CARDIFF UNIVERSITY

The nature of the representations underlying verbal behaviour: The interaction between auditory, visual and motor modalities.

David William Maidment

BSc (Hons), Applied Psychology

MSc, Social Science Research Methods

A thesis submitted for the degree of

Doctor of Philosophy

School of Psychology

Cardiff University

~ 2013 ~

I dedicate this thesis to my wife and sons

Lydia, Jude, Jared & Jesse

DECLARATION

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed ________________________________ (candidate) Date ________________________

STATEMENT 1

This thesis is being submitted in partial fulfilment of the requirements for the degree of PhD.

Signed ________________________________ (candidate) Date ________________________

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references. The views expressed are my own.

Signed ________________________________ (candidate) Date ________________________

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ________________________________ (candidate) Date ________________________

STATEMENT 4: PREVIOUSLY APPROVED BAR ON ACCESS

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loans ***after expiry of a bar on access previously approved by the Academic Standards & Quality Committee.***

Signed ________________________________ (candidate) Date ________________________

ACKNOWLEDGEMENTS

First, and foremost, I would like to thank my supervisor Dr. Bill Macken for his constant support and guidance over the past three years, which have been invaluable in completing this thesis. Further gratitude belongs to Prof. Dylan Jones, for his wealth of knowledge and advice that have also benefited me greatly along the way.

Thank you also for the financial support provided by the Economic and Social Research Council, without which this work would not have been possible.

I wish to thank my wife, Lydia. Without her love, dedication, and patience, I would have probably given up a long time ago. Even when I lost faith in myself, she kept me going. This thesis would almost certainly not be here today if it wasn't for her. Thank you for always believing in me.

Finally, a huge thank you belongs to all my friends and family for everything they have provided during the completion of this work. Special thanks belong to my mother- and father-in law, Catherine and Colin, for all the help and support they have given my family and me during the early stages of my academic career. A huge thank you also belongs to my sister-in law, Hannah who has read this thesis from cover-to-cover. As a result, she is probably worthy of a PhD herself!

THESIS SUMMARY

A fundamental issue in the study of verbal behaviour is whether the underpinning representation of speech, while derived from different modalities, is itself amodal. The current thesis contributes to this debate, utilising two behavioural phenomena to show that verbal performance is not simply limited to representations independent of the modality through which they were derived.

Firstly, similarities in verbal short-term memory performance across presentation modalities have been explained in terms of a phonological level of representation. Namely, both auditory and visual modes of presentation demonstrate similar patterns of performance within the recency portion of the serial position curve. However, it is shown that while recall at the terminal list item for an auditory list is immune to the disruptive effect of task-irrelevant background sound and articulatory suppression, lipread recency is not immune. In addition, although the effect of an auditory suffix on an auditory list is due to the perceptual grouping of the suffix with the list, the corresponding effect with lipread speech is shown to be due to misidentification of the lexical content of the lipread suffix. Furthermore, even though a lipread suffix does not disrupt auditory recency, an auditory suffix does disrupt recency for lipread lists due to attentional capture. These findings are subsequently explained in terms of modality-specific perceptual and motor-speech output mechanisms, rather than to the storage and manipulation at some phonological level of representation.

Secondly, the mechanisms underlying the integration of seen and heard speech is investigated via the McGurk effect in order to understand the stage at which auditory and visual modes of speech come to be integrated. It is shown that concurrently articulating verbal material out loud or silently mouthing speech during syllable identification reduces the McGurk effect, whereas passive listening to task-irrelevant speech or sequential manual tapping does not. On the basis that both concurrent articulation and silent mouthing impede subvocal speech production processes, that both manipulations also disrupt the McGurk effect suggests that subvocal motor mechanisms necessary for speech production are involved in audiovisual integration.

Taken together, if progress is to be made in understanding the underlying representations of verbal behaviour, an approach should be adopted that not only requires an amodal, phonological representational form, but also considers the extent to which modality-specific systems primarily serving perceptual and motor processes contribute to performance.

PUBLICATIONS

Maidment, D. W., Macken, W. J., & Jones D. M. (2013). Modalities of memory: Is reading lips like hearing voices? *Cognition*, *129*, 471-493.

Maidment, D. W., & Macken, W. J. (2012). The ineluctable modality of the audible: Perceptual determinants of auditory verbal short-term memory. *Journal of Experimental Psychology: Human Perception & Performance, 8*(4), 989-997*.*

ORAL PRESENTATIONS

Maidment, D. W., Macken, W. J., & Jones, D. M. (2012). Binding (and unbinding) the contents of working memory. Experimental Psychological Society's Annual Meeting, Bristol.

Maidment, D. W. (2011). What the inner voice tells the inner ear: Engaging the motor system during audio-visual speech perception. Experimental Psychological Society's Annual Meeting, Nottingham.

Maidment, D. W., & Macken, W. J. (2011). Is reading lips like hearing voices? The role of modality in short-term memory performance. The 5th Annual International Conference of Memory, York.

Maidment, D. W. (2010). Audiovisual speech perception: Evidence for sensory-motor integration? British Psychological Society Cognitive Section's 27th Annual Conference, Cardiff.

POSTERS

Maidment, D. W. & Macken, W. J. (2010). Reading lips and hearing voices: Does modality matter? Spanish Experimental Psychological Society's Annual Conference, Spain: Granada.

TABLE OF CONTENTS

CHAPTER 1

CHAPTER 2

INDEX OF FIGURES

CHAPTER 1

General Introduction & Thesis Overview

Our experience as talkers and listeners means that linguistic information can be represented in multiple forms through different senses. As a listener we associate the acoustic patterns of speech with accompanying visual information, such as that derived from mouth movements and facial expressions (e.g., Belin, Bestelmeyer, Latinus, & Watson, 2011; Bishop & Miller, 2009; Summerfield, 1992; Summerfield, MacLeod, McGrath, & Brooke, 1989). When talking out loud, the sound of a speaker's own voice is also correlated with motor actions generated during verbal production (e.g., Hickok, Buchsbaum, Humphries, & Muftuler, 2003; Sato, Buccino, Gentilucci, & Cattaneo, 2010; Skipper, Nusbaum, & Small, 2005). However, a fundamental issue in the study of verbal behaviour is whether the underpinning representation of speech, while derived from different modalities, is exclusively amodal.

Concerns about forms of representation and constraints on their processing hold a central and original place in the very foundations of Cognitive Psychology. The founding assumption of the specifically cognitive approach to language is that verbal performance is underpinned by a level of representation that transcends the physical mode of presentation (e.g., Chomsky, 1959; Chomsky, 2002; Chomsky & Halle, 1968). As such, the basic elements of verbal behaviour are amodal representations, which can be analysed in a modality-independent fashion, to account for the mappings between sounds and their corresponding meaning. Such an approach is still prevalent within accounts of verbal short-term memory, that posit as their basic currency phonological representations processed within bespoke storage systems. These representations occupy a functional and structural status distinct from perceptual input systems and, in turn, supply output systems. Distinctions between heard, read and silently lipread speech, for example, are attributed to mechanisms such as different access routes to phonological representation from heard and seen inputs (e.g., Baddeley, 1992, 2010, 2012; Repovs & Baddeley, 2006), or modality specific features supplementary to the phonological form (e.g., Nairne, 1990; Neath & Nairne, 1995; Penney, 1989; Winkler, Denham, & Nelken, 2009), or different attentional/encoding constraints across modalities (e.g., Burgess & Hitch, 1999; Page & Norris, 1998).

Even so, it should be acknowledged that any language-based cognitive task would generate multiple representations, all of which may or may not contribute to performance. While there may be many levels of representation, including those that do not depend on the derived source of information, such as semantic representations, verbal behaviour is initially derived from perceptual processes that generate sensory representations that are, by definition, modality-specific. As a consequence, our understanding of verbal performance should also consider the extent to which these modality-specific representations are also necessary for accounts of verbal behaviour (e.g., Cheng, 1974; Hickok, Holt, & Lotto, 2009; Jonides et al., 2008; Postle, 2006; Wilson, 2001). For example, within the short-term memory literature there it has been counter argued that the limits to verbal performance do not just arise from constraints associated with representations independent of the modality through which they were derived, or from the motor planning processes through which they may be maintained and output. Rather, verbal performance is additionally dependent on general perceptual, motor, and/or perceptual-motor mapping processes to meet the demands of the particular task (e.g., Hickok, 2009; Hughes, Marsh, & Jones, 2009; Macken, Phelps, & Jones, 2009; Wilson & Fox, 2007).

The current thesis therefore attempts to address the extent to which verbal behaviour can be attributed to the interplay between perceptual and motor processes, as opposed to an account that only utilises an amodal level of representation. As such, two behavioural phenomena are investigated to provide novel evidence that verbal behaviour is constrained by modality-specific, and not phonological, representations.

Firstly, in Chapter 2 the serial recall paradigm is revisited given that functional similarities in verbal short-term memory performance across presentation modalities have been explained in terms of a phonological level of representation. Specifically, the serial recall of auditory and lipread lists shows that, under certain circumstances, recall is enhanced towards the end of the sequence compared to written material – the so-called recency effect (Crowder & Morton, 1969). That lipread recency effects are disrupted by the presentation of a redundant end-of-list lipread or auditory suffix item has been taken to further indicate a common form of representation is shared by the two modalities. However, data across seven experiments reveals that lipread recency is actually underpinned by different mechanisms to auditory recency: auditory recency is immune to the disruptive effect of task-irrelevant, background sound and articulatory suppression – manipulations that impede the speech rehearsal process – whereas lipread recency is not immune.

Furthermore, interactions between lipread and auditory suffix effects on lipread lists are additionally driven by fundamentally different mechanisms. Although the effect of an auditory suffix on an auditory list is due to the perceptual grouping of the suffix with the list, the corresponding effect with lipread speech is due to misidentification of the lexical content of the lipread suffix. In addition, even though a lipread suffix does not disrupt auditory recency, an auditory suffix does disrupt recency for lipread lists. Critically, this effect is due to attentional capture ensuing from the presentation of an unexpected auditory event, and is evident both with verbal and nonverbal auditory suffixes. These findings subsequently add to a growing body of behavioural and neuroscientific evidence showing that, rather than being attributed to the storage and manipulation of phonological representations, verbal short-term memory performance is also constrained by modality-specific perceptual and motor-speech output mechanisms.

The processes underlying the binding between seen and heard speech is then examined in Chapter 3, in order to understand the stage at which auditory and visual modes of speech are integrated. In particular, audiovisual integration is investigated via the McGurk effect, which is demonstrated by seeing the syllable /ga/ mouthed at the same time as hearing /ba/ spoken, the participant will often report the percept /da/. The extent to which individuals are susceptible to this perceptual illusion subsequently provides a means of measuring the degree to which both modalities are bound perceptually. The critical test, however, is that audiovisual integration is further scrutinised in the presence of verbal and non-verbal interference. Specifically, the effect of concurrently articulating task irrelevant verbal material during syllable presentation was compared with passive listening to irrelevant speech, sequential, manual tapping, and silent mouthing, to determine the stage at which auditory and visual speech converge. Crucially, it is shown that concurrently articulating verbal material out loud or silently mouthing speech during syllable identification reduces the McGurk effect, whereas passive listening to task-irrelevant speech does not. This latter observation suggests that speech generated from an irrelevant speaker does not disrupt audiovisual binding. In addition, the disruptive impact of concurrent articulation cannot simply be attributed to the presence of a secondary demanding task, since the non-verbal, dual-task of sequential tapping does not significantly reduce the McGurk effect. On the basis that both concurrent articulation and silent mouthing, by design, impede subvocal speech production processes, and because both manipulations disrupt the McGurk effect, it is suggested that subvocal motor mechanisms necessary for speech production are also involved in binding of auditory-verbal and visual-verbal inputs.

Finally, in Chapter 4 the findings of each experimental chapter are reviewed, and a discussion of the potential implications of the data is presented. Overall it is concluded that, if progress is to be made in understanding the mechanisms underlying verbal behaviour, an approach should be adopted that not only requires an amodal, phonological representational form, but also considers the extent to which verbal performance reflects emergent by-products of modality-specific systems primarily serving perceptual and motor processes.

CHAPTER 2

Modalities of memory: Is reading lips like hearing voices?

Introduction

The assumption that heard and seen speech both gain access to a common representation is suggested by critical similarities in detailed aspects of verbal short-term memory performance. For example, the serial recall of auditory lists shows that under a narrow range of conditions recall is enhanced towards the end of the sequence compared to material that is read – the so-called recency effect (Crowder & Morton, 1969). This enhanced performance is disrupted by the presentation of a redundant end-of-list suffix in the same modality as the sequence – the suffix effect (Crowder & Morton, 1969). However, enhanced recall at recency is evident for lipread material, a stimulus that is inherently visual. There is also evidence of cross-modal interactions in short-term memory between auditory-verbal and lipread-verbal material, including effects of lipread and auditory suffixes on a lipread memory sequences (e.g., Campbell & Dodd, 1980, 1982; de Gelder & Vroomen, 1992; Gathercole, 1987; Greene & Crowder, 1984; Spöehr & Corin, 1978). Such interactions have been taken to indicate that a common form of representation is shared by the two modalities (Campbell & Dodd, 1982; Greene & Crowder, 1984; Spöehr & Corin, 1978).

The motive for revisiting these phenomena in the current series of experiments is based on evidence showing that other aspects of verbal short-term memory phenomena, typically ascribed to processes operating on an amodal level of representation can, under closer scrutiny, be attributed to modality-specific motor and perceptual processes. For example, key evidence establishing the character of phonological representations in verbal short-term memory has been derived from the disruptive effect of task-irrelevant background speech. Namely, when a to-be-ignored sequence of verbal items is presented during the encoding and/or retention of the memory list, serial recall performance for both auditory and visual stimuli is disrupted (e.g., Colle & Welsh, 1976; Jones, Madden, & Miles, 1992; LeCompte, 1996; Neath, 2000; Salamé & Baddeley, 1990; Schlittmeier, Hellbrück, & Klatte, 2008). That the disruptive effect of task-irrelevant sound occurs when the modality of the sequence is either heard or read has been taken to suggest that both the memory items and irrelevant speech occupy the same phonological level of representation (e.g., Baddeley, 1990; Salamé & Baddeley, 1982). Specifically, even when explicitly instructed to ignore the irrelevant sequence, the irrelevant auditory items are automatically converted into phonological representations and enter the phonological store. As such, the degree of resemblance between the phonological identity of the irrelevant sound and the memory items in the phonological store is a key determinant of interference.

However, we now know that the degree of change within the irrelevant speech sequence, and not phonological similarity *per se*, is actually a key determinant of its disruption – termed the changing-state effect (Jones, et al., 1992). For example, a sequence of changing sounds (repetitions of the letter sounds, *a*, *b*, *c*, ...) will disrupt serial recall performance to a greater extent than a sequence of a single repeating sound (e.g., *a*, *a*, *a*, ...). Critically, the changing-state effect can be exploited to show that, even when the phonological identity of the irrelevant sequence is fixed, the magnitude of disruption can be manipulated through changes in perceptual organisation. So, if a repeated sequence of three sound tokens is presented so that it appears to emanate from a single source, then it will disrupt recall performance (see e.g., Ellermeier & Zimmer, 1997). If those same three sounds are presented in the same fixed order, but are perceived as emanating from separate sources (e.g., using stereophonic presentation of the sounds so that the first is assigned to the right ear, the second to the left ear, and third to both ears), this will lead to the perception of three separate repeated sequences and will substantially attenuate their impact on serial recall (Jones, Macken, & Murray, 1993; Jones, Saint-Aubin, & Tremblay, 1999).

Furthermore, another key determinant of the disruptive power of irrelevant sound is the extent to which the verbal short-term memory task requires rehearsal; irrelevant sounds are more disruptive when the memory task involves seriation, such as a test of recall for order that uses a subset of six days of the week (e.g., *Friday*, *Wednesday*, *Sunday*, *Monday*, *Thursday*, *Saturday*) than a task requiring the participant to report the missing day (*Tuesday*). In the latter case, not only is the effect of irrelevant sound attenuated (Beaman & Jones, 1997), it does not exhibit a changing-state effect (Jones & Macken, 1993; Marsh & Jones, 2010). What these findings imply is that interference arises because the irrelevant sequence and memory list both involve sequential processing, which compete for control of the speech-motor planning process (e.g., Jones & Macken, 1993; Jones, Macken, & Nicholls, 2004; Macken, et al., 2009). That is to say, the irrelevant sound effect is due to the obligatory generation of sequential representations, as opposed to amodal representations in phonological storage, which compete for control of motor planning processes that are also engaged in the subvocal rehearsal process (e.g., Jones, Banbury, Tremblay, & Macken, 1999; Jones

& Macken, 1993; Jones, et al., 1993; Macken, Mosdell, & Jones, 1999; Macken, et al., 2009).

The phonological similarity effect is another phenomenon that has also been attributed to a phonological level of representation, but, like the irrelevant sound effect, can in fact be attributed to modality-specific motor and perceptual processes. Specifically, the effect of phonological similarity is the demonstration that sequences of similar sounding verbal items (e.g., the letter sounds, *b*, *c*, *d*, *g*, …) are more poorly recalled than dissimilar sounding sequences (e.g., *f*, *k*, *l*, *q, …*), and occurs whether the sequences are read or heard (e.g., Conrad & Hull, 1964; Crowder & Morton, 1969). That the effect itself seems to transcend presentation modality has again been taken to suggest that it occurs at a representational level that also transcends modality (e.g., Baddeley, 2012); the effect of phonological similarity occurs because similar phonological representations have fewer discriminating features within the phonological store, which ultimately reduces recall accuracy of the memory items. However, when subvocal rehearsal of the to-be-remembered sequence is prevented by concurrent articulatory suppression (i.e., the repetition of irrelevant verbal material during the presentation and/or retention of the to-be-remembered list, such as "*1, 2, 3, 1, 2, 3*…"), the phonological similarity effect is still observed for heard sequences, but not when they are read (e.g., Baddeley, 1986; Jones, et al., 2004). This interaction between similarity, articulation and modality of presentation has typically been ascribed to the joint action of an articulatory loop and phonological store (Larsen & Baddeley, 2003): Concurrent articulation impedes the grapheme-to-phoneme conversion process such that visual-verbal items are no longer represented in the phonological store and are not subject to the effects of phonological similarity. In contrast, the similarity effect remains for auditory sequences since auditory material gains obligatory access to the phonological store, allowing interactions to occur between similar phonological representations. Thus, the role of modality in this interaction is attributed to different modality-based access routes to phonological storage, rather than to differences in the nature of the representations derived from the different modes of presentation.

Nevertheless, recent evidence has shown that there are significant differences in the manifestation of the phonological similarity effect when presented in visual-verbal or auditory-verbal forms (e.g., Jones, Hughes, & Macken, 2006; Jones, et al., 2004; Maidment & Macken, 2012). Although the interaction between similarity, articulation and modality has typically been attributed to modality-based access routes to bespoke phonological storage systems (Larsen & Baddeley, 2003), the critical survival of the phonological similarity effect in the presence of articulatory suppression tends to be localised in the recency portion of the serial position curve for dissimilar sounding sequences (Jones, et al., 2006; Jones, et al., 2004; Maidment & Macken, 2012). This evidence therefore indicates that the (so-called) phonological similarity effect actually has two distinct components, neither which necessitate a representational form that is amodal or phonological in essence. First, as described above, articulatory suppression abolishes the effect of similarity throughout the list for visual presentation, and in all but the terminal item or two for auditory presentation (Jones, et al., 2006; Jones, et al., 2004). This suggests that this specific effect of similarity resides in the rehearsal process; where rehearsal is prevented there is no effect of similarity and points to an effect of articulatory (rather than phonological) similarity – a conclusion supported by the fact that the type of errors made in serial recall are functionally equivalent to those found in natural speech (Acheson & MacDonald, 2009) or when verbal material is read aloud

(Ellis, 1980), and therefore when there are no or minimal demands on a putative phonological short-term memory store. Second, the effect of similarity that survives articulatory suppression is predominantly evident in the recency portion of the serial position curve, and has therefore been attributed to an effect of acoustic, and not phonological similarity (Jones, et al., 2006; Jones, et al., 2004; Maidment & Macken, 2012).

The functional similarities between lipread and auditory serial recall appear to counter an account of recency based on the operation of modality-specific perceptual processes. For example, early accounts of these phenomena were ascribed to specifically auditory mechanisms, such as limited-capacity pre-categorical acoustic storage (PAS) (Crowder & Morton, 1969). According to the original PAS hypothesis, recall of the terminal list items of an auditory list reflects two sources of information – an amodal, phonologically coded (or post-categorical) representation that is also supplemented by a pre-categorical, echoic representation. The PAS account attributes auditory suffix and recency effects to the latter form of representation, the trace of which either decays rapidly or is masked by subsequently presented items. in terms of the PAS account then, a suffix has its impact on recency by masking the representation of the terminal list item within PAS. Such a view is reinforced by a substantial body of work showing that auditory recency and suffix effects are observed only when very specific conditions are met. Specifically, for auditory recency to occur, the memory sequences must consist of dissimilar sounding words or syllables (Crowder, 1971; Frankish, 1996; Jones, et al., 2006). Furthermore, the physical similarity between the suffix and the items of the list is a key determinant of the size of the suffix effect; a suffix will abolish auditory recency if it shares the same acoustic characteristics as the to-be-remembered list. By comparison, if the suffix is spoken in a different voice, comes from a different location, or is temporally delayed, recall at the terminal list item is somewhat disrupted at some mid-point between the intact performance observed during no suffix, control conditions and matched-suffix conditions where it is completely abolished (e.g., Frankish & Turner, 1984; Frick, 1988; Greenberg & Engle, 1983; Greene, 1991; Morton, Crowder, & Prussin, 1971).

Alternative accounts of these phenomena have also evoked specifically auditory processing, but in a different way, primarily by reference to processes of auditory perceptual grouping (e.g., Frankish, 1989, 2008; Frick, 1988). From this perspective, auditory recency occurs because the terminal memory item occupies a distinctive boundary position at the end of the to-be-remembered sequence; the last item in the list is more likely to be recalled in the correct serial position relative to other items within the sequence because of fewer opportunities for transposition (e.g., Harris, 1989; Henson, Norris, Page, & Baddeley, 1996). By extension, the similarity of the suffix to the terminal item promotes grouping of the suffix to the list, modifying the encoding of order. As such, an acoustically similar suffix to the memory items is subject to grouping processes that act to perceptually integrate it with the to-be-remembered sequence, so that it now occupies the distinctive boundary position previously occupied by the terminal list item, disrupting recency as a result (e.g., Bregman, 1990, 1994; Frankish, 2008; Jones, et al., 2006; Jones, et al., 2004; Maidment & Macken, 2012; Nicholls & Jones, 2002). Thus, the reduction in recency from a suffix is not a function of the similarity between the suffix and the memory sequence *per se*, but rather due to whether or not that similarity causes the suffix to be grouped perceptually with the sequence.

Further support for ascribing recency and suffix effects in auditory-verbal short-term memory to essentially perceptual, auditory processes – as opposed to masking within PAS – comes from experiments that exploit an established phenomenon in the study of auditory perceptual organisation known as *auditory capture* (Bregman, 1990). For example, Bregman and Rudnicky (1975) examined memory for the order of tones whereby participants first heard two target tones of slightly different pitches presented rapidly in succession (e.g., *A* and *B* represented in Figure 1), which were then followed after a short interval by two test tones presented in the same or reversed order as the targets. Immediately after presentation of the test tones, participants were required to report whether the test tones were presented in the same order as the target tones. Although participants could discriminate the order of the test tones with relative ease when they were presented in isolation (Figure 1, Panel A), if the test tones were flanked by 'flanker' (*F*) tones, memory for order was substantially impaired (Figure 1, Panel B). However, this impairment was reduced considerably by the presentation of additional 'captor' (*C*) tones similar in pitch to the flankers (Figure 1, Panel C). In much the same way, manipulating the likelihood the suffix belongs to either the to-be-remembered sequence or is 'captured' by an irrelevant auditory sequence presented concurrently with the memory lists will alter the suffix effect appreciably (e.g., Kahneman & Henik, 1981; Maidment & Macken, 2012; Nicholls & Jones, 2002). The spoken word "go", for instance, presented at the end of a spoken sequence of to-be-remembered digits will reduce recency for that sequence compared to conditions where there is no suffix. However, if a second, irrelevant sequence of repetitions of the spoken word "go" is played concurrently with the to-be-remembered sequence, then the effect of the "go" suffix, despite it occupying the same temporal and acoustic relation to the sequence

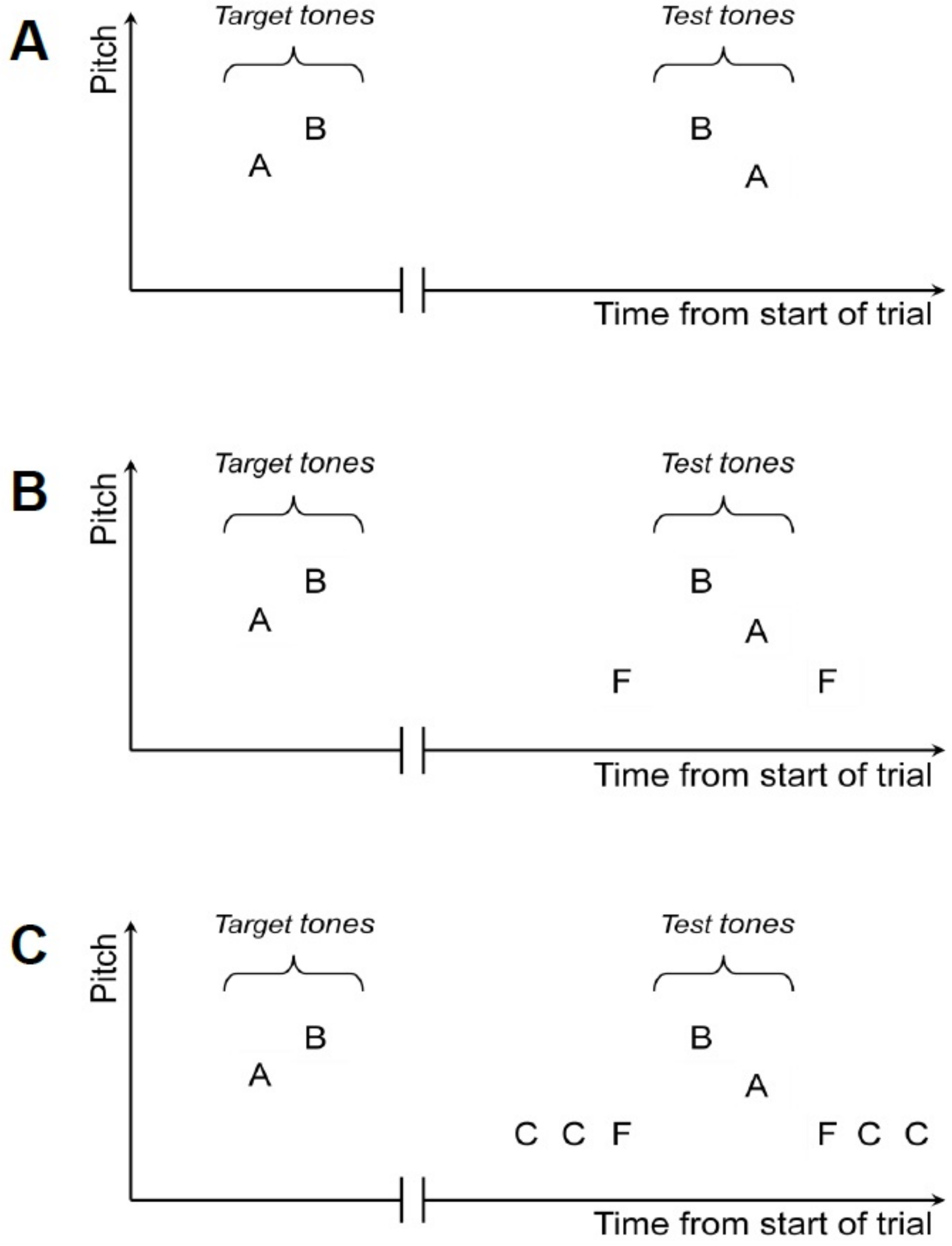

Figure 1. A schematic illustration of the stimuli used by Bregman and Rudnicky (1975). Participants discriminated the order of two test tones relative to the target tones A and B, which were presented either in isolation (Panel A), in the presence of flanker (F) tones (Panel B), or in the presence of flanker and captor (C) tones (Panel C).

