

# **The (Questionable) Role of Neighbourhood Density in Verbal Short-Term Memory**

A thesis submitted for the degree of Doctor of Philosophy

by

David James Greeno

School of Psychology, Cardiff University September 2019



## **Declaration and Statements**

### **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, and the thesis has not been edited by a third party beyond what is permitted by Cardiff University's Policy on the Use of Third Party Editors by Research Degree Students. 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 online in the University's Open Access repository 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 online in the University's Open Access repository and for inter-library loans after expiry of a bar on access previously approved by the Academic Standards & Quality Committee.

Signed ..... (candidate)      Date.....

## **Acknowledgements**

I would like to begin by thanking my supervisors Professor Bill Macken and Professor Dylan Jones. Your help and advice throughout the PhD has been invaluable and I am very grateful for the time you have given. I would also like to thank all the staff (academic and administrative) who have supported me throughout my time at Cardiff University.

Secondly, I would like to thank my parents who have been a source of support during my PhD and willing to help wherever they can. I would also like to thank my partner Kate who has been there for me throughout.

Finally, I would also like to thank all the friends and acquaintances I have met during my time in Cardiff and helped to make it such a good experience.

## Thesis Summary

Neighbourhood density refers to the number of words that can be derived from a given word by changing a single phoneme (phonological neighbours) or letter (orthographic neighbours). Performance in verbal short-term memory (vSTM) tasks (e.g., serial recall/reconstruction) is usually best when to-be-remembered words are from dense rather than sparse neighbourhoods. This is typically used as evidence of short-term storage being supported by networks of long-term lexico-phonological knowledge. The primary aim of the thesis is to assess our current understanding of neighbourhood density and to consider what it really reveals about the nature of vSTM. This was achieved by adopting several approaches. Chapter 2 first identified and then investigated three key variables – modality, task type and word pool size – to assess the parameters required for a dense neighbourhood advantage in vSTM tasks to manifest. However, across 4 experiments, it was revealed that neighbourhood density does not have robust and general effects upon vSTM with an advantage for sparse neighbourhood words elicited in Experiments 3 and 4. The findings raise questions over several models of vSTM. Chapter 3 more closely examined the parameters used by some previous experiments that have investigated neighbourhood density and found that the distribution of onset letters within each word pool is often not controlled. Simulations were used to demonstrate that this oversight can produce analogous effects in serial reconstruction tasks if participants were to use part of the word (e.g., the onset letter) to inform the order of the originally presented sequence. If, in some instances, the distribution of onset letters within word pools was determining vSTM task accuracy then the usefulness of using those results to make specific claims about how neighbourhood density impacts vSTM is brought into question. Chapter 4 considered that neighbourhood density variations might exist because of effort minimisation during the development and evolution of language. It was found that words from denser neighbourhoods tend to consist of more effortful articulations and take longer to vocalise than words from sparser neighbourhoods. This raises the possibility that neighbourhood density distributions impact vSTM because dense neighbourhood words are generally easier to articulate, rather than dense neighbourhood words being better supported by networks of long-term lexico-phonological knowledge. Finally, Chapter 5 attempted to demonstrate that effects analogous to those found when neighbourhood density is manipulated can also be found when only articulatory difficulty is manipulated. However, articulatory difficulty failed to predict the outcome of three experiments. The results of the thesis are considered in relation to several models of vSTM and the concept of neighbourhood density and what role, if any, it plays in vSTM is critically discussed.

## Contents

Declaration and Statements .....	i
Acknowledgements .....	ii
Thesis Summary .....	iii
Contents.....	iv
Chapter 1 Verbal short-term memory – Our current understanding .....	1
1.1 Remembering in the short-term.....	1
1.2 The Working Memory Model .....	1
1.2.1 The Phonological Similarity Effect - Evidence that storage within vSTM is phonological in nature.....	2
1.2.2 Evidence that vSTM is a consequence of decay offset by articulatory rehearsal: Rehearsal strategies, speech rates and the word length effect .....	3
1.3 Remembering in the short-term. Some role for prior (long-term) knowledge? ..	5
1.3.1 The Lexicality Effect.....	6
1.3.2 The Frequency Effect .....	6
1.4 Accepting the role for long-term memory in vSTM .....	6
1.4.1 How does LTM support vSTM? – Redintegration .....	7
1.4.2 A test of redintegration – Removing the vSTM requirement for LTM .....	7
1.5 vSTM – Interference rather than decay? Computational modelling .....	8
1.5.1 Memory for Order .....	10
1.6 A return to articulatory rehearsal – Can it really be claimed to have no causal role for vSTM? .....	10
1.7 ‘Item based’ accounts of vSTM .....	12
1.7.1 Alternative to an ‘item-based’ account of vSTM – An object-oriented approach.....	12
1.8 Neighbourhood Density – What is it and what does it reveal about vSTM? ....	16
1.8.1 The ND effect – Redintegration is reliant upon activation within linguistic networks in LTM .....	17
1.8.2 Neighbourhood density – Further support for the notion that articulatory rehearsal is not a causal predictor of vSTM.....	19
1.8.3 The ND effect – Has articulatory rehearsal really been ruled out as a causal factor? .....	20
1.8.4 The ND effect – A possible relationship with articulatory difficulty? .....	21
1.9 Overview of thesis and rationale for empirical work .....	22
1.9.1 Chapter 2 – Establishing the parameters under which the ND effect manifests .....	23

1.9.2 Chapter 3 – The ND effect in Serial Reconstruction tasks: A consequence of Neighbourhood Density, Orthographic Frequency or a First Letter Confound? .....	23
1.9.3 Chapter 4 - Exploring the Relationship Between Neighbourhood Density and Articulatory Difficulty: Effort-Based and Duration-Based Measures of Articulatory Difficulty .....	24
1.9.4 Chapter 5 – Can manipulating articulatory difficulty produce an analogous effect to the ND effect? .....	24
Chapter 2 Establishing the parameters under which the ND effect manifests .....	25
2.1 Abstract .....	25
2.2 Introduction to Experiments 1, 2, 3 and 4.....	26
2.3 Experiment 1 .....	32
2.3.1 Method.....	32
2.3.2 Results.....	34
2.3.3 Discussion .....	36
2.4 Experiment 2 .....	37
2.4.1 Method.....	38
2.4.2 Results.....	39
2.4.3 Discussion .....	39
2.5 Experiment 3.....	41
2.5.1 Method.....	41
2.5.2 Results.....	43
2.5.3 Discussion .....	44
2.6 Experiment 4.....	46
2.6.1 Method.....	46
2.6.2 Results.....	47
2.6.3 Discussion .....	47
2.7 General discussion .....	48
Chapter 3 Accuracy in Serial Reconstruction tasks: A consequence of Neighbourhood Density, Orthographic Frequency or a First Letter Confound?.....	52
3.1 Abstract.....	52
3.2 Introduction to the simulations .....	53
3.3 Method .....	60
3.3.1 Simulation Assumptions.....	60
3.3.2 Pure Sequence Simulations.....	61
3.3.3 Mixed Sequence Simulations.....	62

3.4 Results .....	63
3.4.1 Jalbert et al. (2011b) Simulation Results. ....	63
3.4.2 Guitard, Saint-Aubin, et al. (2018) Simulation Results .....	67
3.5 General Discussion .....	69
Chapter 4 Exploring the Relationship Between Neighbourhood Density and Articulatory Difficulty: Effort-Based and Duration-Based Measures of Articulatory Difficulty.....	74
4.1 Abstract.....	74
4.2 Introduction to Study 1 and Experiment 5 .....	75
4.3 Study 1 - Articulatory difficulty as measured by an effort-based scoring system .....	77
4.3.1 Study 1a Method.....	80
4.3.2 Study 1b Method.....	83
4.4 Experiment 5 - Articulatory difficulty as measured by spoken duration .....	86
4.4.1 Method.....	90
4.4.2 Results.....	92
4.4.3 Discussion. ....	95
4.5 General Discussion.....	97
Chapter 5 Can manipulating articulatory difficulty produce an analogous effect to the ND effect? .....	102
5.1 Abstract.....	102
5.2 Introduction to Experiments 6 and 7 .....	103
5.3 Experiment 6.....	110
5.3.1 Method.....	110
5.3.2 Results and Discussion .....	113
5.4 Experiment 7.....	115
5.4.1 Experiment 7a Method.....	117
5.4.2 Results and Discussion .....	118
5.4.3 Experiment 7b Method.....	121
5.4.4 Results and Discussion .....	121
5.5 General Discussion.....	123
Chapter 6 General overview of the thesis, theoretical implications and discussion .....	128
6.1 Aims of the thesis and general overview.....	128
6.2 Summary of empirical chapters and key findings.....	129

6.2.1 Chapter 2 – Establishing the parameters under which the ND effect manifests .....	129
6.2.2 Chapter 3 – The ND effect in Serial Reconstruction tasks: A consequence of Neighbourhood Density, Orthographic Frequency or a First Letter Confound? .....	130
6.2.3 Chapter 4 - Exploring the Relationship Between Neighbourhood Density and Articulatory Difficulty: Effort-Based and Duration-Based Measures of Articulatory Difficulty .....	132
6.2.4 Chapter 5 – Can manipulating articulatory difficulty produce an analogous effect to the ND effect? .....	134
6.3 Overall Implications - How reliable is the ND effect? .....	134
6.4 Theoretical Implications: Models of memory .....	136
6.4.1 Redintegrative accounts of vSTM .....	136
6.4.2 Order based account of the ND effect .....	139
6.4.3 Object-oriented account of the ND effect .....	140
6.4.4 Psycholinguistic models of vSTM .....	142
6.5 Further directions and considerations .....	143
6.5.1 Expanding the articulatory difficulty scoring system and using it to create word pools .....	143
6.5.2 Assessing whether different manipulations of co-articulatory fluency can override neighbourhood density.....	144
6.5.3 Separating the potentially independent contributions of phonological and orthographic neighbourhood density.....	145
6.5.4 Considerations beyond orthographic and phonological neighbourhoods	145
6.5.5 Completely forget the Neighbourhood: Modelling the degree of overlap between experimental materials and participants' prior knowledge .....	146
6.6. From theory development to beyond: Greater consideration must be given to the vSTM task.....	148
6.7 Summary and Conclusions .....	152
References.....	153
Appendix A Word pools used in Experiments 1, 2 and 5.....	169
Appendix B Word pools used in Experiments 3 and 4.....	170
Appendix C Vowel Difficulty Scoring .....	172
Appendix D Word pools used in Experiments 6 and 7a .....	175
Appendix E Word pools used in Experiment 7b .....	176



# Chapter 1

## Verbal short-term memory – Our current understanding

### ***1.1 Remembering in the short-term***

Verbal short-term memory (vSTM) refers to the ability by which information presented auditorily (e.g., speech) or visually (e.g., written text) can be stored and, after a short delay, retrieved. vSTM is an extensively studied area within Psychology because, without the ability to temporarily store and retrieve information, a variety of cognitive functions such as sentence comprehension, mathematics and reasoning would not be possible. As such, understanding vSTM should lead to greater understanding of cognition more generally.

Much of our present understanding of vSTM is based upon tasks that are assumed to provide a direct test of a vSTM mechanism (e.g., Baddeley, 2002). One of the most commonly used tasks is serial recall. Here, a sequence of around 5 to 8 items is presented to a participant (usually at a rate of 1-item per second) and then, after a short delay of just a few seconds, they are cued to reproduce (e.g., spoken, written or typed) that sequence of items in the originally presented order. Another commonly used vSTM task is a slight variation upon serial recall known as serial reconstruction. All the to-be-remembered items are re-presented on screen and the participant is required to indicate the order that the re-presented items were originally presented in (e.g., by clicking the items or by re-arranging the items). Despite the vSTM tasks having fairly simple requirements performance is very rarely perfect and high numbers of errors are common. These errors have revealed a wide array of effects that are thought to reveal crucial insights into the workings of a vSTM mechanism and have helped with the development of many models of vSTM. The following sections provide a brief overview of some of these models and their explanations for the related effects.

### ***1.2 The Working Memory Model***

One of the most extensively tested models of short-term memory (STM) is the working memory model (Baddeley & Hitch, 1974; Baddeley, 1986; Baddeley, 2000). The original instantiation of the working memory model (Baddeley & Hitch, 1974; Baddeley, 1986) was comprised of three components. The first of which was a

governing component, the central executive, and then two subsidiary systems, the phonological loop and the visuo-spatial sketchpad. The phonological loop stores verbal material and the visuo-spatial sketchpad stores visual and spatial information.

Of particular relevance to vSTM is the phonological loop which is further divided into two components. Firstly, there is the phonological store which is a limited capacity passive storage system which holds to-be-remembered items in a modality neutral (phonological) form. Within the working memory model capacity is limited because items within the phonological store are considered to be subject to rapid decay with information being lost after just 1-2 seconds (e.g., Schweickert & Boruff, 1986). Because information contained within the phonological store is phonological in nature, visually presented information must first undergo some recoding into phonological form before it can access the store whereas auditory information benefits from obligatory access. The second component, the articulatory rehearsal system, serves two functions. Firstly, it allows for visual information to be recoded into phonological form and gain access to the phonological store. Secondly, it allows material within the phonological store to be revived (via articulatory rehearsal) thus overcoming the detrimental effects of decay upon to-be-remembered information. While the articulatory loop seemingly shares features with general language production it is not considered to be part of the language system but considered a distinct mechanism contained within the phonological loop (e.g., Baddeley & Wilson, 1985).

### **1.2.1 The Phonological Similarity Effect - Evidence that storage within vSTM is phonological in nature**

The *phonological similarity effect* (e.g., Conrad & Hull, 1964) refers to the finding that serial recall performance is more accurate for to-be-remembered sequences containing letters with dissimilar sounds (e.g., *r, k, f*) rather than similar sounds (e.g., *b, p, v*). This is despite dissimilar sounding letters taking the same amount of time to articulate as similar sounding letters (e.g., Schweickert, Guentert, & Hersberger, 1990) and therefore prone to similar levels of decay within the phonological store. The effect is assumed to reflect the phonological nature of storage within the phonological store (e.g., Baddeley, 2002). The phonological similarity effect is thought to arise because acoustically similar items that have become partially decayed within the phonological store are much more difficult to

differentiate from each other than dissimilar sounding items. As such, there is an increased likelihood of confusing similar items within the phonological store and the wrong item being selected at recall.

Additional evidence for the phonological nature of storage within the phonological store comes from observations obtained when participants are instructed to engage in articulatory suppression. Articulatory suppression refers to the instruction for participants to repeatedly recite a word (or words) during the encoding and rehearsal period of a vSTM task. This instruction is assumed to occupy the articulatory loop and prevent the recoding of visually presented items into phonological form. It also prevents any revivification of auditory items that have gained obligatory access to the phonological loop. The phonological similarity effect is abolished under conditions of articulatory suppression but only when to-be-remembered items are presented visually (e.g., Baddeley, Lewis, & Vallar, 1984). The effect persists when presentation of the to-be-remembered items is auditory. Under conditions of articulatory suppression participants are unable to recode items into phonological form. Because of this, the visually presented items never access the phonological store and so the sounds comprising each item cannot be confused with each other. However, with auditory presentation, the items automatically enter the phonological store and the items are just as likely to be confused with each other (because of their similar sounds) as they would be in conditions without articulatory suppression (e.g., Baddeley, 2002).

### ***1.2.2 Evidence that vSTM is a consequence of decay offset by articulatory rehearsal: Rehearsal strategies, speech rates and the word length effect***

To-be-remembered items within the phonological store begin decaying within just 1-2 seconds (e.g., Schweickert, & Boruff, 1986). A basic assumption of the working memory model (Baddeley & Hitch, 1974; Baddeley, 1986; Baddeley, 2000) is that the quicker articulatory rehearsal for those items can take place then the less those items are subject to decay and the more likely they will be recalled correctly. Evidence supporting this assumption comes in several forms.

#### ***1.2.2.1 Observations, self-report and encouraging different rehearsal strategies***

Perhaps the most basic form of evidence for decay offset by rehearsal comes from the simple observation from researchers that participants will very often engage

in some sort of overt rehearsal strategy within a vSTM task setting (e.g., Lewandowsky & Oberauer, 2015). Self-reports have also revealed rehearsal to be a commonly used strategy in STM tasks (e.g., Morrison, Rosenbaum, Fair & Chein, 2016) with task accuracy better for those participants that report using rehearsal (e.g., Bailey, Dunlosky & Hertzog, 2009). If rehearsal has no benefit upon vSTM it could perhaps be considered unusual that so many participants choose to use it. However, just because participants are observed engaging in rehearsal and report adopting rehearsal strategies that does not necessarily mean it has a direct causal impact upon vSTM. Tan and Ward (2008) provided more compelling evidence for rehearsal having a causal link with vSTM by instructing participants to engage in overt rehearsal strategies or to remain silent during a serial recall task. Participants most commonly rehearsed in a cumulative fashion (e.g., item 1, item 1 and item 2, item 1 and item 2 and item 3 etc.). A strong correlation was found between the number of repetitions of the cumulative rehearsal cycle and the overall serial recall accuracy. The more the items were able to be rehearsed in a cumulative fashion then the better serial recall performance was.

#### ***1.2.2.2 The positive relationship between speech rate and memory span***

Memory span refers to the maximum number of items that an individual can store and correctly retrieve from vSTM. It is tested via a span task which is very similar to serial recall. However, instead of repeatedly presenting participants with to-be-remembered sequences of equal length (e.g., 6 items) the to-be-remembered sequences gradually increase in length until the participant is no longer able to recall all the items correctly. Some evidence of vSTM being a direct consequence of decay offset by articulatory rehearsal is the positive relationship between an individual's overt speech rate and their memory span. The faster someone can speak then the larger their memory span is (e.g., Hulme, Thomson, Muir & Lawrence, 1984). Additionally, as children grow older their overt speech rate increases and so does their memory span (Hulme et al., 1984). This suggests that overt speech rate and articulatory rehearsal are very closely related. The faster someone can speak then the faster that they are also able to engage in articulatory rehearsal and therefore the more successful they will be in offsetting decay and sustaining better vSTM.

### **1.2.2.3 The Word Length Effect**

The *word length effect* refers to the finding that serial recall and serial reconstruction is typically better for sequences of short words than sequences of long words (e.g., Baddeley, Thomson & Buchanan, 1975; Jalbert, Neath, Bireta, & Surprenant, 2011a). Within the working memory model the effect can be explained because long words, containing more syllables than short words, take longer to rehearse via the articulatory loop and their representations within the phonological store are prone to higher levels of decay than short words.

The word length effect has also been revealed for words with the same meaning across different languages (e.g., Naveh-Benjamin & Ayres 1986). The average number of syllables comprising the digits zero to nine is lower for English ( $M = 1.2$ ) than Arabic ( $M = 2.25$ ). Articulatory rehearsal of Arabic digits will therefore take longer than for English digits and if vSTM is a consequence of decay offset by rehearsal then, within the phonological loop, the Arabic digits should be more prone to decay than the English digits before they can be revived via articulatory rehearsal. Supporting this prediction, memory span is greater for English compared to Arabic digits. Further evidence for the important role of articulatory rehearsal in overcoming the detrimental effects of decay comes from findings that engaging in articulatory suppression abolishes the word length effect when items are presented both visually and auditorily (e.g., Baddeley et al., 1975). When rehearsal is prevented there can no longer be any variations in the speed in which to-be-remembered materials are rehearsed and therefore no memory differences between long and short words.

### **1.3 Remembering in the short-term. Some role for prior (long-term) knowledge?**

While the original working memory model (Baddeley & Hitch, 1974; Baddeley, 1986) makes very clear predictions based upon articulatory rehearsal time it does not make predictions based upon pre-existing (long-term) knowledge of language. Since the original instantiation of the working memory model, several effects have revealed this to be a critical oversight.

### **1.3.1 The Lexicality Effect**

Serial recall is better for sequences of words (e.g., 'bark', 'peg', 'tart') than non-words (e.g., 'bork', 'peb', 'tark'). This is known as the *lexicality effect* (e.g., Gathercole, Pickering, Hall & Peaker, 2001; Hulme, Roodenrys, Brown & Mercer, 1995; Thorn & Gathercole, 2001). The lexicality effect persists even when the words and non-words have identical overt articulation times (e.g., Thorn & Gathercole, 2001). As such, both the words and non-words should be rehearsed at similar rates and prone to the same levels of decay within the phonological store. The lexicality effect suggests that prior knowledge of language within LTM must have some impact upon vSTM and raises doubts over vSTM simply being a consequence of decay offset by articulatory rehearsal.

### **1.3.2 The Frequency Effect**

Word frequency refers to how commonly each word is used within a given language. The frequency of a word can be calculated by taking counts of how often each word appears within a large corpus of written text or spoken language (e.g., Baayen, Piepenbrock, & van Rijn, 1993). Serial recall is better for sequences containing the words more frequently used within a given language compared to sequences of words used less frequently (e.g., Hulme, Roodenrys, Schweickert, Brown, Martin, & Stuart, 1997). This is referred to as the *frequency effect*. Similarly to the lexicality effect, this effect persists even when the frequently used words have similar overt articulation times to the less frequently used words (e.g., Hulme et al., 1997) and again highlights some role for prior knowledge of language upon vSTM.

## **1.4 Accepting the role for long-term memory in vSTM**

Effects (e.g., lexicality, frequency) based upon prior knowledge within long-term memory (LTM) are thought to highlight the distinct but interactive nature of LTM and STM. As such, researchers now regularly incorporate some mechanism that allows for pre-existing knowledge within LTM to support STM. In light of such effects Baddeley (2000) updated the working memory model to include the episodic buffer. This fourth component is capable of integrating information from each of the original components (the central executive, the phonological loop and visuo-spatial sketchpad) with LTM.

#### **1.4.1 How does LTM support vSTM? – Redintegration**

One way in which LTM is proposed to support STM is via *redintegration* (e.g., Schweickert, 1993). Redintegration is a process whereby partially degraded traces of to-be-remembered items in STM can be repaired by pre-existing representations in LTM. This is usually couched as a process whereby a partially degraded trace is used as a cue in which to select a candidate word contained within LTM. Once a candidate is selected the phonological information from that candidate is used to replace the information that is missing from the degraded trace in STM (e.g., Hulme et al., 1997). The less that a to-be-remembered item has degraded and the more readily accessible and available those candidates in LTM are then the more successful the redintegrative process will be. For example, because words, but not non-words, have pre-existing representations in LTM, only the degraded words in STM can be compared with, and repaired by, information from intact representations in LTM. This underlies the lexicality effect (e.g., Gathercole et al., 2001). Similarly, high frequency to-be-remembered words are more accessible (because of more frequent everyday usage) in LTM than low frequency words and more readily facilitate a successful redintegration process (e.g., Hulme et al., 1997).

#### **1.4.2 A test of redintegration – Removing the vSTM requirement for LTM**

Further support for LTM supporting STM via a redintegration process comes from the observation that effects such as the lexicality effect are robust in serial recall but attenuate or are eliminated in serial recognition (e.g., Gathercole et al., 2001). Unlike serial recall, whereby the participant is required to reproduce the to-be-remembered sequence (e.g., spoken, written, typed), serial recognition presents a to-be-remembered sequence (the *standard* sequence) before re-presenting those items (the *test* sequence) in either the same or different order (via transposition of adjacent items). Because all the to-be-remembered items are re-presented at test, the relative accessibility and availability of those to-be-remembered items in LTM is no longer important for accurate task completion. Instead, the re-presented items can repair any degraded traces in STM obviating the need for LTM. As both the words and non-words are re-presented there are no longer any differences in the level of support provided to the degraded traces of words and non-words in STM and the lexicality effect is attenuated (Gathercole et al., 2001).

