraded speech sounds form weaker physical memory representations, which in turn generate less serial order cues to conflict with those elicited by sub-vocal rehearsal of the TBR items disrupting serial recall less. This account would be in keeping with the CSH, which assumes a sequence in which acoustic changes are removed or attenuated between successive irrelevant auditory items will interfere with serial recall less (Jones et al., 1996).

Surprenant (1999) demonstrated reduced memory for spoken syllables mixed with noise, even though identification performance was equated across syllables. This finding is explained with reference to the memory representations of the syllables formed during their initial encoding. These can be degraded by the addition of noise, even though the physical representation of the sound is left unimpaired or at least occurs at an equivalent level across noise conditions. Degrading the memory representations of these syllables can impair discrimination of these auditory items, thus reducing the amount of information pertaining to their serial order. An indirect inference can be drawn from these results when predicting the effect of task-irrelevant sounds degraded by signal-correlated-noise (SCN) at different signal-to-noise ratios (SNR). A sequence of degraded auditory stimuli would demonstrate less change from item-to-item and would therefore disrupt the maintenance of the order of visual TBR items less as poorly discriminated sounds generate weaker cues to their serial order. The PSH could account for the effect of degradation on the disruptiveness of speech if it was adapted to assume the acoustic patterning of degraded speech sounds would form weaker physical memory representations that would result in less confusion between irrelevant auditory and relevant TBR items.

Jones et al (2000) observed a linear relationship between degradation and serial recall disruption for both speech and non-speech stimuli. Progressive degradation of the speech sounds and cello notes resulted in a reduction in serial recall performance. That is, as degradation of both classes of sound increased, disruption of serial recall decreased. This indicated an equivalent pattern in the reduction of serial recall errors by degradation for speech and non-speech stimuli, as simple cello notes revealed a similar pattern of interference to that produced by speech (Jones et al., 2000). They found continuity in the effect of degradation on serial recall performance for speech sounds. This would not have been observed if a threshold had needed to have been exceeded in order for a speech sound to be identified as a non-word before it significantly became detrimental to immediate memory. Jones et al (2000) argue that this is evidence for the prediction proposed by the CSH, that as adjacent items in an irrelevant sound sequence vary less and therefore become less distinct, the degree of serial recall interference is reduced.

When recalling to-be-remembered (TBR) items in sequential order when irrelevant sounds degraded at various levels are presented, individual differences are found with regards to the number of serial recall errors made. When data from a typical ISE task such as that of Jones et al (2000) is averaged a linear function is observed. Although Jones et al (2000) demonstrated cello notes to be disruptive of serial recall in a functionally similar way to speech; speech when fully degraded is still more disruptive. Fully degraded speech in Jones et al's (2000) experiment was stimuli reduced to amplitude-modulated noise. Recent research has provided direct evidence for the importance of changes in the vowels of successive auditory items in the disruption of serial recall by irrelevant speech (Hughes et al., 2005). Hughes et al (2005) examined which phonetic component changing within a stream of speech sounds acted as the dominant source of serial recall disruption. Hughes et al (2005) manipulated whether the initial consonant, final consonant or vowel portion of a sequence of speech sounds changed from item-to-item and measured their effect on serial recall when presented as irrelevant sound. A sequence of CVC syllables in which all components (initial consonants, vowels and final consonants) of the syllables changed disrupted serial recall performance more than a sequence in which only the final consonant changed. In addition, sequences in which only the vowel changed from syllable-to-syllable produced a larger decrement in serial recall performance than did sequences in which only the initial consonant changed.

The weaker effect on memory performance observed with C-O-C CVC syllables was further emphasised by these syllables causing a level of disruption equivalent to that observed in the presence of a steady-state sequence. For example, disruption in the presence of a steady-state sequence in which a syllable was repeated was similar to disruption in the presence of a sequence in which only the initial consonant changed from item-to-item. Disruption produced in the steady-state condition was also similar to disruption observed in the presence of a sequence in which only the final consonant changed. Hughes et al (2005) explained this finding in light of research on the effect of memory for spoken syllables which also manipulated whether it was the consonants or vowels only that changed in an attended to-be-remembered (TBR) sequence (Surprenant and Neath, 1996). Like Hughes et al (2005), Surprenant and Neath (1996) found V-O-C sequences of syllables were recalled better than their C-O-C counterparts. Hughes et al (2005) inferred that attended-to-auditory items that are more easily recalled will also be more disruptive to the serial recall of visual TBR items when unattended. This is because changing vowels within an irrelevant stream generate more competing cues to serial order than do streams of C-O-C syllables. Hughes et al (2005) further support this inference by considering factors important for the perceptual organisation of speech sounds. Changes on an attribute shared by speech sounds produced by the same speaker, such as fundamental frequency ($f0$) and formant structure are provided by the voiced and therefore periodic vowel sounds. These help the auditory system to integrate speech sounds over time, so as to allow the identification of sounds as coming from the same talker and therefore keeping them in the same stream. The integration of speech sounds over time by the perceptual system therefore serves to maintain the temporal order of streams of speech sounds (Bregman, 1990).

6.2 AIMS AND OBJECTIVES

Experiment 2 was designed to compare disruption of serial recall in the presence of degraded irrelevant CVC sequences in which only the vowels change from syllable-to-syllable with the effect of sequences in which only both the initial and final consonants change. Both classes of irrelevant speech will be subjected to three levels of degradation by randomising a percentage of the sample points in each irrelevant speech sound (0%, 30% and 50% noise). Degraded speech may have been less disruptive of serial recall relative to clear speech in experiment 1b (chapter 5) because the identification of the vowels which changed between the degraded non-words was reduced. Although the consonants of these degraded non-words were misperceived more than the vowels, vowels were still misperceived.

Three possible outcomes can be predicted on the basis of previous research which has identified sequences of V-O-C syllables to be more disruptive of serial recall than sequences of C-O-C syllables (Hughes et al., 2005). If vowel 'changes' are the dominant source of disruption, one outcome is that an equivalent linear function for the relationship between degradation and serial recall performance would be demonstrated by both types of changing irrelevant sequences, but where V-O-C streams would be more detrimental to immediate serial memory than C-O-C streams.

Second, a shallow linear function for the relationship between degradation and serial recall disruption may be observed when sequences in which only the vowels change are implemented as irrelevant sound. This can be predicted on the basis of the resistance of vowels to signal degradation, in that consonant recognition is more sensitive to the degradation of temporal cues than is vowel recognition (Drullman, Festen and Plomp, 1994). A shallower linear function for the relationship between degradation and serial recall interference when V-O-C sequences are degraded by degrees would indicate a higher level of serial recall interference at each level of degradation due to the vowels changing from syllable-to-syllable. In contrast, an irrelevant C-O-C stream would be expected to produce a steep linear degradation function due to the greater effect phonological degradation has on consonant intelligibility as found in experiment 1b (experiment 1b, chapter 5) and because of the suggested greater degradation of the memory representations of consonant phonemes (Surprenant and Neath, 1996). Figure 12 shows a graphical representation of this predicted relationship between stimulus degradation and serial recall performance for V-O-C and C-O-C sequences.

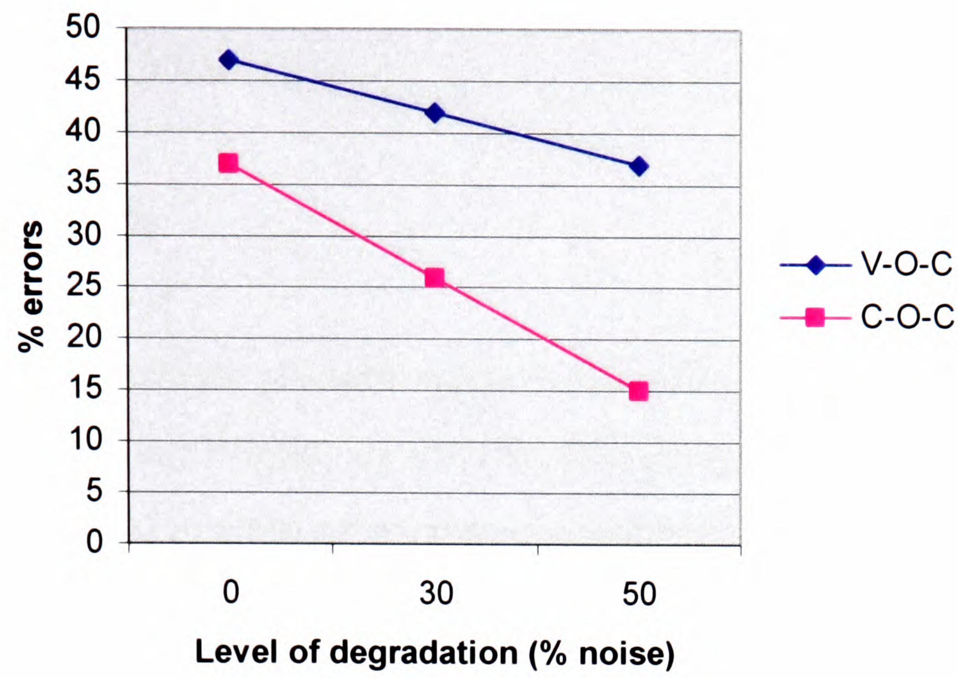


Figure 12. Graphical representation of the predicted relationship between degradation and serial recall disruption for degraded V-O-C and C-O-C sequences. Note that the graph does not represent real data.

The third possible outcome is that as degradation of V-O-C speech increases, the number of serial recall errors made in its presence will decrease. In contrast, no reliable difference in serial recall performance will be found between the number of serial recall errors in the presence of clear and degraded C-O-C speech. This can be predicted on the basis that vowels have been found to be the dominate source of disruption when they change from item-to-item in an irrelevant stream, whereas the size of the ISE in the presence of changing consonants is dramatically reduced (Hughes et al., 2005). Degrading speech reduced the size of the ISE in experiment 1b (chapter 5) and the vowels of the degraded non-words were found to be misperceived, though not to the extent that the consonants were. In light of Hughes et al (2005) research, the finding that degraded speech is less disruptive to serial recall may be due to the effect of degradation being more detrimental to the information provided by vowels, which have been shown to provide the critical changing-state information in speech.

6.3 EXPERIMENT 2: METHODOLOGICAL CONSIDERATIONS

6.3.1 Participants

30 undergraduate student volunteers took part in the study. The participants used had English as their first language and each reported normal hearing and normal or corrected-to-normal vision. They were not paid for their time.

6.3.2 Stimuli

6.3.2.1 <u>Visual stimuli</u>

The digit lists for the serial recall task where constructed in the same way and were presented serially on a VDU as described in chapter 3 (appendix 22).

6.3.2.2 <u>Auditory stimuli</u>

150 CVC non-words were recorded digitally at a sampling rate of 22.5 KHz and to a resolution of 16 bits. The non-words were edited using the sound editing software Cool Edit Pro 1.2 (Syntrillium Software Corporation) as described chapter 3 (p90). For 75 of these non-words, only the vowel changed from item-to-item, for example *gam* (`g{m`), *gim* (`gIm`), *gem* (`gEm`), *gom* (`gQm`), *and garm* (`g£m`). These are referred to as vowel-only-changing (V-O-C) sequences. For the other 75 non-words only the consonants changed from item-to-item and the vowel remained fixed creating consonant-only-changing (C-O-C) sequences, such as *baysh* (`b1S`), *fayv* (`f1v`), *gayd* (`g1d`), *tayn* (`t1n`), and *wayth* (`w1D`) (see appendix 5 for examples of disc phonetic symbols). Both the V-O-C and C-O-C CVC syllables were organised into 15 streams of seven monosyllables

(appendix 23). For each of the V-O-C and C-O-C speech sequences, two additional versions were generated by digital signal processing as was implemented in experiments 1a and 1b (chapter 5, for details see p 98, chapter 4) to form a set of degraded non-words at two levels of degradation for both classes of syllable. The non-words for both types of sound were degraded at a SNR of 0.7 (30% of samples within the signal were randomised) and 0.5 (50% of samples within the signal were randomised). Three levels of degradation were therefore prepared: 0%, 30% and 50%.

6.3.4 Design and procedure

A repeated measures design was used as all participants undertook the recall task in all auditory conditions. There were two repeated measures factors, type of sound (C-O-C and V-O-C speech) and degradation level (0%, 30% and 50% noise). Therefore, participants undertook 6 different conditions. There were 90 trials in total, 15 for each condition. Presentation of conditions was quasi-randomised from trial-to-trial, such that every condition was presented before any one was repeated. As there was a large number of trials the experiment session was split into 3 blocks of 30 trials each with a break after each block controlled by the participant. Participants began each block of trials and moved through trials by pressing the 'space bar'. The presentation of the sounds and digits and the general procedure for the memory task was as detailed in chapter 3.

6.4 RESULTS

Experimental condition	*Mean Errors*	*SD*
Clear V-O-C speech	25.77	17.254
Low degraded V-O-C speech	22.63	16.431
High degraded V-O-C speech	23.33	15.799
Clear C-O-C speech	21.17	15.430
Low degraded C-O-C speech	22.67	17.835
High degraded C-O-C speech	21.10	16.359

Table 10. Descriptive statistics for six experimental conditions; mean number of serial recall errors per condition. N = 30.

Upon examination of the descriptive statistics in table 10, it is clear that more recall errors occur under clear V-O-C speech relative to clear C-O-C speech. The data is summarised in figure 13.

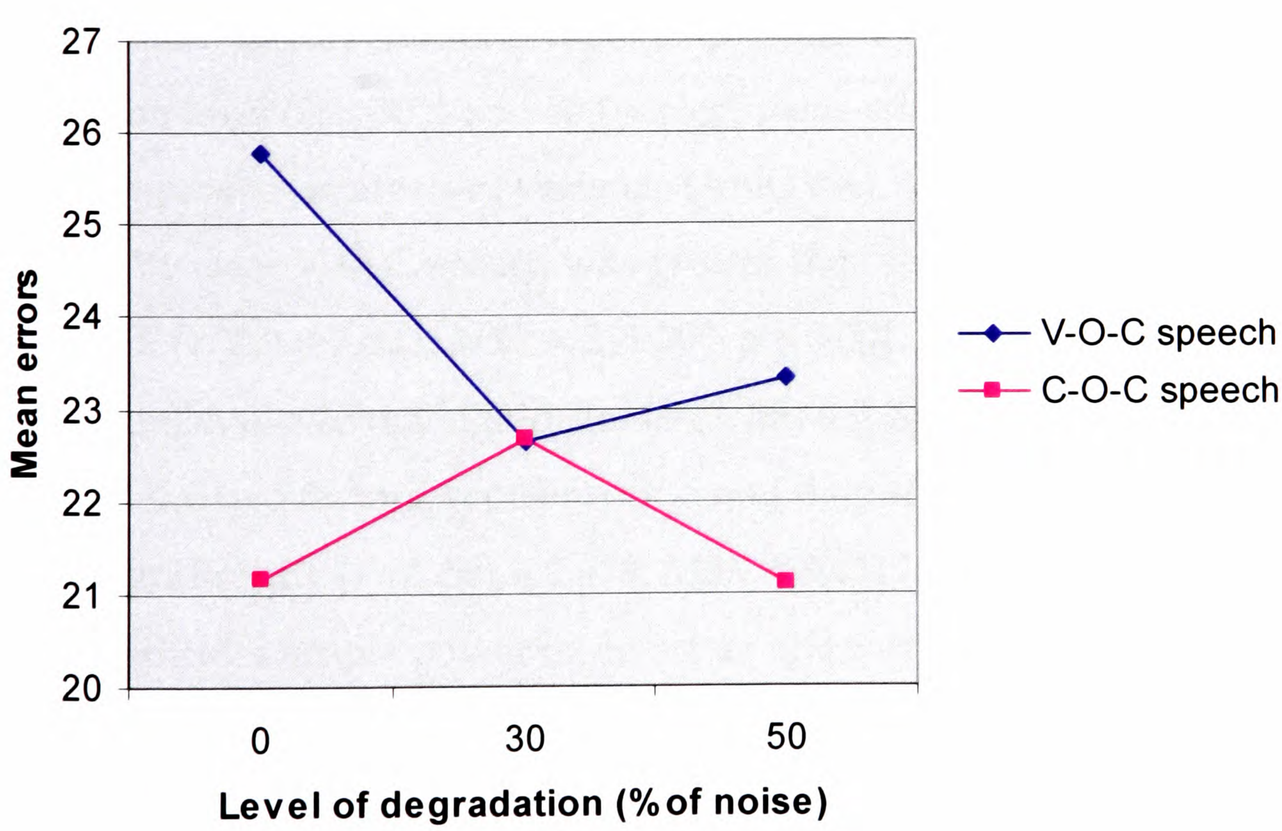


Figure 13. Mean recall errors for Consonant-Only-Changing (C-O-C) and Vowel-Only-Changing (V-O-C) speech sounds as a function of degradation level. Level indicates the percentage (%) of the signal represented by noise.

Inspection of figure 13 clearly shows the line representing the C-O-C speech is relatively flat. That is, it looks as though the level of degradation does not influence recall performance when C-O-C speech acts as irrelevant speech. The line representing the V-O-C speech condition has a negative slope up to the 30% level of degradation, indicating as degradation is increased, the number of recall errors decreases. However, the graph indicates no difference between the disruptive effect of V-O-C speech sounds degraded by 30% and 50% noise. Further, the performance level for serial recall dropped in the presence of V-O-C speech degraded with a SNR of 0.7 (30% noise) to a level that was equivalent to that produced by C-O-C speech. Figure 13 indicates there is a slight trend in one of the predicted directions (see p131). That is the size of the ISE is reduced when V-O-C speech is degraded, whereas degrading C-O-C speech has no effect as changing vowels provide the dominant source of disruption in speech (Hughes et al., 2005).

The two factors of sound type (C-O-C and V-O-C speech) and degradation level (0%, 30% and 50% noise) were subjected to a two-way repeated measures analysis of variance (ANOVA). The disruption produced by clear V-O-C speech was greater than that produced by C-O-C speech [F (1, 29) = 7.822, MSE = 231.200, $p < 0.05$]. The main effect of degradation level was not significant [F (2, 58) = 0.917, MSE = 24.172, $p \leq 0.405$] and the two factors, type of sound and degradation level did not interact significantly [F (2, 58) = 2.479, MSE = 80.517 $p \leq 0.093$]. Power analysis indicates ample power to detect an effect of the type of sound presented (0.771), but not the effect of degradation level on both types of sound (0.201) and the interaction between type of sound and level of degradation (0.479). As revealed by the absence of an interaction between type of sound and degradation level there was no difference between clear V-O-C speech and V-O-C speech degraded at 30% and 50% noise. V-O-C speech degraded at 30% and 50% noise disrupted serial recall to the same extent. Although type of sound and degradation level did not interact significantly, a t-test with bonferroni correction was carried out to see if the difference between the effect of clear V-O-C speech and V-O-C speech degraded at 30% noise (0.7 SNR) approached significance, as is indicated by figure 13. The difference between clear V-O-C speech and V-O-C speech degraded by 30% noise did indeed approach significance ($p \leq 0.062$). The absence of an interaction between type of sound and degradation level also showed that clear C-O-C speech did not differ in its effect on serial memory from C-O-C speech degraded at 0.7 SNR (30% noise) ($p \leq 0.971$) or 0.5 (50% noise) ($p \leq 1.000$). Also, no difference was found between C-O-C speech degraded at 30% and 50% noise (appendix 24).

6.5 THE EFFECT OF CLEAR CONSONANT-ONLY AND VOWEL-ONLY-CHANGING SPEECH COMPARED TO THE POOLED EFFECT OF THEIR DEGRADED VERSIONS.

Analysis of the data examined the effect of clear V-O-C speech and C-O-C speech compared to the pooled effect of the two levels of degradation on both types of sound. The data are summarised in figure 14.

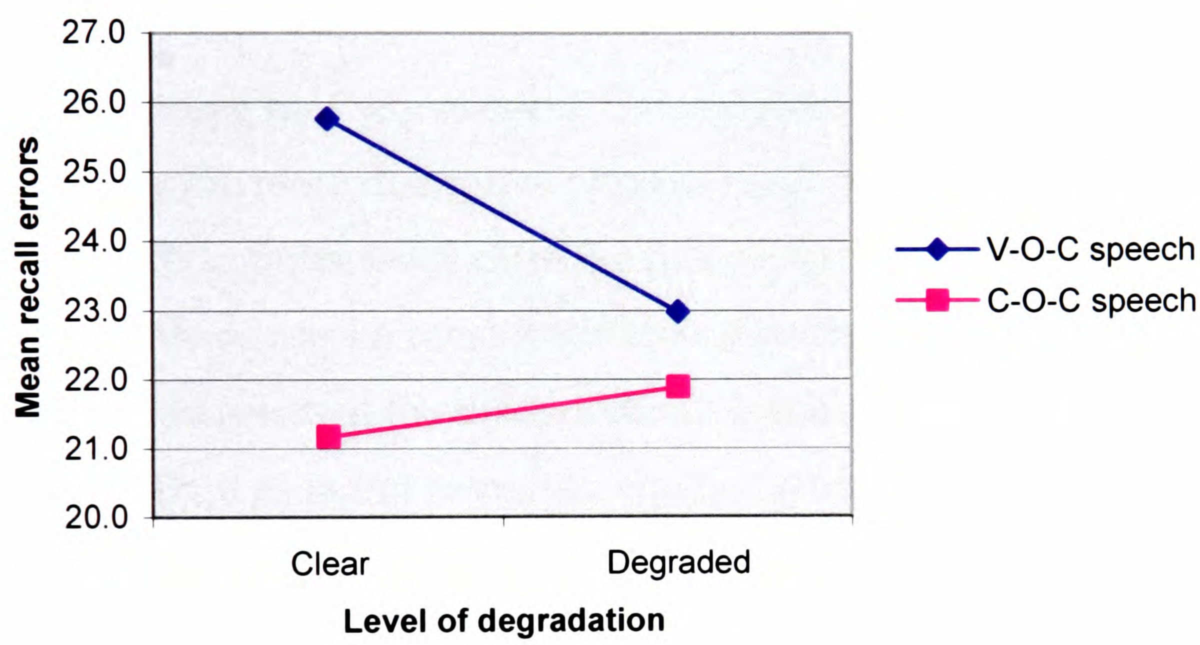


Figure 14. Mean recall errors for clear and degraded consonant-only-changing and vowel-only-changing speech sounds.

A two-way repeated measures ANOVA revealed a significant main effect of sound type [F (1, 29) = 10.205, MSE = 243.675, $p < 0.01$]. The main effect of level of degradation was not significant [F (1, 29) = 1.747, MSE = 32.033, $p \leq 0.197$] and the interaction between type of sound and degradation level approached significance [F (1, 29) = 3.410, MSE = 91.875, $p \leq 0.075$] (appendix 25).

The finding of the interaction between type of sound and degradation level approaching significance may well have been due to a lack of statistical power and therefore, a follow up analysis was conducted using tests of simple main effects for the pooled effect of degradation level on V-O-C and C-O-C speech tokens. A significant difference was found between clear and degraded V-O-C-speech [t (29) = 2.274, $p < 0.05$], but there was no difference between clear and degraded C-O-C speech [t (29) = -0.581, $p \leq 0.566$] (appendix 26). Therefore, it is clear that the effect of degradation is stronger within the V-O-C irrelevant speech condition.

6.6 DISCUSSION

The finding that sequences of CVC syllables that exhibit vowel contrasts only are more disruptive of serial recall than sequences made of only consonant contrasts replicates the findings of Hughes et al (2005). The fact that there was no reliable difference in disruption between the C-O-C speech sequences at the three different levels of degradation can be explained in light of recent research. Hughes et al (2005) found no reliable difference between sequences of CVC syllables that demonstrated a change in the initial or final consonant and a steady-state condition of a repeated CVC. However, a significant difference was observed between the steady-state condition and a silent control. If consonant contrasts within an unattended auditory stream do not produce an ISE different from a steady-state condition then degrading this type of sequence would not produce a reduction in the magnitude of the ISE. If it is assumed that performance was already at ceiling for the presentation of clear C-O-C speech sequences, participant's memory performance would not get any better as the level of degradation of this class of auditory stream increased.

According to the CSH, whether the vowels are changing and the consonants remain fixed or the consonants are changing and the vowels remain fixed, both are changing state stimuli. However, Hughes et al's (2005) data suggests that changing consonants do not seem to exhibit the necessary acoustic changes to produce an ISE that is statistically different from a steady-state condition. In addition, the data of the present experiment show that degrading consonants and thereby reducing the amount of acoustic variation a stream of C-O-C speech conveys, has no effect on the ISE. In contrast, vowel contrasting CVC sequences do seem to exhibit the necessary acoustic changes and thus degradation results in the attenuation of disruption produced by their presentation.

With regards to the phonetic content of speech, vowels seem to be of key importance. The results can be explained by the CSH hypothesis which encompasses an 'interference by process' account (Jones and Tremblay, 2000). The CSH argues the magnitude of the ISE is dictated by the extent to which the irrelevant sound sequences automatically generate information about the order of their sound tokens. It has been shown that in order for a sound stream to elicit serial order cues the sequence must demonstrate acoustic change between adjacent items (e.g. Jones et al., 1992). Further, these acoustic variations must occur on a common ground (Jones et al., 1999a; Jones et al., 1999b). One example emphasising the importance of change occurring on an attribute common to the irrelevant sounds is when the spatial location of the irrelevant sounds is manipulated. If acoustic variation occurs at different spatial locations so that one sound is presented to the left ear, one at the centre of the head and one at the right ear, the level of interference is not only reduced, but reduced to a level that does not differ statistically from a steady-state condition (e.g. Jones and Macken, 1995a). The reduction in the size of the ISE is observed because when sounds are presented such that they traverse different spatial locations, the sequence of sounds segregate and multiple streams of identical repeated (steady-state) sounds are formed. Therefore, if sounds are perceived as coming from the same location in space they will form a coherent changing-state stream, a condition necessary for the observation of an ISE (Jones et al., 1999b).

Hughes et al (2005) suggest the importance of change between adjacent items occurring on an attribute common to the sounds in an irrelevant sequence may explain the different effect vowel and consonant variations have on serial recall. Research into the perceptual organisation of speech sounds indicates that the integration of speech sounds produced by the same voice over time is afforded by a similarity between the periodic phonated vowel sounds (Bregman, 1990). Hughes et al (2005) argue this common ground shared by the irrelevant speech sounds could be provided by an attribute of the changing vowels, for example the fundamental frequency (*f*0) and formant structure of speech tokens produced by the same talker over time.

Further evidence supporting that change on a common ground, afforded by the periodic vowels, aids the perceptual system to maintain the temporal order of sounds is provided by research into the temporal order judgement of syllables. There is evidence for the importance of vowel transitions in maintaining the temporal order of speech sounds in natural speech. It has been demonstrated that listeners are better able to judge the temporal order of a series of auditory loops of Consonant-Vowel (CV) syllables conveying vowel transitions. The frequency transitions into and out of vowels are the product of vocal tract glides from one place of articulation to another (Moore, 2004). Each consonant phoneme exhibits invariant shaped noise, which provides a cue to place of articulation and aids in their perception and is accompanied by a vowel transition. These frequency transitions aid in the perceptual organisation and integration of speech sounds produced by the same voice over time (Moore, 2004). If the consonant noise (energy) is spliced onto the steady-state portion of the CV syllable and the vowel transition is removed then although the sounds when heard in isolation are intelligible, when repeated in an auditory loop the consonant noise segregates from the steady-state vowel after two or three repetitions (Cole and Scott, 1973).

Cole and Scott (1973) played taped loops of 4 sounds to listeners and asked them to judge the order of the sounds. Three types of auditory loops were used (1) consonant noise removed from Consonant-Vowel (CV) syllables (2) CV syllables consisting of consonant noise and the first 75 msec of the vowels, including vowel transitions and (3) CV syllables without the vowel transitions made of consonant noise and 75msec of steady-state vowel. Listeners made more errors in judging the order of tranistionless syllables than they did when presented with normal CV's of the same duration. However, the order of auditory loops composed of consonant noise only was judged most inaccurately. The CSH predicts that it is the ease with which streams of irrelevant sounds generate serial order cues that dictate their disruptive potency (Jones et al., 1999a; Jones et al., 1999b; Jones and Tremblay, 2000). In essence the vowel transitions within the CVC syllable sound stream will help preserve the order of the CVC syllables within the sequence. These transitions hold adjacent segments of sound which differ in their spectral characteristics, such as vowels and consonants. However, although vowel transitions are present within the C-O-C CVC streams of the present experiment, it is reasonable to assume that this perceptual aid to temporal order maintenance is not enough to maintain the magnitude of the ISE since the vowels do not differ from item-to–item and C-O-C sequences are not as disruptive as V-O-C sequences. Therefore, it may be that frequency transitions into and out of vowels that change from item-to-item help the perceptual system to maintain the order of V-O-C items in irrelevant sequences of speech.

Research into the serial recall of V-O-C and C-O-C sequences has shown that V-O-C sequences are recalled in their serial order better than C-O-C sequences (Suprenant and Neath, 1996). Hughes et al (2005) suggest that the recall advantage for sequences of speech items in which the vowels change from item-to-item is the product of stronger serial recall cues elicited when vowels are rehearsed. The pre-attentive obligatory encoding of irrelevant order information is argued to conflict with the retention and maintenance of the order of the TBR items (e.g., Jones and Tremblay, 2000). That vowels have been found to provide more cues to their serial order than do consonants can account for why V-O-C speech is more disruptive than C-O-C speech. The degradation of V-O-C speech may lead to the degradation of serial order cues elicited by changing vowels, which would account for why in the presence of degraded vowels serial recall performance dropped to a level similar to that observed under clear and degraded C-O-C speech.

6.7 ACOUSTIC FEATURES OF CONSONANTS AND VOWELS

The finding that vowel changes are more disruptive of immediate memory than consonant changes may also be explained with reference to the difference in their "acoustic structure" and how they are processed by the auditory system. The acoustic cues that distinguish consonants and vowels differ. The acoustic patterns of vowels vary broadly along steady-state acoustic cues, such as formant frequency and in the case of natural vowels; these additionally vary along rapidly-changing acoustic cues (Schouten and Van Hessen, 1992). Stop-consonants, in contrast, convey rapidly-changing transient acoustic cues and are therefore mostly distinguished by fine temporal distinctions such as Voice-Onset-Time (VOT) and rapid formant transitions (Mirman, Holt and McClelland, 2004). In the case of CVC syllables these rapid formant transitions occur into and out of the vowel portion of the syllable. It follows, that an auditory sequence consisting of only vowel contrasts may produce more interference than sequences depicting contrasts in consonants only because of general differences between processing steady-state and rapidly-changing acoustic cues as well as their subsequent duration as representations in working memory.

6.8 PROCESSING OF RAPIDLY-CHANGING AND STEADY-STATE ACOUSTIC CUES.

The processing of rapidly-changing sounds is lateralised predominantly in the left hemisphere compared to the processing of steady-state sounds, a finding supported by studies into human perception of different types of speech stimuli (e.g. Allard and Scott, 1975) and one that is analogous to that seen with perception of non-speech stimuli (Belin et al., 1998). This auditory processing difference between sounds defined broadly by steady-state information and those defined by rapidly-changing acoustic cues has been explained in terms of different temporal integration windows in the left and right auditory cortices (Poeppel, 2003). Poeppel proposes a short temporal integration window (20-40ms) is used during left auditory cortical processing. In contrast, auditory cortical processing in the right auditory cortex uses a longer temporal integration window (150-300ms). The processing of rapidly-changing cues is left hemisphere dominant as these cues require a shorter temporal integration window, whereas analysis of steady-state information requires a longer temporal integration window and is primarily performed by the right hemisphere.

When pairs of vowels are presented concurrently to both ears they show a left-ear (right-hemisphere) advantage, whereas consonants show a right-ear (left-hemisphere) advantage (Shankweiller and Studdert-Kennedy, 1967). This relates to the work of Hadlington, Bridges and Darby (2004) and Hadlington, Bridges and Beaman (2006) who demonstrated a left-ear disadvantage (LED) for the presentation of irrelevant information, by varying the spatial location in which the irrelevant sounds were presented. The finding that irrelevant tones and spoken consonants presented to the left ear was significantly more disruptive of serial recall than that presented to the right ear, indicates that the right hemisphere plays an important role in the ISE. This LED for speech and non-speech sounds (Hadlington et al, 2004; 2006) may be explained with reference to the fact that these stimuli demonstrated steady-state acoustic cues which have been shown to be processed more efficiently in the right hemisphere (Poeppel, 2003).

For the speech sounds of the present experiment, the changing vowel information provided by V-O-C syllables did consist of acoustically complex steady-state information, and more importantly, changes in steady-state information between syllables. Therefore the critical item-to-item changes occurring on V-O-C syllables would be processed more efficiently by the right hemisphere, and thus more detrimental of serial recall when presented as irrelevant sound. Again, stronger cues to the temporal order of the irrelevant sounds would be generated by vowel changes. These as a consequence would conflict with cues pointing to the order of the visual digits to a higher degree than would cues elicited by consonant changes.

The difference between the disruptive effect of vowel changes compared to consonant changes can be explained with reference to research findings regarding the serial recall of attended-to vowel and consonant contrasting sequences. Sequences of V-O-C syllables have been found to be better recalled than series of C-O-C syllables (Cole, 1973; Surprenant and Neath, 1996). Several theories have been implemented as frameworks in which to attempt to explain the superiority of vowel changes in memory.

6.9 PRECATEGORICAL ACOUSTIC STORE

The precategorical acoustic Store (PAS) (Crowder and Morton, 1969) is a sensory memory system dedicated to the storage of acoustic information. PAS has been used to account for the recall advantage observed for V-O-C syllables in contrast to C-O-C syllables. PAS acts as a sensory acoustic store which holds acoustic information about the last one or two auditory items of a sequence of sounds in an unanalysed, hence 'precategorical' form. Crowder and Morton (1969) say that the analysed acoustic representations are stored for short periods of time, approximately two seconds. Items decay over time and new incoming items that are acoustically similar interfere with those already present within PAS. Better recall of vowels relative to consonants has been suggested to occur because vowels are preserved in acoustic storage more efficiently than are consonants (Crowder, 1973).

The modality effect refers to the difference in recall observed for auditory and visual item lists. When lists of spoken items are presented, participants remember the first few items (primacy effect) and the last few items (recency effect) better relative to the middle list items (Penney, 1989). The finding that the last few items of a verbal sequence are more likely to be recalled when the sequence is presented in the auditory modality as opposed to the visual modality is referred to as the 'auditory recency effect' (Crowder and Morton, 1969). Cole (1973) found a more pronounced recency effect for V-O-C sequences compared to C-O-C sequences. The C-O-C lists varied only in a stop-consonant. In addition, Cole (1973) also found a more pronounced primacy effect for V-O-C sequences relative to C-O-C sequences. When presented with a sequence of syllables, participants can retrieve vowel information of the final syllables from PAS, but cannot do so as readily for consonants as these decay from the acoustic store more rapidly. Crowder (1973) argues that because consonants (exhibiting transient noise) decay at a faster rate in PAS than do vowels (conveying steady-state information), there will be more information pointing to token identity when stimuli are vowels as opposed to consonants. In support of the faster decay of consonants, but not vowels in PAS, Cole, Sales and Haber (1974) found more accurate recall of spoken vowels than consonants following a delay period of 5 or 15sec of mental arithmetic before the recall phase.

PAS has been shown not to provide an adequate account of auditory memory effects. The suffix effect refers to the reduction in the auditory recency effect when an additional item resembling speech occurs at the end of a TBR list, but which participants are told to ignore (Morton, Crowder and Prussin, 1971). Crowder and Morton (1969) propose information held in PAS decays over time and can be overwritten by new items entering the store. The stimulus suffix is argued to overwrite the last few items held in PAS. However, research shows that the memory effects associated with PAS are not acoustically based, as modality effects have been demonstrated with lip-read stimuli (Nairne and Crowder, 1982). In addition, the stimulus suffix effect has been shown to be context-dependent. Neath, Surprenant and Crowder (1993) presented participants with lists of words followed by the stimulus suffix 'baa'. Participants were either instructed that the sound was produced by a human or a sheep. A larger suffix effect was found when the stimulus suffix was interpreted as speech as opposed to non-speech, even though the suffix was the same physical (acoustic) item in both conditions.

6.10 CATEGORICAL PERCEPTION

An alternative explanation for the maintenance of the auditory memory code for vowels has been drawn from theory relating to categorical perception (Surprenant and Neath, 1996). As this theory provides an explanation for why V-O-C syllables are recalled in serial order more accurately than C-O-C syllables, it can be used to account for the greater disruptive effect of irrelevant sequences featuring changing vowels as opposed to consonants between items. Here, participants participate in two tasks, an identification task and a discrimination task. The sounds used are normally synthesised sounds and are made by altering one parameter, such as voice-onset-time (VOT) in a number of equal steps along a continuum between two speech sounds. VOT is the time interval between the first release of air on production of a consonant and the start of voicing, the duration of which is systematically lengthened along the continua. For example, two stop consonants would be arranged at either end of a continuum and small acoustic changes occur from one sound into the next (Pisoni, 1973). During the identification task, listeners are asked to classify each sound into one of two categories. That is, when each acoustically varied sound is presented, it is either heard or categorised as one sound or the other. A discrimination curve is generated on the basis of discrimination judgements regarding whether two sounds are the same or different during the identification task (Pisoni, 1973).

The discrimination task often used is an ABX discrimination task. Listeners hear two sounds, sound A and B and are then asked to decide whether a third sound presented, sound X is the same as sound A or B. Observed discrimination of stimuli is then compared with discrimination performance predicted by categorisation data. Stop consonant continua exhibit sharp categorisation boundaries and their discrimination is well predicted by categorisation (Liberman et al., 1957). It follows that good discrimination performance occurs across the category boundary as these stimuli are always given different category labels. For consonants, continua of sounds demonstrate categorical perception because the sounds within a category (determined during the identification task) are discriminated at chance levels. Therefore, the stimuli along the continua are grouped into different categories and predict poor within-category discrimination performance. This is because all acoustic stimuli of a category are given the same category label (Liberman et al., 1957).

Those sounds classified into different categories, as demonstrated by the perception of vowels, are easily discriminated. Vowels are therefore perceived relatively continuously, even though the physical differences between sounds along the continuum (within and between categories) are the same. Vowels bring about categorisation functions that are not as sharp as that observed for stop consonants. For vowels, observed discrimination of stimuli identified across the stimulus continuum as belonging to different categories exceeds discrimination performance predicted by categorisation (Pisoni, 1975). This is due in part to the ambiguous nature of vowel sounds and so when two vowel sounds are presented they will sometimes be labelled as the same sounds or different sounds (Pisoni, 1973).

The categorical perception of stop consonants that are defined by transient acoustic cues and the relatively continuous perception of vowels which are characterised broadly by steady-state information have been found with non-speech stimuli (Mirman, Holt and McClelland, 2004). It follows that categorisation of non-speech stimuli that vary on a rapidly changing cue demonstrate a sharp category boundary whereas stimuli varying along a steady-state cue are less well categorised, but better discriminated. Non-speech stimuli which varied along both transient and steady-state cues, reminiscent of natural dynamic vowels demonstrated a sharp category boundary and were accurately discriminated (Mirman, Holt and McClelland, 2004).

A dual-code theory (c.f. Pisoni, 1973) explains the categorical perception of consonants relative to the continuous perception of vowels by assuming the existence of two memory codes, an auditory and a phonetic memory code when a speech sound is heard. To allow the listener to discriminate between two speech sounds, the phonetic codes of the stimuli have to be compared first. If the phonetic codes are different the listener successfully discriminates between the two sounds. However, if they are the same, as in the sounds belong to the same category; the listener has to rely on the auditory code to make the discrimination. Consonants are categorically perceived because they convey transient noise and so the auditory store in memory can only make use of phonetic identity cues in order to discriminate between stimuli. In contrast, the steady-state information of vowels is held for longer within the auditory store and is more resilient and useful to a listener when making within-category discriminations.

6.11 THE RELATIONSHIP BETWEEN DISCRIMINABILITY AND SERIAL RECALL PERFORMANCE

Surprenant and Neath (1996) showed that the link between discriminability and how well items are recalled is not straightforward. They presented participants with CVC syllables in which only the vowel changed or a series of syllables in which only the initial stop consonant changed and the vowel and final consonant remained fixed. In addition, due to the acoustic differences between consonants and vowels, a series of silent-centre CVC syllables were presented. The silent-centre syllables were the same as the vowel-only contrasting syllables, but the steady-state part of the speech signal was removed by the centre part of the vowel being replaced by silence. Silent-centre syllables consist of an initial and final consonant and the rapidly changing formant transitions into and out of the vowel, but practically no steady-state information (Surprenant and Neath, 1996). The series of experiments conducted featured an identification phase and a recall phase. No difference between the identification of consonants and silent-centre syllables was found and there was no difference in overall recall of these items. Identification and recall of the vowels was statistically better (Surprenant and Neath, 1996).

The discriminability of consonants and vowels was also manipulated, so that consonants were identified better by manipulating the intensity of noise added to them (Surprenant and Neath, 1996). Serial recall of V-O-C syllables was slightly better than C-O-C syllables, though this difference was not statistically reliable and so recall of V-O-C syllable sequences was statistically equivalent to that of C-O-C syllable sequences. The identification of vowels and consonants was better than identification of the silent-centre syllables, and the recall of the silent-centre syllables was worse than that of V-O-C syllables but did not differ from C-O-C syllables. In addition, Surprenant and Neath (1996) manipulated the stimuli so that the silent-centre syllables would be identified better than the V-O-C syllables. Even so, recall of V-O-C syllables was better than silent-centre syllables, though the difference was not reliable. This lead Surprenant and Neath (1996) to assume memory performance is predicted by the relative discriminabilty of the memory code of an item as opposed to the relative discriminability of the physical stimulus itself, which is the foundation for Nairne's (1990) feature model.

The feature model proposes the TBR items consist of modality-dependent features, which represent presentation modality and acoustic information and modality-independent features brought about by identification and categorisation of the items themselves. The modality-independent features of items presented visually or auditorily will be equivalent. The number of modality-dependent features is assumed not to differ for consonants and vowels (Nairne, 1988). Surprenant and Neath (1996) explain the slightly better serial recall of less discriminable V-O-C syllables relative to C-O-C syllables within the framework of Nairne's (1990) feature model. As far as identification performance goes, with the addition of more noise to an auditory token, its modality-dependent features are degraded. Serial recall of these items is therefore nearly entirely dependent on modality-independent features. As Surprenant and Neath (1996) demonstrated V-O-C syllables were discriminated less well than the C-O-C syllables when more noise was added to these items. However, V-O-C syllables were better serially recalled, though this difference was not reliable. A more resilient representation of items is required in memory. Surprenant and Neath (1996) argue that when identifying the vowel-only contrasting syllables during the identification phase the corresponding memory representations generated, represented by the modality-independent features, were themselves more discriminable. Surprenant and Neath (1996) assume the verbal label or name of the TBR syllables reflect the memory representations and suggest the *verbal labels* of the C-O-C syllables are *less discriminable* than the *verbal labels* of the V-O-C syllables. However, the difference between serial recall performance of V-O-C syllables and C-O-C syllables was not significant. Therefore, the advantage in serial recall found for V-O-C speech streams may be a product of the discreteness demonstrated by the vowel changes which is reflected by the modality-dependent features.

Nairne's (1990) feature model can in turn provide an explanation for the different effect of task-irrelevant auditory V-O-C and C-O-C syllables on immediate serial recall. The same number of modality-dependent features should exist for vowels and stop consonants (Nairne, 1988). But, the similarity of the modality-dependent (physical) features defining stop consonants will be greater, rendering them less discriminable. Vowels on the other hand convey more discrete modality-dependent cues and are better discriminated (Nairne, 1988). This would explain reduced serial recall performance for consonants relative to vowels (Surprenant and Neath, 1996) and explain less serial recall interference by C-O-C speech relative to V-O-C speech. Consonants, which are less discrete, will generate weaker cues to their serial order and therefore conflict less with the seriation of TBR items.

If modality-dependent features of V-O-C syllables that have been degraded by noise are not as useful this may account for why serial recall performance in the presence of degraded V-O-C syllables was reduced to a level that was similar to that observed in the presence of clear C-O-C syllables. Surprenant and Neath (1996) use the feature model to explain the difference in consonant and vowel perception observed in the categorical perception paradigm. Categorical perception of consonants occurs because the physical modality-dependent features they convey are not as useful for discrimination as they are for vowels. A small change acoustically will result in an altered perception of vowel identity, whereas only around the category boundary will a similar change alter the perception of a consonant. Degradation of the vowels in the V-O-C syllables in the present experiment may have degraded the modality-dependent features of the vowels such that they became less distinct and similar to those of consonants which are not as useful for discrimination. It follows that the memory representations of degraded V-O-C syllables may have been less durable in short-term memory (STM) and similar to C-O-C item representations. As a consequence, the CSH would argue that weaker cues to the temporal order of the degraded V-O-C sounds are generated and seriation of the visual digit lists is disrupted at a degree equivalent to that observed in the presence of irrelevant C-O-C syllables. By this argument, it is the acoustic-based discriminabilty of the memory representations which is important in determining the strength of serial order cues elicited by irrelevant sounds.

6.12 SUMMARY

In the context of the present findings, irrelevant auditory sequences exhibiting vowel contrasts only may be more disruptive than sequences depicting only consonant contrasts because the changing steady-state information across the changing vowels in the stream of CVC syllables is processed more efficiently in the right hemisphere (Poeppel, 2003). It is the right hemisphere that research identifies as important in the occurrence of the ISE (Hadlington et al., 2004; 2006).

The greater disruptive potency of changes only in the vowels of CVC syllables can also be accounted for with reference to the change-on-a-common-ground rule. For an irrelevant sound stream to elicit serial order cues the sequence must demonstrate acoustic change between successive sounds (Jones et al., 1992). These acoustic variations must occur on a common ground (Jones et al., 1999a; Jones et al., 1999b). If acoustic variation between irrelevant sounds exceeds a threshold of change, such that the sounds are no longer perceived as one coherent stream of discrete sounds, but as separate streams of identical sounds, recall disruption is attenuated to a level equivalent to that found in a steady-state condition (e.g. Jones and Macken, 1995b). The perceptual system when integrating speech sounds produced by the same speaker over time takes advantage of change occurring on a common ground provided by the periodic vowel sounds and not the noisy aperiodic consonants. This common ground in speech refers to a similarity shared by the vowels, which Hughes et al (2005) suggest is a common fundamental and/or formant structure.

Also within the framework of the feature model, vowel contrasting syllables may result in more durable representations in memory. If V-O-C syllables are clear and not subjected to degradation, Surprenant and Neath (1996) argue they will have more useful modality-dependent (physical) features as they are easily discriminated during identification tasks. That vowels are more easily discriminated may account for the better serial recall of V-O-C sequences relative to C-O-C sequences. Degrading V-O-C syllables may have served to degrade the modality-dependent features of the vowels, rendering them as useful as those of consonants, which are discriminated less well (Surprenant and Neath, 1996).

Degraded V-O-C speech disrupts serial recall performance to a level equivalent to that observed in the presence of both clear and degraded versions C-O-C speech. This suggests that a degradation of the acoustic information of vowels reduces their discreteness and their ability to disrupt serial recall. The finding of no reliable difference between both levels of degraded vowel contrasting sequences can be explained by the fact that vowels are more resistant to degradation by signal-correlated-noise (SCN). As degradation had no effect on consonants, even when the mean serial recall errors for both degraded versions were pooled suggests that serial recall performance was at ceiling and was not improved by degrading C-O-C speech items. However, pooling the data from the degraded V-O-C conditions resulted in a reliable difference between clear and degraded V-O-C sequences. Again, degrading the V-O-C speech reduced its disruptiveness to a level that was similar to that seen in C-O-C speech. Vowels therefore seem to be the most important phonemic component to change within the irrelevant auditory stream and it may be that phonological degradation has an effect on the relative discriminability of the physical features of irrelevant vowel sounds and their non-echoic representations formed in memory.

CHAPTER 7

7 ROLE OF FORMANT CHANGES BETWEEN SPEECH SOUNDS: SERIAL RECALL DISRUPTION BY VOICED SPEECH AND WHISPERED SPEECH

7.1 BACKGROUND

Experiment 2 (chapter 6) replicated the findings of Hughes et al (2005) as vowel-only-changing (V-O-C) syllables disrupted serial recall more than consonant-only-changing (C-O-C) syllables. In addition, degrading V-O-C syllables reduced their disruptive effect on serial recall to a level that was similar to disruption observed with C-O-C sequences. Hughes et al (2005) suggest that in light of the finding that V-O-C syllables are better recalled in serial order than are C-O-C syllable (Surprenant and Neath, 1996), V-O-C syllables are more disruptive of serial recall because they provide stronger cues to their serial order.

The present experiment aims to examine the effect of voiced speech and whispers on immediate serial recall. Whispered speech does not convey the periodic information that voiced speech does. It follows that acoustically, voiced and whispered speech differs in a number of respects. In order to make these acoustical differences apparent, it is useful to consider the mechanisms and physiological structures used during speech production when speech is voiced and whispered.

7.2 SPEECH PRODUCTION

The airways of the mouth, noise and pharynx form the supralaryngeal vocal tract (see figure 15). This acts like a changeable acoustic filter as the speaker changes the shape of their vocal tract during speech production and filters the source of energy (air flow) provided by the subglottal component and workings of the larynx (Yost, 2000). This set of filters are set in motion by a pulsating sound pressure wave due to the workings of the vocal folds (Yost, 2000) and provides speech with its acoustic properties, that of spectral and temporal changes. More acoustic energy is allowed through at particular frequencies as the vocal tract resonates, known as formant frequencies. Formant frequencies are a result of the length and shape of the vocal tract (Morris and Clements, 2002).

Figure 15 has been removed from the digitized thesis for copyright reasons.

7.3 ACOUSTIC CHARACTERISTICS OF VOICED SPEECH

Voiced sounds are periodic sounds. The vocal folds of the larynx modulate the flow of air from the lungs as they vibrate and the *source* of the periodic sound is at the glottis, the horizontal space between the vocal folds (see figure 15). As the vocal folds vibrate the resonant frequency ranges of the vocal tract are excited and therefore the vocal tract acts as a *filter* of the sound (Morris and Clements, 2002). The same pattern repeats regularly and is almost the same throughout the waveform. The 'period' refers to the duration of one complete cycle of the pattern of a periodic waveform. The rate at which the vocal folds open and close as a product of vocal fold vibration at the larynx determines the period, and therefore the fundamental frequency ($f0$) of the air flowing through the glottal constriction (Moore, 2004). Therefore, $f0$ reflects the frequency of the glottal pulses and is the acoustic consequence of vocal fold vibration. The $f0$ is usually the lowest frequency; hence the fundamental in a complex signal and the perceptual correlate of $f0$ is pitch (Moore, 2004). It follows, that as a result of faster glottal pulses, the higher the pitch the higher the $f0$.

The periodic glottal waveform has energy only at the fundamental frequency and its harmonics, which are integer multiples of the fundamental frequency. For example if $f0$ is 100Hz, then the subsequent harmonics would be 200Hz, 300Hz and so forth (Moore, 2004). Formants are the energy peaks of the signal that determine the quality of voiced sounds, for example vowels. Vowel quality, which refers to the perceptible difference between vowel sounds, is determined by the formant frequencies of the vowels (Lieberman, Blumstein, 1988). Formants or formant areas as they are sometimes called are the result of the vocal tract amplifying periodic sound at its resonant frequencies (Morris and Clements, 2002). These formant frequencies can be harmonics, but this is not necessarily the case as formants may be the product of overtones. Whereas harmonics are integer multiples of $f0$, an overtone is any frequency above $f0$ (Moore, 2004). Therefore, all harmonics are overtones, but not every overtone of a complex is a harmonic. In speech sounds, the periodic sections are harmonically complex and include highly variable distributions of energy over the harmonics and formants produced by the vocal tract resonances (Lieberman, Blumstein, 1988).