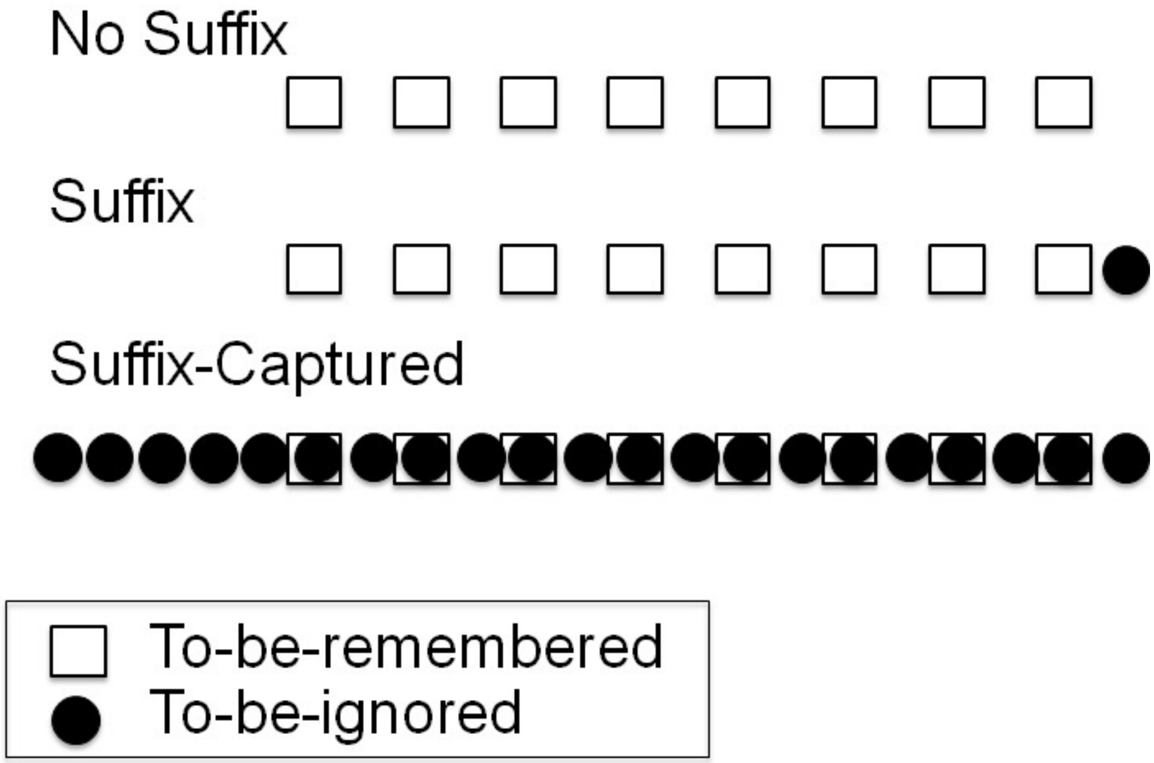

Figure 2. A schematic representation of the sequencing and relative timing of stimuli presented by Nicholls and Jones (2002). The "go" suffix occurred 50ms after the offset of the last to-be-remembered item and occurred in tempo with the proceeding irrelevant auditory "go" sequence when captured. For ease of exposition, the figure does not preserve the scale of item duration or series length.

(see Figure 2), is substantially reduced or eliminated. Furthermore, the extent of this attenuation is determined by how strongly the concurrent sequence forms a coherent stream into which the suffix can be captured (Kahneman & Henik, 1981; Nicholls & Jones, 2002).

The logic that follows from both this and Bregman and Rudnicky's (1975) demonstration is that even when the relation of the suffix to the terminal list item is fixed, the impact of suffix changes as a function of manipulating its relation to surrounding auditory material, which acts to either perceptually isolate it from or incorporate it within a sequence of sounds. However, if recency and suffix effects can be observed when the verbal information is presented in a silent visual (i.e., lipread) form, then the accounts described above based on auditory perceptual organisation are called into question. It could be maintained that silent lipreading gives rise to 'functional sound' due to the tight temporal coupling between heard and seen patterns that the typical individual participant will have been exposed on countless occasions (Crowder, Harvey, Routh, & Crowder, 1983; Engle, Cantor, & Turner, 1989; M. Hall & Bavelier, 2010; Hanson, 1982; Koo, Crain, LaSasso, & Eden, 2008). In support, there is evidence that silently lipread information appears to gain ready access to auditory brain areas (Bernstein et al., 2002; Calvert et al., 1997), implying that visual- and auditory-verbal information converge at a level typically associated with auditory processing (for review see Alais, Newell, & Mamassian, 2010). Even so, this still does not get away from the fact that an account of recency and suffix effects based on auditory processing would at least have to undergo substantial modification in order to account for such effects in the absence of audition. From this perspective, the experimental series that follows explores the functional similarities between auditory and lipread speech in relation to the nature of their respective recency and suffix effects.

Experiment 1

The reason for ascribing recency and suffix effects in auditory-verbal short-term memory to perceptual, auditory processes comes from the findings that these effects are immune to disruption by factors that impede the speech motor processes supporting subvocal rehearsal. Namely, articulatory suppression disrupts short-term serial recall by preoccupying the processes that would otherwise be utilised for subvocal rehearsal of the memory material (e.g., Baddeley, Thomson, & Buchanan, 1975; Jones, et al., 2004; Macken & Jones, 1995). Task-irrelevant sequences of background sound also impact upon the rehearsal process by providing alternative perceptual sequences that compete with the memory sequence for control of the speech motor process (e.g., Macken, et al., 2009). This is evidenced by the fact that such disruption only occurs if the focal task is a serial one and therefore induces a subvocal rehearsal strategy (e.g., Beaman & Jones, 1997; Jones & Macken, 1993), as well as by finding that the disruptive capacity of the irrelevant sound is eliminated if the participant is required to engage in articulatory suppression during the task, again disrupting the role of subvocal rehearsal of the memory sequence (e.g., Jones, et al., 2004; Macken & Jones, 1995). While each of these factors has a substantial detrimental impact on recall performance throughout most of the list for visual-verbal and auditory-verbal sequences, they leave auditory recency intact (Jones, et al., 2004; Maidment & Macken, 2012). On this basis, Experiment 1 investigates whether or not lipread recency is similarly immune to such effects.

Method

Participants

Twenty-eight participants (17 females), aged 18 to 27 years (19.50 mean years), were recruited online from the School of Psychology, Cardiff University, and received course credit for participation. All participants were native English speakers who reported normal (or corrected-to-normal) vision and normal hearing. Pre-screening was used to exclude participants who did not correctly identify 85% or more of the lipread stimuli (in this and all experiments within the current chapter).

Apparatus & Materials

The digits one to nine served as to-be-remembered items. For lipread presentation, in a sound attenuated booth, using a *Sony Handycam HDR-SR11* individual digits one to nine were recorded in black and white from a female, wearing black lipstick to enhance contrast, speaking in a monotone voice (at a F0 of approximately 225Hz). Videos were cropped so that a head-on view of only the upper and lower lips was visible. The sound was digitally sampled with a 16-bit resolution (at a rate of approximately 48 kHz) but removed from the video files to serve as the auditory items. Written digits were presented in black 72-point Times New Roman font. Using *E-Prime* (version 2.0.8.22: Psychological Software Tools, Inc.), visual material was displayed in the centre of a PC screen, while auditory stimuli were presented monophonically over headphones.

Memory sequences were constructed from pseudo-random orderings of eight digits with three constraints: (1) no digit was repeated within a sequence; (2) sequences could not contain more than two digits in an ascending or descending order*;* and (3) a digit could not appear in the same serial position on consecutive sequences. Irrelevant speech was generated by recording the same female speaking the letter sounds *a, b* and *c*. To enhance the ease of distinction between the relevant and irrelevant speech, using *SonicForge 5.0* software (Sonic Foundry, Inc. Madison, WI; 2000) the pitch of the irrelevant tokens was lowered by 3 semitones and compressed to 190ms without further changing pitch.

Design & Procedure

Presentation Modality (auditory, lipread, written), Interference (control, irrelevant speech, articulatory suppression), and Serial Position (1 to 8) were varied within groups, with 12 trials for each of the nine Modality x Interference conditions. Trials were arranged in a pseudo-random order such that no interference type or presentation mode was presented more than twice in succession, and were balanced across all participants. Participants were tested individually in a sound attenuated laboratory and wore headphones throughout the experiment with volume individually adjusted to a comfortable sound level.

To ensure that participants could correctly identify the lipread stimuli, a 20-minute pre-screening phase was administered prior to the experimental procedure. For each trial, a 500ms warning tone (500Hz sinewave) was followed by a central fixation cross presented for 2s before a digit between one and nine was shown for 1s in one of the three presentation formats (i.e., auditory, lipread, written). Participants were then required to identify the presented digit on a response screen displaying the digits one to nine in written form, moving the cursor over the digit they thought had been presented. Following a response, written feedback was provided for 4s indicating the correct identity of the digit, after which the next screening trial commenced. This phase consisted of 81 trials, with each digit being presented three times pseudo-randomly for each presentation modality, with the constraint that no digit or presentation mode was presented in succession.

For experimental trials, the 500ms warning tone was followed by a central fixation-cross presented for 5s before the onset of the first to-be-remembered item. This interval was unfilled in control trials. For irrelevant sound trials, the 5s interval prior to the onset of the to-be-remembered sequence was filled with irrelevant speech that continued without a break in tempo (onset-to-onset = 250ms) throughout presentation of the memory sequence and ending at the offset of the final to-be-remembered item. Participants were instructed to ignore any spoken letters heard during trials. For conditions involving articulatory suppression, participants were instructed to start whispering aloud the letters *a, b, c* at a rate of approximately four per second. The experimenter coached each participant in the correct rate and loudness for articulation before continuing. Articulatory suppression took place from tone offset until the last memory item was presented.

In all conditions, to-be-remembered items were presented with a 1s onset-to-onset interval, with a 100ms central fixation cross interleaved between each item. Immediately after the final to-be-remembered item was presented, participants were cued to recall on the same response screen provided in the pre-screening session. Using the cursor, participants were required to click on eight digits corresponding to the exact order of the presented sequence. Following eight responses, the next trial commenced automatically. The experimental procedure lasted approximately 45-minutes, including an optional 5-minute rest period at the halfway point.

Results

No participant was excluded from the experimental session with pre-screening scores ranging from 91.36% to 100% (mean = 95.68%) lipread items correctly identified.

Responses were scored according to an absolute serial order criterion; an item was scored correct only if it was recalled in its correct position within the sequence. A general analysis of performance within each condition across all serial positions is reported first. The effects on recency of each interference type will then be reported. In general terms, recency measures fall into one of three categories: (1) *absolute*, which takes the accuracy with which the terminal item is recalled; (2) *relative*, based on the change in recall between the terminal position and preterminal position; and (3) *normalised* (or transformed), which is calculated by expressing correct recall at the terminal position as a proportion of the sum of correctly recalled items across all serial positions. Although the serial recall of both auditory and lipread sequences show enhanced recall towards the end of the list compared to written material, correct recall performance across all serial positions is often greater in magnitude for auditory, relative to lipread sequences (e.g., de Gelder & Vroomen, 1992; Turner et al., 1987). Therefore, a relative measure that calculates the improvement in recall between the last and penultimate serial positions is considered most appropriate and will be reported here.

Serial Position Analysis

Figure 3 illustrates the percentage of correct recall in each condition as a function of serial position. A 3 (Interference: control, irrelevant speech, articulatory suppression) x 3 (Modality: auditory, lipread, written) x 8 (Serial Position) repeated measures ANOVA demonstrated that relative to control conditions (Figure 3a) serial recall across all list positions was depressed in the presence of irrelevant speech (Figure 3b) and, to an even greater magnitude, in the presence of articulatory suppression (Figure 3c), shown here as a main effect of interference, $F(2,54) = 74.70$, $MSE = 2085.15$, $p<.001$, $\eta^2 = .74$. There was also a significant main effect of modality, $F(2,54) = 6.18$, $MSE = 2356.34$, $p<.01$, $\eta^2 = .31$, as well as an interaction between modality, interference and serial position, $F(28,756) = 1.48$, $MSE = 166.75$, $p<.05$, $\eta^2 = .05$. All other main effects and interactions also reached significance ($p<.05$). As can be seen in Figure 3, the source of this 3-way interaction appears to reside in the differential effects of modality and interference on the recency portion of the curve. In particular, and critical to the question of whether an amodal representation exists between auditory and lipread inputs, while recency for auditory sequences is robust in the presence of interference, that for lipread sequences is reduced to the same level as that found with written presentation. This picture is supported by the following analysis focussing on recency.

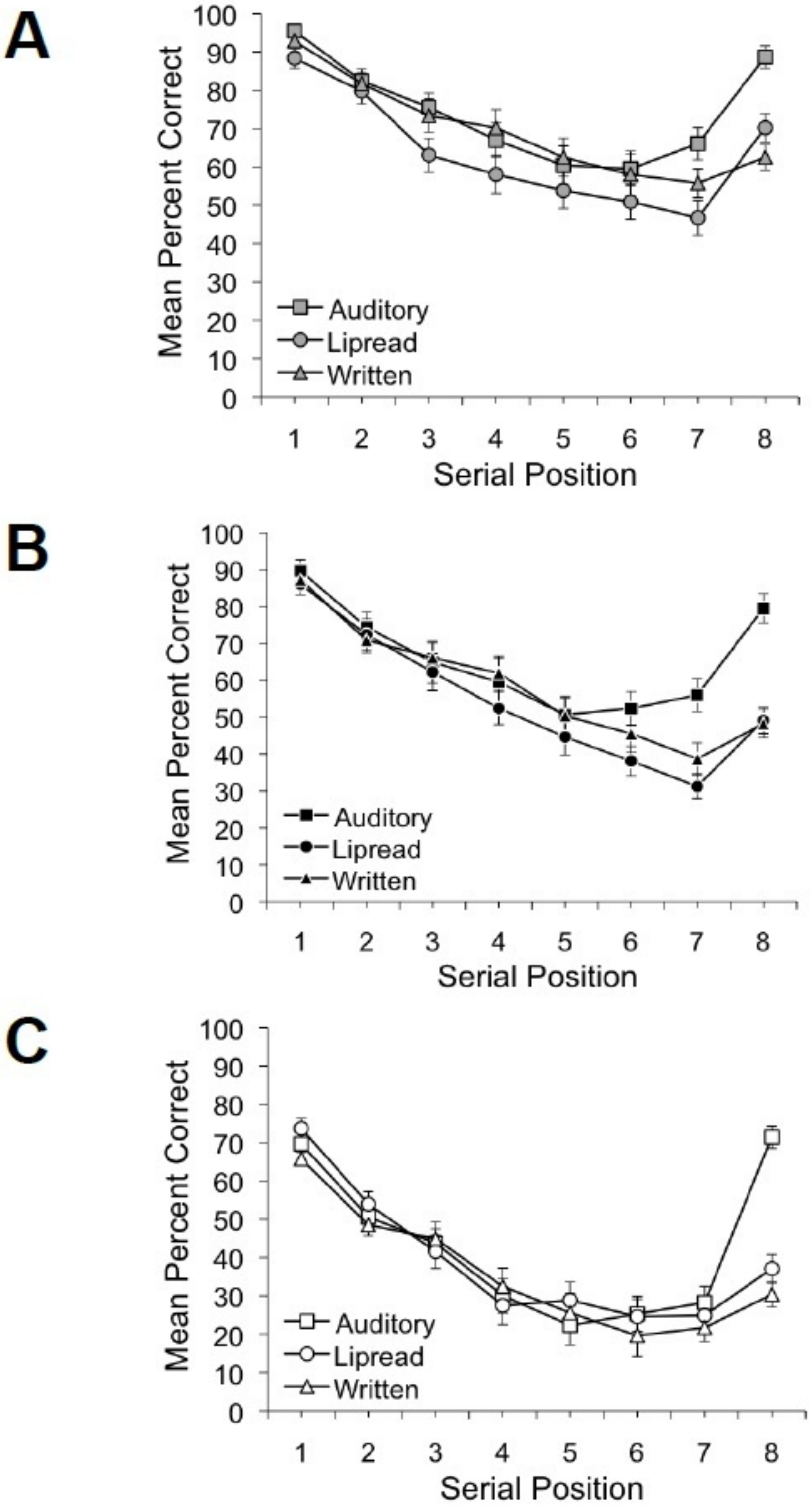

Figure 3. Mean percentage of items correctly recalled across all serial positions for Experiment 1 during auditory, lipread, and written presentation modes. Panel A shows control trials, Panel B irrelevant speech trials, and Panel C concurrent articulatory suppression trials.

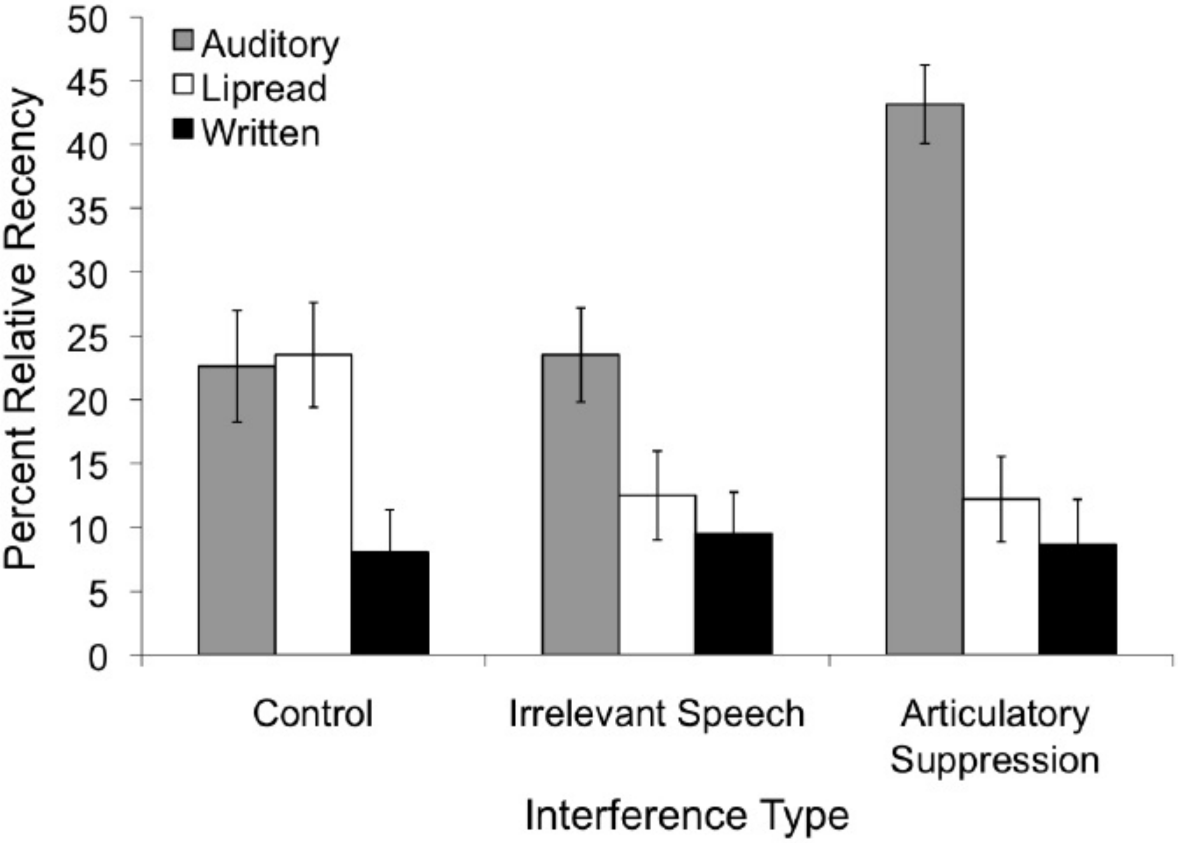

Figure 4. Results of the relative recency measure for Experiment 1 as a function of interference (control, irrelevant speech, articulatory suppression). Scores are expressed in terms of the mean accuracy with which the terminal item was recalled minus recall accuracy at the preterminal serial position for auditory, lipread, and written presentation modes. Error bars represent the standard error of the mean.

Relative Recency Analysis

Figure 4 illustrates the results of the relative recency measure. Overall, in the absence of interference, both auditory and lipread recency effects are equivalent in magnitude. However, while relative auditory recency is undiminished by irrelevant sound and articulatory suppression, lipread recency is reduced.

A 3x3 repeated measure ANOVA on the relative recency measure confirmed that recency was superior for auditory presentation, demonstrated here as a main effect of modality, *F*(2,54) = 27.35, *MSE* = 349.96, *p*<.001, η^2 = .50. Critically, there was also an interaction between modality and interference, *F*(4,108) = 8.19, *MSE* = 253.69, *p*<.001, η^2 = .23, such that recency performance was equivalent for both auditory and lipread presentation in the absence of interference (*p*=.84). Nevertheless, unlike auditory recency, lipread recency was reduced in the presence of both irrelevant speech, *F*(1,27) = 5.69, *MSE* = 298.49, *p*<.05, η^2 = .17, and articulatory suppression, *F*(1,27) = 6.95, *MSE* = 257.75, *p*<.05, η^2 = .21.

Discussion

As expected on the basis of the precedents (e.g., Campbell & Dodd, 1980, 1982; Dodd, Hobson, Brasher, & Campbell, 1983; Spöehr & Corin, 1978), in the absence of interference, when measured in relative terms auditory and lipread recency are equivalent – a finding that, in isolation, points to a common representation for both auditory and lipread speech. However, despite the apparent functional similarities between lipread and auditory speech, auditory recency is evident in the presence of irrelevant sound and articulatory suppression, whereas the same cannot be said for lipread recency.

Clearly then, there are critical functional differences between auditory and lipread recency that go beyond their relative magnitude. Rather, the way in which they are subject to disruption by both articulatory suppression and task irrelevant sound points to a divergence. The results here, as well as elsewhere (Jones, et al., 2006; Jones, et al., 2004; Maidment & Macken, 2012), suggest that the auditory recency advantage is immune to interference from articulatory suppression and irrelevant sound, whereas under these circumstances the shape of the lipread serial position curve mimics that of written performance. This therefore raises the question of the extent to which speech representations derived from visual information are similar to those derived from auditory presentation. It may, despite the present results, be the case that once encoded, lipread and auditory speech share an equivalent mode of representation, but that the path whereby they are encoded is nonetheless distinct. For example, lipread stimuli might gain access to auditory-like representations via associative pathways linking speech production and perception (e.g., Hickok, 2009; Hickok & Poeppel, 2004, 2007; Skipper, Goldin-Medow, Nusbaum, & Small, 2009; Skipper, Van Wassenhove, Nusbaum, & Small, 2007). Both irrelevant sound and articulatory suppression have been shown to have their effects via their impact on speech production processes involved with subvocal rehearsal (Jones & Macken, 1993; Jones, et al., 2004; Macken & Jones, 1995; Macken, et al., 2009), so that disruption of such a pathway may prevent the typical encoding of lipread speech into an auditory-like representation. As such, it may be a shared form of representation that gives rise to functional similarities between auditory and lipread serial recall, but that representational form is encoded directly in the former case and indirectly via speech production mechanisms in the latter.

To investigate this possibility further, Experiments 2 to 5 explored the extent to which other functional similarities between heard and seen speech could be attributed to a shared representational forms. Specifically, as well as exhibiting similar recency effects under control conditions, both auditory and lipread sequences behave similarly in relation to the disruptive effect of an end-of-sequence suffix on that recency performance.

Experiment 2

Critical to our understanding of recency in short-term memory is the existence of the so-called suffix effect (Conrad & Hull, 1964; Crowder & Morton, 1969). Recent evidence has pointed to the fact that auditory perceptual processes play a key role in shaping both recency and suffix effects; the effect of the auditory suffix on auditory recency has been shown to be the result of the tendency for the perceptual system to bind similar acoustic stimuli into coherent sequences – through processes of perceptual streaming – rather than interference due to overwriting or backward masking within acoustic storage (Bregman & Rudnicky, 1975; Kahneman & Henik, 1981; Nicholls & Jones, 2002). However, auditory, lipread and bimodal (auditory and lipread combined) suffixes have been shown to disrupt a lipread list(Spöehr & Corin, 1978), a finding that has been widely replicated (e.g., Campbell & Dodd, 1980; Greene & Crowder, 1984; Nairne & Walters, 1983).

While this interaction seems to indicate that heard and seen speech both gain access to a common representational form, previous studies showing an impact on lipread recency from a lipread and auditory suffixes (de Gelder & Vroomen, 1992, 1994; Greene & Crowder, 1984; Spöehr & Corin, 1978) have failed to show the converse effect whereby a lipread suffix disrupts auditory recency. Furthermore, these studies have typically required the participants to processes the suffix in some way. For example, Spöehr and Corin (1978) and Green and Crowder (1984) presented two different types of lipread suffix: one that acted as a cue to begin recall, and another presented intermediately on catch trials that required a different response such as circling the trial number on the response sheet or writing letters of the alphabet. In contrast, de Gelder and Vroomen (1992) did not require participants to respond to the lipread suffix, but instead presented different coloured circles after the suffix had been presented. For example, a green circle signaled to the participant to begin recall of the memory sequence, whereas when participants saw an occasionally presented red circle they were expected to write a series of crosses. In all instances, these departures from the original procedure were a pragmatic step in order to ensure that attention was paid to the visual suffix. However, this method contravenes the procedure of the classical suffix effect (Crowder & Morton, 1969) in which the suffix is treated as a to-be-ignored item, raising the possibility that its disruptive impact in this setting possibly resides in some form of attentional distraction, rather than in specific interference amongst cross- or amodal verbal representations.

In support of the prediction that auditory and lipread recency and suffix effects may not be attributed to same underlying, amodal representation, comes from the observation that not all studies have deployed fully crossed designs to test the modality specificity of the suffix effect. When all combinations of presentation modality are crossed with suffix modality (tone or a range of lipread, spoken or bimodal speech stimuli), disruption at the terminal list item is typically much greater when suffix modality matches that of the to-be-remembered sequence (e.g., de Gelder & Vroomen, 1992; Gathercole, 1987; Greene & Crowder, 1984). That is to say, in most cases, the impact of a spoken suffix on lipread lists is smaller in magnitude to that produced by a lipread suffix. Thus, previous findings not only raise questions about whether the similarities between lipread and auditory sequences can be attributed to the same source (Turner, et al., 1987), but also highlight that when the modality specificity of the suffix effect has been examined under conditions where list modality has been fully crossed with suffix modality, the impact if the suffix is dependent upon whether or not is shares the same modality as the to-be-remembered sequence.

Consequently, in Experiment 2 the effects of three types of verbal suffix (the word “go”) – auditory, lipread, and bimodal (i.e., compound auditory and lipread), as well as a control condition in which no-suffix was presented – was contrasted on auditory and lipread to-be-remembered sequences, to further test the possibility that a common representation underlies both speech inputs. If auditory and lipread speech converge upon a common form of representation that is organised in much the same way regardless of modality, then the effects of both lipread and auditory suffixes should be of similar magnitude. In relation to the bimodal suffix, only two studies have examined such suffixes with lipread and auditory lists. Both Spöehr and Corin (1978) and deGelder and Vroomen(1992) showed reduced recency (to about the same degree as was shown with a bimodal sequence, see Greene & Crowder, 1984). Assuming then that each element of the compound comprising the bimodal suffix generates a suffix effect, the addition of auditory features to the lipread suffix would not be expected to fundamentally change the nature of its ability to disrupt recency.

Method

Participants

Thirty volunteers (28 female), aged 18 to 25 years (19.27 mean years) were recruited online from Cardiff University's School of Psychology, and received course credit for their participation. All participants, none of whom had taken part in Experiment 1, were native English speakers who reported normal hearing and normal (or corrected-to-normal) vision.

Apparatus & Materials

The apparatus and materials deployed in Experiment 1 were again used in Experiment 2 with the addition that the word "go", spoken by the same female speaker as the to-be-remembered digits, served as the suffix.

Design & Procedure

The factors Modality of to-be-remembered sequence (lipread, auditory), Modality of Suffix (lipread, auditory, bimodal, no-suffix), and Serial Position (1 to 8) were manipulated in a 2x4x8 repeated measures design. Twelve sequences were presented for each of the eight Suffix x Modality conditions, arranged in a pseudo-random order with the constraint that no modality of to-be-remembered sequence or suffix were presented more than twice in succession. In Experiment 2 a faster rate of presentation to that in Experiment 1 was used to promote perceptual streaming of the suffix with the to-be-remembered sequence (see Bregman, 1990; Kahneman & Henik, 1981; Nicholls & Jones, 2002). Each item was 500ms in duration and presented with a 550ms onset-to-onset interval, with a 50ms central fixation cross interleaved between each to-be-remembered item. The offset of the last memory item was followed after 50ms by either a still frame of the speaker with closed lips (no-suffix conditions), or a lipread, auditory, or bimodal suffix. Critically, participants were not required to explicitly report that they had seen the suffix, nor did it act as a cue to recall. Rather, participants were unaware of when a suffix would be present given its random presentation, and were explicitly asked to ignore any item if shown after the last item in the to-be-remembered sequence was presented. As in Experiment 1, immediately after the suffix item was presented, participants were cued to recall when the response screen, displaying the digits one to nine, was presented.

Results

Pre-screening scores ranged from 85.19% to 100% (mean = 92.97%) lipread items correctly identified.