### **1.5 vSTM – Interference rather than decay? Computational modelling**

Participants report using rehearsal strategies in STM tasks (e.g., Bailey et al., 2009) and vSTM accuracy is highly correlated with the number of cumulative rehearsal cycles (e.g., Tan & Ward, 2008) but that does not necessarily mean that rehearsal has any causal benefit upon vSTM by offsetting the detrimental effects of decay. The debate over the need to include decay in models of vSTM has been long running and is still very much at the forefront of research (e.g., Lucidi et al., 2016; Souza & Oberauer, 2018). Because of the lack of consensus over the existence of decay many researchers have chosen to entirely remove it from computational models of vSTM. Despite the removal of decay these models still successfully simulate the variety of vSTM effects discussed so far.

An example of a computational model of vSTM is the feature model (e.g., Nairne, 1990). In this model, to-be-remembered items are comprised of modality dependent and modality independent features. Features such as the presentation voice or visual appearance of the item are modality dependant and features such as the phonemes comprising the item are modality independent. To-be-remembered items are simultaneously encoded into primary memory (containing only currently active representations) and secondary memory (containing all an individual's prior experience). All recall is from secondary memory with the information in primary memory being used as cues in which to sample information from secondary memory. However, within primary memory, information from each subsequent to-be-remembered item has the potential to overwrite some information from the preceding item (retroactive interference). Modality independent information can overwrite modality independent information and modality dependant information can overwrite modality dependant information. If item  $n+1$  of a to-be-remembered sequence shares similar features (e.g., phonemes – a modality independent feature) to item  $n$  then those similar features are removed from the representation of item  $n$ . Correct recall is not solely dependent upon how little or how much information has been overwritten but also dependent upon how successfully the degraded traces in primary memory can still be used as cues to correctly match with undegraded traces in secondary memory. At this point a redintegrative stage (as detailed in Section 1.4.1) helps to repair any degraded traces with the success of redintegration dependent upon the availability and accessibility of the intact representations within secondary memory.



Unlike the working memory model (Baddeley & Hitch, 1974; Baddeley, 1986; Baddeley, 2000) whereby information is lost because of decay, no information is ever truly lost within the feature model because it remains intact within secondary memory. However, direct access to that information, via the cues in primary memory, can become obscured by interference and this decreases the likelihood of the information being successfully retrieved from secondary memory. For example, the word length effect is proposed to occur because longer words are more phonologically complex (Caplan, Rochon & Waters, 1992; Service, 1998) comprising many more phonemes than short words. Subsequently encoded long words are more likely to share similar features (e.g., phonemes) with the preceding items and this increases the likelihood that the preceding items will have those particular features overwritten (Neath & Nairne, 1995). The phonological similarity effect occurs because similar sounding words share more similar features (e.g., Neath & Nairne 1995) which increases the likelihood of information from earlier items being overwritten. Some interference models have tested the effect of incorporating a rehearsal process (e.g., Lewandowsky & Oberauer, 2015) but because decay does not exist within the models the rehearsal mechanism has no measurable benefit upon performance and in some instances can be detrimental to modelled performance.

Just because decay is not needed within computer simulations in order to reproduce vSTM effects that is not evidence for the absence of decay within vSTM. However, a recent series of experiments demonstrated that despite rehearsal commonly being used by participants it seemingly has no causal benefit upon vSTM. Cumulative rehearsal is considered an optimal strategy for rehearsal (e.g., Tan & Ward, 2008) but serial recall was found to be no better when participants were explicitly instructed to cumulatively rehearse compared to a free rehearsal strategy (Souza & Oberauer, 2018). Perhaps more compelling though was that an instructed rehearsal strategy (e.g., rehearse the items re-presented in red text) was no better for serial recall performance than participants being asked to recite 'babibu' after presentation of each to-be-remembered word (Souza & Oberauer, 2018).

### **1.5.1 Memory for Order**

In a serial recall task, there is not only a requirement to correctly reproduce the to-be-remembered items, but also to reproduce them in their correct order. Perhaps the most consistent finding in the vSTM literature is that when serial recall data is plotted to show performance at each of the serial positions there is a distinct curve. Performance is best at serial position 1 (*the primacy effect*) and worsens at each successive serial position. At the final serial position there is usually a small improvement in performance relative to the mid-sequence items (*the recency effect*). A criticism of the working memory model (Baddeley, 2000) is that there is no explicit mechanism in place to explain these common performance curves. This has led some researchers to incorporate explicit mechanisms that treat the encoding of item and order information separately. Some researchers have expanded upon the basic decay principles of the phonological loop by suggesting that positional information is also encoded along with item information (e.g., Burgess & Hitch, 1999, 2006) or proposing that the strength of activation within vSTM for each item decreases at each subsequent serial position (e.g., Page & Norris, 1998). Others eschew the concept of decay and incorporate interference and positional information (e.g., Lewandowsky & Oberauer, 2015).

### **1.6 A return to articulatory rehearsal – Can it really be claimed to have no causal role for vSTM?**

The ability to model vSTM without any need to incorporate decay (e.g., Nairne, 1990; Lewandowsky & Oberauer, 2015), simulations demonstrating that rehearsal can sometimes hinder vSTM performance (e.g., Lewandowsky & Oberauer, 2015) and the demonstration that guided rehearsal strategies are no more beneficial than reciting 'babibu' (e.g., Souza & Oberauer, 2018) seemingly undermine the theoretical relevance of decay offset by articulatory rehearsal. Even though participants report using rehearsal as a strategy in STM tasks (e.g., Bailey et al., 2009) one possibility is that rehearsal is simply an epiphenomenon of vSTM. Rehearsal could be the consequence of strongly encoded memory traces within vSTM rather than a causal mechanism maintaining those traces. However, an alternative is to question the assumption that underpins the models of vSTM discussed so far (irrespective of whether they incorporate decay or interference) that for articulatory rehearsal to have some causal role in vSTM that its impact should

always be beneficial. To demonstrate why this assumption may be incorrect consider the example of a tongue twister. A tongue twister contains words that share many similar phonemes or phoneme repeats (e.g., *'she sells sea shells'*) or have onset consonants that follow an ABBA format (e.g., *'leap note nap lute'*). Repeated reiteration of such tongue twisters is known to induce speech errors (e.g., Wilshire, 1999). If two hypothetical participants took part in a serial recall task and both were very familiar with the items that comprise tongue twisters (e.g., *'she sells sea shells'*) but one regularly practices articulating them whereas the other does not, a cumulative rehearsal strategy (an optimal strategy for vSTM; e.g., Tan & Ward, 2008) is likely to have different implications for them. The experienced participant would possibly make very few errors while cumulatively rehearsing the items because they are well practiced with articulating tongue twisters. When cued to reproduce the items they are likely to all be correct achieving 100% accuracy. Cumulative rehearsal in this instance will have led to perfect performance. The less experienced participant may produce no errors in the first few cumulative rehearsal cycles (e.g., *'she'*, *'she sells'*, *'she sells sea'*, *'she sells sea shells'*). However, after multiple reiterations of the complete sequence there are likely to be errors (e.g., *'she shells she shells'*). When cued to reproduce the sequence these same errors could still be present, and the resultant serial recall accuracy would be 50%. Despite the different outcomes, both instances would be an example of rehearsal having a causal role upon vSTM task performance. However, the first is an example of rehearsal benefitting serial recall whereas the second is an example of rehearsal having a detrimental impact upon serial recall.

Such speech errors (e.g., *'she shells she shells'*) are an example of a spoonerism whereby corresponding sounds between words are switched. Importantly, when considering the role of rehearsal in vSTM, it is not only tongue twisters (and hypothetical vSTM tasks) that are prone to naturally occurring speech errors such as spoonerisms. For example, Ellis (1980) noted that many errors in serial recall tasks are very similar to naturally occurring speech errors. Shattuck-Hufnagel (1992) found that speech errors for tongue twisters were very similar irrespective of whether participants simply had to read them on screen or attempt to retrieve them from memory. Despite naturally occurring errors and errors within vSTM tasks being very similar, the role that naturally occurring speech errors may be having upon vSTM tasks is not considered by the models discussed so far. At most,

the speech errors would be considered a by-product of an error prone vSTM system rather than an actual determinant of vSTM performance.

### **1.7 'Item based' accounts of vSTM**

While far from an exhaustive list, and despite their differences, the models of vSTM introduced so far highlight some key shared concepts that are highly prevalent in much of the vSTM literature. Firstly, recall errors in a vSTM task are dependent upon information pertaining to each individual item being subject to some form of degradation (whether that be via decay or interference). The more decay or interference there is then the less likely that a to-be-remembered item will be successfully recalled. Secondly, successfully overcoming that degradation is reliant upon pre-existing knowledge within LTM regarding each of those items. The more readily accessible and available that pre-existing knowledge within LTM is then the more likely that a degraded item will be successfully reintegrated. As such, each of the accounts described so far could be considered 'item-based' accounts of vSTM. However, because each item, and the processes proposed to operate upon them, is focussed upon in relative isolation the importance of sequence-level effects (e.g., those that may operate upon an entire sequence of items such as naturally occurring speech errors) have been largely overlooked.

#### **1.7.1 Alternative to an 'item-based' account of vSTM – An object-oriented approach**

vSTM tasks are assumed to provide a direct test of a vSTM mechanism (e.g., Baddeley, 2002). Any emerging effects are assumed to be a consequence of processes associated with that vSTM mechanism operating upon each of the to-be-remembered items (e.g., interference, decay, reintegration). However, some accounts of vSTM (e.g., Jones & Macken, 2018; Macken, Taylor & Jones, 2015) make no assumptions that vSTM is a mechanism of itself or about the existence of such processes. Instead, task performance within a vSTM setting is treated as an outcome reliant upon the utilisation of more general processes such as language and perception. Within this framework, any effects that emerge within a vSTM task setting are a consequence of how well participants are able to opportunistically utilise these general skills to complete the task (e.g., Jones & Macken, 2018; Macken et al., 2015).

By utilising the speech motor system humans are highly efficient at reproducing and communicating substantial amounts of structured verbal information (e.g., following familiar grammatical structures). Within a vSTM task setting, rather than some specialised vSTM system, this same speech motor system is also capable of planning the necessary articulatory gestures required to reproduce a sequence of to-be-remembered items. However, the sets of to-be-remembered items are often removed of any familiar grammatical structures. Instead, participants are usually presented with, and then asked to reproduce in an identical order, sequences of seemingly random and unrelated items (very often presented at a rate of 1 item per second). Additionally, before the participant is even permitted to reproduce the items there is usually an experimentally imposed delay. Outside of an experimental setting all these requirements are fairly uncommon and a disjunction arises because reproduction of the experimental materials does not readily map onto the participants' usual speech motor skills. vSTM effects typically ascribed to processes operating upon each to-be-remembered item (e.g., decay, interference, reintegration) instead arise within the speech motor system itself during attempts to overcome this disjunction. More specifically, vSTM effects are thought to arise during the assembly and rehearsal of the articulatory gestures (motor plan) that are required to reproduce the entire to-be-remembered sequence. This assembly of articulatory gestures is known as *segmental recoding* (e.g., Macken, Taylor & Jones, 2014). Analogous to the issues discussed regarding cyclical rehearsal (Section 1.6) segmental recoding is not always a flawless process and naturally occurring errors (e.g., spoonerisms) can arise within the motor plan. Additionally, any attempts to rehearse the articulatory gestures can result in naturally occurring errors arising within the motor plan. The accuracy within a vSTM task setting is therefore determined by the relative fluency in which the articulatory gestures required for output can be prepared and rehearsed. Fluency in this case refers to the smoothness in articulatory transitions between items (co-articulation) and the smoothness of the articulations that comprise the items themselves. The more fluent this process is then the less likely that there will be errors within the motor plan and the more likely that the to-be-remembered sequence will be accurately reproduced (e.g., Macken et al., 2014; Murray & Jones, 2002; Woodward, Macken & Jones, 2008).

Within an object-oriented framework (e.g., Jones & Macken, 2018; Macken et al., 2015) the lexicality effect is proposed to occur because of the relative experience the speech motor system has in reproducing sequences of words compared to non-words. Humans could be considered most experienced in reproducing sequences of words that form sentences. We are less experienced in reproducing sequences of random words, but it is still a skill sometimes required (e.g., repeating a shopping list) and we will at least be experienced with the articulatory configurations comprising the words. However, we could be considered inexperienced in reproducing random sequences of non-words and inexperienced with the articulatory configurations comprising those non-words. Because sequences of non-words will contain many novel articulatory configurations the preparation and rehearsal of the novel articulations will be a far less fluent process than the preparation of the more familiar articulations associated with words. The less fluent the preparation and rehearsal is, then the more likely that there will be errors within the articulatory motor plan and the more likely that there will be errors at output (e.g., Macken et al., 2014). Similarly, more frequently used words have more familiar articulations because they are used more often in everyday speech and afford more fluent preparation of the articulatory gestures required to reproduce them (e.g., Woodward et al., 2008). Performance within a vSTM setting can therefore be viewed as a continuum whereby the degree of overlap between the participants' prior (long-term) linguistic skills and the particular to-be-remembered material determines how accurately the material is reproduced (e.g., Jones & Macken, 2015).

An additional component of the object-oriented account of vSTM (Jones & Macken, 2018; Macken et al., 2015) is that presentation modality is important, not because it varies the route by which material can access the phonological store (e.g., Baddeley, 2002), but because auditory presentation within a vSTM task can reduce the participants' reliance upon the speech motor system. For example, as discussed in Section 1.4.2 the lexicality effect emerges in serial recall but attenuates, or is eliminated, in serial recognition (e.g., Gathercole et al., 2001). The absence of an effect in serial recognition is used to highlight the distinct but interactive nature of LTM and STM (when items are re-presented, LTM is no longer required to repair items in STM). However, Macken et al. (2014) noted that previous examples of the lexicality effect had used both visually and auditorily presented serial recall but almost exclusively used auditorily presented serial recognition. Auditory information

sharing a number of similar features such as timbre, tempo, timing, spatial location etc. tends to fuse into a single perceptual stream or *auditory object* (see, e.g., Bregman, 1990). As such, auditory presentation of to-be-remembered items affords the opportunity for auditory object formation.

Auditory object formation is of particular relevance in serial recognition tasks because an order judgement between two auditory objects can be made without any knowledge of the specific items that comprise that object (e.g., Warren, Obusek, Farmer & Warren, 1969). Participants are able to discriminate whether two sequences comprised of sounds (e.g., a buzz, a hiss, a click, and a vowel sound) are presented in the same or different order. However, they are unable to name the items that comprise those sequences in the correct order until the presentation rates are sufficiently slow (at least 700ms per item). Order judgements can be made even at very fast presentation rates (10ms per item; e.g., Warren, 1974a, 1974b). In that sense, auditory serial recognition for words vs non-words can essentially become a *pattern matching* task whereby successful auditory object comparisons can be made irrespective of whether the auditory object contains words or non-words. This pattern matching judgement can be completed by solely using the auditory perceptual system with no requirement to utilise the speech motor system. For this reason, unless auditory items are presented in such a way that they reduce the likelihood of auditory object formation occurring (e.g., by slowing presentation rates), an attenuated or entirely absent lexicality effect in auditory serial recognition would be expected.

Visual serial recognition cannot be completed by using the auditory perceptual system and serial presentation of items weakens the visual analogous of object formation. This is because, unlike auditory object formation which is reliant upon the temporal parameters of the auditory input, visual object formation is more reliant upon simultaneous presentation of all features comprising the object (see, e.g., Shinn-Cunningham, 2008 for examples of visual vs auditory objects). Therefore, correct judgements in visually presented serial recognition cannot rely upon pattern matching. Instead, a correct comparison is likely to require utilisation of the speech motor system to first recode the items into an articulatory motor plan before later comparing that motor plan with the test sequence (Macken et al., 2014). Likewise, both auditory and visual serial recall require the speech motor system because, irrespective of modality, the to-be-remembered items will need to be reproduced at

test. As such, they must first be recoded into an articulatory motor plan so that they can later be output. The recoding of words is a more fluent process and less prone to errors than the recoding of non-words. In line with this account of vSTM performance Macken et al. were able to successfully elicit a robust lexicality effect in visual and auditory serial recall as well as in visual serial recognition. As in previous demonstrations of the lexicality effect (e.g., Gathercole et al., 2001) it was again absent in auditory serial recognition because the task could be completed via a simple pattern matching process. The findings highlight the importance that a variety of task factors play in determining vSTM accuracy. Item-based accounts of vSTM typically treat modality as providing differential routes for material to access STM (e.g., Baddeley, 2000) with motor processes mediating the output of that material. However, the object-oriented account of memory treats modality as input that will map onto different skills such as those associated with language and perception. Those general skills are the same skills that participants can then use to help sustain successful vSTM performance.

### ***1.8 Neighbourhood Density – What is it and what does it reveal about vSTM?***

The effects discussed so far in the thesis (word length, frequency, lexicality etc.) have been used to inform and develop a wide range of vSTM theories. However, in recent years there has been an increasing interest in another item-level variable known as neighbourhood density. Neighbourhood density refers to the number of similar sounding words (neighbours) that differ from a to-be-remembered word by the substitution, deletion or addition of a single phoneme (phonological neighbours; Luce & Pisoni, 1998) or by substitution of a single letter (orthographic neighbours; Coltheart, Davelaar, Jonasson, & Besner, 1977). Mulatti, Besner, and Job (2003) demonstrated that phonological and orthographic neighbourhoods are positively correlated and, unless specifically controlled for, experiments varying on orthographic neighbourhood density also vary on phonological neighbourhood density and vice versa. As such, in the present thesis, unless there is an important reason to differentiate the two, the term *neighbourhood density* will be used to describe experiments where items vary in either orthographic neighbourhood density or phonological neighbourhood density.



Compared to sequences of words taken from a sparse neighbourhood (e.g., *chef, germ, yarn*), sequences of words from a dense neighbourhood (e.g., *bark, fade, tart*) are more likely to be correctly recalled in span tasks (Roodenrys, Hulme, Lethbridge, Hinton & Nimmo, 2002) and serial recall tasks (Allen & Hulme, 2006) or correctly ordered in serial reconstruction tasks (Clarkson, Roodenrys, Miller & Hulme, 2017; Derragh, Neath, Beaudry, & Saint-Aubin, 2017; Guitard, Gabel, Saint-Aubin, Surprenant & Neath, 2018; Jalbert et al., 2011a; Jalbert, Neath & Surprenant, 2011b). Despite the neighbours of the to-be-remembered words not being presented in the vSTM tasks, and therefore only present within linguistic-networks in LTM, they still seemingly exert some influence upon task performance providing further evidence of the distinct but interactive nature of STM and LTM. This vSTM task advantage for words from a dense neighbourhood will be referred to as the *Neighbourhood Density effect (ND effect)*. The ND effect is of interest to the current thesis because of what it is considered to reveal about the underlying mechanisms of vSTM. For reasons set out in the sections below, it raises doubts over redintegration being reliant upon the relative availability and accessibility of intact representations in LTM (e.g., Hulme et al., 1997; Schweickert, 1993) and over the importance of articulatory rehearsal (e.g., Baddeley, 2000) or articulatory motor planning (e.g., Jones & Macken, 2018; Macken et al., 2014, 2015) having any causal impact upon vSTM.

### ***1.8.1 The ND effect – Redintegration is reliant upon activation within linguistic networks in LTM***

As discussed in Section 1.4 redintegration is usually cast as a process whereby degraded cues in STM are compared with intact representations in LTM (e.g., Hulme et al., 1997). However, the more neighbours that a to-be-remembered word has then the more similar sounding candidates there are in LTM that could be used to repair a degraded trace of that to-be-remembered word in STM. With dense neighbourhood words therefore providing more incorrect candidates for selection this should increase the likelihood of selecting an incorrect, rather than a correct, candidate and therefore hinder the redintegrative process for the dense rather than the sparse neighbourhood words. This makes the ND effect incompatible with the redintegrative account discussed earlier and as such it has been necessary to recast

redintegration as a process reliant upon the overall levels of activation within linguistic networks in LTM.

Although precise accounts of the ND effect are still very much in development a typical explanation is that all neighbours of a target word (e.g., 'Cat') are associatively linked within LTM. A partially degraded target word is still able to activate the target word as well as some its neighbours (e.g., 'Ca', is still an orthographic and phonological neighbour of 'Cab', 'Can', 'Cap' etc.). The more neighbours the original target word has then the more neighbours that can still be activated by a partially degraded version of that target word and the more overall activation there will be in LTM. This activation among neighbours is thought to *boost* activation within LTM of the originally presented word. This boost occurs because essentially the to-be-remembered word (e.g., 'Cat') is the word you would get when you average over its neighbours. Activation of many neighbours therefore increases the level of activation of the original target word to a higher level than other words in LTM and makes it more likely to be selected as a candidate in which to repair the degraded information in STM (e.g., Derragh et al., 2017; Jalbert et al., 2011a, b). Put simply, the more supportive activation that is elicited within a word's linguistic network in LTM then the more likely it is that the redintegration process will be successful for that word.

An account of the ND effect based upon the overall level of supportive activation elicited in LTM is not too dissimilar from Stuart and Hulme's (2000) re-interpretation of the frequency effect. In much the same way that neighbours are considered to be connected in LTM, Stuart and Hulme suggested that LTM will contain more pre-existing associations between the representations of high frequency words than between those of low frequency words. Therefore, high frequency words, because they share more connections, will benefit from higher levels of supporting activation during the redintegrative process. Training participants' LTM by presenting pairs of low frequency words increases the strength of low frequency word associations in LTM. Later recall is enhanced for those familiarised low frequency words compared to non-familiarised low frequency words when they appear in familiarised pairings but not in novel pairings (Stuart & Hulme, 2000). The lexicality effect could also be accommodated in a similar way. Because representations and pre-existing connections already exist within LTM for words, but not non-words, then words will elicit higher levels of supporting activation for the

redintegration process. The ND effect has added to the weight of evidence suggesting that redintegration is a process reliant upon the supportive levels of activation in LTM rather than operating as a process whereby the fewer competing candidates there are in LTM then the more likely there is to be a successful match between degraded cues in STM and intact representations in LTM.

### ***1.8.2 Neighbourhood density – Further support for the notion that articulatory rehearsal is not a causal predictor of vSTM***

Jalbert et al. (2011b) noted that almost all previous demonstrations of the word length effect were confounded by neighbourhood density. Long words tend to have fewer phonological and orthographic neighbours than short words. This raised the question of whether length or neighbourhood density was responsible for previous examples of the word length effect. They tested this by presenting participants with to-be-remembered sequences of long non-words from either a dense or sparse neighbourhood and sequences of short non-words from either a dense or sparse neighbourhood. A serial reconstruction task revealed the usual word length effect when reconstruction accuracy for the short non-words from a sparse neighbourhood was compared with the long non-words from a dense neighbourhood. However, reconstruction accuracy was better for the long non-words from a dense neighbourhood compared to short non-words from a sparse neighbourhood. This finding has important implications for any model of memory whereby performance is proposed to be some consequence of decay offset by articulatory rehearsal (e.g., Baddeley, 2000; Burgess & Hitch, 1999, 2006; Page & Norris, 1998). Irrespective of neighbourhood density, long words will take longer to rehearse than short words and therefore longer, rather than shorter, words should be more prone to the detrimental effects of decay. Long words are also usually more complex in nature (e.g., Caplan, Rochon & Waters, 1992; Service, 1998) so preparation of a motor plan would likely be less fluent for long words. This casts doubt over the ability for an object-oriented account (e.g., Jones & Macken, 2018; Macken et al., 2015) to adequately explain the ND effect.