7.4 ACOUSTIC CHARACTERISTICS OF WHISPERED SPEECH

In contrast to voiced speech, whispers are aperiodic and phonetically voiceless. The pharynx is shaped so that the vocal folds do not vibrate (Ito, Takeda and Itakura, 2005). During voicing, the posterior component of the glottis is closed and phonation is at the anterior section. In whispered speech the posterior section or the whole of the glottis is left open and the source of sound is the noise provided by exhaled air turbulently flowing through the glottal constriction (Lieberman and Blumstein, 1988). The sound source tends to be spread across the lower region of the vocal tract with power, which is 20dB lower than voiced speech (Jovičič and Dordevič, 1996). As the vocal folds do not vibrate the waveform of a whisper does not convey a regular repeating pattern and therefore there is no period, and fundamental frequency $f0$ is eliminated from the signal (Morris and Clements, 2002).

Figure 16a and 16b display the waveforms of the speech signal for the monosyllabic non-word /sof/ (`s5f`) when whispered and voiced (see appendix 5 for examples of disc phonetic symbols). In contrast to voiced speech, the amplitude of vowels is lower than that of consonants in whispered speech. No vocal fold vibration accounts for this reduction in amplitude.

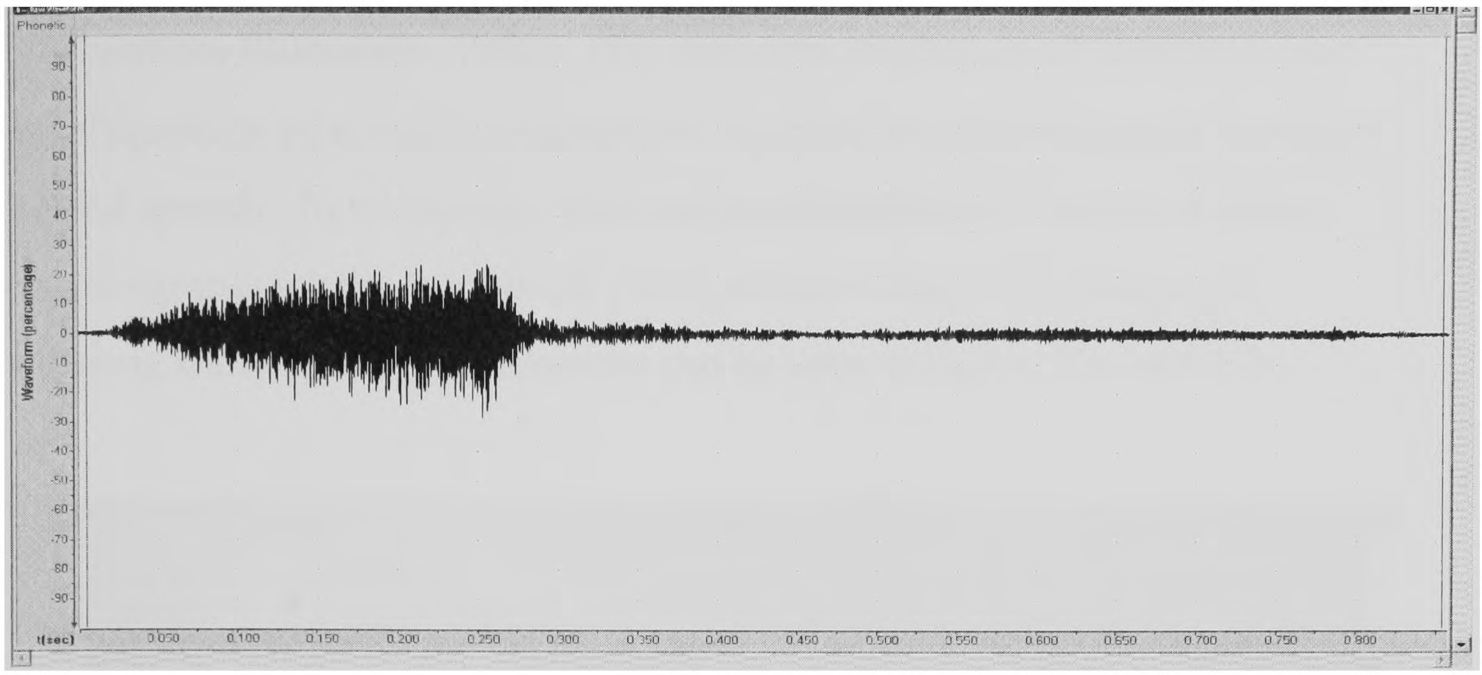


Figure 16a. Waveform of the non-word /sof /when whispered.

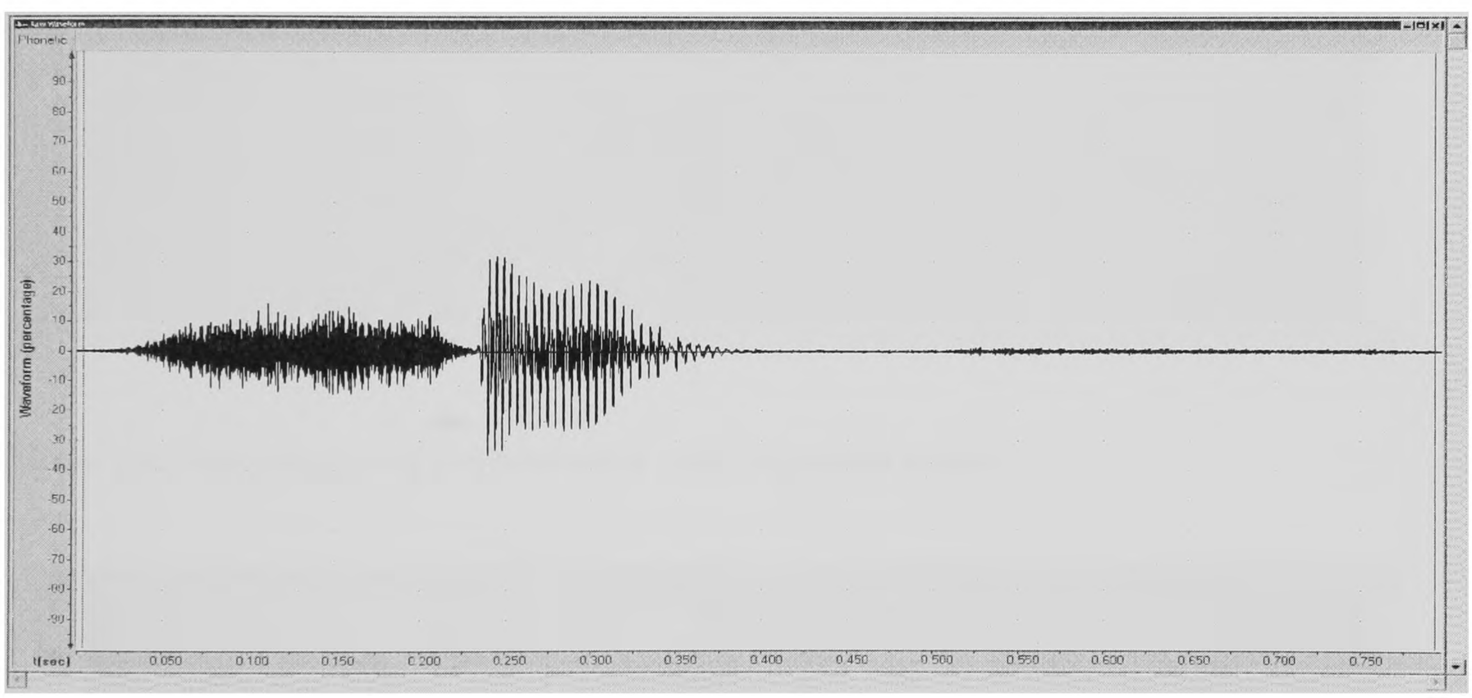


Figure 16b. Waveform of the non-word /sof /when voiced.

As whispers do not convey information regarding pitch, frequency modulation (FM) does not exist. FM refers to fluctuations in pitch within individual tones (Moore, 2004). Instead, the vocal tract shapes broad bands of noise that are excited at its resonant frequencies (formant regions). If the shape of the vocal tract is the same, these formant regions will not change independent of any change in pitch (Lieberman and Blumstein, 1988).

Whispered speech has the formant structure of voiced speech, but not its harmonic fine structure due to the absence of *f*0 information (Lieberman, Blumstein, 1988). The formants in phonated utterances are amplifications of specific frequencies depicted by the horizontal bands in voiced speech. In whispers, these are amplifications of bands of noise. Spectrograms for the non-word /sof/ when voiced and whispered showing these spectral differences can be seen in figure 17a and 17b.

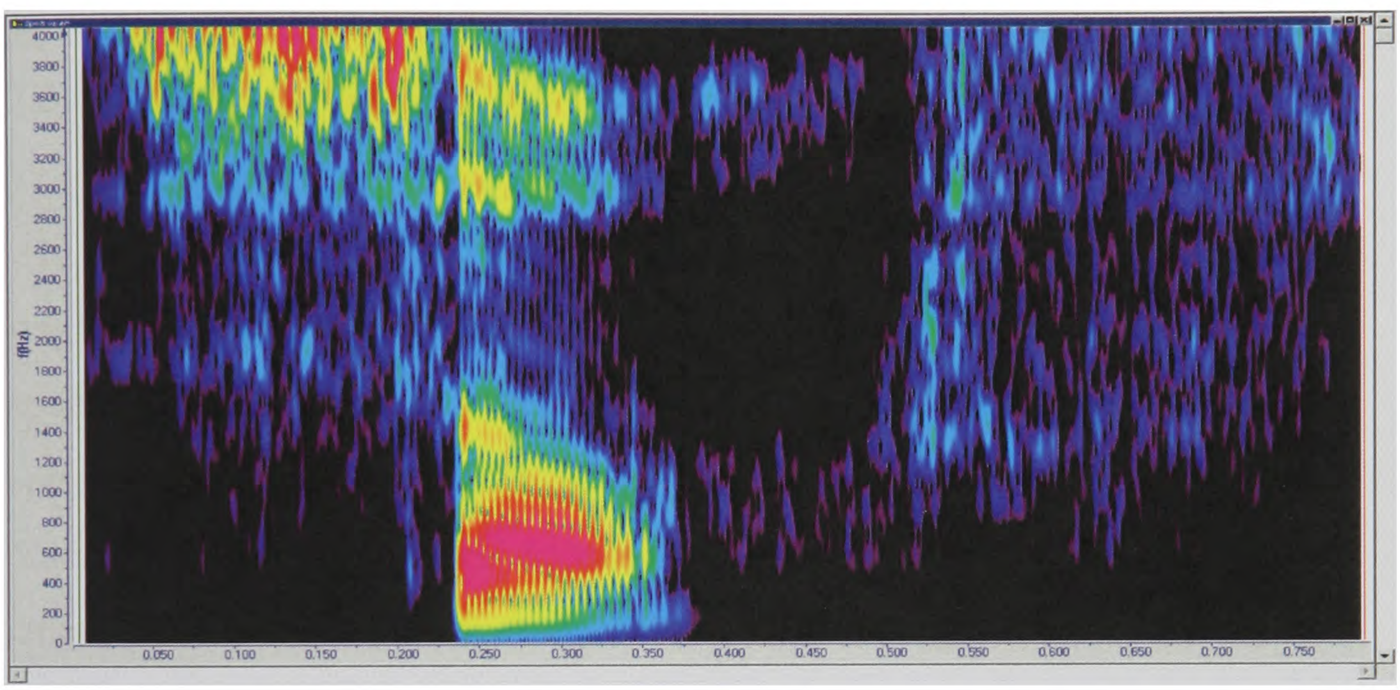


Figure 17a. Spectrogram of the non-word /sof/ in voiced speech.

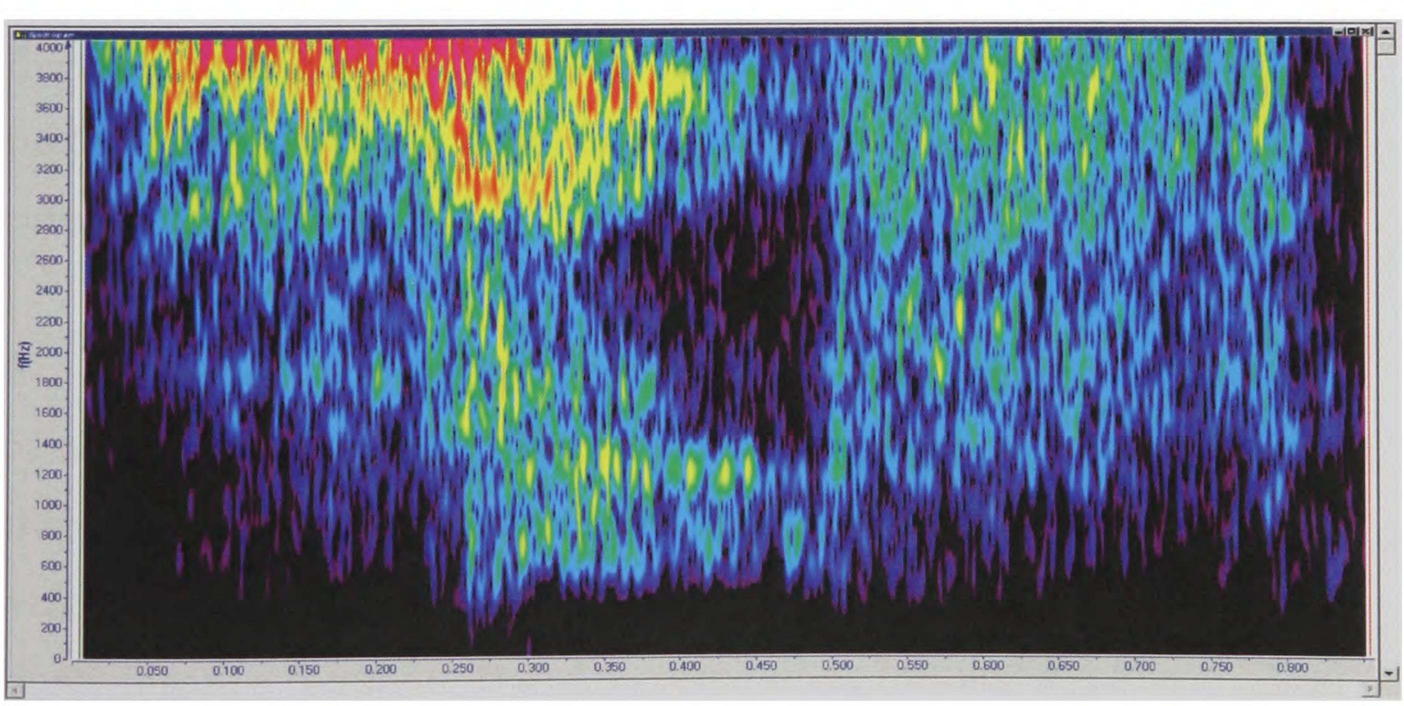


Figure 17b. Spectrogram of the non-word /sof/ in whispered speech.

The absence of $f0$ leaves whispers without pitch information and subsequently, there is no voice characterisation or identity of the speaker. Even though whispered speech is aperiodic and $f0$ is absent, pitch can still be perceived. Variations in perceived pitch have been linked with the formant frequencies and bandwidths (Morris and Clements, 2002). Further, formant shifts, particularly increases in the first formant frequency ($f1$) have been observed (Higashikawa, Nakai, Sakakura and Takahashi, 1996; Jovičić, 1998).

7.5 AIMS AND OBJECTIVES

Voiced speech can be differentiated from whispered speech by it having a richer fine structure due to the quasi-periodic portions of the speech signal produced by phonation. The periodic portions of the speech signal are provided by the vowel sounds and these provide the common ground on which successive items need to change, whilst the consonants provide the aperiodic noisy onsets and offsets of speech sounds. However, the detrimental effect of acoustic changes occurring on this common ground in speech has not been investigated. It is not clear from existing data as to whether the removal of $f0$ information and therefore periodicity of the speech sounds will improve serial recall performance relative to performance observed when periodic sounds are presented.

A sequence of voiced speech sounds would demonstrate more acoustic complexity than a stream of whispered speech sounds. The CSH therefore would predict the acoustic links between distinct adjacent voiced items within the irrelevant stream will be stronger due to change occurring on a common $f0$ and corresponding formant structure shared by the voiced sounds. In contrast, a weakening of the acoustic cues between sounds when the common fundamental of voicing, provided by

$f0$ information, is absent would be predicted when the irrelevant speech sounds are whispered.

The periodic information provided by the vowels is lost when speech sounds are whispered (Morris and Clements, 2002). The higher disruption of serial recall in the presence of V-O-C speech relative to C-O-C speech (see experiment 2, chapter 6; Hughes et al., 2005) was attributed to vowels providing more serial order cues than do consonants. Therefore, if whispered the vowels of speech sounds may be less disruptive of serial recall than those of voiced speech. In addition, whispered speech conveys a weaker formant structure in comparison to voiced speech which has a richer fine structure due to harmonicity as a consequence of vocal fold vibration. Experiment 3 investigates the above predictions of the CSH by examining the importance of $f0$ information. The relative disruptive effect on serial recall of presenting sequences of voiced speech sounds is compared to that observed in the presence of whispered speech sounds.

7.6 EXPERIMENT 3: METHODOLOGICAL CONSIDERATIONS

7.6.1 Participants

30 participants took part in the study. All reported normal hearing and normal or corrected to normal vision. All participants had English as their first language and were not paid for their time.

7.6.2 Stimuli

7.6.2.1 Visual stimuli

Lists of digits to be recalled were constructed as outlined in chapter 3 (appendix 1).

7.6.2.2 Auditory stimuli

Seven non-words for both auditory conditions in this experiment were recorded digitally, edited and presented as detailed in chapter 3 (p90). The non-words for both the voiced and whispered speech conditions are displayed in appendix 27. The same non-words were used in both speech conditions. The intelligibility of the whispered speech sounds was screened in a pilot listening session of which 12 listeners took part. All seven whispered non-words were identified correctly. This experiment examined the effect of manipulating acoustic complexity between irrelevant speech conditions (voiced and whispered speech) whilst maintaining intelligibility. Therefore it was important that all seven whispered non-words were 100% intelligible, so that the intelligibility of stimuli in the voiced and whispered speech condition was matched.

During whispering, the sound source is spread across the lower region of the vocal tract. As a consequence whispers are totally noise excited with 20 dB lower power than their voiced counterparts (Jovičič and Dordevič, 1996; Morris and Clements, 2002). In order to equate the Root Mean Square (RMS) sound level of the voiced and whispered speech sounds, the mean RMS sound level was calculated in 50msec windows for both classes of speech sound. The RMS levels were matched by amplifying the whispers and attenuating the sound level of the voiced sounds, using Cool Edit Pro 1.2 (Syntrillium Software Corporation, 2000).

7.6.3 Design and procedure

The design and procedure was as detailed in the general procedural outline of chapter 3 (p93).

7.7 RESULTS

Experimental condition	*Mean Errors*	*SD*
Voiced speech	41.57	25.570
Whispered speech	43.57	28.527
Silence	31.73	28.654

Table 11. Descriptive statistics for the three experimental conditions; mean number of serial recall errors per condition. N = 30.

The descriptive statistics in table 11 reveals an equivalent disruptive effect of voiced and whispered speech relative to silence. The data are summarised in figure 18.

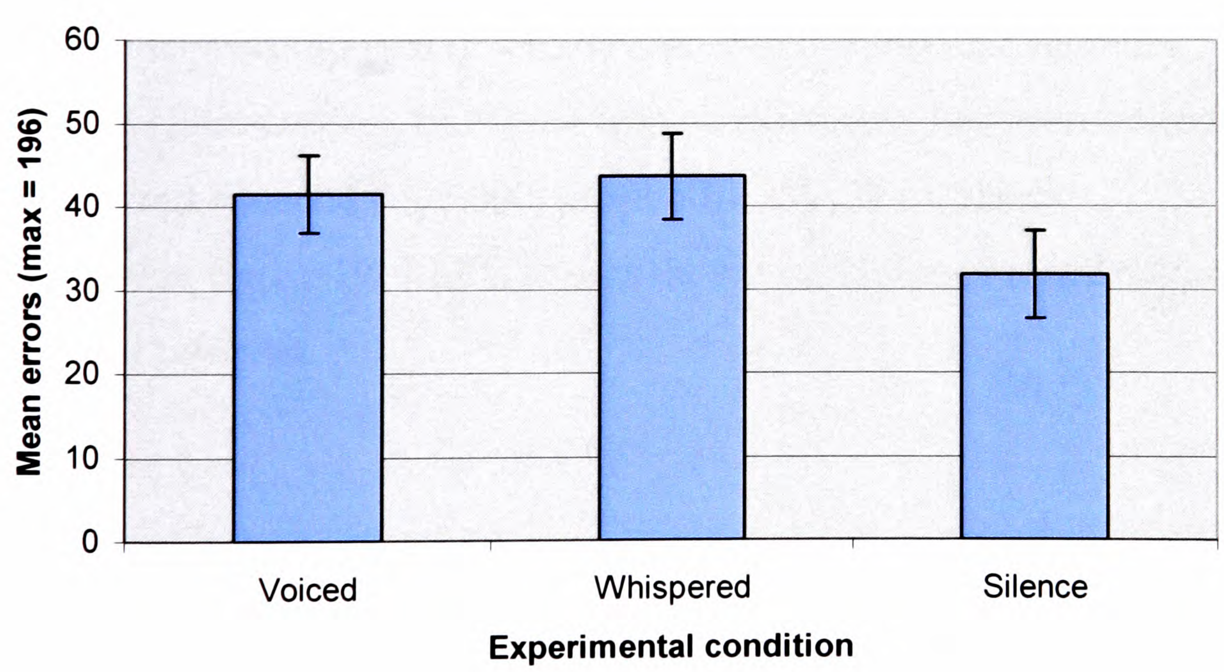


Figure 18. Mean number of serial recall errors for the three experimental conditions. Error bars represent standard error above and below the mean.

	Silence	*Voiced Speech*
Whispered speech	✓ $p < 0.01$	Non-sig $p \leq 1.000$
Voiced Speech	✓ $p < 0.01$	xx

Table 12. Bonferroni corrected pairwise comparisons for the three experimental conditions.

A one factor repeated measures ANOVA was performed on the mean number of digits incorrectly recalled for the three levels of irrelevant sound (voiced speech, whispered speech and silence) (appendix 28). A main effect of sound was found, as serial recall was significantly disrupted by irrelevant speech relative to a silent control (F (2, 58) = 8.213, MSE = 1203.611, $p < 0.01$). Pairwise comparisons with Bonferroni correction as detailed in table 12 show that voiced speech ($p <$ 0.01) and whispered speech ($p < 0.01$) disrupted immediate memory relative to a silent control, but there was no difference between voiced and whispered speech ($p \leq 1.000$) (appendix 28). The data are summarised in figure 19, which shows the overall level of recall collapsed across serial position.

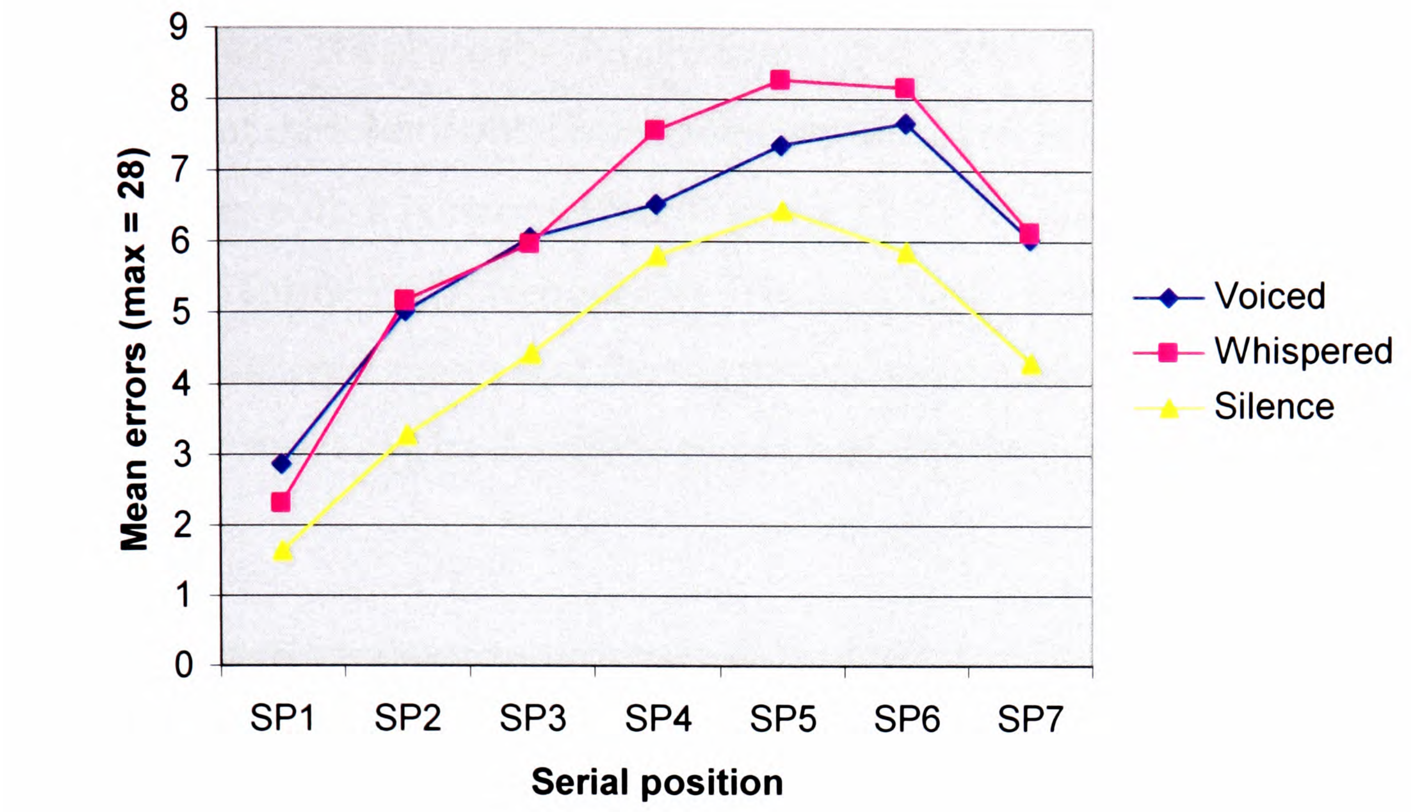


Figure 19. Mean errors for the 3 experimental conditions collapsed across serial position.

7.8 DISCUSSION

The present experiment found that speech whether it is voiced or whispered disrupts immediate memory to the same degree. This finding refutes the prediction of the CSH as whispers demonstrate less acoustic complexity than do voiced speech sounds and yet disrupt memory to an equivalent degree. This result is also inconsistent with the view of Hughes et al (2005) as the absence $f0$ and thus voicing in whispered speech does not attenuate the ISE.

Speech consists of acoustic patterns, which vary over time in frequency and intensity (energy). A spectrogram is a visual representation showing the amount of energy at different frequencies over time and is a plot of frequency over time (Moore, 2004). Figure 20a shows the wideband spectrogram of the voiced monosyllabic non-word /curj/ (k3_) and figure 20b displays the spectrogram for /curj/ when whispered (see appendix 5 for examples of disc phonetic symbols). The dark portions/bands running horizontally represent formants (vocal tract resonances). These are the dominant spectral peaks. The lower prominent dark horizontal band is the first formant frequency ($f1$) and the next two dark horizontal bands above $f1$ are the second formant ($f2$) and third formant ($f3$) frequencies. The spectrogram for /curj/ (k3_) when whispered shows that the signal conveys a weaker formant structure and it can be described as a shadow of its voiced counter part.

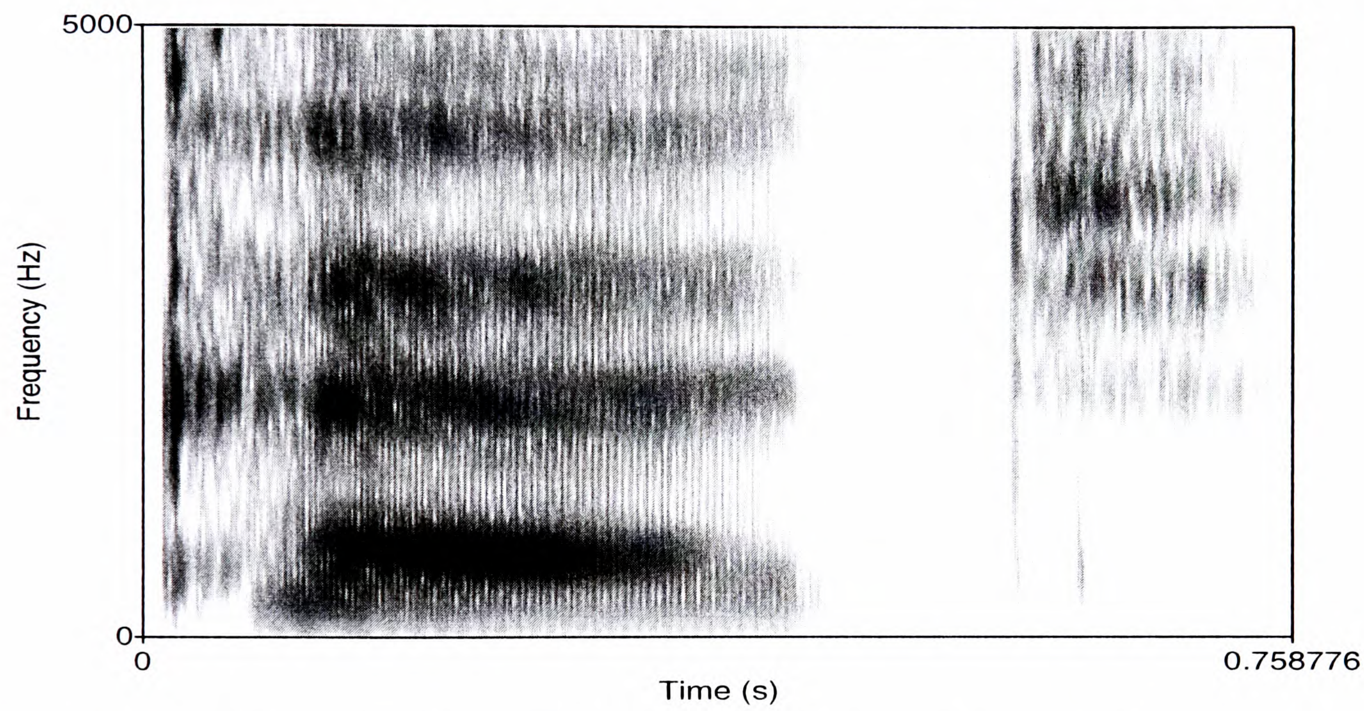


Figure 20a. Wide-band spectrogram of the non-word /curj/ when voiced.

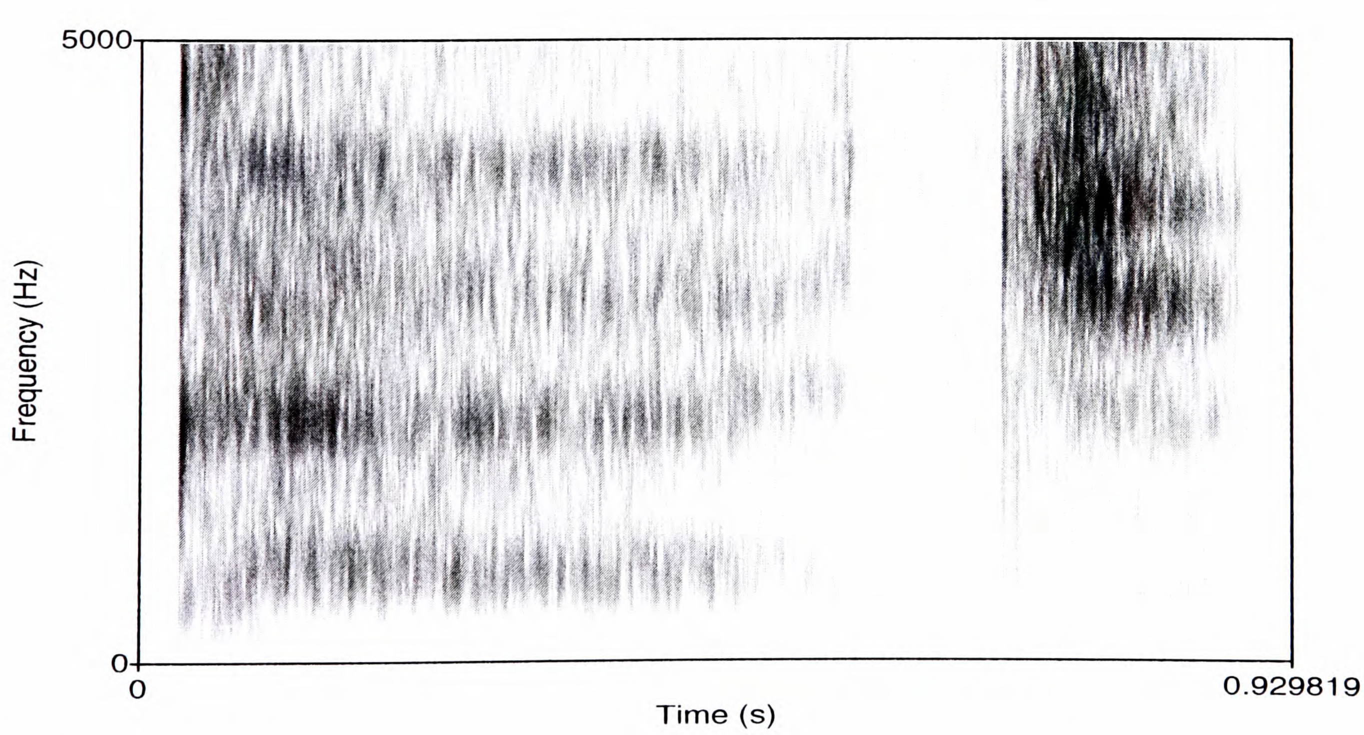


Figure 20b. Wide-band spectrogram of the non-word /curj/ when whispered.

The spectrogram for the voiced version of the non-word /curj/ (figure 20a) shows that due to periodicity of the voiced excitation, harmonics can be seen in the frequency spectrum. The frequency spacing of the harmonics is dictated by the fundamental frequency which could be described as the pitch of the vocal fold vibrations. The $f0$ is depicted by the low frequency energy present, which is indicative of voicing. The spectrogram of /curj/ when whispered (figure 20b) however indicates no low frequency energy due to the vocal folds not vibrating, resulting in the absence of the $f0$.

Upon further inspection of the spectrograms in figures 19a and 19b it is clear that the formant frequencies and their transitions into and out of the vowel are similar, this reflects the fact that the quality of the vowel is the same. This indicates the formants of changing-state speech sounds seem to provide the common ground across which change from item-to-item must occur in order for speech to maintain its higher disruptive power relative to degraded speech (e.g. Jones et al., 2000) and non-speech sounds (e.g. Tremblay et al., 2000; Jones and Macken, 1993). Hughes et al (2005) suggest that changes on a common fundamental and formant structure is important but have not examined whether $f0$ information is necessary in a signal or if the weaker formant structure present when $f0$ is absent is sufficient. The data here suggest that so long as formants in speech and how they vary over time are preserved, whispers will disrupt memory to the same extent as does voiced speech.

Time-varying changes in the frequencies of the lowest three formants have been shown to aid the perception of vowel quality (Strange, Jenkins and Johnson, 1983). Higher formants are present but they are not necessary for the perception of vowel differences (Lieberman and Blumstein, 1988). Hillenbrand (1995) observed lower identification performance for synthesised vowels whose formant pattern was held constant (flattened) over the time course of the vowel than vowels that maintained the natural changes in formant frequencies. A 15 % reduction in identification accuracy for the vowels with *'flattened'* formant tracks was found. This suggests that formant frequency changes are important in the perceptual specification of American English vowels. This corresponds with the intelligibility of the irrelevant sounds of the present experiment. Participants demonstrated 100% correct identification for both the voiced speech sounds and their whispered counterparts.

Assmann and Katz (2000) found that time-varying changes in formant frequencies have similar effects for the intelligibility of voiced and whispered synthesised vowels. The removal of the time-varying formant frequency changes and $f0$ on average lead to a reduction in identification performance by 15% for whispered vowels and 17% for voiced vowels. A similar attenuation in identification accuracy was found for voiced and whispered vowels. In addition, removal of time changes from the formant frequencies led to a substantial reduction in identification accuracy for all speaker groups, for example, males, females and children. Tartter's (1991) data suggests that formant dynamics inherent in the formant pattern can serve to aid the discrimination (disambiguate) between vowels. The time variations in the formant pattern have been maintained in the whispers used in the present experiment, though they are weaker in their structure due to a lack in harmonicity. This is in line with the CSH account of the ISE that sound sequences where adjacent sounds are mismatched and thus distinguishable will interfere with immediate serial recall.

Tartter (1991) compared the intelligibility of voiced and whispered vowels and found that whispered vowels are less intelligible than voiced vowels using natural speech. Tartter (1991) demonstrated that this reduction in identification accuracy was partly due to increased confusions among vowels paired in the acoustic 'vowel space' defined by the first and second formant frequencies. However, pilot intelligibility screening of the present experiment shows that the CVCs in this experiment were 100% intelligible, and they affected serial recall equally. In contrast to the findings of Tartter (1991), the vowels of the whispers were 100% intelligible, as all seven whispered non-words were identified correctly during a pilot listening session. It may be that for the CVC syllables used, the articulation of the initial and final consonants helped cue the identification of the vowels. That whispers and voiced sounds disrupt memory the same and acoustic complexity is not equal but less defined in the whispers suggests an effect of the intelligibility of speech sounds. Intelligibility here refers to the identification of speech sounds, as opposed to speech sounds being comprehendible, since it has been shown that speech in a language unfamiliar to participants is as disruptive of serial recall as speech that is in a language familiar to participants (e.g. Jones et al., 1990). This is evidence against the assumption that coherent sequences of changing-state sounds demonstrating more acoustic change from item-to-item will be more disruptive of serial recall (e.g. Jones et al., 1999a; Tremblay et al., 2000). Whispered sounds are on average found to be less intelligible due to the weaker structure of the formants (Tartter, 1991). It seems as long as changing CVC sounds, whether whispered or voiced, are identified correctly they will be equally as discrete and thus form distinguishable codes in memory. These memory codes are argued to interfere with the rehearsal of TBR items (Jones et al., 1996).

Katz and Assmann (2001) demonstrated that the lower identification accuracy of whispered vowels compared to voiced vowels is not due to the fact that whispered speech, in contrast to voiced speech, has less energy at low frequencies. Data from this series, using a similar experimental design to Assmann and Katz (2000) revealed an equivocal attenuation in intelligibility of unvoiced (noise-excited) vowels which maintained the spectral tilt features of the voiced vowels. The reduction in identification accuracy of whispered vowels was explained as to the removal of periodicity and/or harmonicity. Intelligibility was matched for the whispered and voiced speech sounds of the present experiment. The removal of periodicity and harmonicity in the irrelevant whispered speech did not produce an improvement in serial recall. Therefore, the $f0$ of voiced speech sounds does not provide the common ground on which change needs to occur in order for speech to disrupt serial recall more than non-speech sounds (e.g. Tremblay et al., 2000).

7.9 SUMMARY

The data from this experiment provides further evidence for the suggestion that the magnitude of the ISE may be modulated by the intelligibility of the speech sounds. When speech sounds are whispered and therefore less acoustically complex, as long as they are as intelligible as their voiced counterparts they will disrupt memory to an equivalent degree. This provides evidence against the notion that the size of the ISE is determined by the amount of acoustic variation between successive items in an irrelevant auditory stream proposed by the CSH (e.g. Jones and Macken, 1993). The results also support the finding that vowels, when changing within a sequence, are the dominant source of disruption (c.f. experiment 2; chapter 6; Hughes et al., 2005). The important role of vowel changes would explain the observed equivalent effect of voiced and whispered speech on serial recall performance as the formant frequencies and transitions for both the whispered and voiced speech sounds are similar, indicating preserved vowel quality.

Hughes et al (2005) have suggested that changes carried on an attribute common to speech sounds produced by the same voice over time provide the *common ground* which enables the perceptual system to integrate speech sounds into a coherent stream. Hughes et al (2005) suggest this common ground may be provided by $f0$ and/or the corresponding formant structure. The results of experiment 3 indicate that $f0$ information is not the common attribute of speech sounds that can account for the greater disruptive effect of speech, since when $f0$ information is removed when speech is whispered the ISE is not reduced. It follows that the formant structure of speech sounds produced by the same speaker over time may be the critical attribute on which change from item-to-item in an irrelevant speech stream must occur in order for speech to maintain its ability to disrupt serial recall more than non-speech sounds.

7.10 CONTRASTING THE EFFECTS ON SERIAL RECALL OF VOICED SPEECH AND ALTERNATING BETWEEN VOICED AND WHISPERED SPEECH IN AN IRRELEVANT STREAM

The CSH argues that a sequence of successive sounds that change from item-to-item will be more disruptive than a sequence of identical sounds (e.g. Jones and Macken, 1995a). As the amount of change increases between each successive auditory item, the level of disruption it produces also increases but up to a point, at which the level of change reaches a threshold. Above this threshold of change the sequence of sound is no longer heard as a single coherent changing-state stream, rather it is heard as separate streams of a repeating sound (e.g. Jones and Macken, 1995b; Jones et al 1999a; 1999b). This leads to the suggestion that change between adjacent items must occur on a common ground (in the case of the above example, spatial location). Hughes et al (2005) used the change on a common ground principle to explain their finding that sequences of CVC syllables in which only the vowels change are more disruptive of serial recall than CVC syllables in which only the initial or final consonants change from item-to-item in an irrelevant auditory sequence.

Hughes et al (2005) infer *f*0 information and the formant structure provided by the vowels, which are common to utterances spoken by the same voice, may be the common ground across which change needs to occur. Hughes et al (2005) do not make any predictions as to whether change between adjacent items occurring on both the *f*0 and formant structure of a speaker is necessary for speech to maintain its disruptive potency. Also no predictions are made with regard to whether the formant structure of speech sounds produced by the same speaker alone would be sufficient for speech to be as disruptive as when *f*0 information is present. Experiment 3 demonstrated that change occurring on the common fundamental of $f0$ information does not provide the common ground on which change between adjacent sounds must occur, since whispered speech was as disruptive as voiced speech. This result is inconsistent with the assumption of Hughes et al (2005). Therefore, acoustic change on $f0$ information and the resulting harmonicity and periodicity provided by voiced speech sounds cannot account for the greater disruptive nature of irrelevant changing-state sequences of speech relative to changing-state non-speech sounds, such as sine wave speech, whether perceived as speech or not (Tremblay et al., 2000) or cello notes (Jones et al., 2000).

One inference that can be made from the findings of Experiment 3 is that the maintenance of formant frequency changes over time (provided by the changing vowels) is used to a greater degree by the perceptual system as the common ground on which to temporally organise speech sounds produced by the same speaker than is $f0$ information. This inference is consistent with the interference by process account put forward by the O-OER model which incorporates the CSH (Jones and Tremblay, 2000). That is, the relative interference by irrelevant sound is determined by its ability to pre-attentively and automatically generate cues pointing to the order of its sound components. Acoustic change is required in order for an irrelevant sound sequence to generate competing serial order cues and these changes must occur on a common ground (Jones et al., 1999a; 1999b). Acoustic change, as in the formant frequency changes between successive intelligible whispered speech sounds in experiment 3 would serve to yield the cues required by the perceptual system to maintain their serial order. It follows that these cues to the order of the irrelevant sounds are argued by the O-OER model to conflict with the maintenance of cues that point to the serial order of the TBR items (Jones and Tremblay, 2000).

Contrasting the effects of voiced and whispered sequences of speech sounds is one way of varying the amount of acoustic change within an irrelevant speech stream. Another way is to further manipulate the strength of the acoustic links between distinct sound tokens in an irrelevant stream, which serve to preserve their temporal order by manipulating whether or not successive speech sounds are voiced within an irrelevant sequence.

7.11 AIMS AND OBJECTIVES

Experiment 4 examines the level of serial recall disruption produced by sequences within which voiced and whispered speech sounds are alternated, and sequences consisting of solely voiced speech sounds. Two possible outcomes can be predicted from within the framework of the CSH. First, alternating between voiced and whispered speech would add more change to the irrelevant sequence and therefore produce a larger ISE. This would be predicted as long as the sequence was presented at a rate that would prevent segregation or fission of a sequence into separate auditory streams of identical sounds (e.g. 1 item/second). The auditory sequence would therefore still be perceived as one coherent changing stream. Previous experiments involving normally voiced irrelevant speech have found speech to be more disruptive than music (Salamé and Baddeley, 1989), sequences of tones changing in pitch (Jones and Macken, 1993) and degraded speech (Jones et al, 2000). Varying whether or not successive speech sounds are voiced in a sequence would increase the acoustic changeability of the speech and should render the speech sounds more disruptive of serial recall performance than sequences of voiced speech.

The second outcome which would be predicted by the CSH is that alternating between voiced and whispered speech sounds may act to disrupt the acoustic links between irrelevant speech tokens. A lesser effect of an irrelevant stream in which voiced and whispered speech sounds are alternated would be predicted. These acoustic links are argued by the CSH to be important for the automatic maintenance of the order of speech sounds, a process, which in turn, conflicts with the process of maintaining the order of the TBR visual digits in STM. Fundamental frequency information ($f0$) is absent in whispered speech, as there is no low frequency energy due to the vocal folds not vibrating. Formant structure, however, is maintained though it is weaker than that observed in voiced speech due to noise excitation at an open glottis which excites the resonant frequencies of the vocal tract (Morris and Clements, 2002). Alternating between voiced and whispered speech tokens would convey weaker 'acoustic links' as change across $f0$ information would not be present between adjacent speech sounds. Rather, $f0$ information would be provided by every other speech sound in the sequence. Therefore the common ground provided by $f0$ information would not be common to every speech sound within the irrelevant sequence. Formant structure however would be common to all the distinct speech sounds, though as explained, it would not be as rich as the formants in normally phonated speech. In essence, this experiment aimed to further examine Hughes et al (2005) assumption that change on a common ground (e.g. $f0$ and formant structure) shared by voiced sounds is crucial to modulating the disruptive potency of an irrelevant speech stream on serial recall.

7.12 EXPERIMENT 4: METHODOLOGICAL CONSIDERATIONS

7.12.1 Participants

24 participants volunteered to take part in the study. All reported normal or corrected-to-normal vision and normal hearing. All participants had English as their first language and were not paid for their time.

7.12.2 Stimuli

7.12.2.1 Visual stimuli

A Latin square design was used to construct the to-be-remembered digit lists which consisted of 8 digits from the digit set 1-9 (appendix 29). The digits were constructed with the constraints described in chapter 3, that of there being no upward or downward runs of digits and that no digit appeared in the same position in a successive trial.

7.12.2.2 Auditory stimuli

The non-words for the voiced and alternated (voiced and whispered) speech conditions were */gam/* (`g{m`) and */sof/* (`sQf`). The non-words for both auditory conditions were recorded digitally and presented as described in chapter 3 (p90). The RMS sound level of the voiced and whispered speech sounds in both conditions was equated as described in experiment 3.

7.12.3 Design and procedure

The design and procedure was the same as detailed in chapter 3, with the exception that participants experienced 27 serial recall trials per condition.

7.13 RESULTS

Experimental condition	*Mean Errors*	*SD*
Voiced speech	67.67	27.934
Alternated speech	67.92	26.799
Silence	52.79	29.859

Table 13. Descriptive statistics for the three experimental conditions; mean number of serial recall errors per condition. N = 24.

The descriptive statistics in table 13 indicate that alternating between voiced and whispered speech (alternated speech) disrupts serial recall at a level equivalent to that produced in the presence of a stream of voiced-only speech sounds. The mean errors per experimental condition are summarised in figure 21.

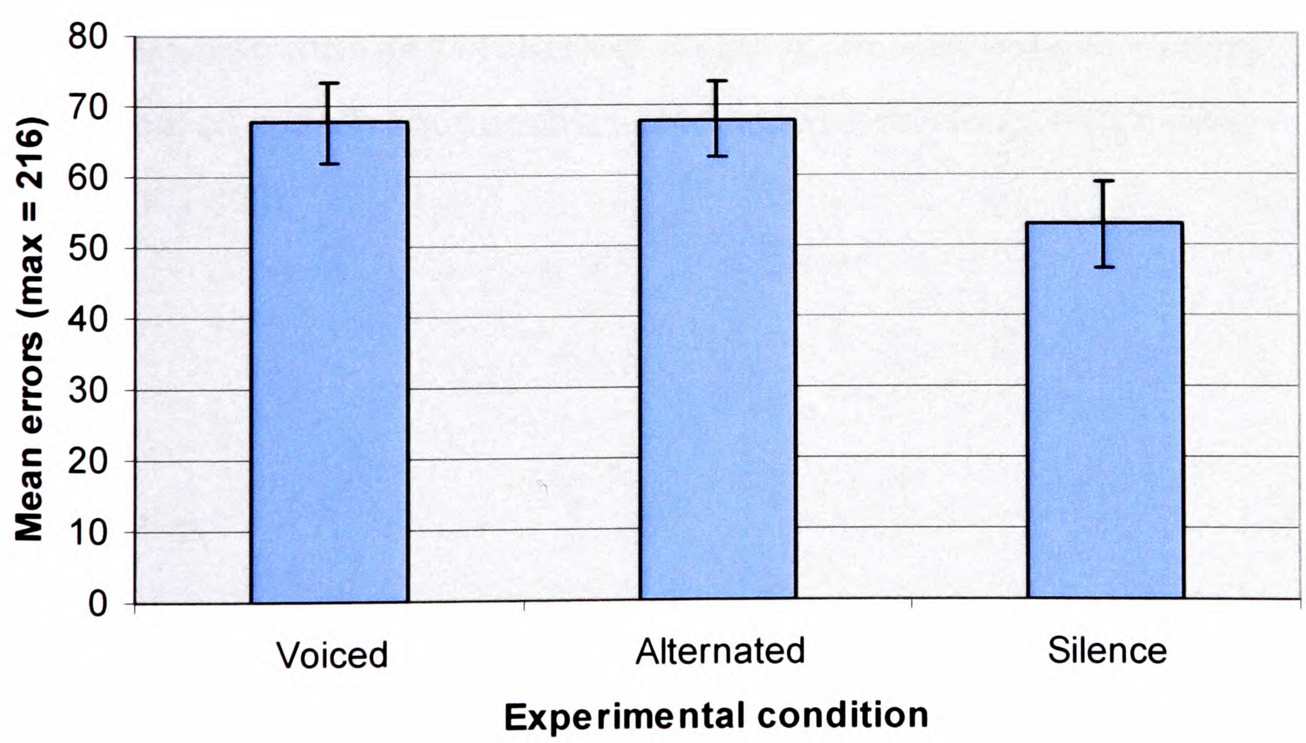


Figure 21. Mean number of serial recall errors for each experimental condition. Error bars represent standard error above and below the mean.

	Silence	*Voiced Speech*
Alternated speech	✓ $p < 0.01$)	Non-sig $p \leq 1.000$
Voiced Speech	✓ $p < 0.01$	xx

Table 14. Bonferroni corrected pairwise comparisons for the three experimental conditions.

A within subjects ANOVA with three levels (speech, alternated speech and silence) found that irrelevant sound significantly impaired serial recall of the digits [F (2, 46) = 8.122, MSE = 1800.375, $p < 0.01$] (appendix 30). The data are summarised in figure 22 which shows the overall level of recall collapsed across serial position. Pairwise comparisons with bonferroni correction as detailed in table 14 revealed that when compared to a silent control speech ($p < 0.01$) and alternated speech ($p < 0.01$) disrupted serial recall (appendix 30). Sequences of voiced speech sounds and sequences which alternated between voiced and whispered speech sounds disrupted serial recall to an equivalent degree ($p \leq 1.000$).