Serial Position Analysis

Figure 5 illustrates the percentage of correct recall in each condition as a function of serial position. A 2 (Modality of to-be-remembered sequence) x 4 (Modality of Suffix) x 8 (Serial Position) repeated measures ANOVA indicated that recall performance was superior when the to-be-remembered list was presented auditorily, shown here as a main effect of modality, *F*(1,29) = 33.95, *MSE* = 2869.81, *p*<.001, η^2 = .54. The disruptive effect of the suffix was restricted to the last one or two serial positions within the memory list, confirmed by a main effect of suffix, *F*(3,87) = 12.65, *MSE* = 376.86, *p*<.001, η^2 = .30, as well as an interaction between suffix and serial

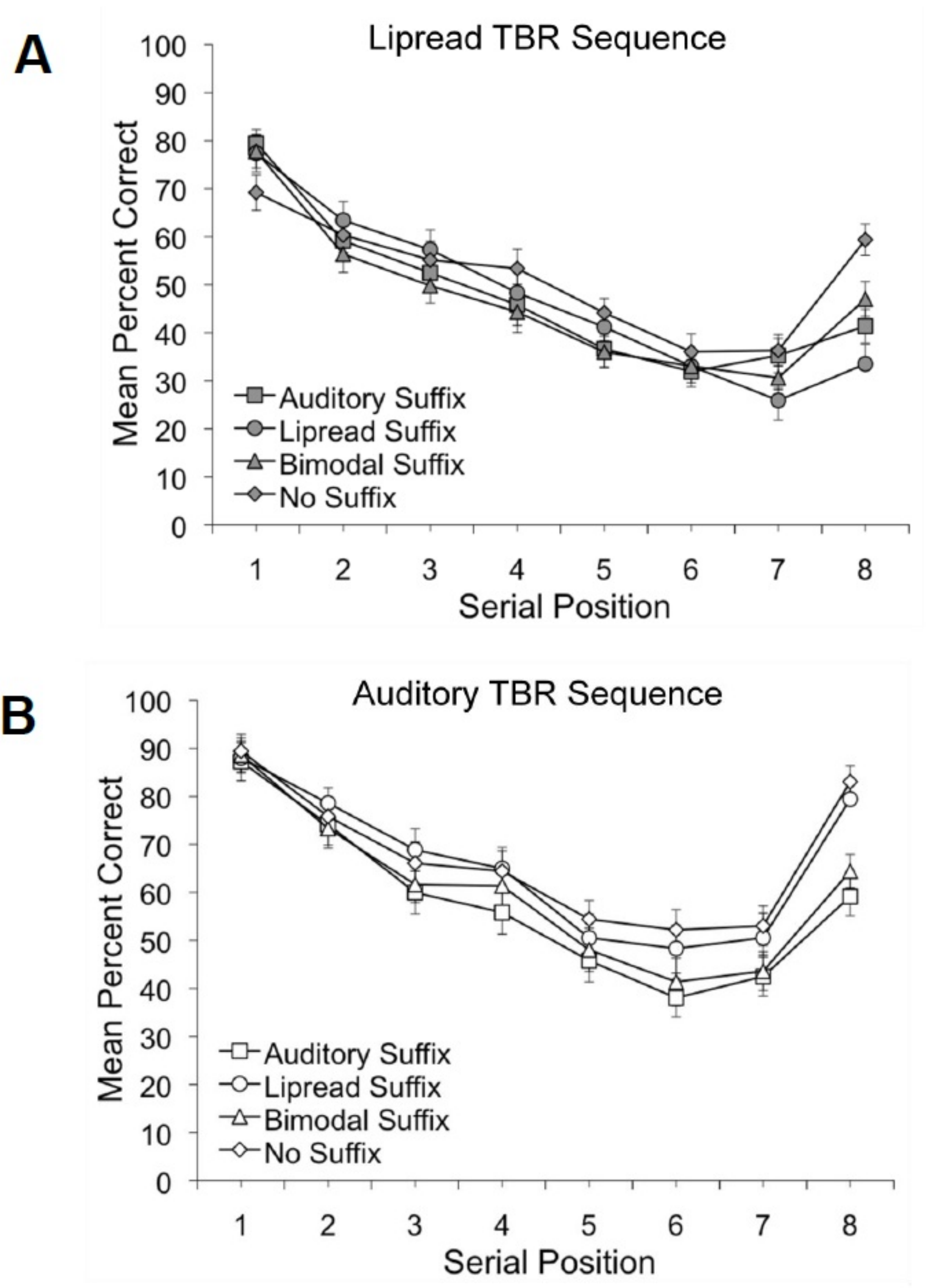

Figure 5. Mean percentage of items correctly recalled for Experiment 2 across all serial positions for lipread, auditory, bimodal, and no-suffix conditions during lipread (Panel A) and auditory (Panel B) to-be-remembered (TBR) sequences.

position, *F*(21,609) = 4.55, *MSE* = 143.84, *p*<.001, η² = .14. Critically, however, there was also an interaction between modality, suffix and serial position, *F*(21,609) = 2.71, *MSE* = 124.11, *p*<.001, η² = .09, such that the effect of each suffix modality differed depending on the modality of the to-be-remembered list. The basis of this is explored in the relative recency analysis.

Relative Recency Analysis

Figure 6 illustrates the results of the relative recency measure and confirms that the effect of each type of suffix differs depending on the modality in which the to-be-remembered sequence is presented. Specifically, to the left of the figure one can see that for lipread presentation, auditory and lipread suffixes reduce lipread recency by almost 20%. Also noticeable is that the drop in recency due to the bimodal suffix is substantially less than for either of the unimodal stimuli which combine to create it. For auditory presentation, shown on the right of Figure 6, the observed suffix effects are different in several ways from this pattern; both auditory and bimodal suffixes reduce recency, while the lipread suffix does not.

This picture was confirmed in a 2x3 repeated measure ANOVA on the relative recency measure which demonstrated a main effect of modality of suffix, *F*(3,87) = 60.98, *MSE* = 207.29, *p*<.001, η² = .68, as well as an interaction between modality of the to-be-remembered sequence and suffix, *F*(3,87) = 20.92, *MSE* = 314.70, *p*<.001, η² = .42, such that relative to no-suffix conditions lipread recency was reduced to an equivalent magnitude in the presence of a lipread suffix, *F*(1,29) = 146.17, *MSE* = 238.26, *p*<.001, η² = .83, and auditory suffix, *F*(1,29) = 154.48, *MSE* = 114.51, *p*<.001, η² = .80, while the effect of a bimodal suffix did not differ significantly (*p*=.10). By comparison, auditory recency was reduced in the presence of an auditory suffix, *F*(1,29) = 7.81, *MSE* = 327.31, *p*<.01, η² =.21, and a bimodal suffix, *F*(1,29) = 4.56, *MSE* = 276.22, *p*<.05, η² = .14, while there was no effect of a lipread suffix when compared to no-suffix trials (*p*=.81).

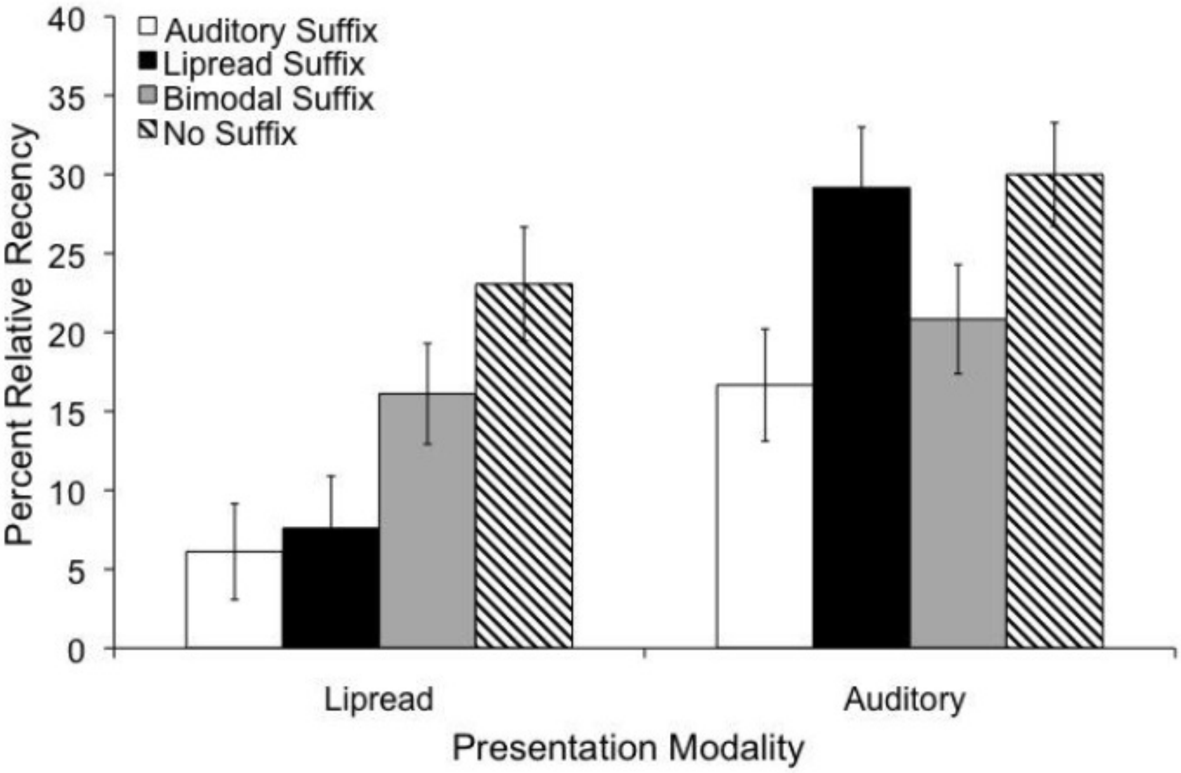

Figure 6. Results of the relative recency measure for Experiment 2 when the last memory item was followed by a lipread, auditory, bimodal, and no-suffix during lipread and auditory presentation modes. Error bars represent the standard error of the mean.

Discussion

The present findings show that, like the precedents in the literature (de Gelder & Vroomen, 1992, 1994), a lipread suffix did not reduce auditory recency. This raises the possibility that a lipread suffix does not gain access to the same representational form as that occupied by the auditory to-be-remembered sequence. The action of the bimodal suffix on a lipread list is also inconsistent with the idea that visual and auditory suffix effects are due to the same mechanisms. The unimodal auditory and lipread suffixes seem equivalent – they both produce equivalent loss of lipread recency. Nevertheless, despite this functional similarity between the action of auditory and lipread suffixes at the end of lipread sequences, a suffix produced by a combination of these two elements, the bimodal suffix, does not reduce lipread recency significantly relative to no-suffix conditions. This then raises the question that if an auditory suffix can interact with a lipread list in much the same way as it does with an auditory list, why is it that a bimodal suffix does not do the same? One possible explanation is that the effect of an auditory suffix on a lipread list, despite the apparent functional similarity to its effect on an auditory list, is actually due to a different mechanism.

While the effect of an auditory suffix on an auditory list is now understood to reside in processes of perceptual streaming (Jones, et al., 2006; Jones, et al., 2004; Maidment & Macken, 2012; Nicholls & Jones, 2002), it may be that the effect of an auditory suffix on a lipread list is not underpinned by the same processes. For lipread lists, the auditory suffix may have act as an attention-capturing event, which, in this case, would be stimulus-driven. There is substantial evidence that the obligatory processing of unexpected auditory events disturbs the efficient execution of the serial memory task by exogenously redirecting attention away from the prevailing focus (e.g., the oddball paradigm, see Hughes, Vachon, & Jones, 2005, 2007; Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001; Vachon, Hughes, & Jones, 2012). If this is true, then it is the status of the auditory suffix as an unexpected acoustic event, rather than its status as a verbal stimulus, which is responsible for its disruptive impact on lipread recency. This possibility is explored in Experiments 3 to 5, where the focus is restricted to the effect of various types of suffix on lipread to-be-remembered sequences, in order better to dissect their underlying mechanisms.

Experiment 3

Thus far, despite there being functional similarities between heard and seen speech stimuli in verbal short-term memory, the findings of Experiments 1 and 2 have identified important distinctions that may undermine the case for a common, amodal form of representation upon which the different modalities of speech converge. Experiment 1 showed that, while auditory recency is immune to interference from articulatory suppression and irrelevant sound, the same is not true for recency in lipread sequences. In Experiment 2, while both lipread and auditory suffixes disrupted recency for lipread sequences, a lipread suffix had no impact on recency for auditory to-be-remembered sequences. These disparities point to a divergence in the mechanisms giving rise to auditory and lipread recency and suffix effects, and therefore point to important distinctions in the nature of the verbal representations derived from those two sources of speech information.

One potential candidate mechanism driving the effect of an auditory suffix on a lipread sequence is attentional capture. That is to say, the auditory suffix acts as a novel stimulus that alters the direction of attention away from current focus because it does not conform to the pattern built-up by the preceding sequence. Critically however, the detection of an unexpected event, and the subsequent attentional reorientation towards that event, involves the organisation of the acoustic environment into perceptual objects, or streams (see e.g., Bregman, 1990; Sussman, 2005). From this viewpoint, the disruptive effect of the auditory suffix on lipread recency could arise because it is registered as an unexpected auditory-verbal event within the otherwise visual-verbal stream. Accordingly, if an alternative auditory stream is provided to perceptually segregate the auditory suffix from the lipread to-be-remembered sequence, this should reduce its ability to detract attention from the processing of the final list items, thereby reducing its effect on recency.

To test this possibility, the effect on serial recall for lipread sequences of the four types of suffix used in Experiment 2 is compared. The critical development in Experiment 3 is that the effect of those suffixes is examined both on their own and in the presence of a to-be-ignored auditory sequence, presented concurrently with the to-be-remembered sequence, which comprised of repetitions of the spoken word "go". This sequence should serve to capture the auditory suffix, binding it into a perceptual stream distinct from the lipread to-be-remembered sequence and thereby abolishing its status as an unexpected auditory event occurring within an unfolding lipread sequence.

Method

Participants

Thirty volunteers (16 female), aged 18 to 26 years (19.80 mean years), recruited from the School of Psychology, Cardiff University, were given course credit for participation. All participants, none of whom had taken part in Experiments 1 and 2, were native English speakers who reported normal (or corrected-to-normal) vision and normal hearing.

Apparatus & Materials

The apparatus and materials deployed in Experiment 2 were again used in Experiment 3.

Design & Procedure

The design involved a repeated measures factorial combination of Concurrent Auditory Sequence (CAS) (absent, present), Modality of Suffix (lipread “go”, auditory “go”, bimodal “go”, no-suffix), and Serial Position. Twelve sequences were presented for each of the four suffix x CAS conditions, arranged in a pseudo-random order. The 5s introductory period prior to the onset of the first to-be-remembered item was filled with either silence in the absence of the CAS or, when the CAS was present, the word “go” was repeatedly heard every 240ms continuing without break in tempo during presentation of the memory sequence. The suffix also occurred in tempo with this preceding CAS, after which participants were visually cued to recall (see Figure 7).

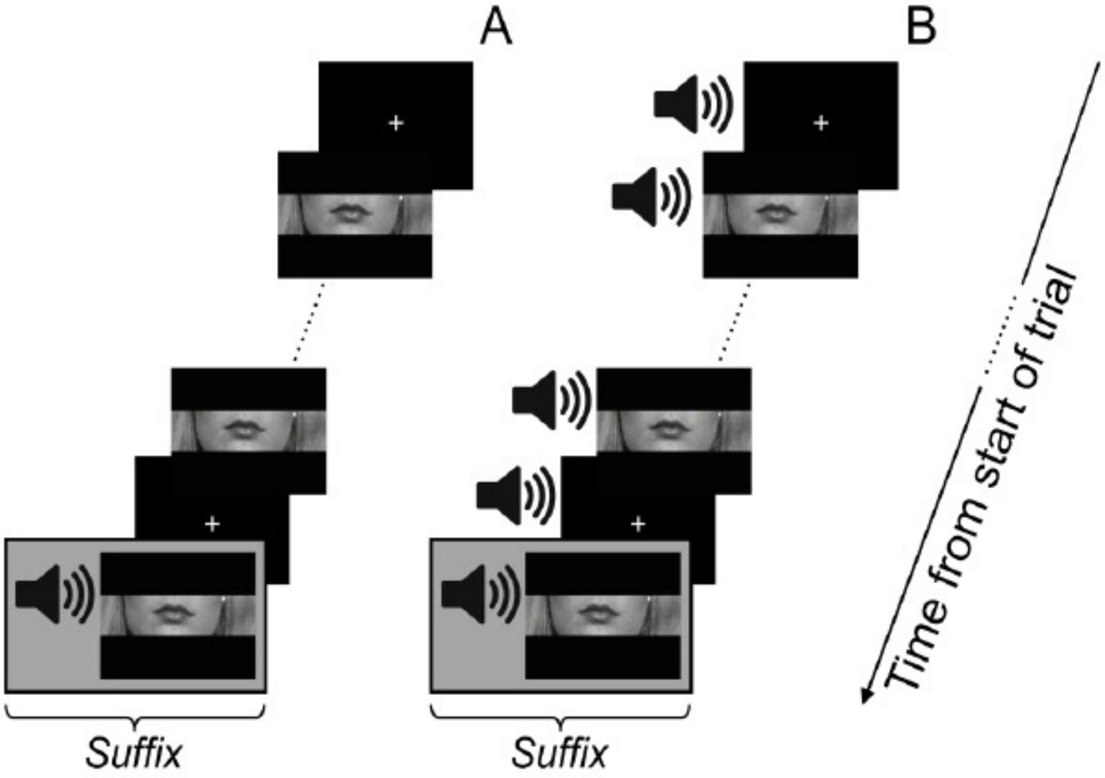

Figure 7. A schematic representation of the sequencing and relative timing of stimuli presented in Experiments 3 to 5. Figure does not preserve the scale of item duration or series length. Panel A shows suffix trials in the absence of a CAS, with Panel B indicating its presence.

Results

Pre-screening scores ranged from 85.19% to 100% (mean = 93.09%) lipread items correctly identified.

Serial Position Analysis

Figure 8 shows the percentage of correct recall in each suffix condition as a function of serial position in the absence (Figure 8a) and presence (Figure 8b) of a CAS. A 2 (CAS) x 4 (Suffix) x 8 (Serial Position) repeated measure ANOVA on the percentage of items correctly recalled across all serial positions demonstrated that serial recall performance was reduced in the presence of a CAS, shown as a main effect, *F*(1,29) = 14.93, *MSE* = 374.19, *p*<.001, η^2 = .34. Serial recall performance was also reduced in the presence of a suffix, *F*(3,87) = 5.33, *MSE* = 375.45, *p*<.01, η^2 = .16, although the disruptive effect of a suffix was restricted to the last serial position within the lipread memory list, confirmed by a main effect of serial position, *F*(7,203) = 70.77, *MSE* = 784.01, *p*<.001, η^2 = .71, as well as an interaction between suffix and serial position, *F*(21,609) = 3.26, *MSE* = 157.71, *p*<.001, η^2 = .10. All other interactions failed to reach significance (*p*>.05).

Relative Recency Analysis

The results of the relative recency measure are shown in Figure 9a. First, to the left the results are shown for suffix effects with no CAS. Replicating the results of Experiment 2, an auditory suffix and a lipread suffix each produce a loss of recency that is roughly the same magnitude. Quantitatively the two suffixes seem equivalent in their action. Also noticeable, however, is that the drop in recency due to the bimodal suffix is considerably less than for either of the unimodal stimuli which combine to create it.

The effect of the CAS on recency, shown on the right of Figure 9a, is different in several ways from this pattern. As expected, the CAS serves to capture the auditory suffix, perceptually segregating it from the lipread sequence and thereby restoring recency, although not fully to the extent found without any suffix. Clearly this is not the case with the lipread suffix; recency is at roughly the same low level as it is in the CAS absent condition. The effect of the CAS on the bimodal suffix is also clear-cut; it makes the bimodal suffix substantially more disruptive of recency relative to the effect of that suffix in the absence of a CAS, and relative to no-suffix conditions.

This pattern was confirmed by a 2x4 repeated measure ANOVA on the relative recency measure which demonstrated a main effect of suffix, *F*(3,87) = 7.19, *MSE* = 201.71, *p*<.001, η^2 = .21, but no main effect of CAS (*p*=.94). Critically, there was an interaction between suffix and CAS, *F*(3,87) = 4.267, *MSE* = 221.80, *p*<.01, η^2 = .13, such that when there was no CAS, lipread recency was reduced by a lipread suffix, *F*(1,29) = 7.81, *MSE* = 327.31, *p*<.01, η^2 = .21, and auditory suffix, *F*(1,29) = 7.92, *MSE* = 336.67, *p*<.01, η^2 = .22, while the bimodal suffix did not reduce recency relative to no-suffix conditions (*p*=.25). However, the loss of recency produced by an auditory suffix when the CAS was absent was reduced significantly in the presence of a CAS, *F*(1,29) = 10.05, *MSE* = 125.36, *p*<.01, η^2 = .26, whereas the lipread suffix still produced a loss of recency of the same magnitude (*p*=.73). The effect of a CAS on a bimodal suffix reduced recency, *F*(1,29) = 7.13, *MSE* = 234.52, *p*<.05, η^2 = .20, and to an identical magnitude as the lipread suffix in the same context (*p*=.65).

Discussion

In the absence of a CAS, the unimodal auditory and lipread suffixes seem equivalent – they both produce equal loss of recency to lipread lists. However, they behave differently in the presence of a CAS. The auditory suffix effect is very much reduced in the context of a CAS, whereas the lipread suffix effect is unaltered. Experiments 4 and 5 will return to the anomaly that, while the lipread suffix and the auditory suffix appear to be functionally equivalent, they are not equally susceptible to the influence of a CAS.

Also apparently anomalous is the action of the bimodal suffix; while it leads to no significant impediment to recency in the absence of a CAS, the addition of the CAS leads the bimodal suffix to impair recency to the same extent as a lipread suffix. One way to account for this is that the auditory component of the bimodal suffix becomes captured by the CAS, thereby isolating the visual component and enabling it to behave in the same way as the unimodal lipread suffix. While this would account for the effect of a CAS in determining the disruptive potency of a bimodal suffix, it does not explain why a bimodal suffix on its own does not significantly affect recency given that each of its components in unimodal form do so. Furthermore, precisely how the lipread suffix has its effect on lipread recency cannot be pinpointed. Although each of these issues will be resolved in Experiments 6 and 7, for the time being the nature of the auditory suffix effect on lipread lists will be established.

If the effect of an auditory suffix on lipread recency is, unlike the typical suffix effect found with auditory sequences, due to attentional capture, then any unexpected auditory event – verbal or otherwise – occurring immediately after the lipread sequence should produce a suffix (or suffix-like) effect. This is tested in Experiment 4 by utilising a nonverbal, noise burst suffix in place of the auditory “go” used in previous experiments.

Experiment 4

Here, the same broad methodology as Experiment 3 is adopted, with the exception that now the auditory suffix stimulus is nonverbal – a broadband noise burst. In all other respects, the approach was the same in that the to-be-remembered sequences were lipread, and the CAS, when present, was a series of the spoken word "go". If previous demonstrations of an effect of an auditory-verbal suffix on lipread recency depend critically on the verbal content of the suffix, then such a noise burst should have no impact. If, on the other hand, the auditory-verbal suffix has its effect because of the attentional capture elicited by an unexpected auditory event, then it should be possible to see reduced lipread recency with a nonverbal suffix, as long as it constitutes an unexpected auditory event occurring at the end of the memory sequence.

Method

Participants

Thirty volunteers (20 female), aged 19 to 32 years (22.73 mean years), recruited from Cardiff University's School of Psychology, were paid for their participation. All participants, none of whom had taken part in Experiment 1 to 3, were native English speakers reporting normal (or corrected-to-normal) vision and normal hearing.

Apparatus & Materials and Design & Procedure

In all respects, the stimuli and conditions presented to participants in Experiment 4 were identical to those in Experiment 3 except that a white noise burst replaced the auditory “go” suffix (but not the “go” CAS). The noise burst was matched subjectively to the spoken “go” suffix in terms of intensity and duration. For the bimodal suffix, using *Final Cut Express* (version 4.0.1, 2002-2008 Apple Inc.), the noise burst was positioned at exactly the same temporal position as the deleted “go” sound so that it was synchronised with the onset/offset of the visual “go” lip movements.

Results

Pre-screening scores ranged from 85.19% to 100% (mean = 90.31%) lipread items correctly identified.

Serial Position Analysis

Figure 8 shows the percentage of correct recall in each suffix condition as a function of serial position in the absence (Figure 8c) and presence (Figure 8d) of a CAS. A 2x4x8 repeated measure ANOVA confirmed that there were main effects of CAS, *F*(1,29) = 6.32, *MSE* = 539.82, *p*<.02, η^2 = .18, suffix, *F*(3,87) = 3.52, *MSE* = 436.54, *p*<.02, η^2 = .11, and serial position, *F*(7,203) = 93.07, *MSE* = 632.79, *p*<.001, η^2 = .76. There was also an interaction between suffix and serial position, *F*(21,609) = 3.72, *MSE* = 151.56, *p*<.001, η^2 = .11, such that the disruptive effect of the suffix was restricted to the last position within the lipread memory list. All other interactions failed to reach significance (*p*>.05).

Relative Recency Analysis

The effect of suffix type and CAS on the relative recency measure is depicted in Figure 9b. In the absence of a CAS, the drop in recency due to the auditory noise burst suffix is minimal compared to the lipread or bimodal suffixes. This is clearly not the case, however, when the same noise burst suffix is presented in the context of a CAS (capturing “go” sequence). In this case, its impact on recency is of equivalent magnitude to that of the lipread suffix. The other clear aspect of the pattern relates to the bimodal (lip movement paired with noise burst) suffix; while the auditory-verbal version of the bimodal suffix had minimal impact on its own, only becoming disruptive of recency in the context of the CAS, the nonverbal version leads to disruption equivalent to that found with a unimodal lipread suffix regardless of the presence of the CAS.

This pattern was borne out in a 2x4 repeated measure ANOVA which demonstrated a main effect of suffix, *F*(3,87) = 13.50, *MSE* = 203.01, *p*<.001, η^2 = .32, but no main effect of CAS (*p*=.30) or an interaction between suffix and CAS (*p*=.93), such that in the absence of a CAS recency was reduced to an equivalent magnitude in the presence of a lipread, *F*(1,29) = 10.22, *MSE* = 342.63, *p*<.01, η^2 = .26, and bimodal suffix, *F*(1,29) = 10.82, *MSE* = 278.14, *p*<.01, η^2 = .27, while the effect of an auditory noise-burst suffix did not differ from no-suffix conditions (*p*=.28). When the CAS was present, the auditory noise burst suffix reduced recency, *F*(1,29) = 5.92, *MSE* = 187.94, *p*<.02, η^2 = .17, to a degree identical to that of both the lipread (*p*=.46) and bimodal suffixes (*p*=.74).

Discussion

It is worth comparing these results in terms of how they differ from those found in Experiment 3, where the auditory element of the suffixes was verbal. Firstly, in the absence of the CAS, the noise burst suffix has minimal impact on recency compared to the auditory-verbal suffix of Experiment 3, which reduced recency as much as the lipread suffix did. However, a noise burst suffix at the end of a verbal CAS did reduce recency. Thus, a nonverbal auditory event can produce the same disruptive effect as a verbal auditory event, so long as it constitutes an unexpected auditory event relative to a predictive model of the preceding environment – in this case, provided by the CAS series of the spoken word “go”.

From this perspective, the CAS plays a critical role. This is because attentional capture is due, not to acoustic novelty, but the violation of a predictive model of the preceding environment. That is to say, attentional capture arises as a consequence of the unexpectedness of the acoustic event relative to perceptual expectations derived from the preceding context, and not its verbal content. So, for example, a repeated sound in an otherwise predictably changing acoustic sequence (e.g., *a,b,a,b,a,b,***b***,a…*) produces attentional capture and concomitant disruption to a focal task (e.g., Hughes, et al., 2007). Such an account of attentional capture is consistent with the effect of an auditory-verbal (i.e., “go”) suffix on lipread recency; the auditory element constitutes as the unexpected event within a previously homogenous visual-verbal stream of digits. In other words, an auditory-verbal suffix that occurs at the end of a visual-verbal sequence constitutes an unexpected and therefore attention-capturing change. In comparison, a noise burst occurring at the end of a lipread sequence, without any preceding acoustic context, does not lead to attentional capture because it does not violate expectations. Rather, in this context a noise-burst suffix is merely novel. However, the nonverbal suffix elicits attentional capture when it violates the expectations built-up by the preceding CAS of the word “go”. Attention is then drawn away from the focal task, thereby reducing lipread recency that would otherwise be evident if the suffix was also the word "go".

The second key contrast between the effect of a verbal and nonverbal stimulus relates to the effect of the bimodal suffix. Despite robust effects with either the auditory or the visual elements on their own, when the bimodal suffix contains congruent auditory-verbal and visual-verbal elements it does not disrupt recency, while in the presence of a CAS it does (see Experiment 3). This latter finding may be due to the CAS perceptually capturing the auditory element, removing it from the otherwise bound auditory-visual percept and allowing the isolated visual element to act in the same way as the unimodal lipread suffix. The disruptive impact in Experiment 4 of the bimodal suffix, regardless of the presence of a CAS, supports this account since the noise burst is unlikely to be bound with the lipread element to form a unified bimodal percept in the first place (Bernstein, Auer Jr, & Moore, 2004; Schwartz, Grimault, Hupé, Moore, & Pressnitzer, 2012; Vatakis & Spence, 2007).

Taken together, the pattern of results for bimodal suffixes across Experiments 3 and 4 provides some explanation of the absence of a suffix effect for the auditory-verbal and visual-verbal bimodal suffix; the binding of auditory and visual elements of a bimodal verbal suffix somehow explains the lack of its disruptive effect. This will be investigated further in Experiments 6 and 7, but Experiment 5 further tests the idea that the impact of an auditory suffix on a lipread list can be attributed to attention capture.