### **1.8.3 The ND effect – Has articulatory rehearsal really been ruled out as a causal factor?**

The work by Jalbert et al. (2011b) raises doubts over models of vSTM that incorporate articulatory rehearsal (e.g., Baddeley, 2000; Burgess & Hitch, 1999, 2006; Page & Norris, 1998) or accounts that highlight the importance articulatory motor planning and rehearsal (e.g., Jones & Macken, 2018; Macken et al., 2015). However, recent work by Guitard, Saint-Aubin, Tehan and Tolan (2018) showed that, by creating pools of short words/non-words with very few orthographic neighbours, Jalbert et al. (2011b) may have introduced a new confound into their materials. According to Guitard, Saint-Aubin, et al. (2018) of the four letter, one-syllable nouns contained in the English language, only 13% of them have fewer than three orthographic neighbours. However, for seven-letter three syllable nouns the percentage reaches 98%. This introduced the risk that the short non-words from a sparse neighbourhood used by Jalbert and colleagues (having zero orthographic neighbours) were non-representative of typical short words and therefore likely to contain unusual orthographic structures. Guitard, Saint-Aubin, et al. controlled for this confound by using n-gram measures. These break a word down into letter combinations and compare the frequency with which those letter combinations appear in other words. For example, the word 'thesis' contains six unigrams ('t', 'h', 'e', 's', 'i', 's'), five bigrams ('th', 'he', 'es', 'si', 'is') and four trigrams ('the', 'hes', 'esi', 'sis'). The frequency at which each of the n-grams appear in other words is calculated and the higher a word's n-gram measure is, then the more common a word's structure is and the more familiar it is assumed to be. Guitard, Saint-Aubin, et al. were able to demonstrate a serial reconstruction advantage for long words over short words. However, this long word advantage only occurred when n-gram measures were higher for the long words than the short words. Once n-gram measures were equated then the usual vSTM advantage for short words re-appeared.

Guitard, Saint-Aubin, et al. (2018) suggested that their findings could be accommodated by a trace decay system supported by redintegration in essentially the same way that the frequency effect is typically accounted for (e.g., Hulme et al., 1997). Higher n-gram frequencies mean that those structures are more frequently used within everyday language usage. The higher frequency of usage means that those orthographic structures are more readily accessible within LTM and are

therefore more likely to promote a successful redintegrative process. Alternatively, unusual orthographic structures are also likely to promote less fluent preparation of an articulatory motor plan (e.g., Jones & Macken, 2018). While the exact implications of these findings in relation to models of vSTM is open to some interpretation it does demonstrate that, despite the fairly strong claims by Jalbert et al. (2011b), articulatory rehearsals causal role in vSTM performance has not been completely ruled out. See Chapter 3 for some further discussion of the research by Guitard, Saint-Aubin, et al. and Jalbert et al.

#### ***1.8.4 The ND effect – A possible relationship with articulatory difficulty?***

The work by Guitard, Saint-Aubin, et al. (2018) shows that the manipulation of neighbourhood density can inadvertently introduce new linguistic confounds that impact vSTM. However, their study specifically focussed on the effect that neighbourhood density has upon the word length effect with their analysis focused on fairly small sub-sets of long and short words. There remains the possibility that some other correlated factor with neighbourhood density is responsible for eliciting the ND effect when words of equal length are used. For example, 'cat' only has many neighbours because the words 'chat', 'bat', 'hat', 'car', 'rat', 'sat' etc. have also been introduced to humans' linguistic inventory. Why 'cat' should have so many neighbours whereas other words (e.g., 'yarn') do not has so far been given very little consideration in the vSTM literature.

According to Lindblom (1990) language is shaped by a pressure to minimise effort while preserving discriminability in perception. These pressures shape languages so that easier to articulate sounds are likely to be the most common sounds within a language. In respect to neighbourhood density this could have resulted in denser clusters of words around easier articulatory configurations and sparser clusters of words around more difficult articulations. The denser the clustering then the more likely that the words within a particular cluster will share a sufficient number of sounds and letters to render them phonological or orthographic neighbours.

Some converging lines of evidence lend support to the suggestion that an effort to minimise articulatory effort may have shaped neighbourhood density distributions. Firstly, words from dense neighbourhoods have higher phonotactic

probabilities (a measure of how often particular phonological segments appear in a given language) and are less prone to speech errors (Vitevich, 2002) than words from sparse neighbourhoods. Secondly, words from denser neighbourhoods are more prone to lenition when spoken in everyday speech (Gahl, Yao, & Johnson, 2012). Lenition is a process whereby articulatory effort is further reduced by not fully articulating consonant and vowel sounds (see, e.g., Honeybone, 2008).

A more direct investigation into the relationship between neighbourhood density and articulatory difficulty was conducted by St. John (2015). St. John created an articulatory difficulty scoring system whereby the articulations required to vocalise 8 consonants were scored on their relative difficulty (e.g., degree of muscular tension or the amount of rotational movement required by the jaw to vocalise the consonants). The scores were then applied to dense and sparse neighbourhood words containing those consonants. A small, but significant, correlation was found between neighbourhood density and articulatory difficulty with sparse neighbourhood words having slightly higher difficulty scores than dense neighbourhood words. Because the analysis only included a small subset of words it is not possible to draw more general conclusions (Chapter 4 attempts to address this problem) but the evidence was suggestive that sparse neighbourhood words may be more difficult to articulate than dense neighbourhood words. If this is the case then, despite claims to the contrary (e.g., Jalbert et al., 2011b), there remains the possibility that the ND effect is a consequence of dense neighbourhood words affording more fluent motor planning than sparse neighbourhood words (e.g., Jones & Macken, 2018).

### ***1.9 Overview of thesis and rationale for empirical work***

Neighbourhood density is the focus of this thesis because of the implications it has for modelling and understanding vSTM. Numerous accounts of vSTM include articulatory rehearsal as a mechanism that offsets decay (e.g., Baddeley, 2000; Burgess & Hitch, 1999, 2006; Page & Norris, 1998). Alternatively, speech motor systems can be used to prepare and rehearse a speech motor plan (e.g., Jones & Macken, 2018). However, the causal role of articulatory rehearsal is disputed and neighbourhood density is another variable currently being used to undermine the value of incorporating rehearsal when explaining vSTM performance (e.g., Jalbert et al., 2011b; Lewandowsky and Oberauer, 2015). Secondly, many models of vSTM

incorporate some link with LTM whereby the availability and accessibility of items within LTM determines the success of a redintegrative process (e.g., Baddeley, 2000; Burgess & Hitch, 1999, 2006; Nairne, 1990; Page & Norris, 1998). The ND effect has been used as evidence to suggest that this redintegrative process is reliant upon the levels of supportive activation elicited within LTM. Despite these claims there are still currently only a handful of studies investigating the effect of neighbourhood density upon vSTM. As such, there are a number of important manipulations, as well as alternative accounts, that have yet to be fully explored.

The aim of the thesis is to expand the currently available research into the ND effect by adopting a variety of different methodologies. Each of the following empirical chapters adopts a slightly different research strategy to scrutinise the neighbourhood density manipulation and the neighbourhood density literature.

### ***1.9.1 Chapter 2 – Establishing the parameters under which the ND effect manifests***

The first empirical chapter tests the conditions under which the ND effect manifests. It explores whether a redintegrative process, reliant upon activation within LTM, can adequately predict the impact of neighbourhood density upon vSTM for 6-item sequences as a function of task type (serial recall vs serial recognition), modality of item presentation (auditory vs visual) and the size of the pool from which the sequences are drawn (48 vs 12). The chapter will introduce and discuss the importance of each of these manipulations before their effect upon vSTM is tested.

### ***1.9.2 Chapter 3 – The ND effect in Serial Reconstruction tasks: A consequence of Neighbourhood Density, Orthographic Frequency or a First Letter Confound?***

Jalbert et al. (2011b) demonstrated that serial reconstruction was better for long non-words from a dense neighbourhood compared to short non-words from a sparse neighbourhood. This finding seemingly suggests that it is unnecessary to incorporate articulatory rehearsal as a causal mechanism for vSTM. As discussed in Section 1.8.3 though, Guitard, Saint-Aubin, et al. (2018) identified orthographic frequency as a possible confound in the materials used by Jalbert et al. which raised doubts over the strength of their conclusions. However, with each experiment controlling for so many possible item-level variables it is possible that others were

overlooked. In a series of simulations, it will be explored whether one such overlooked variable, the distribution of onset letters within each word pool, could yield a similar pattern of data. The rationale for using the simulations and what they might reveal about some of the neighbourhood density literature is discussed.

### ***1.9.3 Chapter 4 - Exploring the Relationship Between Neighbourhood Density and Articulatory Difficulty: Effort-Based and Duration-Based Measures of Articulatory Difficulty***

This chapter investigates whether neighbourhood density is correlated with articulatory difficulty. This is the first step towards establishing whether an object-oriented account (e.g., Jones & Macken, 2018; Macken et al., 2015) can accommodate the ND effect. The chapter expands upon the consonant difficulty scoring system devised by St. John (2015). It will introduce the rationale for the expanded scoring system before testing whether a relationship exists between articulatory difficulty and neighbourhood density. The chapter then investigates the relationship between vocalisation times and neighbourhood density. Differences in vocalisation times between dense and sparse neighbourhood words could be considered reflective of the relative fluency in which those items can segmentally recoded (e.g., Woodward et al., 2008). Woodward et al. (2008) investigated the vocalisation times associated with the lexicality and frequency effect and a similar methodology will be introduced and discussed before establishing what it reveals about neighbourhood density.

### ***1.9.4 Chapter 5 – Can manipulating articulatory difficulty produce an analogous effect to the ND effect?***

Even if a relationship between articulatory difficulty and neighbourhood density exists it is not possible to determine whether neighbourhood density, or articulatory difficulty, is responsible for the ND effect. One way to overcome this is by controlling stimuli for neighbourhood density while varying them on articulatory difficulty. A similar articulatory difficulty manipulation to that used by St. John (2015) was utilised and incorporated within the experimental paradigm discussed in Section 1.7.1 devised by Macken et al. (2014). The methodologies will be introduced and discussed before a series of three experiments investigating the impact of articulatory difficulty upon vSTM task performance are presented.



## Chapter 2

### Establishing the parameters under which the ND effect manifests

#### **2.1 Abstract**

The impact of long-term language knowledge on vSTM performance plays an important role in theories of vSTM. One such aspect, neighbourhood density (i.e., the number of words that can be derived from a given word by changing a single phoneme or single letter), leads to better vSTM task performance when to-be-remembered items are from a dense rather than a sparse neighbourhood. As outlined in Chapter 1, it has been argued that this dense neighbourhood advantage is due to supportive activation being elicited within linguistic networks in LTM. The denser the neighbourhood, then the higher the level of supporting activation and the better vSTM will be. Across 4 experiments the impact of neighbourhood density on vSTM for 6-item sequences as a function of task type (serial recall vs serial recognition), modality of item presentation (auditory vs visual) and the size of the word pool from which the sequences were constructed (48 vs 12) was examined. The effect of neighbourhood density proved highly sensitive to these factors such that the typical vSTM advantage for dense neighbourhood words in serial recall was eliminated when using serial recognition and reversed when using a smaller word pool in both serial recall and serial recognition. The findings raise questions about the viability of typical accounts of vSTM that invoke mutual support from LTM and indicate the critical role played by a variety of task factors in modulating such effects.

## **2.2 Introduction to Experiments 1, 2, 3 and 4**

Chapter 1 highlighted how vSTM is typically better for to-be-remembered words when they are from a dense rather than a sparse neighbourhood (e.g., Allen & Hulme, 2006; Clarkson et al., 2017; Derragh et al., 2017; Guitard, Gabel, et al., 2018; Jalbert et al., 2011a, b; Roodenrys et al., 2002). This ND effect is currently being used to support accounts of vSTM that point towards there being some supporting role for LTM upon vSTM. It is also used to suggest that articulatory rehearsal has no causal impact upon vSTM (e.g., Jalbert et al., 2011b). However, explanations for why exactly dense neighbourhood words sustain better vSTM are still very much in development.

Accounts that posit LTM as supporting vSTM at the retrieval stage often incorporate redintegration as an explanatory mechanism for the ND effect (e.g., Derragh et al., 2017; Jalbert et al., 2011a, b; Roodenrys et al., 2002). As outlined in Section 1.4.1, redintegration is a process whereby degraded information in STM is repaired by pre-existing representations in LTM (Schweickert, 1993). It is usually couched as a process whereby a candidate word contained within LTM is selected and phonological information from that candidate is used to restore the information that is missing from the degraded representation within STM (e.g., Hulme et al., 1997). The fewer competing candidates there are to select from then the more likely the correct one will be chosen. However, the ND effect posed a problem for this type of redintegration account as words with many neighbours provide more incorrect candidates for selection than words with fewer neighbours.

Interestingly, when Roodenrys et al. (2002) first investigated the effects of neighbourhood density on vSTM they predicted that words from sparse neighbourhoods, rather than words from dense neighbourhoods, would undergo a more successful redintegrative process because of less competition from potential candidates in LTM. In light of an unexpected advantage for dense neighbourhood words it was still suggested that a redintegrative process could be used to explain the ND effect. However, as outlined in Section 1.8.1, it was necessary to cast it as a process reliant upon the overall levels of activation elicited within linguistic networks in LTM. The more neighbours the original target word has then the more neighbours that can still be activated by a partially degraded version of that target word. All neighbours within LTM are mutually connected and because the original to-be-remembered word is essentially an average of all its neighbours it is thought to

receive a boost in activation. The boost in activation makes it more likely that the correct word within LTM will be selected as a candidate for the redintegration process (e.g., Derrough et al., 2017; Jalbert et al., 2011a, b).

Redintegration is not the only explanation for the ND effect though. Rather than the ND effect being the consequence of a redintegration process at retrieval, a possible alternative is that LTM provides support during the encoding of items into STM. Clarkson et al. (2017) described two possible processes whereby LTM can provide differential support dependent upon an item's neighbourhood density. One explanation is reliant upon the order encoding hypothesis (DeLosh & McDaniel, 1996) which states that item information and order information are encoded separately. Successful retrieval of an item, in its correct serial position, is dependent upon both the item and order information being correct. However, there are a finite amount of resources available for encoding and this means that there will be a trade-off between the resources available for item encoding and the resources available for order encoding. The fewer resources that are used to encode item information then the more resources that are left available for order encoding and the more likely that both the item identity and corresponding order information will be correct. When to-be-remembered words are presented they automatically activate their linguistic networks in LTM. Words from dense neighbourhoods automatically elicit more activation than sparse neighbourhood words and this increased activation facilitates the encoding of item information. For the dense neighbourhood words this then leaves more resources available for the encoding of serial order and reduces the likelihood of errors at recall. A second explanation for the ND effect offered by Clarkson et al. is that activation elicited within LTM during encoding serves to support item-to-position associations. The stronger the item-to-position association is for a word then the more likely that word will be recalled in its correct serial position. Because dense neighbourhood words elicit more activation within LTM their item-to-position associations will be stronger than those for sparse neighbourhood words. A key difference between the accounts offered by Clarkson et al. and the redintegrative accounts outlined earlier (e.g., Derrough et al., 2017; Jalbert et al., 2011a, b) is that neighbourhood density is considered to have some impact upon order memory (i.e., the items are remembered but they are remembered in the incorrect order) rather than just item memory (i.e., information regarding the identity of the items is degraded).

While the accounts discussed so far differ in their precise explanations of the ND effect a unifying theme is that words from dense neighbourhoods are better remembered than words from sparse neighbourhoods because they elicit more supportive activation within their linguistic networks in LTM. However, there remain a number of important manipulations that are yet to be tested that would help to sustain or raise questions over these accounts of the ND effect. In the upcoming sections three key manipulations - memory task, presentation modality and word pool size – are outlined as factors that should enable further understanding of the ND effect and help to establish what current models of vSTM can best accommodate the ND effect.

Firstly, regarding the vSTM task, while early studies investigating the ND effect used span tasks and serial recall tasks (e.g., Allen & Hulme, 2006; Roodenrys et al., 2002) later studies have opted mainly for serial reconstruction tasks (e.g., Derragh et al., 2017; Clarkson et al., 2017; Guitard, Gabel, et al., 2018; Jalbert et al., 2011a, b). However, span tasks and serial recall tasks involve reproduction of to-be-remembered items (e.g., spoken, written, typed) whereas serial reconstruction does not. Instead, in serial reconstruction, all to-be-remembered items are re-presented on screen and participants are required to indicate the order that the re-presented items were originally presented in (e.g., by clicking the items or by re-arranging the items). While this may appear to be a minor procedural difference in accomplishing the same task (i.e., remembering a sequence of words in the correct order) serial reconstruction has been used to make contradictory claims about the nature of the ND effect. Clarkson et al. used serial reconstruction to investigate the ND effect because they considered re-presentation of the items to reduce the burden upon item memory and make it a pure test of order memory. Because dense and sparse neighbourhood words are all re-presented at test they argued that there is no need for participants to remember the individual items. The only task requirement is for participants to correctly judge the order of the re-presented items. The re-presented items, rather than pre-existing representations in LTM, could perhaps be considered to repair any degraded traces in STM (e.g. Gathercole et al., 2001). Because both dense and sparse neighbourhood items are re-presented the repair process will be successful for both dense and sparse neighbourhood words. If the locus of the ND effect is at a redintegrative stage operating at retrieval, then using serial reconstruction should attenuate the ND effect. However, the ND effect was

also robust in serial reconstruction which was interpreted as the locus of effect being much earlier than a redintegrative process operating at retrieval. This led to their suggestion that it is an effect upon order, rather than just item, memory. However, an issue is that other authors (e.g., Jalbert et al., 2011a, b) have also opted to use serial reconstruction but with a different rationale. In these instances, serial reconstruction was selected because it permits item reproduction times to be controlled. Clicking and dragging times are equivalent for all items in serial reconstruction whereas written or spoken reproduction times in serial recall can vary. The choice of a serial reconstruction task helped to ensure that the ND effect was a consequence of redintegration rather than some consequence of varying spoken or written output times between dense and sparse neighbourhood items.

While there are clear differences in researcher's *a priori* assumptions regarding serial reconstruction and whether it is a test of item or order memory the task does produce comparable data to serial recall when testing a variety of other effects (e.g., word length, irrelevant speech and phonological similarity effects; e.g., see Jalbert et al., 2011a). This suggests that, despite the procedural differences and claims to the contrary (e.g., Clarkson et al., 2017), serial recall and serial reconstruction are fairly similar tasks that produce comparable outcomes. This also suggests that it is not the best task to use in order to make very precise claims about the nature of the ND effect and whether it is an effect upon order rather than item memory. To help sustain this claim it instead seems necessary to identify another suitable vSTM task that is also considered a test of order rather than item memory. One such option is serial recognition. In serial recognition, the originally presented items are also re-presented on screen, but the task is to simply judge whether those re-presented items are in the same or different order to the originally presented items. Re-presentation of the items in serial recognition is again considered to reduce the burden on item memory and make it a test of order memory (e.g., Gathercole et al., 2001). However, unlike serial reconstruction it produces quantitatively different outcomes to serial recall when testing the same effect. For example, the lexicality effect is robust in serial recall but attenuated or eliminated in serial recognition (e.g., Gathercole et al., 2001, but see Macken et al., 2014). Additionally, word length effects are robust in both serial recall and serial reconstruction (e.g., Jalbert et al., 2011a) but attenuate in serial recognition (e.g., Baddeley, Chincotta, Stafford, & Turk, 2002). The attenuation of these effects could

be interpreted as the requirements of a serial recognition task more successfully reducing the burden upon item memory than serial reconstruction does. As such, in these instances, it suggests that the effects are ones upon item rather than order memory. For these reasons serial recognition can perhaps be considered a better test of order memory than serial reconstruction. By directly comparing the ND effect in serial recall and serial recognition it will be possible to further establish whether the ND effect is one upon order memory (e.g., Clarkson et al., 2017) or whether it is predominantly an effect upon item memory (e.g., Derragh et al., 2017; Jalbert et al., 2011a, b).

The second factor to be investigated relates to the modality that the stimuli within a vSTM task are presented in. Unlike many other linguistic variables (e.g., frequency, lexicality), in non-STM tasks neighbourhood density has different effects upon the encoding of words dependent upon their presentation modality. For example, lexical decision times are slower and participants are more likely to misidentify an auditorily presented word if it comes from a dense rather than a sparse phonological neighbourhood (e.g., Luce & Pisoni, 1998). However, for visually presented words, lexical decision times are quicker and identification is more accurate for the dense rather than sparse phonological neighbourhood words (e.g., Yates, Locker & Simpson, 2004). This points towards there being encoding costs associated with auditorily presented dense neighbourhood words and visually presented sparse neighbourhood words. If these encoding costs have some bearing upon vSTM then a consequence of this should be that the ND effect is attenuated or perhaps even reversed when using auditory presentation. Goh and Pisoni (2003) found a span and serial recall advantage for sparse neighbourhood words when using auditory presentation which lends some support to this suggestion. However, their word pools were identified as being confounded by imageability (another item-level variable known to impact vSTM task performance) with the sparse neighbourhood words having significantly higher imageability ratings (Derragh et al., 2017). Additionally, Roodenrys et al. (2002) and Allen and Hulme (2006) have successfully elicited the ND effect in span tasks and serial recall tasks when using auditory presentation. They suggested that whatever process is responsible for the ND effect must therefore be independent of any costs that are associated with the encoding of dense and sparse neighbourhood words (e.g., Luce & Pisoni, 1998; Yates et al., 2004). However, the dense neighbourhood words in their experiments

may still have been more susceptible to misidentification and required longer identification times than the sparse neighbourhood words but the impact of these costs may not have been strong enough to eliminate the ND effect. Since they did not include a visual presentation condition, it is not possible to judge if the ND effect with auditory presentation was attenuated in such a way though. An attenuation of the ND effect when using auditory compared to visual presentation would implicate there being some impact of these encoding costs upon vSTM. Experiments that have used visual presentation (e.g., Clarkson et al., 2017; Derrrough et al., 2017; Jalbert et al., 2011a, b) have used different word pools or different memory tasks (re-presentation of items in serial reconstruction tasks could counteract any deficit arising from misidentification) which means that a direct comparison between the size of the ND effect in visual and auditory presentation is required. Until a direct comparison between vSTM performance using auditorily and visually presented items is made it is not possible to completely rule out an impact of encoding costs/benefits upon the ND effect.