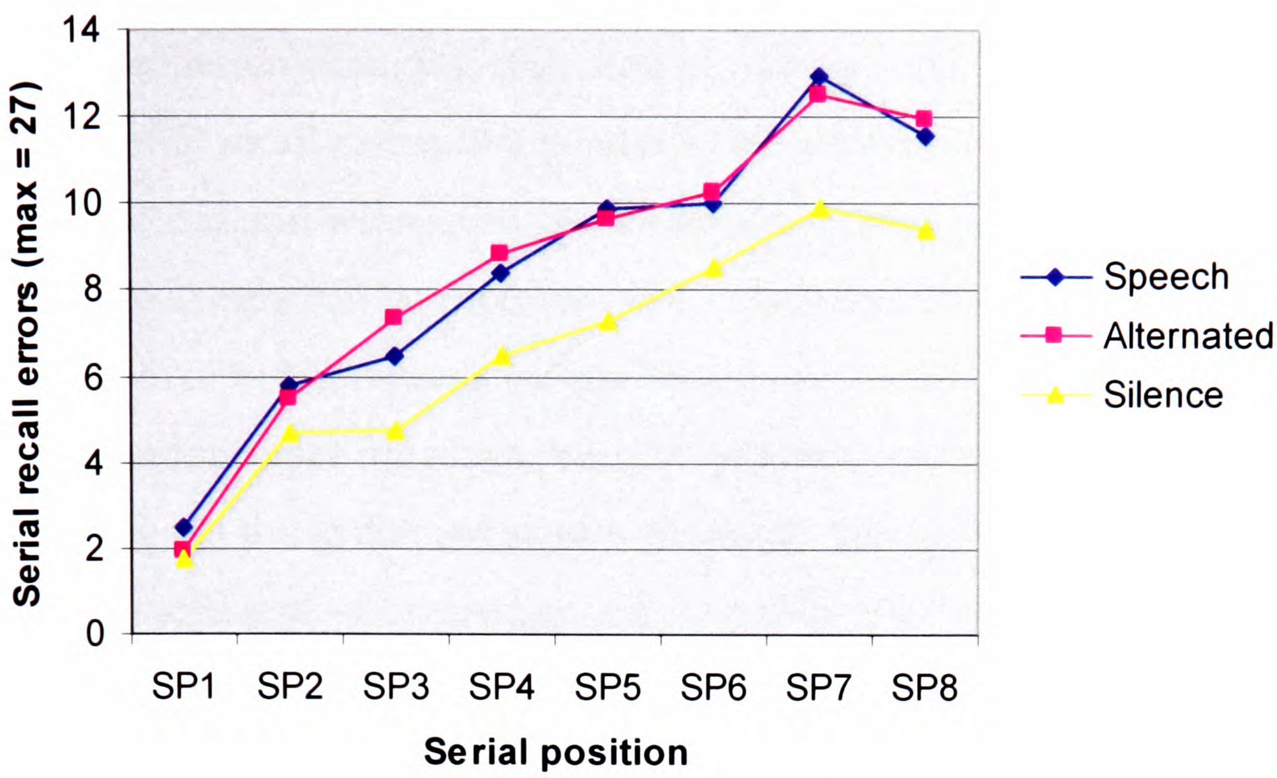


Figure 22. Mean errors for the three experimental conditions collapsed across serial position.

7.14 DISCUSSION

The present experiment compared the relative disruption of serial recall afforded by an irrelevant stream of voiced speech in comparison to a stream that alternated between voiced and whispered speech tokens. No difference in serial recall performance was observed between the voiced and alternated speech conditions and both conditions differed from the silent control, replicating the standard ISE (e.g. Jones and Macken, 1993) which was found when the effect of voiced and whispered speech was contrasted (experiment 3).

The absence of a difference between the speech conditions would not be predicted by the CSH of the object-oriented episodic record (O-OER) model. First, the CSH would predict that alternating between voiced and whispered speech sounds would serve to add more change to an irrelevant auditory sequence and that this would increase the level of serial recall disruption produced. Increased disruption would occur as long as this did not breach a threshold of change which would result in two streams of steady-state items being heard as opposed to the necessary coherent changing-state stream. The absence of a difference between serial recall disruption produced by sequences that alternated between voiced and whispered speech and sequences consisting of voiced speech only can be explained if it is assumed that performance in the presence of voiced speech tokens had reached ceiling. That is, memory performance may have been at its worst and the addition of more change to the irrelevant stream of speech tokens by alternating between voiced and whispered speech sounds would not have acted to further reduce maintenance for their serial order. The addition of further change by alternating between voiced and whispered items may not have produced further disruption as the threshold of maximum interference had already been reached by sequences of changing voiced-only speech.

With regards to the acoustic links which are argued by the CSH to point to the order of speech sounds produced by the same speaker over time, the CSH would have alternatively predicted a reduction in the size of the ISE in the presence of a sequence that alternated between voiced and whispered speech sounds. This is because alternating between whispered and voiced items would be predicted by the CSH to weaken the acoustic links between adjacent speech items. This would be the case as $f0$ information is present in the phonated speech signal, but not in the whispered speech signal. Therefore, $f0$ information would not provide the common ground on which the CVC monosyllables changed as it occurred intermittently in the condition alternating between voiced and whispered sounds. The finding that both speech conditions produced an equal amount of disruption refutes this prediction and it indicates that it is not $f0$ and formant structure together that provide the shared ground on which the syllables must change in order for speech to maintain its disruptive power. Acoustically, whispers are less complex than voiced speech. Voiced speech is produced by a glottal pulse and as such contains a harmonic structure. Whispered speech on the other hand has noise as its source. Even though voiced speech is more complex acoustically, it can be inferred that the spectrally reduced acoustic links between whispered and voiced monosyllabic CVC non-words are sufficient to produce an ISE equivalent to that observed under a stream of only voiced speech. Figure 23a and 23b display the wideband spectrograms for the non-word */gam/* (g{m) when voiced and whispered (see appendix 5 for examples of disc phonetic symbols). The formant frequencies are present within the whispered speech signal of */gam/* as was the case for the whispers in experiment 3. This non-word also featured in the voiced and whispered speech conditions of experiment 3.

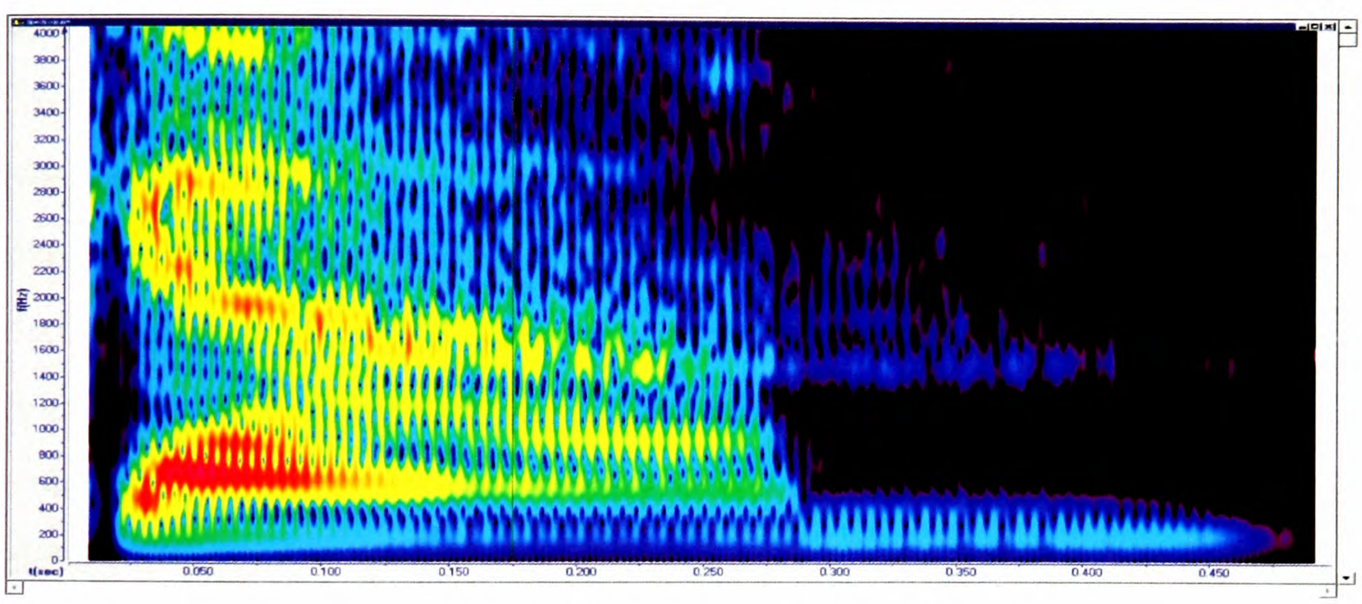


Figure 23a. Spectrogram of monosyllabic non-word /gam/when voiced.

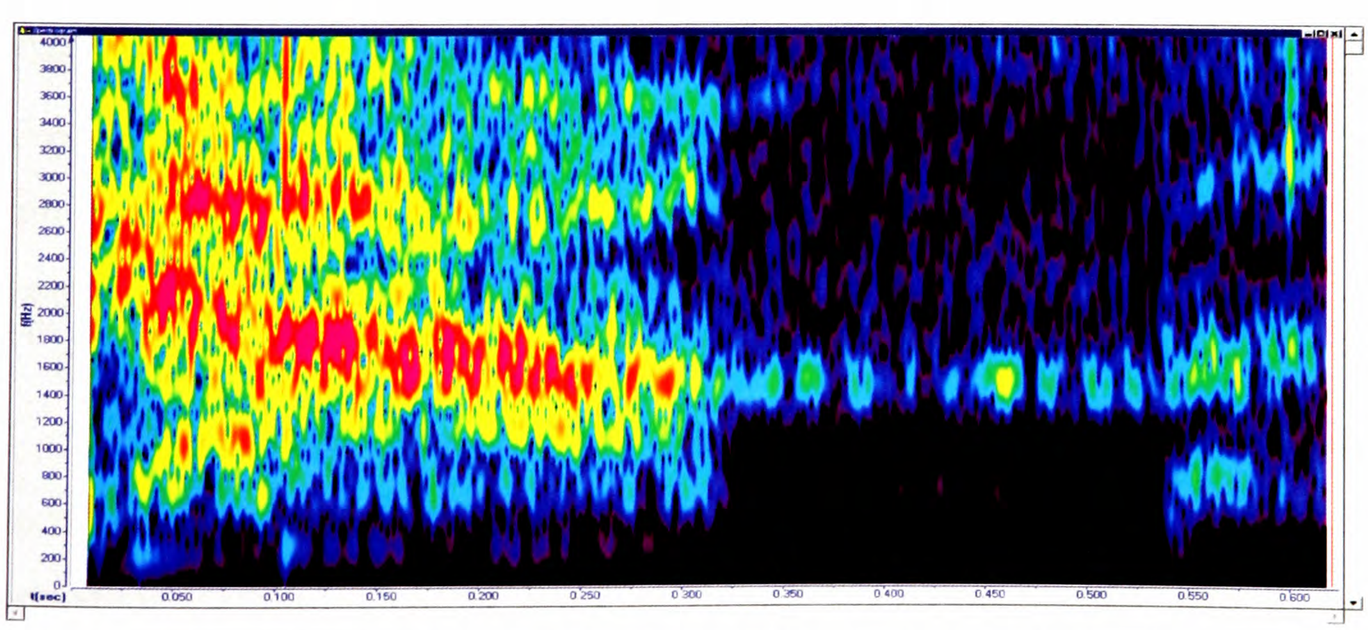


Figure 23b. Spectrogram of monosyllabic non-word /gam/when whispered.

The presence of the formant frequencies of the vowels within the whispered speech indicates that acoustic change, which is required to occur between the syllables for the observation of an ISE, was conveyed by the formant structure of the changing syllables in the absence of $f0$ information in the whispered items of the alternated speech condition. Formant frequency changes are argued to preserve vowel quality and changes on formant structure have been suggested as providing the common ground on which auditory items must change (Hughes et al., 2005).

The upper formants have been shown to be more important for whispered vowel classification than for the classification of voiced vowels. For example Halberstam and Raphael (2004) found the third formant frequency ($f3$) to be more important for the classification of whispered than voiced vowels. It is clear that the upper formants were also maintained in the whispered non-words. As well as the formants being present so are the formant transitions going into and out of the vowel. Cole and Scott (1973) argue vowel transitions are important for temporal order judgment of speech sounds. Cole and Scott (1973) suggest the role of vowel transitions is to hold together adjacent consonant and vowel sounds in speech which demonstrate different spectral characteristics. The presence of the vowel transitions in the whispered CVC syllables can explain why alternating between voiced and whispered speech sounds interfered with serial recall to the same extent as did sequences of only voiced speech.

The findings of experiment 4 can be accounted for with reference to research examining the difference in processing steady-state and transient acoustic cues discussed in chapter 6 (p144). Vowels vary broadly along steady-state cues, such as formant frequency whereas consonants vary broadly along rapidly-changing cues, for example fine temporal distinctions such as voice-onset-time (VOT) (Schouten and Van

Hessen, 1992). Vowel changes are easily discriminated, whereas consonant changes are less well discriminated because the physical steady-state cues of vowels are more discrete and therefore more useful for discrimination. The physical rapidly-changing features of stop consonants however demonstrate greater similarity and are thus less discriminable (Nairne, 1988). The steady-state cues of the vowels remained present in the whispered speech sounds due to the presence of formant frequency changes. Vowels which broadly vary along steady-state cues, such as formant frequency have been found to be processed more in the right hemisphere (Allard and Scott, 1975; Belin et al., 1998). It follows that irrelevant sound is predominately processed in the right hemisphere (Hadlington et al., 2004; 2006).

Changing vowels in an irrelevant stream are argued to be the dominant source of disruption in speech (c.f. experiment 2, chapter 6; Hughes et al., 2005). As the whispered speech sounds were correctly identified during a pilot listening session for experiment 3, the vowels of the two speech sounds (/sof/ and /gam/) in the alternated speech condition would have been as discriminable as the same tokens when voiced only. The match in intelligibility in the absence of a match in acoustic complexity resulted in an equivalence in serial recall disruption for both speech conditions. That alternated and voiced speech did not differ in their disruptive potency is evidence against the assumption that speech is more disruptive than sine-wave speech, perceived as speech or not, because it is spectrally more complex and thus the nature and extent of acoustic change is greater in speech (Tremblay et al., 2000). It seems acoustic change itself cannot account for why speech is found to be the most disruptive of serial recall. Instead, it seems *changes within the formant structure* provided by the *changing vowels* within a speech sequence, which is richer in speech than in sine-wave speech, along with the intelligibility of speech renders speech more disruptive.

7.15 SUMMARY

If irrelevant sounds are perceived as speech and the vowel portions of these changing sounds are discriminable, the subsequent disruption of serial recall in its presence will be at ceiling. Adding more acoustic change through alternating between voiced and whispered speech items will not increase its disruptive effect. In addition, alternating between voiced and whispered speech tokens will not weaken the acoustic links pointing to the order of these irrelevant items to the extent that the ISE will be reduced. This is because these acoustic links are preserved through the presence of formant frequency changes, though these are weaker due to the absence of harmonicity in the whispered speech signal. The findings of experiment 4 provide further evidence that information conveyed in $f0$ is not responsible for the magnitude of serial recall interference produced by irrelevant speech relative to non-speech sounds. Rather, the steady-state feature of formant frequency is a sufficient attribute on which change needs to occur between adjacent spoken items in order for speech to maintain its greater disruption of serial recall.

CHAPTER 8

8 INTERIM SUMMARY

Degradation of irrelevant sequences of speech and cello notes (non-speech) reduces the degree to which they disrupt serial recall (Jones et al., 2000). The improvement in memory performance as a function of the systematic degradation of both speech and cello notes follows a similar linear pattern for both classes of sound. However, clear undegraded speech is more disruptive of serial recall than clear undegraded cello notes (Jones et al., 2000). A higher level of memory interference in the presence of speech sounds as opposed to non-speech sounds has been demonstrated with other types of non-speech stimuli. Speech has been found to disrupt serial recall more than simple tones, both of which form examples of changing-state stimuli (LeCompte et al., 1997). Sine-wave speech, an ambiguous stimulus made from three sinusoids that track the time-varying changes of the first three formants in speech (Remez et al., 1981), has been found to interfere with serial recall less when compared with recall in the presence of natural speech. This is the case whether or not participants are trained to hear sine-wave speech as speech (Tremblay et al., 2000).

The higher interference of serial recall by natural speech relative to non-speech sounds, such as sine-wave speech (Tremblay et al., 2000), cello notes (e.g. Jones et al., 2000) or simple tones (LeCompte et al., 1997) has been explained with reference to the greater acoustic complexity of speech. There is more change in the constituent components of an irrelevant stream of speech sounds (e.g. phonemes) in contrast to a sequence of sine-wave speech, cello notes and simple tones. The fact that when sine-wave speech is heard as speech it does not disrupt memory at a level equivalent to that found with natural speech provides evidence to suggest that it is the ease with which a sound pattern is recognised as speech which is crucial to determining its relative disruption of serial recall. Sine-wave speech is only heard as speech with training, and therefore the acoustic pattern it provides is not sufficient enough for the perception of speech unless the original speech utterance it was constructed from is presented to the listener.

The aim of the experiments in this thesis is to investigate the relative importance of the acoustic-phonetic features of speech in determining how speech-like speech needs to be in order for it to maintain its disruptive power relative to non-speech. In other words, the experiments examine what characteristic(s) of speech give rise to the higher levels of serial recall interference observed in its presence in comparison to the smaller ISE found when non-speech sounds form the irrelevant auditory sequence.

Pilot B (for experiment 1b) (chapter 4) and Experiment 1b (chapter 5) sought to first establish a level of phonological degradation that would result in degraded speech differing reliably from clear speech and also differing from a silent control in the level of serial recall interference it produced. In pilot A (for experiment 1a) (chapter 4), a sequence of monosyllabic consonant-vowel-consonant (CVC) non-words were phonologically degraded by random reversal of a percentage of the samples of each stimulus which turned a percentage of the non-words into noise. The non-words were degraded at a signal-to-noise ratio (SNR) of 0.65 and 0.7. A perceptual identification task produced a range of intelligibility for the non-words at both levels of degradation. Seven non-words of low intelligibility were isolated from the non-word set that was degraded at 0.7 SNR. These non-words formed the irrelevant changing-state sequence of sounds for the degraded speech condition of pilot B (for experiment 1b) whose effect was compared to clear (undegraded) speech and a silent control. Clear speech was found to disrupt serial recall more than degraded speech and only clear speech differed from a silent control.

Recent research by Hughes et al (2005) investigating the relative disruptiveness of vowels and consonants has found that a sequence of CVC syllables where both the vowels and consonants change disrupts serial recall when presented as irrelevant sound more than does a sequence of CVC syllables where only the final consonants change. Further, a sequence of vowel-only-changing (V-O-C) CVC syllables is more disruptive of serial recall than is a sequence of consonant-only-changing (C-O-C) CVC syllables. Both sequences of CVC syllables that only changed in the initial and final consonants disrupted serial recall at a level that did not differ reliably from a steady-state condition consisting of a repeated CVC syllable. This shows that vowels seem to be the most important component of speech that needs to change-in-state in order for speech to maintain its higher disruption of serial recall relative to a steady-state condition.

The perceptual identification data for the seven non-words of the degraded speech condition was analysed and it was evident that the initial and final consonants had been misperceived more than the vowels (see figure 7, chapter 4, p103). Vowels are argued to be the dominant source of disruption in speech (Hughes et al., 2005) and degrading speech reduced its disruption of serial recall. It was concluded that the information in the vowels must have been degraded to a sufficient extent to reduce serial recall interference. That no reliable difference in serial recall performance was found between degraded speech and silence may have been an artifact of the experimental design not being fully counterbalanced. Since vowel changes have been found to be the dominate source of disruption and yet relative to the consonants they have been preserved more by degradation, a reliable difference would have been expected between serial recall disruption by degraded speech and silence. As the sounds for both speech conditions were presented free-field and the presentation order of conditions was not fully counterbalanced it was difficult to make any conclusions with regards to whether a larger SNR would have resulted in a reliable difference between degraded speech and silence.

Experiment 1a (chapter 5) degraded a set of monosyllabic non-words at a SNR of 0.7 and as for the pilot, seven low intelligible non-words were isolated from the intelligibility range obtained from the perceptual identification task and formed the degraded speech condition. Sounds were presented over headphones and the presentation order of conditions was fully counterbalanced. In contrast to pilot B (for experiment 1b) (chapter 4), experiment 1b (chapter 5) itself observed a significant difference between the disruption of serial recall by degraded speech and a silent control. As in pilot B (for experiment 1b), clear speech differed from silence.

Analysis of the identification data in experiment 1a revealed that the initial consonants of the seven non-words forming the degraded speech stream were misperceived more than the vowels and final consonants, for which no difference was found. The finding that vowels were misperceived less and yet degraded speech produced less recall interference when compared to clear speech can be accounted for if it is assumed that important acoustic information within the vowels was distorted to a sufficient degree. This is a plausible explanation as C-O-C sequences have been found not only to differ from V-O-C sequences, but to disrupt serial recall at a level equivalent to that obtained with steady-state sequences formed by a repeating CVC syllable (Hughes et al., 2005).

Experiment 2 (chapter 6) went on to investigate the effect of phonological degradation on the pattern of interference generated in the presence of irrelevant V-O-C and C-O-C non-word sequences. A similar linear pattern of interference has been observed for degraded speech and cello notes. As degradation of both classes of stimuli increased, serial recall performance under both types of sound improved (Jones et al., 2000). It follows that for both classes of sound a monotonic linear relationship was observed between stimulus degradation and its disruption of serial recall. This has been viewed as evidence for the functional equivalence of changing-state speech and non-speech sounds. However, when serial recall data (e.g. Jones et al., 2000) is averaged a linear function is often observed. Experiment 2 compared the linear interference function observed for degraded V-O-C and C-O-C sequences. Different CVC syllables featured in the irrelevant sequences for each condition. Therefore any effect of degradation on V-O-C and C-O-C sequences in terms of their disruption of serial recall could be generalized more to the syllable population.

Three possible outcomes were predicted. First, since vowels have been found to be the dominant source of disruption (Hughes et al., 2005) it was predicted that V-O-C sequences of syllables, degraded at the three levels of 0% noise, 30% noise (0.7 SNR) and 50% noise (50% SNR), will produce a pattern of interference that is equivalent to that found when C-O-C sequences are presented. However, clear V-O-C sequences would be predicted to interfere with serial recall more than clear C-O-C sequences. Second, not only are vowels more disruptive of serial recall (Hughes et al., 2005), they are more redundant to degradation by noise (experiment 1b, chapter 5). Therefore, a shallower linear relationship between stimulus degradation and serial recall interference for the presentation of V-O-C sequences degraded at three levels would be predicted, in comparison to the interference function obtained in the presence of C-O-C sequence as they are degraded. Third, sequences of C-O-C syllables have not only been shown to be less disruptive of serial recall relative to V-O-C sequences of syllables, they also disrupt serial recall at a level equivalent to that observed with a steady-state (repeated) syllables. It was therefore predicted that degrading C-O-C syllables may not influence their effect on serial recall interference. Although the consonants of the degraded non-words in experiment 1b were found to be most effected by degradation as they were misperceived more than vowels (see figure 10, chapter 5, p115), the vowels were still misperceived. As changing vowels produce more disruption (Hughes et al., 2005), the degradation of the vowel portion of the non-words may have accounted for the reduction in the size of the ISE in the presence of degraded speech. Therefore, it may be that as sequences of V-O-C syllables are degraded, the number of serial recall errors made in their presence may decrease.

The results replicated the findings of Hughes et al (2005) by showing that clear V-O-C sequences interfered with serial recall more than did C-O-C sequences. Hughes et al (2005) however did not change both the initial and final consonants in their C-O-C sequences, but instead investigated their effects separately over two experiments. Experiment 2 however showed that even when both consonants were changing in a stream as opposed to the vowels, a C-O-C sequence still disrupted memory less than a V-O-C sequence.

Contrary to the three predicted outcomes, Clear V-O-C sequences did not differ from both degraded versions (0.7 and 0.5 SNR) of the V-O-C sequences in the level of serial recall interference they produced. The difference between clear V-O-C sequences and V-O-C sequences degraded at 0.7 SNR (30% noise) approached significance. When the data from the V-O-C sequences degraded at both 0.7 and 0.5 SNR (50% noise) were pooled a significant difference between the disruptive effect of clear and degraded V-O-C sequences was found. No difference in serial recall disruption was observed between C-O-C sequences at all three levels of degradation. This can be explained if it is assumed that performance was already at ceiling for the clear C-O-C condition. Therefore, subsequent degradation of C-O-C stimuli would not serve to improve memory performance. Changing consonants are not only less disruptive of serial recall relative to changing vowels, but are also found to produce recall interference that does not differ reliably from a steady-state sequence of a CVC syllable (Hughes et al., 2005). It follows that if C-O-C sequences do not produce memory interference equivalent to that produced by a changing-state sequence where the vowels change, degrading them would not result in a reliable improvement in memory performance being observed when they feature as irrelevant sound.

Nairne's (1990) feature model can account for the greater disruptive effect of vowels relative to consonants with reference to the relative discriminability of vowels and consonants. Vowels are easily discriminated (Pisoni, 1973) and this may explain why V-O-C sequences were better serially recalled than sequences of C-O-C sounds. Consonants are discriminated less well and this may account for why sequences featuring only consonant changes are not serially recalled as well as undegraded V-O-C sequences. Neath and Surprenant (1996) degraded the vowels of V-O-C syllables by mixing them with noise. These were identified less well and yet were serially recalled at a level that was equivalent to that found for C-O-C sequences.

The feature model (Nairne, 1990) argues vowels are more discriminable because their modality-dependent features (physical information) are more useful for discrimination than those of consonants and degraded items (Nairne, 1988). Therefore, the equivalent serial recall performance observed in the presence of degraded V-O-C syllables and C-O-C syllables in experiment 2 (chapter 6) may be because the modality-dependent features of the vowels in noise have been degraded and their usefulness equated to that of the modality-dependent features of consonants. It follows that irrelevant clear V-O-C sequences may be more disruptive of serial recall than C-O-C sequences because the modality-dependent features of vowels are more useful. As a result, vowels changing from item-to-item are more discriminable and thus elicit stronger serial order information which conflicts with the process of remembering the order of TBR items.

These findings can be related to conclusions derived from research in categorical perception. Categorical perception involves two tasks. The identification task involves a listener having to match a heard sound to a stored standard. The discrimination task involves the presentation of two stimuli to a listener who has to decide if these differ or not. A small acoustic change between stimuli will result in the altered perception of a vowel. However, only around the category boundary will a small acoustic change result in a listener hearing a different consonant. Consonants are categorically perceived in the discrimination phase because their modality-dependent features are not as useful as those afforded by vowels, which are perceived more continuously (Liberman et al., 1957). The finding that degraded vowels, which are identified less well than consonants and are serially recalled at a level equivalent to that found for consonants is seen as evidence that would predict that degraded vowel stimuli are perceived more categorically than are undegraded vowels (c.f. Neath and Surprenant, 1996). It follows that the vowels of the degraded V-O-C sequences of experiment 2 (chapter 6) may have been perceived more categorically and this may explain why no difference between degraded V-O-C sequences and C-O-C sequences was found.

The difference in the way that vowels and consonants are processed can also account for the greater disruptiveness of vowels. Vowels are defined mainly by steady-state information whereas consonants are defined by rapidly-changing information (c.f. Mirman, Holt and McClelland, 2004). The processing of steady-state information is right hemisphere dominant whereas rapidly-changing cues are processed more in the left hemisphere (Allard and Scott, 1975). This processing difference may be due to vowels requiring longer temporal integration windows which may be used by the right hemisphere (Poeppel, 2003). It follows that irrelevant sound presented to the left ear which is processed in the right hemisphere causes more serial recall interference than when irrelevant sound is presented to the right ear and thus processed in the left hemisphere (Hadlington et al., 2004; 2006). Therefore, vowel changes may bring about more serial recall interference because they are processed more by the right hemisphere which has been shown to be the dominate hemisphere in the processing of unattended task irrelevant sound.

The CSH, which forms an interference by process account of the ISE (Jones and Tremblay, 2000) argues that it is the ease with which an irrelevant auditory sequence automatically yields information pertaining to the serial order of its component sounds which determines the size of the ISE. These automatically encoded cues to the order of irrelevant sounds are argued to conflict with the process of remembering the serial order of the TBR items. Hughes et al (2005) suggest that V-O-C sequences are more disruptive of serial recall than are C-O-C sequences because vowel changes elicit more information with regards to serial order. The idea that changing vowels provide more cues to serial order than do changing consonants is supported by the findings of Surprenant and Neath (1996) who found that serial recall of V-O-C sequences was better than that observed for C-O-C sequences. Hence, sounds which are better recalled when in the focus of attention seem to disrupt memory more when they are irrelevant to the memory task (Hughes et al., 2005).

The idea that vowel changes afford more information as to their serial order seems plausible when the role of vowel changes in the perceptual organization of speech is considered. The ability to integrate speech sounds produced by the same voice over time is argued to be afforded by a similarity common to the periodic vowels (e.g. in a common fundamental and formant structure). It has been suggested that changes between vowels in a common fundamental and/or formant structure have a greater propensity to generate more serial order information than do changes only between consonants and are therefore more likely to interfere with serial recall performance more than do changing consonants (Hughes et al., 2005).

The greater acoustic complexity of irrelevant speech is argued to account for the higher level of serial recall interference found in comparison to when non-speech sounds are presented (e.g. Jones et al., 2000). Experiment 3 (chapter 7) compared the effect of voiced and whispered CVC monosyllabic non-words in order to examine the explanatory power of the idea that an irrelevant auditory sequence whose constituent sounds change more from item-to-item will elicit a larger ISE (e.g. Jones et al., 2000; LeCompte et al., 1997; Tremblay et al., 2000). In whispered speech voicing is removed and hence whispered speech is reduced in its acoustic complexity as it has no harmonic structure and therefore demonstrates less 'acoustic change' from item-to-item in an irrelevant speech sequence. Voiced and whispered speech sounds which differed in acoustic complexity but were matched for intelligibility produced a statistically equivalent ISE (experiment 3, chapter 7). This finding refutes the CSH as more acoustic changes between sounds provided by $f0$ information and the resulting harmonic structure within the voiced speech sequence did not render it more disruptive of serial recall than whispered speech.

The finding that voiced and whispered speech are similar in their disruption of serial recall is at variance with the notion that the acoustic features of a changing-state sequence of sound as opposed to the nature of the sound determine the level of disruption by irrelevant sound (Jones and Macken, 1993; Jones et al., 2000; Tremblay et al., 2000). It seems the nature of the irrelevant stimuli and not just the number and extent of acoustic changes between successive items in an auditory stream dictates how efficiently the perceptual system can automatically encode the temporal order of the sounds, which then conflict with the process of seriating the TBR items in STM memory.

This finding can be explained with reference to the presence of remnants of the formants and how they change over time within the whispered speech signal. Time-varying changes of the lowest three formant frequencies have been shown to be important for vowel perception (Strange, Jenkins and Johnson, 1983). Formants were present in the whispers only as a shadow of the formant frequency changes inherent in the voiced speech due to turbulent noise excitation of the vocal tract's resonant frequencies as opposed to its excitation by a glottal pulse. Formant structure may provide the common fundamental in which speech sounds need to change from item-to-item in an irrelevant stream. This is in line with Hughes et al's (2005) suggestion that changes in a common fundamental and formant structure may help the perceptual system organise speech sounds produced by the same speaker over time. However, no predictions were made by Hughes et al (2005) as to whether $f0$, common to all voiced sounds produced by the same speaker and formant structure are necessary or whether formant structure alone in the absence of $f0$ information would be sufficient to carry change between successive items. Common to speech sounds produced by the same voice is the $f0$ of the voice. However, experiment 3 (chapter 7) shows that when this information is no longer available, formant structure provides the necessary foundation on which change needs to occur between successive sounds in order for the sounds to elicit strong acoustic links pointing to their temporal order. These acoustic links disrupt retention of the serial order of TBR items. It follows that the observation that steady-state and C-O-C irrelevant sequences do not bring about the level of serial recall interference found with changing-state sequences that feature vowel changes (Hughes et al., 2005) can be explained by the absence of a change between successive sounds in the steady-state portion of the vowels inherent within the formants.

Experiment 4 (chapter 7) aimed to analyse further the relative importance of the strength of the acoustic links between adjacent sounds in determining serial recall interference. Sequences of voiced speech sounds were compared to sequences whose sounds alternated between voiced and whispered speech sounds. No difference in serial recall interference was found between the speech conditions. This refutes two predictions of the CSH. First alternating between voiced and whispered speech adds more change to an irrelevant stream, but such a stream does not produced more serial recall interference when compared to the disruptive effect of a stream of voiced speech. This equivalence in disruption can be explained if it is assumed that a maximum threshold of disruption has been reached with a stream of voiced-only speech tokens. It follows that the addition of more change by alternating between voiced and whispered segments would not act to reliably increase the size of the ISE in its presence. Second, alternating between voiced and whispered speech would be argued by the CSH to weaken the acoustic links and therefore interfere with serial recall of TBR items less. This is because $f0$ information, which is suggested to be an important fundamental on which change between successive items needs to occur (Hughes et al., 2005), does not feature between successive items but in every other item. However, the results of experiment 4 (chapter 7) showed that the intermittent presence of $f0$ information does not attenuate the ISE and this provides further evidence in support of the importance of formant structure and changes in formant frequency in determining the size of the ISE.

Although voiced speech sounds are more complex acoustically than whispers, the spectrally reduced acoustic links between whispered speech sounds as well as those between alternated voiced and whispered speech sounds are sufficient to produce an ISE of the same magnitude as found with voiced-only speech. It can be argued that even if the extent of acoustic change between distinct successive speech sounds produced by the same voice is reduced they will disrupt serial recall to the same extent as speech sounds whose acoustic complexity is not reduced. This will be found as long as speech sounds are intelligible (heard as speech) and convey vocal tract resonances (formant frequency changes).

CHAPTER 9

9 CONTRASTING THE DISRUPTIVE EFFECT OF WHISPERED SPEECH AND FINE STRUCTURE REVERSED WHISPERED SPEECH.

9.1 BACKGROUND

The finding that speech, whether played forwards or backwards (reversed) as well as speech in an unfamiliar language disrupt serial recall to an equivalent degree (Jones et al., 1990) has been viewed as evidence that it is not the meaning of sound that determines the size of the ISE. Rather, it is the acoustic changes between irrelevant sounds that determine the degree to which they disrupt serial recall (Jones, 1999). The acoustic spectral detail of reversed speech is preserved, but it is not the same phonologically as speech played forwards. This is because the reversed signal has a different rise and decay time, as speech offsets become speech onsets and therefore reversed speech cannot be articulated (Scott and Wise, 2004). Although these sounds cannot be articulated and are perceived as unfamiliar by listeners (Scott and Wise, 2004) arguably they still sound speech-like, particularly if the maintenance of spectral detail is considered. As speech in an unfamiliar language and reversed speech have a similar effect on memory, it seems that as long as sounds are perceived as speech, even if they are not understood, they will have an impact on memory equivalent to that observed with fully intelligible irrelevant speech sounds.

9.2 AIMS AND OBJECTIVES

Experiment 5 examined the effect on serial recall performance of whispers whose fine structure was temporally reversed, whilst maintaining the amplitude envelope in contrast to normal whispered speech. Whispers whose fine structure was temporally reversed are referred to as fine structure reversed (FSR) whispers. Figure 24 shows the speech waveform for the voiced non-word /larb/ (l£b). The amplitude (waveform) envelope is the smooth curve that would be observed by drawing a line that would join the spectral peaks in the waveform (Moore, 2004).

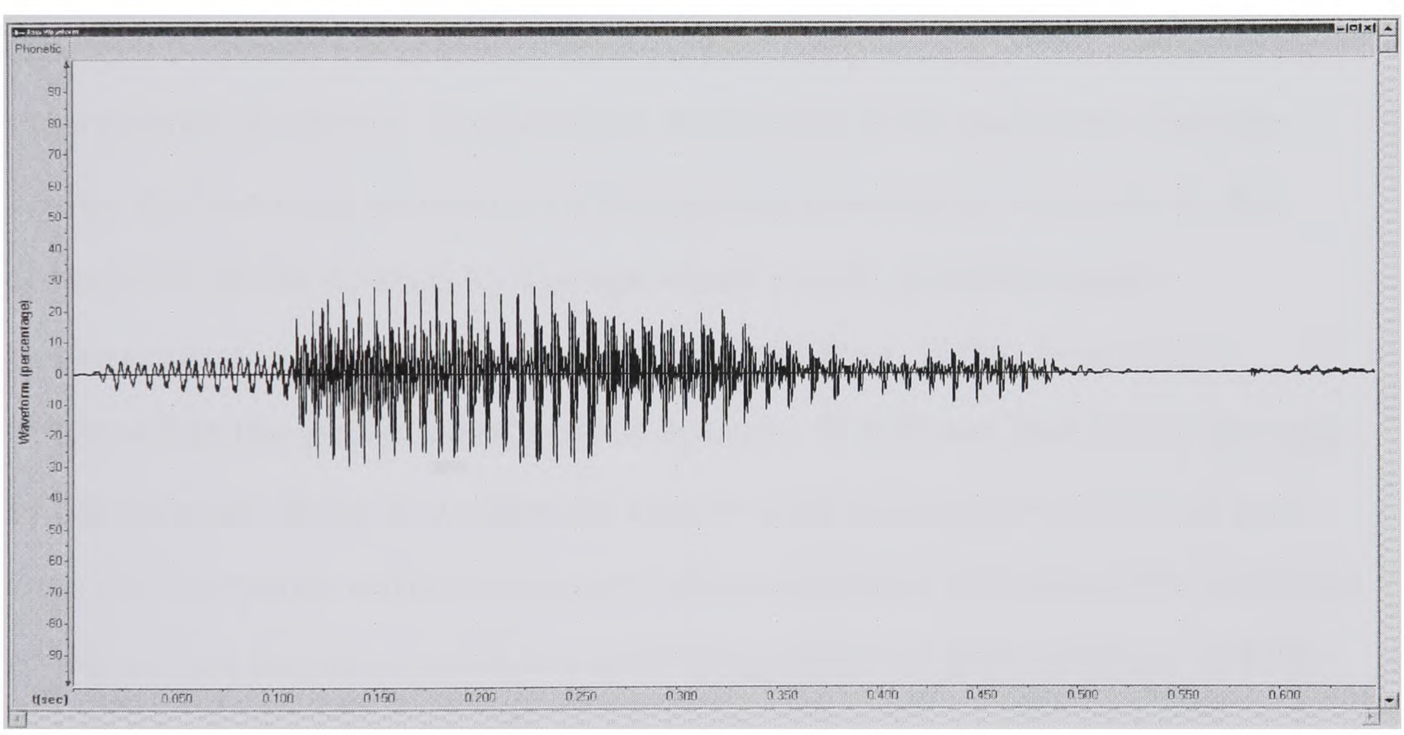


Figure 24. Waveform envelope for the voiced non-word /larb/ produced by a male speaker.

Reversing the fine structure of whispered speech, whilst preserving the original sound patterning provided by the waveform envelope, acts to reverse the spectral-temporal detail within the fine structure of the speech. Changes in the lowest three formant frequencies over time are observed to aid the perception of vowels (Strange et al., 1983) and vowel changes are found to be more disruptive of memory than consonant changes (Hughes et al., 2005). It follows that temporal reversal of the spectral information within the fine detail of whispered speech may weaken the acoustic links between sounds that are encoded by the perceptual system. It has been suggested that these acoustic links, provided by changes in formants over time and/or in $f0$ preserve the temporal order of irrelevant sounds produced by the same speaker, thereby affording their perceptual coherence as sounds forming a single changing stream (Bregman, 1990; Hughes et al., 2005). Jones et al (1990) observed no difference between serial recall disruption by reversed speech (speech played backwards) and speech played forwards. Reversing speech maintains the spectral detail of speech which conveys information regarding $f0$ and formants. It may be that reversing the fine structure of whispers will have a more detrimental effect on its already reduced spectral detail. In whispered speech, the suggested important changes between successive items would only occur in the formants as $f0$ information is absent. Experiment 3 (chapter 7) showed that changes within the formant structure of whispered sounds is sufficient in the absence of $f0$ for speech to disrupt serial recall, as serial recall performance in the presence of whispers did not differ from that observed in the presence of voiced speech. It follows that if the strength of the acoustic links is weakened by reversal of the spectral detail and thus the formants within whispers, this would act to reduce the seriation of the sound sequence and its conflicting effect on the seriation of TBR items.

The aim of Experiment 5 was to investigate whether maintaining the acoustic complexity of whispered speech, whilst destroying its intelligibility by reversing its spectral detail, would render it less disruptive of serial recall relative to normal whispered speech. In addition, whether weakening the acoustic links between successive whispered speech sounds by reversing their spectral detail in the time domain will reduce the ISE is also examined.

FSR whispered speech which maintains the original amplitude envelope of whispered speech will produce unfamiliar sounds that do not occur in naturally spoken speech, as is the case when the complete speech signal is reversed in the time domain (c.f. Jones et al., 1990). However, reversing the spectral-temporal information within the fine structure of whispers (e.g. formant frequency changes) would destroy the intelligibility of the phonemes more than when this information is reversed in normal speech, which is spectrally more complex due to its quasi-periodic and harmonic structure. Further, as the patterning of information within the amplitude (waveform) envelope, which is suggested to provide information regarding phonemes (Moore, 2004), would not match the temporally reversed spectral detail this should also generate FSR whispers that are not intelligible as speech.

9.3 EXPERIMENT 5: METHODOLOGICAL CONSIDERATIONS

9.3.1 Participants

30 participants took part in the study. All reported normal hearing and normal or corrected-to-normal vision. All participants had English as their first language and were not paid for their time.

9.3.2 Stimuli

9.3.2.1 Visual stimuli

Lists of digits to be recalled were constructed in the same way as detailed in chapter 3 (appendix 1).

9.3.2.2 Auditory stimuli

The non-words presented for the whispered and FSR whispered speech conditions were /gam/ (g{m) and /rarn/ (r£n) (see appendix 5 for examples of disc phonetic symbols). The whispers for both auditory conditions were recorded digitally and edited as described in chapter 3 (p90). For the FSR whispered speech, the fine structure was temporally reversed, but the original amplitude envelope was preserved using custom software. Fine structure reversal resulted in sounds that were matched to the normal whispered sounds on overall spectral content, duration, intensity and acoustic complexity. Therefore, FSR whispers only differed from whispers by the removal of their intelligibility. The RMS sound levels of both types of sounds were equated as in experiment 3 and 4 (chapter 7). The presentation and duration of these irrelevant sounds was as described in chapter 3.

9.3.3 Pilot listening test

12 participants listened to seven FSR whispered non-words (appendix 31). Each non-word was played once and participants were asked to describe what they heard. Two of the reversed non-words were not identified as non-words; however participants reported hearing these sounds as strange, scary and unfamiliar sounds "produced by a voice". Participants referred to these two reversed non-words as sounding like sounds you would here in a horror film and that they sounded like they came from a human voice.

9.3.4 Design and procedure

The design and procedure was the same as detailed in chapter 3.

9.4 RESULTS

Experimental condition	*Mean Errors*	*SD*
Whispers	38.70	27.921
FSR whispers	39.53	31.898
Silence	23.60	21.586

Table 15. Descriptive statistics for 3 auditory conditions; mean number of serial recall errors per condition. N = 30

The descriptive statistics in table 15 indicate that whispered and reversed whispered speech has a similar effect on serial recall performance and that serial recall performance is poorer under both conditions relative to that observed under the silent control. The data is summarised in figure 25.

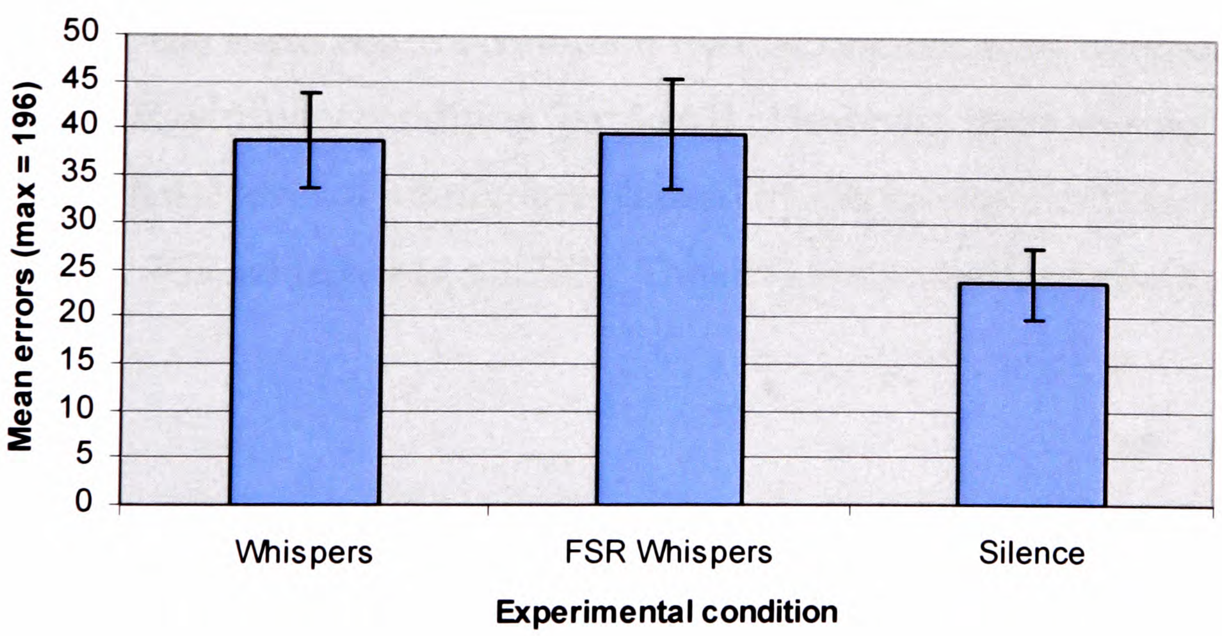


Figure 25. Mean number of serial recall errors for the 3 experimental conditions. Error bars represent standard error above and below the mean.

	Silence	*Whispers*
FSR Whispers	✓ $p < 0.001$	Non-sig $p \leq 1.000$
Whispers	✓ $p < 0.001$	xx

Table 16. Bonferroni corrected pairwise comparisons for the three experimental conditions. (FSR whispers = fine structure reversed whispers).

A within-subjects ANOVA with whispers, FSR whispers and silence as the levels for the factor of irrelevant sound (appendix 32) was performed on the data. Mauchy's test revealed that the assumption of sphericity had been violated ($\chi^2(2) = 12.266, p < 0.01$); therefore the degrees of freedom were corrected using Huynh-Feldt estimates of sphericity. The ANOVA showed that irrelevant sound disrupted immediate serial recall relative to the silent control [F (1.536, 44.558) = 18.911, MSE = 3140.766, $p < 0.001$]. Pairwise comparisons following Bonferroni correction (appendix 32) as shown in table 16 revealed that the silent control condition differed significantly from the whispers condition ($p < 0.001$). The silent control condition was also significantly different from the FSR whispers condition ($p < 0.001$). However, there was no difference in the level of interference caused by the presence of irrelevant whispers or FSR whispers ($p \leq 1.000$). The data is summarised also in figure 26.

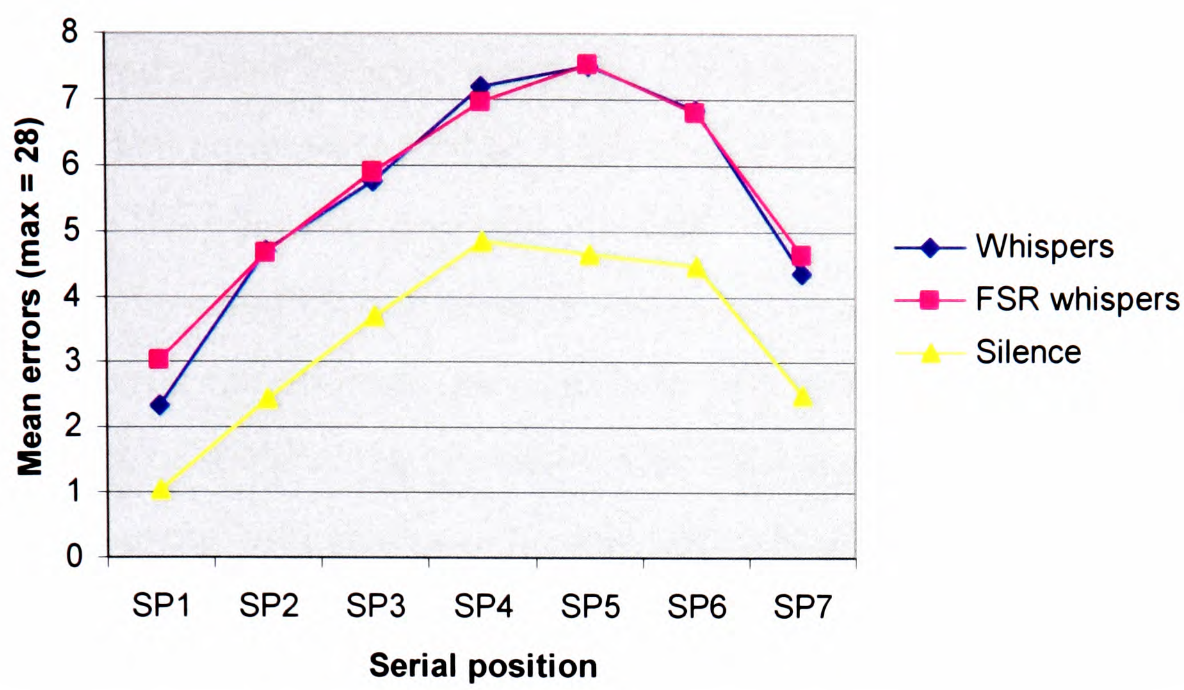


Figure 26. Mean serial recall errors for each experimental condition collapsed across serial position.

9.5 DISCUSSION

The finding of no difference in the disruptive effect of whispers and FSR whispers may be accounted for by the presence of some phonetic information within the reversed fine structure. When the whole speech signal is reversed, some phonetic detail remains present (Scott and Wise, 2004). Therefore, time reversal of the fine structure is not sufficient enough to remove the phonemic information. Amplitude modulation in speech conveyed by the waveform envelope was maintained and thus the rise and fall of amplitude was unaltered. Information relevant to the perception of speech sounds is carried by structures within the speech signal (e.g. formants) and the onsets of phonemes (Scott and Wise, 2004). It follows that features inherent within vowel onsets determine the rhythm of speech sounds. When speech is reversed in time the temporal-spectral structure of this information is distorted. The temporal structure of a phoneme determines the how much variation it conveys. It is for this reason that most of the steady-state information for vowels will remain unchanged (Scott and Wise, 2004). One inference that can be drawn from this finding is that the changing-state information provided by the reversed steady-state information in the FSR whispers was sufficient to produce an ISE equivalent to that produced by normal whispers. Although in the pilot listening task, the FSR whispered speech sounds were not understood by participants who were unable to repeat them back, it seems some phonetic features were still present within the signal. Maintenance of some of the phonetic information, although diminished relative to normal whispers, can be inferred as FSR whispers were described by participants as noise produced by a voice. In addition no frequency-specific spectral features, which are argued to be important for speech perception (Shannon et al., 1995), were removed from the FSR whispered sounds. Therefore, although FSR whispers cannot be articulated, it seems reversing the fine structure of whispers creates phonotactically illegal sounds, which is apparent for speech where the whole signal is reversed in time (Scott and Wise, 2004). Intelligibility of speech sounds, in terms of being able to accurately repeat them back and hence be able to accurately identify individual phonemes does not seem to be important to the ISE. Rather distinct stimuli, such as FSR whispers in a changing sequence need to be perceived as 'speech-like' in order for them to disrupt memory to the extent that normal whispered speech and voiced speech does. Speech-likeness would still be perceived if sounds are heard as being produced by a voice.