Experiment 5

The results of Experiments 3 and 4 suggest that an attentional capture mechanism gives rise to the effect of an auditory-verbal suffix on a lipread list. The finding that lipread recency is similarly disrupted by a nonverbal suffix, so long as it constitutes an unexpected and therefore attention-capturing change, supports this. Namely, the status of the noise burst as an unexpected event depends on the preceding acoustic context being uniformly verbal, as with the “go” CAS in Experiment 4. On this basis, if the terminal noise burst no longer violates the predictive model built-up by the preceding acoustic context it should no longer disrupt recency. To test this, the broad methodology of Experiment 4 was again adopted, but this time, the CAS was a sequence of noise bursts rather than the spoken word “go”.

Method

Participants

Thirty volunteers (23 female), aged 18 to 58 years (22.65 mean years), recruited from Cardiff University’s School of Psychology, were paid for their participation. All participants, none of whom had taken part in Experiment 1 to 4, were native English speakers reporting normal (or corrected-to-normal) vision and normal hearing.

Apparatus & Materials and Design & Procedure

The same sample of lipread digits and suffixes used in Experiment 4 were presented in Experiment 5. This time, however, the CAS comprised repetitions of the auditory noise burst stimulus used in Experiment 4.

Results

Pre-screening scores ranged from 85.19% to 98.15% (mean = 89.69%) lipread items correctly identified.

Serial Position Analysis

Figure 8 shows the percentage of correct recall in each suffix condition as a function of serial position in the absence (Figure 8e) and presence (Figure 8f) of a CAS. A 2x4x8 repeated measure ANOVA confirmed that there were main effects of suffix, *F*(3,87) = 5.57, *MSE* = 302.84, *p*<.01, η^2 = .16, and serial position, *F*(7,203) = 80.80, *MSE* = 632.36, *p*<.001, η^2 = .74, but no effect of CAS (*p*>.05). There was also an interaction between suffix and serial position, *F*(21,609) = 5.12, *MSE* = 151.61, *p*<.001, η^2 = .16, such that the disruptive effect of the suffix was restricted to the last one or two serial positions within the lipread memory list. All other comparisons failed to reach significance (*p*>.05).

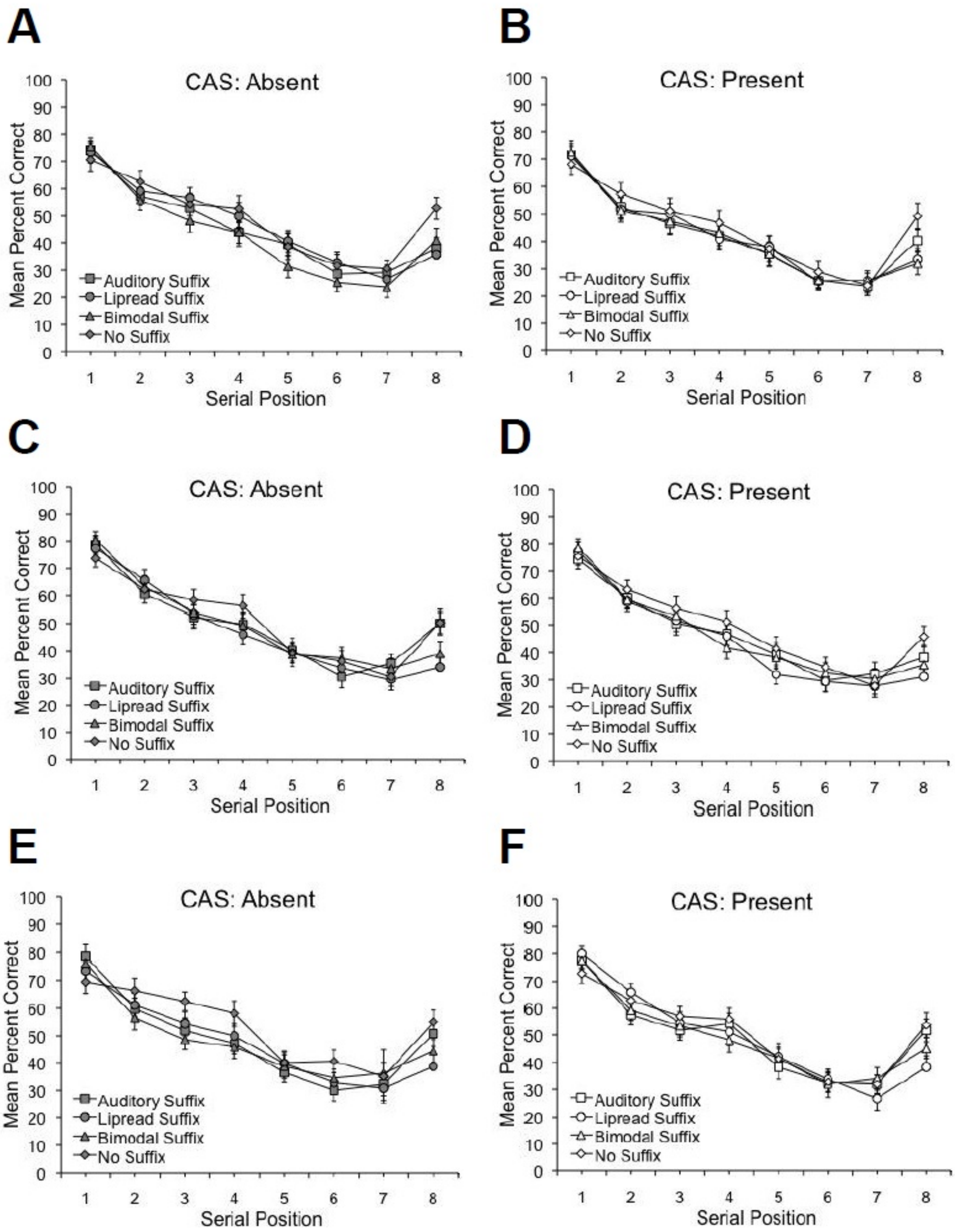

Figure 8. Mean percentage of lipread items correctly recalled across all serial positions when the last memory item was followed by an auditory, lipread, bimodal, or no-suffix, in the absence and presence of a concurrent auditory sequence (CAS) for Experiment 3 (Panel A and B), Experiment 4 (Panel C and D) and Experiment 5 (Panel E and F).

Relative Recency Analysis

The results are in line with the idea that the effect of the noise burst shown in Experiment 4 is due to attentional capture (see Figure 9c). If the noise burst suffix appears at the end of a sequence of similar sounds, its effect is the same as that found when it is presented in isolation, namely, that there is no-suffix effect.

A 2x4 repeated measure ANOVA demonstrated that there was a main effect of suffix, *F*(3,87) = 5.77, *MSE* = 312.45, *p*<.001, η^2 = .17, but no main effect of CAS (*p*=.41) or an interaction between suffix and CAS (*p*=.93), such that in the absence of a CAS recency was reduced to an equivalent magnitude in the presence of a lipread, *F*(1,29) = 7.44, *MSE* = 301.09, *p*<.02, η^2 = .20, and bimodal suffix, *F*(1,29) = 7.01, *MSE* = 291.19, *p*<.02, η^2 = .20, while the effect of the auditory noise burst suffix did not differ from no-suffix conditions (*p*=.64). Critically, in the presence of the CAS the same trends were observed; the effect of an auditory suffix did not differ from no-suffix conditions (*p*=.95), while recency performance was reduced in the presence of a lipread, *F*(1,29) = 4.14, *MSE* = 251.44, *p*<.05, η^2 = .13, and bimodal suffix, *F*(1,29) = 6.43, *MSE* = 184.23, *p*<.02, η^2 = .18.

Discussion

In comparison to Experiment 4, Experiment 5 demonstrated that if the terminal noise burst does not violate the predictive model built-up by a preceding stream of noise burst sounds it no longer disrupts recency. Previous demonstrations that an auditory-verbal suffix disrupts recency for lipread lists, just as it does for auditory lists, has been taken, in general terms, to point to a shared form of verbal

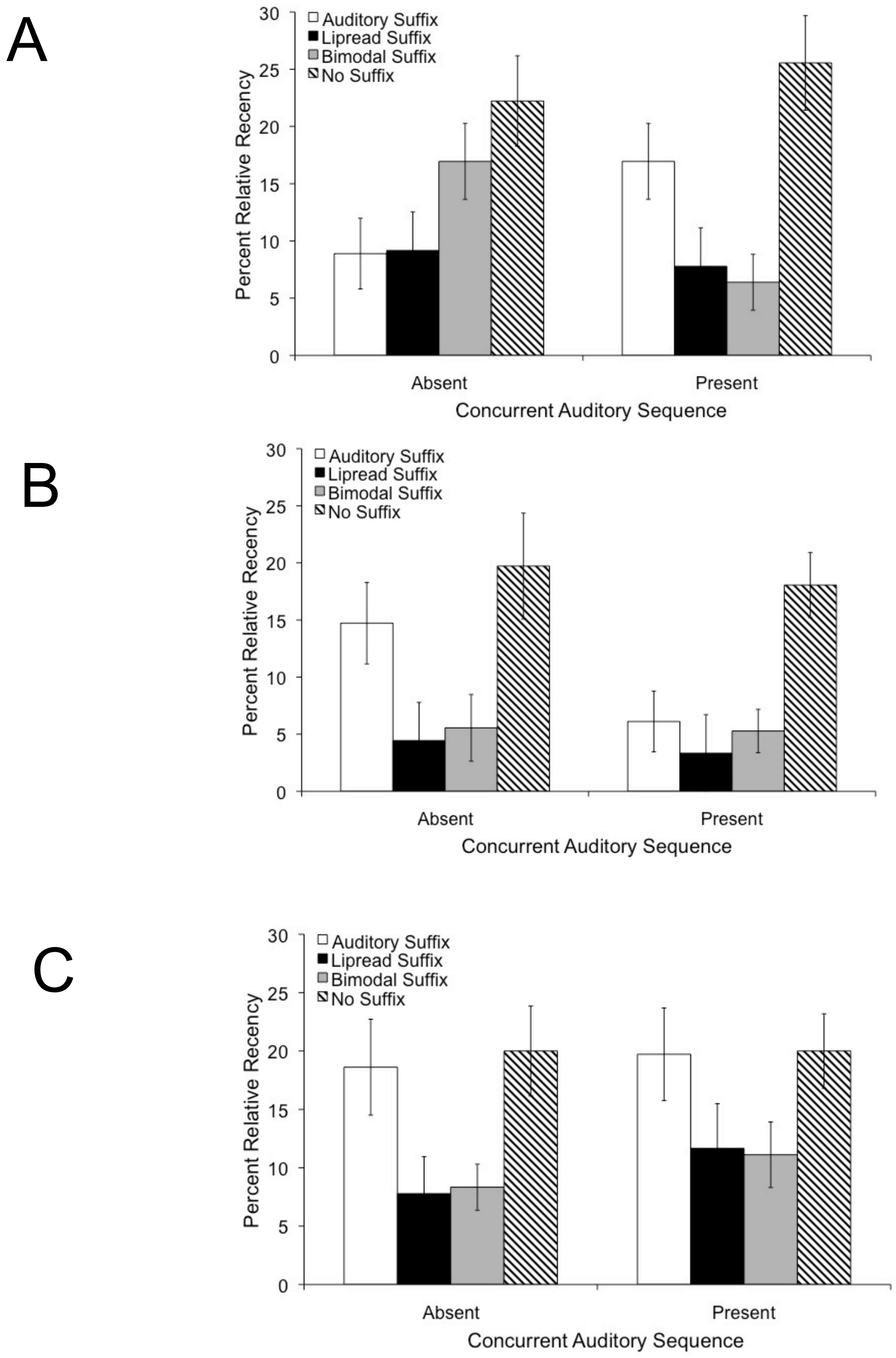

Figure 9. Measure of relative recency for auditory, lipread, bimodal, and no-suffix conditions in the absence and presence of a concurrent auditory sequence for Experiment 3 (Panel A), Experiment 4 (Panel B) and Experiment 5 (Panel C). Error bars represent the standard error of the mean.

representation (Campbell & Dodd, 1980, 1982; de Gelder & Vroomen, 1992; Gathercole, 1987; Greene & Crowder, 1984; Spöehr & Corin, 1978). However, the results of Experiments 3, 4 and 5 point to a different mechanism underpinning the disruptive effect of an auditory-verbal suffix on lipread recency; one that resides within general processes of attentional capture as opposed to interactions amongst central verbal representations. On this basis, an auditory-verbal suffix disrupts lipread recency because it constitutes an unexpected auditory-verbal event within the otherwise homogeneously visual-verbal context.

This is a fundamentally different mechanism to that underpinning the effect of an auditory suffix on auditory recency, which is one of perceptual organisation – the suffix becomes part of the perceptual stream constituting the to-be-remembered sequence, thereby displacing the terminal memory item from its perceptually privileged position at the end of the stream (Nicholls & Jones, 2002; Warren, 1999). Indeed, as the auditory suffix becomes less perceptually similar (and therefore less 'expected') to the to-be-remembered sequence, its disruptive impact on recency declines (e.g., Crowder, 1971; Frankish & Turner, 1984; Frick, 1988; Greene, 1991; Morton, et al., 1971) indicating that it is not an effect of attentional capture but of perceptual grouping. A broadband noise burst suffix, which is very dissimilar to the lipread memory sequence, disrupts lipread recency when it occurs unexpectedly against an auditory-verbal context (Experiment 4), and to a similar extent to that found with an unexpected auditory-verbal suffix (Experiment 3). Critically however, if the auditory-verbal suffix is deprived of its attention-capturing properties by preceding it with a capturing sequence of auditory tokens similar to the suffix itself, its disruptive effect is eliminated (Experiment 3 and 5). The importance of this is that, despite its capacity to disrupt recency in both auditory and lipread to-be-remembered sequences, the mechanism whereby it does so is distinct in each case. As such, the impact of an auditory suffix on lipread and auditory lists is not evidence of a shared, amodal form of representation between the two modalities, but rather of modality-general, attentional capturing processes in the former, and modality-specific perceptual organisation processes in the latter.

Having provided an account of the mechanism underpinning the effect of an auditory suffix on lipread lists, attention is now redirected to understanding the action of lipread and bimodal suffixes. As noted above, the pattern reported so far is perplexing; while both unimodal auditory and lipread suffixes disrupt lipread recency, the bimodal suffix does not. Experiments 6 and 7 seek to resolve this.

Experiment 6

The disruptive effect of an auditory suffix on a lipread to-be-remembered sequence is due to attentional capture. What, then, is the mechanism whereby a lipread suffix disrupts lipread recency? To begin to answer this question, Experiment 6 starts from the robust finding in auditory recency that the similarity between the auditory suffix and the to-be-remembered sequence is a key determinant of its disruption – the less similar the suffix to the list, the less it disrupts recency. If similar processes operating on similar representations are also involved with lipread material, then it might be expected that the perceptual similarity of the suffix to the sequence modulate its effect. To this end, Experiments 6a and 6b manipulated the perceptual properties of the lipread suffix to determine whether this is indeed the case.

Experiment 6a

If the disruptive effect of a lipread suffix on lipread recency depends upon its perceptual similarity and consequent grouping of the suffix with the sequence (as does the effect of an auditory suffix on auditory recency), then reducing that perceptual similarity should reduce the size of the suffix effect. Obviously, in the absence of any acoustic content, identical manipulations to those employed in the auditory setting cannot be implemented. However, reversing the contrast of the black and white video between the sequence and the suffix could be argued to be a reasonable analogue to altering a feature such as voice or pitch. The effect of this manipulation was tested in Experiment 6a.

Method

Participants

Thirty volunteers (24 female), aged 18 to 25 years (19.07 mean years), recruited online from Cardiff University's School of Psychology, were given course credit for their participation. All participants, none had taken part in Experiments 1 to 5, were native English speakers reporting normal (or corrected-to-normal) vision and normal hearing.

Apparatus & Materials

The same lipread digits and the lipread “go” suffix used in Experiments 2 to 5 were adopted for Experiment 6a. Using *Final Cut Express* software, the lipread “go” suffix was reversed in contrast by switching the black and white areas on the display.

Design & Procedure

The design involved a repeated measures factorial combination of Suffix Type (standard lipread “go”, reversed contrast lipread “go”, no-suffix) and Serial Position. Twelve sequences were presented for each of the three suffix conditions, arranged in a pseudo-random order. All other aspects of the procedure were identical to that deployed in Experiments 2 to 5.

Results

Pre-screening scores ranged from 85.19% to 100% (mean = 92.90%) lipread items correctly identified.

Serial Position Analysis

Figure 10 shows the percentage of correct recall for each suffix type as a function of serial position. A 3x8 repeated measure ANOVA demonstrated that there were main effects of suffix, *F*(2,58) = 12.89, *MSE* = 457.62, *p*<.001, η^2 = .31, and serial position, *F*(7,203) = 61.29, *MSE* = 282.49, *p*<.001, η^2 = .68, as well as an interaction between suffix and serial position, *F*(14,406) = 3.20, *MSE* = 150.63, *p*<.001, η² = .10, such that, relative to no-suffix conditions, recency performance was depressed in the presence of a standard lipread suffix, *F*(1,29) = 16.22, *MSE* = 575.53, *p*<.001, η² = .36, and reversed contrast suffix, *F*(1,29) = 21.45, *MSE* = 388.53, *p*<.001, η² = .62, the effects of which did not differ (*p*=.79).

Relative Recency Analysis

The results for Experiment 6a (Figure 11) are clear in that, although the reversed contrast clearly signalled to the participant that the suffix was the last, non-memory item, the suffix effect was nevertheless undiminished. A repeated measure ANOVA on the percentage of items correctly recalled at the terminal position relative to the preterminal item confirmed that there was a main effect of suffix, *F*(2,58) = 4.25, *MSE* = 198.44, *p*<.05, η² = .13. Post hoc comparisons further revealed that, relative to no-suffix conditions recency was reduced in the presence of a standard lipread, *F*(1,29) = 4.76, *MSE* = 248.88, *p*<.05, η² = .14, as well as a reversed contrast suffix, F(1,29) = 8.25, MSE = 162.20, p<.01, η² = .22, the effects of which did not differ (p=.88).

Discussion

Again, it appears that the lipread suffix effect on lipread recency is different from the apparently analogous effect in audition; altering the perceptual similarity between the sequence and the suffix has no impact on its disruptive capacity. It could be argued, however, that reducing the superficial, physical similarity between

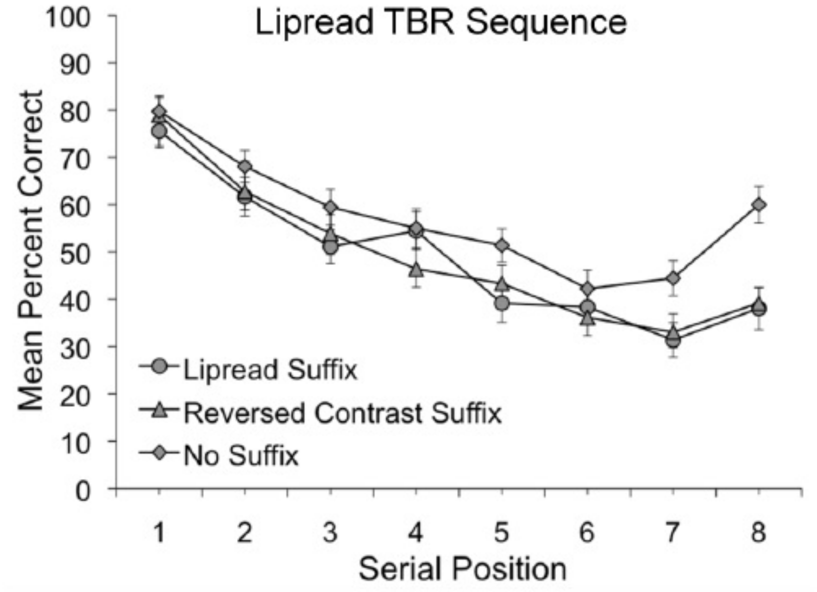

Figure 10. Mean percentage of items correctly recalled for Experiment 6a across all serial positions for standard lipread, reversed contrast and no-suffix trials during lipread to-be-remembered (TBR) sequence presentation.

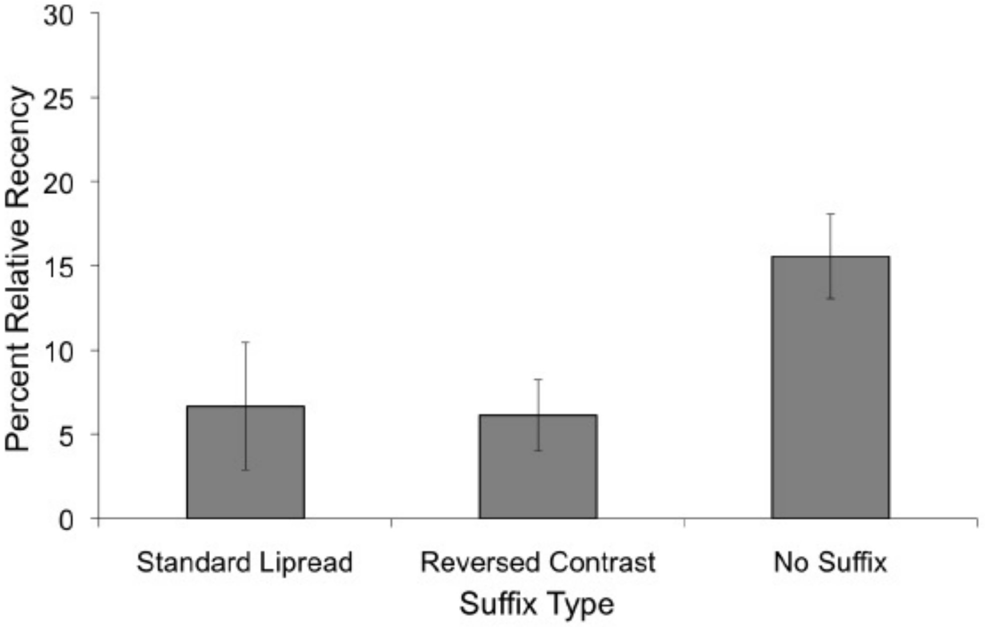

Figure 11. Outcome of the relative recency measure for Experiment 6a. Scores are expressed in terms of percentage correct for standard, reversed contrast and no-suffix trials. Standard error bars are shown.

suffix and sequence might not be the best way to investigate parallels between visual-verbal and auditory-verbal materials. That is to say, the mere contrast reversal does not constitute a fair comparison with the types of perceptual change that have been implemented in the auditory domain (voice, pitch etc.). As a result, it might still be the case that the critical interaction between the suffix and the sequence resides at the amodal verbal level typically proposed to account for cross-modality similarities. This possibility was explored further in Experiment 6b.

Experiment 6b

In Experiment 6b the suffix was modified in ways that should alter its verbal (i.e., lexical) identity in order to reduce the similarity between the lip movements in the suffix and those in the sequence. In particular, the effect of standard lipread and auditory “go” suffixes were contrasted with suffixes constructed by either temporally reversing the video so that it was played backward or by rotating it through 180-degrees to render it upside-down.

Method

Participants

Thirty volunteers (25 female), aged 18 to 23 years (20.05 mean years), recruited online from Cardiff University’s School of Psychology, were given course credit for their participation. All participants, none had taken part in Experiments 1 to 6a, were native English speakers reporting normal (or corrected-to-normal) vision and normal hearing.

Apparatus & Materials

Using *Final Cut Express* software, a backward lipread suffix was achieved by editing the original "go" image so that it was played backwards, while an inverted suffix was created by rotating the original version of the "go" item through 180-degrees so that it appeared upside-down (see Figure 12).

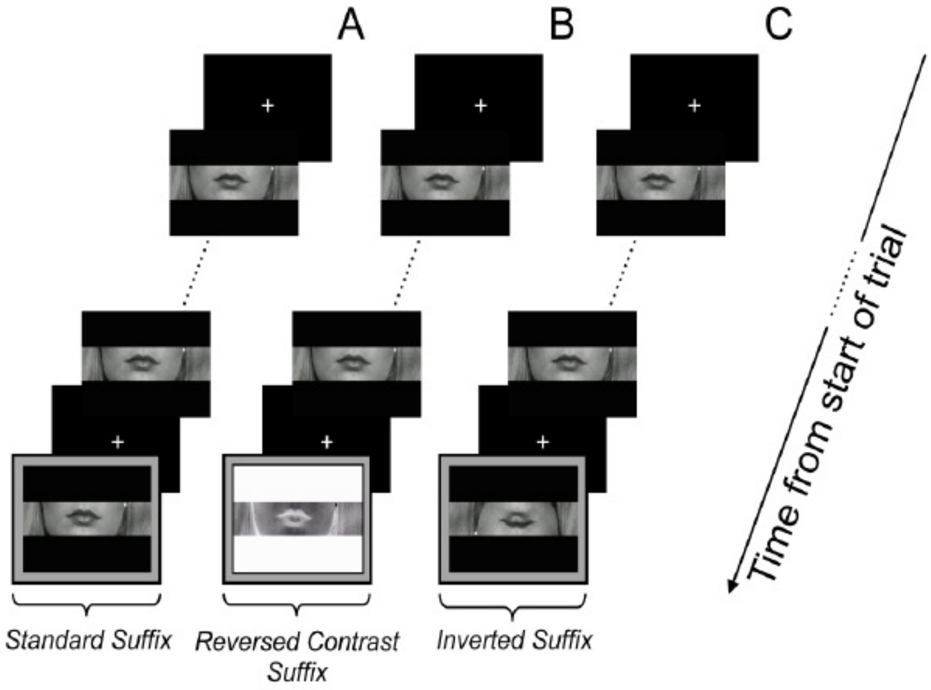

Figure 12. A schematic representation of the sequencing and relative timing of stimuli presented in Experiments 6. Figure does not preserve the scale of item duration and series length. Panel A shows standard lipread "go" suffix, Panel B reversed contrast lipread "go" suffix (Experiment 6a), and Panel C inverted lipread "go" suffix (Experiment 6b).

Design & Procedure

The design involved a repeated measures factorial combination of Suffix Type (no-suffix, standard lipread “go”, backward lipread “go”, inverted lipread “go”, auditory “go”) and Serial Position. Twelve sequences were presented for each of the five suffix conditions, arranged in a pseudo-random order. All other aspects of the procedure were identical to that deployed in Experiment 6a.

Results

Pre-screening scores ranged from 85.19% to 100% (mean = 92.56%) lipread items correctly identified.

Serial Position Analysis

Figure 13 shows the percentage of correct recall in each suffix condition as a function of serial position. A 5x8 repeated measure ANOVA confirmed that there were main effects of suffix, *F*(4,116) = 11.94, *MSE* = 387.78, *p*<.001, η^2 = .29, and serial position, *F*(7,203) = 75.08, *MSE* = 366.98, *p*<.001, η^2 = .72. There was also an interaction between suffix and serial position, *F*(28,812) = 2.96, *MSE* = 149.85, *p*<.001, η^2 = .09, such that, relative to no-suffix conditions, recency was depressed in the presence of a standard lipread suffix, *F*(1,29) = 38.59, *MSE* = 315.29, *p*<.001, η^2 = .57, backward suffix, *F*(1,29) = 29.83, *MSE* = 277.02, *p*<.001, η^2 = .51, and auditory suffix, *F*(1,29) = 26.49, *MSE* = 425.09, *p*<.001, η^2 = .48, the effects of which did not differ (*p*>.05). Just as notable was the absence of any effect for an inverted lipread suffix, with recency performance identical to that of the no-suffix control condition (*p*=.65), and significantly different from both a standard lipread suffix, *F*(1,29) = 17.21, *MSE* = 300.27, *p*<.001, η² = .37, backward suffix, *F*(1,29) = 6.25, *MSE* = 440.79, *p*<.05, η² = .18, and auditory suffix conditions, *F*(1,29) = 11.90, *MSE* = 385.212, *p*<.01, η² = .29.

Relative Recency Analysis

The results of Experiment 6b (Figure 14) are clear-cut, with each method rendering the suffix unintelligible – playing the digital recording of the suffix backwards or spatially inverting it so that the lips are upside-down – having a different effect. When compared to no-suffix conditions, recency was reduced in the presence of a standard lipread suffix, *F*(1,29) = 4.83, *MSE* = 622.97, *p*<.05, η² = .14, backward suffix, *F*(1,29) = 5.81, *MSE* = 538.63, *p*<.05, η² = .17, and an auditory "go" suffix, *F*(1,29) = 10.38, *MSE* = 290.11, *p*<.01, η² = .26, the effects of which did not differ (*p*>.05). Just as notable was the absence of any effect for an inverted lipread suffix relative to the no-suffix condition (*p*=.65), which was also significantly different from the disruptive effects of a standard lipread suffix, *F*(1,29) = 5.33, *MSE* = 401.54, *p*<.05, η² = .16, a backward suffix, *F*(1,29) = 5.07, *MSE* = 442.37, *p*<.05, η² = .15, as well as the auditory suffix, *F*(1,29) = 5.67, *MSE* = 377.59, *p*<.05, η² = .16, at recency.