The final factor to be investigated relates to the size of the word pools that the to-be-remembered items are drawn from. Utilising a small word pool is suggested to make a vSTM task more heavily dependent on order rather than item memory (e.g., Goh & Pisoni, 2003; Saint-Aubin & Poirier, 1999). Studies have found a robust ND effect when using relatively small word pools of 16 items or fewer (e.g., Allen & Hulme, 2006; Jalbert et al., 2011a, b; Roodenrys et al., 2002) and when using larger pools of 47 items or more (e.g., Clarkson et al., 2017; Derrrough et al., 2017). However, Derrrough et al. did report that their ND effect was smaller than other examples in the literature. This is suggestive that the ND effect may be impacted by word pool size with the effect size possibly reducing as the reliance upon item memory increases. However, as most of the experiments have used entirely different sets of items this casts doubt over the usefulness of making any direct comparisons between experiments. Goh and Pisoni (2003) is the only example where the same items have been included in both a large and small word pool. There was an advantage for sparse neighbourhood words when using large pools of dense and sparse neighbourhood words (68 items per pool) but when using a subset of those same words (8 items per pool) the effect was eliminated. However, as already discussed, their materials were confounded by imageability so the relevance of their findings in relation to neighbourhood density is not entirely clear. In the present

experiments an investigation into the ND effect using a large word pool (48 items) before selecting a subset (12 items) of those words will help provide some further indication of whether the ND effect is impacted by changes in word pool size. An attenuation of the ND effect when using a small word pool in serial recall (considered by some to make the vSTM task a test of order memory; e.g., Goh & Pisoni, 2003; Saint-Aubin & Poirier, 1999), as well as an attenuation in serial recognition (also considered a test of order memory; e.g., Gathercole et al., 2001), would cast doubt over the claim that the ND effect is one upon order rather than just item memory (e.g., Clarkson et al., 2017).

To summarise, three important factors - modality, task-type and word pool size - have been highlighted that require further exploration before it is possible to make robust claims about the precise way in which neighbourhood density impacts vSTM. So far, the relative importance of each of these factors and what they could reveal about the locus of the ND effect has been largely overlooked in the vSTM literature. In the 4 experiments that follow the impact of these three factors upon vSTM was tested by manipulating the type of vSTM task (recall vs. recognition), the modality of the presented items (visual vs. auditory) and the word pool size (48 vs 12 items).

## **2.3 Experiment 1**

Experiment 1 assessed the effects of modality on the ND effect by testing vSTM of 6-item sequences constructed of words drawn from 48-item pools of either dense or sparse neighbourhood words. Sequences were presented visually and auditorily and memory for those sequences was tested via serial recall.

### **2.3.1 Method**

#### **2.3.1.1 Participants**

30 participants (mean age 19 years, 23 female and 7 male) were recruited from the Cardiff School of Psychology participation panel and awarded course credits for their participation. Participants were required to have normal/corrected vision and hearing. All stages of the experiment were conducted in accordance with the Cardiff School of Psychology ethics procedures.



### 2.3.1.2 Materials

Stimuli were the 94 single syllable consonant-vowel-consonant words used by Clarkson et al. (2017). They obtained these words and their corresponding phonological neighbourhood densities from the Celex database (Baayen et al., 1993). The phonological neighbourhood densities were used to create two conditions, with 47 words forming a dense neighbourhood word pool and 47 words forming a sparse neighbourhood word pool. In the present experiment, two additional words ('nut' & 'rib') were also added from the Celex database (Baayen et al., 1993). This ensured that all words could be presented an equal number of times. 'nut' was added to the dense neighbourhood word pool and 'rib' was added to the sparse neighbourhood word pool. With the additional stimuli this gave a 48-item dense neighbourhood word pool (*Mean phonological neighbourhood density* = 22.81, *SD* = 4.97, *Mean orthographic neighbourhood density* = 10.90, *SD* = 5.77) and a 48-item sparse neighbourhood word pool (*Mean phonological neighbourhood density* = 9.73, *SD* = 2.78, *Mean orthographic neighbourhood density* = 3.75, *SD* = 3.53) (see Appendix A). *t*-tests revealed that the pools significantly differed on phonological neighbourhood density,  $t(94) = 15.91$ ,  $p < .001$ , and orthographic neighbourhood density,  $t(94) = 7.32$ ,  $p < .001$ . There was no significant difference between the pools on word frequency,  $t(94) = 1.25$ ,  $p = .21$ , or between words where imageability ratings could be obtained<sup>1</sup>,  $t(62) = 0.06$ ,  $p = .95$ .

Dense and sparse neighbourhood density sequences were constructed by selecting 6 items at random, without replacement, from the appropriate word pool until the pool was depleted. Once depleted, all items were returned to the pools and a further set of 6-item sequences were constructed. This process was repeated until a total of 24 sparse neighbourhood and 24 dense neighbourhood sequences were constructed. To familiarise participants with the experimental procedure an additional 6 (3 dense neighbourhood/3 sparse neighbourhood) practice sequences were constructed using the same method.

<sup>1</sup>Imageability values were obtained from N-Watch (Davis, 2005). This includes imageability values taken from the MRC Psycholinguistic Database (Coltheart, 1981) and imageability values collected by Bird et al. (2001). However, not all words have imageability values assigned to them. Imageability values were obtained for 33 words from the dense neighbourhood word pool and 31 words from the sparse neighbourhood word pool.

For the auditory stimuli, each item was recorded in a monotone male voice at a sample rate of 44.1 kHz/16-bit using a condenser microphone. Each item was digitized using Audacity® (version 2.0.1, 2012) software and edited to a duration of 350ms using the 'adjust tempo' option which preserves the pitch of edited items. Items were then assembled into 6-item sequences (650ms ISI) and exported as 16-bit WAV files.

### **2.3.1.3 Design and procedure**

The experiment used a 2x2x6 within-subject, repeated measures design, with modality (auditory, visual), neighbourhood density (dense, sparse) and serial position (1-6) as factors. The experiment took place in a sound attenuated booth with visual sequences presented on a computer monitor and auditory sequences presented via Sennheisser HD280 headphones. For both modalities the onset of trials was initiated by the participant pressing the spacebar. A fixation cross then appeared on screen for 1000ms followed by presentation of the first item. Each item was presented for 350ms separated by a 650ms interval which consisted of a blank screen (visual presentation) or silence (auditory presentation). 2000ms after presentation of the final item, spoken recall in both modalities was cued by the appearance of a centrally located question mark. Participants were instructed to verbally recall the items in the correct order and to replace any missing items with the word "blank". The spoken responses were recorded for later scoring. Participants indicated that they had finished recalling by pressing the spacebar. The visual and auditory stimuli were presented separately in two blocks of 48 trials. The blocks were counterbalanced across participants and for each modality the 24 6-item dense neighbourhood sequences and 24 6-item sparse neighbourhood sequences were presented randomly without replacement.

### **2.3.2 Results**

An item was scored as correct if both its identity and serial position was correct. At each serial position the percentage correct was calculated for the four combinations of modality and neighbourhood density. Serial position curves are shown in Figure 1.

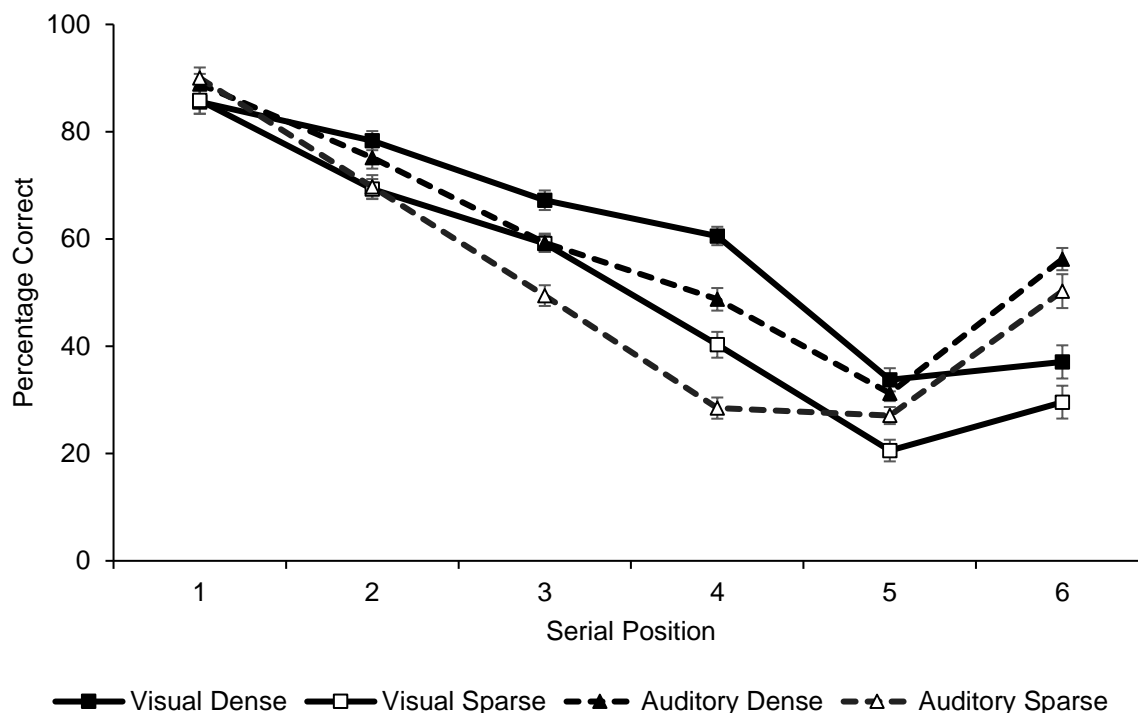


Figure 1. Mean serial position curves for recall of 6-item sequences. Error bars denote within-subject standard error (cf. Cousineau, 2005). Items were scored as correct if both the item identity and serial position were correct.

To assess the effects of neighbourhood density (dense, sparse), modality (auditory, visual) and serial position (1-6) on the serial recall task a 2x2x6 repeated measures ANOVA was conducted. Modality, neighbourhood density and serial position were all within-subject factors. Where necessary, any violations of the sphericity assumption were adjusted using the Greenhouse-Geisser statistic. There was a significant main effect of neighbourhood density  $F(1, 29) = 70.86$ ,  $MSE = 184.8$ ,  $p < .001$ ,  $\eta_p^2 = .71$ , with dense neighbourhood sequences recalled more accurately than sparse neighbourhood sequences. The main effect of serial position,  $F(2.63, 76.24) = 139.47$ ,  $MSE = 774.3$ ,  $p < .001$ ,  $\eta_p^2 = .83$ , was also significant. This reflected a general decline in performance across serial positions until position 6. The main effect of modality,  $F(1, 29) = 0.81$ ,  $MSE = 838.7$ ,  $p = .78$ ,  $\eta_p^2 < .01$ , was not significant. The key interaction of interest between modality and neighbourhood density,  $F(5, 145) = 1.61$ ,  $MSE = 135$ ,  $p = .21$ ,  $\eta_p^2 = .05$ , was not significant suggesting that the size of the ND effect was similar in both modalities. The ND effect in each of the presentation modalities, collapsed across serial position, is

shown in Figure 2. Despite the main effect of modality not being significant there was a significant interaction between modality and serial position,  $F(3.14, 91.11) = 29.09$ ,  $MSE = 207.5$ ,  $p < .001$ ,  $\eta_p^2 = .5$ , revealing both the typical auditory advantage in recency and accompanying visual advantage in medial serial positions (see Macken, Taylor, Kozlov, Hughes & Jones, 2016). The interaction between neighbourhood density and serial position,  $F(5, 145) = 20.29$ ,  $MSE = 67.5$ ,  $p < .001$ ,  $\eta_p^2 = .41$ , was also significant with the size of the ND effect largest at serial positions 3 and 4 but reduced at serial positions 1, 2 and 6. The remaining three way interaction,  $F(3.42, 99.09) = 1.91$ ,  $MSE = 82.6$ ,  $p = .13$ ,  $\eta_p^2 = .06$ , was not significant.

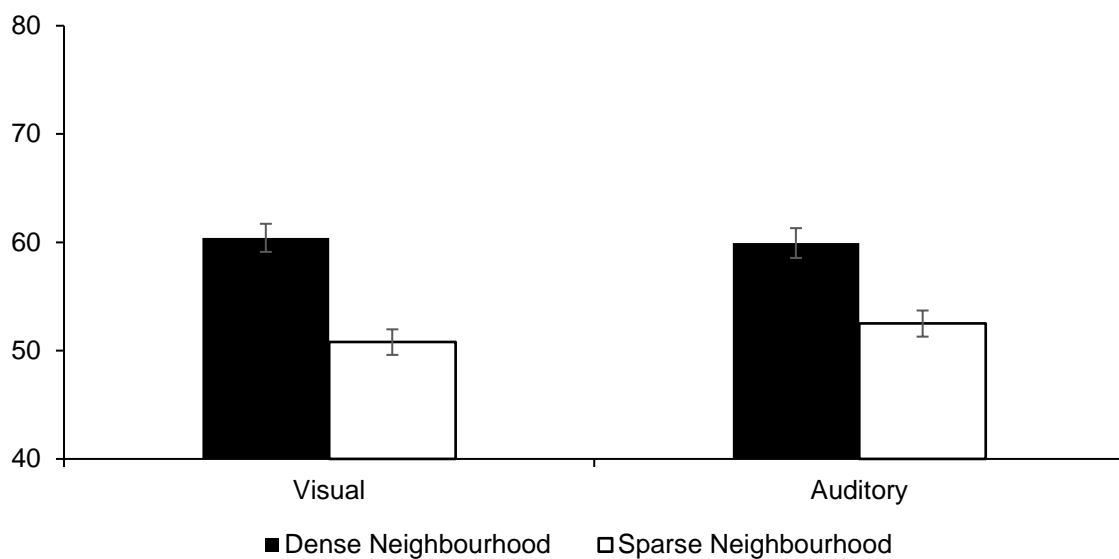


Figure 2. Mean percentage correct in each condition collapsed across serial position. Error bars denote within-subject standard error (cf. Cousineau, 2005).

### 2.3.3 Discussion

Results of Experiment 1 demonstrate that words from a dense neighbourhood are more accurately recalled than words from a sparse neighbourhood. This adds to previous demonstrations of the ND effect (e.g., Allen & Hulme, 2006; Clarkson et al., 2017; Derragh et al., 2017; Guitard, Gabel, et al., 2018; Jalbert et al., 2011a, b; Roodenrys et al., 2002). However, a critical addition to previous research is the finding that the size of ND effect is similar irrespective of whether the serial recall task is presented visually or auditorily. This is of importance because dense neighbourhood words presented auditorily have increased lexical decision and item identification times (e.g., Luce & Pisoni, 1998) whereas there is a facilitative effect for

visually presented dense neighbourhood words (e.g., Yates et al., 2004). If these differences are considered reflective of there being encoding costs/benefits for dense and sparse neighbourhood words dependent upon their presentation modality (e.g., Clarkson et al., 2017) then any differences that there may have been at encoding seem to have no impact upon the outcome of the task. This finding is to be expected if the ND effect is considered to be independent of processes operating at encoding (e.g., Allen & Hulme, 2006; Derragh et al., 2017; Jalbert et al., 2011a, b; Roodenrys et al., 2002). However, it does have important implications for the suggestion by Clarkson et al. (2017) that the ND effect may result from a trade-off between the resources available for item and order encoding. Because of the encoding costs thought to be associated with auditorily presented dense neighbourhood words there should have been an increase in the amount of resources being used for item encoding leaving fewer resources available for order encoding. As such, an attenuation or reversal of the ND effect among the auditorily presented sequences would be predicted. However, this was not the case with a similarly sized ND effect in both modalities.

## **2.4 Experiment 2**

Experiment 2 tested auditory and visual serial recognition of the same to-be-remembered sequences presented in Experiment 1. This not only provided a further test of whether modality differentially impacts the ND effect but also whether it is impacted by the type of task used to investigate it. Unlike serial recall, where participants reproduce a to-be-remembered sequence, in a serial recognition task participants are presented with a to-be-remembered sequence (the standard sequence) and after a short delay presented with a test sequence that is either identical or different (via transposition of two adjacent items) to the standard sequence. Compared to serial recall, serial recognition has been argued to reduce the burden on item memory making it more of a test of order memory (e.g., Gathercole et al., 2001). By using the same sets of words used in Experiment 1 it was possible to establish whether the ND effect also emerges in serial recognition. A ND effect in serial recognition would lend weight to the suggestion that the ND effect is one upon order rather than just item memory (e.g., Clarkson et al., 2017).

## **2.4.1 Method**

### **2.4.1.1 Participants**

30 participants (mean age 19 years, 27 female and 3 male) who had not participated in Experiment 1 were recruited from the same demographic and awarded course credits for participating.

### **2.4.1.2 Materials**

The visual and auditory stimuli were the same as those used in Experiment 1.

### **2.4.1.3 Design and procedure**

The experiment used a 2x2 within-subject, repeated measures design, with modality (auditory, visual) and neighbourhood density (dense, sparse) as factors. In both modalities each trial involved the sequential presentation of a standard sequence using the same temporal parameters as Experiment 1. Following the 2000ms interval, a test sequence was then presented that was either identical to the standard sequence or differed from the standard sequence by transposing two adjacent items. Transpositions in the *different* sequences could occur between serial positions 2 and 3, 3 and 4, and 4 and 5 and each transposition occurred an equal number of times across the trials. Half of the test sequences were identical to the standard sequence and half of the test sequences were different to the standard sequence. 2000ms after presentation of the test sequence a centrally located question mark prompted participants to provide a same/different response via the keyboard ('z' for same and 'm' for different). To ensure that participants understood which key corresponded to which response, 'Same' appeared at the bottom left of the computer screen and 'Different' at the bottom right (locations directly above the corresponding keys). The two physical keys were also labelled with a "Same" or "Different" sticker. The counterbalancing of conditions and randomisation of trials was the same as Experiment 1.

### 2.4.2 Results

Data were collapsed across the *same/different* trials and a percentage correct was calculated for each neighbourhood density in each modality (see Figure 3). A 2x2 repeated measures ANOVA was conducted to assess the effects of modality (auditory, visual) and neighbourhood density (dense, sparse) on serial recognition performance. The main effect of neighbourhood density was not significant,  $F(1, 29) = 0.2$ ,  $MSE = 84.94$ ,  $p = .66$ ,  $n_p^2 = .01$ , indicating that serial recognition accuracy did not differ between the dense and sparse neighbourhood sequences. The main effect of modality was significant,  $F(1, 29) = 14.31$ ,  $MSE = 116.28$ ,  $p = .001$ ,  $n_p^2 = .32$ , with better serial recognition for the auditory sequences compared to visual sequences. The interaction between neighbourhood density and modality was not significant,  $F(1, 29) = 0.63$ ,  $MSE = 80.93$ ,  $p = .44$ ,  $n_p^2 = .02$ .

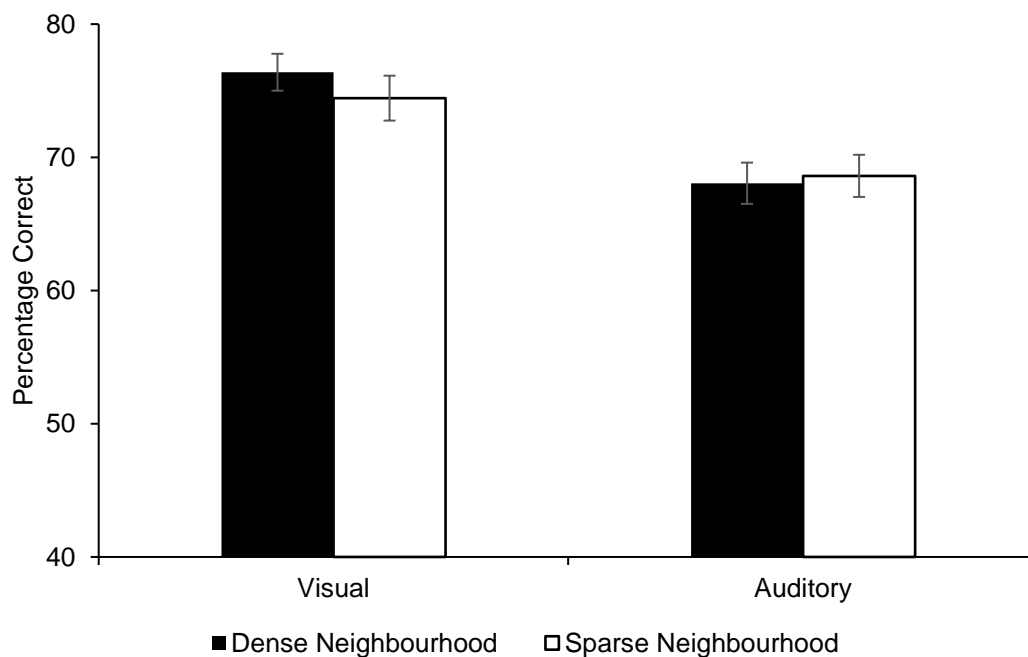


Figure 3. Mean percentage correct in each condition. Error bars denote within-subject standard error (cf. Cousineau, 2005).

### 2.4.3 Discussion

The results of the current experiment demonstrate that the ND effect is eliminated in both auditorily and visually presented serial recognition. However, the ND effect is robust in both auditorily and visually presented serial recall (Experiment

1). The results of the two experiments so far can be best accommodated by accounts of the ND effect that incorporate a process of redintegration at the retrieval stage (e.g., Derragh et al., 2017; Jalbert et al., 2011a, b; Roodenrys et al., 2002). Irrespective of modality, a robust ND effect emerges in visually presented and auditorily presented serial recall. However, when those same words are re-presented in serial recognition the ND effect is abolished. These findings are to be expected if re-presentation of items in serial recognition obviates the need for LTM with the re-presented items serving to repair any degraded items in STM (e.g., Gathercole et al., 2001).

The present findings are more difficult to accommodate within the accounts based upon order encoding offered by Clarkson et al. (2017). No interaction between modality and neighbourhood density was found in either experiment. If longer lexical decision times and the increased likelihood of misidentification of dense neighbourhood words (e.g., Luce & Pisoni, 1998) are considered to reflect there being an encoding cost associated with auditorily presented dense neighbourhood words then the encoding costs have very little (if any) bearing upon vSTM. This raises doubts over the suggestion made by Clarkson et al. that the ND effect is caused by a trade-off between the resources available for item and order encoding. Additionally, a failure to elicit the ND effect in serial recognition, considered a test of order memory (e.g., Gathercole et al., 2001), raises doubts over the suggestion that the ND effect is a consequence of dense neighbourhood words benefitting from the encoding of stronger item-to-position associations (e.g., Clarkson et al.). Weaker item-to-position associations among the sparse neighbourhood words should have led to more incorrect order judgements among the sparse neighbourhood sequences. However, this was not the case with similar recognition performance irrespective of neighbourhood density.

Unlike Experiment 1 there was a main effect of modality in Experiment 2 with serial recognition accuracy higher for the visually presented items. The presentation parameters of the to-be-remembered sequences in serial recall and the standard sequences in serial recognition were identical across the two experiments. Any differences in modality across the two experiments could possibly have arisen as a consequence of what participants were expected to do with the to-be-remembered information (i.e., reproduce items in serial recall vs order judgement in serial recognition). A key difference between the two tasks is that serial recognition can, in



some instances, be completed by only remembering the first letter of each item in the standard sequence - a *first letter strategy*. The order of those first letters can then be compared with the first letters in the test sequence. It is not possible to complete a serial recall task by only remembering the first letter as an item can only be scored as correct if it is accurately reproduced. Whether such an approach would be more likely to be adopted among visual rather than auditory items is unclear but ensuring that word pools are matched on onset consonant frequency, as well as making the onset consonants phonologically similar, is thought make the first letter strategy a less desirable option for participants (e.g., Baddeley et al., 2002). This was incorporated into Experiments 3 and 4 to address this concern.