The spectrograms for the whispered and FSR whispered non-word /rarn/ are shown in figure 27a and 27b. Figure 27b shows reversing the fine structure preserves the formants.

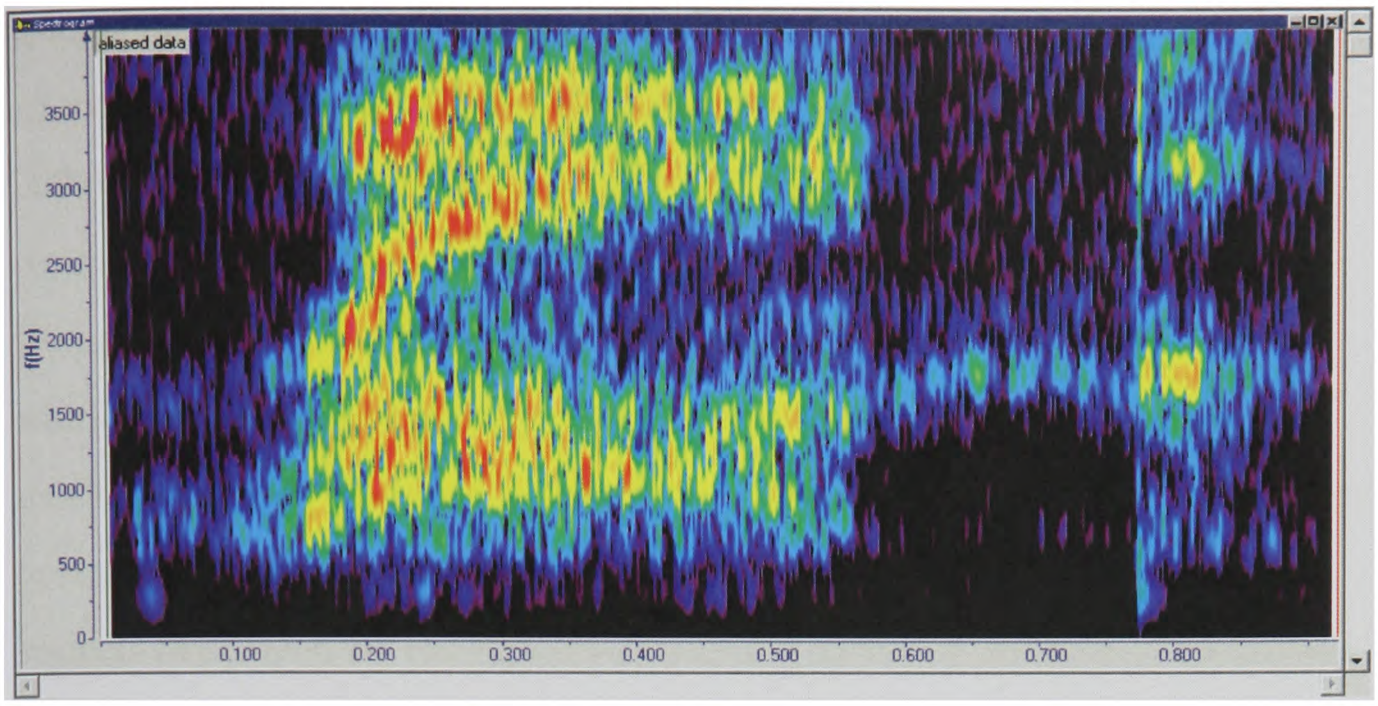


Figure 27a. Spectrogram of the non-word /rarn/ whispered by a female voice.

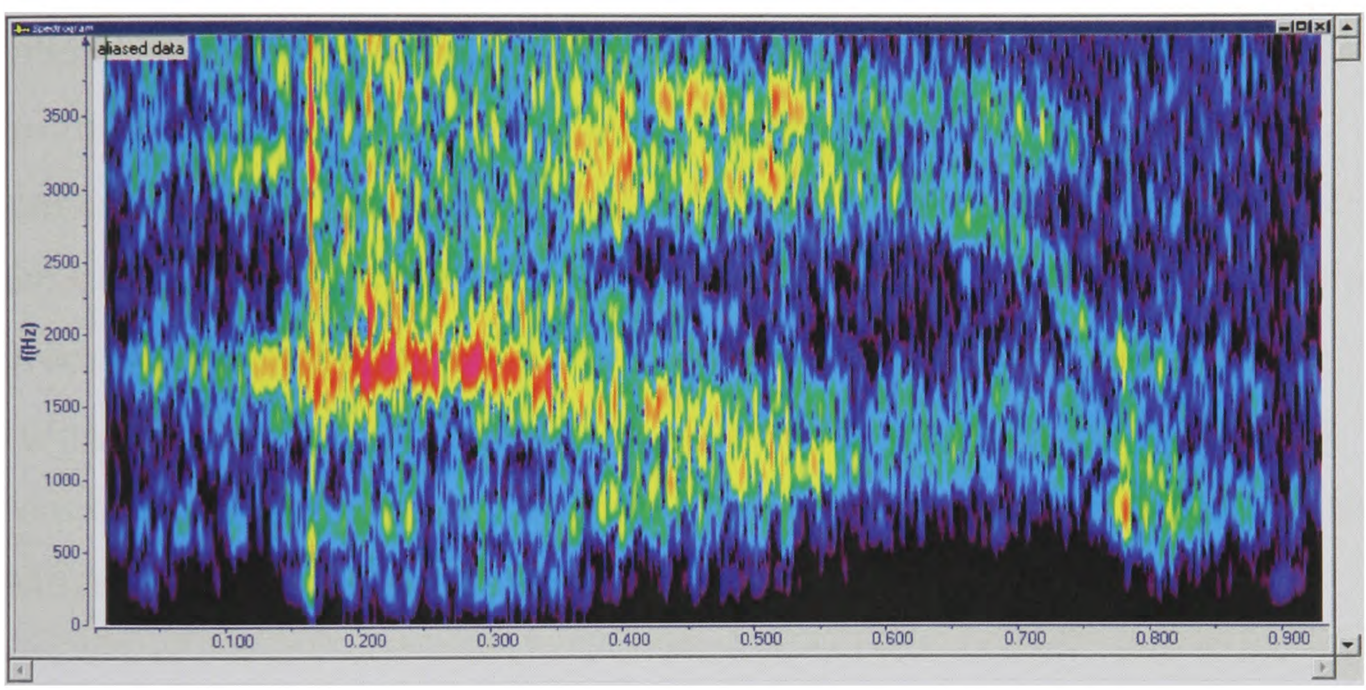


Figure 27b. Spectrogram of the FSR non-word /rarn/ whispered by a female voice.

Experiment 3 (chapter 7) found no difference between serial recall performance when voiced and whispered speech was presented as irrelevant sound. Even with *f*0 information missing from whispered speech, time-varying changes in the formants of the irrelevant whispers were present and these seem to have afforded the coherence of the temporal order of irrelevant whispers. Change on a common fundamental, such as *f*0 and formant structure is argued to help the perceptual system maintain the temporal order of speech sounds (Bregman, 1990). That changing vowels are more disruptive of memory than changing consonants has been explained by the importance of vocal tract changes superimposed on $f0$ and formant structure between successive sounds produced by the same voice over time (Hughes et al., 2005). Even though the formants of the FSR whispered speech were reversed, which distorted their structure in the time domain, no frequency-specific information was removed from the signal of these sounds. It seems the presence of the reversed formants and generally unaltered steady-state portions of the changing vowels still provided the perceptual system with the information required to integrate overtime sounds produced by the same voice into a coherent stream. This would lead to the temporal order of the FSR whispers being maintained. It follows that the object-oriented episodic record (O-OER) model's interference by processes account, an extension of the CSH, can account for these findings if it assumes the FSR whispers would still provide a similar amount of serial order cues as does whispered and normal speech whose component vowels change from item-to-item. The interference by process account argues it is the ease with which the perceptual system automatically encodes the temporal order of changing-state unattended sounds that determines the number of serial order cues that will compete with the seriation of the TBR items in short-term memory (STM).

The maintenance of the abrupt changes in the sound patterning inherent within the amplitude envelope of the whispered speech sounds may also account for why no difference was observed between the two whispered speech conditions. The waveform envelope is argued to convey information about the phonemes within speech (Moore, 2004). The perceptual importance of amplitude modulation is not specific to speech. The patterning of amplitude as it rises and falls has been shown to be important for the perception of the attack time of musical tones (Gordon, 1987) and is observed to be asymmetric in the auditory system (Irino and Patterson, 1996). In the irrelevant sound paradigm, highly reverberated speech has been found to be less disruptive than normal speech (Beaman and Holt, 2007). This has been explained in terms of reverberation smoothing the profile of the speech waveform and thus reducing significant variations in the amplitude envelope. Passive listening showed that highly reverberated speech was still intelligible, unlike other forms of degraded speech (e.g. Jones et al., 2000 and experiment 1b, chapter 5). However, no formal screening of the intelligibility of the reverberated speech was performed. As well as soothing the profile of the speech waveform, reverberation acts also to smear the harmonicity of the speech signal (Wu and Wang, 2006). Periodicity of the signal is removed by increasing room reverberation time (Roman and Wang, 2005). Formant frequencies seem to be the important carriers of change between successive sounds in an irrelevant sequence as indicated by experiment 3 and 4 (chapter 7). Formants can be harmonics, although not all harmonics are formants and are instead overtones (Moore, 2004). If reverberation serves to corrupt the harmonic structure of a speech signal; it may also smear formant structure which may account for the reduction in the ISE observed under high levels of reverberation.

Research shows that spectral and temporal cue distortion affects vowel and consonant perception differently. Temporal smearing of the speech signal has been found to have a larger detrimental effect on consonants than vowels (Drullman et al., 1994). Fine structure reversal of the whispered signal only distorted spectral detail in the time domain and thus a greater effect of this manipulation on consonant recognition as opposed to vowels would account for the equivalent serial recall performance observed under whispers and reversed whispers. Experiments with signal-correlated-noise (SCN) have removed all spectral detail and only maintained the broadband temporal envelope. Recognition of consonants was high in the absence of spectral detail (Shannon et al., 1995; Turner et al., 1995). Shannon et al (1995) filtered speech to produce a high-pass and a low-pass band divided at 1500Hz. The envelope from the low-pass band was then used to modulate the envelope of a low-pass band of noise and vise versa. This generated two bands of speech modulated noise. Consonants were correctly recognised in SCN with only two bands of modulated noise and so it is argued that contrasts in these consonants can be perceived well even with little spectral detail. However, more spectral detail was needed for accurate vowel and consonantal place of articulation (Shannon et al., 1995). This indicates that the correct identification of vowels is determined more by spectral detail than it is by temporal cues, such as fine distinctions in the onset time of vowels.

9.6 SUMMARY

Reversing the fine structure of whispered speech does not act to reduce the ISE observed with normal whispers. There are two accounts of the data. First the O-OER model's interference by process account of the ISE (c.f. Jones and Tremblay, 2000) can explain the present findings if the formant structure of FSR whispers is considered. Although the spectral information pertaining to the formants is reduced in whispered speech, further distortion by temporal reversal leaves some of the steady-state information inherent within the formants of vowels relatively unaltered (Scott and Wise, 2004). Formant frequencies give vowels their identity. Formant structure, provided by the vowel sounds spoken by the same voice over time, is a fundamental shared by the speech sounds. Change on a common attribute of speech sounds (e.g. formant structure) is argued to help the perceptual system maintain the temporal order of speech sounds (Bregman, 1990). It has been argued that change between the vowels of speech sounds that is carried on a common attribute (such as formant structure) determines the disruption of serial recall by irrelevant speech. This is because vowel-only-changing (V-O-C) irrelevant speech sounds have been found to produce more serial recall interference than consonant-only-changing (C-O-C) speech sounds (c.f. Hughes et al., 2005; Experiment 2, chapter 6). The interference by process account of the O-OER model argues that vowel changes yield more cues to the serial order of irrelevant sounds than do consonant changes. It is suggested that these cues to the order of the speech sounds are automatically encoded and as a consequence conflict with the cues pointing to the serial order of the TBR digits, thereby reducing serial recall performance (Hughes et al., 2005). That temporal reversal does not distort formant structure to a sufficient degree that would remove the important changes on this common fundamental may be due to the considerable spectral complexity in the patterning of the formants which are resistant to temporal reversal.

Second, although the FSR whispers were not understood or heard as non-words by listeners, they were still heard as "sounds produced by a voice". Whispers whose fine structure is reversed form a poor comparison to normal whispers in terms of phonemic information. This can be explained with reference to the steady-state information related to the vowels being maintained in the FSR whispered signal. Although the two classes of whispered sounds were matched for acoustic complexity only, the fact that some phonological information was still perceptible indicates that if a distorted sound is perceived as coming from a voice it will have the same impact on memory as speech whose constituent phonemes are intelligible. The steady-state information within the changing vowels in an irrelevant speech stream is important. When speech is degraded (e.g. experiment 1b, chapter 4 and experiment 2, chapter 6) information in the vowels is degraded and serial recall performance in its presence is reduced relative to serial recall performance in the presence of clear speech. The importance of information within changing vowels was also demonstrated by V-O-C sequences being more disruptive than C-O-C sequences. Further, degradation had an effect on V-O-C sequences. As V-O-C sequences were degraded, serial recall performance improved. However, Degradation had no effect on C-O-C sequences (experiment 2, chapter 6).

It can be concluded that if the steady-state information inherent within the formants of the vowels of irrelevant speech is preserved, it will be perceived as sound produced by a voice. As such, it seems the intelligibility of speech in terms of being able to accurately repeat it back and identify individual phonemes is not important. Rather, the formant frequency information provided by the changing vowels in irrelevant speech seems to be a factor which may be responsible for the observed greater serial recall disruption by speech relative to non-speech sounds (e.g. Jones et al., 2000).

CHAPTER 10

10 PRELIMINARY INVESTIGATION OF MEMORY DISRUPTION BY SPEECH AND NON-SPEECH: MATCHING ACOUSTIC COMPLEXITY

10.1 BACKGROUND

Experiment 5 (chapter 9) examined whether reversing the spectral detail within the fine structure of whispered speech in the time domain whilst maintaining its original amplitude envelope would render whispered speech less disruptive of serial recall than when it is presented in its original form. Reversing the fine structure of whispered stimuli resulted in a signal conveying the same acoustic complexity as the original whispered stimuli. Both classes of whispered stimuli had the same long-term average spectrum and exhibited an equivalent amount of overall amplitude modulation. However, reversing the spectral detail (fine structure) of whispers disrupted the temporal structure of whispers and generated stimuli that could no longer be articulated. It follows that the fine structure reversed (FSR) whispers were not intelligible to listeners. During intelligibility screening, listeners were unable to understand and repeat back the reversed whispered stimuli. More importantly, listeners did not hear these as non-words or any other type of word.

Despite the FSR whispered stimuli of experiment 5 not being heard as non-words, they disrupted recall to the same extent as the same words in their original whispered form. Although reversed speech exhibits reduced phonetic information (as sound sequences are produced that do not occur naturally in spoken speech and are thus heard as unfamiliar by listeners) some phonetic detail remains in the signal (Binder et al., 2000). Changing vowels in an irrelevant speech stream are more disruptive of serial recall than are changing consonants (see experiment 2, chapter 6; Hughes et al., 2005). Therefore, it can be argued that FSR whispers are as disruptive of serial recall as normal whispers because spectral detail pertaining to the changing vowels, in particular energy at the formants is preserved. During the intelligibility screening session listeners reported hearing the FSR whispered stimuli as "sounds produced by a voice", though they were unable to repeat them back. This makes evident that the FSR whispers still exhibited speech-like characteristics. In support of this idea, Binder et al (2000) found listeners were able to extrapolate phonetic information from reversed speech during a transcription task and assumed phonetic categorisation processes may have been in operation.

Many phonemes are relatively temporally symmetrical (fricatives and long duration vowels). Examination of the spectrograms (figures 27a and 27b, p209) in chapter 9 for the non-word /rarn/ (r£n) in its original and FSR form show roughly mirror reversal of formant transition structure into and out of the vowel portion of the signal. Although reversing spectral detail in the time domain reduces phonetic information as no non-words are heard by listeners, the fact that the phonetic information present allows them to be heard as sounds produced by a voice may explain why they disrupted recall in the same way as did untransformed stimuli.

Intelligibility in terms of comprehension and the decoding of words from a signal is therefore not of importance with regards to determining the disruption caused by irrelevant speech sounds. Rather, sounds which are heard as speech-like, and thus sounds produced by a voice seem to be as disruptive to serial recall as sounds readily intelligible as words or non-words. It seems formants are the important acoustic-phonetic characteristics of speech that carry the necessary changes between successive speech-like stimuli that result in eliciting the strong serial order cues suggested to disrupt the seriation of TBR items (c.f. Hughes et al., 2005). In essence, intelligibility of speech sounds in the ISE paradigm refers to the speech-likeness of sounds.

Speech is an extremely complex acoustic signal, consisting of multidimensional variations. Speech has a complex changeable temporal structure. The presence or absence of voicing creates the characteristic quasi-periodic and aperiodic (noise) components of the speech signal. It exhibits relatively continuous amplitude and frequency modulations and conveys a complex spectral structure signified by the formants which are generated by the movement of the articulators (Narain et al., 2003). It follows that a constantly varying distribution of energy over time is seen over the spectrum due to vocal tract resonances (speech formants).

The acoustic-phonetic cues of speech need to be processed in order for the signal to be perceived as intelligible, leading to the decoding of words (Narain et al., 2003). No single acoustic feature of the speech signal determines its intelligibility, due to there being no simple mapping between acoustic features and phonetic identity (Bailey and Summerfield, 1980), therefore the speech signal is resistant to degradation. Skilled listeners find intelligibility to be good for a signal highly spectrally and temporally degraded, whilst a certain amount of spectral-temporal modulation is crucial (Shannon et al., 1995). However, signal-correlated-noise (SCN) consisting of only two speech modulated bands of noise allows for consonantal contrasts to be well distinguished, whilst vowel recognition is poor. This indicates that accurate recognition of vowels requires more spectral detail (Shannon et al., 1995).

Experiment 1b (chapter 5) showed how phonological degradation of irrelevant speech reduced its intelligibility and resulted in it interfering with serial recall performance less. However, phonological degradation not only reduced the intelligibility of the speech sounds, but also its acoustic complexity. Turning a proportion of the signal into noise would have resulted in distorting the structure of the formants, which have been shown to be important carriers of change between successive items, as far as the disruption of serial recall by irrelevant speech is concerned (c.f. experiment 3 and 4 in chapter 7 and experiment 5 in chapter 9). Therefore, designing sounds that are as acoustically complex as speech, but lack phonetic features and are unintelligible is difficult, given that acoustic and phonetic features are interlinked in the speech signal.

10.2 AIMS AND OBJECTIVES

Consideration of the complex acoustic structure of speech is important when looking at serial recall interference of irrelevant speech, since the number of changes from item-to-item in an irrelevant stream is argued to modulate the magnitude of the ISE (Jones et al., 1996). Experiment 5 (chapter 9) controlled for acoustic complexity across two auditory conditions (whispers and FSR whispers), but did not adequately control for phonetic information as the relationship among formants was preserved. Pilot experiment 6 aims to examine the relative disruptiveness of irrelevant sound matched for acoustic complexity, but not intelligibility by implementing the established signal processing technique of spectral rotation (Blesser, 1972) to remove the intelligibility of speech by destroying the relationship between formants.

10.3 SPECTRALLY ROTATED SPEECH

Spectrally rotating speech around a centre frequency of 2 kHz maintains the acoustic complexity of the speech signal by preserving its temporal and spectral structure, whilst at the same time it is unintelligible (Narain et al., 2003; Scott et al., 2000; Scott and Wise, 2004). The spectral and temporal properties of the speech signal are maintained, but now occur in different frequency channels. For example, the high frequency amplitude modulations are at lower frequencies, and the low frequency amplitude modulations are at higher frequencies. It follows that whilst this transformation preserves the range of frequencies, it does not maintain the relative spacing of the formants (Lachs and Pisoni, 2004). Spectrally rotated speech is only found to be intelligible by some listeners after extensive training pertaining to weeks or months (Narain et al., 2003) and has been described by participants as sounding like an *alien language* produced by articulators distinct from those of the human vocal tract (Blesser, 1972). It follows that spectrally rotated speech sounds do not exhibit intelligible phonetic features. It can be construed therefore that whilst spectrally rotated speech maintains some of the acoustic features that might reflect the acoustic correlates of phonetic information it is successful in destroying intelligibility (Scott and Wise, 2004).

The dynamic pitch variation conveyed by the original speech signal is also maintained. The spectrum of untransformed speech is characterised by frequency components (harmonics) which are integer multiples of the fundamental frequency ($f0$) of the speaker. Speech transformed by spectral rotation still conveys equally spaced component frequencies, however these are no longer multiples of the $f0$ and the true periodicity of the signal is destroyed (Scott et al., 2000). Because, the range of frequencies due to vocal fold vibration is preserved they still elicit a fairly strong sensation of perceived pitch (Blesser, 1972; Scott et al., 2000).

10.4 PILOT EXPERIMENT 6: METHODOLOGICAL CONSIDERATIONS

10.4.1 Participants

24 participants took part in the study. All reported normal hearing and normal or corrected to normal vision. All participants had English as their first language and were not paid for their time.

10.4.2 Stimuli

10.4.2.1 Visual stimuli

Lists of 7 TBR digits were constructed and displayed as described in chapter 3 (appendix 1).

10.4.2.2 Auditory stimuli

The same two monosyllabic CVC words were used for the speech and rotated speech condition, these were /birds/ and /tree/. Research shows, meaningful speech is no more disruptive of serial recall than meaningless speech (Buchner, Irmen, and Erdfelder, 1996). The original speech sounds were spliced from a low-pass filtered (3.8 kHz) version of the spoken sentence "the birds sang from the tree" recorded by a male speaker. This sentence was taken from the IHR ASL sentence list (Scott, Rosen, Wickham and Wise, 2004). The non-words were spliced from the spectrally rotated version of the original sentence. The original sentence, sampled at 22.05 kHz, was low passed filtered at 3.8 kHz and then spectrally rotated around 2 kHz using a digital version of Blesser's (1972) simple modulation method. The spectrally rotated signal was first passed through an equalising filter that was essentially a high-pass filter in order to equate the long term spectra of the spectrally rotated signal to that of the original signal. The equalised signal was then amplitude modulated by a sinusoid at 4 kHz and then low passed filtered at 3.8 kHz. The sounds were presented as described in chapter 3.

10.4.3 Pilot listening test

The original and spectrally rotated versions of the two words /birds/and /tree/were each presented twice to 10 participants who were instructed to write down a description of what they heard. The non-words were not heard as speech by any of the participants, whereas the original speech versions were 100% intelligible and all participants were able to accurately understand and repeat back what they heard.

10.4.4 Design and Procedure

The design and procedure used was as detailed in chapter 3 (p93).

10.5 RESULTS

Experimental condition	*Mean Errors*	*SD*
Speech	40.04	29.443
Rotated speech	33.33	27.508
Silence	24.67	31.873

Table 17. Descriptive statistics for 3 experimental conditions; mean number of serial recall errors per condition. N = 24 (rotated speech= spectrally rotated speech).

The descriptive statistics in table 17 shows that speech was slightly more detrimental to immediate memory than spectrally rotated speech and both speech and spectrally rotated speech disrupted memory relative to a silent control. The data are summarised in figure 28. Figure 28 shows a trend is evident in the data which is in support of the prediction that speech sounds will be more disruptive than their spectrally rotated counterparts.

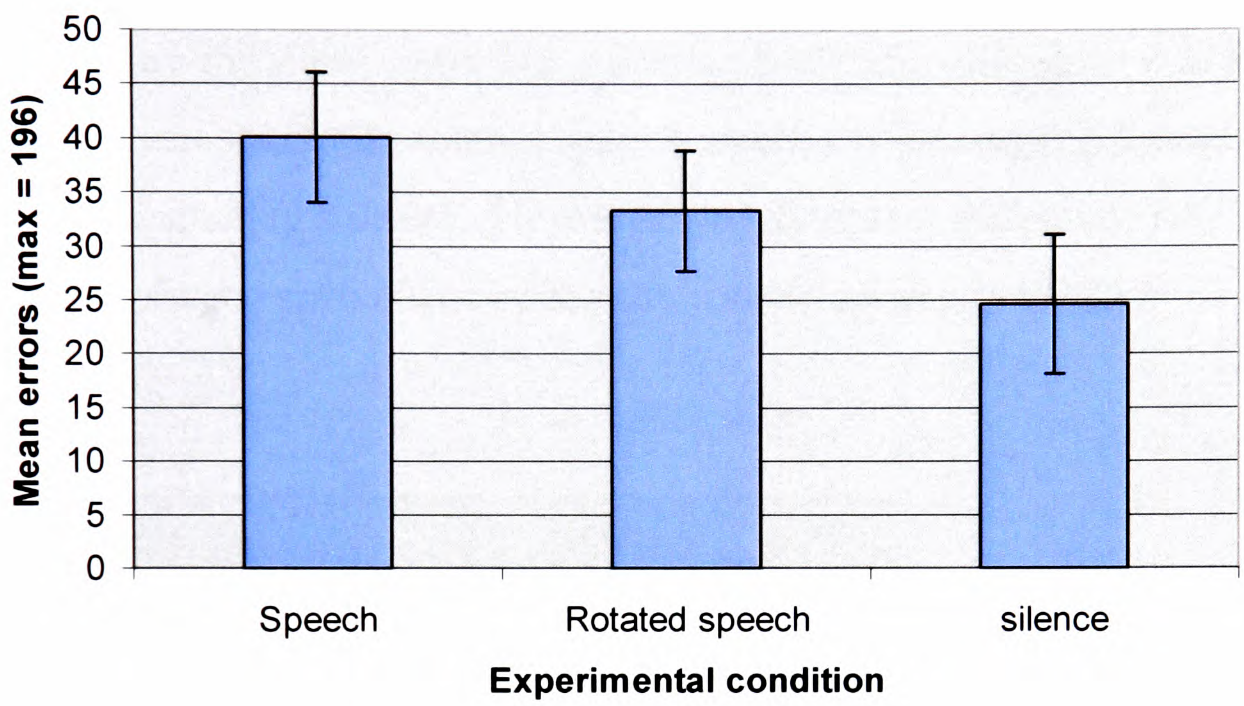


Figure 28. Mean errors per list for the three experimental conditions. Error bars represent standard error above and below the mean (rotated speech= spectrally rotated speech).

	Silence	*Speech*
Rotated speech	✓ $p \leq 0.046$	Non-sig $p \leq 0.120$
Speech	✓ $p < 0.001$	xx

Table 18. Bonferroni corrected pairwise comparisons for the three experimental conditions (rotated speech = spectrally rotated speech).

The mean number of digits incorrectly recalled for each experimental condition were subjected to a repeated measures ANOVA with 3 levels (speech, spectrally rotated speech and silence). A main effect of irrelevant sound on serial recall performance was found as irrelevant sound impaired serial recall relative to a silent control [F (2, 46) = 12.193, MSE = 1426.014, $p < 0.001$] (appendix 33). Figure 29 shows the overall level of errors in recall collapsed across serial position. Pairwise comparisons with Bonferroni correction as displayed in table 18 were performed on the data to identify which conditions differed statistically (appendix 33). It was found that the speech condition was significantly different from the silent control ($p < 0.001$). A reliable difference was also found between spectrally rotated speech and the silent control, though this was marginal ($p \leq 0.046$). However, no significant difference was observed between speech and spectrally rotated speech ($p \leq 0.120$).

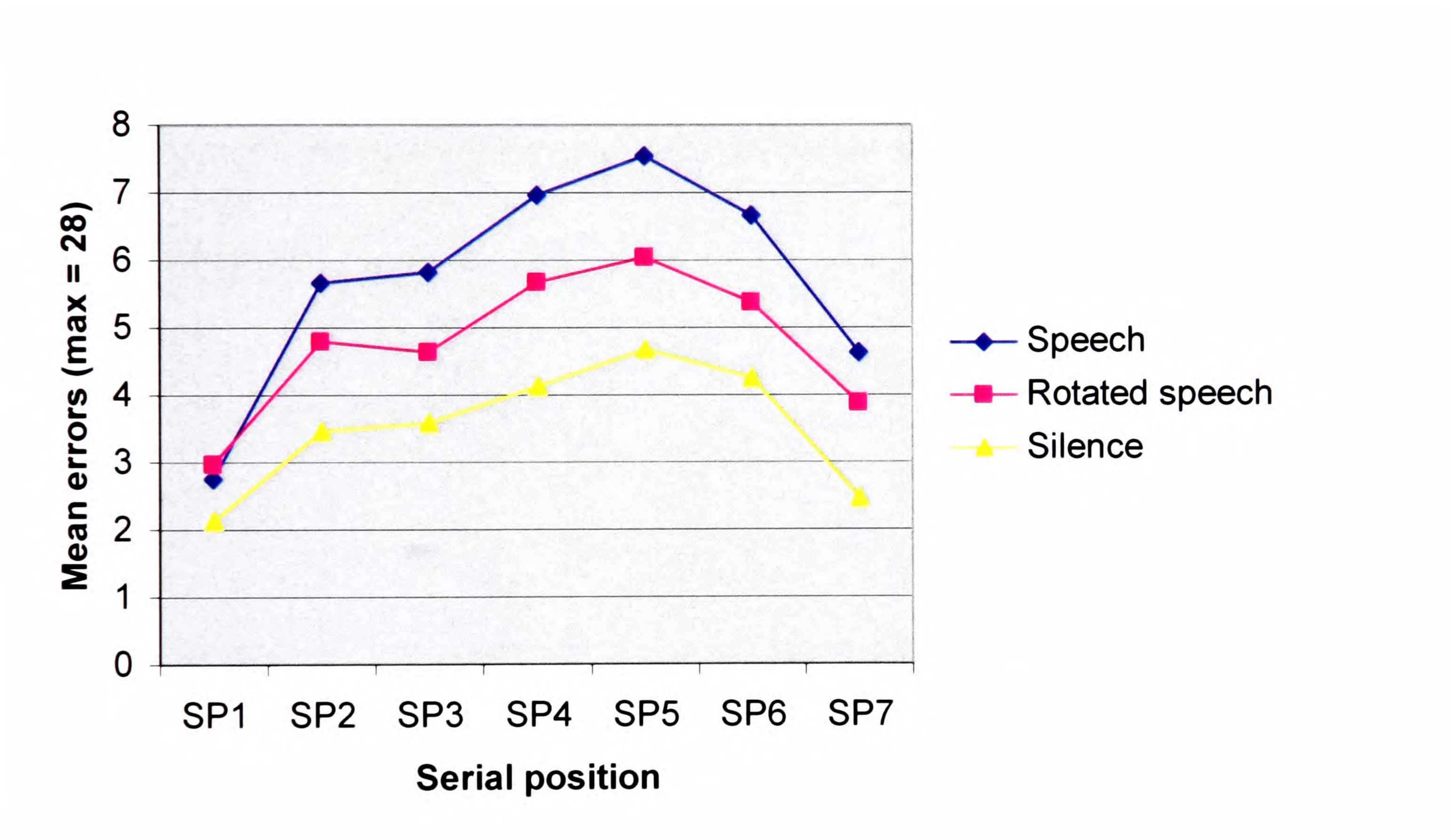


Figure 29. Mean errors for the 3 experimental conditions collapsed across serial position (rotated speech= spectrally rotated speech).

10.6 DISCUSSION

The equivalent level of serial recall disruption produced by speech and spectrally rotated speech is consistent with the CSH, which argues the degree of acoustic variation determines disruption. As spectrally rotated speech maintains the spectral and temporal structure of speech it conveys the same degree of acoustic variation as untransformed speech. The CSH would therefore predict that stimuli exhibiting the same degree of change from item-to-item would impair memory for serial order to an equivalent degree.

The absence of a reliable difference between speech and spectrally rotated speech indicates that intelligibility may not play a role in modulating the effect of irrelevant sound, since spectrally rotated speech maintains the acoustic complexity of speech but is not heard as speech. Figures 30a and 30b display the spectrograms for the untransformed and spectrally rotated version of the word /tree/ (rotated around 2 kHz). The spectrograms show that the spectrally rotated version of /tree/ is a mirror image of the untransformed version of the word.

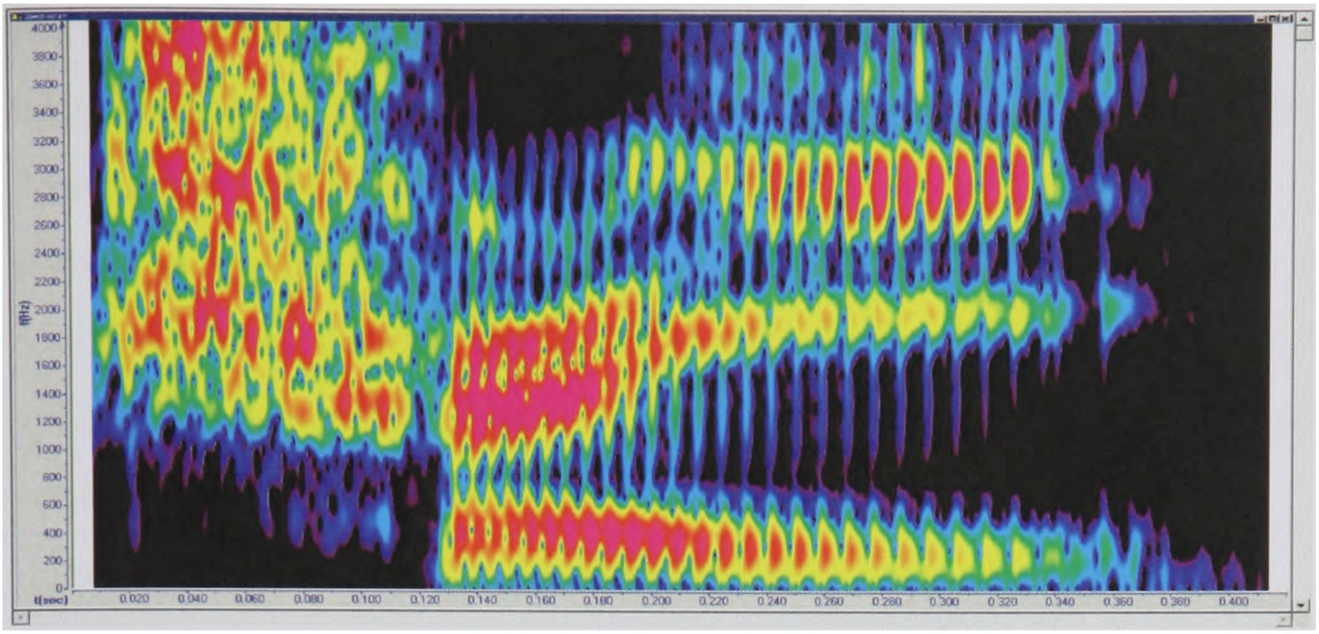


Figure 30a. Spectrogram of the untransformed version of the word /tree/spoken by a male voice.

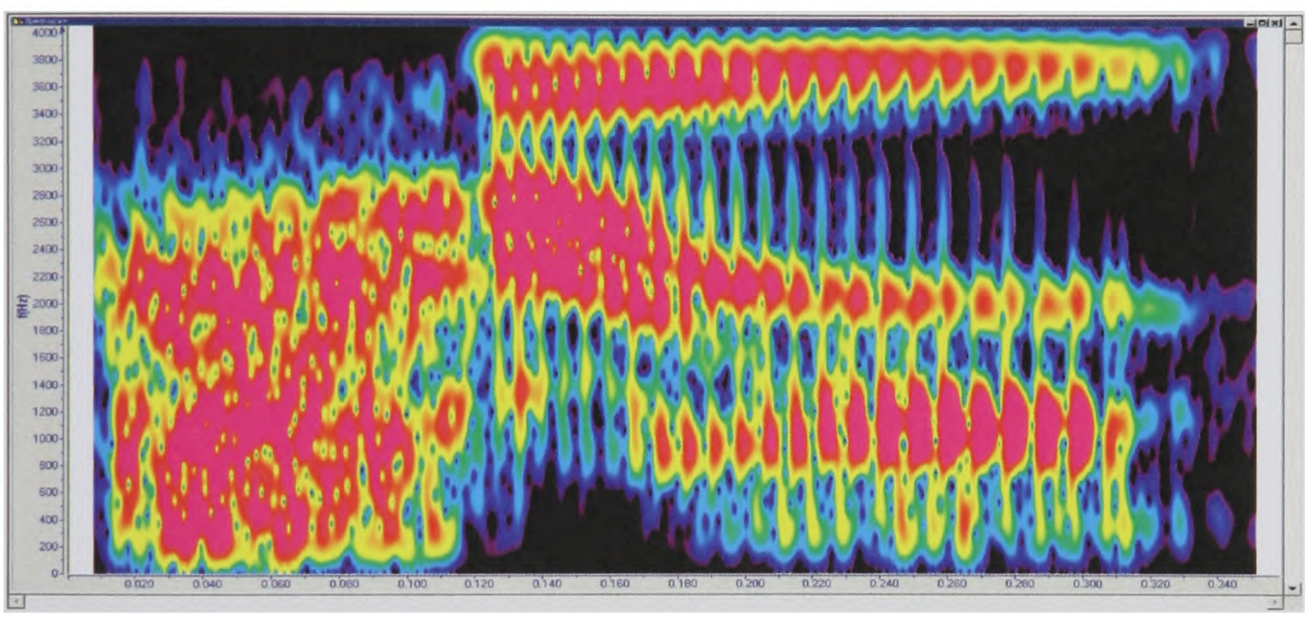


Figure 30b. Spectrogram of the spectrally rotated version of the word /tree/ spoken by a male voice.

Although no reliable difference was found between the disruptive effect of speech and spectrally rotated speech, both figure 28 and 29 indicate a clear trend in the predicted direction. Table 17 shows there is a numerical difference between the mean number of serial recall errors in speech (40.04) and spectrally rotated speech (33.33). The spectrogram of the untransformed version of the word /tree/ in figure 30a shows energy across the frequency domain is distributed and so when spectrally rotated, energy remained present at low frequencies as evident in figure 30b. The distribution of spectral information in the frequency domain provides critical speech pattern information (Shannon et al., 1998).

Although when the word /tree/ was spectrally rotated it was no longer heard as speech, it may be that the distribution of energy in the frequency domain may not have been distorted sufficiently for the perceptual system to treat the word as a non-speech pattern. It is difficult to draw conclusions considering the words used in both speech conditions were spliced from a sentence and thus acoustic complexity was not adequately controlled. Further investigation with stimuli that are adequately matched for acoustic complexity is therefore required.

CHAPTER 11

11 MEMORY DISRUPTION BY SPEECH AND NON-SPEECH: MATCHING ACOUSTIC COMPLEXITY

11.1 BACKGROUND

Shannon et al (1998) showed that speech pattern recognition is not robust to all distortions in the tonotopic pattern. Distorting spectral detail of speech by shifting and warping the spectral distribution of envelope cues reduced vowel recognition more than consonant recognition. Consonant recognition was mostly high in these conditions but vowel recognition was reduced to a level observed under single spectral band conditions. Research showing that vowel perception is poorer than consonant perception under conditions where the spectral distribution of envelope cues is shifted or warped suggests that the removal of speech-likeness may be achieved if the spectral distribution of envelope cues is distorted to the extent that vowels as well as consonants are no longer perceived. Pilot experiment 6 (chapter 10) found no difference between the effect of speech and spectrally rotated speech on serial recall and yet the relative spacing of the formants was not preserved. This would not be expected in light of the findings of Shannon et al (1998) that shifting and warping the spectral distribution of envelope cues reduced vowel perception. It is changing vowels as opposed to consonants in an irrelevant speech stream that determine the disruptive power of speech (Hughes et al., 2005). The absence of a difference between speech and spectrally rotated speech is not consistent with research indicating the importance of the relationship among formants. Lachs and Pisoni (2004) observed that spectral rotation of speech, which destroys the patterning of formants over time, removes the information needed for word recognition. Poor word identification was found for spectrally rotated words. This was seen as evidence that word recognition required the information carried in the spectral structure of formants as they vary over time.

11.2 AIMS AND OBJECTIVES

The aim of the present experiment is to re-examine the disruptive effect of speech and spectrally rotated speech using sounds that are better controlled acoustically. The sound streams in the previous pilot experiment (chapter 10) were spliced from a recording of a sentence produced by a male speaker. Splicing the words from the sentence meant that acoustically the two words in the speech stream and their spectrally rotated counterparts were not adequately matched. In experiment 6 non-words spoken by a male speaker in neutral intonation were transformed in order to destroy intelligibility whilst maintaining acoustic variation.

11.3 EXPERIMENT 6: METHODOLOGICAL CONSIDERATIONS

11.3.1 Participants

30 participants took part in the study. All reported normal hearing and normal or corrected-to-normal vision. All participants had English as their first language and were not paid for their time.

11.3.2 Stimuli

11.3.2.1 Visual stimuli

Lists of digits to be recalled were constructed in the same way as described in chapter 3 (appendix 1).

11.3.2.2 Auditory stimuli

11 non-words (appendix 34) were recorded and edited as detailed in chapter 3 (p90). The spectral rotation transformation (Blesser, 1972) was applied to the sounds as described in pilot experiment 6 (chapter 10, p220). An intelligibility screening test was used to isolate non-words that were not heard as speech. 12 participants were required to listen to the 11 non-words over headphones. Each non-word was presented twice and participants were instructed to provide a written description of what they heard on a response sheet. The non-words /teash/ (tiS) and /forb/ (f$b) were consistently heard as non-speech by all 12 listeners. Listeners typically described these sounds as complex noise and more importantly, no references to the speech-likeness of sounds were made. The two spectrally rotated non-words /teash/ and /forb/ not heard as speech were isolated for the spectrally rotated speech condition. The non-words /marv/ (m£v) and /curj/ (k3_) made up the speech condition (see appendix 5 for examples of disc phonetic symbols).

The decision to use different non-words in the two sound conditions was informed by the literature which shows that the phonological content of stimuli does not determine the degree of serial recall disruption. Jones et al (1990) demonstrated that speech played forwards, reversed speech and speech in a foreign language (welsh) disrupt serial recall to the same degree. Second, Blesser (1972) reports that spectrally rotated speech can become intelligible with training on the order of weeks. The term intelligibility in research examining the potential for spectrally rotated speech to become intelligible (Blesser, 1972) refers to the identification of words. Experiment 5 demonstrated that FSR whispered non-words that were not heard as non-words but were heard as sounds produced by a voice disrupted memory to the same extent as whispered non-words. It may be the case that spectrally rotated speech sounds can be recognised as sounds produced by a voice after training over a much shorter period than the amount of time needed for spectrally rotated words to be identified. As this has not yet been investigated, this also informed the decision to use different non-words in the speech and spectrally rotated speech conditions. Using different non-words in the sound conditions avoids any possibility of those participants presented with the speech condition before the spectrally rotated speech condition recognising the spectrally rotated non-words as voiced sounds.

The non-words in both speech conditions were constructed from different phonemes to allow maximum acoustic variation between stimuli. A silent condition acted as the control as in previous experiments. Figure 31a and 31b show the spectrograms for the original and spectrally rotated version of the non-word /teash/ and figure 31c and 31d show the spectrograms for both versions of the non-word /forb/. It is clear from the spectrograms that the spectrally rotated versions of /teash/ and /forb/ are mirror images of their untransformed counterparts.

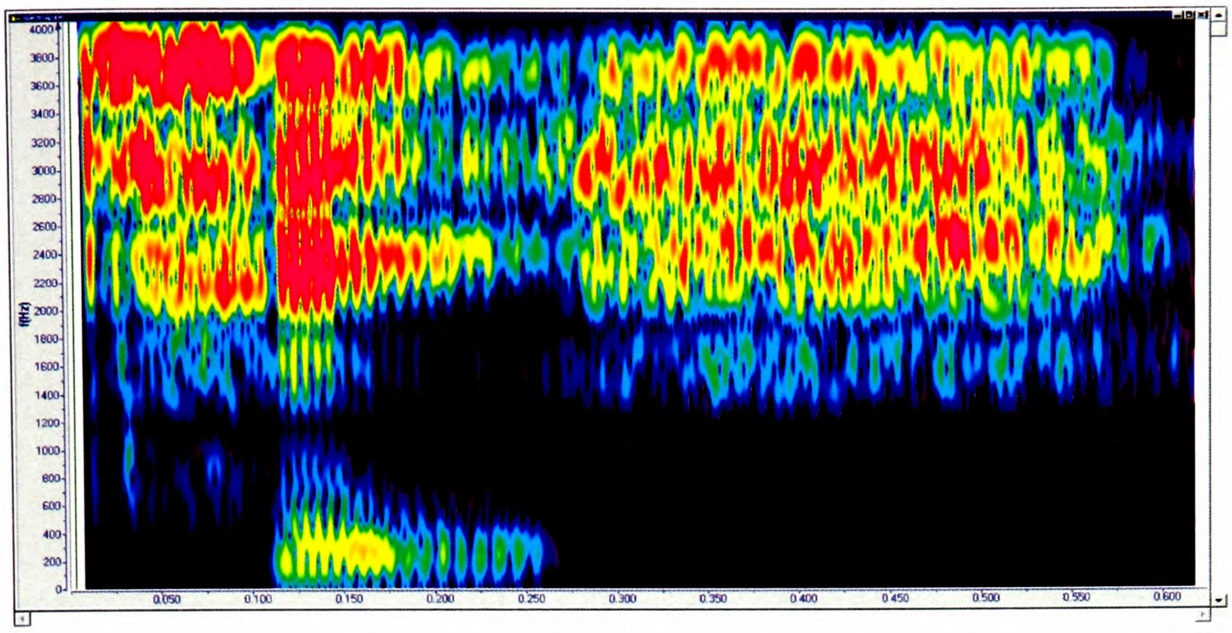


Figure 31a. Untransformed version of the non-word /teash/ produced by a male speaker.

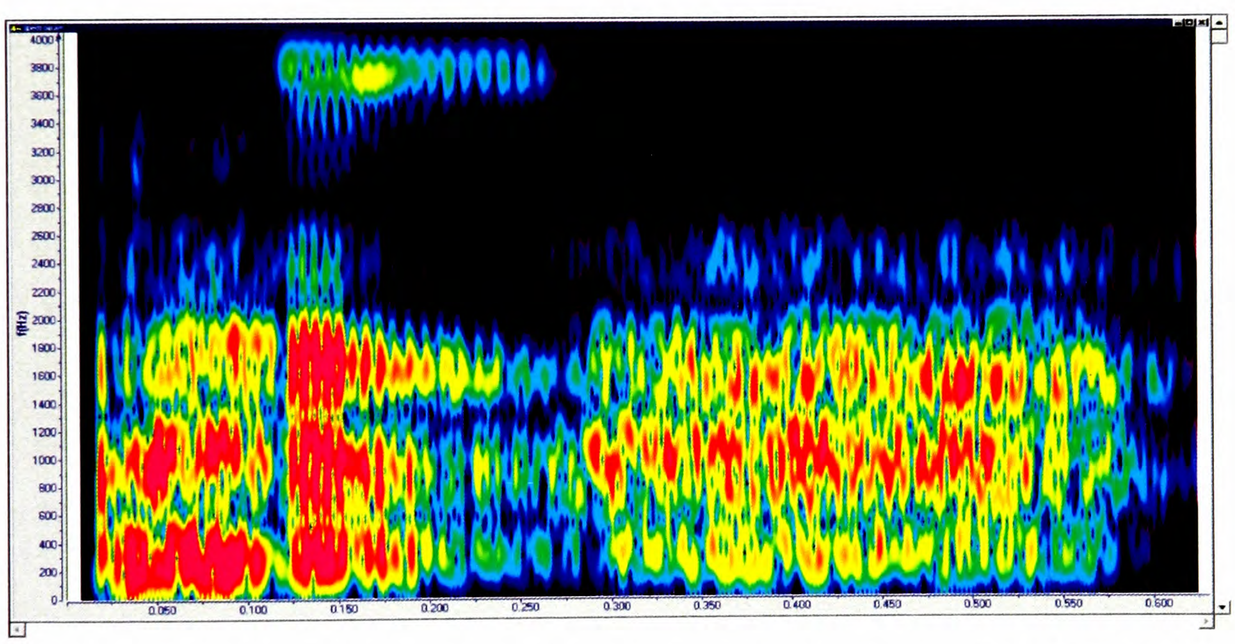


Figure 31b. Spectrally rotated version of the non-word /teash/ produced by a male speaker.

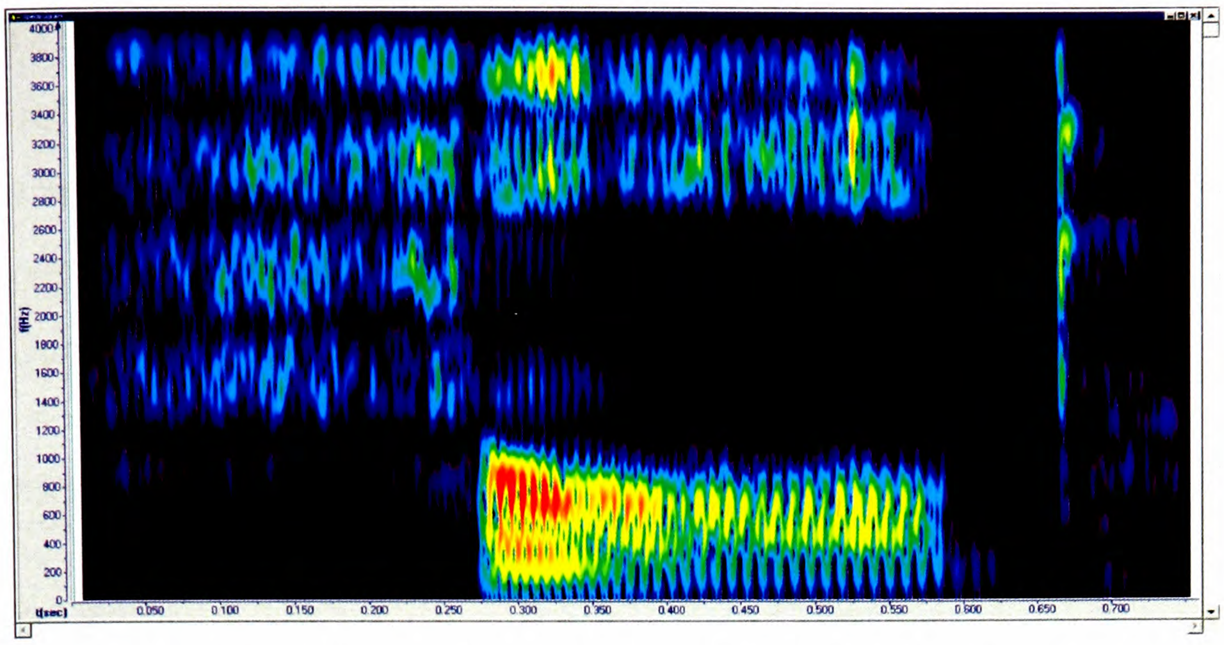


Figure 31c. Untransformed version of the non-word /forb/ produced by a male speaker.