Discussion

Here it is shown that presenting a lipread "go" suffix rotated 180-degrees eliminates its effect on lipread recency. This might be taken to point to evidence that

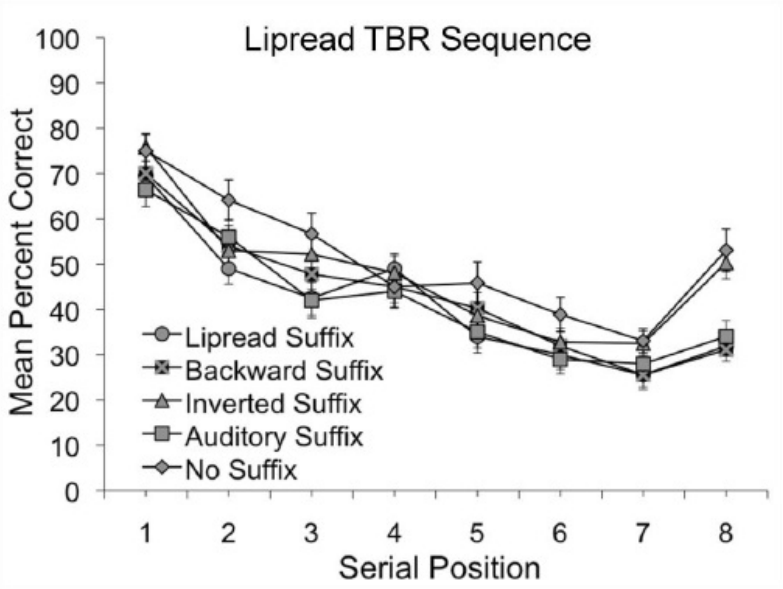

Figure 13. Mean percentage of items correctly recalled for Experiment 6b across all serial positions for standard lipread, backward, inverted, auditory and no-suffix conditions during lipread to-be-remembered (TBR) sequence presentation.

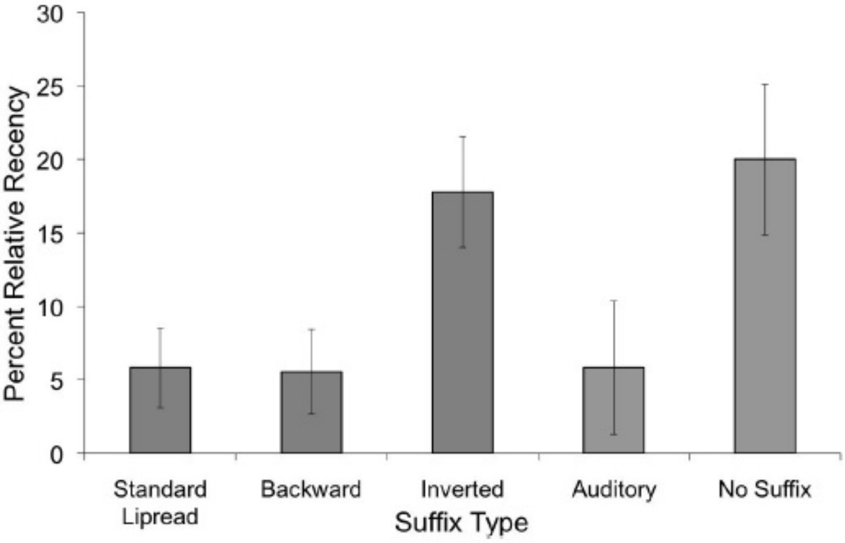

Figure 14. Outcome of the relative recency measure for Experiment 6b. Scores are expressed in terms of percentage correct for standard, backward, inverted, auditory and no-suffix trials. Standard error bars are shown.

the lipread suffix effect is modulated by the similarity between the suffix and the to-be-remembered sequence. In the auditory domain, superficial, acoustic aspects of the suffix modulate its effect even when the suffix contains verbal-lexical information congruent with the to-be-remembered sequence (e.g., Bloom, 2006; Crowder & Morton, 1969; Frankish, 2008; Nairne, 1990). Nevertheless, reducing the superficial, similarity between lipread suffix and lipread sequence does not alter the effect of the suffix, since neither inverting the contrast (Experiment 6a) nor temporally reversing the verbal lip movements so the verbal content is played backwards (Experiment 6b) reduces its disruptive effect. Rather, quite drastic distortions to the lipread suffix have to be implemented, namely, rotating the suffix 180-degrees, implying that different mechanisms are responsible for lipread and auditory suffix effects.

The action of the lipread suffix may instead reside within the perceived lexical identity of the suffix. While throughout these experiments participants were tested on their ability to identify the individual memory items, it is not clear how the lipread "go" is interpreted in the setting of a sequence of lipread digits. The speculative hypothesis here is that participants might actually be 'misreading' the lipread suffix; a lipread suffix is interpreted as a digit, which leads participants to output whatever digit they interpret it to be as the final item in their recall protocol, depressing recall for the actual terminal item as a result. This would explain why such drastic distortion measures are needed to eliminate the lipread suffix effect – if the lip movements no longer afford any lexical identity that could be construed as a digit, they will not disrupt recency. This possible basis of the lipread suffix effect is tested in Experiment 7.

Experiment 7

Here, a novel approach to understanding the impact of lipread suffixes on lipread recency is taken by directly assessing how participants interpret the lipread suffixes, rather than evaluating their impact on memory. Participants were presented with instances of the suffixes used in previous experiments – the standard lipread "go" suffix, the reversed contrast suffix, the backward suffix and the spatially rotated, inverted suffix. The analysis focused on which digit would be chosen when each suffix item was presented in isolation. This provided the opportunity to gauge whether the results in the lipread suffix conditions for previous experiments were the result of the over-inclusion of the suffix as a digit in the to-be-remembered list and, if so, which digit. At the same time, the pattern of perception/misperception should map onto the pattern of suffix effects in Experiment 6a and 6b; the greater the likelihood of misinterpreting the various forms of suffix as a particular digit, the greater the 'suffix effect'.

Method

Participants

Twenty volunteers (19 females), aged 18 to 25 years (21.50 mean years), recruited from Cardiff University's School of Psychology, were given course credit for their participation. All participants, none had taken part in Experiments 1 to 6, were native English speakers reporting normal (or corrected-to-normal) vision and normal hearing.

Apparatus & Materials

The same lipread digits (1-9), as well as the standard lipread “go”, reversed contrast, backward, and inverted suffixes used in Experiments 2 to 6 were presented.

Design

Thirteen items (lipread suffixes: standard, reversed contrast, backward, inverted; digits: one to nine) were shown on five occasions pseudo-randomly, with the constraint that no item was presented more than twice in succession. A total of 65 trials were administered to all participants. The dependent measure was the frequency that each suffix item was chosen as each digit. It should be noted that participants had to provide a response in terms of a digit (i.e., there was no ‘not a digit’ response option). This meant that if there was no systematic pattern of attribution of a specific digit label to the lip movements, no errors in lexical misidentification of the suffix as a digit would be expected.

Procedure

Participants were tested individually in a sound-attenuated laboratory and wore headphones throughout the experiment. A 500ms warning tone signalled the start of each trial, followed by a fixation cross presented for 1s before the onset of the first item. Participants were then asked to identify that item on a forced-choice response screen in which the digits one to nine were shown. The experiment lasted approximately 15-minutes.

Results

Pre-screening scores ranged from 86.67% to 100% (mean = 89.11%) lipread digits correctly identified.

The results (in Figure 15) show that the suffixes varied in their likelihood of being chosen as a particular digit. The standard lipread “go” suffix used in Experiment 2 to 6 was most often identified as the digit ‘two’, as was the reversed video suffix used in Experiment 6a. Similarly, the temporally reversed suffix used in Experiment 6b was more likely to be chosen as the digit ‘one’. The only type of suffix that did not give rise to a systematic error pattern was the inverted (upside-down) suffix, the very condition that failed to show a suffix effect in Experiment 6b.

A 4x9 repeated measure ANOVA, with the factors of suffix (standard, reversed contrast, backward, inverted) and digit identified (1-9), confirmed that there was a main effect of digit identified, *F*(8,152) = 20.69, *MSE* = 327.07, *p*<.001, η^2 = .52. Critically, there was also an interaction between suffix and digit identified, *F*(24,456) = 11.08, *MSE* = 288.21, *p*<.001, η^2 = .37, such that the standard lipread suffix was most frequently misidentified as the spoken digit ‘two’, *F*(8,152) = 18.85, *MSE* = 251.58, *p*<.001, η^2 = .50. This was also the case for the reversed contrast suffix, *F*(8,152) = 14.47, *MSE* = 395.00, *p*<.001, η^2 = .43. By comparison, the backward suffix was more commonly misidentified as the spoken digit ‘one’, *F*(8,152) = 20.72,

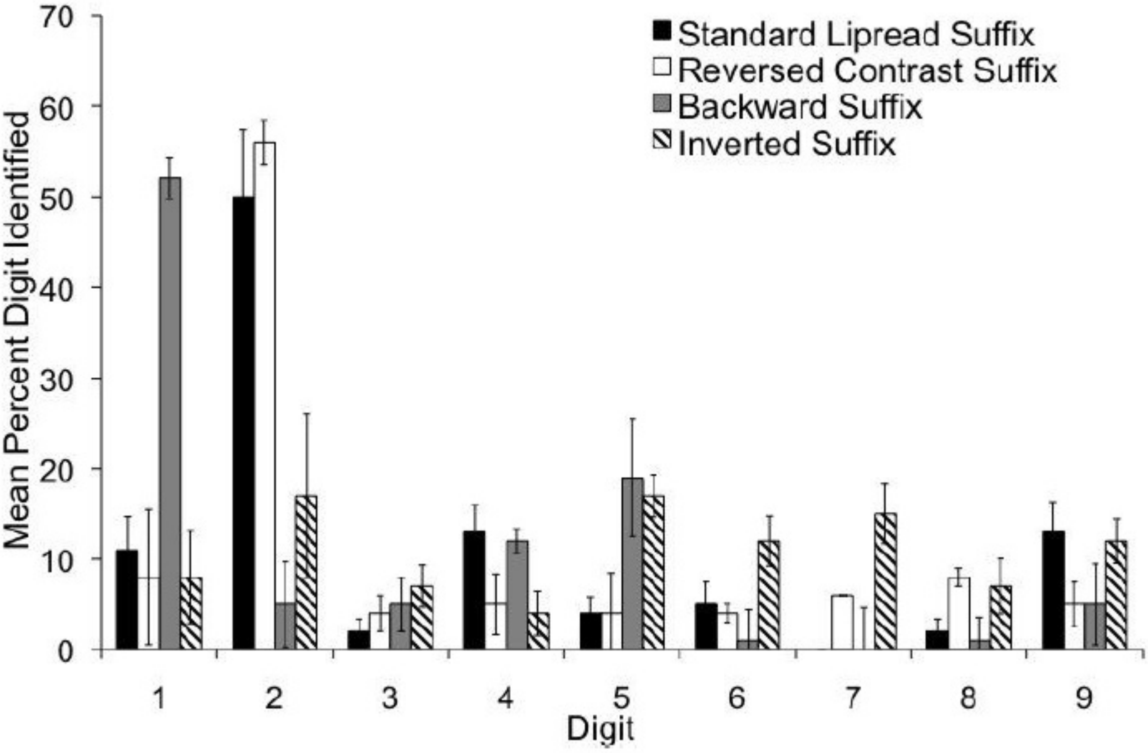

Figure 15. The mean percentage digit (1-9) identified for Experiment 7 across standard, reversed contrast, backward, and inverted lipread suffix types. Standard error bars are shown.

MSE = 262.37, *p*<.001, η² = .52. Just as notable, for the inverted suffix there was no clear preference (*p*>.05).

From the analysis done so far, it seems reasonable to conclude that the lipread suffixes used in Experiment 2 to 6 seem capable of being interpreted (with a frequency of perhaps as high as 50% of the time) as digits, not as the word "go". However, one potential caveat of Experiment 7 is that participants were not able to report that a lipread "go" item was not a digit. Given that the pre-screening scores shown here, as well as in Experiments 1 to 6, show that participants were very accurate at identifying single digits, it might have been expected that participants would have been just as accurate in discriminating digits from non-digits. Furthermore, the word “go” arguably involves a similar motor movement as the digit two in terms of lip protrusion into a rounded shape. As a result, given the constraints of having to always choose between digits regardless of whether the stimulus actually was a digit, participants may have preferentially selected the digit two for the standard and reverse contrast suffixes because it looks most like this digit.

To test whether such preferential selection occurs, it might be necessary to re-administer the experiment so that it included some type of "not a digit" option; if including this option eliminated the preference for the digit two, this would demonstrate that it is not lexical misidentification. However, it was decided that revisiting the serial recall data of previous experiments provides a sufficient alternate means of verifying whether the effect of the lipread suffix can be explained in terms of being misidentified as a candidate memory list item, as opposed to some response bias given the constraints of the forced-choice task. Consequently, in the next section the frequency with which the digit ‘two’ was recalled as the last item in Experiments 6a and 6b was noted for lipread and no-suffix conditions, specifically excluding trials when the digit ‘two’ was actually presented at the terminal position. If the digit two was recalled more frequently at the end of the list when a lipread “go” suffix was presented in comparison to no suffix conditions, this would certainly make an interpretation based on lexical misidentification more plausible.

Further Analysis

Figure 16 shows that the incidence of reporting the digit ‘two’ at the terminal list item was in fact much higher in a lipread list that was followed by the lipread suffix “go” than with either an auditory “go” suffix or no-suffix.

For Experiment 6a (Figure 16a), a one-way ANOVA confirmed that relative to no-suffix condition the spoken digit ‘two’ was more frequently recalled at the terminal serial position for standard lipread “go”, *F*(1,29) = 19.64, *MSE* = 536.80, *p*<.001, η^2 = .25.

For Experiment 6b (Figure 16b), the spoken digit ‘two’ was also more frequently recalled at the terminal serial position for standard lipread “go” relative to both auditory “go”, *F*(1,29) = 10.81, *MSE* = 423.41, *p*<.01, η^2 = .27, and no-suffix conditions, *F*(1,29) = 8.03, *MSE* = 555.12, *p*<.01, η^2 = .22, which did not differ (*p*=.93).

Discussion

It seems probable that what was witnessed across Experiments 1 to 6 as a lipread suffix effect was in fact a case of mistaken identity. Although the lip movements should have signified the redundant suffix item “go”, it was actually interpreted as a digit and reported as such as the last item in the sequence. The finding that the digit ‘two’ was more frequently chosen by participants when the lipread “go” suffix was presented supports this hypothesis. Furthermore, this may have been made more likely when presented immediately after the to-be-remembered list because the suffix was seen in the context of a string of prior

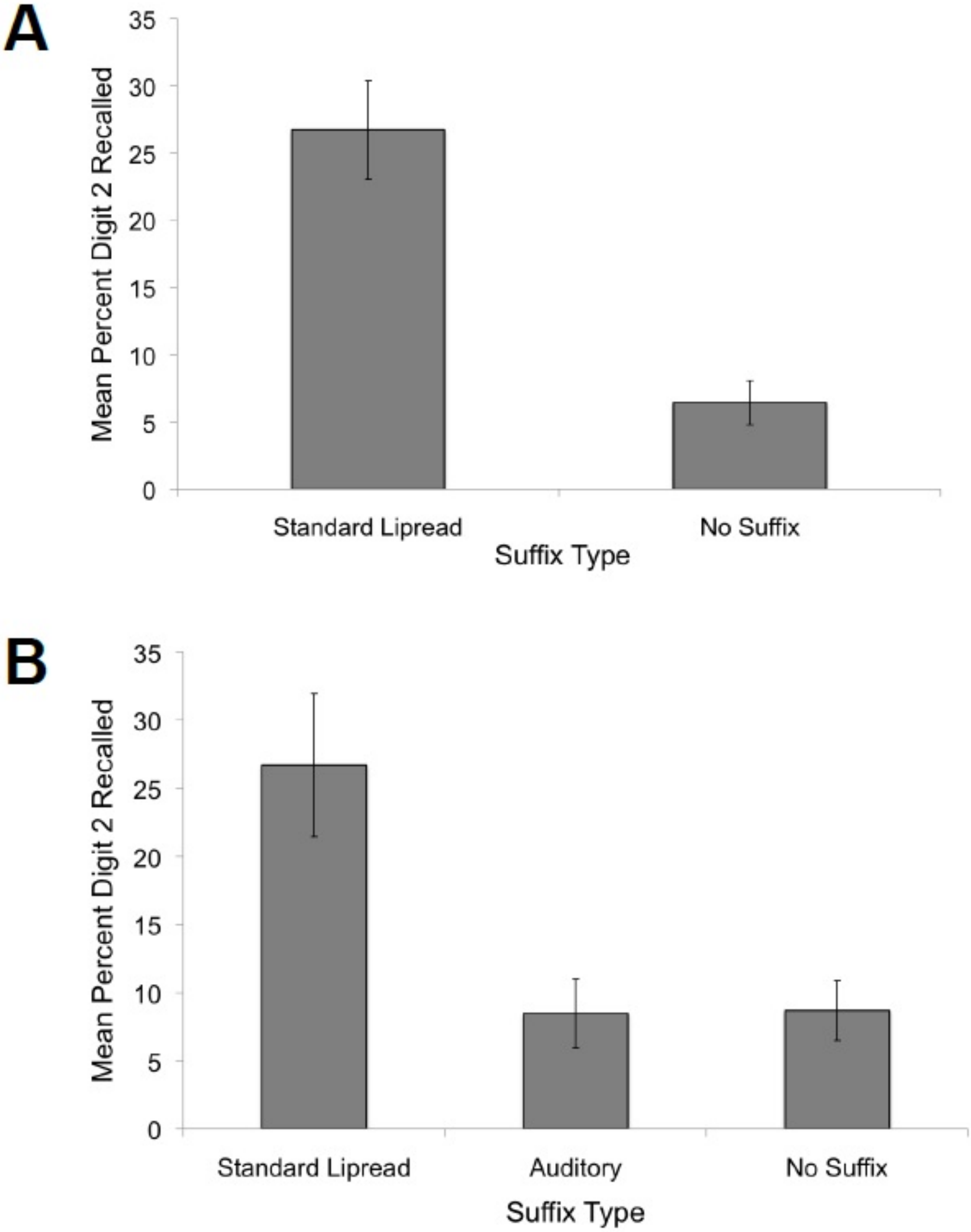

Figure 16. Mean percentage of items identified as the digit ‘two’ at the terminal position during Experiment 6. Panel A shows standard lipread and no-suffix trials (Experiment 6a), and Panel B standard lipread, auditory, and no-suffix trials (Experiment 6b). Standard error bars are shown.

lipread digits.

This discovery may also provide the solution to the question concerning the effect of the bimodal suffix. That is, while both unimodal auditory and lipread suffixes disrupt recency, in bimodal combination there is no disruption (Experiment 2 and 3). The fact that the lipread suffix effect is actually an effect of misinterpretation of the lexical content of the lip movements means that when auditory information is bound perceptually with those lip movements within a bimodal suffix, it serves to unambiguously identify the item as the word “go”, eliminating its ability to disrupt recency. That is to say, including disambiguating auditory information so that the lip movements are no longer misinterpreted as a digit can eliminate the disruptive effect of a lipread suffix on lipread recency.

Furthermore, because the auditory element is bound with the visual lip movements, the unexpected shift from one modality to the other that occurs when a unimodal auditory-verbal suffix is presented at the end of a lipread sequence attenuates its disruptive effect. This is because a stream of visual-verbal information continues from the to-be-remembered sequence through to the bimodal suffix. Thus, the disruptive effect of an auditory suffix on lipread recency – the result of attentional capture – can be eliminated when it is bound to a verbal event that is also visual.

General Discussion

Although the current chapter presents a relatively complex set of findings, overall the evidence provided points to one clear and general conclusion; the apparent similarities between short-term memory for auditory-verbal and lipread-verbal material do not provide evidence that a shared level representation can account for the functional similarities in short-term memory performance across the two modalities.

Firstly, the elements of auditory perceptual processing are immune to the effects of manipulations that impede the speech rehearsal process, as evidenced by enhanced recency for heard over written material in the presence of articulatory suppression or irrelevant speech. However, whatever gives rise to lipread recency is not immune (Experiment 1). Secondly, cross-modal interactions between lipread and auditory suffixes on lipread lists are actually driven by fundamentally different mechanisms in each case (Experiments 2 to 7). Specifically, the disruptive effect of an auditory suffix on lipread recency is due to attention being captured away from the lipread sequence by an unexpected auditory event, rather than interactions amongst common verbal representations (Experiment 3 to 5). In comparison, the impact of a lipread suffix on a lipread list is attributable to the misidentification of the lipread suffix as a specific digit, and the output of that misidentified digit in the subsequent recall protocol (Experiment 6 and 7). Therefore, in view of these findings it may be inappropriate to refer to just one recency and suffix effect, which relies on the operation of an amodal level of representation. Rather, it appears that there are in fact several recency and suffix effects – one underpinned by processes supporting auditory-verbal information and the other supporting visual-verbal (i.e., lipread) information.

In terms of auditory recency, elsewhere it has been argued that this effect is attributed to auditory perceptual organisation (Jones, et al., 2006; Jones, et al., 2004; Maidment & Macken, 2012; Nicholls & Jones, 2002). Key evidence for this position stems from a re-examination of the phonological similarity, which has typically been ascribed to the manipulation and storage of phonological representations. However, the similarity effect is not due to some phonological representation, but has been shown to emerge as a result of the combined roles of auditory perceptual organisation and speech-motor processes (Jones, et al., 2006; Jones, et al., 2004; Maidment & Macken, 2012; Nicholls & Jones, 2002). First, articulatory suppression abolishes the effect of similarity throughout the list for visual presentation, and in all but the terminal item for auditory presentation (Jones, et al., 2006; Jones, et al., 2004). This suggests that this specific effect of similarity resides in the rehearsal process; with the exception of auditory recency, there is no effect of similarity for written or auditory presentation when rehearsal is prevented. Second, the effect of similarity that survives articulatory suppression is predominantly evident in the recency portion of the serial position curve (Jones et al., 2004), with sequences of dissimilar sounding items, relative to similar-sounding sequences, showing enhanced recency when presented auditorily. What this suggests is that this similarity effect is independent of the rehearsal process, and should therefore be attributed to processes supporting auditory recency.

Critical to our understanding of the mechanisms underpinning auditory recency is the existence of the auditory suffix effect. The effect of an auditory suffix has been argued to disrupt auditory recency by interfering with acoustic storage of the final items via backward interference, or masking, based on similarity between the content of the memory list and the suffix (e.g., Bloom, 2006; Crowder & Morton, 1969; Frankish, 2008; Nairne, 1990). However, it has been shown that auditory recency is actually restored when an auditory suffix is partitioned, or 'captured', into an alternative perceptual stream to that formed by the memory sequence (Bregman & Rudnicky, 1975; Kahneman & Henik, 1981; Nicholls & Jones, 2002). For example, a spoken word (e.g., "go"*)* presented at the end of a spoken verbal sequence will reduce recall performance for the terminal list item, but auditory recency will not be disrupted in the presence of that same suffix if the digit sequence is concurrently accompanied by a repeated series of the spoken suffix word (Nicholls & Jones, 2002). This indicates that the mere presence of an auditory suffix on an auditory memory list does not necessarily result in a suffix effect, rather, the suffix effect reflects the operation of auditory perceptual organisation, where the suffix must also be part of the perceptual stream formed by the memory sequence (Maidment & Macken, 2012; Nicholls & Jones, 2002). Therefore, the key process determining whether or not a suffix will disrupt auditory recency is an auditory perceptual one, which determines whether or not the suffix will be perceptually grouped with the to-be-remembered list (Kahneman & Henik, 1981; Nicholls & Jones, 2002).

In comparison, despite both auditory and lipread lists showing enhanced recall at the terminal list item, the findings throughout the current chapter argue that the lipread recency effect can be dissociated from the mechanisms underpinning auditory recency. Specifically, while it could be maintained that lipread recency arises because visual speech information generates an auditory-like representation, this is called into question by the observation that lipread recency is diminished in the presence of articulatory suppression and irrelevant sound. As a result, lipread recency, unlike auditory recency, appears to be dependent on the rehearsal process necessary for short-term maintenance of verbal information. Alternatively however, it might be that, rather than directly modifying activity within the auditory system, lipread speech information activates the auditory system via motor pathways linking speech production and perception (Hickok, 2009; Hickok & Poeppel, 2004, 2007; Skipper, et al., 2009; Skipper, et al., 2007). On this basis, the typical encoding of lipread speech into an auditory-like representation is prevented in the presence of articulatory suppression and irrelevant sound because the motor pathways linking visual and auditory speech are disrupted.

Nevertheless, the fact that a lipread suffix does not disrupt auditory recency suggests that it is not susceptible to processes of auditory perceptual organisation underpinning auditory recency and suffix effects. Likewise, the lipread recency effect, and its interaction with auditory and lipread suffixes, cannot be attributed to the same perceptual processes supporting auditory recency and suffix effects. Rather, although the effect of an auditory suffix on an auditory list is due to the perceptual grouping of the suffix with the list, the effect of an auditory suffix on a lipread list is actually due to attentional capture ensuing from the presentation of an unexpected auditory event. The finding that lipread recency is disrupted by either a verbal or nonverbal auditory suffix supports this conclusion.

Furthermore, the effect of lipread suffix on a lipread list is due to misidentification. That is to say, the lipread suffix is actually interpreted as a digit and is subsequently reported as the last item in the sequence. This conclusion is grounded in the observation that rotating the lipread “go” suffix through 180-degrees, so that it is displayed upside-down, not only restores lipread recency, but is also the only lipread suffix that is not consistently identified as a specific digit. Taken together, despite their apparent similarities, lipread and auditory recency effects, and their interactions with auditory and lipread suffixes, appear to be due to distinct mechanisms.

However, whether the tendency of a lipread suffix to be misidentified as a list item is a universal one that appears in all experiments of this kind is debatable. Previous studies, which most closely resemble the procedures here (e.g., de Gelder & Vroomen, 1992, 1994; Greene & Crowder, 1984; Spöehr & Corin, 1978), have typically required the participants to signal to the experimenter that they had perceived, and therefore encoded, the lipread suffix. This departure from the original procedure in classical demonstrations using auditory lists was a necessary step to ensure that participants did not look away or close their eyes during the presentation of a lipread suffix – otherwise it would have no effect. Thus, these studies used an array of conditions to ensure that attention was always paid to the suffix. Nevertheless, although this may have been a necessary safeguard   it is argued that by using lipread lists throughout, so that attention was already oriented at the screen in preparation for any visual stimulus, this step was not necessary in the present series. This was likely an improvement in the current methodology, minimising the risk that the suffix effect would be confounded with the differential engagement of executive processes of checking that a suffix had been presented in the different list/suffix modality combinations. In any case, in this regard the methods deployed in the current chapter are more in line with the procedure of the classical suffix effect in which signaling the presence of the suffix was not required (Crowder & Morton, 1969). As a result, the current findings provide a more suitable comparison, showing that the processes supporting lipread recency and suffix effects are not the same as those supporting analogous effects in the auditory modality.

Overall, the current chapter focussed on distinct ways in which verbal material from different modalities interact in short-term memory in order to better understand the fundamental mechanism, or mechanisms, underpinning verbal performance. Critically, the current findings provide significant evidence against an account of verbal behaviour that relies on an amodal level of representation, which has typically been the object of bespoke storage and manipulation (e.g., Baddeley, 1992; Burgess & Hitch, 1999; Nairne, 1990; Neath & Nairne, 1995; Page & Norris, 1998; Penney, 1989; Winkler, et al., 2009). Rather, this series of experiments adds to a growing body of evidence that, in the short-term, verbal performance is actually constrained by modality-specific representations. This therefore provides serious implications for accounts of verbal behaviour that rely on the operation of phonological representations, which still continue to be broadly influential within the realm of Cognitive Psychology.

CHAPTER 3

What the inner voice tells the inner ear: Motor mediation in audiovisual speech perception.

Introduction

When auditory and visual information are presented simultaneously, stimuli in one modality often influence the perception and comprehension of ambiguous stimuli in the other, second modality. The most commonly cited example of this in the study of verbal behaviour is the enhanced intelligibility, localisation and discrimination of heard speech when accompanied by congruent visual information, such as that derived from facial expressions and lip movements (e.g., Bishop & Miller, 2009; McGettigan et al., 2012; Sánchez-García, Alsius, Enns, & Soto-Faraco, 2011; Summerfield, et al., 1989). The McGurk effect (McGurk & MacDonald, 1976) further demonstrates that when the auditory signal /ba/ is simultaneously presented with seeing the talker say /ga/, the resulting 'heard' (or illusionary) percept is /da/ (for recent discussion, see Jiang & Bernstein, 2011). What this evidence suggests is that there is a point at which auditory and visual speech information converge, that is, the features from these different systems come to be bound in some way (for review, see Price, 2012). However, the experiments reported in Chapter 2 revealed important distinctions in the nature of verbal representations derived from heard and seen speech, undermining the case for the involvement of amodal functions in the processing of both sources of linguistic information. Consequently, the concern of the current chapter is to continue to understand the nature of the representations underpinning both modalities, specifically examining how and where auditory-verbal and visual-verbal inputs come to be integrated.

One of the primary motives for revisiting the mechanisms underlying audiovisual integration here is that the stage at which both modalities are bound has been debated for some time, resulting in the generation of two theoretical standpoints. Firstly, the hierarchical view of audiovisual integration conceives that seen and heard speech inputs are processed independently in each modality prior to their integration at ‘higher’ levels of multimodal processing (e.g., Felleman & Van Essen, 1991). That is, modality-specific representations of speech are integrated at high levels of processing supporting both modalities. Secondly, the hierarchical view of multisensory integration has been challenged by more recent accounts proposing that visual-verbal and auditory-verbal information converge at a level typically associated with unimodal, auditory processing (for review, see Alais, et al., 2010). This view stems from neuroscientific evidence showing that silently lipread information gains ready access to brain areas typically associated with auditory processing, including the primary auditory cortex (Bernstein, et al., 2002; Calvert, et al., 1997). Consequently, these findings have been interpreted to suggest that auditory and visual representations of speech share a common ‘auditory’, as opposed to amodal, form (for review, see Green, 1998; Schwartz, Grimault, et al., 2012).