## **2.5 Experiment 3**

Experiment 3 extended the findings of Experiments 1 and 2 by once again assessing the effects that modality and task type have on the ND effect. This was tested by again using serial recall for both auditorily and visually presented sequences. However, a critical addition to Experiments 1 and 2 is that that 6-item to-be-remembered sequences were drawn from a 12-item rather than a 48-item word pool. Smaller word pools are suggested to reduce the burden upon item memory and effectively make the memory task a test of order memory (e.g., Goh & Pisoni, 2002; Saint-Aubin & Poirier, 1999). Along with the results of Experiment 2 this provided a further test of whether the ND effect impacts order rather than item memory (e.g., Clarkson et al., 2017).

### **2.5.1 Method**

#### **2.5.1.1 Participants**

30 participants (mean age 20 years, 24 female and 6 male) who had not participated in either of the previous experiments were recruited from the same demographic and awarded course credits for participating.

#### **2.5.1.2 Materials**

Stimuli were a 24-item subset of the 96 single syllable consonant-vowel-consonant words used in Experiments 1 and 2. By calculating the percentage correct

that each word achieved in Experiment 1, it was possible to construct dense and sparse neighbourhood word pools on the basis that they had contributed to the ND effect in Experiment 1. Additionally, as the same word pools were used in Experiment 4, to discourage participants from using a first letter strategy in serial recognition, it was also ensured that all the onset consonants were phonologically similar (e.g., Baddeley et al., 2002) and that each word pool had the same dispersion of onset consonants (see Appendix B). A 12-item dense neighbourhood word pool (*Mean Percentage Correct in Experiment 1* = 60.69, *SD* = 16.16, *Mean phonological neighbourhood density* = 24.5, *SD* = 3.78, *Mean orthographic neighbourhood density* = 12.42, *SD* = 6.39) and a 12-item sparse neighbourhood word pool (*Mean Percentage Correct in Experiment 1* = 53.1, *SD* = 13.55, *Mean phonological neighbourhood density* = 8.75, *SD* = 1.82, *Mean orthographic neighbourhood density* = 3, *SD* = 3.59) was constructed (see Appendix B). *t*-tests revealed that the word pools significantly differed on phonological neighbourhood density,  $t(22) = 13.02, p < .001$ , and orthographic neighbourhood density,  $t(22) = 4.45, p < .001$ . There was no significant difference between pools on word frequency,  $t(22) = 0.04, p = .97$  or imageability ratings,  $t(22) = 0.17, p = .87$ . An independent-samples *t*-test also established that, despite attempts to do so, there was not a significant difference between the percentage correct achieved between each word pool in Experiment 1,  $t(22) = 1.25, p = 0.23, d = 0.51$ . Interpretation of the *p* value should be treated with some caution though as a power analysis indicated that, with an effect size of  $d = 0.51$ , there was only a 22% chance of finding a significant effect below the conventional  $\alpha < .05$ . However, the descriptive statistics are indicative that the subset of words chosen contributed towards the ND effect found in Experiment 1<sup>2</sup>.

Dense and Sparse neighbourhood sequences were constructed using the same methodology as Experiment 1. This process produced a total of 24 sparse neighbourhood sequences and 24 dense neighbourhood sequences. 8 (4 dense neighbourhood/4 sparse neighbourhood) practice sequences were constructed to familiarise participants with the experimental procedure.

<sup>2</sup>A possible confound when selecting a subset of words based upon the percentage correct achieved in Experiment 1 is that because the selected words appeared in different serial positions then any differences in the descriptive statistics may have resulted from which serial position the selected words most often appeared in. If one word pool had more words that predominantly appeared in earlier serial positions, then that pool would likely have achieved a higher percentage correct in Experiment 1 simply because of primacy effects. Appendix B outlines how this was controlled for.

### 2.5.1.3 Design and procedure

The experimental design and procedure was identical to Experiment 1.

### 2.5.2 Results

Items were scored and analysed using the same methodology as Experiment 1. Serial position curves are shown in figure 4.

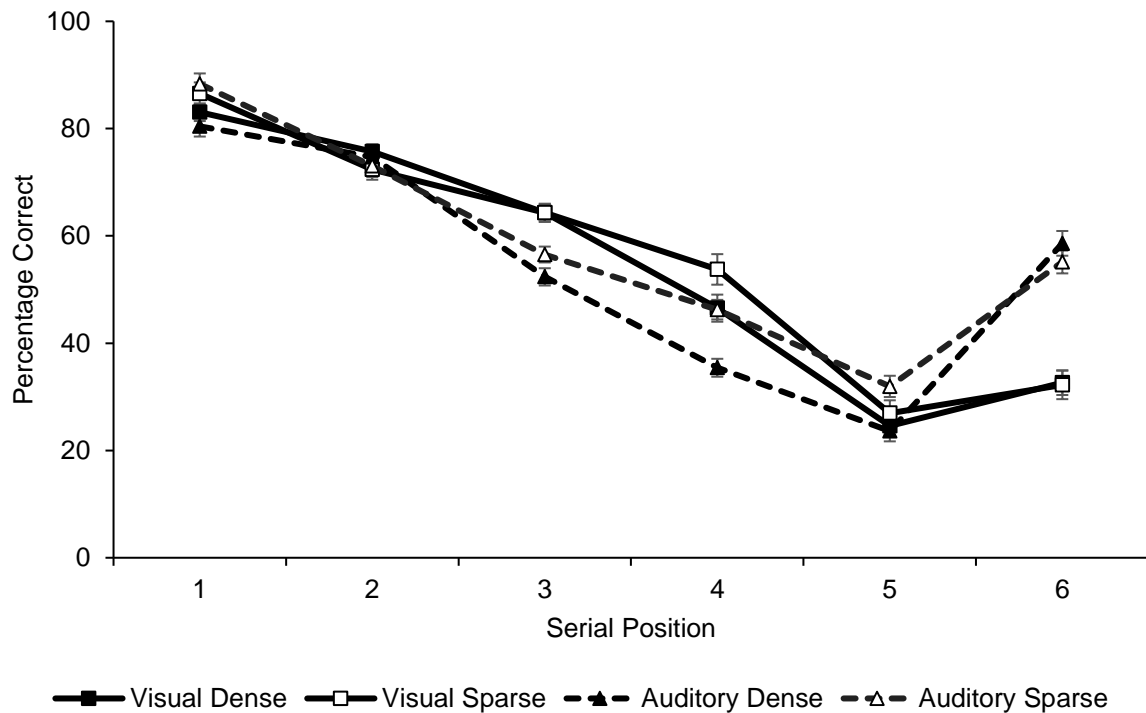


Figure 4. Mean serial position curves for recall of 6-item sequences. Error bars denote within-subject standard error (cf. Cousineau, 2005). Items were scored as correct if both the item identity and serial position were correct.

The main effect of neighbourhood density,  $F(1, 29) = 5.29$ ,  $MSE = 296.5$ ,  $p = .03$ ,  $n_p^2 = .15$ , was significant but unlike Experiment 1 this was due to sparse neighbourhood sequences being recalled more accurately than dense neighbourhood sequences. The effect of serial position,  $F(3.3, 95.65) = 173.26$ ,  $MSE = 471.3$ ,  $p < .001$ ,  $n_p^2 = .86$ , was significant reflecting a general decline in performance across serial positions until position 6. The main effect of modality,  $F(1, 29) = 0.31$ ,  $MSE = 732.6$ ,  $p = .58$ ,  $n_p^2 = .01$ , was not significant. The key interaction of interest between modality and neighbourhood density,  $F(5, 145) = 2.01$ ,  $MSE =$

172.76,  $p = .17$ ,  $n_p^2 = .07$ , was once again not significant suggesting that the advantage for sparse neighbourhood words was not differentially affected by modality. The sparse neighbourhood advantage in each of the presentation modalities, collapsed across serial position, is shown in Figure 5. Despite the main effect of modality not being significant, the interaction between modality and serial position,  $F(3.29, 95.33) = 35.47$ ,  $MSE = 200.2$ ,  $p < .001$ ,  $n_p^2 = .55$ , was significant. This again revealed the typical auditory advantage in recency and visual advantage in medial serial positions (see Macken et al., 2016). The interaction between neighbourhood density and serial position,  $F(3.77, 109.44) = 10.44$ ,  $MSE = 79.7$ ,  $p < .001$ ,  $n_p^2 = .27$ , was also significant with the advantage for sparse neighbourhood words appearing largest at serial positions 1, 4 and 5. The remaining three-way interaction,  $F(3.76, 109.1) = 1.1$ ,  $MSE = 92.5$ ,  $p = .36$ ,  $n_p^2 = .04$ , was not significant.

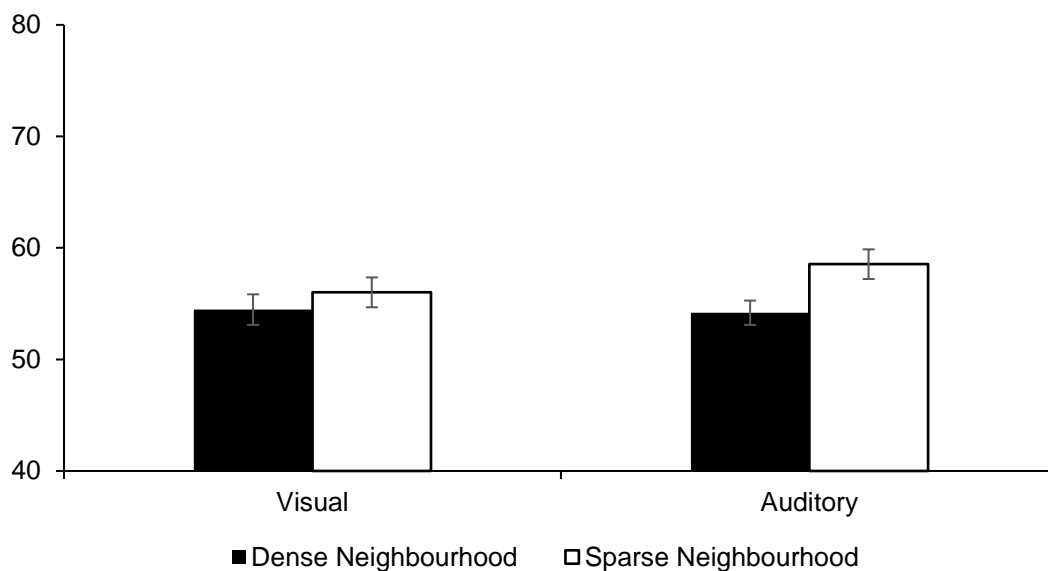


Figure 5. Mean percentage correct in each condition collapsed across serial position. Error bars denote within-subject standard error (cf. Cousineau, 2005).

### 2.5.3 Discussion

Results of Experiment 3 yielded a reversal of the ND effect with more accurate recall of words from a sparse rather than a dense neighbourhood. Other than Goh and Pisoni (2003) this is the only demonstration of a serial recall advantage for sparse neighbourhood words. However, unlike Goh and Pisoni the materials in the current experiment are not confounded by differences in imageability

and were selected because they contributed towards the ND effect in Experiment 1. The ND effect has been argued to arise because dense neighbourhood words elicit more supportive activation in LTM than sparse neighbourhood words do (e.g., Clarkson et al., 2017; Derragh et al., 2017; Jalbert et al., 2011a, b; Roodenrys et al., 2002). However, if to-be-remembered words do elicit supportive activation within their linguistic networks in LTM then the dense neighbourhood words in the present experiment should also have elicited more activation in LTM than the sparse neighbourhood words. As such, vSTM should have been better for the dense rather than sparse neighbourhood words. The finding that words from a sparse neighbourhood, supposedly eliciting less supportive activation in LTM, were better remembered raises doubts over vSTM being dependent upon the levels of supporting activation within LTM.

Although Experiment 3 yielded a reversal of the ND effect, a similarity to the results of Experiments 1 and 2 is that the test of the interaction between modality and neighbourhood density was not significant. The increased lexical decision and item identification times for auditorily presented dense neighbourhood words (e.g., Luce & Pisoni, 1998) and facilitative effects for visually presented dense neighbourhood words (e.g., Yates et al., 2004) once again seem to have no direct impact upon subsequent recall accuracy. With Experiments 1, 2 and 3 all suggesting that modality does not interact with neighbourhood density further doubts are raised over the suggestion made by Clarkson et al. (2017) that the ND effect may be some consequence of there being a trade-off between the resources available for item and order encoding.

A possibility for the reversal of the ND effect is that the reduction in word pool size, similarly to using serial recognition, reduces the burden on item memory (e.g., Goh & Pisoni, 2003; Saint-Aubin & Poirier, 1999). Because some researchers consider the ND effect to be one upon item rather than order memory (e.g., Derragh et al., 2017; Jalbert et al., 2011a, b; Roodenrys et al., 2002) then an ND effect would not be expected in a vSTM task for order memory. As such, the ND effect may have been successfully eliminated in the present experiment with the significant sparse neighbourhood advantage possibly being reflective of a Type 1 error rather than a true effect.

## **2.6 Experiment 4**

Experiment 4 tested auditory and visual serial recognition of the same sequences as presented in Experiment 3. The aim was to provide a test of whether the ND effect emerges in serial recognition when drawing items from a small word pool (both variables that are considered to reduce the burden upon item memory; e.g., Gathercole et al., 2001; Goh & Pisoni, 2003; Saint-Aubin & Poirier, 1999). If any effect of neighbourhood density (whether that be an advantage for sparse or dense neighbourhood words) is eliminated, then this suggests that the sparse neighbourhood advantage in Experiment 3 could possibly have been due to a Type 1 error and that neighbourhood density impacts item rather than order memory. However, if the sparse neighbourhood advantage remains in serial recognition, then this suggests that the particular set of sparse neighbourhood words somehow sustains better vSTM. This would indicate that vSTM task performance can be independent of the supportive levels of activation that are supposedly elicited within LTM (e.g., Derragh et al., 2017; Jalbert et al., 2011a, b; Roodenrys et al., 2002). As such, it would raise further doubts over accounts of vSTM that claim memory will be best for items eliciting higher levels of supportive activation within linguistic networks in LTM.

### **2.6.1 Method.**

#### **2.6.1.1 Participants**

30 participants (mean age 19 years, 24 female and 6 male) who had not participated in any of the previous experiments were recruited from the same demographic and awarded course credits for participating.

#### **2.6.1.2 Materials**

The visual and auditory stimuli were the same as those used in Experiment 3.

#### **2.6.1.3 Design and procedure**

The experimental design and procedure was identical to Experiment 2.

### 2.6.2 Results

The analysis procedure was identical to Experiment 2. Data were collapsed across the *same/different* trials (see Figure 6) and modality (auditory, visual) and neighbourhood density (dense, sparse) were built in to a 2x2 ANOVA. The main effect of neighbourhood density was significant,  $F(1, 29) = 7.66$ ,  $MSE = 92.6$ ,  $p = .01$ ,  $n_p^2 = .21$ , and, similarly to the serial recall performance in Experiment 3, this was because serial recognition was more accurate for the sparse rather than the dense neighbourhood sequences. Similarly to Experiment 2, the main effect of modality was once again significant,  $F(1, 29) = 11.05$ ,  $MSE = 158.4$ ,  $p = .002$ ,  $n_p^2 = .28$ , with serial recognition more accurate for the visually presented sequences compared to auditorily presented sequences. Similarly to Experiments 1, 2 and 3 the interaction between neighbourhood density and modality was not significant,  $F(1, 29) = 0.63$ ,  $MSE = 80.93$ ,  $p = .44$ ,  $n_p^2 = .02$ , suggesting that the advantage for sparse neighbourhood words did not differ as a consequence of modality.

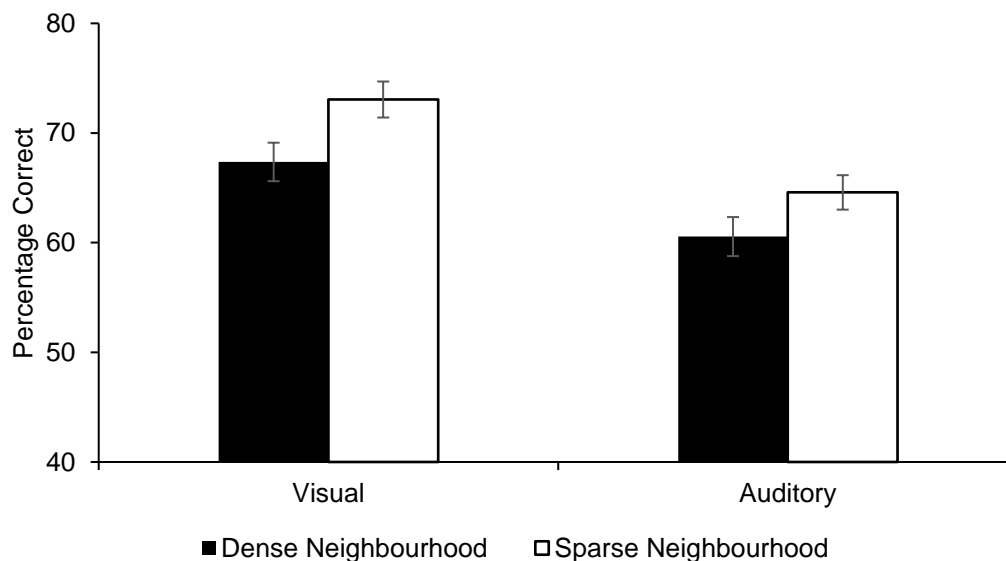


Figure 6. Mean percentage correct in each condition. Error bars denote within-subject standard error (cf. Cousineau, 2005).

### 2.6.3 Discussion

Unlike Experiment 2 the current experiment successfully elicited a neighbourhood density effect in serial recognition. However, similarly to Experiment 3, the advantage was for the sparse rather than the dense neighbourhood sequences. Serial recognition is generally considered to be a pure test of order

memory (e.g., Gathercole et al., 2001) which suggests that whatever drove the effects in Experiment 4 could perhaps be considered as having some effect upon order memory. However, irrespective of whether the task is considered a test of order or item memory the advantage was again for sparse rather than dense neighbourhood words. Once again, if words are considered to elicit activation within their linguistic networks in LTM (e.g., Derragh et al., 2017; Jalbert et al., 2011a, b; Roodenrys et al., 2002) then the dense neighbourhood words should have elicited more supportive activation. However, it was the sparse neighbourhood words, supposedly eliciting less supportive activation that sustained better vSTM. Along with Experiment 3, the results of the present experiment raise doubts over claims that accuracy in a vSTM task is dependent on the levels of supportive activation elicited within linguistic networks in LTM.

## **2.7 General discussion**

The current experiments identified three key parameters - modality, word pool size and memory task type - that required further exploration before making more robust claims about the precise way in which neighbourhood density impacts vSTM. In Experiment 1 the ND effect was successfully elicited in both auditorily and visually presented serial recall when drawing sequences from a 48-item word pool. In Experiment 2, the ND effect was eliminated in visually and auditorily presented serial recognition for the same sets of to-be-remembered sequences. The findings mirror those of other linguistic effects (e.g., the lexicality effect, word length effect) whereby there is typically a robust effect in serial recall but a smaller or eliminated effect in serial recognition (e.g., Baddeley et al., 2002; Gathercole et al., 2001). However, in Experiment 3, when the size of the dense and sparse neighbourhood word pools was reduced to just 12 items, serial recall was better for the sparse rather than the dense neighbourhood words. Additionally, this reversal of the ND effect was also found in Experiment 4 when using visually and auditorily presented serial recognition.

The results of these experiments are problematic for accounts of vSTM whereby the success of vSTM is suggested to be dependent on how much supportive activation is elicited within LTM. Firstly, the data are incompatible with accounts of the ND effect that posit a redintegrative process at retrieval (e.g.,



Derrough et al., 2017; Jalbert et al., 2011a, b; Roodenrys et al., 2002). Degraded words from dense neighbourhoods are considered to elicit more activation within LTM than degraded words from sparse neighbourhoods. The more activation there is then the more successful redintegration is likely to be. Even when drawn from a smaller word pool the dense neighbourhood words in Experiments 3 and 4 should still arguably have been able to elicit more supportive activation within LTM (the ND effect has been found elsewhere when using small word pools of 16 items; Jalbert et al., 2011a, b). Secondly, the findings are incompatible with accounts suggesting that increased activation within LTM helps to support the encoding of order information (e.g., Clarkson et al. 2017). The dense neighbourhood words should have elicited more activation during encoding than the sparse neighbourhood words and therefore benefitted from superior order encoding. Overall, irrespective of the precise mechanism that is suggested to underly successful vSTM, words from denser neighbourhoods are thought to elicit more supportive activation within LTM. The more supportive activation that can be elicited within LTM, then the more accurate vSTM task performance will be. However, this activation failed to predict the recall and recognition performance in Experiments 3 and 4 and instead, performance was better for the sparse neighbourhood words supposedly eliciting less supportive activation in LTM.

An issue touched upon in Section 1.8.4 and that will be explored in more depth in Chapter 4 is that phonological neighbourhood distributions are possibly a consequence of pressures to minimise articulatory effort during the natural development and evolution of language (e.g., Lindblom, 1990). Sparse neighbourhood words may have fewer neighbours because generally they consist of more effortful articulations than dense neighbourhood words. This would mean that dense neighbourhood words should be able to be rehearsed more quickly and possibly be more likely to offset decay (e.g., Baddeley, 2002). Alternatively, they may afford more fluent segmental recoding (e.g., Jones & Macken, 2018; Macken et al., 2015). However, irrespective of whether the ND effect is driven by variations in articulatory difficulty or differing levels of supportive activation within LTM, vSTM should still be better for dense neighbourhood words. As such, the results of Experiments 3 and 4 seemingly undermine the usefulness of considering that vSTM can be impacted by variations in articulatory difficulty. However, one possibility for the sparse neighbourhood advantage is that while the onset letters were controlled

for in order to minimise a first letter strategy in serial recognition (e.g., Baddeley et al., 2002) the phonological onsets were not controlled for. As such, the dense neighbourhood word pool had words beginning with 5 phonemes (/b/, /k/, /p/, /t/ & /θ/) whereas the sparse neighbourhood word pool had words beginning with 7 phonemes (/b/, /f/, /tʃ/, /k/, /p/, /t/ & /θ/). With such small word pools, it would have meant that there was a lot of repetition of phonological onsets within to-be-remembered sequences. However, there would have been the most repetition in sequences drawn from the dense neighbourhood word pool.

The difference in phonological onset repetitions between the dense and sparse neighbourhood words pools is possibly an important oversight as words sharing initial phoneme onsets are considered to act as strong *lexical competitors* to one another (e.g., Sevald & Dell, 1994). Sevald and Dell presented participants with pairs of words that either shared onsets (e.g., PICK-PIN) or shared offsets (e.g., PICK-TICK). They found that participants were able to repeat pairs of words faster when the word pairs differed in their onsets. Additionally, when onsets repeated this encouraged more speech errors to be produced later in the word. In Experiments 3 and 4 it is likely that there were more word pairs sharing onset phonemes in the dense neighbourhood to-be-remembered sequences compared to the sparse neighbourhood to-be-remembered sequences. Instead of Experiments 3 and 4 only manipulating neighbourhood density they may also have inadvertently manipulated lexical competition. As a result of lexical competition, the dense neighbourhood words may have taken longer to rehearse and have been more prone to decay (e.g., Baddeley et al., 2002) or may have involved more disfluent segmental recoding and introduced more speech errors into a motor plan (e.g., Jones & Macken., 2018; Macken et al., 2015). Such lexical competition effects would be less likely to occur in larger word pools because of the wider range of phonological onsets and decreased likelihood of word pairs with shared onsets appearing in the to-be-remembered sequences. While only a speculative account of performance the results of Experiments 3 and 4 do highlight that neighbourhood density may not always predict vSTM performance. This not only raises questions about the usefulness of using neighbourhood density values to advance theories of vSTM but also highlights how very specific combinations of task variables (e.g., word pool size with phonological onsets) may drastically alter how to-be-remembered materials interact with the mechanisms that are proposed to help remember them. A similar issue is returned to

in Chapter 3 whereby the onsets of other experimental word pools used to measure the ND effect are investigated more closely.