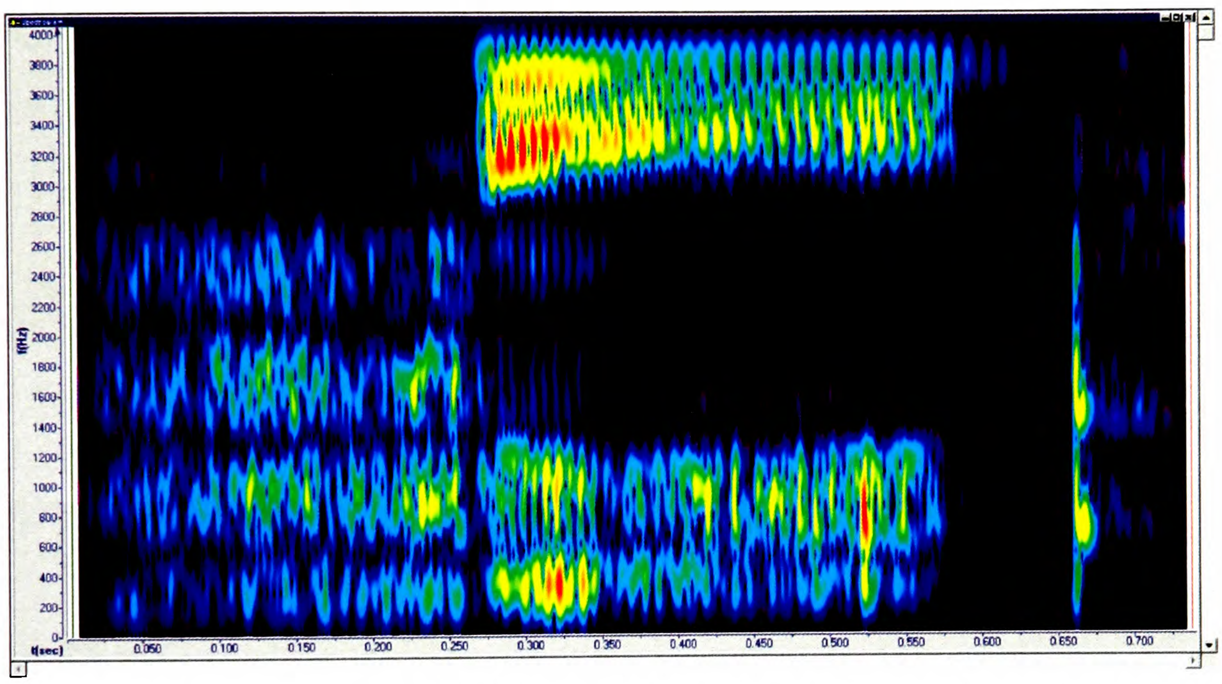


Figure 31d. Transformed version of the non-word /forb/ produced by a male speaker.

Scott, Blank, Rosen and Wise (2000) reported that participants heard speech sounds transformed by spectral rotation as an "alien language"; however, descriptions of participants in the present experiment bore no reference to the speech-likeness of the non-words or an alien language. This may be due in part to the fact that the non-words occurred in isolation, whilst previous experiments have used spoken sentences (Scott, Blank, Rosen and Wise, 2000). The remaining 16 non-words were either heard as speech-like or partially as speech, for example, in some instances vowels were heard. Also, some non-words were at times described as vocalised sounds. One participant described the non-word /hoc/ as a sound produced by a bullfrog. A look at the spectrograms in figure 32a and 32b reveals that for the non-word /hoc/ the formant patterns are strong and the range of acoustic energy across the frequency domain is more distributed and so even when the spectrum is rotated (inverted) the phonemes can still be identified.

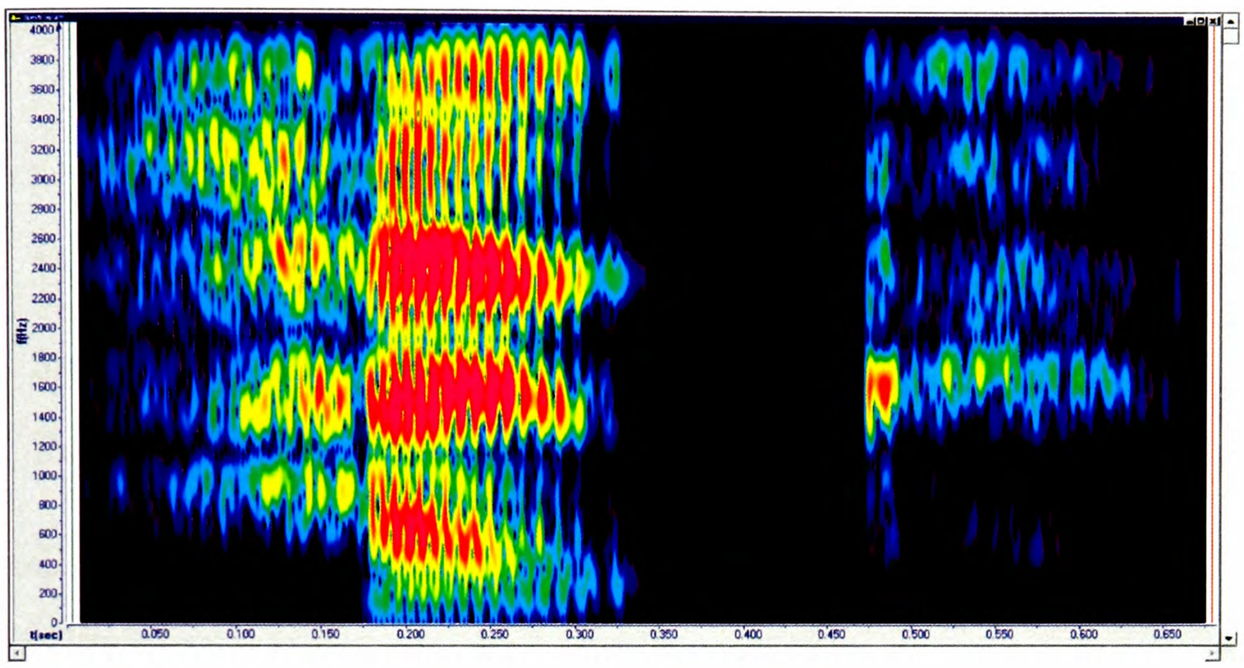


Figure 32a. Untransformed version of the non-word /hoc/.

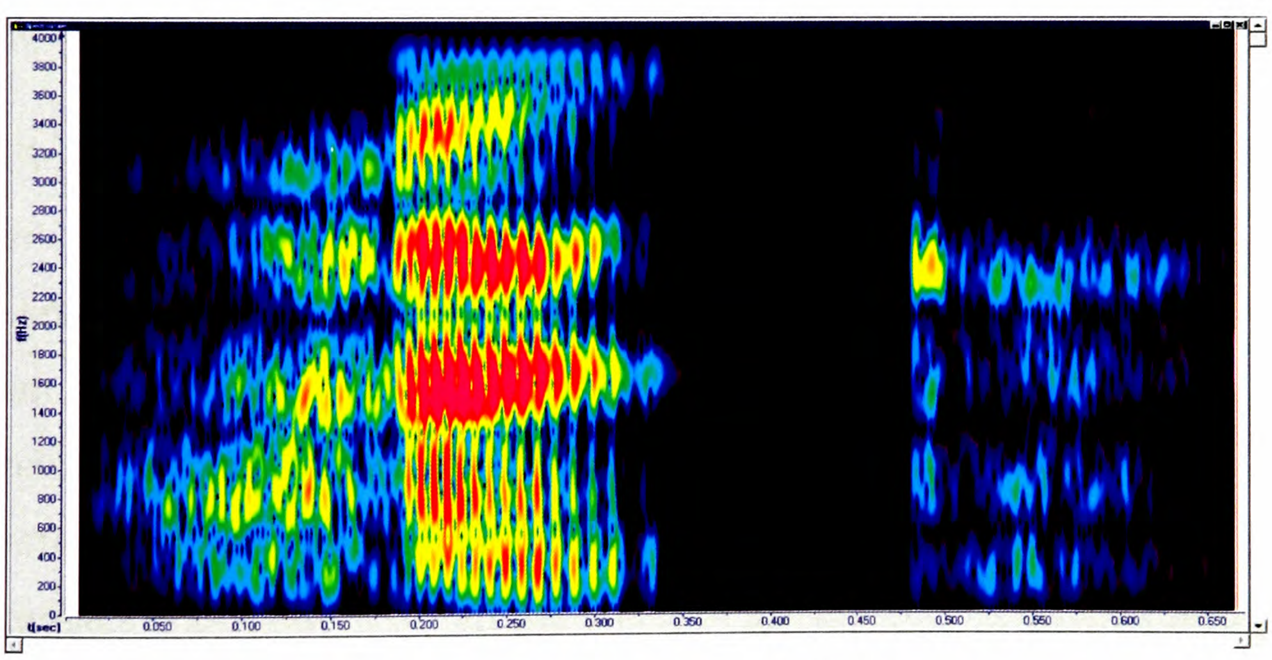


Figure 32b. Spectrally rotated version of the non-word /hoc/.

11.3.3 Design and procedure

The design and procedure was the same as that used for the previous pilot experiment (chapter 10) and as detailed in the general procedural outline of chapter 3 (p92-94).

11.4 RESULTS

Experimental condition	*Mean Errors*	*SD*
Speech	37.93	22.12
Rotated speech	33.13	23.03
Silence	28.83	22.55

Table 19. Descriptive statistics for the 3 experimental conditions; mean number of serial recall errors per condition. N = 30 (rotated speech = spectrally rotated speech).

The descriptive statistics in table 19 indicate a replication of the findings of the previous pilot experiment (chapter 10). More recall errors were made under speech than silence, thus replicating the robust irrelevant sound effect (ISE). A small numerical difference between the mean number of serial recall errors produced under speech and spectrally rotated speech indicates speech produced slightly more interference than did spectrally rotated speech. A small numerical difference is evident between the levels of interference produced under spectrally rotated speech relative to that observed under silence. The data are summarised in figure 33.

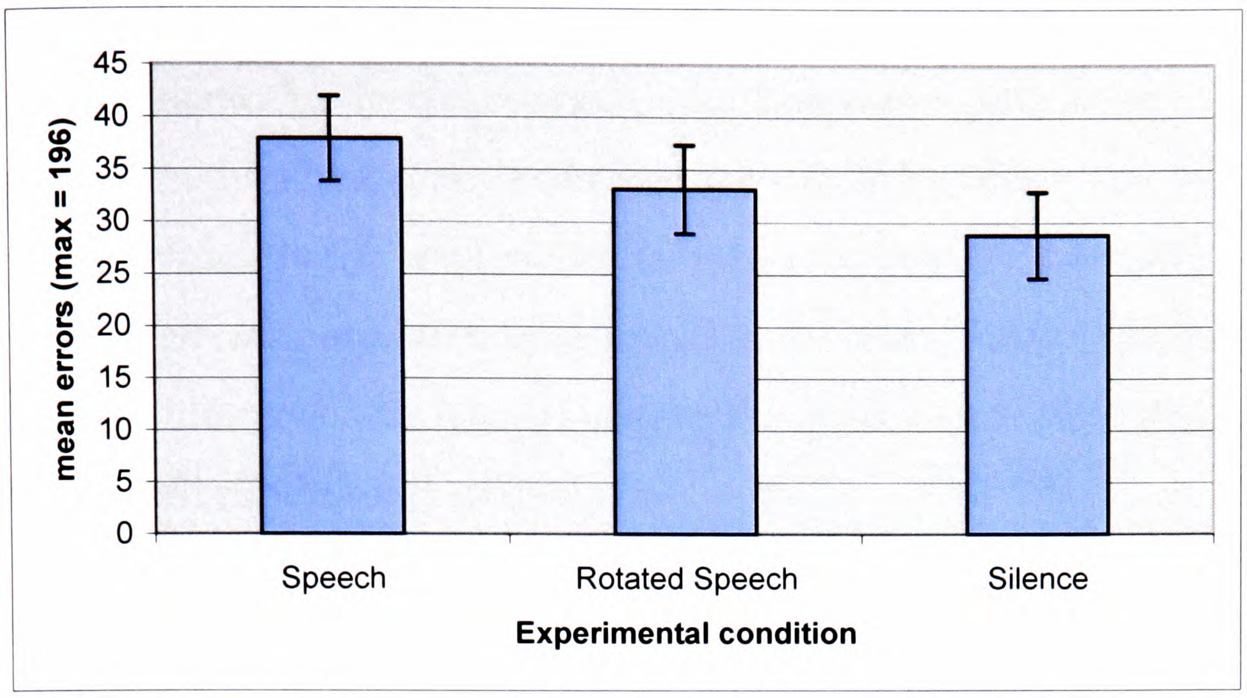


Figure 33. Mean number of recall errors for the three experimental conditions. Error bars represent standard error above and below the mean (rotated speech = spectrally rotated speech).

	Silence	*Speech*
Rotated speech	Non-sig $p \leq 0.574$	Non-sig $p \leq 0.394$
Speech	✓ $p < 0.05$	XX

Table 20. Bonferroni corrected pairwise comparisons for the three experimental conditions (rotated speech = spectrally rotated speech).

A repeated measures ANOVA on three levels (speech, spectrally rotated speech and silence) was performed on the mean number of digit recall errors for each experimental condition. Irrelevant sound disrupted memory relative to a silent control [F (2, 58) = 4.190, MSE = 621.700, $p <$ 0.05] (appendix 35). Figure 34 summarised the mean number of errors collapsed across serial position. Table 20 displays the pairwise comparisons with Bonferroni correction for the three experimental conditions. Pairwise comparisons revealed the main effect of irrelevant sound stemmed only from the effect of clear speech (appendix 35). Serial recall performance in the clear speech condition was significantly different from the silent control condition ($p < 0.05$) but there was no reliable difference in the level of interference produced by spectrally rotated speech compared to that observed in the silent control ($p \leq 0.574$). No reliable difference was found between the speech and spectrally rotated speech condition ($p \leq 0.394$).

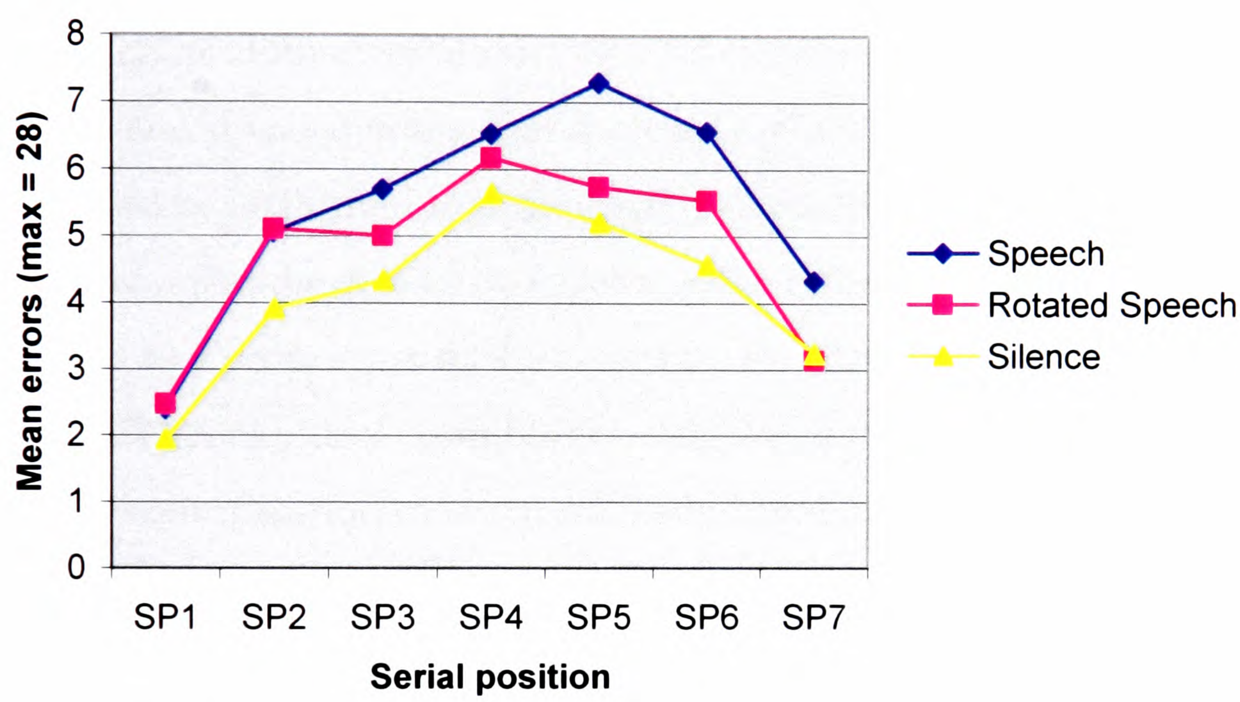


Figure 34. Overall number of serial recall errors for the three experimental conditions collapsed across serial position (rotated speech = spectrally rotated speech).

The data depicted in figures 33 and 34 shows a trend in the predicted direction. No difference between the two speech conditions and between spectrally rotated speech and silence was found as there were only small differences between the sample means of the auditory conditions, which are insufficient for a reliable difference in their disruption of immediate serial recall to be found. The majority of participants performed well at the serial recall task and only a few made a large number of errors in all three sound conditions. In particular, participant scores in one or both speech conditions differ only slightly from errors made in the silent control condition. If the difference between the mean scores under two auditory conditions is the same, then a greater number of participants making only a few errors in the speech conditions in comparison to only a few participants making a greater number of errors would lower the average error score. Thus, no reliable difference between the mean scores for speech and spectrally rotated speech may well be due to participants, in general making few serial recall errors.

There are three forms of variance in data, the variance attributable to the effect, random error and individual differences. The raw data was examined and it was evident that there was a lot of variability in the data which could be attributed to individual differences in memory ability. The data was standardised by expressing the difference between conditions as a proportion of the error rate for the silent control. Hence, data in the silent control condition was treated as the baseline. This was a suitable standardisation to factor out individual differences in memory ability. The number of errors made in the speech condition by each participant was standardised by calculating the difference between the number of errors in the speech condition and those made in silence for each participant. Then the difference in errors was divided by the number of errors made in silence and this figure was multiplied by 100 (see formula in appendix 36). Scrutiny of the data set for the silent control found that participant 21 had scored zero errors. Therefore 1 was added to the denominator (error score) for each participant so as to avoid dividing by zero. The number of errors observed in the spectrally rotated speech condition was standardised in the same way. Thus, the difference between the number of errors in the rotated condition and the silent condition was calculated for each participant. The difference in errors was then divided by the number of errors made in silence for that participant. Again, this figure was multiplied by 100. A paired-samples t-test (appendix 37) was applied to the standardised data to compare the mean number of errors in the speech condition and the mean number of errors in the spectrally rotated speech condition. A significant difference between speech and spectrally rotated speech was found [t (29) = 2.464, $p \leq 0.020$] which is consistent with the trend in the data evident in figures 33 (p230) and 34 (p231).

11.5 DISCUSSION

The present experiment provided a direct test of the CSH by contrasting speech and non-speech (spectrally rotated speech) stimuli matched for acoustic complexity. The acoustic characteristics of spectrally rotated speech have the same spectral-temporal structure as untransformed speech, but are not intelligible. The difference found between speech and spectrally rotated speech following standardisation of the data is inconsistent with the predictions of the CSH, which suggests the extent and nature of acoustic changes in the irrelevant stream modulates the size of the ISE (Jones et al., 1996). Changing-state speech is argued to exhibit greater acoustic changes in an unfolding stream than changing-state non-speech stimuli, such as sine-wave speech and simple tones. The complex acoustic change between successive items leads to the formation of stronger cues to their serial order which by direct entry into short–term memory (STM) conflict with the seriation of the TBR items entering verbal working memory by sub-vocal rehearsal (e.g. Jones and Tremblay, 2000).

The greater disruptive effect of speech relative to spectrally rotated speech can be explained with reference to the processing of the spectral detail of speech in the right hemisphere. Although the left hemisphere is specialised for most language functions, the analysis of the speech signal has been shown to be performed bilaterally in the superior temporal cortex (Hickok and Poeppel, 2000). Zatorre and Belin (2001) argue that the left auditory cortex is specialised for temporal processing, whereas the right auditory cortex is specialised for processing the spectral detail of sounds. In the literature, the terms spectral and pitch are often used interchangeably (Scott and Wise, 2004). Poeppel (2003) suggests that different temporal integration windows in the processing of the speech signal may explain the bilateral activation seen with intelligible speech. Processing transient acoustic cues such as those of consonants require a shorter temporal integration window and are processed by the left hemisphere. In contrast, spectral information, such as that provided by speech formants requires a longer temporal integration window. It follows that spectral detail and thus information pertaining to pitch variation is analysed by the right hemisphere. This is consistent with studies concerning the hemispheric processing of irrelevant sound, which have shown that irrelevant sound is processed predominately by the right hemisphere (Hadlington et al., 2004; 2006).

Spectrally rotated speech maintains the pitch variation of the untransformed speech signal (Beaman et al., 2007; Narain et al., 2003; Scott et al., 2000) as it maintains the acoustic complexity but not the intelligibility of speech. Beaman et al (2007) view the acoustic correlates of pitch variation in the speech signal as representing the 'changing states' in an irrelevant speech sequence and argue the right hemisphere analyses the acoustic features that make up the 'changing-states' of the irrelevant sounds. Since spectrally rotating speech only maintains pitch variation, neural processing observed in its presence as unattended sound will be associated with the processing of information pertaining to pitch variation within the signal.

Scott et al., (2004, submitted, cited in Beaman et al., 2007) contrasted spectrally rotated speech with signal-correlated-noise (SCN) to examine the neural activity related to ignoring the pitch dynamics of unattended speech and activity associated with processing the ignored lexical and semantic features of speech. Subtracting activity related to SCN from that produced by spectrally rotated speech lead to more activation in the right superior temporal gyrus (STG). This supports the left ear disadvantage (LED) found be Hadlington et al (2004; 2006). The larger ISE reported by Hadlington et al (2004; 2006) when speech and tones are presented to the left ear only may be due to the right hemisphere's analysis of pitch change information as opposed to processing the intelligibility of the speech signal. This perspective is consistent with the established finding that non-speech sounds, such as cello notes, pitch shifted simple tones and sine-wave speech, are sufficient to produce an ISE (Jones et al., 2000; Macken and Jones, 1993; Tremblay et al., 2000). However, disruption of serial recall is normally only observed in non-speech streams when successive sounds exhibit abrupt variations in pitch (Jones et al., 1992). Beaman et al (2007) argues the ISE observed with speech and non-speech to be behaviourally different from the lexical semantic effects seen with irrelevant speech as meaningless sounds are sufficient to produce the standard ISE (Buchner et al., 1996).

Although a LED is reported for both irrelevant speech and non-speech and they are found to disrupt serial recall in a similar way, speech was found to be more disruptive of serial recall than non-speech. Therefore, acoustic complexity pertaining to pitch variation cannot account for why speech is more disruptive of memory than spectrally rotated speech which preserves pitch variation, but not the intelligibility of the signal.

The absence of a reliable difference between spectrally rotated speech and silence however would not be predicted within the framework of the CSH, which argues for the importance of acoustic changes over time between successive items in an irrelevant stream. However, due to the observed low error rate it is difficult to infer why spectrally rotated speech did not differ from the silent control and conclusions would be more plausible with a replication of this finding as well as replicating the difference between speech and non-speech. The fact that only seven digits featured in each list can account for the reduction in the number of errors made, as this suggests the use of a 'grouping' memory strategy by participants. The seven digits per list were constructed from the digit set 1-7. Therefore, participants may have come to realise that by remembering the order of the first six out of the seven digits in each list would mean that they could work out the seventh digit, since the digits went from 1-7 in random order in each trial. Visual inspection of figure 34 (p236) shows a more pronounced recency effect than is observed in ISE data as the serial position curves are relatively symmetrical. It seems reasonable to assume that increasing the number of errors by using a more demanding series of digit lists will result in a reliable difference between speech and spectrally rotated speech, and a serial position curve which is more fitting to the standard primacy and recency effect observed in the ISE (e.g. Jones et al., 1990).

11.6 EXPERIMENT 7: AIMS AND OBJECTIVES

It would seem that the use of only seven digits from the digit sequence 1-7 does produce the robust irrelevant sound effect as demonstrated in the previous experiments in this experimental series. However, due to the low error rate reported in Experiment 6 the differences between speech conditions and rotated speech could be numerically observed, but were statistically unreliable. Standardisation of the data, with the silent condition as the baseline lead to the two speech conditions differing reliably. The low error rate may have been due to participants adopting a grouping strategy, where participants remembered the first 6 digits, and from this, could identify the final seventh digit of lists constructed from random perturbations of the digit set 1-7. Experiment 7 aimed to increase the error rate by using 8 digits from the digit set 1-9. It investigated whether increasing the error rate will bring out reliably the difference between the effect of speech and rotated speech and whether or not a difference between spectrally rotated speech and the silent control would be observed.

11.7 METHODOLOGICAL CONSIDERATIONS

11.7.1 Participants

24 undergraduate psychology students volunteered to participate in the study. All reported normal hearing and normal or corrected to normal vision. All participants had English as their first language and were not paid for their time.

11.7.2 Visual and auditory stimuli

The to-be-remembered (TBR) digit lists consisted of 8 digits from the digit set 1-9 (appendix 29) as used in experiment 4 (chapter 7). The same non-words used in experiment 6 were presented concurrently with the digits in both auditory conditions (/teash/ (tiS) and /forb/ (f$b) for the spectrally rotated speech condition and /marv/ (m£v) and /curj/ (k3_) for the speech condition (see auditory stimuli section of experiment 6 on page 237 for explanation of why different non-words were used in the speech and spectrally rotated speech conditions.

11.7.3 Design and procedure

The design and procedure was the same as experiment 6, but this time participants undertook 27 serial recall trials per auditory condition.

11.8 RESULTS

Experimental condition	*Mean Errors*	*SD*
Speech	64.54	31.248
Rotated speech	48.58	28.264
Silence	41.88	27.603

Table 21. Descriptive statistics for 3 experimental conditions; mean number of serial recall errors per condition. N = 24 (rotated speech = spectrally rotated speech).

.

The data are summarised in figure 35. Inspection of the sample means displayed in table 21 shows that the error rate for both speech conditions has indeed been increased with the addition of an eighth digit in the TBR digit list.

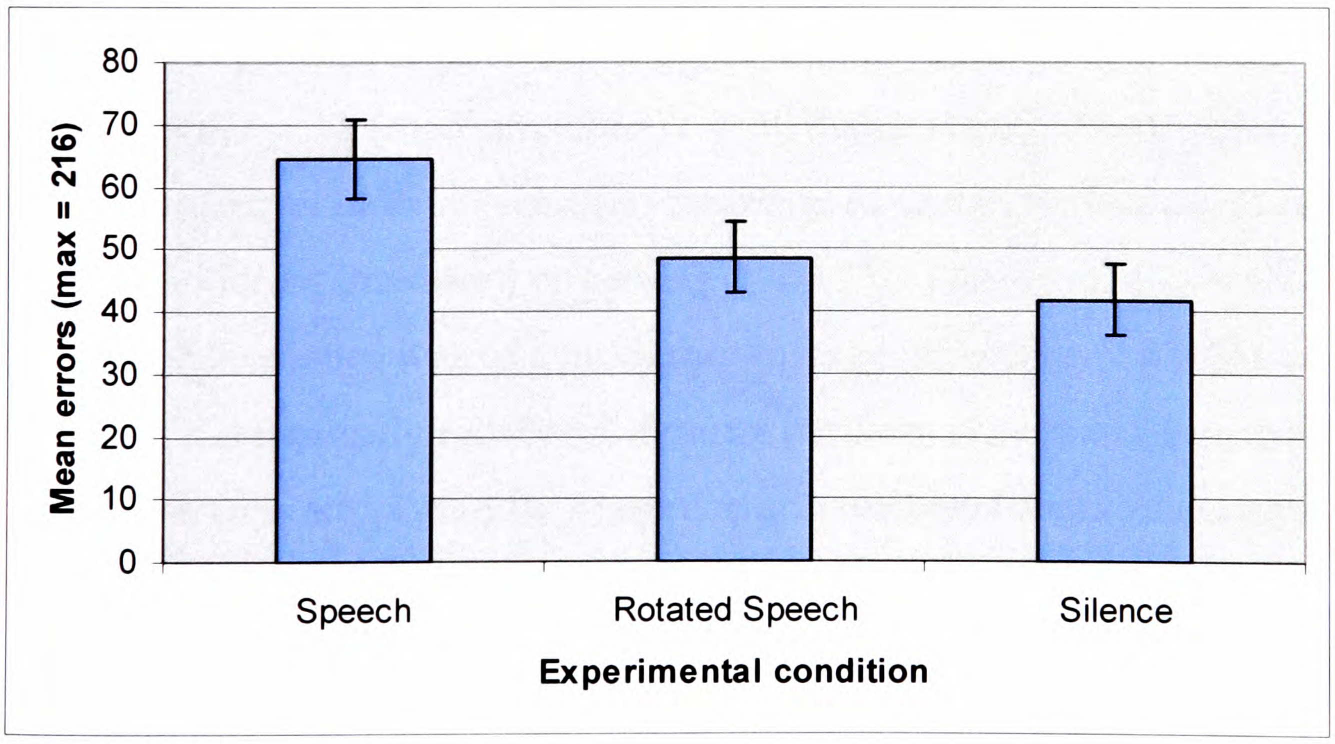


Figure 35. Mean number of recall errors for the three experimental conditions. Error bars represent standard error above and below the mean (rotated speech= spectrally rotated speech).

	Silence	*Speech*
Rotated speech	Non-sig $p \leq 0.507$	✓ $p < 0.01$
Speech	✓ $p < 0.001$	XX

Table 22. Bonferroni corrected pairwise comparisons for the three experimental conditions (rotated speech= spectrally rotated speech).

.

A one-factor repeated measures ANOVA found there to be a main effect of irrelevant sound [F (2, 46) = 11.754, MSE = 3253.792, $p < 0.001$] (appendix 38). Bonferroni corrected pairwise comparisons of the means as shown in table 22 revealed a reliable difference between speech and spectrally rotated speech ($p < 0.01$). Speech also differed significantly from silence ($p < 0.001$) (appendix 37). However, there was no difference between the number of errors produced by spectrally rotated speech and silence ($p \leq 0.507$). Contrasts were performed comparing the 3 possible combinations of conditions in order to calculate r as a measure of effect size (appendix 39). Cohen (1988) suggests the benchmarks of r = .10 (small effect), r =.30, (medium effect) r = .50 (large effect). The greater memory impairment of speech in comparison to spectrally rotated speech was of a moderate (medium) effect size (r = 0.33). The effect of speech in comparison to silence was of a moderate to large effect size (r = 0.48), and the lack of a statistically reliable difference between speech and spectrally rotated speech is supported by r not reaching the benchmark of a small effect (r = 0.08). The data collapsed across serial position is evident in figure 36.

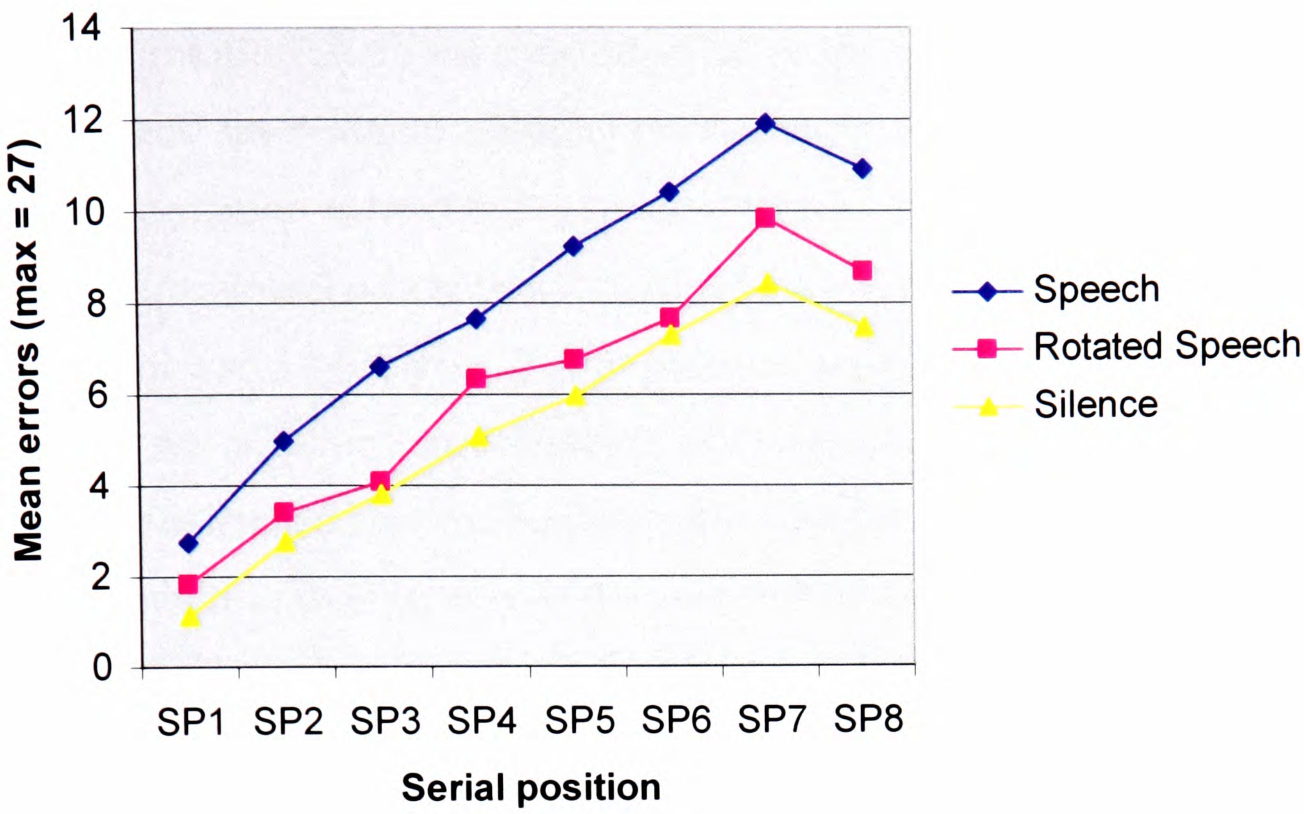


Figure 36. Number of recall errors collapsed across serial position for the three experimental conditions (rotated speech = spectrally rotated speech).

11.9 DISCUSSION

That speech is more disruptive than spectrally rotated speech is inconsistent with the CSH as acoustic complexity was controlled for between the two speech conditions. This indicates an important role for the intelligibility of speech in mediating the distinction between speech and non-speech in their disruption of serial recall. Spectrally rotated speech sounds were not heard as speech and cannot be produced by a talker (Blesser, 1972). In a pilot listening session, participants reported that the spectrally rotated non-words used in the spectrally rotated condition sounded like computer generated complex noise, or noise from a computer game. The observation of a difference in serial recall disruption between spectrally rotated speech and speech indicates the distribution of critical speech pattern information may explain why intelligible speech is more disruptive than non-speech matched for acoustic complexity. Energy in frequency regions not necessarily present in speech is a consequence of frequency (spectral) rotation. Thus, frequency rotation alters the spectral detail of speech, in particular the temporal and spectral patterning of the formants (c.f. Lachs and Pisoni, 2004). Information carried in the spectral structure of formants is important for vowel perception (Strange, Jenkins and Johnson, 1983). Experiment 3 and 4 (chapter 7) demonstrated whispered speech, which has no $f0$, but which conveys remnants of speech formants is as disruptive as voiced speech. Sequences of distinct irrelevant CVC speech tokens exhibiting change only in the vowels, which is carried by the formant structure, have been found to be more disruptive of serial recall than sequences conveying change between consonants of successive items only (Hughes, et al., 2005; experiment 2, chapter 6). Spectral rotation destroys the original relationship between formants, but maintains the range of frequencies across the frequency domain (Blesser, 1972) as is evident in the spectrograms in figures 31a-31d (pp226 and 227) for the speech sounds spectrally rotated. As frequency rotation destroys the original relationship between formants, vowel intelligibility is destroyed.

In the speech recognition paradigm research has focussed on the recognition of speech pattern information under conditions of both spectral and temporal distortion and attenuation. In particular, this work has been of considerable use to the hearing impaired and cochlear implant users who experience reduced spectral detail in speech. Research has looked at the consequences of degrading the speech signal, in terms of consonant and vowel identification. Drullman *et al.* (1994) showed that temporal smearing had a more adverse effect on consonant rather than vowel identification. Also, Shannon et al (1995) systematically attenuated spectral detail in a speech signal to one, two, three or four bands of signal-correlated-noise (SCN). Each band of noise was modulated by the envelope of the original spectral band in the speech signal. When speech was reduced to one band of modulated noise, preserving only the broadband temporal envelope, whilst removing all spectral detail, recognition of consonants in contrast to vowels was relatively good. Therefore preserving the temporal envelopes within the spectrum of the speech signal is insufficient for vowel recognition.

The findings demonstrate that when intelligibility is manipulated but the acoustic complexity of speech and non-speech sounds is matched; speech is significantly more disruptive than sounds perceived as non-speech. This finding, however, is problematic for the assumptions of the changing-state-hypothesis (CSH). According to the CSH, disruption by irrelevant sound is a function of the amount of acoustic change that an acoustic stimulus demonstrates, and that as the number and extent of acoustic change is increased, up until a point, the level of disruption will increase (e.g. Jones et al., 1999a). Thus, the ISE is explained purely in terms of acoustic 'change', but what characteristic(s) need to change between distinct items has not been explained. Previous research using sine-wave speech as irrelevant sound has shown that regardless of whether subjects are trained to perceive or are left unaware of the stimuli's speech status, there is no reliable difference in level of disruption produced. But, crucially, when compared to both sine-wave speech conditions, natural speech is found to be more disruptive. This was explained with reference to the more complex acoustic variation within the speech signal. But, when acoustics are equated, sounds heard as speech are still more disruptive of serial recall, as the present experiment and experiment 6 shows. It seems that the spectral detail pertaining to the formants as they evolve over time as well as their relative spacing is important in rendering natural speech more disruptive of serial recall.

The present findings can be viewed in light of other research emphasising the importance of formant structure. As discussed, formant frequencies, in particular those of the first two or three formants, are important in the perception of vowels (Strange, Jenkins and Johnson, 1983). Studies which have suppressed formant movement have shown a decrease in identification accuracies (Assmann and Katz, 2000). Although the formants are still present within the spectrum of the speech signal, spectral rotation results in them occurring in the wrong frequency channels. As the patterning of formant resonances are the product of vocal tract shape, when formant frequencies are inverted the sounds produced are ones which the human vocal tract could not produce and they are no longer heard as speech-like. Experiment 4 (chapter 7) temporally reversed the spectral detail (fine structure) of whispered speech resulting in the samples exhibiting slow onsets and rapid offsets. However, the frequency information, though reversed, was still in the correct frequency channels and the fine structure reversed (FSR) whispered speech was perceived as a vocalised sound, unlike spectrally rotated speech. This may account for why, in contrast to FSR speech, spectrally rotated speech affected memory to a lesser degree than did speech.

Speech which is highly reverberated produces a level of disruption which is statistically indistinct from that produced in silence (Beaman and Holt, 2007). This has been accounted for by the effect of reverberation on the speech signal. Reverberation acts to smear the temporal patterning (profile) of the speech signal and as a consequence reduced the extent of acoustic variation in the irrelevant sound stream (Beaman and Holt, 2007). Smoothing of the profile of the sound acts to prevent sufficient segmentation of the signals constitute parts, in this case words from connected speech. In addition, the peaks and troughs of the signal are suppressed which acts to lessen the number and extent of acoustic variation (Beaman and Holt, 2007). However, highly reverberated speech is heard as speech and yet when presented as unattended sound the irrelevant sound effect (ISE) is removed. This supports the notion of the CSH that it is the amount of acoustic change between elements within a stream and not their nature or intelligibility that is important in determining the ISE.

Although unlike highly reverberated speech, spectrally rotated speech preserves the spectral and temporal patterning of speech, spectrally rotating speech removes the ISE observed with speech as does highly reverberating speech. Spectral-temporal modulation is important in speech pattern recognition (Shannon et al., 1995). Reverberation not only acts to smooth the temporal patterning of speech but also smears and thus corrupts its harmonic structure (Roman and Wang, 2005; Wu and Wang, 2006). Harmonicity of the signal does not seem to be important in the ISE since whispered speech produces an ISE which is equivalent to that found with voiced speech (experiment 3, chapter 7). However, as reverberation smears and corrupts the harmonic structure of speech, it will also corrupt the structure of the formants.

Spectral rotation of the speech signal results in formants occurring in frequency regions not naturally found in speech and destroys the relative spacing between the formants. Maintenance of the relationships among formants as they evolve over time is reported to be important for word recognition (Lachs and Pisoni, 2004). In addition, verbal serial recall has been found to be heavily disrupted by the to-be-ignored information of changing vowels carried in the formant structure (experiment 2, chapter 6; Hughes et al., 2005). Experiment 5 (chapter 9) demonstrated that intelligibility of speech in terms of being able to accurately repeat back and identify speech does not explain why speech is more disruptive of serial recall than non-speech stimuli. Participants were unable to repeat back or identify phonemes of whispered non-words whose fine structure was reversed. Rather, participants heard these sounds as strange sounds produced by a voice. Fine structure reversed (FSR) whispered non-words were found to disrupt serial recall at a level that was indistinct from whispered non-words. Highly reverberated speech is difficult to comprehend, one can think of the difficulty of trying to understand announcements at train stations. Beaman and Holt (2007) state that passive listening of reverberated stimuli indicated intelligibility to be good by participants; however no formal testing was carried out to screen the intelligibility of the speech in terms of the identifiability of the component words of the highly reverberated stream of speech. As high reverberation would smear the formant structure it can be argued that participants would not have been able to accurately recognise and repeat the highly reverberated speech back. However, participants heard highly reverberated speech as speech. As reverberation and spectral rotation distort the spectral structure of formants, it may be that although highly reverberated speech was heard as speech, because its formant structure was distorted this may explain why an ISE was not observed in its presence. This provides further evidence that formant structure and the relative spectral distribution and spacing of formants is crucial to speech maintaining its higher power to interfere with serial recall relative to non-speech sounds.

The lack of a difference between spectrally rotated speech and silence is unexpected, given that although the rotated stimuli are not heard as speech, speech and rotated speech convey spectral complexity to an equivalent degree, as temporal and spectral variation in the signal is left relatively unaltered (Scott et al., 2000; Narain et al., 2003). Serial recall performance in sine-wave speech (Tremblay et al., 2000) and degraded speech (e.g. experiment 1b, chapter 5), which have reduced acoustic complexity compared to natural speech, have both been found to differ from a silent control condition. In addition, bursts of broadband noise that change in band-pass frequency produce an ISE (Tremblay et al., 2001), but spectrally rotated speech does not. The bursts of broadband noise that changed in centre frequency used were described by Tremblay et al (2001) as sounding more like noise, exhibiting a low level of tonality.

As spectral rotation does not produce an ISE it seems this transformation of the non-words removed the tonal quality that band-pass noise exhibits.

Examination of the spectrograms in Figures 31c and 31d (p 232) show that rotating the non-word /forb/around a centre frequency of 2 kHz introduced energy at high frequencies into the signal not produced by the vocal tract. For the non-word /teash/ spectral rotation introduced high concentrations of low frequency energy as is evident when comparing figures 31a and 31b (p231). This distortion of the spectral distribution of energy would have distorted the tonality of these sounds, making them sound 'brighter' (Blesser, 1972). This suggests changes in tonal quality are needed to produce the standard ISE, as is observed with tones shifted in pitch (Hadlington et al., 2004; Jones and Macken, 1993). The idea that tonal quality is of importance is consistent with reports that irrelevant speech and non-speech are predominantly processed by the right hemisphere, since information presented to the left ear produces the largest ISE. The LED reported with sequences of changing-state speech and pitch-shifted tones is evidence that it is pitch variation that provides the changing-state information necessary for the ISE, as it is the right hemisphere that processes the pitch dynamics of speech (Hadlington, et al., 2004; 2006).

When acoustic complexity is controlled between speech and non-intelligible speech stimuli by contrasting speech and spectrally rotated speech a peak of activation is observed in the right superior temporal gyrus (STG) (Beaman et al., 2007). This is indicative of the considerable processing of the pitch dynamics of sound by the right hemisphere. In light of this increased right STG activation in the presence of to-be-ignored spectrally rotated speech a difference between spectrally rotated speech and silence would have been expected. Pitch and tonal quality are related. It may be that the tonal quality of the non-words used in the rotated condition of the present experiment was affected more by frequency rotation than the rotated samples used by Beaman et al (2007). That is, higher concentrations of low frequency energy may have been introduced at high frequencies. Therefore, distortion of the tonal quality of the spectrally rotated non-words may account for the removal of the ISE in its presence.

11.10 SUMMARY

Experiment 7 identifies an important role for the intelligibility of speech sounds. Spectrally rotated non-words preserve the temporal and spectral patterning of speech, but are not heard as speech and cannot be produced by the vocal tract (Blesser, 1972; Scott et al., 2000). Spectral rotation alters the information relating to speech articulation afforded by the temporal and spectral patterning of the formants (c.f. Lachs and Pisoni, 2004). The finding that speech is more disruptive of serial recall than acoustically matched spectrally rotated speech is evidence against the CSH, as both irrelevant sound conditions conveyed the same amount of acoustic variation between successive items. This emphasises the importance of critical speech pattern information (c.f. Blesser, 1972) in particular the distribution of spectral envelope cues. Vowels have been seen to be the dominant source of disruption in the ISE paradigm (experiment 2, chapter 6; Hughes et al., 2005) and it is the distortion of the distribution of spectral detail, (formants) provided by the vowels which is reported to be more detrimental to vowel identification (Lachs and Pisoni, 2004).

CHAPTER 12

12 SUMMARY OF EXPERIMENTS

12.1 EXPERIMENT 1

A perceptual identification task was used to screen the intelligibility of degraded non-words for pilot measurements for experiment 1a (chapter 4). It aimed to find a level of degradation that would result in a reliable difference between the serial recall disruption produced by speech and degraded speech, but where memory performance in degraded speech differed significantly from that in a silent control. A better range of intelligibility for non-words degraded at a signal-to-noise-ratio (SNR) of 0.7 as opposed to 0.65 SNR was established. Seven low intelligible non-words degraded at 0.7 SNR were isolated for the degraded speech condition in pilot B (for experiment 1b) (chapter 4) which compared the effect of clear speech and degraded speech on serial recall performance. The clear speech sequences differed from the degraded speech sequences with regards to both phonetic content and auditory complexity, both of which were reduced in degraded speech. Clear speech interfered with serial recall more than did degraded speech. However, degraded speech did not differ from the silent control. The perceptual identification task showed that initial and final consonants of the degraded non-words were misperceived more than the vowels. Vowels as opposed to consonants have been shown to provide important changing-state information in an irrelevant stream (Hughes et al., 2005). Although the consonants of the degraded non-words were misperceived more than the vowels, it may be that important changing-state information for the ISE was removed by degrading the non-words at a SNR of 0.7. As the presentation order of conditions was not fully counterbalanced and the sounds were presented free-field using a single speaker it was difficult to draw any reliable conclusions from this pilot data.

Experiment 1a (chapter 5) screened the intelligibility of a set of non-words in a perceptual identification task as in the pilot. The non-words were degraded at a SNR of 0.7, since this SNR produced a better range of intelligibility for the non-words in the pilot. As before, seven non-words were isolated for the degraded speech condition of experiment 1b (chapter 5). Clear speech sounds were found to impair serial recall performance significantly more than degraded speech sounds and memory performance in the degraded speech condition differed reliably from performance in the silent control condition. The perceptual identification task of experiment 1a revealed that for the seven low intelligible non-words forming the irrelevant degraded speech sounds, the initial consonants were misperceived more than the vowels, but there was no reliable difference between the numbers of vowels and final consonants misperceived. This indicated that although preserved relative to the initial consonants, the vowels were damaged by degradation. As these have been identified as providing critical changing-state information within an irrelevant sequence of spoken utterances (Hughes et al., 2005), the fact that they were damaged explains the reduction in the size of the ISE in their presence.

12.2 EXPERIMENT 2

The components of the non-words that changed-in-state within an irrelevant sequence was manipulated in experiment 2 (chapter 6), in order to generate vowel-only-changing (V-O-C) and consonant-only-changing (C-O-C) sequences. These irrelevant sequences were phonologically degraded at a SNR of 0.7 (30% noise) and 0.5 (50% noise). The effect of the clear and degraded versions of V-O-C and C-O-C sequences on serial recall was examined. Although no test of the intelligibility of the non-word stimuli was performed, it can be assumed that their intelligibility was reduced. This assumption can be inferred on the premise that phonological degradation reduced the intelligibility of non-word stimuli in Experiment 1a.

A linear relationship has been observed between serial recall performance and the degradation of an irrelevant sequence. That is, as the degradation of speech stimuli increases, serial recall performance decreases (Jones et al., 2000). However, the linearity of this relationship is called into question when the components changing in the speech sequence is manipulated. Clear V-O-C sequences were more disruptive of serial recall than C-O-C sequences. Serial recall interference was reduced in the presence of V-O-C sequences degraded at 0.7 SNR, to a level observed with C-O-C sequences, both clear and degraded. However, the difference in serial recall disruption by V-O-C sequences that were clear or degraded at 0.7 SNR was not significant. Also, there was no reliable difference between serial recall impairment by V-O-C sequences degraded at an SNR of 0.7 and 0.5. In contrast, degradation had no effect on serial recall interference produced by C-O-C sequences. When the number of serial recall errors in both degraded V-O-C sequence conditions was pooled, a reliable difference was found between the effect of clear and degraded V-O-C sequences on serial recall. However, no reliable difference was found between the effect of clear and degraded C-O-C sequences. Therefore, in the presence of degraded V-O-C sequences, serial recall impairment was reduced leading to an improvement in serial recall performance. In contrast clear C-O-C sequences are less disruptive of memory relative to clear V-O-C sequences and the degradation of C-O-C sequences has no effect on serial recall performance.

12.3 EXPERIMENT 3

Comparing the effect of whispered stimuli and voiced stimuli in experiment 3 (chapter 7) provided a more objective test of the importance of the intelligibility of speech sounds. Voiced and whispered speech sounds were matched for intelligibility, but not acoustic complexity. The observed equivalent effect of both conditions on serial recall performance can be explained by the fact that both speech conditions were matched for intelligibility. Experiment 3 shows that not all acoustic information inherent within the signal needs to change between successive items. The absence of fundamental frequency ($f0$) in the whispered stimuli, which leads to it exhibiting no harmonic structure, does not render whispers less disruptive of serial recall relative to the same stimuli when voiced. It follows that the $f0$ of a speaker producing the sounds occurring in the irrelevant auditory stream is not the attribute on which change from item-to-item needs to occur to make speech more disruptive than non-speech sounds. Rather, it is the presence of formant structure common to speech sounds produced by the same voice over time which carries the disruptive acoustic changing information from item-to-item. The presence of formant structure can account for why the whispered speech stimuli were as intelligible as their voiced counterparts. Information within the structure of formants has been found to be important for vowel perception (Strange, Jenkins and Johnson, 1983). Formants are provided by the vowel portion of the CVC non-words and it is these as opposed to consonants which are more disruptive of memory when changing in an irrelevant sequence (experiment 2, chapter 6; Hughes et al., 2005).

12.4 EXPERIMENT 4

Experiment 4 (chapter 7) set out to test the notion that it is the strength of the acoustic links between irrelevant auditory items which preserve their temporal order that determine the magnitude of the ISE. These acoustic links between irrelevant sounds are argued to afford the obligatory maintenance of their serial order which acts to conflict with the seriation of TBR items (Jones and Tremblay, 2000). Experiment 4 contrasted the effect of irrelevant sequences made of only voiced speech sounds with sequences within which voiced and whispered speech sounds were alternated. Alternating between voiced and whispered speech sounds would have weakened the acoustic links as change across $f0$ was only conveyed by every other irrelevant sound and was not common to all the sounds in the sequence. In contrast, the formant structure of the sounds was common to all the sounds. No reliable difference was observed between both speech conditions and both conditions differed reliably from the silent control condition.