However, there is also a growing body of evidence showing visual speech information gains ready access to auditory-like representations via associative pathways linking speech production and perception (Hickok, 2009; Hickok & Poeppel, 2004, 2007; Skipper, et al., 2009; Skipper, et al., 2007). On this basis, lipread and auditory speech may share an equivalent auditory mode of representation, but the path whereby they are encoded is distinct. As such, it might be the case that visual speech does not directly modify activity within the auditory system, but rather, is encoded into an auditory-like form via processes necessary for speech production. As a result, some speech production mechanism may be essential when integrating both modalities. The findings reported in Chapter 2 speak directly to this matter, with Experiment 1 showing that both auditory and lipread inputs result in equivalent levels of performance within the recency portion of the serial position curve. However, while these findings have been used to support the notion that both seen and heard inputs lead to an amodal level of representation, auditory recency was shown to be immune to the disruptive effects of articulatory suppression and irrelevant sound, whereas lipread recency was not immune. Given that both articulatory suppression and irrelevant sound disrupt the rehearsal process (Jones & Macken, 1993; Jones, et al., 2004; Macken & Jones, 1995; Macken, et al., 2009), it could be maintained that lipread recency is, unlike auditory recency, dependent upon subvocal speech production processes. Thus, when the speech production pathway linking vision and audition is disrupted, the typical encoding of visual speech into an auditory-like representation is prevented so that an auditory-like recency effect found with lipread lists is no longer observed.

Critical to this argument, linguistic information not only has acoustic and visual properties, since the sound of the speaker’s voice – in the form of self-generated auditory feedback – is commonly correlated with articulatory motor actions generated during verbal production. In support, speech production mechanisms appear to be necessary for the perception of speech. For example, behavioural data demonstrates that the perception of one's own voice influences the planning and execution of articulatory gestures (for review, see Casserly & Pisoni, 2010; Villacorta, Perkell, & Guenther, 2007), with articulatory constraints imposed upon the speaker, such as mechanically altering jaw movements, modifying the perception of heard speech material (Nasir & Ostry, 2009). The motor system's involvement in auditory speech perception has also be garnered from studies utilising transcranial magnetic stimulation (TMS): when TMS is applied to face areas corresponding to the lip and mouth regions of the primary motor cortex, motor-evoked potentials (MEPs) recorded from these facial areas are enhanced when passively listening to linguistic information. That is, TMS to motor tongue areas increases MEPs in the tongue when listeners hear partially ambiguous speech sounds that, if produced, would require tongue movement (e.g., /t/). In comparison, stimulation to the motor lip areas increases MEPs in the lips when listening to or watching speech sounds formed by lip closure (e.g., /b/) (D'Ausilio et al., 2009; Fadiga, Craighero, Buccino, & Rizzolatti, 2002; Murakami, Restle, & Ziemann, 2011; Watkins, Strafella, & Paus, 2003).

Nevertheless, while the evidence discussed above provides strong support for the view that speech production mechanisms are also involved in speech perception, the question that arises is whether motor mechanisms necessary for speech production also play a role in the processing and integration of auditory-verbal and visual-verbal information more generally. Again, the findings of Chapter 2 appear to address this issue, with Experiments 2 to 7 demonstrating that the effects of lipread and auditory suffixes on lipread and auditory recency effects do not appear to be supported by the same mechanisms. Elsewhere, it has been argued that the effect of an auditory suffix on auditory recency reflects the operation of auditory perceptual organisation, whereby the suffix must be part of the perceptual stream formed by the memory sequence to disrupt recall at the terminal list item (Maidment & Macken, 2012; Nicholls & Jones, 2002). If, as it has been argued, lipread information generates an auditory-like representation, a lipread suffix might also be expected to disrupt auditory recency in much the same way as an auditory suffix – via processes of perceptual organisation. Critically however, this was not the case, since a lipread suffix did not disrupt auditory recency. Furthermore, auditory and lipread suffix effects on lipread lists were actually driven by fundamentally different mechanisms, whereby the disruptive effect of an auditory suffix on lipread recency was due to attentional capture, while the impact of a lipread suffix on a lipread list was attributable to the misidentification of the lipread suffix as a specific digit. On this basis, although lipread information appears to generate auditory-like representations in terms of similar levels of performance at recency, effects found with visual speech can actually be dissociated from those supporting auditory-verbal information.

From this perspective, the experimental series that follows attempts to further understand the nature of the representations underlying seen and heard inputs. Specifically, the main objective was to understand how and where auditory and visual speech inputs interact, investigating the extent to which both modalities converge via articulatory-motor processes necessary for speech production. In the first experiments of this series the mechanisms underlying the integration of seen and heard speech was behaviourally investigated using the McGurk effect. The degree to which the reported percept (e.g., /da/) deviates from what is either seen (/ga/) or heard (/ba/) indicates the degree to which both modalities are bound perceptually. Audiovisual binding was measured by contrasting one of three syllable presentation formats: (1) *auditory-only*, where a spoken syllable (e.g., /ga/ or /ba/) was heard in the absence of visual information; (2) *audiovisual congruent*, in which matching auditory (e.g., /ba/) and visual (e.g., /ba/) speech material were presented; and (3) *audiovisual incongruent*, consisting of the McGurk pair of an auditory /ba/ accompanied with seeing the talker say /ga/. Crucially, a McGurk effect was expected in this latter condition relative to auditory-only and audiovisual congruent conditions.

The critical test here was that audiovisual integration was further scrutinised in the presence of different interference manipulations, which were contrasted against a no task, control condition. The effect of concurrently articulating task irrelevant verbal material during syllable presentation was compared with passive listening to irrelevant speech (Experiment 8), sequential, manual tapping (Experiment 9), or silent mouthing (Experiment 10). While the reasons for deploying each of these manipulations will be discussed in turn, the specific aim of Experiments 8 to 10 was to determine the stage at which auditory and visual speech converge. That is, each experiment was designed with the intention of verifying whether the integration of seen and heard inputs requires a pathway linking speech production and perception (see Hickok, 2009; Skipper, et al., 2009). Importantly, this was measured via the McGurk effect, whereby a reduction in the illusionary (\da\)

percept as a result of any interference manipulation would provide an indication of the mechanism(s) involved during audiovisual integration.

Experiment 8

In Experiment 8 the integration of seen and heard speech was examined in the presence of concurrent articulation and irrelevant speech. Part of the impetus for examining the impact of concurrent articulation and irrelevant speech on audiovisual binding stems from the finding that both manipulations have been shown to disrupt speech production mechanisms supporting the subvocal rehearsal of verbal information in short-term memory (e.g., Jones & Macken, 1993; Jones, et al., 2004; Macken & Jones, 1995; Macken, et al., 2009). That is, with the exception of auditory recency, both concurrent articulation and irrelevant speech reduce serial recall throughout the to-be-remembered list when presented in auditory-verbal or visual-verbal form.

Perhaps most critically is the finding that, while auditory and lipread lists show enhanced performance at the terminal list item, lipread recency is disrupted when subvocal rehearsal is impeded by concurrent articulation and irrelevant speech respectively (see Experiment 1). This consequently points to the possibility that an auditory-like verbal performance found with lipread lists arises because visual representations of speech generate auditory representations via a speech production pathway linking both modalities.

Therefore, the motive of the current experiment was to show that if auditory and visual representations of speech interact via the speech production pathway, manipulations that disrupt speech production – namely, concurrent articulation and irrelevant speech – should also prevent the integration of seen and heard inputs. Subsequently, it was predicted that the presence of concurrent articulation and irrelevant speech would reduce the likelihood for participants to integrate auditory and visual inputs, as indexed by a reduction in the McGurk effect.

Method

Participants

Twenty-five volunteers (17 female), aged 19 to 49 years (23.56 mean years), recruited online from the School of Psychology, Cardiff University, were paid for participation. All participants were native English speakers who reported normal (or corrected-to-normal) vision and normal hearing.

Apparatus and Materials

Using a *Canon Legria HF200* high-definition camcorder, /ba/ and /ga/ syllables were recorded in a sound attenuated laboratory by a male speaking in a monotone voice (at a F0 of approximately 150Hz). Using *Final Cut Express*, audiovisual incongruent (McGurk) syllables were edited by replacing the /ga/ sound from the audiovisual /ga/ video clip with the sound /ba/. The /ba/ sound was positioned at exactly the same temporal position as the /ga/ sound, so that it was synchronised with the onset of the visual /ga/ lip movements. For auditory-only presentation, /ba/ and /ga/ sounds were dubbed onto a video of a still frame of the speaker's face with closed lips. Video clips were approximately one second in duration and cropped so that only a head-on view of the upper and lower lips was visible.

Irrelevant speech was generated by the same male speaking the digit sounds *1*, *2* and *3*. Using *SonicForge 5.0* software (Sonic Foundry, Inc., Madison, WI; 2000), the pitch of each item was three semitones higher than that of the target syllables and was compressed digitally to 190ms without further changing pitch. Visual stimuli were displayed in the centre of a white PC screen and auditory stimuli monophonically over headphones using *E-Prime*.

Design

Syllable Presentation (auditory-only, audiovisual congruent, audiovisual incongruent) and Interference (no task, irrelevant speech, concurrent articulation) were varied within-groups, with twelve trials undertaken for each of the nine Syllable Presentation x Interference conditions. Trials were arranged in a pseudo-random order with no interference type or syllable presentation shown more than twice in succession, and were balanced across all participants.

The number of auditory responses across trials (expressed as a percentage) for each syllable condition was taken as the dependent measure, since acoustic stimuli were presented across conditions. Furthermore, this dependent variable permits the degree of audiovisual integration to be quantified objectively; for audiovisual incongruent presentation, more auditory responses indicate a weaker

McGurk effect suggesting that the visual stimuli did not influence the participant's judgement and/or an illusionary, \da\ percept was not generated (Munhall, Gribble, Sacco, & Ward, 1996).

Procedure

Participants were tested individually in a sound attenuated laboratory and wore headphones throughout the experiment where the sound level was individually adjusted to a comfortable level. A 500ms warning tone (500Hz sinewave) signalled the start of each trial, followed by a fixation cross presented for 5s before the onset of the first syllable. This introductory period was filled with either silence in the case of the no task and concurrent articulation conditions or a period of irrelevant speech (approximately 20 tokens presented with an onset-to-onset interval of 250ms, with four items being presented every second) that continued without a break in tempo during syllable presentation. When irrelevant speech was present, participants were instructed to ignore any spoken numbers heard over the headphones during that trial. For concurrent articulation conditions, following warning tone offset, participants were required to whisper aloud the numbers *1*, *2* and *3*. The experimenter coached each participant in the correct rate (approximately four items per second) and loudness of articulation, remaining in the laboratory to ensure compliance with instructions.

Both irrelevant speech and concurrent articulation were repeated until participants were visually cued to identify the presented syllable via a forced-choice response screen displaying the syllables ba, ga, and da in written form. Using the cursor, participants were required to click over the syllable that corresponded to the one they thought had been presented. Once a response had been registered, the next trial commenced automatically. The experiment lasted approximately 30-minutes, including an optional 5-minute rest period at the halfway point.

Results

Figure 17a shows the rate of auditory responses averaged across all participants for no task, irrelevant speech, and concurrent articulation conditions as a function of syllable presentation (auditory-only, audiovisual congruent, audiovisual incongruent). A 3 (Syllable Presentation) x 3 (Interference) repeated measure ANOVA confirmed that fewer auditory responses were reported for audiovisual incongruent presentation when compared to both auditory-only and audiovisual congruent conditions, shown as a main effect of Syllable Presentation, *F*(2,48) = 214.08, *MSE* = 657.37, *p*<.001, η^2 = .90. Critically, however, there was also a main effect of Interference, *F*(2,48) = 15.74, *MSE* = 96.89, *p*<.001, η^2 = .40, as well as an interaction between Interference and Syllable Presentation, *F*(4,96) = 17.95, *MSE* = 96.63, *p*<.001, η^2 = .43.

As can be seen in Figure 17a, the source this interaction appears to reside in the differential effects of interference on the auditory response rate for audiovisual incongruent presentation. In particular, and critical to the question of whether audiovisual binding is mediated by articulatory motor processes necessary for speech production, when the auditory and visual stimuli were incongruent significantly more auditory responses were reported in the presence of concurrent articulation relative to both no task, *F*(1,24) = 20.64, *p*<.001, η^2 = .46, and irrelevant speech manipulations, *F*(1,24) = 26.54, *p*<.001, η^2 = .53, which did not differ (*p*=.25).

Discussion

As expected, in comparison to auditory-only and audiovisual congruent conditions, a McGurk effect was present during audiovisual incongruent presentation, as measured by a reduced likelihood for individuals to report what was auditorally presented. However, despite it being shown that concurrent articulation and irrelevant speech both disrupt speech production mechanisms supporting subvocal processes during serial recall (e.g., Jones & Macken, 1993; Macken & Jones, 1995; Macken, et al., 2009), the presence of concurrent articulation significantly reduced this McGurk effect by ~20%, whereas the same cannot be said for irrelevant speech, which did not differ from control, no task conditions. Nevertheless, it is also worth highlighting that this observed reduction in the presence of concurrent articulation was only partial. Why this was the case will be investigated further in Experiment 11, but, for the time being, the mechanisms responsible for this partial reduction will now be discussed.

The finding that to-be-ignored, irrelevant speech did not reduce the McGurk effect contradicts the expected hypothesis that both concurrent articulation and irrelevant speech would disrupt the McGurk effect. Even so, an explanation for the lack of an effect in the presence of irrelevant speech may be found from a closer inspection of the origin of its disruption in verbal short-term memory. As argued in the previous chapter, the disruptive effect of irrelevant speech on serial short-term memory arises because the task-relevant information (i.e., the to-be-remembered list) and the task-irrelevant speech involve sequential processing, which compete for control of the speech motor mechanisms that are also engaged in the subvocal rehearsal process (Jones & Macken, 1993; Jones, et al., 2004; Macken, et al., 2009). This contrasts with an effect of concurrent articulation, which disrupts serial recall by explicitly suppressing the articulatory processes necessary for subvocal rehearsal. Subsequently, the lack of disruption by irrelevant speech may have arisen because the target speech in the present experiment did not generate a sequential representation. That is to say, because syllable identification did not require sequential, subvocal rehearsal of verbal stimuli, there was no effect of irrelevant speech.

By comparison, given that concurrent articulation, by design, interferes with speech production processes, that it also interferes with audiovisual binding suggests that speech production processes may be involved in the integration of auditory and visual speech. In spite of this, an alternative explanation as to why concurrent articulation reduced the McGurk effect, while irrelevant speech did not, could be attributed to the possibility that passive listening to speech may be less cognitively demanding than actually producing speech out loud. Critical to this argument is evidence showing that the McGurk effect is reduced during dual-task situations. For example, the McGurk effect is reduced by ~10 to 20% – a similar reduction observed in the current experiment in the presence of concurrent articulation – during dual-task conditions (Alsius, Navarra, Campbell, & Soto-Faraco, 2005; Alsius, Navarra, & Soto-Faraco, 2007; Buchan & Munhall, 2012; Tiippana, Andersen, & Sams, 2004). For example, in Alsius et al.'s (2005) study participants were required to monitor a screen displaying the face of a speaker producing spoken words, some of which were dubbed to produce McGurk stimuli, at irregular intervals – with an average inter-word interval of ~21 seconds. Immediately after each word had been shown, participants were instructed to repeat it out loud. Critically, when the McGurk effect was measured in this way, the authors found that it was reduced when participants were also required to complete an additional task involving visual images (Experiment 1) or environmental sounds (Experiment 2). In their first experiment, a sequence of line drawings was presented at a rate of approximately two items per second (~240 ms offset to onset), which were superimposed on the display, but did not conceal the speaker's lips or jaw, and rotated at various angles. In the second experiment, the procedure was identical except the pictures were replaced by sounds, such as a telephone ringing or a dog barking. In both instances, participants were asked to indicate when two images or two sounds were identical via button press.

According to Alsius, et al. (2005), a reduction in the McGurk effect during dual-task conditions can be explained in terms of the demand to perform another secondary task during audiovisual integration, which depletes the attentional resources required for the binding of auditory and visual features into a coherent speech object (see also Sarmiento, Shore, Milliken, & Sanabria, 2012; Treisman & Gelade, 1980; Zvyagintsev, Nikolaev, Sachs, & Mathiak, 2011). In support, the McGurk effect is also reduced when syllable identification is paired with another cognitively demanding, verbal short-term memory task (e.g., Buchan & Munhall, 2012). However, a further consideration of the dual-task methodology discussed above is whether audiovisual integration can also be disrupted when attentional demands are imposed on a sensory domain that is not involved in speech perception (Alsius, Navarra, & Soto-Faraco, 2007)). Consequently, Alsius, et al. (2007) replicated their initial procedure using non-verbal, tactile stimuli. In this study, the secondary task required the participant to place two fingers from each hand on a pair of buttons that vibrated every 1.2 seconds. Each interval consisted of two of the four buttons vibrating for 30 ms. The button pairs were always selected at random, with participants signalling via a foot pedal whether two fingertips were subsequently followed by stimulation of their opposite counterparts. Again, Alsius et al. (2007) found that the McGurk effect was reduced during dual-task conditions involving touch.

Taken together therefore, the observation that the McGurk effect was reduced in Experiment 8 in the presence of concurrent articulation might also be explicable in terms of the cognitive load placed upon attentional resources required to integrate auditory and visual speech during dual-task conditions. Consequently, it might not be the requirement to produce and perceive speech simultaneously that disrupts audiovisual binding during dual-task conditions, *per se*, but the demand to perform another secondary task. This prediction will now be explored in Experiment 9.

Experiment 9

The aim of Experiment 9 was to determine whether the partial reduction of the McGurk effect when participants engaged in concurrent articulation shown in

Experiment 8 was the result of the demands placed upon participants when required to administer a secondary task, as opposed to an effect arising as a result of impeding speech production processes. To investigate this, the impact of a secondary, non-verbal task on audiovisual binding was contrasted with the effect of concurrent articulation. Specifically, sequential, manual tapping was deployed – a manipulation that has also been shown to disrupt cognitive task performance to an equivalent magnitude as concurrent articulation (e.g., Alloway, Kerr, & Langheinrich, 2010; D. Hall & Gathercole, 2011; Henson, Hartley, Burgess, Hitch, & Flude, 2003; Jones, Farrand, Stuart, & Morris, 1995). Furthermore, prior to experimentation it was also shown that, relative to control conditions, concurrent articulation, irrelevant speech, and sequential, manual tapping all significantly reduced serial verbal short-term memory performance (see Appendix A). As such, it was deemed that sequential, manual tapping was a suitable non-verbal comparison.

It was predicted that if audiovisual binding were disrupted by the presence of this secondary task, both verbal (i.e., concurrent articulation) and non-verbal (i.e., sequential tapping) tasks should be expected to reduce the McGurk effect. On the other hand, if the partial reduction of the McGurk effect in the presence of concurrent articulation observed in Experiment 8 was the result of suppressing the speech production mechanisms involved in the binding of auditory and visual speech, sequential tapping should have no impact on audiovisual integration.

Method

Participants

Twenty-five right-handed volunteers (17 females), aged 20 to 39 years (23.83 mean years), recruited online from Cardiff University's School of Psychology, were paid for participation. All participants, none of whom had taken part in Experiment 8, were native English speakers who reported normal (or corrected-to-normal) vision and hearing.

Apparatus & Materials and Design & Procedure

In all respects, the stimuli presented to participants in Experiment 9 were identical to those in Experiment 8, with the factors Syllable Presentation (audiovisual congruent, audiovisual incongruent) and Interference (no task, sequential tapping, concurrent articulation) manipulated in a repeated measures design. When sequential tapping was expected participants were instructed to tap with their right hand the left, down and right arrow keys in order with their index finger, middle finger and ring finger, from warning tone offset until presentation of the forced-choice response screen. Participants were required to tap the computer keys at a rate of four items per second – matching the number of tokens and rate of concurrent articulation.

Results

Figure 17b shows the auditory response rate for no task, sequential tapping, and concurrent articulation conditions as a function of syllable presentation (audiovisual congruent, audiovisual incongruent). Critically, for audiovisual incongruent pairs the auditory response rate is approximately 20% greater in the presence of concurrent articulation relative to both no task and sequential tapping conditions.

This picture was confirmed by a 2x3 (Syllable Presentation x Interference) repeated measure ANOVA, with fewer auditory responses for audiovisual incongruent presentation when compared to audiovisual congruent presentation, shown as a main effect, $F(1,24) = 133.66$, $MSE = 128.61$, $p<.001$, $\eta^2 = .85$. There was also a main effect of Interference, $F(2,48) = 6.96$, $MSE = 198.90$, $p<.01$, $\eta^2 = .23$, as well as an interaction between Syllable Presentation and Interference, $F(2,48) = 8.33$, $MSE = 178.30$, $p<.001$, $\eta^2 = .262$, such that for audiovisual incongruent presentation more auditory responses were reported in the presence of concurrent articulation relative to both no task, $F(1,24) = 10.32$, $p<.01$, $\eta^2 = .30$, and sequential tapping conditions, $F(1,24) = 8.65$, $p<.01$, $\eta^2 = .27$, which did not differ ($p=.34$).

Discussion

The findings of Experiment 9 replicate those shown in Experiment 8: the McGurk effect was partially disrupted in the presence of concurrent articulation.

While this partial effect will be explored in Experiment 11, crucially sequential tapping did not significantly reduce the McGurk effect relative to control, no task conditions. These findings therefore suggest that speech production processes are involved in audiovisual binding, as opposed to an effect arising as a consequence cognitive demands placed upon participants when administering two tasks simultaneously.

Nevertheless, like the effect of irrelevant speech, elsewhere it has been argued that the disruptive effect of manual tapping arises as a result of sequential processing; concurrently tapping a sequence during the encoding and/or retention of the to-be-remembered list disrupts the sequential processes that are also inherent to the subvocal rehearsal process (e.g., Alloway, et al., 2010; D. Hall & Gathercole, 2011; Henson, et al., 2003; Jones, et al., 1995). Consequently, sequential tapping may not have disrupted audiovisual integration because the target speech in the current experiment did not require sequential processing. In support of this argument, although both irrelevant speech and sequential finger tapping reduced the serial recall of visual-verbal sequences, the disruptive effect of concurrent articulation was much greater in magnitude (see Appendix A). On this basis, it could still be maintained that concurrent articulation partially reduced the McGurk effect because it was more demanding than irrelevant speech or sequential tapping, a possibility that will now be studied further in Experiment 10.

Experiment 10

Experiment 10 continues to investigate whether the effect of concurrent articulation was the result of impeding the apparatus necessary for speech production. As a result, the disruptive effect of concurrent articulation was contrasted with silent mouthing in order to confirm whether impeding speech production processes subvocally is sufficient to disrupt audiovisual integration. Critically, silent mouthing does not require the overt vocalisation of irrelevant speech material, such that, unlike concurrent articulation, it does not require the additional task demands of either engaging the vocal tract or ignoring the irrelevant speech produced by the speaker in the form of self-generated auditory feedback.

The main objective therefore, was to further verify the level at which speech production processes are involved in audiovisual integration. That is, if concurrent articulation specifically disrupts the subvocal apparatus necessary for the integration of seen and heard speech, it was predicted that a partial reduction of the McGurk effect should also be observed to in the presence of silent mouthing. If, however, the effect of concurrent articulation can be attributed to additional demands – specifically those placed upon participants when producing speech out loud – audiovisual integration should not be disrupted to an equivalent magnitude by a manipulation that arguably does not require these additional demands.

Method

Participants

Thirty volunteers (25 females), aged 18 to 47 years (20.30 mean years), recruited online from the School of Psychology, Cardiff University, were given course credit for participation. All participants, none of whom had taken part in Experiment 8 or 9, were native English speakers who reported normal (or corrected-to-normal) vision and normal hearing.

Apparatus & Materials

The apparatus and materials deployed in Experiment 8 and 9 were again used in Experiment 10.

Design & Procedure

The factors Syllable Presentation (audiovisual congruent, audiovisual incongruent) and Interference (no task, silent mouthing, concurrent articulation) were manipulated in a repeated measures design. Twelve sequences were presented for each of the six Syllable Presentation x Interference conditions, arranged in a pseudo-random order. For silently mouthed trials, participants were instructed to mouth the numbers *1*, *2* and *3* without vocalising any sounds from warning tone offset until presentation of the forced-choice response screen.

Results

Figure 17c shows the auditory response rate for no task, silent mouthing, and concurrent articulation conditions as a function of syllable presentation (audiovisual congruent, audiovisual incongruent). The critical comparison here is for audiovisual incongruent presentation where the auditory response rate is approximately 20% greater in the presence of both concurrent articulation and silent mouthing relative to no task conditions.

A 2x3 (Syllable Presentation x Interference) repeated measure ANOVA confirmed that fewer auditory responses were reported for audiovisual incongruent presentation in comparison to audiovisual congruent presentation, shown as a main effect, *F*(1,29) = 101.97, *MSE* = 1559.39, *p*<.001, η² = .78. There was also a main effect of Interference, *F*(2,58) = 7.49, *MSE* = 145.90, *p*<.001, η² = .21, as well as an interaction between Syllable Presentation and Interference, *F*(2,58) = 10.33, *MSE* = 134.22, *p*<.001, η² = .26. Critically, the source of this interaction can be attributed to the differential effects of interference on the auditory response rate for audiovisual incongruent presentation; significantly fewer auditory responses were reported during no task conditions relative to both silent mouthing, *F*(1,29) = 6.02, *p*<.02, η² = .17, and concurrent articulation manipulations, *F*(1,29) = 16.07, *p*<.001, η² = .36, which did not differ (*p*=.18).

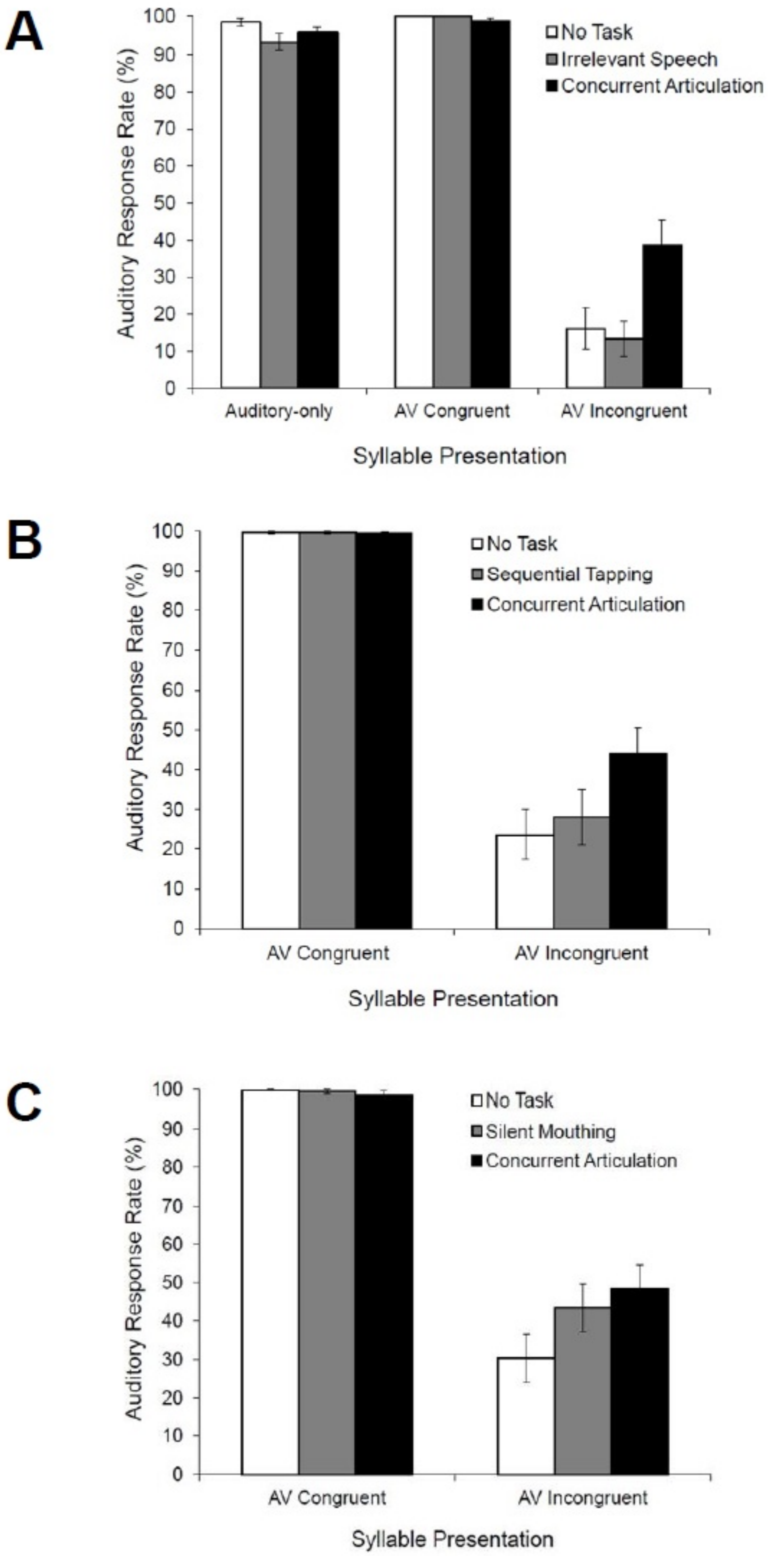

Figure 17. Mean percentage of auditory responses during auditory-only, audiovisual (AV) congruent and incongruent syllable presentation conditions at each level of interference for Experiments 8 (Panel A), 9 (Panel B) and 10 (Panel C).