Despite best efforts to match the word pools on known item-level variables it is also possible that some other unknown item-level confound was responsible for the sparse neighbourhood advantage in Experiments 3 and 4. However, the word pools did vary on neighbourhood density (both orthographic and phonological). Therefore, dense neighbourhood words should still have elicited more supportive activation in LTM. Additionally, the word pools were equated on imageability which is a confound suggested by Derragh et al. (2017) to have produced the only other example of a vSTM advantage for sparse neighbourhood words (Goh and Pisoni, 2002). To claim that the results were caused by some other confounding item-level variable would first require identification of another item-level variable that is somehow able to override the supportive activation elicited by the neighbours of a word. While this is a possibility it is very speculative. Neighbourhood density has been used to reverse the word length effect (e.g., Jalbert et al., 2011b) with long words from dense neighbourhoods sustaining better vSTM than short words from sparse neighbourhoods (although see Guitard, Saint-Aubin, et al., 2018 & Chapter 3 of current thesis). This would implicate neighbourhood density as having very strong impacts upon vSTM even when to-be-remembered materials are not equated on all item-level variables. Therefore, it seems unlikely that the results of the present experiments were a consequence of some other item-level confound. This issue is returned to in much more depth though (Section 6.4.1) in light of findings across the remainder of the thesis.

Overall, the results of the experiments contained within this chapter cannot be easily accommodated by accounts of vSTM that claim task accuracy to be dependent upon the level of supportive activation elicited within linguistic networks in LTM. As such, processes that have previously been suggested to be responsible for the ND effect (e.g., redintegration or more robust encoding) are called into question. Instead, giving greater consideration to aspects such as the underlying variability of neighbourhood density or the impact that changes in task variables (e.g., reducing word pools and the specific combinations of word onsets) can have upon vSTM may provide a clearer understanding of why neighbourhood density impacts vSTM.

## Chapter 3

### **Accuracy in Serial Reconstruction tasks: A consequence of Neighbourhood Density, Orthographic Frequency or a First Letter Confound?**

#### **3.1 Abstract**

The results of Chapter 2 raise doubts over the ND effect being a consequence of the levels of supportive activation being elicited within linguistic networks LTM. As such, more consideration of what could be driving examples of the ND effect, beyond the neighbourhood density manipulation, is warranted. Research by Jalbert et al. (2011b) and Guitard, Saint-Aubin, et al. (2018) has been used to inform current understanding of the ND effect. However, even when controlling many item-level variables there remains the risk of others being overlooked. Serial reconstruction is a task that can be accurately completed with only partial knowledge of what the to-be-remembered words were. For example, one effective strategy involves only remembering the onset letters of to-be-remembered words and using them to inform the order in which to click the re-presented items. As such, unless onset letter distributions are matched across word pools then effects found in serial reconstruction could be influenced by variations in the distribution of onset letters rather than the item-level variable of interest. In a series of simulations, it was explored whether the particular distribution of onset letters in the to-be-remembered word pools used by Jalbert et al. and Guitard, Saint-Aubin, et al. could have yielded similar patterns of data to those found by the researchers. The simulations modelled performance by assuming a task completion strategy whereby only the onset letters of each to-be-remembered item are remembered. At reconstruction, those onset letters are then used as cues in which to decide the order that the sequence should be reconstructed. Across 8 simulations, the patterns of data were very similar to those found by both Jalbert et al. and Guitard, Saint-Aubin, et al. This raises questions whether their results were really a consequence of their item-level manipulations or whether the effects were perhaps some consequence of the word pools simply not being matched on onset letter frequency.

### **3.2 Introduction to the simulations**

In Experiments 3 and 4 (Chapter 2) the ND effect was reversed with better vSTM for sparse rather than dense neighbourhood words. These findings are incompatible with typical accounts of the ND effect that suggest performance to be a consequence of supportive activation being elicited within LTM (e.g., Roodenrys et al., 2002; Jalbert et al., 2011a, b; Clarkson et al., 2017; Derragh et al., 2017). A possibility is that some other factor (whether an item-level confound or perhaps some particular combination of the onset phonemes contained in the dense and sparse neighbourhood word pools) determined accuracy. However, rather than only questioning the ‘seemingly’ unusual results, it is also important to remain open-minded that previous research into the ND effect may also have been driven by some other variables besides from neighbourhood density. Work by Jalbert et al. (2011b) and Guitard, Saint-Aubin, et al. (2018) have both been used to inform our current understanding of the ND effect and are the subject for closer inspection in the current chapter for reasons that will be set out below.

As discussed in Chapter 1 (Section 1.8.2) Jalbert et al. (2011b) noted that almost all previous experimental demonstrations of the word length effect were confounded by neighbourhood density. Long words typically have fewer phonological and orthographic neighbours than short words and previous experiments investigating word length had not controlled for neighbourhood density. The confound raised the question of whether length or neighbourhood density was responsible for previous examples of the word length effect. Jalbert and colleagues conducted three experiments that explored the impact of neighbourhood density upon the serial reconstruction of words and non-words. In Experiment 1, participants were presented with sequences of 6 to-be-remembered consonant-vowel-consonant (CVC) words taken from either a dense or sparse neighbourhood. Additionally, participants were also presented with to-be-remembered mixed sequences that were comprised of both dense and sparse neighbourhood words. Experiment 2 followed an identical procedure, but the to-be-remembered sequences consisted of single syllable non-words, rather than words. They were once again taken from either a dense or sparse neighbourhood. After each to-be-remembered sequence had been presented all the items were re-presented on screen. Participants reconstructed the order of the originally presented items by clicking on each item in the order that they were first presented. Participants completed the tasks in silence or were instructed to

engage in articulatory suppression during the sequence presentation stage (required to repeatedly say 'A-B-C-D-E-F-G' out loud). In both experiments the reconstruction accuracy was better for dense neighbourhood items compared to sparse neighbourhood items and this effect was removed by articulatory suppression. Reconstruction accuracy for the mixed sequences was similar for dense and sparse neighbourhood items. These findings mirror the effects typically found in the word length effect literature. The word length effect is usually abolished under articulatory suppression (e.g., Baddeley et al., 1975) and when using mixed sequences (e.g., Bireta, Neath & Surprenant, 2006). Because similar effects were found when length was equated, but only neighbourhood density varied, Jalbert et al. suggested that this was evidence of the word length effect essentially being another example of the ND effect.

In their Experiment 3, Jalbert et al. (2011b) factorially manipulated length and neighbourhood density. Participants completed a serial reconstruction task with to-be-remembered sequences of long (2 syllable) or short (single syllable) non-words taken from either a dense or sparse neighbourhood. Results revealed no overall main effect of word length but a significant main effect of neighbourhood density. Reconstruction accuracy for the short non-words from a sparse neighbourhood was better than it was for the long non-words from a dense neighbourhood. Additionally, there was a reversal of the word length effect with more accurate reconstruction for the long non-words from a dense neighbourhood compared to the short non-words from a sparse neighbourhood. These findings are particularly problematic for any accounts of vSTM that incorporate decay offset by rehearsal (e.g., Baddeley, 2000; Burgess & Hitch, 1999, 2006; Page & Norris, 1998). Long words, irrespective of neighbourhood density, should take longer to rehearse than short words and therefore be more susceptible to decay within STM. The findings are also difficult for an object-oriented account (e.g., Jones & Macken, 2018; Macken, Taylor & Jones, 2015) to accommodate. Long words are typically phonologically more complex than short words (e.g., Caplan et al., 1992; Service, 1998) and could be considered to involve more disfluent segmental recoding than short words. The findings by Jalbert et al. raise serious questions over a causal role for rehearsal upon vSTM. However, the findings can be readily accommodated by accounts of vSTM that incorporate a redintegrative process reliant upon the levels of supportive activation in LTM (e.g., Derrahg et al., 2017; Jalbert et al., 2011a, b). Irrespective of length, words with

more neighbours will elicit more supportive activation within linguistic networks in LTM and more readily undergo a successful redintegrative process.

Guitard, Saint-Aubin, et al. (2018) highlighted a possible confound in Jalbert et al. (2011b). Single syllable words typically have many neighbours and, in English, only around 13% of 4 letter single syllable words have fewer than three neighbours. As such, short words/non-words with relatively few neighbours could be considered atypical and nonrepresentative of more common single syllable words. They also noted that those atypical words are more likely to contain unusual orthographic structures (e.g., the orthographic structure of the sparse neighbourhood words *grief* and *shriek* is less typical than the orthographic structure of the dense neighbourhood words *pain* and *beam*). Therefore, the orthographic structures of short words/non-words from a sparse neighbourhood used within an experiment are very likely to be non-representative of the orthographic structures for more typical short words. Rather than variations in neighbourhood density, variations in orthographic structure may be responsible for the ND effects found by Jalbert et al.

Guitard, Saint-Aubin, et al. (2018) highlighted the importance of orthographic structures by controlling neighbourhood density but varying orthographic frequency (the number of times that segments of words appear in a given language). Orthographic frequency was controlled for by using n-gram measures that take unigram, bigram and trigram frequency into account (see Section 1.8.3 for more detail). Guitard, Saint-Aubin, et al. were able to demonstrate a serial reconstruction advantage for long four syllable words (e.g., *concubinage*) over short two syllable words (e.g., *chlamyde*) but only when the long words had more frequent orthographic structures than the short words did. When long and short words were matched on both neighbourhood density and orthographic structure the usual vSTM advantage for short words over long words consistently appeared and was abolished under conditions of articulatory suppression (Experiments 1, 2, 3, 5 and 6). Guitard, Saint-Aubin, et al. concluded that most of their results could be easily accommodated by trace decay offset by rehearsal (e.g., Baddeley, 2000). The reversal of the word length effect by Jalbert et al. (2011b), and the reversal of the word length effect in their Experiment 4, was likely due to the orthographic structures of the to-be-remembered material. They suggested that such an advantage for long words over short words, when orthographic frequency is not controlled for, can still be accommodated by a trace decay system if a redintegration process is considered

to operate at retrieval. They suggested that because common orthographic structures are more frequently used within everyday language that those same orthographic structures will also be more readily accessible within LTM. When it comes to repairing degraded traces in STM the more readily accessible and available information is within LTM then the more likely that the redintegrative process will be successful (e.g., Hulme et al., 1997).

Jalbert et al. (2011b) and Guitard, Saint-Aubin, et al. (2018), despite differing in their view over rehearsal having any causal impact upon vSTM, both share the view that vSTM benefits from support from LTM. Their results are used as evidence to further their theoretical viewpoint. However, recently there has been an increased interest in the ways that different vSTM tasks can be completed by participants. For example, Morrison et al. (2016) asked participants to complete a variety of vSTM tasks (e.g., serial recall, serial recognition, span tasks) and afterwards asked participants to select, from a list, which strategy they had chosen to use (e.g., rehearsal, mental imagery, other). They found that even among the same task (e.g., serial recall) there were a wide variety of completion strategies reported by participants. Additionally, despite only asking participants whether they rehearsed, the authors acknowledge that there are likely to be numerous rehearsal strategies that are used, dependent upon the task. For example, in serial recall participants may only rehearse one item at a time (single word rehearsal) or rehearse at least two items in succession (cumulative rehearsal). The research suggests that, despite experimenters having some a priori theory about what is happening during a vSTM task (e.g., items are encoded into the phonological loop; Baddeley, 2000), participants may actually be opting to use a wide variety of strategies (possibly not considered by the researchers) to complete the task. Additionally, these strategies may be entirely independent of any theoretical vSTM mechanisms. Morrison et al. suggest that the great variation in observed strategies raises questions about the very existence of a common underlying vSTM system. The effects and specific processes that emerge within a given task environment may be better understood as a consequence of the materials, the output requirements of the task and the particular skills and knowledge that are possessed by the participant (e.g., Jones & Macken, 2018; Macken et al., 2015).

The finding that participants adopt such a wide variety of strategies in vSTM tasks (Morrison et al., 2016) enables some further consideration of what the data



found by Jalbert et al. (2011b) and Guitard, Saint-Aubin, et al. (2018) reveals. Unlike serial recall, serial reconstruction provides a particular task setting whereby participants can successfully reconstruct the sequences without having to have correctly remembered the word. Firstly, the task can be completed by adopting a first letter strategy. Instead of attempting to memorise an entire sequence of to-be-remembered items a first letter strategy involves participants opting to only remember the onset letter of each to-be-remembered item. At the reconstruction stage, when all items are re-presented, the order of the onset letters can be used as cues to decide the order that each item should be clicked. A first letter strategy is most effective when all items in a to-be-remembered sequence begin with a different first letter and its effectiveness reduces as the number of duplicate onset letters increase. Participants may choose to adopt a first letter strategy because vSTM for letters is generally more accurate than it is for words. At 6 and 7-item sequence lengths recall accuracy is around 50% for an entire string of to-be-remembered letters but less than 25% for an entire sequence of to-be-remembered 3-letter words such as *ask*, *cab*, *cry*, etc. (e.g., Crannell & Parrish, 1957). As the to-be-remembered words get longer then accuracy is only likely to further decrease (e.g., Baddeley et al., 1975). A reduction in errors may be enough motivation for participants to switch to a first letter strategy. Participants may also wish to give the impression to the experimenter that they are performing well in the memory task (similar to *social desirability bias* whereby participants in an experiment wish to present an idealised version of themselves; e.g., Grimm, 2010).

A second possibility is that even if participants do attempt to remember items in their entirety (e.g., perhaps via rehearsal) the results of serial reconstruction could still be dependent on the distribution of onset letters within the to-be-remembered sequences. Speech errors among sequences of to-be-remembered consonant-vowel-consonant (CVC), consonant-consonant-vowel (CCV) and vowel-consonant-consonant (VCC) words are not random and portions of the words tend to behave as separate units (e.g., Treiman & Danis, 1988). The onset letters of to-be-remembered words are usually maintained, but the words most often break at the C/ VC boundary, at the CC/V boundary or at the VC/C boundary. They then recombine with the same unit from another to-be-remembered word. These types of errors are not just confined to single syllable words though and similar patterns are also found among those segments in disyllabic and trisyllabic words (e.g., Treiman, Fowler,

Gross, Berch, & Weatherston, 1995). Errors within vSTM tasks such as serial recall are very similar to those contained within speech (e.g., Ellis, 1980).

While serial reconstruction does not require vocalisation of the items it still requires the sequence to be reconstructed. Some researchers assume that the re-presented items repair any degraded traces in STM, essentially making serial reconstruction a pure test of order memory (e.g., Clarkson et al. 2017). However, if successful reconstruction is perhaps dependent on first retrieving degraded items from vSTM with attempts to repair them resulting in errors akin to speech errors (e.g., Page & Norris, 1998) or dependent upon the contents of a motor plan which is prone to errors (e.g., Jones & Macken, 2018; Macken et al., 2015) then a serial reconstruction task would not necessarily capture those errors. Instead, any experimental conditions that allow for overcoming errors (e.g., by having more unique onset letters) would likely facilitate better reconstruction performance. For example, a to-be-remembered sequence (e.g., 'cat', 'dot', 'sit') could end up having three VC errors that maintain the onset letters (e.g., 'cot', 'dit', 'sat'). However, this could afford more accurate reconstruction than a to-be-remembered sequence (e.g., 'pit', 'pat', 'tat') with just two VC errors that maintain the onset letters (e.g., 'pat', 'pit', 'tat'). This is despite the first example containing no correct words and the second example still containing all three. At reconstruction, if the remembered words are then used as cues in which to decide which items to click then the second example would result in an order error for the first two items and the third item being scored as correct. However, in the first example, the entirely incorrect words could still be used to accurately complete the serial reconstruction task. The participant could first remember the incorrect word 'cot' and attempt to find it on screen. However, because it would not be available for selection a reasonable guess then might be to select another word with the same onset letter (e.g., the correct word 'cat'). The same process of using the onset letter to inform which item to click would then also work for the all the remaining incorrectly remembered words (e.g. 'dit' must have been 'dot' and 'sat' must have been 'sit'). Irrespective of the exact mechanisms required for remembering and reconstructing a sequence of items, if the results of the vSTM task are not actually a true reflection of what the participant has remembered then they cannot be used to reliably inform the development of any vSTM theories.

Any effects found when using serial reconstruction cannot be attributed to a first letter strategy if the word pools corresponding to different conditions/variables are matched on onset letter frequency or if the pools are sufficiently large to ensure fairly even distributions of unique and duplicated onset letters across all to-be-remembered sequences. In these cases, if participants use first letters as cues in which to reconstruct the sequence, task accuracy would be similar between each word pool irrespective of any item-level manipulations. The ND effect has been found in serial reconstruction with 47-item word pools (Clarkson et al., 2017) and open word pools (Derragh et al., 2017). Additionally, it has been found in span (e.g., Roodenrys et al., 2002) and serial recall tasks (e.g., Allen & Hulme, 2006) whereby the first letter strategy would be of little use - choosing only to remember the first letter would not be particularly helpful when the task requires the entire item to be reproduced. Such tasks also enable speech errors to be captured. This suggests that generally the ND effect is not a consequence of onset letter confounds. However, a reversal of the word length effect (whether that be due to neighbourhood density or orthographic frequency) has only been demonstrated using serial reconstruction for sequences drawn from small word pools (16 items or fewer). Additionally, neither Jalbert et al. (2011b) nor Guitard, Saint-Aubin, et al. (2018) controlled their word pools for the number of onset letters.

Given the theoretical importance that the work by Jalbert et al. (2011b) and Guitard, Saint-Aubin, et al. (2018) has, the aim of the following section is to assess what performance would have been like in their experiments had participants used the first letters as cues in which to inform the order that items should be reconstructed. This was achieved by running a series of simulations that closely followed their experimental procedures but where sequence reconstruction was based purely on a first letter strategy. While simulations cannot provide evidence for a particular strategy being used, if the patterns of data are different to those obtained by the experiments, then it at least demonstrates that the experimental effects could not have been driven by variations in onset letter distributions across word pools and participants opting to use a first letter strategy. However, if the patterns of data are similar between experiments and simulations then it raises the possibility that the experimental results were a consequence of the onset letter distributions within each word pool rather than the item-level manipulations of neighbourhood density or orthographic frequency.

### **3.3 Method**

A simulation for each experiment was programmed and run in PsychoPy (Peirce, 2007). Data was obtained for  $n = 100$  in each simulation. This sample size was decided because it provided a situation whereby the simulations were highly powered (>80%) to detect effect sizes of the magnitude originally found by the authors. It therefore helps to reduce the possibility of any of the simulation results being a Type 1 error. Larger sample sizes also provide more accurate estimates of effect size (e.g., Coe, 2002) so the impact of a first letter strategy upon task accuracy could be more precisely calculated.

Each word from the experimental word pools used by Jalbert et al. (2011b) and Guitard, Saint-Aubin, et al. (2018) was replaced with the word's onset letter. The simulations then closely followed their experimental procedures with 15 trials in each experimental condition and to-be-remembered sequences consisting of 6 or 7 onset letters.

#### **3.3.1 Simulation Assumptions**

The simulations assumed perfect memory of sequences of 6 or 7 letters that corresponded to the onset letters of the words/non-words used by Jalbert et al. (2011b) and Guitard, Saint-Aubin, et al. (2018). Although this may not be representative of real-life performance (e.g., serial recall accuracy in Crannell & Parrish, 1957 for 6-item sequences of letters was around 50%) the aim of the simulations is to assess whether a first letter strategy would reveal a similar pattern of results to those found originally by the authors. If the simulations did not assume perfect memory for the sequences of onset letters, then several further assumptions would have to be made about the types of errors and how those errors are dealt with during reconstruction. It does mean that the percentage correct in the simulations is likely to be far higher than those found in the experiments though. However, as the aim is not to provide an exact data fit but just to establish whether the same general pattern of data will emerge, this was not considered to be an issue. The simulations then assumed that only those onset letters informed the order in which to reconstruct the sequence of those letters. Whenever there were duplicate onset letters a guess (at chance level) was made by the simulation as to the original order in which they were presented. Within a serial reconstruction task, a participant is likely to have

more information regarding the original to-be-remembered item in which to inform a reconstruction decision. However, incorporating this within a simulation would once again involve several further assumptions. Finally, no attempt was made to model serial position curves or to model performance under conditions of articulatory suppression. This was decided upon in order to avoid the inclusion of any theoretical vSTM mechanisms.

### **3.3.2 Pure Sequence Simulations**

For each experimental condition, to-be-remembered onset letters were drawn at random from the relevant word pools. For the Jalbert et al. (2011b) and Guitard, Saint-Aubin, et al. (2018, Experiment 6) simulations this meant drawing 6 onset letters. For the Guitard, Saint-Aubin, et al. (Experiments 2-5) simulations this meant drawing 7 onset letters. The simulations assumed perfect memory for the order of these onset letters. As such, any unique letters contained in the drawn sequence were automatically scored as correct by the simulation. This is because unique onset letters can be used to inform accurate reconstruction of those sequences of letters. For example, a to-be-remembered sequence containing the six onset letters *A-C-B-D-G-F* can then be used to inform the order in which to click the re-presented onset letters (e.g., *A* followed by *C* followed by *B* and so on). Any trial comprised solely of unique onset letters always scored 100%. However, when duplicated onset letters were drawn (e.g., *A-B-D-G-F-A*) only some of the onset letters (e.g., *B-D-G-F*) could be used to inform accurate reconstruction of the sequence. In this example, it means that four of the onset letters are guaranteed to be accurately reconstructed (i.e., *B* will be clicked second, *D* clicked third, *G* fourth and *F* fifth) and the trial will score a minimum of 66.67%. For the remaining two onset letters though it becomes necessary to guess which re-presented onset letters correspond to the originally drawn onset letter (i.e., there are two re-presented *As* but which of them corresponds to serial position 1 and which to serial position 6 is not known). In these situations, all possible permutations of the reconstruction order for identical onset letters was made available to the simulation and one of those permutations was selected by the simulation at random. With two duplicated onset letters, they can either both be correct, or both be incorrect (2 permutations). With sets of three duplicated onset letters there are 6 possible permutations of order and with four duplicated letters

there are 24 possible permutations. Returning to the example, both As could be placed in the correct order and the trial would score 100% or both As could be placed in the incorrect order and the trial would score 66.67%. There was a 50% chance of the simulation selecting either outcome. This basic logic formed the basis for every simulation trial. For each experimental condition, the percentage correct for each corresponding simulation trial was totalled and then an average calculated to provide an overall simulated condition mean.