This finding is inconsistent with the predictions of the object-oriented episodic record (O-OER) model's changing-state-hypothesis (CSH). First, alternating between voiced and whispered sounds within an irrelevant auditory sequence would have increased the amount of acoustic changeability in the sequence; however this did not serve to increase the magnitude of the ISE. The CSH argues that the addition of change within a sequence will act to increase the size of the irrelevant sound effect (ISE) as long as this does not lead to the sequence segregating into separate steady-state streams of identical items. As the sequence was perceived as a coherent changing-state stream it seems memory performance under voiced speech had reached ceiling. Therefore the addition of more change by alternating between voiced and whispered speech did nothing to increase the level of serial recall interference.

Second, a reduction in the ISE was not found in the presence of alternating voiced and whispered speech which would be argued to weaken the acoustic links between the irrelevant sounds. This is inconsistent with the notion that $f0$ and formant structure need to be common to the sounds in a changing-state irrelevant sequence (Hughes et al., 2005). Rather, formant structure alone, which was common to both voiced and whispered speech sounds produced by the same speaker, was a feature of the irrelevant sound on which change occurred from-item-to-item. Formants in the whispered speech items were observed as a shadow of those present in the voiced speech items due to the absence of harmonicity and periodicity in the signal. However, due to the presence of energy at the formant frequencies, formants were still present in whispered speech.

12.5 EXPERIMENT 5

Experiment 5 (chapter 9) examined whether preserving the complex acoustic structure of speech whilst damaging its intelligibility would lessen the impairment of serial recall by irrelevant speech. Experiment 5 compared the effect of whispers and fine structure reversed (FSR) whispers on serial recall performance. This served to test the effect of damaging the intelligibility of speech sounds whilst maintaining their acoustic complexity. The experiment examined whether reversing the fine structure of whispers whilst maintaining their original amplitude envelopes would attenuate their disruptive effect on serial recall in comparison to whispers played forwards. The FSR whispers provided a signal with the same acoustic complexity and acoustic characteristics along with the same long-term average spectrum as the original whispers. However, the temporal structure of the fine structure in FSR whispers is distorted and so they cannot be articulated and provide little phonetic detail. Listeners did not hear the two FSR whispers as words or non-words and were unable to repeat back or understand what they heard. However, listeners did report hearing the FSR whispered stimuli as sounds produced by a voice. Despite the fact that non-words were not construed from the signal of FSR whispers, whispers and FSR whispers disrupted serial recall to the same extent. This shows that as long as distorted speech is heard as a vocalised sound, even if the non-words can no longer be perceived, it will interfere with serial recall of TBR items in the same way as clear (un-distorted) speech.

12.6 EXPERIMENT 6

Pilot measurements for experiment 6 (chapter 10) examined the disruptive effect of spectrally rotated speech sounds in contrast to speech sounds. The spectrally rotated sounds conveyed the acoustic complexity of speech, but were not heard as speech. Rather, these sounds were heard as complex noise. Spectrally rotated speech distorts the distribution of spectral cues which have been shown to be important for vowel recognition (Shannon et al., 1998). Speech and spectrally rotated speech disrupted memory to an equivalent degree, which was evidence that when acoustic complexity between speech and non-speech is controlled, the intelligibility of sounds as vocalised sounds is not the factor which renders speech more disruptive of serial recall than non-speech sounds. However, a trend in the predicted direction was evident in the data. Inspection of the spectrograms of the untransformed speech sounds showed energy at low frequency regions. However, energy across the frequency domain was distributed so that when spectrally rotated there was still energy at low frequency regions in the spectrally rotated speech. It follows that although these were not heard as speech, the distribution of spectral information in the frequency domain which provides critical speech pattern information may not have been distorted enough for the perceptual system to treat the spectrally rotated sounds as a non-speech pattern.

Experiment 6 (chapter 11) itself used spectrally rotated speech sounds that conveyed energy that was less distributed across the frequency domain in comparison to those used in the pilot. Two non-words heard as non-speech sounds were isolated for experiment 6 during a pilot listening session. Serial recall performance in the speech condition differed reliably from performance in silence. However, there was no difference found between the effect of speech and spectrally rotated speech on serial recall performance and there was no reliable difference between spectrally rotated speech and the silent control condition. It was evident upon examination of the data that there was large variability in the data attributable to individual differences in memory performance. The majority of participants performed well at the serial recall task. Only a few made a large number of errors in the speech conditions relative to the silent control. In order to remove this variability the data was standardised by calculating the difference between the number of errors in the speech and spectrally rotated speech conditions as a proportion of the silent control. After standardising the data a reliable difference was found between the disruptive effect of speech and spectrally rotated speech.

12.7 EXPERIMENT 7

In general, participants had made few errors in both speech conditions relative to the silent control in experiment 6. Experiment 7 (chapter 11) increased the error rate by using 8 digit lists constructed from the digit set 1-9 as opposed to 7 digits from the digit set 1-7. It is plausible to assume that participants may have adopted the strategy of only remembering the first 6 digits. As the digit lists were constructed from the set 1-7 in random order this would mean the digit that was not included in the first six would be the seventh digit. Experiment 7 found that spectrally rotated speech disrupted memory less than did speech. This observed difference between the disruptive effect of speech and spectrally rotated speech can be explained by the fact that spectral rotation of the speech signal destroyed the relationship between the formants. The relationship between the formants as they change over time has been shown to be important for word identification (Lachs and Pisoni, 2004). As the formants have been identified as providing the critical changing information, the fact that their distribution is distorted can explain why spectrally rotated speech was not as disruptive of serial recall as was speech. However, although spectrally rotated speech and speech were matched in acoustic complexity and thus conveyed the same amount of acoustic variation, serial recall performance in the spectrally rotated speech condition did not reliably differ from performance in the silent control condition.

CHAPTER 13

13 GENERAL DISCUSSION

This thesis shows that for speech to maintain its ability to disrupt serial recall at a high level it is not enough that it is heard as speech. Degraded non-words are heard as speech, though the constituent phonemes of the non-words are inaccurately identified. However, degraded speech disrupts serial recall less than un-degraded (clear) speech. What is required is that the vowel portion of the spoken utterances is preserved. If the vowels of speech are damaged by degradation the disruptive power of that speech is reduced relative to un-distorted speech (experiment 1a and 1b, chapter 5). The importance of vowels was further emphasized by the finding that vowel-only-changing (V-O-C) sequences of irrelevant sounds disrupt serial recall more than consonant-only-changing (C-O-C) sequences (experiment 2, chapter 6), replicating the findings of previous research (Hughes et al., 2005). Further, when the V-O-C sequences were degraded serial recall performance improved in their presence. However, degradation of C-O-C sequences did not render them less disruptive when their effect on serial recall was compared to clear C-O-C sequences. This shows information regarding change offered by V-O-C sequences as opposed to C-O-C speech sequences provide the critical changing-state information.

Critical pattern information provided by the formants pertaining to the vowels is important in preserving the temporal order of the sounds. The importance of vocal tract resonances (formants) and their relationship in the patterning of the speech signal over time to the effect of irrelevant speech on serial recall was highlighted by the finding that it is not fundamental frequency information (*f*0) that carries the important changing-state information. Instead, it is the structure of the other formants that carries important changing-state information from item-to-item in the irrelevant stream (experiments 3 and 4, chapter 7). When the fine structure of whispers is reversed, whispers can no longer be articulated or understood but they are still heard as produced by a voice. Both whispers and fine structure reversed (FSR) whispered speech reduced serial recall to the same extent (experiment 5, chapter 9). Reversing the fine structure of whispers does not distort formant structure in a way which reduces the disruptive effect of speech on serial recall. This provided evidence that it is the maintenance of formant structure over time that accounts for the level of serial recall interference observed in the presence of speech. When the relationship among formants was distorted by spectral rotation of the speech sounds, the intelligibility and 'speech-likeness' of the speech was completely destroyed. Consequently, spectrally rotated speech was heard as non-speech and spectrally rotating speech resulted in a reduction in the size of the ISE (experiment 6 and 7, chapter 11). This demonstrates that spectral rotation distorted speech pattern information critical to the speech/non-speech distinction observed in previous studies in the ISE paradigm (e.g. LeCompte et al., 1997; Jones et al., 2000).

It can be concluded that for irrelevant speech to preserve its disruptive power relative to non-speech sounds, speech sounds need to be perceived as vocalized sounds, hence sounds produced by a vocal tract (experiment 5, chapter 9 and 6 and 7, chapter 11). Therefore, vocal tract resonances (speech formants), the acoustic characteristics of voicing, seem to provide the critical changing-state information in the irrelevant speech stream.

The aim of this thesis was to examine whether it is the acoustic complexity or the phonetic detail of the speech signal which accounts for the higher serial recall interference observed in its presence in comparison to the effect of non-speech sounds. The research provided a direct test of the explanatory power of the changing-state-hypothesis (CSH), which advocates the heightened disruption of serial memory by irrelevant speech to the fact that speech exhibits more changing-state information than non-speech sounds. It is therefore important to evaluate the present findings, in particular the speech/non-speech distinction observed when acoustic complexity was preserved between irrelevant speech and non-speech sounds (experiments 6 and 7, chapter, 11) with reference to how far they can be explained by within the framework of the object-oriented episodic record (O-OER) model, and also from the perspective of other models addressed in the literature review (chapter 1).

13.1 IMPLICATIONS FOR MODELS OF THE ISE

13.1.1 The Working Memory Model (WMM)

The effect of irrelevant speech on serial recall is addressed by the working memory model displayed in figure 37 through the interaction of its phonological store and loop. The phonological store is a temporary store in which items decay over a short period of time, approximately three seconds. Information can enter the store using one of two routes. Irrelevant auditory verbal information has direct obligatory access from the auditory perceptual system, whereas the to-be-remembered (TBR) visual verbal information has indirect access to the store. Indirect access is achieved through the sub-vocal rehearsal of visual TBR stimuli which involves the grapheme-to-phoneme conversion by the articulatory loop (Baddeley and Hitch, 1974, Baddeley, 1986). As the irrelevant sounds automatically enter the store, they compete with the visual TBR items, which are also of a phonological representational format. Therefore, the ISE is argued to be the result of confusion between phonological codes in memory (Salamé and Baddeley, 1982). This assumption is represented as the more general phonological store hypothesis (PSH).

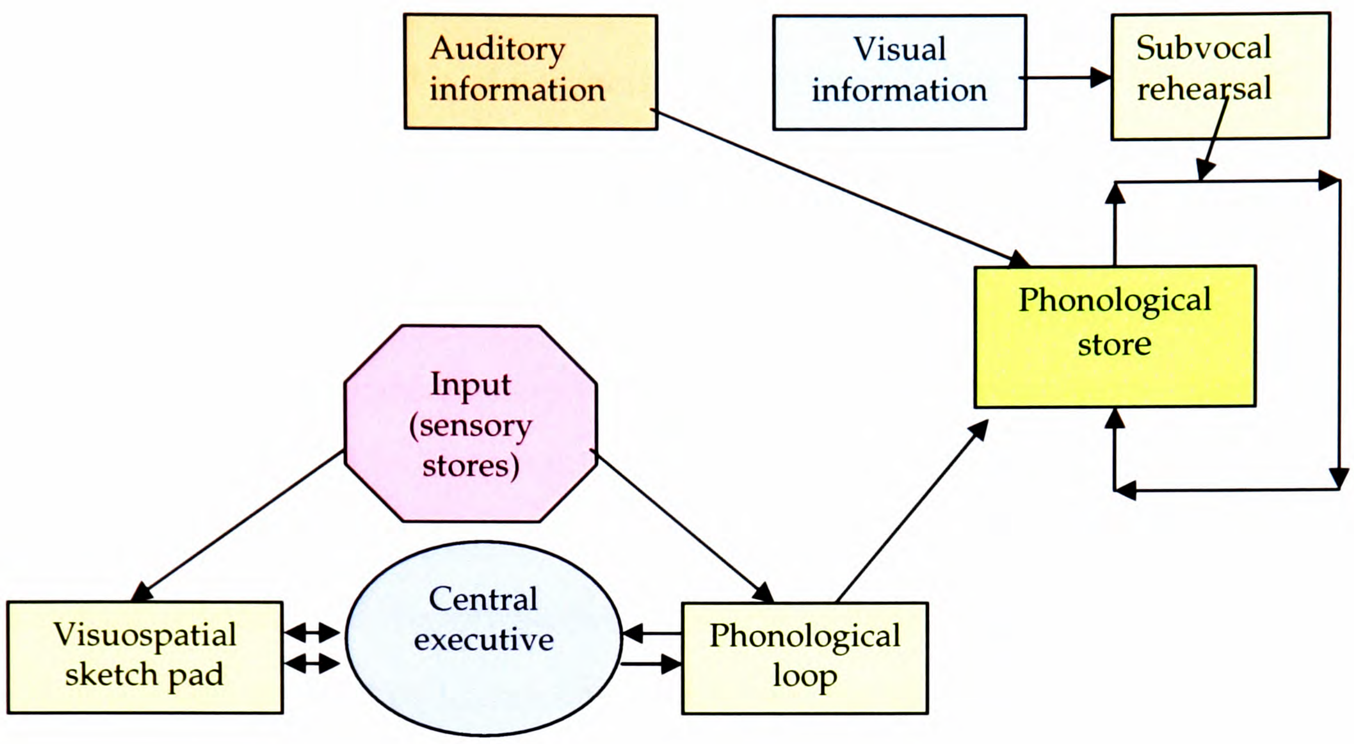


Figure 37. Simplified representation of Baddeley and Hitch's (1974) working memory model. Adapted from Baddeley (1990, p71).

The PSH would account for the finding that speech is more disruptive than non-speech sounds matched for acoustic complexity (experiment 6 and 7, chapter 11) because non-speech sounds cannot be converted into a phonological code and therefore only speech would be predicted to have access to the phonological store. In experiment 7 (chapter 11), no ISE was observed with the spectrally rotated speech perceived as non-speech, which is evidence for this assumption of the PSH. However, different types of changing non-speech stimuli have been shown to produce an ISE, such as sine-wave speech (Tremblay et al., 2000), pitch glides (e.g. Jones et al., 1993), band-pass noise (Tremblay et al., 2001b) and simple tones shifted in pitch (Jones and Macken, 1993). This body of evidence demonstrates the generality of the ISE and that irrelevant sounds do not need to be phonological in nature and that any sound exhibiting change between successive items will disrupt serial recall relative to a silent control.

However, when acoustic complexity is matched between speech and non-speech sounds (experiment 6 and experiment 7, chapter 11) speech is still more disruptive of serial recall than non-speech. It follows that it is the phonological nature of speech that is of importance in rendering it more disruptive, if not being the determinant of the ISE itself. Nevertheless, because non-speech items produce the standard ISE this shows the short-term memory (STM) store is not a store which holds only phonological representations and would better be characterized as an auditory store where confusion between auditory and visual items is at the level of the physical features of the items rather than at the level of phonological representation. This is appropriate also because the phonological identity of the phonemes making up non-words that act as irrelevant speech are not of importance. Rather the signal can be distorted by fine structurally reversing the signal so that acoustic-phonetic information is not removed but is distorted leading to non-words being perceived as sounds produced by a voice, but which participants cannot repeat back or understand. Hence, the phonetic coherence of the speech sounds is significantly reduced. Experiment 5 (chapter 9) showed that FSR whispered speech disrupted memory to the same extent as did whispered speech.

An alternative argument is that when auditory and visual items enter short-term memory, the memory representations generated for these items may be linked by temporal markers. It may be the confusion between two sets of temporal markers that results in the ISE. Speech may generate stronger temporal markers than non-speech, and this may explain why it is found to be more disruptive of serial recall (e.g. LeCompte et al., 1997; Jones et al., 2000 and Experiment 6 and 7). Evidence to support this notion comes from research demonstrating that the temporal order judgment of sounds can be affected by the type of sounds employed. Cole and Scott (1973) found that the order of CVC syllables which exhibit vowel transitions within an auditory loop is better judged than the order of auditory loops consisting of consonant noise only. The irrelevant speech sequence for experiment 6 and 7 comprised changing vowels, whereas the spectrally rotated speech stimuli were perceived as non-speech and information pertaining to the vowels was distorted. Therefore, irrelevant spectrally rotated speech would generate weaker temporal markers than irrelevant speech, leading to less confusion in short-term memory between irrelevant auditory items and visual TBR items. This may explain the greater disruption of serial recall by irrelevant speech observed in experiment 6 and 7.

The finding that distorted non-words which can no longer be perceived as non-words, but are still heard as vocal sounds, disrupt memory to the same extent as does undistorted speech, along with the finding that speech disrupts memory more than non-speech stimuli matched for acoustic complexity suggests that it is the biological nature of vocalisations in terms of their acoustic pattern which render them more disruptive of serial recall. It may be that biological sounds, hence vocalisations, are more disruptive of memory because they distract attention leading to the re-allocation of cognitive processing resources. Frequency-changing tones, changing cello notes and sine-wave speech not heard as speech produce an ISE, but are not as disruptive of serial memory as irrelevant changing speech (LeCompte et al., 1997; Jones et al., 2000 and Tremblay et al., 2000). It may be that the re-directing of processing resources is applied to speech because speech sounds are of behavioural relevance. Speech, in contrast to tones, may signal information about the environment that needs to be attended to. It is therefore plausible to assume that an attentional component needs to be specified in the working memory model in order for it to provide a more fitting account of the ISE.

The central executive is identified as an attentional component, however it is argued not to be involved in STM due to the assumed modular nature of the WMM and therefore the serial recall of TBR items cannot be affected by attentional modulation (Baddeley and Logie, 1999). However, secondary tasks that impede more on central executive functions have been found to disrupt serial recall more than secondary tasks that impede less on central executive functioning. This has lead Meiser and Klauer (1999) to argue that the working memory model could provide a better account of the ISE if the central executive was adopted as a mechanism which coordinates and supervises cognitive processing resources.

13.1.2 The Object-Oriented Episodic Record (O-OER) Model

The O-OER model like the WMM stipulates that attention does not play a role in the ISE. Unlike the PSH however all auditory information gains entry to an amodal short-term-memory (STM) store where both auditory and visual stimuli are represented by objects (Jones et al., 1996). The O-OER model posits that interference in memory by irrelevant sound can be understood in an *auditory streaming* framework that takes into account the role played by the perceptual organisation of sounds in the representation of the order of objects. Organisational factors lead to the segregation of concurrent objects into streams. The serial order of the objects representing the TBR items is encoded by cues constructed by their articulation. In terms of the irrelevant sounds, the pre-attentive and automatic processing of unattended sound involves the analyses of stimulus distinctiveness which determines the amount of information relating to the order of the sounds (Jones and Tremblay, 2000). The serial cues connecting the irrelevant sound objects are automatically generated and maintain the serial order of the auditory items. The automatically generated cues to the serial order of the irrelevant sounds thus compete with the cues pointing to the serial order of the TBR items. These competing order cues act to interfere with the rehearsal of links between

TBR objects STM. It can therefore be argued that the ISE is the result of a conflict of process not content (Jones and Tremblay, 2000).

Derived from the O-OER model is the CSH, which posits that it is not the nature of the sound, but the nature and extent of acoustic changes within an irrelevant stream which determines the degree of serial recall disruption (c.f. Tremblay and Jones, 1999). The number of changing-states in the sound sequence, which are broadly characterized by rapid variation in frequency and amplitude, determine the number of competing cues to serial order that will be formed (Jones and Tremblay, 2000).

Degradation of CVC non-words was found to effect consonant identification more than vowel identification (experiment 1a, chapter 5), but degrading V-O-C sequences resulted in them disrupting serial recall at a level that was equivalent to serial recall disruption in the presence of clear C-O-C sequences. Degradation however had no effect on the disruption produced by C-O-C sequences. As clear V-O-C sequences were found to disrupt memory more than clear C-O-C sequences, the reduced recall performance in the presence of degraded V-O-C sequences indicates critical changing-state information relating to changes in the vowels of successive irrelevant sounds is damaged. It can be concluded that the effect of signal degradation has its locus in the vowel portion of the CVC syllables as opposed to the initial or final consonants (experiment 2, chapter 6).

The finding that C-O-C sequences whether clear or degraded are less disruptive of serial recall than V-O-C sequences can be accounted by the CSH as consonants are assumed to provide less seriation information (Hughes et al., 2005). Changing vowels on the other hand are argued to elicit more serial order cues, an assumption based on the serial recall advantage observed for attended to V-O-C sequences as opposed to C-O-

C sequences (Surprenant and Neath, 1996). Hughes et al (2005) argue that items that are better serially recalled are more disruptive of serial recall. This is indeed the case as V-O-C sequences have been found to disrupt serial recall more than C-O-C sequences (Hughes et al., 2005), a finding replicated in experiment 2 (chapter 6).

That degradation has no effect on the disruption produced by C-O-C sequences refutes the principle assumption of the CSH, that is consonants changing from item-to-item should represent a change-in-state and so degradation C-O-C sequences should act to reduce serial recall interference in their presence. Although C-O-C sequences disrupt serial recall less than V-O-C sequences, they still have been shown by previous research to differ reliably from a silent control (Hughes et al., 2005). However, experiment 2 (chapter 6) did not feature a silent control condition and so it is not known whether or not the clear and degraded C-O-C sequences produced an ISE. Therefore if C-O-C sequences met the criteria of a changing-state sequence, as Hughes et al (2005) observed an ISE with C-O-C sequences, degradation of the changing-state information conveyed in C-O-C sequences should have resulted in a reduction in serial recall performance.

The biological nature of speech sounds, in that they are produced by a vocal tract, would not be predicted by the CSH to effect the changing-state information of the auditory objects as the nature of the irrelevant sounds is argued not to be important. On the basis of this assumption, the CSH would predict that no reliable difference would be observed between the effect of speech and non-speech (spectrally rotated speech) on serial recall with acoustic complexity controlled between auditory conditions. Matching the acoustic complexity between speech and non-speech sounds would mean that both would exhibit the same number of acoustic changes. Several studies have shown that bottom-up factors, such as acoustic change and primitive streaming, rather than phonological or semantic variables result in the ISE (c.f. Jones, 1999) and that semantic factors do not modulate disruption at least when the task involves serial rehearsal (Buchner et al., 1996). However, the greater disruption produced by speech relative to non-speech (spectrally rotated speech) is evidence that for speech, top-down variables such as hearing *irrelevant sound as speech* is responsible for serial recall disruption. It seems that the phonetic detail conveyed by speech, as in hearing sounds as being produced by a vocal tract, modulates the size of the ISE.

The O-OER model could be adjusted to explain the higher disruptive effect of speech if it is argued that attention is required for the seriation of TBR objects to take place (see figure 38). Along with the competing cues to serial order in STM resulting from the automatic seriation of irrelevant sounds, attention may also be distracted by auditory objects that may be of behavioural relevance, such as speech. Vocalised sounds may provide information about the environment and attentional resources may be re-directed to processing these sounds, as it may be important for these sounds to be attended to.

In terms of how serial order cues are derived for speech and non-speech sounds, if more cognitive processing resources are applied to the seriation of vocalised sounds this would lead to stronger cues to the serial order of these irrelevant sounds being generated in contrast to weaker cues being generated for the seriation of non-speech sounds. This adjustment of the O-OER model was put forward by Buchner et al (2006) to account for why emotionally negative and positive distractors disrupted serial recall more than did neutral distractors and why negatively valent distractors interfered with serial recall more than positively valent distractors. This finding can be viewed as evidence in support of a role of semantics in the ISE. However, the majority of experiments examining the influence of meaning on the size of the ISE have found that meaning does not play a role in the disruption of serial recall by task-irrelevant sound (e.g. Jones et al., 1990, and Buchner et al., 1996). Experiments that have examined the influence of the emotional valence of irrelevant sounds differ from those that have investigated the effect of the meaningfulness of sounds because emotionally valent speech sounds signal behavioural demands. Therefore emotionally valent speech sounds provide information in the environment that may need to be attended to (Buchner et al., 2004 and Buchner et al., 2006). It may be the case that all vocalised sounds signal possible behavioural demands. This may explain why Jones et al. (1990) who manipulated the meaningfulness of speech sounds found no difference in the level of disruption produced by speech played forwards, reversed speech and speech in a foreign language. In contrast, varying the emotional valence of speech would serve to increase the disruptive power of speech, by signaling different types of behavioural demands. For example, negatively valent speech sounds may be more disruptive than positively valent speech sounds because they may signal a possible threat in the environment (Buchner et al., 2004 and Buchner et al., 2006).

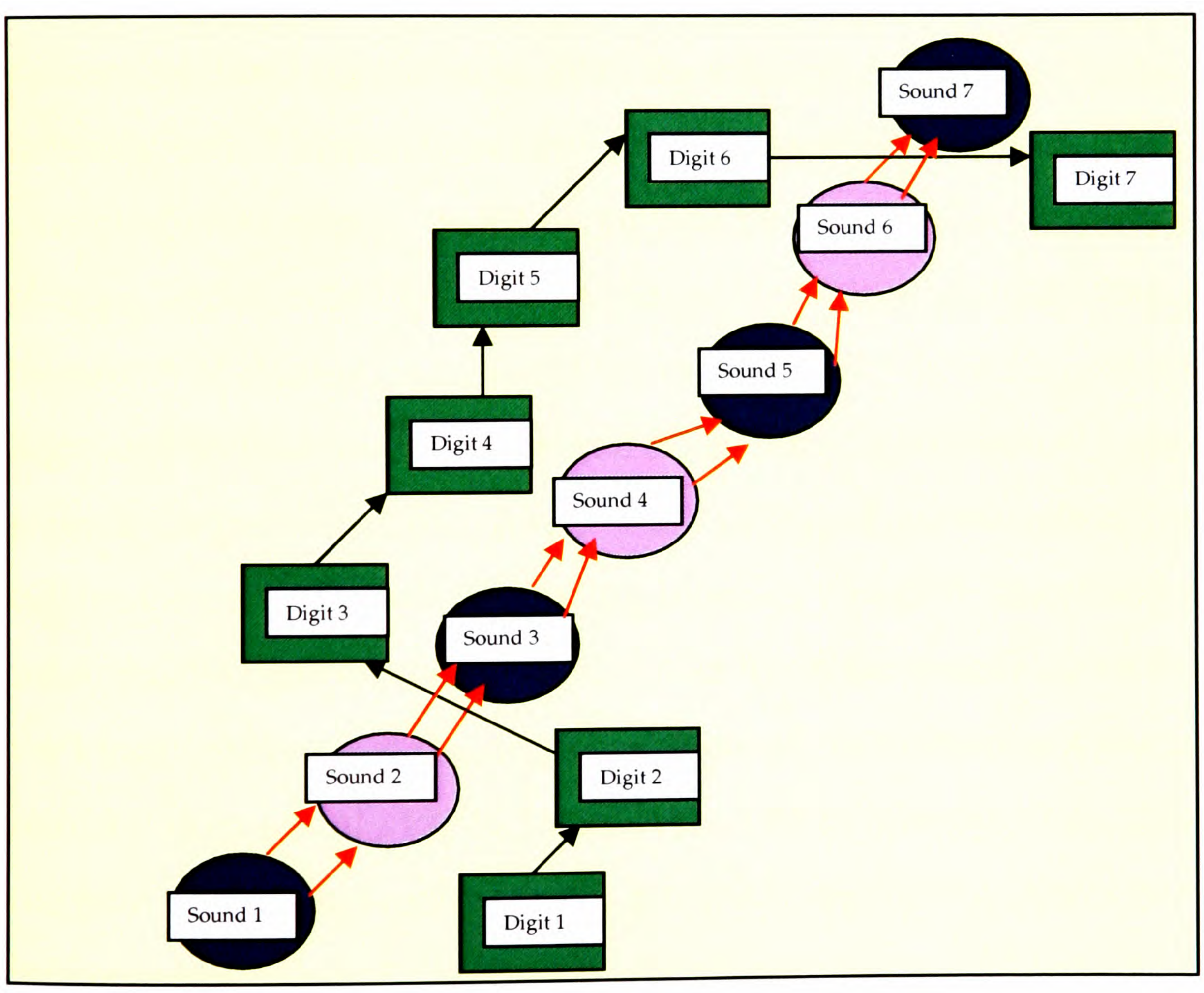


Figure 38. Schematic diagram of the interference of serial recall by task irrelevant vocalized sounds upon the episodic record including an attentional component. The red arrows connecting the irrelevant sounds represent the re-allocation of more cognitive resources to processing the serial order of vocalised (speech) sounds, due to them distracting attention. Two arrows connect the sounds, as the re-allocation of more cognitive resources to processing their order results in the generation of stronger serial order cues.

The O-OER model also posits that the effect of acoustic change is further determined by the perceptual organization of the irrelevant sounds for both speech and non-speech stimuli. Pitch is an attribute in an irrelevant sequence that can be varied from item-to-item to produce an ISE. For example, if an utterance (e.g. a consonant) is repeated, thus maintaining the phonological identity of the sounds, but the consonants differ in pitch, an ISE is found (Jones et al., 1999b). In terms of non-speech, irrelevant sequences of simple tones that change in pitch produce an ISE (Jones and Macken, 1993). Pitch is also an attribute of sounds that can be modulated to determine whether one or two streams of sounds are heard. This is because objects adjacent in a stream are compared and the level of stimulus distinctiveness determines the amount of information pertaining to their order. Hence, as the degree of change between successive items increases so does the amount of seriation (order information). However, as stimulus mismatch is increased a coherent changing-state stream is still perceived, but only up to a threshold of change, which might be described as a 'binding threshold'. Beyond this threshold of change, the changing-state sounds within the sequences segregate so that multiple streams of unchanging (identical repeated) sounds are perceived (e.g. Jones et al., 1999b). For example, if the pitch difference between two alternating tones or vowels is increased, initially serial recall disruption increases, but as the difference in pitch breaches the binding threshold the degree of memory interference is markedly reduced (Jones et al., 1999a; Jones et al., 1999b; Macken et al., 2003). Thus the relationship between acoustic variation and disruption is non-monotonic.

If consonants and their pitch are fixed in a changing irrelevant stream but the rate of their presentation is speeded up, serial recall disruption is reduced. This reduction in disruption is the result of the presentation rate exceeding the binding threshold and as a consequence the alternating sounds within the irrelevant stream segregate and two unchanging streams of identical items are perceived (Macken et al., 2003). These unchanging streams would constitute a steady-state sequence whose constituent sounds change less. Steady-state sequences produce less serial recall interference than do changing-state sequences featuring sounds that change more physically (Jones and Macken, 1993; Jones and Macken, 1995a; Neath, Surprenant, and LeCompte, 1998). Both speech and spectrally rotated speech sequences would have been heard as coherent streams of alternating sounds as the sounds in both auditory conditions were produced in a monotone and were presented at a rate of one item per second. Therefore the observed difference in serial recall disruption between both auditory conditions cannot be explained by a difference in streaming.

The finding that spectrally rotated speech did not differ in its disruption of serial recall from that observed in silence would not be predicted by the CSH, since spectral rotation of speech preserves the pitch of speech whilst destroying intelligibility. Bursts of broadband noise that change in band-pass frequency have been found to produce an ISE (Tremblay, et al., 2001). Bursts of broadband noise were described by Tremblay et al (2001) as sounding like noise, conveying a low level of tonality. Spectrally rotating the non-words *'teash'* (`tiS`) and *'forb'* (`f$b`) may have distorted or removed the tonal quality that broadband noise changing in band-pass frequency conveyed (see appendix 5 for examples of disc phonetic symbols). Therefore, it may be that tonal quality is required in stimuli in order for an ISE to be observed.

The fact that the speech sounds themselves were produced in a mono-tone by a male speaker may account for why spectrally rotated speech did not differ from the silent control (experiment 6 and 7, chapter 11). Irrelevant sequences of speech sounds (e.g. consonants) for which pitch and phonological identity was fixed only disrupt serial recall because their presentation rate is speeded up which results in the presentation rate exceeding the binding threshold (Macken et al., 2003). As a consequence, these alternating sounds segregate into separate steady-state streams. The spectrally rotated sounds in the present research were generated from two different non-words produced in a monotone. Therefore, it was the phonetic content which varied between the two non-words, the intelligibility of which was destroyed by spectral rotation. As the non-words were produced in a monotone the non-words would not have changed in pitch. Hence, although the pitch of the non-words themselves is preserved, this was not the physical attribute of speech which was varied between successive sounds. Instead the phonological identity and acoustic characteristics of the non-words varied. The formant structure of the vowels of speech sounds was found to be the attribute common to the speech sounds that carried the critical changing information between successive items (experiment 3 and 4, chapter 7). As spectral rotation destroys the relationship among the formants of vowels, this important carrier of change is also destroyed. Therefore, when phonetic detail is lost by spectrally rotating speech, the maintenance of its spectral and temporal structure and thus acoustic complexity may not be sufficient for a reliable effect of irrelevant sound.

The right hemisphere has been shown to play a critical role in the analyses of prosodic and melodic changes of sounds (Zatorre et al., 1992). Hadlington et al (2004; 2006) found a left ear disadvantage (LED) as irrelevant sounds presented to the left ear only produced a larger ISE than sounds presented to both ears and the right ear only. The LED has been explained by the suggested conflict between two processes of seriation (Hadlington et al., 2004). They argue the obligatory seriation of 'changing-state' auditory sequences as well as the seriation of TBR items may be the responsibility of the right hemisphere. Hence a conflict between seriation processes is suggested to occur in the right hemisphere; an argument which is consistent with the notion that disruption is caused by a conflict of process as proposed by the O-OER model. As the right hemisphere is argued to be specialised in processing the temporal and prosodic detail in a sound stream (Searleman, 1977) the automatic seriation of the irrelevant sounds would result in cognitive load in the right hemisphere being high. In addition the seriation of the TBR items by rehearsal is required and thus both concurrent processes of seriation would result in high cognitive load in the right hemisphere (Hadlington et al., 2004). Further, the fact that speech and non-speech both produce a LED indicates that the cognitive system processes irrelevant auditory stimuli in the same manner, regardless of the nature of the irrelevant sound (Hadlington et al., 2004). However, that the right hemisphere is specialised in processing pitch variation does not account for the greater serial recall interference produced by speech relative to non-speech (Spectrally rotated speech) (experiment 6 and 7, chapter 11) since spectral rotation preserves the pitch of speech (Beaman et al., 2007; Scott et al., 2000). This provides further support for the PSH, which argues that it is the nature of the irrelevant sound that determines its disruptive effect.

The greater disruption of serial recall by speech compared to non-speech found in experiments 6 and 7 (chapter 11) can be explained if the role of the right hemisphere in processing the steady-state information defining the changing vowels as opposed to the pitch dynamics of speech is considered. Vowels are broadly defined by steady-state information, such as formant frequencies, whereas consonants are broadly defined by rapidly-changing cues, such as fine distinctions in voice-onset-time (VOT) (Mirman, Holt and McClelland, 2004). The right hemisphere has been shown to be dominant in processing steady-state information and the left hemisphere has been found to be specialised in processing rapidly-changing information. Poeppel (2003) explained this hemispheric difference in processing steady-state and rapidly-changing information with reference to temporal integration windows. Rapidly-changing information is processed in the left hemisphere because it requires a shorter temporal integration window. Processing steady-state information requires a longer temporal integration window and is thus processed predominantly in the right hemisphere.

It can be argued that the greater serial recall disruption observed in the presence of irrelevant speech in contrast to non-speech is not due to the greater acoustic complexity of speech with regards to pitch variation as would be predicted by the CSH. Instead, since the right hemisphere seems to be the dominant hemisphere in processing steady-state information, the irrelevant speech/non-speech distinction may be due to the perception and processing of the formant frequencies of vowels changing-in-state from item-to-item in the irrelevant speech stream. Vowels have a richer acoustic structure than do non-speech stimuli such as tones and therefore provide more steady-state cues. This is plausible, considering changing vowels as opposed to changing consonants interfere more with serial recall (experiment 2, chapter 6; see also Hughes et al., 2005). Further, no important acoustic information is removed from the signal during spectral rotation (Blesser, 1972), but the relationship among the formants is destroyed which resulted in the non-words being completely unintelligible.

Hadlington et al (2004) did not directly compare the effect of speech and simple tones in a single experiment and it may be that although both types of sound produced a LED, speech may have produced a greater LED than the simple tones. This can be predicted since when speech and simple tones are presented binaurally, speech is more disruptive of memory (LeCompte et al., 1997). Further, considering the speech/non-speech distinction of the present research and the role of the right hemisphere in the processing of vowels, a greater LED for changing speech can be predicted (Poeppel, 2003) in contrast to a smaller LED for spectrally rotated speech sounds that do not change in pitch. Consequently, when spectrally rotated speech sounds are produced in a monotone, (and therefore do not change in pitch), the information regarding phonological identity which once provided the changing-state information in the speech is destroyed. It follows that the acoustic variation important for the observation of an ISE is no longer conveyed between the spectrally rotated speech sounds. Cognitive load in the right hemisphere would be predicted by the CSH to be higher in the presence of speech than non-speech stimuli if it is assumed that speech exhibits more order information, creating a greater conflict between the seriation of the irrelevant sounds and the TBR digits.

Alternatively, as the phonetic detail is present within the speech condition this should have resulted in bilateral activation of the STG, which has been observed with intelligible, forward speech (Scott et al., submitted, cited in Beaman et al., 2007). Spectrally rotated speech would not have demonstrated bilateral activation, rather right hemisphere activation would have been observed as it is not intelligible and conveys no semantic information. Right hemisphere activation has been observed for unattended spectrally rotated speech (Scott et al., submitted, cited in Beaman et al., 2007). Therefore, it may be that speech is more disruptive than non-speech stimuli due to the phonetic detail of the changing-state speech sounds, which is destroyed by spectral rotation.

13.1.3 An integrated model of attention and memory

Cowan's (1995; 1999) integrated model of attention and memory provides a general framework within which the effect of irrelevant sound can be explained. Immediate memory is viewed as the activated part of a more long-term store. It is argued that only one part of activated memory is attended to at a time. The rehearsal of TBR items keeps them activated and in the focus of attention. The ISE is explained with reference to it diverting attention from the task of rehearsing the TBR items. The observed distinction in memory performance in the presence of vocal (speech) and non-vocal (non-speech) sounds may be explained by the biological nature of vocalizations in that they may potentially provide information about the environment, since speech and vocalizations are important in communicating behaviourally relevant environmental states (Moore, 2000). It may be that irrelevant sounds processed by the auditory perceptual system as vocalized sounds are treated as requiring attention and therefore attract more processing resources away from the memory task.

Evidence that it is the possible behavioural relevance of task-irrelevant vocalizations that may account for the amount of disruption in their presence is provided by the finding that the emotional valence of an irrelevant sound can modulate the size of the ISE. Buchner et al (2004; 2006) found that positively and negatively valent irrelevant sounds interfere with serial recall performance more than neutral irrelevant sounds. Further, negatively valent irrelevant sounds cause more interference than positively valent irrelevant sounds. Cowan's (1999) conception of working memory explains the greater disruptive effect of emotionally valent distractors by arguing these distractors attract processing resources from the task of memorizing the order of the TBR items. Negatively valent distractors are more disruptive of memory than positively valent distractors because they may signal danger in the environment. This supports the notion that the nature of irrelevant sounds and not simply how much they change physically can determine their power to disrupt immediate memory. As irrelevant vocalized sounds may carry information relating to the environment then it can be argued that the cognitive system would re-allocate some attentional resources as it may be necessary for these sounds to be attended to.

The finding that low frequency irrelevant words cause more serial recall interference than high frequency words provides further evidence that the nature of sounds can modulate the size of the ISE (Buchner and Erdfelder, 2005). Cowan's model would assume that processing less frequent words would divert more attention from the task of rehearsing the TBR items than would more frequent words.

Cowan's (1999) model of working memory can also explain the observed difference between the effect of vocalized and non-vocalised sounds in the same way it has been suggested to account for the effect of emotional valence on disruption by irrelevant sound (Buchner et al., 2004; 2006). The amount of cognitive processing resources is represented by an attentional parameter. The focus of attention would be on the rehearsed TBR digits, which acts to keep the TBR items at a level of activation. Vocalised sounds may signal possible behaviourally relevant information and as a consequence would automatically attract attention away from the TBR items. This re-allocation of attentional resources in order to process the arguably ignored sounds would reduce the activation levels of the TBR digits and lead to a reduction in serial recall performance. Therefore, the vocalized/non-vocalised distinction can be explained if it is assumed that vocalizations attract more attentional resources than non-vocalisations.

Although this integrated model of attention and memory does not make any predictions regarding the relative disruptive potency of vowels and consonants it can offer an explanation as to why V-O-C sequences are more disruptive of serial recall than C-O-C sequences if it assumes that changing vowels attract more attention than changing consonants. The fact that C-O-C sequences have been found to produce an ISE (Hughes et al., 2005) and yet when they are degraded, disruption in their presence does not attenuate (experiment 2, chapter 6) indicates that changing vowels are more demanding of attention for their processing.

Although this model provides an account of the effect of vocalization of sounds, it is unable to account for other empirical findings within this paradigm. First, that a 'changing-state' sequence of sounds is more disruptive of serial recall than a 'steady-state' sequence cannot be accounted for by Cowan's (1999) model (Jones et al., 1992). Initially an attempt to explain this finding was based on the participant habituating to the unattended steady-state sequence leading to less attentional resources being allocated to process the sequence (Cowan, 1995). However, participants do not habituate to a changing sequence of sounds over numerous trials and the ISE is observed even when experimental sessions are days apart (Ellermeier and Zimmer, 1997). Another constraint is that this model does not incorporate a mechanism that controls the perceptual organization of sounds, which plays an important role in mediating the degree of serial recall interference (e.g. Jones et al., 1999a and 1999b).

13.1.4 The feature model

The feature model (Nairne, 1990) stipulates that TBR items are represented in memory as a set of features. The successful recall of items is determined by the match between the primary and secondary memory features of the TBR items. Primary memory representations consist of two types of features. Modality-dependent features encode the physical information conveyed by each item, whereas modality-independent features encode the internal responses to an item, such as the verbal label of an item (Neath, 2000). Auditory stimuli have more modality-dependent features than visual stimuli and modality-dependent features of an item are only overwritten by similar features of the items following in a list. Since auditory and visual stimuli do not share any modality-dependent features, interference of serial recall by irrelevant sound is related to the corruption of the modality-independent features of the TBR items (Neath, 2000). The sounds in an irrelevant sequence would add modality-independent features to the representations of the TBR items in memory. This would attenuate the likelihood of a successful match between an item's primary and secondary memory representation

(Neath, 2000). Stimulus interference is therefore argued to be responsible for the disruption of serial recall. Like the PSH, the feature model is constrained by its assumption that interference is dependent on the similarity of items in the irrelevant and relevant streams. However, it is the dissimilarity between items in the irrelevant and relevant streams which is important as irrelevant sounds which do not rhyme with the TBR items are more disruptive than when sounds in the irrelevant stream rhyme with the TBR items (Jones and Macken, 1995a).

The emphasis on the importance of the similarity of item identity is in direct contrast to the O-OER model which argues that serial recall interference is a product of the similarity in the serial processing of both the unattended and attended to streams. As the notion of feature adoption refers to the modality-independent features of the irrelevant sounds being added to those of the TBR items, variations in pitch could not result in feature adoption. If successive sounds varied in only pitch, as this is a modality-dependent (physical) feature of sound an ISE would not be predicted. However, changes in the pitch of both speech and non-speech sounds produce an ISE (e.g. Jones et al., 1999a).

In addition the feature model cannot account for the ISE observed with non-speech stimuli in previous studies (e.g. Jones and Macken, 1993). It argues that non-speech stimuli cannot be subject to feature adoption and views disruption in the presence of non-speech stimuli as being representative as a different effect, one that is independent from the disruptive effect of irrelevant speech. However, speech and non-speech sounds have been shown to be functionally similar in their disruption of serial recall. For example, the relationship between interference and token-set size is the same for speech non-speech stimuli. As the number of different tokens in the irrelevant stream increases from one to two, memory disruption also increases. However, further increases do not result in a reliable increase in serial recall disruption (Tremblay and

Jones, 1998). Also, as discussed when considering the O-OER model, a non-monotonic relationship between stimulus mismatch and disruption is observed for speech and non-speech stimuli when the effects of factors influencing the perceptual organisation of sound, such as pitch are investigated (Jones et al., 1999a; Jones and Macken, 1995b; Jones et al., 1999b).

The feature model can account for the greater disruptive effect of speech compared to non-speech if it is argued that speech exhibits modality-independent features that are more similar to the modality-independent features of the TBR items. Hence, the internal encoded response to both the irrelevant and irrelevant stimuli would be verbal in nature (LeCompte et al., 1997). However, another problematic constraint of this model is that it predicts that irrelevant sounds need to be concurrent with the TBR items either at encoding or at rehearsal in order for interference by feature adoption to take place. Each irrelevant sound in the present experiments was synchronised with the presentation of a visual TBR digit. However, the ISE has been observed when irrelevant sounds are presented concurrently with the TBR items and during a retention interval (Jones et al., 1992). As sub-vocal rehearsal takes place as the sounds are presented it is difficult to test whether this constraint is ever met (Buchner et al., 2006).

Surprenant and Neath (1996) explain the better serial recall of V-O-C sequences compared to C-O-C sequences, when the identification of V-O-C sequences is reduced by the addition of noise, to a level below that observed for un-degraded C-O-C sequences within the framework of the feature model. They argue the modality independent features that encode the verbal labels of vowels may form more durable memory representations than those of consonants. Therefore, when identifying the V-O-C sequences the memory representations generated reflected by the verbal label would have been more discriminable leading to them being better serial recalled. However, the difference between the serial recall of the V-O-C and C-O-C sequences was not reliable. It can be argued that the recall advantage for sequences of V-O-C sounds must be determined by the modality-dependent features of the vowels. When V-O-C sequences were better identified than C-O-C sequences they were subsequently recalled better in their serial order (Surprenant and Neath, 1996). This was explained by the modality-dependent features of vowels being more useful than those of consonants. The similarity of the modality-dependent features defining stop consonants is suggested to be greater, making them less discriminable. Vowels on the other hand are argued to comprise modality-dependent features that are less similar and hence more discriminable (Surprenant and Neath, 1996).

Since Jones and Tremblay (2000) suggest the ISE is a product of the conflict between seriation processes as opposed to the similarity of content between irrelevant and relevant sequences, then the physical information encoded by the modality-dependent features might be of importance in determining the magnitude of the ISE rather than modality-independent features. That the modality-dependent features of vowels are thought to be more discriminable than those of consonants might address the higher interference of serial recall observed in the presence of irrelevant V-O-C relative to C-O-C sequences (experiment 2, chapter 6; see also Hughes et al., 2005) if it is assumed that because vowels are more discriminable they will in turn generate more cues to their serial order. The fact that V-O-C sequences are better recalled when attended to also suggests that they provide more order information (Surprenant and Neath, 1996). More cues to serial order would mean more cues which would compete with the cues pointing to the order of the Visual TBR items. The finding that degrading V-O-C sequences resulted in them producing an ISE that was equivalent to that obtained by C-O-C sequences, whether clear or degraded suggests that degradation acts to reduce the discriminability of the changing vowels.

Since the modality- dependent features of consonants are more similar and thus less discriminable, degradation of these sequences will have no effect on the level of serial recall.

Although auditory and visual stimuli do not share any modality-dependent cues (Neath, 2000), greater memory interference by speech in contrast to changing non-speech stimuli (spectrally rotated speech) may simply be because the modality-dependent features of non-speech sounds are similar to those of consonants in that they are less discriminable due to the greater similarity between the modality-dependent features of consonants. It follows that the modality-dependent features of speech that convey vowel changes are better discriminated and thus generate more competing serial order cues than spectrally rotated speech sounds, because the physical information pertaining to the vowels is distorted. This seems plausible as the perceptual system is argued to integrate speech sounds by using a similarity shared by the vowel sounds (Bregman, 1990). This similarity seems to be provided by the formant structure of the vowels, since when this is destroyed serial recall performance under spectrally reversed speech is improved relative to performance in speech (experiment 7, chapter 11).

Evidence supporting the notion that the modality-independent features are not of importance to the ISE is provided by the finding that fine structure reversed (FSR) whispers, which are heard as being produced by a voice but cannot be repeated back or understood, produce an ISE equivalent to that produced by normal whispers (experiment 5, chapter 9). FSR whispers have the same acoustic information as normal whispers and importantly the relative spacing of the formants of the vowels is not distorted as the spectral detail is only reversed in time. Hence, critical speech pattern information, in particular information relating to the vowels, which leads to the perception of sounds as emanating from a voice, is still present. As the non-words which form the FSR whispers are no longer perceived, arguably a verbal label cannot have been attached to the FSR whispered stimuli and thus the modality-independent information would not have been as useful as that of the normal whispers and yet no reliable difference between the disruptive effects of these stimuli was observed (experiment 5, chapter 9). A diverse range of non-speech stimuli has produced an ISE (e.g. Jones et al., 1993; Tremblay et al., 2000). This indicates it is the physical acoustic (modality-dependent) features of irrelevant sounds that are critical to the ISE being observed.

If it is assumed that the physical attributes of sounds as opposed to their modality-independent features are responsible for serial recall disruption this would result in the feature model being able to account for more of the empirical findings, in particular the ISE observed with non-speech stimuli. This assumption would provide a framework from within which the effects of bottom-up variables, such as changes in pitch between successive speech and non-speech stimuli could be explained.

However, as the feature model sees irrelevant sound as adding modality-independent items to those of the TBR items it makes no predictions regarding a conflict between seriation processes in the irrelevant and relevant streams. Instead, it proposes that the level of seriation required by a memory task is not responsible for the magnitude of the ISE and that the nature of the items is important as opposed to the rehearsal strategy adopted. The model would need to incorporate a mechanism for seriation. The model does however include an attentional parameter and this has been used to account for the changing-state effect. The greater disruptive effect of changing items relative to repeated items is argued to be found because a sequence of changing items is harder to ignore than a sequence of repeated items. The harder it is to ignore a sequence of sounds, the more attention will be directed away from the seriation of the TBR items (Neath, 2000). The notion of an orientating response to irrelevant sounds has been used to explain why repeated sounds interfere with serial recall less. It is argued that over time, the amount of attention directed to repeated sounds is reduced (c.f. Cowan, 1995). Contrary to this assumption, research has shown that participants do not habituate to the effects of irrelevant sound, as the ISE is observed over blocks of trials (e.g. Tremblay and Jones, 1998). However, that attention does not habituate over time even for repeated sequences is emphasised by the notion that incoming auditory information is automatically processed and it makes sense that the degree of processing resources diverted to the irrelevant sound would not attenuate over time. Hence, it may well be that less processing is required for an unchanging auditory sequence in contrast to changing auditory sequences. If the model assumed that the more discriminable modality-dependent features of changing-state speech, in particular the features reflected by the vowels, detracted more processing resources from the TBR items than those of changing non-speech stimuli it could account for greater disruptive power of irrelevant speech.

13.2 IMPORTANCE OF PATTERN RECOGNITION

That spectral rotation destroyed the mismatch in phonological identity between the speech sounds, which provided changing-state information, is one account of why the acoustically matched changing-state information of the spectrally rotated speech disrupted memory less (experiment 6 and 7, chapter 11). An alternative account is one based on what the spectral rotation manipulation does to critical information in the speech pattern. Spectral rotation of speech preserves the temporal and spectral structure of speech (Blesser, 1972) and thus important information inherent within the first three formants of speech is still conveyed in the signal, though at higher frequencies. As spectrally rotated speech is not as disruptive of serial recall, this indicates destruction of the relationship among formant frequencies within the signal distorts critical speech pattern information important in rendering speech more disruptive than non-speech. Therefore, the top-down processing of speech afforded by the familiarity of critical speech pattern information may account for the greater disruptive effect of speech. It may be that the top-down processing of speech sounds may result in a more durable code for speech sounds in memory than non-speech stimuli.