Discussion

Concurrent articulation and silent mouthing both appeared to reduce the McGurk effect to an equivalent magnitude. As a result, because concurrent articulation and silent mouthing both impede the deployment of subvocal speech production processes, that both manipulations disrupted audiovisual binding to a similar extent supports the prediction that the origin of disruption resides at a subvocal level. Taken together therefore, the findings of Experiments 8, 9 and 10 provide strong support for the view that subvocal processes are involved in the integration of auditory and visual speech inputs.

In spite of this, there are a number of fundamental issues with the findings presented throughout Experiments 8 to 10: Firstly, while the McGurk effect was reduced in the presence of concurrent articulation and silent mouthing to a similar magnitude to previous dual-task studies (e.g., Alsius, et al., 2005; Alsius, et al., 2007; Tiippana, et al., 2004), this reduction was only partial. Although it is feasible that these manipulations were sub-optimal in impeding speech productions processes, it is also possible that speech production processes are not entirely necessary for audiovisual integration. Secondly, the auditory response rate was at ceiling for audiovisual congruent presentation. As a result, it may be inappropriate to compare the rate of auditory responses between audiovisual incongruent and congruent presentation since it merely indicates whether the auditory stimulus was or was not perceived. In the latter case this would be expected nearly 100-percent of the time, unless the auditory signal was degraded, say by background noise. This therefore begs an examination of audiovisual integration via some form of continuous variable, as opposed to dichotomous variable deployed in the preceding experiments.

Experiment 11

Experiment 11 intended to explore the tolerance of the McGurk effect in the presence of interference when the incongruent visual and auditory attributes of the speech signal were discordant in time. The reason for examining this was twofold: Firstly, to replicate the findings of Experiments 8 to 10, showing that concurrent articulation interferes with audiovisual binding because it disrupts speech production processes involved in the integration of auditory and visual speech. Secondly, to explicitly examine the point in time, and therefore stage, at which concurrent articulation disrupts this binding process. Consequently, time was used as a continuous dependent variable – to remove the dichotomy between the congruent and incongruent conditions in Experiments 8 to 10 – in order to verify whether concurrent articulation specifically disrupts audiovisual integration as opposed to speech processing more generally.

Critically, we now know that audiovisual binding does not require precise alignment, but is maintained over a large temporal window ranging from approximately 40ms audio lead to 240ms audio lag (e.g., Conrey & Pisoni, 2006; Massaro, Cohen, & Smeele, 1996; Munhall, et al., 1996; van Wassenhove, Grant, & Poeppel, 2007; Wiersinga-Post et al., 2010). Similarly, the McGurk effect is also evident when heard and seen speech are presented at different temporal synchronies, although illusionary, \da\ responses are observed over a smaller temporal window ranging from 30ms audio lead to 170ms audio lag (Munhall, et al., 1996; van Wassenhove, et al., 2007; Wiersinga-Post, et al., 2010). Subsequently, the current experiment examined the binding of McGurk syllables at different stimulus-onset asynchronies (SOAs).

On the basis of the precedent literature (e.g., van Wassenhove, et al., 2007), it was expected that as asynchrony increased, the McGurk effect would decrease. Most importantly, however, the McGurk effect was also expected to be reduced in the presence of concurrent articulation relative to control conditions during the stage in time at which both inputs are integrated, specifically within the temporal window of integration for audiovisual incongruent pairs – i.e., between 30ms audio lead to +170ms audio lag (van Wassenhove, et al., 2007). By comparison, given that the mere presence of task-irrelevant background speech in Experiment 8 did not disrupt the audiovisual integration process, the pattern of auditory responses was not expected to differ in the presence of irrelevant speech across all SOAs.

Method

Participants

Sixty volunteers (40 females), aged 18 to 51 years (23.39 mean years), recruited online from Cardiff University's School of Psychology, were paid for participation. All participants, none of whom had taken part in Experiments 8 to 10, were native English speakers who reported normal (or corrected-to-normal) vision and normal hearing. Thirty participants were assigned at random to either Experiment 11a or Experiment 11b.

Apparatus & Materials, Design, & Procedure

Using *SonicForge 5.0* software the sound of each audiovisual video clip was displaced with respect to the visual attributes at the following increments: 250ms, 100ms, 50ms audio lead, and 250ms, 100ms, 50ms audio lag. The factors consisted of Syllable Presentation (audiovisual congruent, audiovisual incongruent) Interference (no task, concurrent articulation, irrelevant speech), and SOA (250ms, 100ms, 50ms audio lead, 0ms, 50ms, 100ms, 250ms audio lag), with 12 trials for each of the Syllable Presentation x Interference x SOA conditions. Trials were arranged in a pseudo-random order and were balanced across all participants.

It should also be noted that because audiovisual congruent presentation does not permit the degree of integration between auditory and visual inputs to be quantified, the auditory response rate was analysed for audiovisual incongruent presentation only. Audiovisual congruent syllables were presented in order to replicate the design of Experiment 8 to 10, as well as prevent participants deducing the nature of the experimental situation.

Results

For the ease of explanation, two 2x7 (Interference x SOA) repeated measure ANOVAs were administered for Experiment 11a and 11b:

Figure 18a illustrates the auditory response rate for each interference condition (no task, concurrent articulation) in Experiment 11a. Analyses confirmed that as synchrony between the auditory and visual speech signals decreased, the auditory response rate also decreased, shown here as a main effect of SOA, *F*(6,174) = 6.15, *MSE* = 196.98, *p*<.001, η^2 = .18. Critically, however, relative to no task conditions, significantly more auditory responses were reported in the presence of concurrent articulation across all SOAs, shown here as a main effect of Interference, *F*(1,29) = 4.16, *MSE* = 1739.08, *p*<.05, η^2 = .15. Although there was no interaction between Interference and SOA (*p*=.76), planned comparisons confirmed that the rate of auditory responses only differed significantly between no task and concurrent articulation conditions at SOAs of 0ms, 50ms and 100ms audio lag only (*p*<.05), there were no significant difference between conditions for any of the auditory lead conditions (-250ms, -100ms, -50ms) or at 250ms audio lag (*p*>.05).

Figure 18b illustrates the auditory response rate for each interference condition (no task, irrelevant speech) in Experiment 11b as a function of SOA for audiovisual incongruent presentation. Again, analyses demonstrated that as synchrony between the auditory and visual speech signals decreased, the auditory response rate also decreased, shown here as a main effect of SOA, *F*(6,174) = 6.15, *MSE* = 196.98, *p*<.001, η^2 = .18. Unlike Experiment 11a, however, there was no main effect of Interference (*p*=.97), and no Interference x SOA interaction (*p*=.39), such that the auditory response rate did not differ significantly between irrelevant speech and no task conditions at any SOA.

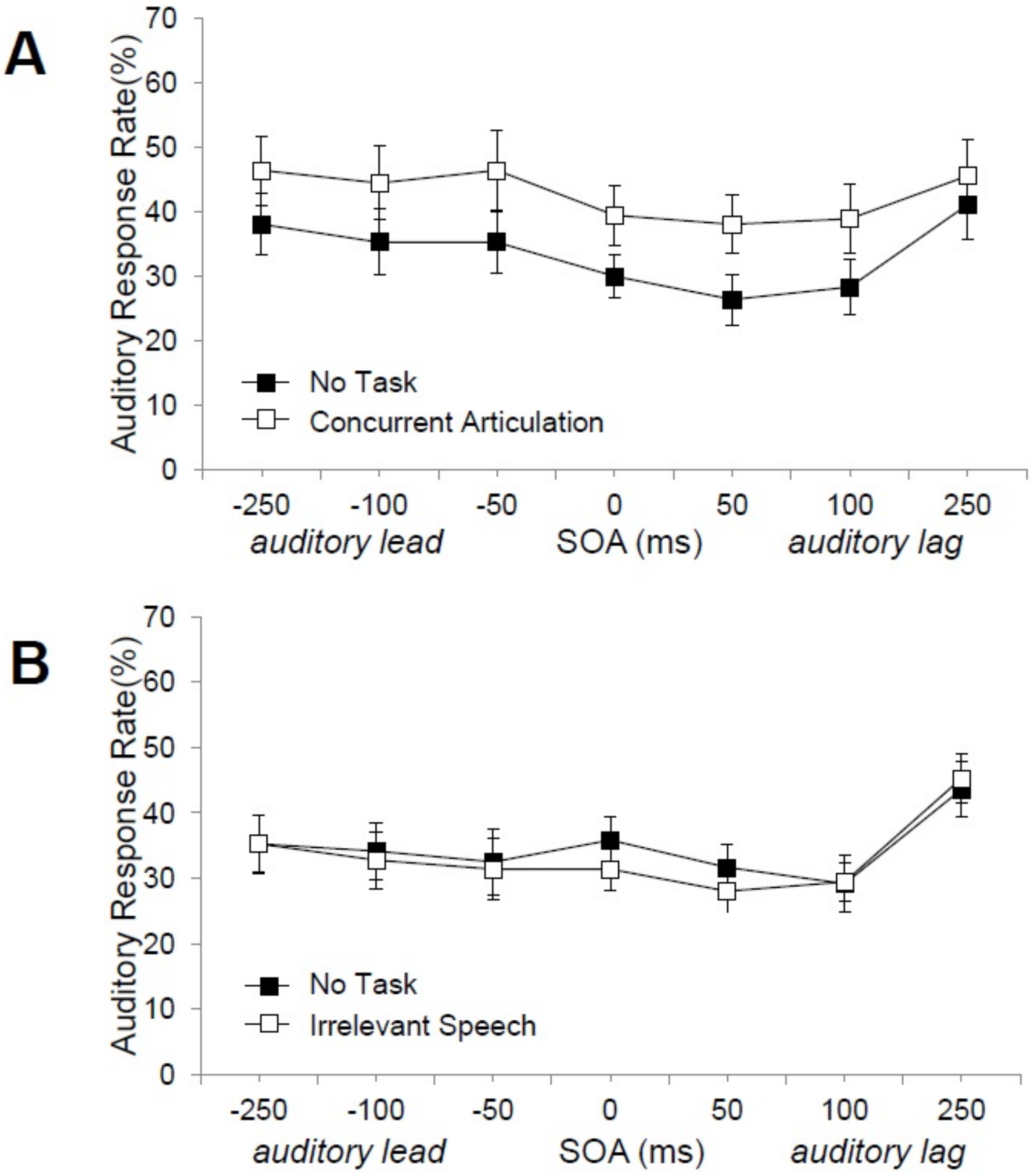

Figure 18. Mean percentage of auditory responses during Experiment 11 for audiovisual incongruent pairs as a function of stimulus-onset asynchrony (SOA). Panel A shows no task and concurrent articulation conditions (Experiment 11a), with Panel B indicating no task and irrelevant sound conditions (Experiment 11b).

Discussion

The present findings show that, relative to no task, control conditions, the McGurk effect was significantly reduced in the presence of concurrent articulation at 0, 50 and 100ms SOAs. In comparison, the McGurk effect did not differ between control and irrelevant speech conditions at any SOA measured. Arguably therefore, concurrent articulation appears to disrupt the integration of auditory and visual speech that approximately correspond to the temporal window at which incongruent auditory and visual speech come to be bound – ranging between 30ms audio lead to +170ms audio lag (see van Wassenhove, et al., 2007).

in terms of the temporal asynchrony observed here, as well as elsewhere (van Wassenhove, et al., 2007), during the integration of seen and heard speech, this could be understood in terms of the physical and neural differences between both modalities. For example, although sounds travel through the air more slowly than visual signals, there is evidence to suggest that the brain processes auditory stimuli much more rapidly (for review see Vroomen & Keetels, 2010). On this basis, according to Vroomen and Keetels (2010), in order for synchrony to be perceived between audition and vision, the visual-verbal signal should be presented slightly before the auditory-verbal signal in order to compensate for the slower neural transmission time of visual, relative to auditory inputs. Subsequently, this explanation may help to explain the current results, whereby the integration of auditory and visual inputs was disrupted by concurrent articulation when the auditory signals were temporally presented after visual speech signal.

Thus, these findings supplement those obtained from Experiments 8 to 10, strengthening the conclusion that concurrent articulation not only impedes the subvocal apparatus necessary for speech production, but also appears to specifically disrupt audiovisual binding. What this suggests is that speech production processes are therefore involved in the integration of auditory and visual speech inputs.

General Discussion

Taken together, the present findings across four experiments provide evidence to suggest that our understanding of audiovisual binding must not only incorporate influences from audition and vision, but should also consider the extent to which knowledge about how to produce speech is also involved in their integration. This conclusion stems from the robust observation that concurrently articulating irrelevant verbal material during syllable identification reduced the McGurk effect. By comparison, the presence of task-irrelevant speech did not disrupt audiovisual integration (Experiments 8 and 11). However, relative to irrelevant speech, the effect of concurrent articulation, which was in similar magnitude to that observed in previous dual-task studies (e.g., Alsius, et al., 2005; Alsius, et al., 2007; Buchan & Munhall, 2012; Tiippana, et al., 2004), could have alternatively been attributed to the cognitive demands placed on participants when administering two tasks simultaneously. This possibility was nonetheless refuted on the basis that the non-verbal task of sequential, manual finger tapping did not reduce the McGurk effect comparative to control conditions (Experiment 9).

Furthermore, concurrent articulation did not require overt vocalisation (Experiment 10) or precise temporal alignment (Experiment 11) to disrupt audiovisual integration. That is, silent mouthing reduced the McGurk effect to an equivalent extent as concurrent articulation, which also disrupted audiovisual binding within a temporal window of 0 to 100 ms – specifically when the integration of incongruent seen and heard inputs has been shown to temporally occur (van Wassenhove, et al., 2007). Taken together, both concurrent articulation and silent mouthing, by design, impede the subvocal apparatus necessary for speech production. That these manipulations also disrupt the McGurk effect suggests that speech production mechanisms are involved in audiovisual integration.

Nevertheless, the effect of both concurrent articulation and silent mouthing only partially reduced the McGurk effect in both experiments. Although these manipulations may have been sub-optimal in disrupting speech production processes, there is an increasing body of evidence showing that auditory-verbal and visual-verbal inputs come together via two pathways. Firstly, auditory and visual speech inputs appear to be processed independently in each modality prior to their integration at some higher stage of multisensory processing (Hickok & Poeppel, 2004, 2007; Okada & Hickok, 2009). As such, audiovisual binding might still be possible even when speech production processes are prevented because the integration between both modalities cannot be disrupted via this ‘direct’ pathway. Secondly, an ‘indirect’ sensory-motor (dorsal) integration pathway has been shown to provide a region of overlap between speech input and output systems (Hickok, et al., 2009; Okada & Hickok, 2009; Schwartz, Basirat, Ménard, & Sato, 2012). For Hickok (2009), however, this sensory-motor pathway is not essential for verbal recognition, but allows knowledge of speech production to exert a top-down influence during perception of sensory signals. In particular, this pathway may be required during adverse listening conditions, such as when speech sounds are severely distorted by the presence of background noise or when the auditory and visual signals do not correspond (Schwartz, Basirat, et al., 2012; Schwartz, Grimault, et al., 2012). On this basis, it may be that concurrent articulation disrupted this indirect pathway during the perception of McGurk syllables, thereby reducing, but not eliminating the integration of incongruent seen and heard speech signals.

Such an account contrasts sharply with theories that assume an essential role for articulatory motor areas in the perception of seen and heard speech signals (see e.g., Motor Theory of speech perception: Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman, Delattre, & Cooper, 1952; Liberman, Delattre, Cooper, & Gerstman, 1954; Liberman & Mattingly, 1985). Furthermore, this view challenges the existence of a mirror neuron system in humans (Craighero, Metta, Sandini, & Fadiga, 2007; Molenberghs, Cunnington, & Mattingley, 2012; Rizzolatti & Sinigaglia, 2010). Specifically, the mirror neuron account stems from the finding that regions responsible for motor planning and execution in the macaque monkey play a role in the perception and comprehension of goal-directed action, with a selection of neurons in motor regions discharging both when an action is performed, as well as during the perception of another person performing that action (e.g., Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Kohler et al., 2002; Rizzolatti, Fogassi, & Gallese, 2002). Critically, the mirror neuron system has been generalised to encompass a communicative function in humans and has been used to suggest that motor representations necessary for speech planning and production are essential when integrating auditory and lipread inputs during speech perception (Skipper, et al., 2007; Wilson, 2009). However, the evidence provided in the current chapter – namely, the observation that the McGurk effect was only partially reduced on the presence of concurrent articulation – as well as that presented elsewhere (Hickok, 2010; Hickok & Hauser, 2010; Lotto, Hickok, & Holt, 2009), suggests that speech production processes may not be entirely necessary during audiovisual binding.

The dual-pathway account also has the capacity to explain the relatively complex set of findings obtained in Chapter 2. That is, while verbal short-term memory performance for lipread sequences resembles that of auditory sequences within the recency portion of the serial position curve, this auditory-like behaviour is dependent upon speech production processes – as evidence by a reduced lipread recency effect in the presence of concurrent articulation or irrelevant sound (see Experiment 1). These findings point to an effect whereby lipread information appears to gain access to an auditory-like representation via the speech production pathway, providing evidence in favour of the existence of an indirect pathway linking auditory and visual aspects of speech. So, if this pathway is otherwise engaged in speech production of additional material, auditory-like performance observed with lipread speech will be disrupted. In spite of this, lipread and auditory suffix effects on lipread and auditory lists were shown to be driven by fundamentally different mechanisms. These findings might therefore be used to support the notion that modality-specific perceptual, motor and perceptual-motor processes, as well as modality-general attentional mechanisms, support these effects in each case (see Experiments 2 to 7).

These findings might therefore be understood in terms of verbal performance being based, not on the operation of an amodal level of representation, but rather on the opportunistic co-opting of attentional, modality-specific perceptual, motor and perceptual-motor processes to meet the demands of the particular verbal task. In support, verbal short-term memory performance may be an emergent by-product of the sensory-motor integration network (Aboitiz & Garcı́a V, 1997; Buchsbaum & D'Esposito, 2008; Hickok, 2012; Hickok, et al., 2003; Jacquemot & Scott, 2006). According to Hickok (2009), auditory and visual speech inputs are processed within modality-specific perceptual systems, with speech production mechanisms supporting their maintenance. However, when required, the sensory-motor integration network may mediate the relationship between these systems. For this account then, verbal short-term memory performance is underpinned, not by bespoke storage systems operating at an amodal level of representation, but by the joint action of systems involved in sensory and motor processes, as well as an additional network involved in sensory-motor integration.

In conclusion, the evidence provided throughout the current chapter contributes to our understanding of the mechanisms underpinning the integration of seen and heard speech, as well as verbal performance more generally. That is, the finding that both concurrent articulation and silent mouthing reduce the McGurk effect suggests that subvocal mechanisms necessary for speech production are involved in audiovisual binding. Nevertheless, the data does not completely refute the possibility that auditory and visual speech inputs are also integrated obligatorily, since audiovisual binding was still evident even when the motor system was engaged by the production of irrelevant verbal material. Critically then, the present findings seem to suggest that auditory-verbal and visual-verbal information is processed independently prior to integration at some higher level of multisensory processing, as well as, under certain circumstances, via a speech production pathway.

CHAPTER 4

Thesis Discussion

The current thesis utilised two behavioural paradigms to show that verbal performance is based, not on the operation of an amodal, phonological level of representation, but modality-specific representations. A summary of each experimental chapter will be presented first, which will then be followed by a discussion of the potential implications of the present data.

Firstly, in Chapter 2 the functional similarities between auditory and lipread speech in verbal short-term memory was examined. It was shown that, despite there being similarities between the serial recall of auditory and lipread sequences, different mechanisms actually gave rise to superficially similar effects across modalities. That is, the elements of auditory perceptual processing, as manifested in enhanced recency for heard over written material, were immune to manipulations that impede the speech-rehearsal process (i.e., articulatory suppression and irrelevant speech). By comparison, whatever gave rise to lipread recency was not immune. Furthermore, cross-modal interactions between lipread and auditory suffix effects on lipread lists were shown to be driven by fundamentally different mechanisms. The effect of an auditory suffix on lipread recency was due to attention being captured away from the lipread sequence by an unexpected auditory event, whereas the impact of a lipread suffix on a lipread list was attributable to the misidentification of the lipread suffix. These disparities therefore point to a divergence in the mechanisms giving rise to recency and suffix effects found in both modalities. This not only highlights important distinctions in the nature of the verbal representations derived from seen and heard speech, but also undermines the case for a common, amodal form of representation upon which these modes of speech come together.

Secondly, in Chapter 3 the McGurk effect was utilised in order to understand how and where auditory and visual modes of speech come to be bound. The critical test here was that audiovisual binding was evaluated in the presence of verbal and non-verbal interference to investigate the degree to which motor processes necessary for speech production are also involved in the integration of both inputs. The reason for exploring this stemmed from the findings of Chapter 2, as well evidence also showing that, while seen and heard speech information appear to converge at a level typically associated with auditory processing (for review, see Alais, et al., 2010), visual speech might actually gain access to auditory-like representations via the speech production pathway (Hickok, 2009; Hickok & Poeppel, 2004, 2007; Skipper, et al., 2009; Skipper, et al., 2007).

Critically, support for this latter prediction was garnered from the finding that the McGurk effect was only reduced when participants concurrently articulated or silently mouthed irrelevant verbal material during syllable identification. Furthermore, to-be-ignored, task irrelevant speech, and the non-verbal, dual-task of sequential tapping did not significantly reduce the McGurk effect relative to control, no task conditions. As a result, because concurrent articulation and silent mouthing impede subvocal speech production processes, that these manipulations disrupted the McGurk effect to an equivalent magnitude suggests that subvocal processes are also involved in the integration of auditory and visual speech. Nevertheless, the McGurk effect was only partially reduced in the presence of concurrent articulation, leading to the conclusion that seen and heard speech inputs not only come together via a speech production pathway, but may also be processed independently prior to integration at some higher level of multisensory processing.

Overall, the evidence presented throughout this thesis can be interpreted in terms of an increasingly plausible view of verbal performance as being based on modality-specific functions. This account consequently goes further than considerations of linguistic behaviour that just require an amodal form of representation, in maintaining that verbal performance is not simply underpinned by mechanisms that are essentially amodal in nature. In support, a range of behavioural and neuroscientific findings also seems to point to this.

For example, the irrelevant sound effect provides one of the most robust demonstrations in support of an account of verbal performance as being based, not on the action of mechanisms operating on amodal, phonological representations, but on the operation of systems supporting perceptual organisation and motor planning. The disruptive effect on verbal short-term memory from task-irrelevant verbal material has typically been ascribed to the degree of resemblance between the phonological identity of the irrelevant sounds and the to-be-remembered items in bespoke mnemonic storage systems (e.g., LeCompte, 1996; LeCompte, Neely, & Wilson, 1997). However, the disruptive effect of task-irrelevant non-verbal material, such as pitch glides (Jones, et al., 1993; Klatte, Kilcher, & Hellbrück, 1995), tones (Divin, Coyle, & James, 2001; Jones, Alford, Bridges, Tremblay, & Macken, 1999; Jones, et al., 1992; Little, Martin, & Thomson, 2010; Neath, Surprenant, & LeCompte, 1998), sine-wave speech (Tremblay, Nicholls, Alford, & Jones, 2000), bandpass noise (Tremblay, Macken, & Jones, 2001), and music (Alley & Greene, 2008; Schlittmeier, et al., 2008), is functionally equivalent to that found with task-irrelevant verbal material. Consequently, as opposed to its physical similarity to the memory list, the degree of change within the irrelevant sequence is actually a key determinant of disruption, termed the changing-state effect (Jones, et al., 1992). The other key determinant is the extent to which rehearsal is necessary; irrelevant speech is more disruptive when the task involves memory for order. The changing-state effect of irrelevant speech has therefore been explained in terms of a conflict between two seriation processes: one involved in rehearsal of the to-be-remembered items and the other a by-product of organising the irrelevant elements into a single sound stream (Beaman & Jones, 1997; Hughes, et al., 2005; Jones, Hughes, & Macken, 2010; Jones & Macken, 1993). Thus, rather than relying on the operation of a phonological level of representation, the disruptive effect of irrelevant speech points to two distinct components – one underpinned by speech production mechanisms utilised to perform subvocal rehearsal, and the other arising within auditory perceptual sequence processing. These behavioural data accord with results from neuroimaging studies which have also demonstrated that neural networks that are engaged during auditory verbal tasks, as well as being engaged during speaking and passive listening to speech, show similar patterns of activation when listening to non-speech sounds (Chang, Kenney, Loucks, Poletto, & Ludlow, 2009; Hickok, et al., 2009).

The findings reported throughout the current thesis add to this emerging picture, strengthening an account of verbal processing that locates performance within modality-specific perceptual, motor, and perceptual-motor processes. Namely, it is shown that an understanding of verbal performance must not only incorporate influences from auditory and visual perceptual processes, but should also consider the extent to which motor processes necessary for speech production are also involved in linguistic behaviour. The key findings in this respect are that auditory-like behaviour in verbal short-term memory found with lipread presentation appears to be dependent upon speech production processes – as evidenced by a reduced lipread recency effect in the presence of concurrent articulation or irrelevant sound. This result subsequently points to an effect whereby lipread information appears to gain ready access to an auditory-like representation via speech production processes – a prediction supported by the reduction of the McGurk effect in the presence of concurrent articulation. Nevertheless, not only was the observed reduction of the McGurk effect partial, it was also shown that lipread and auditory suffix effects on lipread and auditory lists were driven by fundamentally different mechanisms. As such, it is concluded that similarities in verbal performance across auditory and lipread presentation modalities, which have been explained in terms of an amodal level of representation, like the effect of irrelevant sound, actually reflect emergent by-products of systems primarily serving modality-specific perceptual, motor, and perceptual-motor processes.

Critically, such an account of verbal behaviour will have critical implications for understanding the aetiology of disorders associated with deficits in verbal performance (for review seeHickok, Houde, & Rong, 2011). Conduction aphasia, for instance, has typically been characterised as a disorder involving the storage of phonological representations in verbal short-term memory, whereby patients can comprehend speech but have difficulty repeating it verbatim (for a detailed discussion of the relation between verbal short-term memory and conduction aphasia, see Buchsbaum & D'Esposito, 2008). As a result, conduction aphasia has been explained in terms of patients inability to retain the phonological details of an utterance, so that they paraphrase or make frequent phonemic errors and repeated self-correction attempts (Buchsbaum et al., 2011; Damasio, 1992; Damasio & Geschwind, 1984; Goodglass, 1992). More recently, however, it has been shown that sensory-motor integration processes, as opposed to an amodal level of representation, provide a more accurate account of this disorder. This view stems from the finding that damage to brain regions sensitive to auditory, visual and motor modes of speech are the source of this deficit (Buchsbaum, et al., 2011; Hickok, 2012; Hickok, et al., 2011). These findings, coupled with the current data, subsequently provide a novel framework for future studies to examine the contribution of mechanisms involved in the integration of sensory (i.e., modality-specific perceptual) and motor processes, as well as its expected connection with disorders associated with verbal processing.

Taken together, a fundamental concern in the study of linguistic behaviour has been whether verbal performance requires a modality-general, amodal level representation or can be explained in terms of modality-specific perceptual and motor-output processes. This thesis subsequently addressed this issue, presenting findings from two behavioural paradigms to show that an amodal representational form cannot adequately explain all aspects of verbal performance. Rather, the current data provides novel evidence that adds to a growing body of work showing that modality-specific perceptual and motor processes (e.g., Hickok, 2009; Hughes, et al., 2009; Macken, et al., 2009; Wilson & Fox, 2007), as well as attentional mechanisms (Hughes, et al., 2005, 2007; Vachon, et al., 2012), determine verbal performance. As such, it is concluded that rather than relying on an amodal, phonological level of representation, what have typically been regarded as 'peripheral' aspects of performance (i.e., perceptual and speech planning processes) should also be considered when understanding the mechanism, or mechanisms, involved in verbal behaviour.

REFERENCES

Aboitiz, F., & García V, R. (1997). The evolutionary origin of the language areas in the human brain. A neuroanatomical perspective. *Brain Research Reviews, 25*(3), 381-396.

Acheson, D. J., & MacDonald, M. C. (2009). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological bulletin, 135*(1), 50-68.

Alais, D., Newell, F. N., & Mamassian, P. (2010). Multisensory processing in review: From physiology to behaviour. *Seeing and Perceiving, 23*(1), 3-38.

Alley, T. R., & Greene, M. E. (2008). The relative and perceived impact of irrelevant speech, vocal music and non-vocal music on working memory. *Current Psychology, 27*(4), 277-289.

Alloway, T. P., Kerr, I., & Langheinrich, T. (2010). The effect of articulatory suppression and manual tapping on serial recall. *European Journal of Cognitive Psychology, 22*(2), 297-305.

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology, 15*, 839-843.

Alsius, A., Navarra, J., & Soto-Faraco, S. (2007). Attention to touch weakens audiovisual speech integration. *Experimental Brain Research, 183*(3), 399-404.