### **3.3.3 Mixed Sequence Simulations**

Guitard, Saint-Aubin, et al. (2018) did not use mixed sequences but in Jalbert et al. (2011b) there were two types of mixed sequence - Dense, Sparse, Dense, Sparse, Dense, Sparse (DS) or Sparse, Dense, Sparse, Dense, Sparse, Dense (SD). The simulations for mixed sequences were very similar to the pure sequence simulations except that information pertaining to the scores achieved at each serial position were held by the simulation. The overall correct for dense neighbourhood words was then calculated by combining the scores from the 1<sup>st</sup>, 3<sup>rd</sup> and 5<sup>th</sup> letters from all DS trials with the scores from the 2<sup>nd</sup>, 4<sup>th</sup> and 6<sup>th</sup> letters from all SD trials. An overall average was then calculated to provide the mean percentage correct for dense neighbourhood words. Overall correct for sparse neighbourhood words was calculated by combining the scores from the 1<sup>st</sup>, 3<sup>rd</sup> and 5<sup>th</sup> letters from all SD sequences and the scores from the 2<sup>nd</sup>, 4<sup>th</sup> and 6<sup>th</sup> letters from all the DS sequences. An overall average was then calculated to provide the mean percentage correct for sparse neighbourhood words. Each type of sequence (DS or SD) therefore contributed 50% to the overall score achieved by each neighbourhood density.

### 3.4 Results

#### 3.4.1 Jalbert et al. (2011b) Simulation Results.

Three simulations, corresponding to each of the three experiments, were used to simulate the data that Jalbert et al. (2011b) would have obtained had participants only adopted a first letter strategy.

##### 3.4.1.1 Simulation 1

Figure 7 shows the data obtained by Jalbert et al. (2011b, Experiment 1) and the data obtained by the simulation.

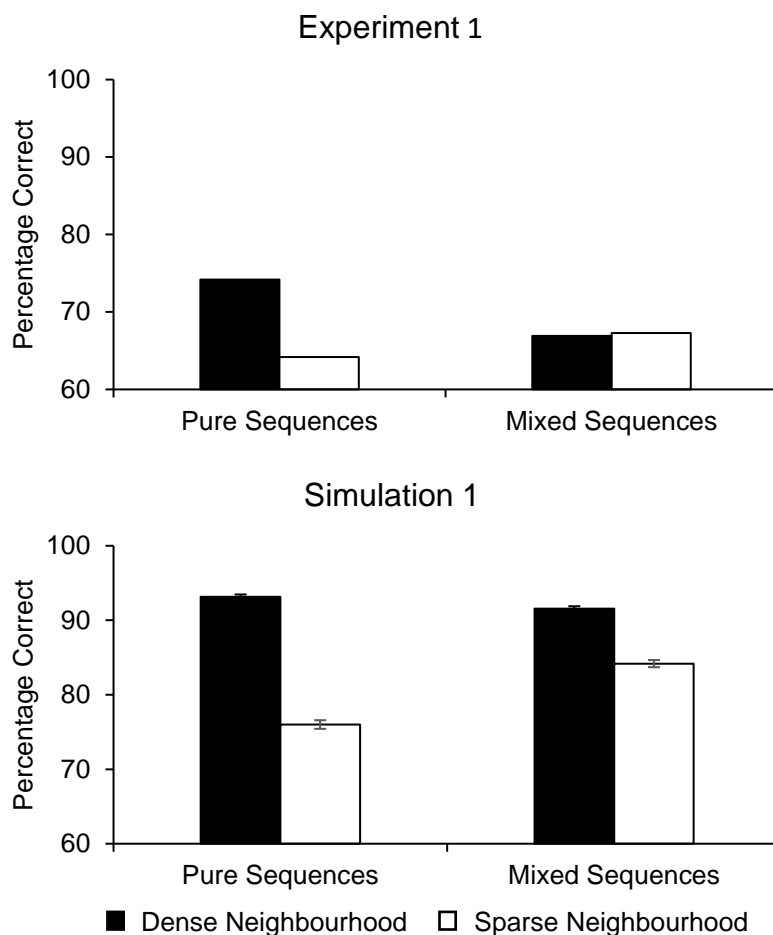


Figure 7. Mean percentage correct in each experimental condition in Jalbert et al. (2011b, Experiment 1) and Simulation 1. From the data provided by Jalbert et al. error bars could not be calculated for experimental data but for the simulated data they show standard errors of the means.

In Jalbert et al. (2011b, Experiment 1) they reported a significant effect ( $p < 0.05$ ) of neighbourhood density in the pure sequences with better reconstruction accuracy for sequences from a dense rather than a sparse neighbourhood. There was no significant difference found in mixed sequences. The data from the simulation was analysed using a 2x2 repeated measures ANOVA that included sequence type (pure, mixed) and neighbourhood density (dense, sparse). The main effects of sequence type,  $F(1, 99) = 46.5$ ,  $MSE = 23.3$ ,  $p < .001$ , and neighbourhood density,  $F(1, 99) = 848.1$ ,  $MSE = 17.8$ ,  $p < .001$ , were significant. Similarly to Jalbert et al., performance was better for the dense neighbourhood sequences. However, unlike Jalbert et al. this advantage seemed irrespective of whether the sequences were mixed or pure. Importantly however is that the interaction between neighbourhood density and sequence type was also significant,  $F(1, 99) = 146.1$ ,  $MSE = 16.3$ ,  $p < .001$ . Pairwise comparisons (two-tailed) suggest that this is a consequence of there being better reconstruction accuracy for the dense neighbourhood sequences compared to sparse neighbourhood sequences in the pure sequences,  $t(99) = 26.5$ ,  $p < 0.001$ ,  $d = 2.65$ , but a reduced effect size among the mixed sequences,  $t(99) = 14.4$ ,  $p < 0.001$ ,  $d = 1.44$ . Although the simulation for mixed sequences did not eliminate the advantage for dense neighbourhood sequences it did attenuate it somewhat.

### **3.4.1.2 Simulation 2.**

Figure 8 shows the data obtained by Jalbert et al. (2011b, Experiment 2) and the data obtained by the simulation.



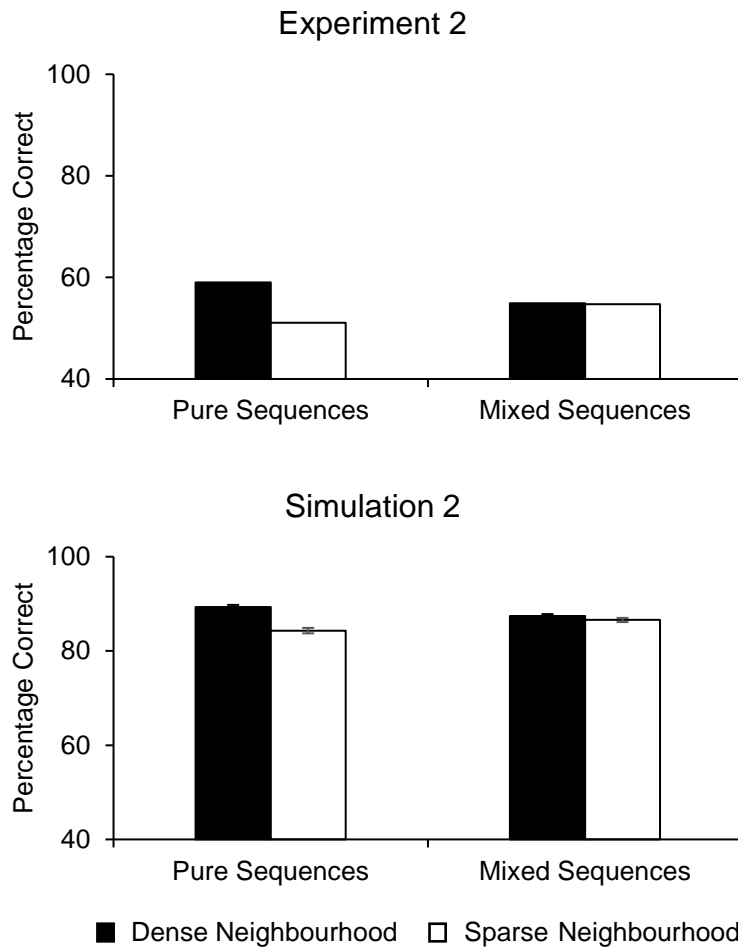


Figure 8. Mean percentage correct in each experimental condition in Jalbert et al. (2011b, Experiment 2) and Simulation 2. From the data provided by Jalbert et al. error bars could not be calculated for experimental data but for the simulated data they show standard errors of the means.

In Jalbert et al. (2011b, Experiment 2), similarly to their Experiment 1, there was once again a significant effect ( $p < 0.05$ ) of neighbourhood density in pure sequences but no difference in mixed sequences. The data from the simulation was again analysed using a 2x2 repeated measures ANOVA that included sequence type (pure, mixed) and neighbourhood density (dense, sparse). The main effect of neighbourhood density was significant,  $F(1, 99) = 44.59$ ,  $MSE = 19.22$ ,  $p < .001$ . The main effect of sequence type was not significant,  $F(1, 99) = 0.1$ ,  $MSE = 25.45$ ,  $p = .72$ . However, the interaction between neighbourhood density and sequence type was significant,  $F(1, 99) = 24.87$ ,  $MSE = 17.64$ ,  $p < .001$ . The interaction was driven by a dense neighbourhood advantage among pure sequences but not mixed sequences. Pairwise comparisons (two-tailed) confirmed this interpretation with a

significant effect of neighbourhood density in pure sequences,  $t(99) = 6.88$ ,  $p < 0.001$ ,  $d = 0.69$ , but not in mixed sequences,  $t(99) = 1.84$ ,  $p = .07$ ,  $d = .18$ . This general pattern of results is very similar to that found by Jalbert et al. with a significant ND effect in pure but not mixed sequences.

### 3.4.1.3 Simulation 3

Figure 9 shows the data obtained by Jalbert et al. (2011b, Experiment 3) and the data obtained by the simulation.

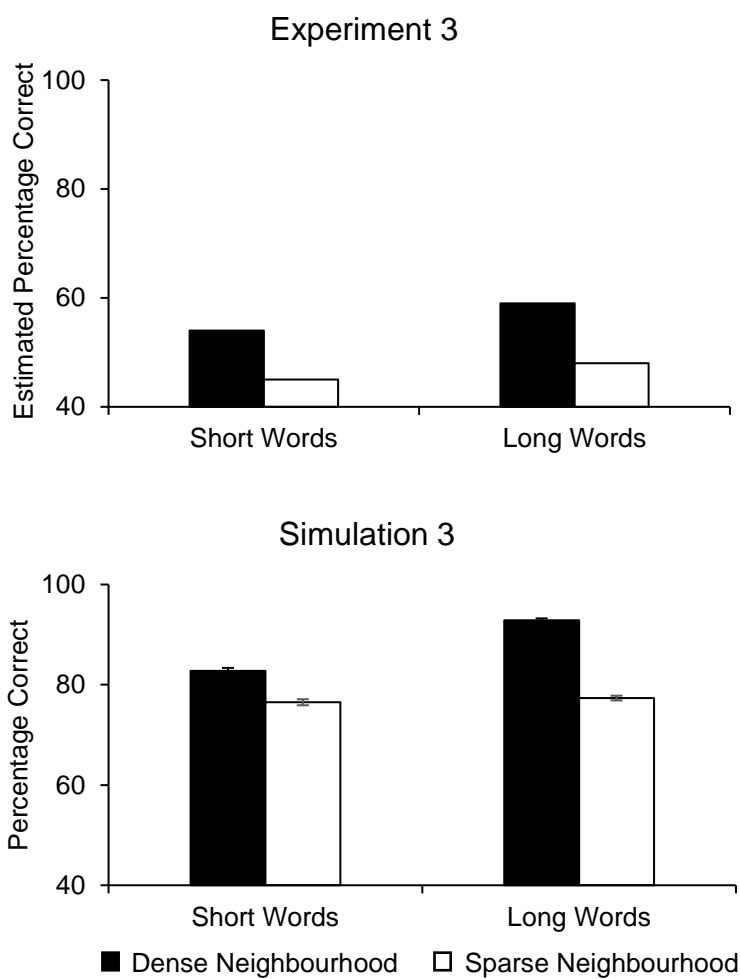


Figure 9. Mean percentage correct in each experimental condition in Jalbert et al. (2011b, Experiment 3) and Simulation 3. Error bars for the simulated data show standard errors of the means. Note. Because precise values were not provided by Jalbert et al. the means were estimated from the figure they provided in the paper.

In Jalbert et al. (2011b, Experiment 3) there was a main effect of neighbourhood density,  $F(1, 15) = 25.371$ ,  $MSE = .006$ ,  $p < .001$ ,  $n_p^2 = .628$ , with more accurate reconstruction for non-words from a dense neighbourhood, compared to non-words from a sparse neighbourhood. The main effect of length was not significant,  $F(1, 15) = 3.209$ ,  $MSE = .009$ ,  $p = .09$ ,  $n_p^2 = .389$ , although there was a general trend identified for the long non-words to be more accurately reconstructed than the short non-words. The interaction between length and neighbourhood density was not significant,  $F < 1$ . Their results suggest that only neighbourhood density had a measurable effect upon accuracy in the reconstruction task. In the simulations there was also a main effect of neighbourhood density,  $F(1, 99) = 553.5$ ,  $MSE = 21.5$ ,  $p < .001$ ,  $n_p^2 = .85$ , with more accurate reconstruction for words from a dense neighbourhood regardless of length. The main effect of length,  $F(1, 99) = 99.4$ ,  $MSE = 30.1$ ,  $p < .001$ ,  $n_p^2 = .5$ , was also significant and is reflective of a trend, also identified by Jalbert et al., for long non-words to be more accurately reconstructed than short non-words. Unlike Jalbert et al., the interaction between neighbourhood density and length was also significant,  $F(1, 99) = 78.9$ ,  $MSE = 27.2$ ,  $p < .001$ ,  $n_p^2 = .44$ . Pairwise comparisons (two-tailed) suggest that this was driven by more accurate reconstruction among the dense neighbourhood long non-words compared to dense neighbourhood short non-words,  $t(99) = 14.51$ ,  $p = .07$ ,  $d = 1.45$ , but no significant difference between reconstruction accuracy for sparse neighbourhood short non-words compared to sparse neighbourhood long non-words,  $t(99) = 1.04$ ,  $p = .3$ ,  $d = .1$ . Once again, the general pattern of results is very similar to those found by Jalbert et al., with neighbourhood density, rather than length, being the main determinant of task accuracy.

### **3.4.2 Guitard, Saint-Aubin, et al. (2018) Simulation Results**

Five simulations were used to simulate the serial reconstruction data obtained by Guitard, Saint-Aubin, et al. (2018). Figure 10 shows the results of the experiments and the simulations.

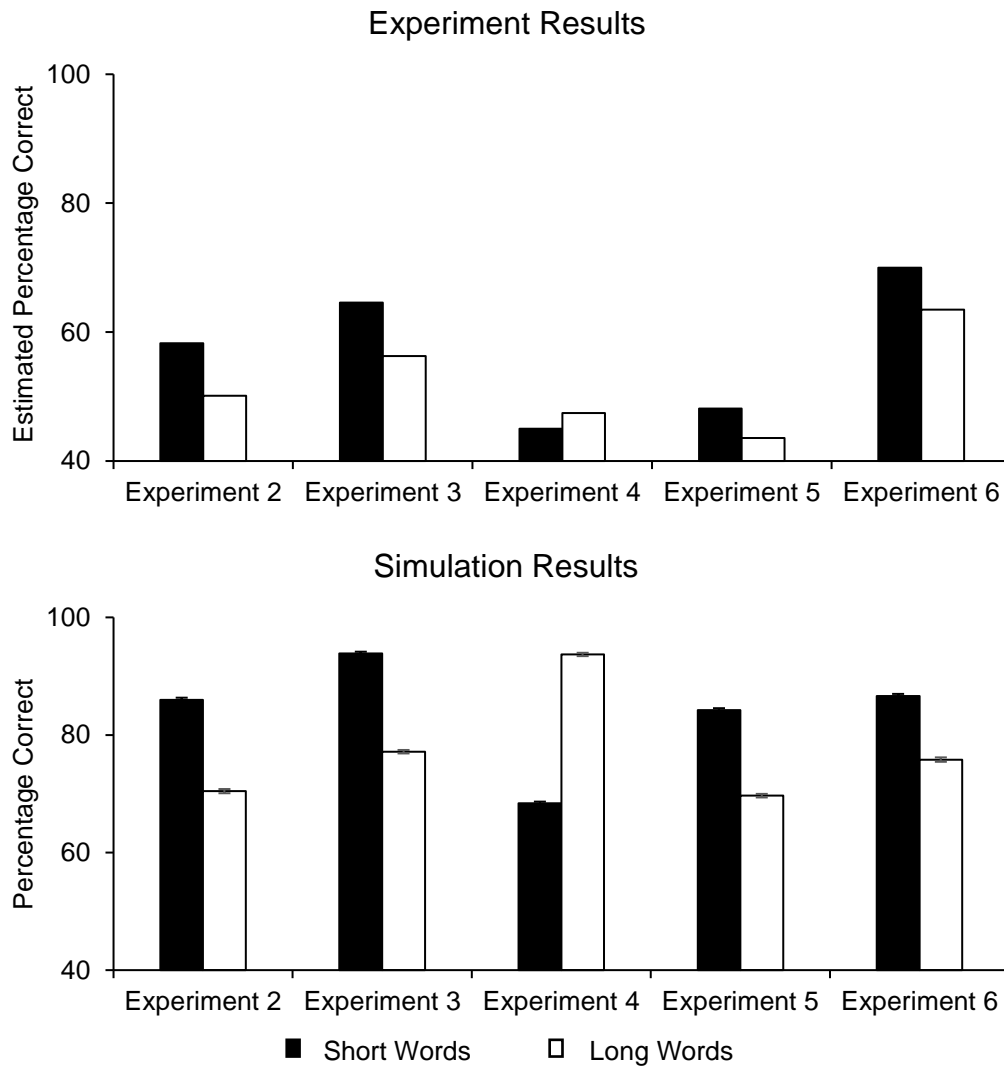


Figure 10. Mean percentage correct in each serial reconstruction experiment conducted by Guitard, Saint-Aubin, et al. (2018) and the corresponding simulation for each experiment. Error bars show standard errors of the means. Note. Because precise values were not provided by Guitard, Saint-Aubin, et al. the means were estimated from the figures they provided in the paper.

Guitard, Saint-Aubin, et al. (2018) reported a reliable word length effect in Experiment 2,  $F(1, 21) = 10.71, p = .004, n_p^2 = .34$ , Experiment 3,  $F(1, 21) = 25.54, p < .001, n_p^2 = .55$ , Experiment 5,  $F(1, 21) = 8.84, p = .007, n_p^2 = .30$ , and Experiment 6,  $F(1, 21) = 9.79, p = .005, n_p^2 = .32$ . In Experiment 4 the word length effect was reversed with better reconstruction accuracy for the long words rather than the short words,  $F(1, 21) = 4.55, p = .045, n_p^2 = .30$ . The simulations produced an almost identical pattern of data with a reliable word length effect in the simulations that corresponded

to Experiments 2, 3, 5, and 6,  $t(99) = 21.7$ ,  $p < 0.001$ ,  $d = 2.17$ ,  $t(99) = 28.2$ ,  $p < 0.001$ ,  $d = 2.82$ ,  $t(99) = 23.9$ ,  $p < 0.001$ ,  $d = 2.39$  and  $t(99) = 14.4$ ,  $p < 0.001$ ,  $d = 1.44$ , respectively. The simulation corresponding to Experiment 4, similarly to Guitard, Saint-Aubin, et al., yielded a significant reversal of the word length effect,  $t(99) = 44.5$ ,  $p < 0.001$ ,  $d = 4.45$ .

### **3.5 General Discussion**

The simulations established what the patterns of data could have been like had the participants in Jalbert et al. (2011b) and Guitard, Saint-Aubin, et al. (2018) all opted to use a first letter strategy. Across all 8 simulations there was the same general pattern of results as those found by the authors. The simulations of Jalbert et al., Experiments 1 and 2, found a significant ND effect in pure sequences but, for mixed sequences, the effect attenuated in Simulation 1 and was eliminated in Simulation 2. In Simulation 3, an effect of neighbourhood density was found independent of word length. Non-words from a dense neighbourhood were reconstructed more accurately than non-words from a sparse neighbourhood regardless of whether they were long or short. The results of the simulations are contrary to a claim made by Jalbert et al. that, had participants used a first letter strategy in their Experiment 3, there would have been an elimination of the ND effect. The simulations also produced a pattern of data very similar to that obtained by Guitard, Saint-Aubin, et al. The simulation corresponding to Guitard, Saint-Aubin, et al. (Experiment 4) produced the same advantage for long words over short words and all other simulations produced the usual advantage for short words over long words. Because all the simulated data is similar to the corresponding experimental data it raises questions whether the experimental effects were a consequence of the item-level manipulations or whether they were instead some consequence of having uneven onset letter distributions among the word pools.

There were some clear differences between the simulated data and the experimental data. Firstly, in Simulation 1, unlike Experiment 1 of Jalbert et al. (2011b), the ND effect was not entirely removed in mixed sequences. However, in the simulation, the interaction between neighbourhood density and sequence type was significant with the size of the ND effect in mixed sequences being far smaller than the size of the ND effect in pure sequences. This suggests that the onset letter

distributions among the mixed sequences still went some way towards attenuating the ND effect.

A second difference between the simulations and experimental data is that Jalbert et al. (2011b) did not find a significant effect of length in their Experiment 3. However, their data did hint towards an effect of length with a tendency for long non-words to be better remembered than short non-words. With only 16 participants in their study it was less powered than the simulation to detect smaller effect sizes. Therefore, finding a significant effect of length in the simulation, but there only being a general trend towards an effect of length in the experiment, is not particularly surprising. Additionally, the significant effect of length in the simulation still fits with the general pattern of results obtained by the authors.

A third difference is that the significant interaction between neighbourhood density and word length in Simulation 3 is not entirely consistent with the data obtained by Jalbert et al. (2011b). In the simulation it was driven by better performance for long non-words from a dense neighbourhood compared to short non-words from a dense neighbourhood. There was no strong evidence of a similar pattern emerging in their experiment. However, the simulation assumed perfect memory of onset letter order and guesses at chance level when there were duplicated onset letters. As outlined earlier, the aim of the simulations was not to provide a perfect fit for the experimental data but simply to establish whether a similar general pattern of data would emerge if a first letter strategy were to have been used. It is very clear from the far higher levels of accuracy across all the simulations that they provide a simplistic account of performance. Even if participants adopted to use a first letter strategy it seems unlikely that they would have used it in every trial. Furthermore, even if they did use it in every trial it is even more unlikely that they would have had perfect memory for every sequence of onset letters (e.g., serial recall accuracy for 6-item sequences is only around 50%; Crannell & Parrish, 1957). What the simulation results do indicate however, is that given the distribution of onset letters in the experiments, there would have been very strong and robust effects of using a first letter strategy upon the results obtained. Even if a first letter strategy were to be adopted on just a handful of trials it is likely to have had a measurable impact upon the overall percentage correct achieved in each experimental condition. A possibility is that the experimental data were driven by a combination of some participants almost exclusively adopting a first letter strategy,

some participants occasionally adopting a first letter strategy and other participants attempting to remember the items in their entirety. A combination of all those approaches would reduce the overall percentage correct in the experiments and dilute some effects (e.g., the interaction between neighbourhood density and word length found in Simulation 3) that would otherwise have been caused by only using a first letter strategy.