In light of the differential disruptive effect of speech and spectrally rotated speech, the finding that sine-wave speech, whether perceived as speech or not, is less disruptive than natural speech can be explained (Tremblay et al., 2000). Sine-wave speech can be perceived as speech with training as it is constructed from three sinusoids that track the first three formants of speech (Remez et al., 1981). The first three formants have been argued to be important for speech recognition (Moore, 2004) and the perception of vowel quality (Strange et al., 1983). However, the sinusoids would not have the complex structure that the formants in natural speech exhibit. In particular, the steady-state information which broadly defines vowels would be significantly reduced. Therefore, sine-wave speech is only perceived as speech with training, due to its reduced spectral detail and ambiguous nature relative to natural speech. Thus, the perception of sine-wave speech as speech after training can be described as a *problem solving process* as not all the information inherent within the natural speech pattern is present. Accordingly, one inference that can be made is that top-down processing of speech due to the familiarity with the properties of speech sounds may account for higher memory interference by irrelevant speech relative to non-speech.

Speech is over learned in terms of the recognition of its pattern. Functional imaging studies provide evidence that the auditory system will attempt to process any sound as if it were intelligible speech if the signal exhibits acoustic-phonetic information. For example, research has demonstrated equivalent bilateral activation of the superior temporal gyrus (STG) by speech, reversed speech and words in an unfamiliar language (e.g. Binder et al., 2000). This suggests that the STG processes the acoustic complexity of speech and since all these stimuli exhibit some phonetic information, this region may process the phonetic detail of speech as well (Binder et al., 2000). It can be argued that if the auditory system processes sound as speech, and this is attributable to acoustic-phonetic processing, then sounds conveying acoustic-phonetic information will be processed similarly independent of their intelligibility (Scott and Wise, 2004). That sounds conveying natural acoustic-phonetic information are processed similarly may account for the equivalence in serial recall interference found with whispered and FSR whispered speech sounds. However, sine wave speech contains no acoustic-phonetic information. The three sinusoids it consists of only track the patterning of the first three formants in time and therefore convey some of the temporal detail of speech but not its spectral complexity over time. Therefore, sine wave speech is not processed as speech and can only be heard as such through training.

If the perceptual system is able to process sound as speech, without the need of training, it may be more disruptive of serial recall because it is the ease with which the auditory perceptual system *recognises the pattern* of an incoming auditory signal which dictates how disruptive it will be of memory. Since speech is an over learned pattern, the perceptual system will preattentively integrate and decode the sounds at a faster rate than non-speech due to top-down processing. Following from the account of the differential effect of speech and spectrally rotated speech, top-down processing of speech may result in more durable representations of speech items being encoded in memory. As a consequence it may be that the seriation of TBR items during rehearsal is disrupted more by speech than spectrally rotated speech because more durable memory representations would result in more seriation. If when non-speech sounds are presented less durable memory representations are generated, then they would decay at a faster rate. As a consequence, the degree of seriation at either the encoding or rehearsal stage would be less, reducing the conflict produced between the automatic seriation of unattended sounds and the attended to visual TBR items.

The CSH makes no predictions regarding the effects of pattern recognition on the relative disruption by irrelevant speech compared to non-speech. It would need to be adjusted to make predictions regarding the duration of the objects in memory representing the irrelevant sounds, and the effect this has on the seriation of the TBR items.

13.3 CONCLUSIONS

Speech sounds need to be perceived as being produced from a vocal tract and the relationship between vocal tract resonances (speech formant) provided by the vowels must be preserved, in order for speech to remain more disruptive than non-speech. Therefore, the speech/non-speech distinction observed in experiment 6 and 7 (chapter 11) is better characterised as a distinction between irrelevant vocal and non-vocal sounds. The absence of equivalent serial recall interference in the presence of speech and non-speech (spectrally rotated speech) matched for acoustic complexity is problematic for the CSH, which argues disruption is determined only by the pre-attentive processing of bottom-up acoustic factors (Jones et al., 1996; also see Jones et al., 2004). This suggests the O-OER model would need to be adapted to include an attentional parameter which could account for the greater serial recall interference observed in the presence of sounds perceived as speech as opposed to non-speech sounds.

The fact that speech is not more disruptive than non-speech once acoustic complexity is controlled between auditory conditions gives support to the notion that the nature of irrelevant sounds is important. The speech/non-speech distinction may therefore be better accounted for with reference to top-down processing due to speech pattern recognition. Since speech is an auditory stimulus used in communication, it is an over learned stimulus. There is a top-down component to speech processing because of its over learned nature and it may be that its pattern leads to more durable representations in memory allowing for greater interference of serial recall. Spectrally rotating speech distorts the pattern information of speech which seems critical in distinguishing the effect of speech and non-speech sounds. That is speech is no longer perceived as speech or sound produced by a voice. This is due to the relative spacing of the formant frequencies of the vowels being distorted. Destroying the intelligibility of speech leaves a spectrally altered pattern which cannot be processed as speech, and which is less disruptive of serial recall. This suggests that it is the characteristics of voicing which is of key importance in rendering irrelevant speech the most disruptive sound.

The possible behavioural relevance of sounds conveying vocal characteristics which are reflected in the natural speech pattern may also serve to attract more processing resources from the memory task at hand. Since spectrally rotated speech is perceived as complex noise, it would not have any behavioural relevance and thus may have diverted less processing resources from the task of remembering the order of the TBR items.

13.4 FUTURE WORK

Future work is required to test the notion that it is the formants of vocalised sounds and the possible behavioural relevance of these biological sounds which lead them to attract more processing resources than non-vocalised sounds. If the speech/non-speech distinction is a product of processes relating to pattern recognition as opposed to 'speech being special', any type of vocalised sound exhibiting a formant structure, regardless of its acoustic complexity, would be predicted to disrupt serial recall to the same extent as speech, as long as it formed a changing-state sequence. Any vocal sound with a formant structure such as an animal cry would convey steady-state information (e.g. formant frequencies) as do vowels, and it is this changing information which produces the most interference (experiment 2; chapter 6; see also Hughes et al., 2005). This is because animal cries contain many of the acoustic characteristics of human speech (Moore, 2000). This can also be predicted on the grounds that FSR whispers disrupted memory to the same extent as did normal whispers. FSR whispers could not be articulated, however their formant structure was preserved and so they were heard as sounds produced by a vocal tract, though unintelligible.

Further research is also needed to examine whether the importance of the formant structure in irrelevant speech concerns the information it provides for the maintenance of order information or if it is the general sensory processing of formant structure which may instead lead to attentional distraction which accounts for why speech is more disruptive than non-speech. If formant structure is important solely because it provides information about the order of sounds, a changing-state sequence of sounds produced by different voices should be less disruptive than a changing-state sequence of sounds produced by a single speaker. According to the change on a common ground principle (e.g. Jones et al., 1999a and 1999b), auditory items produced by one speaker would have a common formant structure, and thus change between successive items would be carried on the common carrier of the formant structure. A sequence of changing sounds alternating from speaker to speaker would not exhibit a common formant structure and so change between adjacent items would not be carried on a common attribute. If however, manipulating the number of voices within a changing sequence of sounds does not modulate the size of the ISE produced by irrelevant speech, this would suggest that attentional distraction in terms of the recruitment of processing resources away from the memory task was responsible for the degree of serial recall disruption.

Serial recall disruption was found to be reduced in the presence of degraded V-O-C sequences. Because only two levels of degradation were used in experiment 2 (chapter 6), it is not clear where the point of inflection is in the relationship between V-O-C item degradation and disruption. Hence, the threshold of degradation beyond which degraded V-O-C items begin to disrupt serial recall less. Degrading V-O-C items along a continuum of degradation from 0 to 100% would allow a precise measurement of the threshold of vowel degradation beyond which serial recall disruption is reduced. Alongside a more systematic and parametric measure of the function relating V-O-C item degradation and disruption, a measure of the discriminability of degraded V-O-C sequences would be of importance. Measuring the discrimination of V-O-C items at each level of degradation along the continuum used in the ISE measure would allow for investigating whether or not vowels beyond the threshold of degradation in an ISE paradigm produce less disruption because they are discriminated less well.

The above measures would test the assumption of the CSH (e.g. Hughes et al., 2005) that V-O-C items are more disruptive because not only are they recalled better in their serial order than C-O-C items (Surprenant and Neath, 1996), but they also have implications for the temporal integration of speech sounds by the perceptual system (Bregman, 1990). It would also provide a test of the relative importance of the modality dependent (physical) features of irrelevant V-O-C and C-O-C items. If V-O-C sequences are more disruptive because they are better discriminated, vowels degraded at a level beyond the degradation threshold (were degraded sounds disrupt memory less) should be less discriminable. Furthermore, the discrimination of these degraded vowels should differ reliably from the level of discrimination observed for degraded vowels that do not exceed the discrimination threshold. If this is the case, it would show that degrading vowels at a certain level renders them less discriminable. It may be that vowels degraded at this level are discriminable at a level that is equivalent to the level of discriminability observed for consonants. This can be predicted on the basis that consonants have been found to be less discriminable than vowels. The better discriminability of clear V-O-C irrelevant sequences may account for why these sequences disrupt serial recall more than C-O-C sequences. The effect of degradation on C-O-C sequences along a continuum of degradation from 0 to 100% on serial recall disruption would need to be measured as well as the discrimination of the degraded C-O-C items. Only then could the possibility that degraded vowels are as disruptive as clear C-O-C sequences because their discrimination is similar to that observed with C-O-C items be examined. This would provide a way of mapping discrimination onto serial recall performance in the presence of irrelevant speech.

13.5 PRACTICAL APPLICATIONS

Workplace environments involving tasks that require short-term memory are particularly susceptible to the disruptive effects of irrelevant sound, such as pilot cockpits and open plan offices. In terms of open plan offices, common complaints include distraction from people talking, phones ringing, office machinery and air conditioning (Banbury et al., 2001). The few studies which have examined the disruptive effects of irrelevant sound on office related tasks show that tasks, especially those with a high demand on seriation are sensitive to interference by irrelevant sound (Banbury and Berry, 1997; 1998). Research into the effects of background sound on cognitive performance has implications for the design of open-plan offices where workers are subjected to irrelevant sound, in particular the disruptive effects of background speech, which may be of no relevance or importance to them but whose acoustic characteristics will act to disrupt their cognitive performance (Banbury et al., 2001).

One way to lessen the disruptive effect of irrelevant speech is to mask the changing-state information within the speech signal. The 'babble' effect reported by Jones and Macken (1995c) demonstrated that when changing sounds are presented monaurally, manipulating the number of voices occurring concurrently modulates the magnitude of its interference on memory. As the number of voices presented increased from one to two and also from two to three, an increase in the level of disruption was observed. However, above three voices, the degree of disruption was attenuated and when six voices were simultaneously presented, disruption was significantly reduced. However, if each of the voices was presented from a different location in space, their power to disrupt was restored (Jones and Macken, 1995c). This has been explained in terms of the effect 'babble' has on the signal. The amount of change in energy at the boundary of sounds is related to the observed level of interference. Babble masks the energy at the boundaries of spoken utterances and therefore serves to reduce the cues available for the segmentation of the speech sounds making up a stream of speech and thus an irrelevant auditory stream no longer represents a changing-state stream. This research indicates that workers in smaller offices are more likely to be adversely effected by irrelevant speech as it is less likely that the irrelevant speech sounds of co-workers will mask each other and thus reduce the cues to segmentation sufficiently to lessen memory disruption.

Highly reverberating speech has been shown to remove the ISE, an effect which has been explained with reference to reverberation smearing the profile of the speech signal (Beaman and Holt, 2007). Smoothing the profile of the sound smears the boundaries between adjacent words by suppressing the peaks and troughs in the irrelevant stream. This would act to possibly prevent sufficient segmentation of the words in the irrelevant speech stream. Therefore, reverberation acts to lessen the number and extent of acoustic variation in an irrelevant speech (Beaman and Holt, 2007).

That spectral rotation does not reduce the number or extent of acoustic changes within speech, and yet speech is still more disruptive than spectrally rotated speech (experiment 7, chapter 11) suggests that it is change in the spectral information and how it evolves over time which is important in rendering speech more disruptive than non-speech. Reverberation not only smoothes the profile of irrelevant speech, it also acts to smear and corrupt its harmonic structure (Roman and Wang, 2005; Wu and Wang, 2006). If the harmonic structure is corrupted so will be the formant structure. As it is the formant structure which seems to be the necessary common carrier of changing-state information between successive speech utterances (experiment 3 and 4, chapter 7), its corruption by reverberation offers an alternative account of why the ISE is removed in the presence of highly reverberated speech. Regardless of whether it is the smoothing of the temporal patterning of the irrelevant sound, or the corruption of its formant structure, reverberation acts to improve cognitive performance. This is in direct contrast to the construction of open plan offices. Manufacturers design and fit acoustic ceilings which serve to absorb rather than reflect sound. These act to attenuate the degree of echo experienced (Beaman and Holt, 2007). However, the research by Beaman and Holt (2007) suggests that reducing reverberation does not serve to reduce the distraction by irrelevant sound experienced. The research of this thesis adds support to this finding as it shows that the formant structure of vowels produced by the same voice is an important common carrier of changing-state information. Since reverberation is found to smear the harmonic structure of speech it would also act to corrupt the formant structure of irrelevant speech. Therefore, large open plan offices require acoustic ceilings that will increase reverberation and therefore reduce the disruption of cognitive performance.

In pilot cockpits, the allocation of a number of tasks to automated systems has left a number of cognitive tasks to be undertaken by pilots. The increase in automated systems has also resulted in an increase in irrelevant sound experienced in the cockpit (Banbury et al., 2001). Conjoined with voice communication within and between aircrafts, the amount of background sound in cockpits has been further increased by automated auditory messages. Not all the sound received by the flight crew is of relevance to them and it can occur at irregular intervals (Banbury et al., 2001). Research looking at performance on visual-spatial tasks in the presence of irrelevant sound has implications for investigating the detrimental effect of irrelevant sound in pilot cockpits as these tasks are representative of those carried out on the flight deck. Banbury, Jones and Emery (1999) showed that performance on visual-spatial tasks is adversely affected by task-irrelevant speech. Recall of a moving target's track history on a radar display was found to be reduced in the presence of irrelevant speech. In addition, Banbury et al (1999) examined the effect of irrelevant cockpit sound on memory for navigation information regarding longitude and latitude. Participants were presented with an incoming auditory message, which they were instructed to retain in memory for a brief period. Participants were then asked to recall the message in written form. Irrelevant auditory messages from other aircraft were presented during the retention interval which the participants were instructed to ignore. The recall of navigation information was significantly disrupted by the irrelevant auditory messages relative to recall performance in a silent control or in the presence of irrelevant ambient noise. Therefore, in order for flight crew to accurately monitor system displays over time, the temporal order of information must be maintained and this is significantly disrupted by irrelevant spoken auditory messages. Hence, sufficient system monitoring requires seriation processes to be un-disrupted by irrelevant speech.

Banbury et al (2001) argue situational awareness is important for not only the immediate comprehension of aircraft system states but also for the prediction of future system states. Errors made inputting navigation-based information is not the only system for which inaccurate performance may be detrimental. Banbury et al (2001) intuitively point out that the cockpits of military aircraft would also benefit from research into background sound and cognitive performance. Auditory alerts for situations of low importance may result in errors being made by flight crew inputting flight coordinates into systems delivering weapons. Therefore, the sound experienced on flight decks needs to be controlled in a way that allows more accurate cognitive performance. Banbury et al (2001) suggest the use of digital storage which would allow the timing of non-critical automated auditory messages to be controlled, which would serve to reduce error rates in performance during critical system analysis.

The finding that distorting the acoustic pattern of the speech signal renders spectrally rotated speech not as disruptive as speech has implications for the design of cochlear implants. Cochlear implant listeners have a limited capacity for processing the speech signal and there is evidence that the reduced cues they have available are more sensitive to distortion of the spectral detail than the temporal information of speech. Shannon et al (1998) showed that spectral shifting and warping of the frequency information in a speech signal reduced to four frequency bands had a more disruptive effect on vowel recognition as opposed to consonant recognition. Vowel recognition was often found to be reduced to that observed with single spectral channels. Further, not only was vowel recognition poor in comparison to consonant recognition which was relatively good, sentence recognition was completely effected. Shannon et al (1998) inferred that this indicates either consonant and vowel recognition must be at a certain level before words can be construed from the speech signal or vowels as opposed to consonants are critical for sentence recognition. Drullman et al (1994) found in contrast that smearing the temporal cues of speech had a greater disruptive effect on consonant recognition than vowel recognition. Therefore, when speech is reduced to the minimum spectral representation that results in good speech recognition, the distortion of temporal and spectral cues affects consonants and vowels differently (Shannon et al., 1998).

In the present research, spectral rotation preserved the long-term features of the speech signal, but destroyed the patterning of formant structure in time (Lachs and Pisoni, 2004). Hence, the spectral cues were distorted as spectral rotation resulted in destroying the relationship among the formants. Lachs and Pisoni (2004) showed that spectrally rotating speech had a detrimental effect on word recognition. Consistent with this, the present results provide further support that spectral information is critical for word recognition and that distorting the relative spacing between the formants of speech has a detrimental effect on intelligibility. Thus, as the spectral representation of the speech signal is reduced in cochlear implant listeners, designers of cochlear implants need to ensure that the tonotopic (spatial) distribution of spectral envelope cues in the speech signal is preserved to allow for successful retrieval of lexical information.

References

Allard, F., and Scott, B. L. (1975). Burst cues, transition cues, and hemispheric specialization with real speech sounds. *Quarterly Journal of Experimental Psychology, 27,* 487-497.

Allport, D. A. (1989). Visual attention. In: M. I. Posner (Ed.), *Foundations of Cognitive Science,* (pp.631-682). Cambridge, MA: MIT Press.

Assmann, P. F., and Katz, W. F. (2000). Time-varying spectral change in the vowels of children and adults. *Journal of the Acoustical Society of America, 108,* 1856-1866.

Baddeley, A. D. (1986). *Working memory.* Oxford: Oxford University Press.

Baddeley, A. D. (1990). *Human memory: Theory and Practice,* (pp.70-73). London: Lawrence Erlbaum Associates.

Baddeley, A. D. (1996). The concept of working memory. In: S. E. Gathercole (Ed.), *Models of short-term memory,* (pp. 1-27). Hove, UK: Psychology Press.

Baddeley, A. D. (2003). Working memory: Looking back and looking forwards. *Nature Reviews Neuroscience 4,* 829-839.

Baddeley, A. D., and Salamé, P. (1986). The unattended speech effect: perception or memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 525-529.

Baddeley, A. D., and Hitch, G. J. (1974). Working Memory. In: G. Bower (Ed.) *The Psychology of Learning and Motivation,* (Vol. 8, pp. 47-90). New York: Academic Press.

Baddeley, A. D., and Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake and P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28-61). New York: Cambridge University Press.

Bailey, P. J., and Summerfield, Q. (1980). Information in Speech: observations in perception. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 536-563.

Banbury, S. and Berry, D. (1997). Habituation and dishabituation to office noise. *Journal of Experimental Psychology: Applied, 3,* 181-195.

Banbury, S. and Berry, D. (1998). Disruption of office-related tasks by speech and office noise. *British Journal of Psychology, 89,* 499-517.

Banbury, S. P., Jones, D. M., and Emery, L. (1999). Extending the "irrelevant sound effect": The effects of extraneous sound on aircrew performance. In D. Harris (Ed.), *Engineering psychology and cognitive ergonomics: Transportation systems, medical ergonomics, and training* (Vol.3, pp. 199-206). Aldershot, England: Ashgate and Town.

Banbury, S. P., Macken, W. J., Tremblay, S., and Jones, D. M. (2001). Auditory Distraction and short-term memory: Phenomena and Practical Implications. *Human Factors, 43,* 12-29.

Baum, S. R., Pell, M. D., Leonard, C. L., and Gordon, J. K. (1997). The ability of right and left hemisphere damaged individuals to produce and interpret prosodic cues marking phrasal boundaries. *Language and speech, 40,* 313-330.

Beaman, C.P. (2000). Computational explorations of the irrelevant sound effect in serial short-term memory. In: L. R. Gleitman and A. K. Joshi (Eds.), *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society*, (pp. 37-41). Mahwah, NJ: Erlbaum

Beaman, C.P. (2000). The irrelevant sound phenomenon revisited: what role for working memory capacity? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 1106-1118.

Beaman, C.P. (2004). The irrelevant sound phenomenon revisited: What role for working memory capacity? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 1106-1118.

Beaman, C.P., Bridges, A.M., and Scott, S.K. (2007). From dichotic listening to the irrelevant sound effect: A behavioural and neuroimaging analysis of the processing of unattended speech. *Cortex, 43,* 124-134.

Beaman, C. P. & Holt, N. J. (2007). Reverberant auditory environments: The effects of multiple echoes on distraction by 'irrelevant' speech. *Applied Cognitive Psychology*. 21, 1077-1090.

Beaman, C. P., and Jones, D. M. (1997). Role of serial order in the irrelevant speech effect: Tests of the changing state hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 459-471.

Beaman, C. P., and Jones, D. M. (1998). Irrelevant sound disrupts order information in free as in serial recall. *The Quarterly Journal of Experimental Psychology, 51A,* 615-636.

Belin, P., Zilbovicius, M., Crozier, S., Thivard, L., Fontaine, A., Masure, M. C., and Samson, Y. (1998). Lateralization of speech and auditory temporal processing. *Journal of Cognitive Neuroscience, 10,* 536-540.

Bever, T. (1980). Broca., and Lashley were right: Cerebral dominance is an accident of growth. In: J. Kaplan and C. Chomsky (Eds.), *Biology and Language*. Cambridge, MA: MIT Press.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Springer, J. A., Kaufman, J. N., and Possing, E. T. (2000). Human temporal lobe activation by speech and non-speech sounds. *Cerebral Cortex, 10,* 512-528.

Blesser, B. (1972). Speech perception under conditions of spectral transformation: I. Phonetic characteristics. *Journal of Speech and Hearing Research, 15,* 5-41.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound,* (pp. 143-545). Cambridge, MA: MIT Press.

Bridges, A.M., and Jones, D. M. (1996). Word dose in the disruption of serial recall by irrelevant speech: Phonological similarity or changing state? *Quarterly Journal of Experimental Psychology, 49A,* 919-939.

Broadbent, D.E., and Ladefoged, P. (1959). Auditory perception of temporal order. *Journal of the Acoustical Society of America, 31,* 1539.

Brown, G. D., Preece, T., Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review, 107,* 127-181.

Buchner, A. and Erdfelder, E. (2005). Word frequency of irrelevant speech distractors affects serial recall. *Memory and Cognition, 33,* 86-97.

Buchner, A., Irmen, L., and Erdfelder, E. (1996). On the irrelevance of semantic information for the 'irrelevant speech' effect. *Quarterly Journal of Experimental Psychology, 49A,* 765-779.

Buchner, A., Mehl, B., Rothermund, K., and Wentura, D. (2006). Artificially induced valence of distractor words increases the effects of irrelevant speech on serial recall. *Memory and Cognition, 34,* 1055-1062.

Buchner, A., Rothermund, K., Wentura, D., and Mehl, B. (2004). Valence of distractor words increases the effects of irrelevant speech on serial recall. *Memory and Cognition, 32,* 722-731.

Buchner, A., Steffens, M. C., Irmen, L., and Wender, K. F. (1998). Irrelevant auditory material effects counting. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24,* 48-67.

Burani, C., Vallar, G., and Bottini, G. (1991). Articulatory coding and phonological judgments on written words and pictures: The role of the phonological output buffer. *European Journal of Cognitive Psychology, 3,* 379-398.

Burgess, N., and Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review, 106,* 551-581.

Campbell, T., Beaman, C. P., and Berry, D. C. (2002). Auditory memory and the irrelevant sound effect: Further evidence for changing-state disruption. *Memory, 10,* 199-214.

Campbell, R., and Dodd, B. (1984). Aspects of hearing by eye. In: H. Bouma and D.G. Bounhuis (Eds.), *Attention and performance x,* (pp. 300-311. Hove: Lawrence Erlbaum Associates, Ltd.

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,* 975-979.

Clark, H. H. (1973). The language-as-fixed effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behaviour, 12,* 335-359.

Cole, R.A. (1973). Different memory functions for consonants and vowels. *Cognitive Psychology, 4,* 39-54.

Cole, R., Sales, B.D., and Haber, R.N. (1974). Mechanisms of aural encoding: VII. Differences in consonant and vowel recall in a Peterson and Peterson short-term memory paradigm. *Memory and Cognition, 2,* 211-214.

Cole, R.A. and Scott, B. (1973). Perception of temporal order in speech: The role of vowel transitions. *Canadian Journal of Psychology/Rev. Canad. Psychol., 27,* 441-449.

Colle, H. A. (1980). Auditory encoding in visual short-term recall: Effects of noise intensity and spatial location. *Journal of Verbal Learning and Verbal Behaviour, 19,* 722-735.

Colle, H. A., and Welsh, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning and Verbal Behavior, 15,* 17-31.

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

Cowan, N. (1995). Attention and memory. *An integrated framework.* Oxford: Oxford University Press.

Cowan, N. (1999). An embedded-processes model of working memory. In: A. Miyake and P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control,* (pp. 62-101). New York: Cambridge University Press.

Crowder, R.G., and Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception and Psychophysics, 5,* 365-373.

Divin, W., Coyle, K., and James, D. T. T. (2001). The effects of irrelevant speech and articulatory suppression on the serial recall of silently presented lipread digits. *British Journal of Psychology, 92,* 593-616.

Drullman, R., Festen, J. M., and Plomp, R. (1994). Effect of temporal envelope smearing on speech perception. *Journal of the Acoustical Society of America, 95,* 1053-1064.

Elliot, E. M. (2002). The irrelevant-speech effect and children: Theoretical implications of developmental change. *Memory and Cognition, 30,* 478-487.

Elliot, E. M., and Cowan, N. (2005). Coherence of the irrelevant sound effect: Individual profiles of short-term memory and susceptibility to task-irrelevant materials. *Memory and Cognition, 33,* 761-767.

Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors, 37,* 65-84.

Ellermeier, W., and Hellbrück, J. (1998). Is level irrelevant in 'irrelevant speech'? Effects of loudness, signal-to-noise ratio, and binaural masking. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1406-1414.

Ellermeier, W.,and Wolski, U. (1998). Effects of frequency-modulated tones on serial recall performance. In: W. Ellermeier, and J. Hellbrück. (1998). Is level irrelevant in 'irrelevant speech'? Effects of loudness, signal-to-noise ratio, and binaural masking. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1406-1414.

Ellermeier, W., and Zimmer, K. (1997). Individual differences in susceptibility to the 'irrelevant speech' effect. *Journal of the Acoustical Society of America, 102,* 2191-2199.

Frankish, C. (1996). Auditory short-term memory and the perception of speech. In: S. E. Gathercole (Ed.), *Models of Short-term memory,* (pp. 179-207). Hove, UK: Psychology Press.

Gisselgård, J., Petersson, K. M., Baddeley, A., and Ingvar, M. (2003). The irrelevant speech effect: a PET study. 1899-1911.

Gisselgård, J., Petersson, K. M., and Ingvar, M. (2004). The irrelevant speech effect and working memory load. *NeuroImage, 22,* 1107-1116.

Glenberg, A. M. and Swanson, N. C. (1986). A temporal distinctiveness theory of recency and modality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 3-15.

Gordon, J. W. (1987). The perceptual attack time of musical tones. *Journal of the Acoustical Society of America, 82,* 88-105.

Hadlington, L.J., Bridges, A. M., and Darby, R. J. (2004). Auditory Location in the irrelevant sound effect: The effects of presenting auditory stimuli to either the left ear, right ear or both ears. *Brain and Cognition, 55,* 545-557.

Hadlington, L.J., Bridges, A.M., and Beaman, C.P. (2006). A left-ear disadvantage for the presentation of irrelevant sound: Manipulations of task requirements and changing-state. *Brain and Cognition, 61,* 159-171.

Halberstam, B. and Raphael, L. J. (2004). Vowel normalization: The role of fundamental frequency and upper formants. *Journal of phonetics, 32,* 423-434.

Hellbrück, J., Kuwano, S. and Namba, S. (1995). Irrelevant background speech and human performance: Is there long-term habituation? *Journal of the Acoustical Society of Japan, 17,* 239-247.

Henson, R. N., Burgess, N., and Frith, C. D (2000). Recoding-storage , rehearsal and grouping in verbal short-term memory: An FMRI study. *Neuropsychologia, 38,* 426-440.

Henson, R., Hartley, T., Burgess, N., Hitch, G., and Flude, B. (2003). Selective interference with verbal short-term memory for serial order information: A new paradigm and tests of a timing-signal hypothesis. *The Quarterly Journal of Experimental Psychology, 56A,* 1307-1334.

Hickok, G. and Buchsbaum, B. (2001). Temporal lobe speech perception systems are part of the verbal working memory circuit: Evidence from two recent FMRI studies. *Behavioural and Brain Sciences, 26,* 740-741.

Hickok, G., and Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences, 4,* 131-138.

Higashikawa, M., Nakai, K., Sakakura, A., Takahashi, H. (1996). Perceived pitch of whispered vowels – relationship with formant frequencies: A preliminary study. *Journal of Voice, 10,* 155-158.

Hillenbrand, J. (1995).Identification of vowels resynthesized from /hVd/ utterances: Effects of formant contour. *Journal of the Acoustical Society of America, 97,* 3245(A).

Hirano, S., Naito, Y., Okazawa, H., Kojima, H., Honjo, I., Ishru, K., Yenokura, Y., Nagahama, Y., Fukuyama, H., and Konishi, J. (1997). Cortical activation by monaural speech sound stimulation demonstrated by positron emission tomography. *Experimental Brain Research, 113,* 75-80.

Hughes, R.W., and Jones, D.M. (2005). The impact of order incongruence between a task-irrelevant auditory sequence and a task-relevant visual sequence. *Journal of Experimental Psychology: Human perception and performance, 31,* 316-327.

Hughes, R. W., Tremblay, S., and Jones, D. M. (2005). Disruption by speech of serial short-term memory: The role of Changing-state Vowels. *Psychonomic Bulletin and Review, 12,* 886-890.

Irino, T., and Patterson, R. D. (1996). Temporal asymmetry in the auditory system. *Journal of the Acoustical Society of America, 99,* 2316-2331.

Ito, T., Takeda,K., and Itakra, F. (2005). *Analysis and recognition of whispered speech. Speech Communication, 45,* 139-152.

Jancke, L., Wustenberg, T., Schulze, K., and Heinze, H. J. (2002). Asymmetric hemodynamic responses of the human auditory cortex to monoaural and binaural stimulation. *Hearing research, 170,* 166-178.

Jones, D. M. (1993). Objects, streams and threads of auditory attention. In A. D. Baddeley and L. Weiskrantz (Eds.), *Attention: Selection, awareness and control* (pp. 167-198). Oxford: Clarendon Press.

Jones, D.M. (1994). Disruption of memory for lip read lists by irrelevant speech: Further support for the changing state hypothesis. *The Quarterly Journal of experimental psychology, 47 A,* 143-160.

Jones, D. M. (1999). The cognitive psychology of auditory distraction: The 1997 BPS Broadbent Lecture. *British Journal of Psychology, 90,* 167-187.

Jones, D. M., Alford, D., Bridges, A., Tremblay, S., and Macken, W. J. (1999a). Organisational factors in selective attention: The interplay of acoustic distinctiveness and auditory streaming in the irrelevant sound effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 464-473.

Jones, D. M., Alford, D., Macken, W. J., Banbury, S. P., and Tremblay, S. (2000). Interference from degraded auditory stimuli: Linear effects of changing-state in the irrelevant sequence. *Journal of the Acoustical society of America, 108,* 1082-1088.

Jones, D. M., Beaman, C. P., and Macken, W. J. (1996). The object-oriented episodic record model. In: S. Gathercole (Ed.), *Models of short-term memory,* (pp 209-238). London: Lawrence Erlbaum Associates.

Jones, D.M., Farrand, P., Stuart, G., and Morris, N. (1995). Functional equivalence of verbal and spatial information in serial short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1008-1018.

Jones, D. M., and Macken, W. J. (1993). Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 369-381.

Jones, D. M., and Macken, W. J. (1995a). Phonological similarity in the irrelevant speech effect: Within-or between-stream similarity? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 103-115.

Jones, D. M., and Macken, W. J. (1995b). Organizational factors in the effect of irrelevant speech: The role of spatial location and timing. *Memory and Cognition, 23,* 192-200.

Jones, D. M., and Macken, W. J. (1995c). Auditory babble and cognitive efficiency: The role of number of voices and their location. *Journal of Experimental Psychology: Applied, 1,* 216-226.

Jones, D. M., Macken, W. J., and Mosdell, N. (1997). The role of habituation in the disruption of recall performance by irrelevant sound. *British Journal of Psychology, 88,* 549-564.

Jones, D. M., Macken., W. J., and Murray, A. C. (1993). Disruption of visual short-term memory by changing state auditory stimuli: The role of segmentation. *Memory and Cognition, 21,* 318-328.

Jones, D. M., Madden, C., and Miles, C. (1992). Privileged access by irrelevant speech to short-term memory: The role of changing state. *The Quarterly Journal of Experimental Psychology, 44A,* 645-669.

Jones, D. M., Macken, W. J., and Nicholls, A. P. (2004). The phonological store of working memory: Is it phonological and is it a store? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 656-674.

Jones, D. M., Miles, C., and Page, J. (1990). Disruption of reading by irrelevant speech: Effects of attention, arousal, or memory? *Applied Cognitive Psychology, 4,* 89-108.

Jones, D. M., Saint-Aubin, J., and Tremblay, S. (1999b). Modulation of the irrelevant sound effect by organizational factors: Further evidence from streaming by location. *Quarterly Journal of Experimental Psychology, 52A,* 545-554.

Jones, D. M., and Tremblay, S. (2000). Interference in memory by process or content? A reply to Neath (2000). *Psychonomic Bulletin and Review, 7,* 550-558.

Jovičič, S.T., Dordevič, M. M. (1996). Acoustic features of whispered speech. *Acustica-acta acustica, 82,* S228.

Jovičič, S.T. (1998). Formant feature differences between whispered and voiced sustained vowels. *Acustica-acta acustica, 84,* 739-743.

Katz, W. F., and Assmann, P. F. (2001). Identification of children's and adult's vowels: Intrinsic fundamental frequency, fundamental frequency dynamics, and presence of voicing. *Journal of Phonetics, 29,* 23-51.

Kimuru, D. (1961a). Some effects of temporal-lobe damage on auditory perception. *Canadian Journal of Psychology, 15,* 156-165.

Kimuru, D. (1961b). Cerebral dominance and the perception of verbal stimuli. *Canadian Journal of Psychology, 15,* 166-171.

Klatte, M., Kilcher, H. and Hellbrück, J. (1995). The effects of temporal structure of background noise on working memory. *Zeitschrift für Experimental Psychologie, 42,* 517-544.

Kowalski, N., Depireux, D., and Shamma, S. (1996). Analysis of dynamic spectra in ferret primary auditory cortex: Characteristics of single unit responses to moving ripple spectra. *Journal of Neurophysiology, 76,* 3503-3523.

Lachs, L., & Pisoni, D. B. (2004). Cross-modal source information and spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 378-396.

Larsen, J. D., and Baddeley, A. D. (2003). Disruption of Verbal STM by irrelevant speech, articulatory suppression, and manual tapping: Do they have a common source? *The Quarterly Journal of Experimental Psychology, 56A,* 1249-1268.

Larsen, J. D., Baddeley, A.D., and Andrade, J. (2000). Phonological similarity and the irrelevant speech effect: Implications for models of short-term verbal memory. *Memory, 8,* 145-157.

LeCompte, D. C. (1994). Extending the irrelevant speech effect beyond serial recall. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20,* 1396-1408.

LeCompte, D. C. (1995). An irrelevant speech effect with repeated and continuous background speech. *Psychonomic Bulletin and Review, 2,* 391-397.

LeCompte, D. C. (1996). Irrelevant speech, serial rehearsal, and temporal distinctiveness: A new approach to the irrelevant speech effect. *Journal of Experimental Psychology: Learning, memory and cognition, 22,* 1154-1165.

LeCompte, D. C., Neely, C. B., and Wilson, J. R. (1997). Irrelevant speech and irrelevant tones: The relative importance of speech to the irrelevant speech effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 472-483.

LeCompte, D. C., and Shaibe, D. M. (1997). On the irrelevance of phonological similarity to the irrelevant speech effect. *Quarterly Journal of Experimental Psychology, 50A,* 100-118.

LeCompte, D. C., and Watkins, M. J. (1993). Similiarity as an organizing principle in short-term memory. *Memory, 1,* 3-11.

LeCompte, D. C., and Watkins, M. J. (1995). Grouping in Primary memory: The case of the compound suffix. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 96-102.

Levitin, D. J. (1999). Memory for musical attributes. In: P. R. Cook (Ed.) *Music, Cognition and Computerized Sound: An Introduction to Psychoacoustics,* (pp. 209-227). MIT Press, Cambridge, MA.

Liberman, A.M., Harris, K.S., Hoffman, H.S., and Griffith, B.C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54,* 358-368.

Liberman, A. M., and Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences, 4,* 187-196.

Lieberman, P., and Blumstein, S. E. (1988). *Cambridge studies in speech science and communication: Speech physiology, speech perception, and acoustic phonetics,* (pp 3-14). New York: Cambridge University Press.

Logie, R. H., Venneri, A., Della Sala, S., Redpath, T. W., and Marshall, I. (2003). Brain activation and the phonological loop: The impact of rehearsal. *Brain and Cognition, 53,* 293-296.

Luce, P.A., Feustel, T.C., and Pisoni, D.B. (1983). Capacity demands in short-term memory for synthetic and natural speech. *Human factors, 25,* 17-32.

Macken, W. J., and Jones, D. M. (1995). Functional Characteristics of the inner voice and the inner ear: Single or double agency? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 436-448.

Macken, W. J., Mosdell, N., and Jones, D. M. (1999). Explaining the irrelevant sound effect: Temporal distinctiveness or changing-state? *Journal of Experimental Psychology: Learning, Memory and Cognition, 25,* 810-814.

Macken, W. J., Tremblay, S., Houghton, R. H., Nichols, 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,* 43-51.

McElree, B., and Dosher, B.A. (1989). Serial position and set size in short-term memory: The time course of recognition. *Journal Experimental Psychology: General, 118,* 346-373.

McGurk, H., and McDonald, J. W. (1976). Hearing lips and seeing voices. *Nature, 264,* 746-748.

Meiser, T., and Klauer, K. C. (1999). Working emmeory and changing-state hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1272-1299.

Menon, V., Levitin, D. J., Smith, B. K., Lembke, A., Krasnow, B. D., Glazer, D., Glover, G. H., and McAdams, S. (2002). Neural correlates of timbre change in harmonic sounds. *NeuroImage, 17,* 1742-1754.

Miles, C., Jones, D. M., and Madden, C. A. (1991). Locus of the irrelevant speech effect in short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 578-584.

Mirman, D., Holt, L.L, and McClelland, J.L. (2004). Categorization and discrimination of nonspeech sounds: Differences between steady-state and rapidly-changing acoustic cues. *Journal of the Acoustical Society of America, 116,* 1198-1207.

Moore, B. C. J. (1989). *Introduction to the Psychology of Hearing,* (pp. 280-281) 3rd edition. London: Academic Press.

Moore, B. C. J. (2004). *An Introduction to the Psychology of Hearing.* 5th edition, (pp. 305-307). Elsevier academic press. London.

Moore, D. R. (2000). Auditory Neuroscience: Is speech special? *Current Biology, 10,* R362-R364.

Morris, R. W., and Clements, M. A. (2002). Reconstruction of speech from whispers. Medical Engineering and Physics, *24,* 515-520.

Morris, N. and Jones, D. M. (1990). Habituation to the irrelevant speech effect: Effects on a visual short-term memory task. *Perception and Psychophysics, 47,* 291-287.

Morton, J., Crowder, R.G., and Prussin, H. (1971). Experiments with the stimulus suffix effect. *Journal of Experimental Psychology, 91,* 169-190.

Nairne, J.S. (1988). A framework for interpreting recency effects in immediate serial recall. *Memory and Cognition, 16,* 343-352.

Nairne, J.S. (1990). A feature model of immediate memory. *Memory and Cognition, 18,* 251-269.

Nairne, J.S., and Crowder, R.G. (1982). On the locus of the stimulus suffix effect. *Memory and Cognition, 10,* 350-357.

Narain, C., Scott, S. K., Wise, R. J. S., Rosen, S., Leff, A., Iversen, S. D., and Matthews, P. M. (2003). Defining a left-lateralized response specific to intelligible speech using FMRI. *Cerebral Cortex, 13,* 1362-1368.

Neath, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin and Review, 7,* 403-423.

Neath, I., Farley, L. A., and Surprenant, A. (2003). Directly assessing the relationship between irrelevant speech and articulatory suppression. *Quarterly Journal of Experimental Psychology, 56A,* 1269-1278

Neath, I., Surprenant, A. M., and Crowder, R. G. (1993). The context-dependent stimulus suffix effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 698-703.

Neath, I., Surprenant, A. M., and LeCompte, D. C. (1998). Irrelevant speech eliminates the word length effect. *Memory and Cognition, 26,* 343-354.

Neely, C. B., and LeCompte, D. C. (1999). The importance of semantic similarity to the irrelevant speech effect. *Memory and Cognition, 27,* 37-44.

Neumann, O. (1987). Beyond Capacity: A functional view of attention: In: H. Hever and A. F. Sanders (eds.), *Perspectives on perception and action,* (pp. 361-393). Hillsdale, N. J: Erlbaum.

Oswald, C. J. P; Tremblay, S., and Jones, D. M. (2000). Disruption of comprehension by the meaning of irrelevant sound. *Memory, 8,* 345-350.

Page, M. P. A., and Norris, D. G. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105,* 761-781.

Paulesu, E., Frith, C. D., and Frackowiak, R. S. (1993). The neural correlates of the verbal component of working memory. *Nature, 362,* 342-345.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., and Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron, 36,* 767-776.

Penny, C.G. (1989). Modality effects and the structure of short-term verbal memory. *Memory and Cognition, 17,* 398-422.

Perham, N., Banbury, S., Jones, D. M. (in press). Do realistic reverberation levels reduce auditory distraction. *Applied Cognitive Psychology (DOI: 10.1002/acp. 1300, published online 29/09/2006).*

Pisoni, D. B. Auditory and Phonetic memory codes in the discrimination of consonants and vowels. Perception and Psychophysics, 13, 253-260.

Pisoni, D. B. (1975). Auditory short-term memory and vowel perception. *Memory and Cognition, 3,* 7-18.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time, *Speech Communication, 41,* 245-255.

Pratto, F., and John, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology, 61,* 380-391.

Rabbitt, P.M.A. (1991). Mild hearing loss can cause apparent memory failures which increase with age and reduce with IQ. *Acta Otolaryngology supplement, 476,* 167-176.

Remez, R.E., and Rubin, P.E (1990). On the perception of speech from time-varying acoustic information: Contributions of amplitude variations. *Perception and Psychophysics, 48,* 313-325.

Remez, R. E., Rubin, P. E., Pisoni, D. B. and Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science, 212,* 947-950.

Roman, N., and Wang, D. L. (2005). A pitch-based model for separation of reverberant speech. *Interspeech, September, 4-8, Lisbon, Portugal,* 2109-2112.

Rothermund, K., Wentura, D., and Bak, P. M. (2001). Automatic attention to stimuli sugnalling chances and dangersL Moderating effects of positive and negative goal and action contexts. *Cognition and Emotion, 15,* 231-248.

Salamé, P., and Baddeley, A. D. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior, 21,* 150-164.

Salamé, P., and Baddeley, A. D. (1987). Noise, unattended speech and short-term memory. *Ergonomics, 30,* 1185-1194.

Salamé, P., and Baddeley, A. D. (1989). Effects of background music on phonological short-term memory. *Quarterly Journal of Experimental Psychology, 41A,* 107-122.

Salamé, P., and Baddeley, A. D. (1990). The effects of irrelevant speech on immediate free recall. *Bulletin of the Psychonomic Society,* 28, 540-542.

Schouten, M. E., and Van Hessen, A. J. (1992). Modeling phoneme perception: I Categorical perception. *Journal of the Acoustical Society of America, 92,* 1841-1855.

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory and Cognition, 21,* 168-175.

Scott, B., and Cole, R. (1972). Auditory illusions as caused by embedded sounds. *Journal of the Acoustical Society of America, 51,* (1A), 112 (abstract).

Scott, S. K. and Wise, R. J. S. (2004). The functional neuroanatomy of prelexical processing in Speech perception. *Cognition, 92,* 13-45.

Scott, S. K., Blank, C. C., Rosen, S. and Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain, 123,* 2400-2406.

Scott, S. K, Rosen, S, Davis, J. and Beaman, C. P. and Wise, R. J. (submitted). The neural processing of unattended speech: dual mechanisms in the cocktail party effect? In: C.P. Beaman., A.M. Bridges., and S.K. Scott. (2007). From dichotic listening to the irrelevant sound effect: A behavioural and neuroimaging analysis of the processing of unattended speech. *Cortex, 43,* 124-134.

Scott, S. K., Rosen, S., Wickham, L., and Wise, R. J. (2004). A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. *Journal of Acoustical Society of America, 115,* 813-821.

Searleman, A. (1977). A review of right hemisphere linguistic capabilities. *Psychological Bulletin, 84,* 503-528.

Shankweiller, D., and Studdert-kennedy, M. (1967). Identification of consonants and vowels presented to left and right ears. *The Quarterly Journal of Experimental Psychology, 19,* 59-63.

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

Shannon, R. V., Zeng, F-G., & Wygonski, J. (1998). Speech recognition with altered spectral distribution of envelope cues. *Journal of the Acoustical Society of America, 104,* 2467-2476.

Smith, A. P., and Jones, D. M. (1992). Noise and Performance. In: D. M. Jones and A. P. Smith (Eds.), *Handbook of Human Performance,* (Vol.1, pp. 1-28). London: Academic Press.

Surprenant, A. M. (1999). The effect of Noise on memory for spoken syllables. *International Journal of Psychology, 34,* 328-333.

Surprenant, A. M., LeCompte, D. C., & Neath, I. (2000). Manipulations of irrelevant information: Suffix effects with articulatory suppression and irrelevant speech. *Quarterly Journal of Experimental Psychology, 53A,* 325-348.

Surprenant, A. M., and Neath, I. (1996). The relation between discriminability and memory for vowels, consonants, and silent-center syllables. *Memory and Cognition, 24,* 356-366.

Strange, W., Jenkins, J. J., Johnson, T. L. (1983). Dynamic specification of coarticulated vowels. *Journal of the Acoustical Society of America, 74,* 695-705.

Tartter, V. C. (1991). Identifiability of vowels and speakers from whispered syllables. *Percept. Psychophys, 49,* 365-372.

Tolan, G. A., and Tehan, G. (2002). Testing Feature Interaction: Between-stream irrelevant speech effects in immediate recall. *Journal of Memory and Language, 46,* 562-585.

Tremblay, S., and Jones, D. M. (1998). The role of habituation in the irrelevant sound effect: Evidence from the effects of token set size and rate of transition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 659-671.

Tremblay, S., and Jones, D. M. (1999). Change of intensity fails to produce an irrelevant sound effect: Implications of the representations of unattended sound. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1005-1015.

Tremblay, S., Macken, W. J., and Jones, D. M. (2001). The impact of broadband noise on serial memory: Changes in band-pass frequency increase disruption. *Memory, 9,* 323-331.

Tremblay, S., Nicholls, A. P., Alford, D. and Jones, D. M. (2000). The irrelevant sound effect: Does speech play a special role? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1750-1754.

Turner , M. L., and Engle, R. W. (1989). Is working memory capacity task-dependent. *Journal of Memory and Language, 28,* 127-154.

Turner, L. W., Souza, P. E., and Forget, L. N. (1995). Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners. *Journal of the Acoustical Society of America, 97,* 2568-2576.

Voyer, D., and Flight, J. I. (2001). Reliability and Magnitude of auditory laterality effects: The influence of attention. *Brain and Cognition, 46,* 397-413.

Warren, R. M. and Obusek, C. J. (1972). Identification of Temporal order within auditory sequences. *Perception and Psychophysics, 12,* 86-90.

Warren, R. M. and Obusek, C. J., Farmer, R. M., and Warren, R. P. (1969). Auditory sequence: Confusion of patterns other than speech or music. *Science, 164,* 586-587.

Wentura, D., Rothermund, K., and Bak, P. (2000). Automatic vigilance: The attention-grabbing power of approach-and avoidance-related social information. *Journal of Personality and Social Psychology. 78,* 1024-1037.

Wu, M., and Wang, D. L. (2006). A pitch-based method for the estimation of short reverberation time. *Acta Acustica United With Acustica, 92,* 337-339.

Yost, W. A. (2000). *Fundamentals of Hearing. An Introduction.* 4th edition, (pp. 207-225). London: Academic Press.

Zahorian, S. A., and Jagharghi, A. J. (1993). Spectral-shape features versus formants as acoustic correlates for vowels. *Journal of the Acoustical Society of America, 94,* 1966-1982.

Zatorre, R. J., and Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex, 11,* 946-953.

Zatorre, R. J., Evans, A. C., Meyer, E., and Gjedde, A. (1992). Lateralisation of phonetic and pitch discrimination in speech processing. *Science, 256,* 846-849.

APPENDIX 1: DIGIT LISTS (EXPERIMENTS 3, 5, PILOT 6 AND EXPERIMENT 6).

Speech condition 1

Trial							
1	2	7	4	3	6	5	1
2	7	3	1	6	4	2	5
3	4	1	5	7	2	6	3
4	6	4	3	5	7	1	2
5	3	5	7	2	1	4	6
6	5	2	6	1	3	7	4
7	1	6	2	4	5	3	7
8	2	1	5	3	4	7	6
9	3	6	7	1	5	2	4
10	5	4	2	6	7	3	1
11	4	2	6	5	3	1	7
12	1	7	4	2	6	5	3
13	6	5	3	7	1	4	2
14	7	3	1	4	2	6	5
15	3	5	7	2	4	1	6
16	2	3	5	7	6	4	1
17	4	6	1	3	5	2	7
18	7	2	4	5	1	6	3
19	1	4	3	6	2	7	5
20	6	7	2	1	3	5	4
21	5	1	6	4	7	3	2
22	1	4	3	5	2	6	7
23	5	1	7	3	6	2	4
24	3	2	4	1	7	5	6
25	7	6	2	4	1	3	5
26	4	7	6	2	5	1	3
27	2	3	5	6	4	7	1
28	6	5	1	7	3	4	2

Speech condition 2

Trial							
1	7	3	5	6	1	2	4
2	4	1	3	7	2	5	6
3	3	5	1	2	4	6	7
4	6	2	4	5	7	1	3
5	1	7	6	3	5	4	2
6	5	6	2	4	3	7	1
7	2	4	7	1	6	3	5
8	7	5	1	6	3	2	4
9	6	1	3	2	5	4	7
10	2	7	5	4	6	1	3
11	5	3	4	1	2	7	6
12	4	2	6	3	7	5	1
13	1	6	2	7	4	3	5
14	3	4	7	5	1	6	2
15	1	5	4	2	6	7	3
16	5	3	1	6	2	4	7
17	6	7	5	3	4	2	1
18	3	1	2	7	5	6	4
19	4	2	7	5	1	3	6
20	7	4	6	1	3	5	2
21	2	6	3	4	7	1	5
22	6	1	4	5	2	3	7
23	3	4	5	2	6	7	1
24	4	6	1	7	3	5	2
25	7	2	3	6	5	1	4
26	2	5	7	3	1	4	6
27	5	7	2	1	4	6	3
28	1	3	6	4	7	2	5

Silent condition

Trial							
1	2	5	4	1	3	6	7
2	1	2	7	3	5	4	6
3	4	6	3	5	1	7	2
4	7	1	6	2	4	3	5
5	5	4	1	7	6	2	3
6	6	3	2	4	7	5	1
7	3	7	5	6	2	1	4
8	5	4	2	7	1	3	6
9	1	7	4	6	3	5	2
10	7	3	1	2	5	6	4
11	6	5	3	1	2	4	7
12	3	6	7	5	4	2	1
13	2	1	5	4	6	7	3
14	4	2	6	3	7	1	5
15	2	3	5	1	4	6	7
16	7	2	4	3	6	1	5
17	3	5	7	6	1	4	2
18	1	4	3	5	7	2	6
19	5	1	6	2	3	7	4
20	6	7	2	4	5	3	1
21	4	6	1	7	2	5	3
22	3	4	7	2	5	6	1
23	6	1	5	7	2	4	3
24	1	6	4	5	3	7	2
25	4	5	2	3	7	1	6
26	5	7	3	6	1	2	4
27	2	3	6	1	4	5	7
28	7	2	1	4	6	3	5

APPENDIX 2: MEMORY TASK STANDARD INSTRUCTIONS

Welcome

This is a short experiment to test your memory for numbers. Numbers will appear on the screen, one after the other. A short series of numbers will appear. This is called a 'trial'.