Baddeley, A. (1986). *Working Memory* (Vol. 11). Oxford: Oxford University Press, Clarendon Press.

Baddeley, A. (1990). *Human memory: Theory and practice*. London: Lawrence Erlbaum Associates.

Baddeley, A. (1992). Is working memory working? The fifteenth Bartlett lecture. *The Quarterly Journal of Experimental Psychology, 44*(1), 1-31.

Baddeley, A. (2010). Working memory. *Current Biology, 20*(4), R136-R140.

Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annual Review of Psychology, 63*, 1-29.

Baddeley, A., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*(6), 575-589.

Beaman, C. P., & 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 & Cognition, 23*(2), 459-471.

Belin, P., Bestelmeyer, P. E., Latinus, M., & Watson, R. (2011). Understanding voice perception. *British Journal of Psychology, 102*(4), 711-725.

Bernstein, L. E., Auer Jr, E. T., & Moore, J. K. (2004). Audiovisual speech binding: Convergence or association? In G. Calvert, C. Spence & B. E. Stein (Eds.), *Handbook of Multisensory Processing* (pp. 202-203). Cambridge, MA: MIT Press.

Bernstein, L. E., Auer Jr, E. T., Moore, J. K., Ponton, C. W., Don, M., & Singh, M. (2002). Visual speech perception without primary auditory cortex activation. *Neuroreport, 13*(3), 311-315.

Bishop, C. W., & Miller, L. M. (2009). A multisensory cortical network for understanding speech in noise. *Journal of Cognitive Neuroscience, 21*(9), 1790-1804.

Bloom, L. C. (2006). Two-component theory of the suffix effect: Contrary evidence. *Memory & Cognition, 34*(3), 648-667.

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

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT press.

Bregman, A. S., & Rudnicky, A. I. (1975). Auditory segregation: Stream or streams? *Journal of Experimental Psychology: Human Perception and Performance, 1*(3), 263-267.

Buchan, J. N., & Munhall, K. G. (2012). The effect of a concurrent working memory task and temporal offsets on the integration of auditory and visual speech information. *Seeing and Perceiving, 25*(1), 87-106.

Buchsbaum, B. R., Baldo, J., Okada, K., Berman, K. F., Dronkers, N., D'Esposito, M., et al. (2011). Conduction aphasia, sensory-motor integration, and phonological short-term memory–an aggregate analysis of lesion and fMRI data. *Brain and Language, 119*(3), 119-128.

Buchsbaum, B. R., & D'Esposito, M. (2008). The search for the phonological store: from loop to convolution. *Journal of Cognitive Neuroscience, 20*(5), 762-778.

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review, 106*(3), 551-581.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., et al. (1997). Activation of auditory cortex during silent lipreading. *Science, 276*(5312), 593-596.

Campbell, R., & Dodd, B. (1980). Hearing by eye. *Quarterly Journal of Experimental Psychology, 32*(1), 85-99.

Campbell, R., & Dodd, B. (1982). Some suffix effects on lipread lists. *Canadian Journal of Psychology, 36*(3), 508-514.

Casserly, E. D., & Pisoni, D. B. (2010). Speech perception and production. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(5), 629-647.

Chang, S.-E., Kenney, M. K., Loucks, T. M., Poletto, C. J., & Ludlow, C. L. (2009). Common neural substrates support speech and non-speech vocal tract gestures. *NeuroImage, 47*(1), 314-325.

Cheng, C.-M. (1974). Different roles of acoustic and articulatory information in short-term memory. *Journal of Experimental Psychology, 103*(4), 614-618.

Chomsky, N. (1959). On certain formal properties of grammars. *Information and Control, 2*(2), 137-167.

Chomsky, N. (2002). *Syntactic Structures*. Berlin: de Gruyter Mouton.

Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper and Row.

Colle, H. A., & Welsh, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning and Verbal Behavior, 15*(1), 17-31.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology, 55*(4), 429-432.

Conrey, B., & Pisoni, D. B. (2006). Auditory-visual speech perception and synchrony detection for speech and nonspeech signals. *The Journal of the Acoustical Society of America, 119*(6), 4065-4073.

Craighero, L., Metta, G., Sandini, G., & Fadiga, L. (2007). The mirror-neurons system: Data and models. In C. v. Hofsten & K. Rosander (Eds.), *Progress in Brain Research* (Vol. 164, pp. 39-59). Oxford: Elsevier.

Crowder, R. G. (1971). Waiting for the stimulus suffix: Decay, delay, rhythm, and readout in immediate memory. *The Quarterly Journal of Experimental Psychology, 23*(3), 324-340.

Crowder, R. G., Harvey, N., Routh, D., & Crowder, R. (1983). The Purity of Auditory Memory [and Discussion]. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 302*(1110), 251-265.

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

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology, 19*(5), 381-385.

Damasio, A. R. (1992). Aphasia. *New England Journal of Medicine, 326*(8), 531-539.

Damasio, A. R., & Geschwind, N. (1984). The neural basis of language. *Annual Review of Neuroscience, 7*(1), 127-147.

de Gelder, B., & Vroomen, J. (1992). Abstract versus modality-specific memory representations in processing auditory and visual speech. *Memory & Cognition, 20*(5), 533-538.

de Gelder, B., & Vroomen, J. (1994). A new place for modality in a modular mind. *Current Psychology of Cognition, 13*, 84-91.

Divin, W., Coyle, K., & James, D. (2001). The effects of irrelevant speech and articulatory suppression on the serial recall of silently presented lipread digits. *British Journal of Psychology, 92*(4), 593-616.

Dodd, B., Hobson, P., Brasher, J., & Campbell, R. (1983). Deaf children's short-term memory for lip-read, graphic and signed stimuli. *British journal of developmental psychology, 1*(4), 353-364.

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

Ellis, A. W. (1980). Errors in speech and short-term memory: The effects of phonemic similarity and syllable position. *Journal of Verbal Learning and Verbal Behavior, 19*(5), 624-634.

Engle, R. W., Cantor, J., & Turner, M. (1989). Modality effects: Do they fall on deaf ears? *The Quarterly Journal of Experimental Psychology, 41*(2), 273-292.

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience, 15*(2), 399-402.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1*(1), 1-47.

Frankish, C. (1985). Modality-specific grouping effects in short-term memory. *Journal of Memory and Language, 24*(2), 200-209.

Frankish, C. (1989). Perceptual organisation and Precategorical acoustic storage. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15*(3), 469-479.

Frankish, C. (2008). Precategorical acoustic storage and the perception of speech. *Journal of Memory and Language, 58*(3), 815-836.

Frankish, C. (Ed.). (1996). *Auditory short-term memory and the perception of speech*: Psychology Press.

Frankish, C., & Turner, J. (1984). Delayed suffix effect at very short delays. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(4), 767-777.

Frick, R. W. (1988). The role of memory in attenuations of the suffix effect. *Memory & Cognition, 16*(1), 15-22.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119*(2), 593-609.

Gathercole, S. E. (1987). Lip-reading: Implications for theories of short-term memory. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: Experimental studies in the psychology of lipreading* (pp. 67-82). Hildale, NJ: Lawrence Erlbaum Associates, Inc.

Goodglass, H. (1992). Diagnosis of conduction aphasia. In S. E. Kohn (Ed.), *Conduction Aphasia* (pp. 39-49). Hillsdale, NY: Lawrence Erlbaum Associates, Inc.

Green, K. P. (1998). The use of auditory and visual information during phonetic processing: Implications for theories of speech perception. . In R. Campbell, B. Dodd & D. Burnha (Eds.), *Hearing By Eye II: Advances in the Psychology of Speechreading and Auditory-visual Speech* (pp. 3-25). Hove, UK: Psychology Press. .

Greenberg, S. N., & Engle, R. W. (1983). Voice change in the stimulus suffix effect: Are the effects structural or strategic? *Memory & Cognition, 11*(5), 551-556.

Greene, R. L. (1991). Serial recall of two-voice lists: Implications for theories of auditory recency and suffix effects. *Memory & Cognition, 19*(1), 72-78.

Greene, R. L., & Crowder, R. G. (1984). Modality and suffix effects in the absence of auditory stimulation. *Journal of Verbal Learning and Verbal Behavior, 23*(3), 371-382.

Hall, D., & Gathercole, S. E. (2011). Serial recall of rhythms and verbal sequences: Impacts of concurrent tasks and irrelevant sound. *The Quarterly Journal of Experimental Psychology, 64*(8), 1580-1592.

Hall, M., & Bavelier, D. (2010). Working memory, deafness and sign language. In M. Marschark & P. E. Spencer (Eds.), *The Handbook of Deaf Studies, Language and Education* (Vol. 2, pp. 458-487). Oxford: Oxford University Press.

Hanson, V. L. (1982). Short-term recall by deaf signers of American Sign Language: implications of encoding strategy for order recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*(6), 572-583.

Harris, R. W. (1989). The stimulus suffix effect and positional uncertainty. *Canadian Journal of Psychology/Revue canadienne de psychologie, 43*(1), 74-87.

Henson, R., Hartley, T., Burgess, N., Hitch, G., & Flude, B. (2003). Selective interference with verbal short-term memory for serial order information: A new paradigm and tests of a timing-signal hypothesis. *Quarterly Journal of Experimental Psychology Section A, 56*(8), 1307-1334.

Henson, R., Norris, D., Page, M., & Baddeley, A. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *The Quarterly Journal of Experimental Psychology: Section A, 49*(1), 80-115.

Hickok, G. (2009). The functional neuroanatomy of language. *Physics of Life Reviews, 6*(3), 121-143.

Hickok, G. (2010). The role of mirror neurons in speech perception and action word semantics. *Language and Cognitive Processes, 25*(6), 749-776.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience, 13*, 135-145.

Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience, 15*(5), 673-682.

Hickok, G., & Hauser, M. (2010). (Mis) understanding mirror neurons. *Current Biology, 20*(14), R593-R594.

Hickok, G., Holt, L. L., & Lotto, A. J. (2009). Response to Wilson: What does motor cortex contribute to speech perception? *Trends in cognitive sciences, 13*(8), 330-331.

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron, 69*(3), 407-422.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition, 92*(1-2), 67-99.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*(5), 393-402.

Hughes, R. W., Marsh, J. E., & Jones, D. M. (2009). Perceptual–gestural (mis) mapping in serial short-term memory: The impact of talker variability. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(6), 1411-1425.

Hughes, R. W., Vachon, F., & Jones, D. M. (2005). Auditory attentional capture during serial recall: Violations at encoding of an algorithm-based neural model? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(4), 736-749.

Hughes, R. W., Vachon, F., & Jones, D. M. (2007). Disruption of short-term memory by changing and deviant sounds: Support for a duplex-mechanism account of auditory distraction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(6), 1050-1061.

Jacquemot, C., & Scott, S. K. (2006). What is the relationship between phonological short-term memory and speech processing? *Trends in cognitive sciences, 10*(11), 480-486.

Jiang, J., & Bernstein, L. E. (2011). Psychophysics of the McGurk and other audiovisual speech integration effects. *Journal of Experimental Psychology: Human Perception and Performance, 37*(4), 1193-1209.

Jones, D. M., Alford, D., Bridges, A., Tremblay, S., & Macken, B. (1999). Organizational 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*(2), 464-473.

Jones, D. M., Banbury, S. P., Tremblay, S., & Macken, W. J. (1999). *The effect of task-irrelevant sounds on cognitive performance.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Jones, D. M., Farrand, P., Stuart, G., & Morris, N. (1995). Functional equivalence of verbal and spatial information in serial short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 1008-1018.

Jones, D. M., Hughes, R. W., & Macken, W. J. (2006). Perceptual organization masquerading as phonological storage: Further support for a perceptual-gestural view of short-term memory. *Journal of Memory and Language, 54*(2), 265-281.

Jones, D. M., Hughes, R. W., & Macken, W. J. (2010). Auditory distraction and serial memory: The avoidable and the ineluctable. *Noise and Health, 12*(49), 201-209.

Jones, D. M., & 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*(2), 369-381.

Jones, D. M., Macken, W. J., & Murray, A. C. (1993). Disruption of visual short-term memory by changing-state auditory stimuli: The role of segmentation. *Memory & Cognition, 21*(3), 318-328.

Jones, D. M., Macken, W. J., & 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*(3), 656-674.

Jones, D. M., Madden, C., & Miles, C. (1992). Privileged access by irrelevant speech to short-term memory: The role of changing state. *The Quarterly Journal of Experimental Psychology, 44*(4), 645-669.

Jones, D. M., Saint-Aubin, J., & Tremblay, S. (1999). Modulation of the irrelevant sound effect by organizational factors: Further evidence from streaming by location. *The Quarterly Journal of Experimental Psychology: Section A, 52*(3), 545-554.

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology, 59*, 193-224.

Kahneman, D., & Henik, A. (1981). Perceptual organization and attention. In M. Kubovy & J. R. Pomerantz (Eds.), *Percetual organization*. Hilldale, NJ: Erlbaum.

Klatte, M., Kilcher, H., & Hellbrück, J. (1995). Wirkungen der zeitlichen Struktur von Hintergrundschall auf das Arbeitsgedächtnis und ihre theoretischen und praktischen Implikationen. *Zeitschrift für experimentelle Psychologie, 42*(4), 517-544.

Kohler, E., Keysers, C., Umilta, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science, 297*(5582), 846-848.

Koo, D., Crain, K., LaSasso, C., & Eden, G. F. (2008). Phonological Awareness and Short-Term Memory in Hearing and Deaf Individuals of Different Communication Backgrounds. *Annals of the New York Academy of Sciences, 1145*(1), 83-99.

Larsen, J. D., & Baddeley, A. (2003). Disruption of verbal STM by irrelevant speech, articulatory suppression, and manual tapping: Do they have a common source? *Quarterly Journal of Experimental Psychology Section A, 56*(8), 1249-1268.

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*(5), 1154-1165.

LeCompte, D. C., Neely, C. B., & 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*(2), 472-483.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*(6), 431-461.

Liberman, A. M., Delattre, P., & Cooper, F. S. (1952). The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *The American Journal of Psychology, 65*, 497-516.

Liberman, A. M., Delattre, P. C., Cooper, F. S., & Gerstman, L. J. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs: General and Applied, 68*(8), 1-13.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*(1), 1-36.

Little, J. S., Martin, F. H., & Thomson, R. H. (2010). Speech versus non-speech as irrelevant sound: Controlling acoustic variation. *Biological psychology, 85*(1), 62-70.

Lotto, A. J., Hickok, G. S., & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences, 13*(3), 110-114.

Macken, W. J., & 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*(2), 436-448.

Macken, W. J., & Jones, D. M. (2003). Reification of phonological storage. *Quarterly Journal of Experimental Psychology Section A, 56*(8), 1279-1288.

Macken, W. J., Mosdell, N., & Jones, D. M. (1999). Explaining the irrelevant-sound effect: Temporal distinctiveness or changing state? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(3), 810-814.

Macken, W. J., Phelps, F. G., & Jones, D. M. (2009). What causes auditory distraction? *Psychonomic Bulletin & Review, 16*(1), 139-144.

Maidment, D. W., & Macken, W. J. (2012). The ineluctable modality of the audible: Perceptual determinants of auditory verbal short-term memory. *Journal of Experimental Psychology: Human Perception and Performance, 38*(4), 989-997.

Marsh, J. E., & Jones, D. M. (2010). Cross-modal distraction by background speech: What role for meaning? *Noise and Health, 12*(49), 210-216.

Massaro, D. W., Cohen, M. M., & Smeele, P. M. (1996). Perception of asynchronous and conflicting visual and auditory speech. *The Journal of the Acoustical Society of America, 100*, 1777-1786.

McGettigan, C., Faulkner, A., Altarelli, I., Obleser, J., Baverstock, H., & Scott, S. K. (2012). Speech comprehension aided by multiple modalities: Behavioural and neural interactions. *Neuropsychologia, 50*(5), 762-776.

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

Molenberghs, P., Cunnington, R., & Mattingley, J. B. (2012). Brain regions with mirror properties: a meta-analysis of 125 human fMRI studies. *Neuroscience & Biobehavioral Reviews, 36*(1), 341-349.

Morton, J., Crowder, R. G., & Prussin, H. A. (1971). Experiments with the stimulus suffix effect. *Journal of Experimental Psychology Monograph, 28*(1), 169-190.

Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Attention, Perception, & Psychophysics, 58*(3), 351-362.

Murakami, T., Restle, J., & Ziemann, U. (2011). Effective connectivity hierarchically links temporoparietal and frontal areas of the auditory dorsal stream with the motor cortex lip area during speech perception. *Brain and Language, 122*, 135-141.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001). 'Primitive intelligence'in the auditory cortex. *Trends in Neurosciences, 24*(5), 283-288.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18*(3), 251-269.

Nairne, J. S., & Walters, V. L. (1983). Silent mouthing produces modality-and suffix-like effects. *Journal of Verbal Learning and Verbal Behavior, 22*(4), 475-483.

Nasir, S. M., & Ostry, D. J. (2009). Auditory plasticity and speech motor learning. *Proceedings of the National Academy of Sciences, 106*(48), 20470-20475.

Neath, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic bulletin & review, 7*(3), 403-423.

Neath, I., & Nairne, J. S. (1995). Word-length effects in immediate memory: Overwriting trace decay theory. *Psychonomic Bulletin & Review, 2*(4), 429-441.

Neath, I., Surprenant, A. M., & LeCompte, D. C. (1998). Irrelevant speech eliminates the word length effect. *Memory & Cognition, 26*(2), 343-354.

Nicholls, A. P., & Jones, D. M. (2002). Capturing the suffix: Cognitive streaming in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(1), 12-28.

Okada, K., & Hickok, G. (2009). Two cortical mechanisms support the integration of visual and auditory speech: A hypothesis and preliminary data. *Neuroscience letters, 452*(3), 219-223.

Page, M., & Norris, D. (1998). The primacy model: a new model of immediate serial recall. *Psychological Review, 105*(4), 761-781.

Penney, C. G. (1989). Modality effects in delayed free recall and recognition: Visual is better than auditory. *The Quarterly Journal of Experimental Psychology, 41*(3), 455-470.

Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience, 139*(1), 23-38.

Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage, 62*(2), 816-847.

Repovs, G., & Baddeley, A. (2006). The multi-component model of working memory: Explorations in experimental cognitive psychology. *Neuroscience, 139*(1), 5-22.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2002). Motor and cognitive functions of the ventral premotor cortex. *Current Opinion in Neurobiology, 12*(2), 149-154.

Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nature Reviews Neuroscience, 11*(4), 264-274.

Salamé, P., & Baddeley, A. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior, 21*(2), 150-164.

Salamé, P., & Baddeley, A. (1990). The effects of irrelevant speech on immediate free recall. *Bulletin of the Psychonomic Society, 28*(6), 540-542.

Sánchez-García, C., Alsius, A., Enns, J. T., & Soto-Faraco, S. (2011). Cross-Modal Prediction in Speech Perception. *PloS one, 6*(10), e25198.

Sarmiento, B. R., Shore, D. I., Milliken, B., & Sanabria, D. (2012). Audiovisual interactions depend on context of congruency. *Attention, Perception, & Psychophysics*, 1-12.

Sato, M., Buccino, G., Gentilucci, M., & Cattaneo, L. (2010). On the tip of the tongue: Modulation of the primary motor cortex during audiovisual speech perception. *Speech Communication, 52*(6), 533-541.

Schlittmeier, S. J., Hellbrück, J., & Klatte, M. (2008). Does irrelevant music cause an irrelevant sound effect for auditory items? *European Journal of Cognitive Psychology, 20*(2), 252-271.

Schwartz, J.-L., Basirat, A., Ménard, L., & Sato, M. (2012). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics, 25*(5), 336-354.

Schwartz, J.-L., Grimault, N., Hupé, J.-M., Moore, B. C., & Pressnitzer, D. (2012). Multistability in perception: binding sensory modalities, an overview. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1591), 896-905.

Skipper, J. I., Goldin-Medow, S., Nusbaum, H., & Small, S. L. (2009). Speech-associated gestures, Broca's areas and the human mirror system. *Brain and Language, 101*(3), 260-277.

Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2005). Listening to talking faces: motor cortical activation during speech perception. *Neuroimage, 25*(1), 76-89.

Skipper, J. I., Van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007). Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex, 17*(10), 2387-2399.

Spöehr, K. T., & Corin, W. J. (1978). The stimulus suffix effect as a memory coding phenomenon. *Memory & Cognition, 6*(6), 583-589.

Summerfield, Q. (1992). Lipreading and audio-visual speech perception. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 335*(1273), 71-78.

Summerfield, Q., MacLeod, A., McGrath, M., & Brooke, M. (1989). Lips, teeth, and the benefits of lipreading. In A. W. Young & H. D. Ellis (Eds.), *Handbook of research on face processing* (pp. 223-233). Amsterdam: Elsevier Science Publishers.

Sussman, E. S. (2005). Integration and segregation in auditory scene analysis. *The Journal of the Acoustical Society of America, 117*(3), 1285-1298.

Tiippana, K., Andersen, T., & Sams, M. (2004). Visual attention modulates audiovisual speech perception. *European Journal of Cognitive Psychology, 16*(3), 457-472.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97-136.

Tremblay, S., Macken, W. J., & Jones, D. M. (2001). The impact of broadband noise on serial memory: Changes in band-pass frequency increase disruption. *Memory, 9*(4-6), 323-331.

Tremblay, S., Nicholls, A. P., Alford, D., & Jones, D. M. (2000). The irrelevant sound effect: Does speech play a special role? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(6), 1750-1754.

Turner, M., LaPointe, L., Cantor, J., Reeves, C. H., Griffeth, R. H., & Engle, R. W. (1987). Recency and suffix effects found with auditory presentation and with mouthed visual presentation: They're not the same thing. *Journal of Memory and Language, 26*, 138-164.

Vachon, F., Hughes, R. W., & Jones, D. M. (2012). Broken expectations: Violation of expectancies, not novelty, captures auditory attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition; Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(1), 164-177.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia, 45*(3), 598-607.

Vatakis, A., & Spence, C. (2007). Crossmodal binding: Evaluating the "unity assumption" using audiovisual speech stimuli. *Attention, Perception, & Psychophysics, 69*(5), 744-756.

Villacorta, V. M., Perkell, J. S., & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *The Journal of the Acoustical Society of America, 122*, 2306-2319.

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: a tutorial review. *Attention, Perception, & Psychophysics, 72*(4), 871-884.

Warren, R. M. (1999). *Auditory perception: A new analysis and synthesis*. Cambridge: Cambridge University Press.

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia, 41*(8), 989-994.

Wiersinga-Post, E., Tomaskovic, S., Slabu, L., Renken, R., de Smit, F., & Duifhuis, H. (2010). Decreased BOLD responses in audiovisual processing. *Neuroreport, 21*(18), 1146-1151.

Wilson, M. (2001). The case for sensorimotor coding in working memory. *Psychonomic Bulletin & Review, 8*(1), 44-57.

Wilson, M. (2009). Speech perception when the motor system is compromised. *Trends in Cognitive Sciences, 13*(8), 329-330.

Wilson, M., & Fox, G. (2007). Working memory for language is not special: Evidence for an articulatory loop for novel stimuli. *Psychonomic bulletin & review, 14*(3), 470-473.

Winkler, I., Denham, S. L., & Nelken, I. (2009). Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences, 13*(12), 532-540.

Zvyagintsev, M., Nikolaev, A. R., Sachs, O., & Mathiak, K. (2011). Early attention modulates perceptual interpretation of multisensory stimuli. *Neuroreport, 22*(12), 586-591.

APPENDICES

Appendix A

Method

Participants

Thirty volunteers (25 female), aged 18 to 24 years (19.27 mean age), recruited online from Cardiff University's School of Psychology, were given course credit for their participation. All participants were right-handed, native English speakers who reported normal (or corrected-to-normal) hearing and vision.

Apparatus & Materials

The digits one to nine served as the to-be-remembered stimuli and were constructed from pseudo-random orderings of eight items, with three constraints: (1) no digit was repeated more than once within a sequence, (2) sequences could not contain more than two digits in an ascending or descending order (e.g. "*3 4*" or "*7 8*") and (3) a digit could not appear in the same serial position on consecutive sequences. Using *E-Prime* software, digits were presented in black, 72-point Times New Roman font, in the centre of a white PC screen. Each item was presented with a one-second onset-to-onset interval. Irrelevant items (the letters a, b, c) were recorded in a sound attenuated laboratory by a male speaking in a monotone voice (at approximately 150Hz). Using SonicForge 5.0 software (Sonic Foundry, Inc

Madison, WI; 2000), the pitch of each item was lowered by three semitones and compressed digitally to 190ms, without further changing pitch

Design

Interference (no task, irrelevant speech, concurrent articulation, sequential tapping) and Serial Position were manipulated in a repeated measures 4x8 factorial design, with the percentage of items correctly recalled across all serial positions being taken as the dependent measure. Twelve sequences were presented for each of the four interference conditions, arranged in a pseudo-random order that was balanced across all participants with the constraint that no condition was presented more than twice in succession. A total of 48 experimental trials were administered to all participants, as well as 8 practice sequences preceding the test phase.

Procedure

Participants were tested individually in a sound-attenuated laboratory and wore headphones throughout the experiment where the sound level was adjusted to a comfortable level (approximately 65 dB(A)). A 500ms warning tone (500Hz sinewave) signalled the start of each trial, followed by a fixation cross, presented for 5s before the onset of the first to-be-remembered item. This introductory period was filled with either silence, in the case of no task, concurrent articulation, and sequential tapping conditions, or a period of irrelevant speech (20 tokens of irrelevant speech) that continued, without a break in tempo, during presentation of the memory sequence.

Concurrent articulation and sequential tapping was expected from warning tone onset until the offset of the last memory item. For concurrent articulation participants were asked to start whispering aloud the letters *a*, *b*, *c*, while for sequential tapping trials, participants were instructed to tap, with their right hand, the left, down and right arrow keys in order with their index finger, middle finger and ring finger. The experimenter coached each participant in the correct rate (approx. four per second) and loudness for articulation and tapping and remained in the laboratory to ensure compliance with instructions. Immediately after the offset of the last memory item, participants were visually cued to recall via a response screen displaying the digits one to nine in written form. Using the mouse, participants were required to move the cursor and click over eight digits corresponding to the exact order of the presented lipread sequence (i.e., strict serial order). Following eight responses, the next trial commenced automatically. The experiment lasted approximately 30-minutes, including an optional 5-minute rest period at the halfway point.

Results

The data were scored according to strict serial order criterion and subjected to a repeated measure ANOVA on the percentage of items recalled correctly across all serial positions for each interference condition (Figure A1). A difference was evident across all interference manipulations, *F*(3,27) = 43.62, *p*<.001, η^2 = .83, such that, relative to no task conditions, recall performance was disrupted in the presence of concurrent articulation, *F*(1,29) = 132.73, *p*<.001, η^2 = .82, as well as to smaller, but still significant extent during sequential tapping, *F*(1,29) = 25.40, *p*<.001, η² = .47, and irrelevant speech, *F*(1,29) = 8.67, *p*<.01, η² = .23.

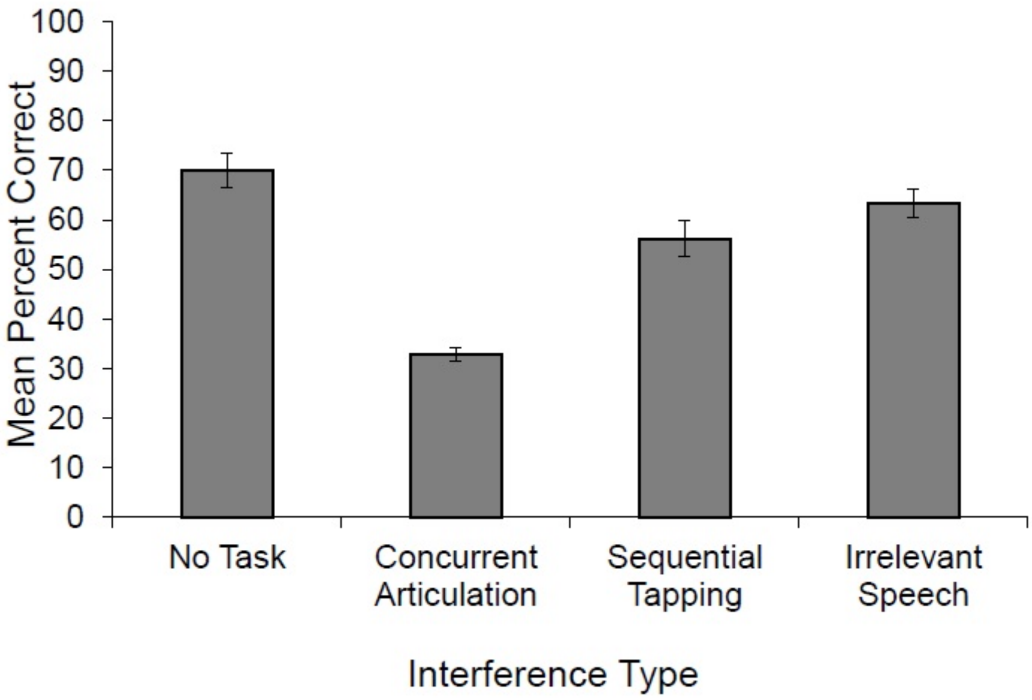

Figure A1. Mean percentage of items correctly recalled across all serial positions for each interference condition (no task, irrelevant speech, concurrent articulation, sequential tapping).