While the numbers are appearing, sound will be played through your headphones.

After each trial try and recreate the number list on the response sheet.

Push 'SPACEBAR' to move onto the next trial.

There are 28 trials in each condition

Any questions? Good luck!

Push SPACEBAR to begin experiment.

APPENDIX 3: CONSENT FORM.

Bath Spa University
Department of Psychology

You are invited to participate in a study investigating immediate memory for short sequences of digits. If you decide to participate you will be asked to wear a set of headphones and learn a series of digits presented on screen. Each digit will be presented one after the other on screen and there will be 7 digits per sequence. Once all the digits in each sequence have appeared you will be prompted to recall these digits in the order they were presented in, in written form on a response sheet. This is known as a 'trial', there are 28 trials in each condition and there will be three conditions, during two of which you will hear sound presented over headphones. You will be asked to ignore any sound you hear as it is irrelevant to the memory task and you will not be tested on any aspect of the sounds you hear. The experiment will last approximately 30 minutes.

Any information or personal details gathered in the course of the study are confidential. No individual will be identified in any publication of results. Only the experimenter and the PhD supervisory team will have access to the data and your anonymity will be protected. If you decide to participate, you are free to withdraw your consent and to discontinue participation at any time without having to give a reason and without consequence. You are free to withdraw your data from any future analysis and/or publication.

Do you confirm you have/are:

(1) Normal or corrected to normal vision ☐

(2) Normal hearing ☐

(3) Native English speaker ☐

I,...have read and understand the information above and any questions I have asked have been answered to my satisfaction. I agree to participate in this research, knowing that I can withdraw at any time without consequence. I have been given a copy of this form to keep.

Participant's Name: ....
(Block letters)

Participant's Signature: ....Date...................

Investigator's Name: ....

Investigator's Signature: ....Date.................

The ethical aspects of this study have been approved by Bath Spa University's Research Ethics Committee. If you have any complaints or reservations about any ethical aspect of your participation in this research, you may contact the primary supervisor of this research Dr Nigel Holt (Tel: 01225 876111 email: n.holt@bathspa.ac.uk). Any complaint you make will be treated in confidence and investigated, and you will be informed of the outcome.

APPENDIX 4: NON-WORDS AND THEIR DISC FORMAT FOR PILOT A (FOR EXPERIMENT 1)

Non-word	Disc	Non-word	Disc
chung	JVN	shoob	Sub
fich	fIJ	tarb	t£b
chaf	J{f	thorg	T$g
muj	mV_	darb	d£b
gec	gEk	darj	d£_
hoch	hQJ	nop	nQp
jorb	_$b	jarm	_£m
pum	pVm	kuys	k2s
yem	jEm	meaz	miz
zas	z{s	dach	d{J
cheav	Jiv	jarv	_£v
sach	s{J	weath	wiD
lurb	l3b	rayf	r1f
lurj	l3_	veap	vip
toosh	tuS	bown	b6n
faz	f{z	nurb	n3b
feash	fiS	rarch	r£J
pab	p{b	baysh	b1S
gowch	g6J	poth	pQT
shuf	SVf	thayc	T1k
jarb	_£b	theaz	Tiz
vosh	v5S	yoom	j9m
yong	jQN	garl	g£l
losh	l5S	nairz	n8z
howt	h6t	zom	zQm

APPENDIX 5: DISC PHONETIC SYMBOLS

Disc symbols for English vowels

DISC	example
I	pip
E	vet
{	tat
V	putt
Q	tot
U	put
i	seam
£	barn
$	torn
u	spoon
3	turn
1	may
2	buy
4	toy
5	no
6	brow
7	peer
8	fair
9	poor

Note: /£/replaces #

Disc symbols for English consonants

DISC	example
p	pat
b	bad
t	tap
d	dad
k	cap
g	gain
N	fang
m	map
n	nap
l	lap
r	rat
f	fat
v	vap
T	thin
D	the
s	sat
z	zap
S	show
j	yank
h	had
w	why
J	cheat
_	jeep

APPENDIX 6: PILOT A (FOR EXPERIMENT 1A) STANDARD INSTRUCTIONS

- 'Respond' will appear on VDU after the presentation of each non-word. Please write down what you think you heard on the score sheet provided. Please make sure you write your answer in the space that corresponds to the trial.
- After 50 non-words have been presented there will be a 5min break. You may continue onto the next trial if you wish.
- The experiment will last for approximately 10mins.

APPENDIX 7: PILOT A (FOR EXPERIMENT 1A) AND EXPERIMENT 1A CONSENT FORM

BATH SPA UNIVERSITY COLLEGE
DEPARTMENT OF PSYCHOLOGY

PERCEPTUAL IDENTIFICATION TASK

You are invited to participate in a study investigating the perceptual identification of non-words. The study is designed to establish the intelligibility of a series of non-words. The study is being conducted by Marie Cahillane a PhD student who can be contacted by email : m.cahillane@bathspa.ac.uk. If you decide to participate, you will be asked to wear a set of headphones through which a list of 100 non-words will be presented. Each word will be presented once. After the last presentation of each non-word you will be asked to write down what you heard on a response sheet. The experiment will last approximately 20 mins.

Any information or personal details gathered in the course of the study are confidential. No individual will be identified in any publication of the results. Only the experimenter and the PhD supervisory team will have access to the data and your anonymity will be protected. If you decide to participate, you are free to withdraw your consent and to discontinue participation at any time without having to give a reason and without consequence. You are free to withdraw your data from any future analysis and/or publication.

Do you confirm you have/are:

(i) Normal or corrected to normal vision ☐

(ii) Normal hearing ☐

(iii) Native English speaker ☐

I,.................................... have read and understand the information above and any questions I have asked have been answered to my satisfaction. I agree to participate in this research, knowing that I can withdraw at any time without consequence. I have been given a copy of this form to keep.

Participant's Name: ....…
(block letters)

Participant's Signature: ....Date.........................

Investigator's Name: ....

Investigator's Signature: ....Date.........................

The ethical aspects of this study have been approved by the department of Psychology Ethics Review Committee. If you have any complaints or reservations about any ethical aspect of your participation in this research, you may contact the primary supervisor of this research Dr Nigel Holt (Tel: 01225 876111 email: n.holt@bathspa.ac.uk). Any complaint will be treated in confidence and investigated, and you will be informed of the outcome.

APPENDIX 8: INTELLIGIBILITY RANGE: NUMBER OF PARTICIPANTS CORRECTLY IDENTIFYING EACH OF THE 50 NON-WORDS DEGRADED AT 0.65 SNR AND 0.7 SNR FOR PILOT A (FOR EXPERIMENT 1A)

0.65 SNR	Disc Format	Range	0.7 SNR	Disc Format	Range
Fich	fIJ	22	Chaf	J{f	21
Chung	JVN	21	Fich	fIJ	21
Chaf	J{f	20	Chung	JVN	20
Muj	mV_	20	Yem	jEm	19
Gec	gEk	19	Cheav	Jiv	18
Hoch	hQJ	18	Hoch	hQJ	18
Jorb	_$b	18	Jorb	_$b	17
Pum	pVm	18	Lurb	l3b	17
Yem	jEm	17	Pum	pVm	17
Zas	z{s	17	Toosh	tuS	17
Cheav	Jiv	16	Gec	gEk	16
Sach	s{J	15	Lurj	l3_	16
Lurb	l3b	14	Muj	mV_	16
Lurj	l3_	14	Shoob	Sub	16
Toosh	tuS	14	Feash	fiS	15
Faz	f{z	13	Gowch	g6J	14
Feash	fiS	13	Sach	s{J	14
Pab	p{b	13	Tarb	t£b	14
Gowch	g6J	12	Yong	jQN	14
Shuf	SVf	12	Zas	z{s	14
Jarb	_£b	11	Faz	f{z	12
Vosh	v5S	11	Jarb	_£b	11
Yong	jQN	11	Losh	l5S	11
Losh	l5S	10	Nop	nQp	10
Shoob	Sub	10	Shuf	SVf	10
Tarb	t£b	10	Darj	d£_	9
Thorg	T$g	8	Jarv	_£v	9
Darb	d£b	7	Pab	p{b	9
Darj	d£_	7	Darb	d£b	8
Nop	nQp	7	Jarm	_£m	8
Jarm	_£m	6	Rayf	r1f	8
Kuys	k2s	5	Howt	h6t	7
Meaz	miz	5	Vosh	v5S	7
Dach	d{J	4	Dach	d{J	6
Jarv	_£v	4	Kuys	k2s	6
Weath	wiD	4	Thorg	T$g	6

Rayf	r1f	3	Zom	zQm	5
Veap	vip	3	Nurb	n3b	4
Bown	b6n	2	Meaz	miz	3
Nurb	n3b	2	Veap	vip	3
Rarch	r£J	2	Poth	pQT	2
Baysh	b1S	1	Weath	wiD	2
Howt	h6t	1	Baysh	b1S	1
Zom	zQm	1	Bown	b6n	1
Garl	g£l	0	Garl	g£l	1
Nairz	n8z	0	Thayc	T1k	1
Poth	pQT	0	Nairz	n8z	0
Thayc	T1k	0	Rarch	r£J	0
Theaz	Tiz	0	Theaz	Tiz	0
Yoom	j9m	0	Yoom	j9m	0
Total		461			494

APPENDIX 9: PILOT B (FOR EXPERIMENT 1B): NON-WORDS FOR THE CLEAR AND DEGRADED SPEECH CONDITION.

Clear speech condition

Non-word	Disc Format
Shoob	Sub
Gowch	g6J
Faz	f{z
Darj	d£_
Veap	vip
Muj	mV_
Thayc	T1k

Degraded speech condition

Non-word	Disc Format
Bown	b6n
Theaz	Tiz
Rayf	r1f
Nairz	n8z
Yoom	j9m
Garl	g£l
Kuys	k2s

APPENDIX 10: PILOT B (FOR EXPERIMENT 1B): ONE FACTOR REPEATED MEASURES ANOVA ON THREE LEVELS OF NON-WORD COMPONENTS (INITIAL CONSONANTS, VOWELS, AND FINAL CONSONANTS)

Tests of within-subjects effects

		Type III Sum of Squares	df	Mean Square	F	Sig.
Sound	Sphericity Assumed	35.565	2	17.783	18.153	0.000
Error (sound	Sphericity Assumed	43.101	44	0.980		

Bonferroni corrected pairwise comparisons for initial consonants, vowels and final consonants

(I) Sound	(J) Sound	Mean Difference (I-J)	Std. Error	Sig.[a]
Initial Consonant	Vowel	1.348*	0.285	0.000
Initial Consonant	Final Consonant	0.304	0.284	0.887
Final Consonant	Vowel	1.652*	0.305	0.000

* The mean difference is significant at the 0.05 level.
a = Adjustment for multiple comparisons: Bonferroni.

APPENDIX 11: DIGIT LISTS FOR PILOT B (FOR EXPERIMENT 1B

Silent condition

Trial							
1	6	2	7	5	8	4	1
2	9	5	8	6	2	7	1
3	6	2	8	5	7	1	4
4	4	6	9	2	3	1	8
5	5	1	3	7	4	8	6
6	2	6	1	4	8	3	5
7	3	6	1	8	2	5	7
8	6	8	4	7	1	5	9
9	4	6	2	7	3	9	1
10	1	9	2	5	7	3	8
11	1	3	9	4	7	5	2
12	7	5	9	1	3	8	4
13	3	7	2	9	5	1	8
14	1	6	5	9	3	8	2
15	4	2	6	8	1	3	7
16	5	3	7	9	2	6	1
17	7	4	6	2	8	2	9
18	5	1	3	7	9	2	4
19	4	1	6	8	5	2	9
20	9	4	1	5	7	2	8
21	8	4	1	9	6	3	7
22	3	8	6	9	7	4	1
23	2	9	4	7	5	1	8
24	8	3	9	5	7	1	4
25	7	4	8	6	3	9	1

Degraded speech condition

Trial							
1	5	3	8	6	9	2	1
2	9	7	1	3	8	2	6
3	5	1	9	7	2	4	8
4	2	6	3	5	9	7	1
5	9	7	4	2	6	8	1
6	7	5	1	8	4	6	2
7	8	6	2	4	9	7	3
8	5	3	6	2	4	9	7
9	1	3	9	2	7	5	8
10	5	1	4	9	6	2	8
11	8	5	7	2	9	6	1
12	2	6	4	1	7	3	8
13	6	1	9	5	3	7	2
14	1	7	2	6	3	9	5
15	9	6	2	4	1	3	8
16	5	4	8	6	2	7	1
17	5	2	4	7	1	8	3
18	9	7	2	4	8	3	6
19	4	8	6	3	1	9	2
20	3	5	1	7	9	2	8
21	6	8	1	9	2	4	7
22	9	4	2	8	1	7	5
23	1	8	6	3	9	2	7
24	5	2	6	8	1	4	9
25	6	9	1	4	2	7	3

Clear speech condition

Trial							
1	6	9	1	3	8	4	2
2	8	7	2	6	4	1	9
3	2	7	4	1	3	9	6
4	4	1	8	5	3	7	9
5	3	7	5	2	6	9	4
6	1	9	2	5	7	3	4
7	5	3	1	7	4	8	6
8	1	3	8	6	2	9	7
9	4	6	3	1	9	5	7
10	2	4	1	8	3	5	9
11	4	6	3	9	5	1	8
12	1	6	2	5	3	7	9
13	9	3	7	4	8	6	2
14	8	1	4	2	6	3	9
15	4	7	2	6	8	1	3
16	6	2	4	1	8	3	9
17	3	7	4	6	8	9	1
18	5	1	4	9	2	8	7
19	9	1	5	3	6	8	2
20	6	2	5	1	3	7	9
21	3	1	8	4	7	9	6
22	5	3	6	9	7	1	8
23	4	1	6	8	3	5	9
24	7	2	4	9	6	3	1
25	6	8	1	5	3	9	2

APPENDIX 12: STANDARD INSTRUCTIONS FOR PILOT B (FOR EXPERIMENT 1B)

- This is a short experiment to test your memory for numbers.
- Numbers will appear on the screen, one after the other. A short series of numbers will appear. This is called a 'Trial'.
- While the numbers are appearing sound will be played through your headphones.
- After each 'Trial' try and recreate the number list on the response sheet.
- There are 25 trials in each condition.
- Would you like to ask any questions? Good Luck!

APPENDIX 13: PILOT B (FOR EXPERIMENT 1B) CONSENT FORM

Bath Spa University
Department of Psychology

You are invited to participate in a study investigating immediate memory for short sequences of digits. If you decide to participate you will be asked to wear a set of headphones and learn a series of digits presented on screen. Each digit will be presented one after the other on screen and there will be 7 digits per sequence. Once all the digits in each sequence have appeared you will be prompted to recall these digits in the order they were presented in, in written form on a response sheet. This is known as a 'trial', there are 25 trials in each condition and there will be three conditions, during two of which you will hear sound presented over headphones. You will be asked to ignore any sound you hear as it is irrelevant to the memory task and you will not be tested on any aspect of the sounds you hear. The experiment will last approximately 30 minutes.

Any information or personal details gathered in the course of the study are confidential. No individual will be identified in any publication of results. Only the experimenter and the PhD supervisory team will have access to the data and your anonymity will be protected. If you decide to participate, you are free to withdraw your consent and to discontinue participation at any time without having to give a reason and without consequence. You are free to withdraw your data from any future analysis and/or publication.

Do you confirm you have/are:

(1) Normal or corrected to normal vision ☐

(2) Normal hearing ☐

(3) Native English speaker ☐

I,..have read and understand the information above and any questions I have asked have been answered to my satisfaction. I agree to participate in this research, knowing that I can withdraw at any time without consequence. I have been given a copy of this form to keep.

Participant's Name: ....
(Block letters)

Participant's Signature: ....Date...................

Investigator's Name: ....

Investigator's Signature: ....Date..................

The ethical aspects of this study have been approved by Bath Spa University's Research Ethics Committee. If you have any complaints or reservations about any ethical aspect of your participation in this research, you may contact the primary supervisor of this research Dr Nigel Holt (Tel: 01225 876111 email: n.holt@bathspa.ac.uk). Any complaint you make will be treated in confidence and investigated, and you will be informed of the outcome.

APPENDIX 14: PILOT B (FOR EXPERIMENT 1B): ONE FACTOR REPEATED MEASURES ANOVA WITH THREE LEVELS OF IRRELEVANT SOUND (SPEECH, DEGRADED SPEECH AND SILENCE)

Mauchly's test of sphericity

Within Subjects Effect	Mauchly' s W	Approx. Chi-Square	df	Sig.
Sound	0.628	7.482	2	0.024

Tests of within-subjects effects

		Type III Sum of Squares	df	Mean Square	F	Sig.
Sound	Sphericity Assumed	21153.444	2	10576.722	12.133	0.000
	Huynh-Feldt	21153.444	1.558	13581.485	12.133	0.000
Error (sound	Sphericity Assumed	29639.889	34			
	Huynh-Feldt	29639.889	26.478	1119.422		

Bonferroni corrected pairwise comparisons for speech, degraded speech and silence

(I) Sound	**(J) Sound**	**Mean Difference (I-J)**	**Std. Error**	**Sig.[a]**
Clear Speech	Degraded speech	31.722*	11.925	0.049
Clear Speech	Silence	47.611*	10.352	0.001
Silence	Degraded Speech	15.889	6.421	0.073

* The mean difference is significant at the 0.05 level.
a = Adjustment for multiple comparisons: Bonferroni

APPENDIX 15: NON-WORDS AND THEIR DISC FORMAT FOR EXPERIMENT 1A

Non-word	DISC
baysh	b1S
gel	gEl
nairz	n8z
rarch	r£J
rarl	r£l
thet	TEt
yoom	j9m
beath	biT
rorl	r$l
thayc	T1k
theaz	Tiz
thurt	T3t
vung	vVN
chuyz	J2z
goom	g9m
veen	v7n
fain	f8n
warch	w£J
shuf	SVf
weath	wiD
rayf	r1f
jarv	_£v
poth	pQT
thin	DIn
vod	vQd
zom	zQm
jarm	_£m
roth	r5T
tarch	t£J
veat	vit
garl	g£l
hayv	h1v
loj	l5_
meaz	miz
neash	niS
bown	b6n
heash	hiS
ning	nIN
seash	siS
coyd	k4d
mooz	muz

pum	pVm
toosh	tuS
vosh	v5S
warv	w£v
borch	b$J
cheav	Jiv
darb	d£b
darv	d£v
fam	f{m
jorn	_$n
kuys	k2s
tarb	t£b
feash	fiS
howt	h6t
nurb	n3b
veap	vip
caysh	k1S
dog	d5g
huyj	h2_
losh	l5S
lurj	l3_
cheen	J7n
lurb	l3b
marl	m£l
pov	pQv
vowt	v6t
zam	z{m
zas	z{s
duj	dV_
faz	f{z
jorb	_$b
sab	Z{b
shoob	Sub
dach	d{J
jarb	_£b
chung	JVN
durm	d3m
gowch	g6J
sach	s{J
cayb	k1b
chaf	J{f
cuyb	k2b
darj	d£_
nop	nQp

pab	p{b
pud	pVd
thorg	T$g
zog	zQg
hoch	hQJ
yong	jQN
fis	fIs
garr	g£R
gec	gEk
lich	lIJ
muj	mV_
pas	p{s
wij	wI_
yem	jEm
fich	fIJ

APPENDIX 16: EXPERIMENT 1A STANDARD INSTRUCTIONS

Welcome

You will be asked to listen to a series of sounds.

After each sound you'll be asked to write down what you think you heard.

Do you have any questions?

Please push any key to begin; the whole procedure will take about 20 minutes.

Good luck!

APPENDIX 17: INTELLIGIBILITY RANGE: NUMBER OF PARTICIPANTS CORRECTLY IDENTIFYING EACH OF THE 100 NON-WORDS DEGRADED AT 0.7 SNR FOR EXPERIMENT 1A

Non-words	Disc Format	Total Correct
baysh	b1S	0
gel	gEl	0
nairz	n8z	0
rarch	r£J	0
rarl	r£l	0
thet	TEt	0
yoom	j9m	0
beath	biT	1
rorl	r$l	1
thayc	T1k	1
theaz	Tiz	1
thurt	T3t	1
vung	vVN	1
chuyz	J2z	3
goom	g9m	3
veen	v7n	3
fain	f8n	4
warch	w£J	4
shuf	SVf	5
weath	wiD	5
rayf	r1f	6
jarv	_£v	7
poth	pQT	7
thin	DIn	7
vod	vQd	7
zom	zQm	7
jarm	_£m	9
roth	r5T	9
tarch	t£J	9
veat	vit	9
garl	g£l	10
hayv	h1v	10
loj	l5_	10
meaz	miz	10
neash	niS	10
bown	b6n	11
heash	hiS	11
ning	nIN	11

seash	siS	11
coyd	k4d	12
mooz	muZ	12
pum2	pVm	12
toosh	tuS	12
vosh	v5S	12
warv	w£v	12
borch	b$J	13
cheav	Jiv	13
darb	d£b	13
darv	d£v	13
fam	f{m	13
jorn	_$n	13
kuys	k2s	13
tarb	t£b	13
feash	fiS	14
howt	h6t	14
nurb	n3b	14
veap	vip	14
caysh	k1S	15
dog	d5g	15
huyj	h2_	15
losh	l5S	15
lurj	l3_	15
cheen	J7n	16
lurb	l3b	16
marl	m£l	16
pov	pQv	16
vowt	v6t	16
zam	z{m	16
zas	z{s	16
duj	dV_	17
faz	f{z	17
jorb	_$b	17
sab	Z{b	17
shoob	Sub	17
dach	d{J	18
jarb	_£b	18
chung	JVN	19
durm	d3m	19
gowch	g6J	19
sach	s{J	19
cayb	k1b	20
chaf	J{f	20

cuyb	k2b	20
darj	d£_	20
nop	nQp	20
pab	p{b	20
pud	pVd	20
thorg	T$g	20
zog	zQg	20
hoch	hQJ	21
yong	jQN	21
fis	fIs	22
garr	g£R	22
gec	gEk	22
lich	lIJ	22
muj	mV_	22
pas	p{s	22
wij	wI_	22
yem	jEm	22
fich	fIJ	23
TOTAL		**1241**

APPENDIX 18: NON-WORDS FOR EXPERIMENT 1B

Clear Speech Condition

Non-word	Disc Format
toosh	tuS
fis	fIs
jarm	_£m
beath	biT
cheen	J7n
coyd	k4d
hayv	h1v

Degraded Speech Condition

Non-word	Disc Format
nairz	n8z
thet	TEt
vung	vVN
warch	w£J
rorl	r$l
baysh	b1S
goom	g9m

APPENDIX 19: EXPERIMENT 1B: ONE FACTOR REPEATED MEASURES ANOVA ON THREE LEVELS OF NON-WORD COMPONENTS (INITIAL CONSONANTS, VOWELS, FINAL CONSONANTS)

Mauchly's test of sphericity

Within Subjects Effect	Mauchlys W	Approx. Chi-Square	df	Sig.
Sound	0.711	7.170	2	0.028

Tests of within-subjects effects

		Type III Sum of Squares	df	Mean Square	F	Sig.
Sound	Sphericity Assumed	77.565	2	38.783	33.393	0.000
	Huynh-Feldt	77.565	1.647	47.093	33.393	0.000
Error (sound	Sphericity Assumed	51.101	44	1.161		
	Huynh-Feldt	51.101	36.235	1.410		

Bonferroni corrected pairwise comparisons for initial consonants, vowels and final consonants

(I) Sound	(J) Sound	Mean Difference (I-J)	Std. Error	Sig.[a]
Initial Consonant	Vowel	2.000*	0.361	0.000
Initial Consonant	Final Consonant	2.435*	0.216	0.000
Final Consonant	Vowel	0.435	0.355	0.700

* The mean difference is significant at the 0.05 level.
a = Adjustment for multiple comparisons: Bonferroni.

APPENDIX 20: DIGIT LISTS FOR EXPERIMENT 1B

Clear speech condition

Trial							
1	3	2	4	1	7	5	6
2	2	1	5	3	4	7	6
3	3	5	7	2	1	4	6
4	1	4	3	6	2	7	5
5	6	7	2	1	3	5	4
6	5	1	7	3	6	2	4
7	7	3	1	6	4	2	5
8	4	6	1	3	5	2	7
9	3	5	7	2	4	1	6
10	4	1	5	7	2	6	3
11	1	6	2	4	5	3	7
12	6	5	1	7	3	4	2
13	5	2	6	1	3	7	4
14	3	6	7	1	5	2	4
15	6	4	3	5	7	1	2
16	4	2	6	5	3	1	7
17	4	7	6	2	5	1	3
18	7	3	1	4	2	6	5
19	1	7	4	2	6	5	3
20	5	1	6	4	7	3	2
21	2	3	5	6	4	7	1
22	2	3	5	7	6	4	1
23	7	2	4	5	1	6	3
24	6	5	3	7	1	4	2
25	1	4	3	5	2	6	7
26	7	6	2	4	1	3	5
27	5	4	2	6	7	3	1
28	2	7	4	3	6	5	1

Degraded speech condition

Trial							
1	6	7	5	3	4	2	1
2	1	7	6	3	5	4	2
3	5	6	2	4	3	7	1
4	7	5	1	2	4	6	3
5	3	4	5	2	6	7	1
6	4	2	7	5	1	3	6
7	7	3	5	6	1	2	4
8	6	1	3	4	7	2	5
9	2	5	7	3	1	4	6
10	1	5	4	2	6	7	3
11	1	6	2	3	5	7	4
12	3	1	2	7	5	6	4
13	5	7	2	1	4	6	3
14	2	7	5	1	3	4	6
15	7	2	3	6	5	1	4
16	1	3	6	4	7	2	5
17	4	6	1	7	3	5	2
18	5	3	1	6	2	4	7
19	4	2	6	5	1	3	7
20	7	4	6	1	3	5	2
21	2	4	7	1	6	3	5
22	5	3	4	7	6	1	2
23	6	1	4	5	2	3	7
24	6	2	4	5	7	1	3
25	3	5	1	2	4	6	7
26	2	6	3	4	7	1	5
27	3	4	7	6	2	5	1
28	4	1	3	7	2	5	6

Silent condition

Trial							
1	3	7	5	6	2	1	4
2	4	2	6	3	7	1	5
3	2	3	6	1	4	5	7
4	2	3	5	1	4	6	7
5	6	7	2	4	5	3	1
6	3	6	7	5	4	2	1
7	1	6	4	5	3	7	2
8	5	7	3	6	1	2	4
9	6	3	2	4	7	5	1
10	6	5	3	1	2	4	7
11	4	5	2	3	7	1	6
12	5	4	1	7	6	2	3
13	5	4	2	7	1	3	6
14	7	1	6	2	4	3	5
15	6	1	5	7	2	4	3
16	3	4	7	2	5	6	1
17	4	6	1	7	2	5	3
18	2	5	4	1	3	6	7
19	1	4	3	5	7	2	6
20	7	2	4	3	6	1	5
21	1	2	7	3	5	4	6
22	7	2	1	4	6	3	5
23	1	7	4	6	3	5	2
24	4	6	3	5	1	7	2
25	2	1	5	4	6	7	3
26	3	5	7	6	1	4	2
27	7	3	1	2	5	6	4
28	5	1	6	2	3	7	4

APPENDIX 21: EXPERIMENT 1B: ONE FACTOR REPEATED MEASURES ANOVA WITH 3 LEVELS OF IRRELEVANT SOUND (SPEECH, DEGRADED SPEECH AND SILENCE)

Tests of within-subjects effects

		Type III Sum of Squares	df	Mean Square	F	Sig.
Sound	Sphericity Assumed	6057.800	2	3028.900	26.518	0.000
Error (sound	Sphericity Assumed	6624.867	58	114.222		

Bonferroni corrected pairwise comparisons for speech, degraded speech and silence

(I) Sound	(J) Sound	Mean Difference (I-J)	Std. Error	Sig.[a]
Clear Speech	Degraded Speech	8.300*	2.288	0.003
Degraded Speech	Silence	11.700*	2.751	0.001
Silence	Clear Speech	20.000*	3.169	0.000

* The mean difference is significant at the 0.05 level.
a = Adjustment for multiple comparisons: Bonferroni

APPENDIX 22: EXPERIMENT 2 DIGIT LISTS

Block 1

Trial							
1	2	7	4	3	6	5	1
2	7	3	1	6	4	2	5
3	4	1	5	7	2	6	3
4	6	4	3	5	7	1	2
5	3	5	7	2	1	4	6
6	5	2	6	1	3	7	4
7	1	6	2	4	5	3	7
8	2	1	5	3	4	7	6
9	3	6	7	1	5	2	4
10	5	4	2	6	7	3	1
11	4	2	6	5	3	1	7
12	1	7	4	2	6	5	3
13	6	5	3	7	1	4	2
14	7	3	1	4	2	6	5
15	3	5	7	2	4	1	6
16	2	3	5	7	6	4	1
17	4	6	1	3	5	2	7
18	7	2	4	5	1	6	3
19	1	4	3	6	2	7	5
20	6	7	2	1	3	5	4
21	5	1	6	4	7	3	2
22	1	4	3	5	2	6	7
23	5	1	7	3	6	2	4
24	3	2	4	1	7	5	6
25	7	6	2	4	1	3	5
26	4	7	6	2	5	1	3
27	2	3	5	6	4	7	1
28	6	5	1	7	3	4	2
29	7	4	2	5	1	3	6
30	5	3	6	2	4	1	7

Block 2

Trial							
1	7	3	5	6	1	2	4
2	4	1	3	7	2	5	6
3	3	5	1	2	4	6	7
4	6	2	4	5	7	1	3
5	1	7	6	3	5	4	2
6	5	6	2	4	3	7	1
7	2	4	7	1	6	3	5
8	7	5	1	2	4	6	3
9	6	1	3	4	7	2	5
10	2	7	5	1	3	4	6
11	5	3	4	7	6	1	2
12	4	2	6	5	1	3	7
13	1	6	2	3	5	7	4
14	3	4	7	6	2	5	1
15	1	5	4	2	6	7	3
16	5	3	1	6	2	4	7
17	6	7	5	3	4	2	1
18	3	1	2	7	5	6	4
19	4	2	7	5	1	3	6
20	7	4	6	1	3	5	2
21	2	6	3	4	7	1	5
22	6	1	4	5	2	3	7
23	3	4	5	2	6	7	1
24	4	6	1	7	3	5	2
25	7	2	3	6	5	1	4
26	2	5	7	3	1	4	6
27	5	7	2	1	4	6	3
28	1	3	6	4	7	2	5
29	6	2	4	7	5	3	1
30	2	5	3	6	7	4	1

Block 3

Trial							
1	2	5	4	1	3	6	7
2	1	2	7	3	5	4	6
3	4	6	3	5	1	7	2
4	7	1	6	2	4	3	5
5	5	4	1	7	6	2	3
6	6	3	2	4	7	5	1
7	3	7	5	6	2	1	4
8	5	4	2	7	1	3	6
9	1	7	4	6	3	5	2
10	7	3	1	2	5	6	4
11	6	5	3	1	2	4	7
12	3	6	7	5	4	2	1
13	2	1	5	4	6	7	3
14	4	2	6	3	7	1	5
15	2	3	5	1	4	6	7
16	7	2	4	3	6	1	5
17	3	5	7	6	1	4	2
18	1	4	3	5	7	2	6
19	5	1	6	2	3	7	4
20	6	7	2	4	5	3	1
21	4	6	1	7	2	5	3
22	3	4	7	2	5	6	1
23	6	1	5	7	2	4	3
24	1	6	4	5	3	7	2
25	4	5	2	3	7	1	6
26	5	7	3	6	1	2	4
27	2	3	6	1	4	5	7
28	7	2	1	4	6	3	5
29	2	6	7	3	5	1	4
30	6	1	5	2	7	4	3

APPENDIX 23: EXPERIMENT 2 NON-WORDS

Vowel-only-changing speech condition

Non-word	Disc	Non-word	Disc	Non-word	Disc	Non-word	Disc	Non-word	Disc
chaym	J1m	chem	JEm	chim	JIm	chom	JQm	cham	J{m
theas	Tis	thees	T7s	this	TIs	thuys	T2s	thurs	T3s
marv	m£v	mav	m{v	mev	mEv	miv	mIv	meav	miv
burs	b3s	bees	b7s	bis	bIs	bars	b£s	bais	b8s
rarn	r£n	rorn	r$n	royn	r4n	rown	r6n	roon	r9n
forb	f$b	fayb	f1b	fuyb	f2b	furb	f3b	feb	fEb
gam	g{m	gim	gIm	gem	gEm	gom	gQm	garm	g£m
nart	n£t	nayt	n1t	nurt	n3t	noyt	n4t	nowt	n6t
cuz	kVz	caz	k{z	coz	k5z	coyz	k4z	cayz	k1z
laych	l1J	lowch	l6J	loch	lQJ	lich	lIJ	luch	lVJ
suyf	s2f	sif	sIf	sof	sQf	suf	suf	saf	s{f
zooc	zuk	zeac	zik	zic	zIk	zurc	z3k	zayc	z1k
toyd	t4d	towd	t6d	taid	t8d	tud	tVd	tid	tId
poyd	p4d	pod	p5d	powd	p6d	pud	pVd	ped	pEd
vorz	v$z	vayz	v1z	vurz	v3z	voyz	v4z	voz	v5z

Consonant-only-changing speech condition

Non-word	Disc	Non-word	Disc	Non-word	Disc	Non-word	Disc	Non-word	Disc
ching	JIN	thiz	DIz	shib	SIb	thil	TIl	sig	ZIg
jong	_QN	bot	bQt	dom	dQm	foz	fQz	hon	hQn
juyn	_2n	kuys	k2s	huyb	h2b	duyt	d2t	wuyv	w2v
gowch	g6J	hown	h6n	lowz	l6z	rows	r6s	thowt	D6t
noov	nuv	yoof	juf	gooz	guz	joob	_ub	boosh	buS
charc	J£k	sharf	S££	jarm	_£m	harz	h£z	garl	g£l
baysh	b1S	fayv	f1v	gayd	g1d	tayn	t1n	wayth	w1D
saj	s{_	tash	t{S	vap	v{p	zam	z{m	dac	d{k
goyd	g4d	boyf	b4f	hoyz	h4z	loyt	l4t	moyl	m4l
curj	k3_	murch	m3J	thurz	T3z	burl	b3l	gurc	g3k
job	_5b	shov	S5v	thod	T5d	fos	f5s	hoc	h5k
vorr	v$R	torp	t$p	lorc	l$k	horb	h$b	worsh	w$S
chung	JVN	shud	SVd	jum	_Vm	buv	bVv	vul	vVl
chev	JEv	fet	fEt	hez	hEz	zek	zEk	wep	wEp
teash	tiS	beath	biT	jead	_id	yeas	jis	neam	nim

APPENDIX 24: EXPERIMENT 2: 2 x 3 (TYPE AND LEVEL) REPEATED MEASURES ANOVA

		Type III Sum of Squares	df	Mean Square	F	Sig.	Observed Power[a]
Type	Sphericity Assumed	231.200	1	231.200	7.822	0.009	0.771
Error (Type)	Sphericity Assumed	857.133	29	29.556			
Level	Sphericity Assumed	48.344	2	24.172	0.917	0.405	0.201
Error (Level)	Sphericity Assumed	1528.989	58	26.362			
Type*Level	Sphericity Assumed	161.033	2	80.517	2.479	0.093	0.479
Error (Type*Level)	Sphericity Assumed	1883.633	58	32.476			

* Type = Vowel-Only-Changing (V-O-C) and Consonant-Only-Changing (C-O-C) Speech

* Level = Degradation level: clear (0% noise), 0.7 SNR (30% noise), 0.5 SNR (50% noise)

Bonferroni corrected pairwise comparisons for Type (V-O-C and C-O-C speech) and Level of degradation (clear, 0.7 and 0.5 SNR)

Type	(I) Level	(J) Level	Mean Difference	Std. Error	Sig.[a]
V-O-C	Clear	0.7 SNR	3.133	1.280	0.062
	Clear	0.5 SNR	2.433	1.520	0.361
	0.7 SNR	0.5 SNR	0.700	1.381	1.000
C-O-C	Clear	0.7 SNR	1.500	1.494	0.971
	Clear	0.5 SNR	6.667E-02	1.329	1.000
	0.7 SNR	0.5 SNR	1.567	1.383	0.799

* 0.7 SNR = 30% noise
* 0.5 SNR = 50% noise
*V-O-C = Vowel-Only-Changing
*C-O-C = Consonant-only-Changing

APPENDIX 25: EXPERIMENT 2: 2 x2 (TYPE AND LEVEL) REPEATED MEASURES ANOVA

		Type III Sum of Squares	df	Mean Square	F	Sig.
Type	Sphericity Assumed	243.675	1	243.675	10.205	0.003
Error (Type)	Sphericity Assumed	692.450	29	23.878		
Level	Sphericity Assumed	32.033	1	32.033	1.747	0.197
Error (Level)	Sphericity Assumed	531.842	29	18.339		
Type*Level	Sphericity Assumed	91.875	1	91.875	3.410	0.075
Error (Type*Level)	Sphericity Assumed	781.250	29	26.940		

* Type = Vowel-Only-Changing (V-O-C) and Consonant-Only-Changing (C-O-C) Speech

* Level = Degradation level: clear (0% noise), degraded: pooled effect of V-O-C and C-O-C speech degraded at 0.7 (30% noise) and 0.5 SNR (50% noise).

APPENIDX 26: EXPERIMENT 2: TESTS OF SIMPLE MAIN EFFECTS

Paired-Samples	Mean	Std. Deviation	Std. Error Mean	t	df	Sig. (2-tailed)
Clear and Degraded V-O-C speech	2.78	6.704	1.224	2.274	29	0.031
Clear and Degraded C-O-C speech	-0.72	6.754	1.233	-0.581	29	0.566

* V-O-C speech = Vowel-Only-Changing speech
* C-O-C speech = Consonant-Only-Changing speech

APPENDIX 27: EXPERIMENT 3: NON-WORDS FOR THE SPEECH AND WHISPERED SPEECH CONDITION

Non-word	Disc
curj	k3_
duyt	d2t
gam	g{m
rarn	r£n
sof	sQf
powd	p6d
worsh	w$S

APPENDIX 28: EXPERIMENT 3: ONE FACTOR REPEATED MEASURES ANOVA WITH 3 LEVELS OF IRRELEVANT SOUND (VOICED SPEECH, WHISPERED SPEECH AND SILENCE)

Tests of within subjects effects

		Type III Sum of Squares	df	Mean Square	F	Sig.
Sound	Sphericity Assumed	2407.222	2	1203.611	8.213	0.001
Error (sound	Sphericity Assumed	8500.111	58	146.554		

Bonferroni corrected pairwise comparisons for voiced speech, whispered speech and silence

(I) Sound	(J) Sound	Mean Difference (I-J)	Std. Error	Sig.[a]
Voiced Speech	Whispered Speech	2.000	3.274	1.000
Voiced Speech	Silence	9.833*	2.802	0.004
Silence	Whispered Speech	11.833*	3.277	0.003

* The mean difference is significant at the 0.05 level.
a = Adjustment for multiple comparisons: Bonferroni

APPENDIX 29: DIGIT LISTS FOR EXPERIMENTS 4 AND 7

List 1

Trial								
1	9	7	5	1	4	8	2	6
2	6	4	2	7	1	5	8	3
3	3	1	8	4	7	2	5	9
4	2	9	7	3	6	1	4	8
5	5	3	1	6	9	4	7	2
6	1	8	6	2	5	9	3	7
7	4	2	9	5	8	3	6	1
8	7	5	3	8	2	6	9	4
9	8	6	4	9	3	7	1	5
10	1	9	6	3	8	5	7	2
11	3	2	8	5	1	7	9	4
12	4	3	9	6	2	8	1	5
13	8	7	4	1	6	3	5	9
14	9	8	5	2	7	4	6	1
15	7	6	3	9	5	2	4	8
16	2	1	7	4	9	6	8	3
17	6	5	2	8	4	1	3	7
18	5	4	1	7	3	9	2	6
19	2	5	8	1	6	3	7	9
20	4	7	1	3	8	5	9	2
21	9	3	6	8	4	1	5	7
22	7	1	4	6	2	8	3	5
23	3	6	9	2	7	4	8	1
24	1	4	7	9	5	2	6	8
25	5	8	2	4	9	6	1	3
26	8	2	5	7	3	9	4	6
27	6	9	3	5	1	7	2	4

List 2

Trial								
1	3	6	7	1	9	2	4	8
2	9	3	4	7	6	8	1	5
3	4	7	2	5	1	6	8	3
4	1	4	8	2	7	3	5	9
5	8	2	6	9	5	7	3	4
6	2	5	9	3	8	1	6	7
7	5	8	3	6	2	4	9	1
8	7	1	5	8	4	9	2	6
9	6	9	1	4	3	5	7	2
10	1	8	4	9	6	7	2	3
11	5	3	8	4	1	2	6	7
12	8	6	2	7	4	5	9	1
13	3	1	6	2	8	9	4	5
14	7	5	1	6	3	4	8	9
15	9	7	3	8	5	6	1	2
16	2	9	5	1	7	8	3	4
17	4	2	7	3	9	1	5	6
18	6	4	9	5	2	3	7	8
19	1	7	2	6	8	5	9	3
20	8	5	9	4	6	3	7	1
21	4	1	5	9	2	8	3	6
22	2	8	3	7	9	6	1	4
23	6	3	7	2	4	1	5	8
24	5	2	6	1	3	9	4	7
25	9	6	1	5	7	4	8	2
26	3	9	4	8	1	7	2	5
27	7	4	8	3	5	2	6	9

List 3

Trial								
1	1	7	2	5	9	3	8	6
2	7	4	8	2	6	9	5	3
3	3	9	4	7	2	5	1	8
4	2	8	3	6	1	4	9	7
5	6	3	7	1	5	8	4	2
6	4	1	5	8	3	6	2	9
7	5	2	6	9	4	7	3	1
8	8	5	9	3	7	1	6	4
9	9	6	1	4	8	2	7	5
10	7	3	5	9	2	6	1	4
11	5	7	3	4	9	1	8	2
12	8	1	6	7	3	4	2	5
13	3	5	7	2	4	8	6	9
14	2	4	9	1	6	7	5	8
15	6	8	1	5	7	2	9	3
16	9	2	4	8	1	5	3	6
17	4	9	2	6	8	3	7	1
18	1	6	8	3	5	9	4	7
19	1	9	4	8	7	2	5	3
20	7	6	1	5	4	8	2	9
21	9	5	3	4	6	7	1	8
22	4	3	7	2	1	5	8	6
23	3	8	6	7	9	1	4	2
24	8	4	2	6	5	9	3	7
25	2	7	5	9	8	3	6	1
26	6	2	9	1	3	4	7	5
27	5	1	8	3	2	6	9	4

APPENDIX 30: EXPERIMENT 4: ONE FACTOR REPEATED MEASURES ANOVA WITH THREE LEVELS OF IRRELEVANT SOUND (VOICED SPEECH, ALTERNATED SPEECH AND SILENCE)

Tests of within-subjects effects

		Type III Sum of Squares	df	Mean Squares	F	Sig.
Sound	Sphericity Assumed	3600.750	2	1800.375	8.122	0.001
Error (sound)	Sphericity Assumed	10196.583	46	221.665		

Bonferroni corrected pairwise comparisons for voiced speech, alternated speech and silence

(I) Sound	(J) Sound	Mean Difference (I-J)	Std. Error	Sig.[a]
Voiced Speech	Alternated Speech	0.250	4.071	1.000
Alternated Speech	Silence	15.125*	4.447	0.007
Silence	Voiced Speech	14.875*	4.367	0.007

* The mean difference is significant at the 0.05 level.
a = Adjustment for multiple comparisons: Bonferroni
Note: Alternated speech = voiced and whispered speech alternated in an irrelevant auditory sequence

APPENDIX 31: EXPERIMENT 5: FINE STRUCTURE REVERSED NON-WORDS FOR PILOT LISTENING TEST

Non-word	Disc
curj	k3_
duyt	d2t
gam	g{m
powd	p6d
rarn	r£n
sof	sQf
worsh	w$S

APPENDIX 32: EXPERIMENT 5: ONE FACTOR REPEATED MEASURES ANOVA WITH THREE LEVELS OF IRRELEVANT SOUND (WHISPERED SPEECH, FINE STRUCTURE REVERSED (FSR) WHISPERED SPEECH AND SILENCE)

Mauchly's Test of Sphericity

Within Subjects Effect	Mauchly' s W	Approx. Chi-Square	df	Sig.
Sound	0.645	12.266	2	0.002

Tests of Within Subjects Effects

		Type III Sum of Squares	df	Mean Square	F	Sig.
Sound	Sphericity Assumed	4825.756	2	2412.878	18.911	0.000
	Huynh-Feldt	4825.756	1.536	3140.766	18.911	0.000
Error (sound)	Sphericity Assumed	7400.244	58	127.590		
	Huynh-Feldt	7400.244	44.558	166.080		

Bonferroni corrected pairwise comparisons for whispered speech, reversed whispered speech and silence

(I) Sound	(J) Sound	Mean Difference (I-J)	Std. Error	Sig.[a]
Whispered Speech	FSR Whispered Speech	0.833	2.034	1.000
FSR Whispered Speech	Silence	15.933*	3.593	0.000
Silence	Whispered Speech	15.100*	2.911	0.000

* The mean difference is significant at the 0.05 level.
a = Adjustment for multiple comparisons: Bonferroni

APPENDIX 33: PILOT EXPERIMENT 6: ONE FACTOR REPEATED MEASURES ANOVA WITH THREE LEVELS OF IRRELEVANT SOUND (SPEECH, SPECTRALLY ROTATED SPEECH AND SILENCE)

Tests of within-subjects effects

		Type III Sum of Squares	df	Mean Squares	F	Sig.
Sound	Sphericity Assumed	2852.028	2	1426.014	12.193	0.000
Error (sound)	Sphericity Assumed	5379.972	46	116.956		

Bonferroni corrected pairwise comparisons for speech, spectrally rotated speech and silence

(I) Sound	(J) Sound	Mean Difference (I-J)	Std. Error	Sig.[a]
Speech	Spectrally Rotated speech	6.708	3.082	0.120
Spectrally Rotated speech	Silence	8.667*	3.313	0.046
Silence	Speech	15.375*	2.960	0.000

* The mean difference is significant at the 0.05 level.
a = Adjustment for multiple comparisons: Bonferroni.

APPENDIX 34: EXPERIMENT 6: NON-WORDS AND THEIR DISC FORMAT FOR PILOT LISTENING TEST

Non-word	Disc
borch	b$J
feash	fiS
howt	h6t
marv	m£v
sheav	Siv
zol	zQl
forb	f$b
forch	f$J
lowch	l6J
mowz	m6z
teash	tiS

APPENDIX 35: EXPERIMENT 6: ONE FACTOR REPEATED MEASURES ANOVA WITH THREE LEVELS OF IRRELEVANT SOUND (SPEECH, SPECTRALLY ROTATED SPEECH AND SILENCE)

Tests of within-subjects effects

		Type III Sum of Squares	df	Mean Square	F	Sig.
Sound	Sphericity Assumed	1243.400	2	621.700	4.190	0.020
Error (sound)	Sphericity Assumed	8605.267	58	148.367		

Bonferroni corrected pairwise comparisons for speech, spectrally rotated speech and silence

(I) Sound	(J) Sound	Mean Difference (I-J)	Std. Error	Sig.[a]
Speech	Spectrally rotated speech	4.800	3.092	0.394
Spectrally rotated speech	Silence	4.300	3.215	0.574
Silence	Speech	9.100*	3.127	0.021

* The mean difference is significant at the 0.05 level.
a = Adjustment for multiple comparisons: Bonferroni

APPENDIX 36: FORMULA FOR STANDARDISATION OF DATA FOR THE SPEECH AND SPECTRALLY ROTATED SPEECH CONDITION.

Formula for standardising number of errors made per participant in the speech condition:

$$100 \times \frac{C1 - C3}{(C3 + 1)}$$

* C1 = speech condition. C3 = silent control.

Formula for standardising number of errors made per participant in the spectrally rotated speech condition:

$$100 \times \frac{C2 - C3}{(C3 + 1)}$$

* C2 = spectrally rotated speech condition. C3 = Silent control.

APPENDIX 37: EXPERIMENT 6: PAIRED SAMPLES T-TEST BETWEEN SPEECH AND SPECTRALLY ROTATED SPEECH

Paired-Samples	Mean difference (between pairs of values)	Std. Deviation	Std. Error Mean	t	df	Sig. (2-tailed)
Speech - spectrally rotated speech	61.87	137.522	25.108	2.464	29	0.020

Paired Samples Statistics

	Mean	N	Std. Deviation	Std. Error Mean
Speech	125	30	241	44
Spectrally rotated speech	62.63	30	156.166	28.512

APPENDIX 38: EXPERIMENT 7: ONE FACTOR REPEATED MEASURES ANOVA WITH THREE LEVELS OF IRRELEVANT SOUND (SPEECH, SPECTRALLY ROTATED SPEECH AND SILENCE)

Tests of within-subjects effects

		Type III Sum of Squares	df	Mean Squares	F	Sig.
Sound	Sphericity Assumed	6507.583	2	3253.792	11.754	0.000
Error (sound)	Sphericity Assumed	12733.750	46	276.821		

Bonferroni corrected pairwise comparisons for speech, spectrally rotated speech and silence

(I) Sound	(J) Sound	Mean Difference (I-J)	Std. Error	Sig.[a]
Speech	Spectrally Rotated speech	15.958*	4.764	0.008
Spectrally Rotated speech	Silence	6.708	4.725	0.507
Silence	Speech	22.667*	4.918	0.000

* The mean difference is significant at the 0.05 level.
a = Adjustment for multiple comparisons: Bonferroni

APPENDIX 39: TESTS OF WITHIN-SUBJECTS CONTRASTS FOR THE THREE IRRELEVANT CONDITIONS (SPEECH, SPECTRALLY ROTATED SPEECH AND SILENCE)

Source	Sound	Type III Sum of Squares	df	Mean Square	F	Sig.
Sound	Speech and Silence	12330.667	1	12330.667	21.245	0.000
	Spectrally Rotated Speech and Silence	1080.042	1	1080.042	2.016	0.169
	Speech and Spectrally Rotated Speech	6112.042	1	6112.042	11.220	0.003
Error (sound)	Speech and Silence	13349.333	23	580.406		
	Spectrally Rotated Speech and Silence	12322.958	23	535.781		
	Speech and Spectrally Rotated Speech	12528.958	23	544.737		

FORMULA FOR EFFECT SIZE: *r*

$$r = \sqrt{\frac{F\,(1, dfR)}{F\,(1, dfR) + dfR}}$$