As touched upon in the current chapter's introduction, even if participants did attempt to remember items in their entirety, the onset letter confound still leaves the data obtained by Jalbert et al. (2011b) and Guitard, Saint-Aubin, et al. (2018) open to re-interpretations not considered by the authors. Because errors in vSTM tasks are very similar to those that arise within natural speech (e.g., Ellis, 1980), if participants were to adopt a cumulative rehearsal strategy then this could lead to item errors akin to naturally occurring speech errors (e.g., a recombination of word segments that have broken at the C/VC, CC/V or VC/C boundary; Treiman & Danis, 1988). In a serial recall task, the incorrect items would be output by the participant (e.g., spoken or written) and therefore scored as incorrect. However, in serial reconstruction it is not possible to detect if an incorrect item has been remembered by the participant. Instead, before the experiment moves onto the next trial, participants must still decide on some order to click the items even if the re-presented items do not match their memory of the originally presented items. One way to complete the trial would be to simply click any remaining items without giving any consideration to their order (i.e., complete guesses). However, another way would be to use any incorrectly remembered items to inform the order in which to click the re-presented items. One useful cue could be the onset letter. The reason for this suggestion is due to the special status that onsets seem to have in language production. For example, when required to transpose consonants that are contained in pairs of disyllabic words, participants are fastest and most accurate when transposing the onset consonants compared to later consonants (Fowler, Treiman & Gross, 1993). If any speech errors in serial reconstruction maintain the onset consonant (e.g., 'cat', 'dot' 'sit' becomes 'cot', 'dit', 'sat) then those onset consonants could perhaps be more readily available than the offsets and used to accurately reconstruct the sequence.

Most of the to-be-remembered materials in Jalbert et al. (2011b) and Guitard, Saint-Aubin, et al. (2018) began with onset consonants rather than vowels. It could therefore be suggested that whenever a participant remembered an incorrect item,

the onset consonants of those incorrect items were more readily available to participants than the later consonants (e.g., Fowler et al., 1993). This could possibly have encouraged the use of onset letters as a cue for deciding which re-presented items to click in the serial reconstruction tasks. In Jalbert et al. (Experiment 3) and Guitard, Saint-Aubin, et al. (Experiment 4) it may be the case that the long to-be-remembered items, because they are phonologically more complex than short items (e.g., Caplan et al., 1992; Service, 1998), all afforded less fluent rehearsal which resulted in more speech errors. In a task requiring reproduction of those items (e.g., serial recall, span task) there may have been more errors among the long items. However, due to the nature of the serial reconstruction task and the onset letter confound (the long word pools sustaining better vSTM also contained more unique onset letters) these errors could have been overcome if the onset letters were maintained and then used as cues to decide which items to click. Overall, the possibility raised is that irrespective of whether participants decide to adopt a first letter strategy or opt to rehearse the items in their entirety, failing to match word pools on the distribution of onset letters may, in some instances, produce results that are not actually a true reflection of what the participant has remembered.

Finally, it may be that the research by Jalbert et al. (2011b) and Guitard, Saint-Aubin, et al. (2018) were successful demonstrations of neighbourhood density or orthographic frequency impacting vSTM. The onset letter confounds, and potential issues discussed, may have had no bearing upon their results. However, at the very least, the simulations highlight a broader issue of vSTM research. While the focus is typically on understanding some underlying mechanism of performance a far less discussed issue is what exactly a vSTM task is testing and whether participants may be utilising a variety of alternative strategies to complete the task. Morrison et al. (2016) demonstrated that participants self-report a wide range of strategies for a single task. Therefore, to assume that tasks are likely to have some common underlying strategy and that the task instructions will be enough to ensure that participants will complete the task in the manner intended by the experimenter is perhaps unwise. Memory for letters is better than memory for words (e.g., Crannell & Parrish, 1957) so, even if instructed to memorise the entire words, if a first letter strategy can be used it does not seem unreasonable to suggest that participants might opt to use that strategy. In a similar vein, Macken et al. (2014) demonstrated how auditory serial recognition, rather than requiring any explicit knowledge of the



items comprising a to-be-remembered sequence, can in some instances be completed via pattern matching. This is despite serial recognition, irrespective of presentation modality, usually being considered a task whereby participants encode each item that comprises the to-be-remembered sequences into STM (e.g., Baddeley et al., 2002; Gathercole et al., 2001). As another example, in tasks such as free recall participants are encouraged to reproduce the items in any order. However, at to-be-remembered sequence lengths of up to 7 items there is a tendency to try and reproduce the words in their originally presented serial order (e.g., Ward, Tan & Grenfell-Essam, 2010). This suggests that, despite their different instructions, participants often decide to complete free recall and serial recall in similar ways. A failure to consider what kinds of strategy might be utilised by participants in order to complete vSTM tasks is possibly providing some misleading results and generating incorrect theories. While it will never be possible to fully control the strategy that participants decide to adopt, it is possible to at least minimise the risk of alternative strategies yielding significant effects by giving even more consideration to the task variables. For example, ensuring that the distribution of onset letters in each word pool are matched or using larger word pools minimises the impact that a first letter strategy is likely to have upon the results in serial reconstruction.

Overall, the simulations have highlighted that, by attempting to control for many item-level variables and using serial reconstruction tasks, the data by Jalbert et al. (2011b) and Guitard, Saint-Aubin, et al. (2018) may have been a consequence of the uneven onset letter distributions among their word pools rather than the item-level manipulations of interest. As such, until similar effects are demonstrated in other vSTM tasks or, at the very least, demonstrated in serial reconstruction when the word pools are matched for onset letter frequency, it seems best to remain cautious before using the data obtained by Jalbert et al. and Guitard, Saint-Aubin, et al. to inform our understanding of vSTM.

## Chapter 4

### **Exploring the Relationship Between Neighbourhood Density and Articulatory Difficulty: Effort-Based and Duration-Based Measures of Articulatory Difficulty**

#### **4.1 Abstract**

Performance in vSTM tasks (e.g., serial recall/reconstruction) is typically better when to-be-remembered words are from dense rather than sparse neighbourhoods. However, it is rarely considered why exactly words vary on neighbourhood density and what impact this may have upon vSTM. In the present chapter, Study 1 and Experiment 5 both explore the possibility that neighbourhood density distributions are a consequence of pressures to minimise articulatory effort during the natural development and evolution of language (e.g., Lindblom, 1990). Words from dense neighbourhoods may consist of easier (less effortful) articulations than words from sparse neighbourhoods and these differences in articulatory difficulty, rather than neighbourhood density per se, may determine how successful vSTM is. Study 1 investigated articulatory difficulty by quantifying the difficulty of the 24 English consonant sounds based upon three dimensions – their manner of articulation, their place of articulation and their voicing (e.g., St. John, 2015). Experiment 5 investigated articulatory difficulty via the time taken to vocalise dense and sparse neighbourhood words in isolation and within sequences (e.g., Woodward et al., 2008). In Study 1 and Experiment 5 there was evidence that dense neighbourhood words are easier to articulate than sparse neighbourhood words. This was indicated by a regression analysis showing that words from denser neighbourhoods tend to score lower on articulatory difficulty than words from sparser neighbourhoods (Study 1a). Additionally, word pools that have been used to investigate the ND effect also vary on articulatory difficulty with the dense neighbourhood word pools tending to consist of easier articulations (Study 1b). Experiment 5 found that dense neighbourhood words take less time to vocalise in isolation and in sequences than sparse neighbourhood words do. The findings suggest a close relationship between neighbourhood density distributions and articulatory difficulty. As such, more consideration should be given to the role of articulatory difficulty when interpreting what the ND effect reveals about vSTM.

## **4.2 Introduction to Study 1 and Experiment 5**

As already extensively discussed in the thesis, neighbourhood density impacts vSTM with recall and reconstruction typically better when to-be-remembered sequences are comprised of words from a dense rather than a sparse neighbourhood (e.g., Allen & Hulme, 2006; Clarkson et al., 2017; Derragh et al., 2017; Guitard, Gabel, et al., 2018; Jalbert et al., 2011a, b). Words with many neighbours are suggested to elicit more supportive activation within linguistic networks in LTM than words with relatively few neighbours. The more activation there is within those networks in LTM then the more LTM is able to assist with the reintegration of any degraded traces in STM (e.g., Derragh et al., 2017; Jalbert et al., 2011a, b; Roodenrys et al., 2002) or support the maintenance of item order in STM (e.g., Clarkson et al., 2017). However, Chapter 2 raised questions over these interpretations with Experiments 3 and 4 finding a serial recall and serial recognition advantage for words from a sparse rather than a dense neighbourhood. Additionally, by highlighting that similar effects to those found by Jalbert et al. (2011b) and Guitard, Saint-Aubin, et al. (2018) could be achieved had participants used the first letters of to-be-remembered words as cues in which to reconstruct their order, Chapter 3 raised doubts over the conclusion offered by Jalbert et al. that the ND effect is unlikely to be some consequence of articulatory rehearsal. These findings in the thesis suggest that further consideration of what could be causing the ND effect is warranted. What has received little consideration in the vSTM literature so far is why exactly some words have more neighbours than others. This is a potentially important oversight because the reason for the neighbourhood density distributions, rather than the distributions themselves, may be a key determinant of the ND effect.

According to Lindblom (1990) language is shaped by pressures to minimise effort in production while preserving discriminability in perception. The consequence of these pressures is that easier to articulate sounds are likely to be the most common sounds within a language. Sounds that are harder to articulate will be far less common. For example, speech sounds such as voiced sibilant affricates (e.g., /dʒ/ /dʒ/) that require high air pressure to articulate tend to be avoided in languages and are entirely absent in English (e.g., Zygis, Fuchs & Koenig, 2012). In respect to neighbourhood density it is possible that, in order to minimise articulatory effort, the natural development and evolution of language has resulted in denser clusters of words around easier articulatory configurations and sparser clusters of words around

more difficult articulatory configurations. The denser the clustering then the more likely that the words within a particular cluster will share a sufficient number of sounds and letters to render them phonological or orthographic neighbours.

If dense neighbourhood words generally tend to be comprised of less effortful articulations, then it is possible that some examples of the ND effect, considered to be a consequence of varying levels of support in LTM, are actually some consequence of variations in articulatory effort. The reason for this suggestion is that variations in articulatory effort, rather than item-level manipulations, have successfully been demonstrated elsewhere to impact vSTM. For example, when to-be-remembered words have small changes in the place of articulation between items they are remembered better than words with larger changes in place of articulation (e.g., Murray & Jones, 2002, see Section 5.2 for more detail). The easier (more fluently) that a word or sequence of words can be articulated then the better vSTM is. If dense neighbourhood words generally require less articulatory effort than sparse neighbourhood words then differences in articulatory effort, rather than neighbourhood density per se, may be determining the relative success of vSTM.

Research has highlighted how short words from a sparse neighbourhood tend to consist of fairly atypical structures (Guitard, Saint-Aubin, et al., 2018) and sparse neighbourhood words tend to have lower phonotactic probabilities (a measure of how often particular phonological segments appear in a given language) than dense neighbourhood words (e.g., Storkel, 2004; Vitevitch, Luce, Pisoni, & Auer, 1999). If effort minimisation is crucial as languages develop and evolve (e.g., Lindblom, 1990), then the most frequently used sounds should also be the easiest to articulate. However, it is not possible to claim with any certainty that those more frequently used speech sounds are any easier to articulate than the less frequently used sounds. To overcome this issue, it is important to first quantify what exactly is meant by articulatory difficulty. The current chapter adopts two different approaches for quantifying articulatory difficulty. Firstly, in Study 1a, articulatory difficulty is quantified by scoring the 24 English consonant sounds on the relative difficulty required to articulate them based upon three dimensions – their manner of articulation, their place of articulation and their voicing (e.g., St. John, 2015). The scores are then applied to the onset and offset of English CVC words to establish whether any relationship exists between a word's neighbourhood density and its assigned difficulty scores. In Study 1b this scoring system is used to establish

whether word pools that have been used to investigate the ND effect also vary in articulatory difficulty. In Experiment 5, spoken duration measures are used to indicate the relative ease in which dense and sparse neighbourhood words can be vocalised (e.g., Woodward et al., 2008). Sequences tend to be vocalised fastest when they are comprised of less effortful articulations. Each method for quantifying difficulty is outlined in detail in the corresponding experiment introduction.

#### **4.3 Study 1 - Articulatory difficulty as measured by an effort-based scoring system**

It is possible that some examples of the ND effect have been a consequence of variations in articulatory difficulty. Dense neighbourhood words contain more common phonological segments (e.g., Storkel, 2004; Vitevitch et al., 1999) than sparse neighbourhood words and according to Lindblom (1990) the most common sounds should also be the easiest to articulate. However, to sustain the suggestion that the ND effect is really some consequence of variations in articulatory difficulty it is first important to quantify what exactly constitutes articulatory difficulty and then demonstrate that, not only do denser neighbourhood words tend to consist of more difficult articulations but also that word pools manipulating neighbourhood density have also varied in articulatory difficulty.

In the present experiment, articulatory difficulty was quantified by scoring the 24 English consonant sounds on their relative difficulty according to three dimensions – their manner of articulation, their place of articulation and their voicing (e.g., St. John, 2015). Each dimension is described below:

**Dimension 1 - Manner of Articulation.** The articulation of stop consonants demands less articulatory precision than the articulation of fricative consonants. This is because the vocalisation of stop consonants requires complete blockage of airflow through the oral cavity. Fairly coarse movements can be made to achieve the blockage (e.g., Ladefoged & Maddieson, 1996). Fricatives require far more precise movements of the active articulators. For example, the tongue tip must be held in place to achieve the /s/ speech sound (e.g., Stevens, 1971). If the tongue tip were to overshoot its target, then full blockage of airflow would prevent correct vocalisation of /s/. Affricates begin with a stop sound and end with fricative so could be considered to require equally precise tongue movements to fricatives and therefore of similar difficulty. Nasal, and approximants could be considered to require a degree of

articulatory precision that lies somewhere between stop consonants and affricates/fricatives. They do not quite require the precision of fricatives but cannot be achieved with the same coarse movements of stop consonants. Some support for these patterns of difficulty comes from developmental language data. Stop consonants are the first to emerge during language development and are all acquired by around 3 years of age (McLeod & Crowe, 2018). They are also often used instead of fricatives during early language development (e.g., 'sad' pronounced as 'tad'; Oller, Wieman, Doyle & Ross, 1975) suggesting that they are the easiest consonants to articulate. Nasal and approximant consonants begin to emerge at 1-2 years and 1 to 3 years of age respectively with the slightly later development suggesting that they are more difficult to articulate than stop consonants. Finally, affricates emerge between 3-4 years of age but /tʃ/ and /dʒ/ may not be fully mastered until around 6 years of age and fricatives emerge between 1-4 years, but some (e.g., /θ/) do not begin to emerge until around 6 years of age (McLeod & Crowe, 2018). Because affricates and fricatives are the final consonant sounds to be mastered it suggests that they are also the most difficult to articulate.

For the reasons outlined above, stop consonants were categorised as the least difficult to articulate and received a difficulty score of 0. Nasals and approximants were categorised as more difficult receiving a difficulty score of 1 and affricates and fricatives were categorised as the most difficult and received a difficulty score of 2.

**Dimension 2 - Voicing.** In order to produce a voiceless consonant sound the vocal folds must become elongated and highly tensed (Zemlin, 1998) and the glottis must be abducted in order to occlude airflow (Kirchner, 1998). This requires a higher degree of transglottal muscular tension than is involved with the production of voiced consonants. Children tend to replace voiceless consonants with voiced consonants (e.g., Oller et al., 1975). As such, voiced consonants could be considered easier to articulate than voiceless consonants. The current scoring system categorised voiced consonants as easier than voiceless consonants with voiced consonants receiving a difficulty score of 0 and voiceless consonants receiving a difficulty score of 2.

**Dimension 3 - Place of Articulation.** The lower lip, tongue tip and tongue body (active articulators) are all attached to the lower jaw. The jaw can be used to assist with articulations by moving the active articulators towards the upper lip, velum and alveolar ridge (passive articulators). However, because the jaw operates by

rotating around a posterior pivot it can assist with anterior articulations more efficiently than posterior articulations (e.g., Mooshammer, Poole & Geumann, 2007). As articulation moves further back in the mouth more rotational movement from the jaw is required which can be considered more effortful. Some developmental support for this claim is that labial consonants are typically acquired earliest (Edwards & Shriberg, 1983; McLeod & Crowe, 2018) and children often replace coronal and dorsal onset consonants with labials (Oller et al., 1975). Additional evidence for dorsal sounds being more effortful comes from the finding that across a variety of languages they tend to be the least frequent sounds produced in babbling (e.g., Kern et al., 2014). The current scoring system categorised labial consonants as the easiest to articulate and gave them a difficulty score of 0. Coronal consonants were rated as more difficult and assigned a difficulty score of 1 and dorsal consonants were rated as the most difficult and assigned a difficulty score of 2.

One previous attempt to quantify the articulatory difficulty of dense and sparse neighbourhood words was conducted by St. John (2015) using similar dimensions to those outlined above. A very small ( $R^2 = 0.03$ ), but significant, correlation was found between neighbourhood density and articulatory difficulty with words from sparse neighbourhoods having slightly higher difficulty scores than words from dense neighbourhoods. There was also some indication that the experimental word pools of Roodenrys et al. (2002), Allen and Hulme (2006) and Clarkson et al. (2017) were confounded by articulatory difficulty with the sparse neighbourhood words having slightly higher difficulty ratings. However, a critical issue with the work by St. John is the very limited application the results have. Of the 24 English single consonant sounds only 8 were given a difficulty score. Additionally, the scores were only applied to the onset consonants which means that articulations located later in the word, whether easy or difficult, were entirely ignored. It is possible that any relationship between articulatory difficulty and neighbourhood density will disappear when all 24 consonant sounds are used or when the difficulty of the entire word is considered. The analysis of the experimental word pools was also very limited as it could only include a limited selection of the words contained within the word pool (i.e., those beginning with the 8 consonant sounds given a difficulty score). This may have provided misleading results because a subset of words from any given word pool is not representative of the entire word pool. Analysis of the remaining words could possibly reveal that there is no difference in articulatory difficulty between the

dense and sparse neighbourhood words or even reveal that the dense neighbourhood words contain more difficult articulations than the sparse neighbourhood words.

The current investigation overcame the outlined limitations and provided a significant advance upon the research conducted by St. John (2015) by expanding the articulatory difficulty scoring system so that it accommodates all the 24 English single consonant sounds and then applies those scores to both the onset and offset of CVC words (Study 1a). Study 1b then applies difficulty scores to all the items contained within a selection of dense and sparse neighbourhood CVC word pools that have been used to demonstrate the ND effect to establish whether those word pools are confounded by articulatory difficulty.

### **4.3.1 Study 1a Method**

#### **4.3.1.1 Procedure**

Each of the 24 consonants were assigned three difficulty scores (see Table 1) that correspond to each of the three dimensions - Manner of Articulation, Voicing and Place of Articulation - outlined in the experiment introduction.

The MRC database was then searched for English CVC words. This yielded an initial pool of 1653 words. The Linguistics Program N-Watch (Davis, 2005) was used to calculate phonological neighbourhood density values for each word. Phonological neighbourhood density values were obtained for 1218 words. For each of the 1218 words a difficulty score for each dimension was assigned to the onset consonant and the offset consonant. The sum of the two scores in each dimension provided an articulatory difficulty score for that word in that dimension. This gave each word three final overall difficulty scores corresponding to each of the three dimensions.



Table 1

*Articulatory difficulty scores assigned to each consonant (C) in each dimension*

C	Manner of Articulation					Voicing		Place of Articulation		
	Stop	Affricate	Approximant	Nasal	Fricative	Voiced	Voiceless	Labial	Coronal	Dorsal
b	0					0		0		
tʃ		2					2		1	
d	0					0			1	
f					2		2	0		
g	0					0				2
h			1				2			2
dʒ		2				0			1	
k	0						2			2
l			1			0			1	
m				1		0		0		
n				1		0			1	
ŋ				1		0				2
p	0						2	0		
r			1			0			1	
s					2		2		1	
ʃ					2		2		1	
t	0						2		1	
θ					2		2		1	
ð					2	0			1	
v					2	0		0		
w			1			0		0		
j			1			0			1	
z					2	0			1	
ʒ					2	0			1	

*Note.* '0' indicates a least difficult feature, '1' indicates a difficult feature and '2' indicates a most difficult feature.

#### **4.3.1.2 Results and Discussion**

A multiple linear regression was performed to investigate the relationship between neighbourhood density and the three difficulty scores that were calculated for each word based upon each of the three dimensions. The articulatory difficulty scores predicted a significant amount of variance in phonological neighbourhood density distributions,  $F(3, 1214) = 90.193$ ,  $p < .001$ ,  $R^2 = .18$ . The relationships between manner of articulation and neighbourhood density ( $p < .001$ ) and voicing and neighbourhood density ( $p < .001$ ) were significant. However, the relationship between place of articulation and neighbourhood density was not significant ( $p = .48$ ).

The results of Study 1a provide some further evidence to St. John (2015) that neighbourhood density and articulatory difficulty are closely related. CVC words from sparser neighbourhoods tend to consist of more difficult articulations than CVC

words from denser neighbourhoods. These findings are important because articulatory difficulty is considered to impact vSTM with to-be-remembered sequences consisting of less effortful articulations sustaining better vSTM than sequences containing more effortful articulations (e.g., Murray & Jones, 2002; Woodward et al., 2008). As such, the current findings suggest that manipulating neighbourhood density may impact vSTM because denser neighbourhood words tend to consist of easier articulations. However, despite the general trend identified in Study 1a it is entirely possible that dense and sparse neighbourhood words selected by researchers do not vary on articulatory difficulty. This is because there will be some words from a sparse neighbourhood comprised of relatively easy articulations and some words from a dense neighbourhood comprised of more difficult articulations. Study 1b explores whether the specific word pools used by researchers are also confounded by articulatory difficulty.

One issue with the current scoring system that first needs to be addressed is that the place of articulation dimension was not significant. One possible explanation for this is that the majority (> 66%) of the CVC words used in the analysis contained onset and offset articulations that were located in different areas in the mouth (e.g., labial onset and coronal offset or vice versa). As such, because the place of articulation score for each word was calculated by adding the score at the onset and offset this may have cancelled out the effectiveness of the measure in the current experiment. For example, if the earlier analysis were to be repeated but with place of articulation split into separate scores for onset and offset consonants then both appear to be significant ( $p < .05$ ) measures. However, place of articulation for onset consonants is negatively correlated with neighbourhood density ( $\beta = -.61$ ) with difficulty decreasing as neighbourhood density increases. Place of articulation for offset consonants is positively correlated with neighbourhood density ( $\beta = 1.21$ ) with difficulty increasing as neighbourhood density increases. This suggests that combining both onset and offset measures simply serves to cancel the other out. This issue further highlights the importance of considering the articulatory difficulty of the entire word rather than just the onset. In St. John's (2015) scoring system only the onset consonant was scored and had that also been the case in the current system then difficulty scores associated with place of articulation would have seemingly predicted neighbourhood density distributions in the expected direction. However, this would have failed to take articulatory difficulty located later in the word

into account. Because of the tendency for consonants in CVC words to have two different places of articulation it raises doubts over the usefulness of including place of articulation, in its current form, in the scoring system. Because of this, the place of articulation measure was dropped in Study 1b.

#### **4.3.2 Study 1b Method**

Study 1a demonstrated that denser neighbourhood CVC words tend to consist of easier articulations than sparser neighbourhood CVC words. However, just because there is a general trend that does not mean that experimental word pools varying on neighbourhood density also vary on articulatory difficulty. To establish whether this is also the case, and whether some examples of the ND effect may be a consequence of variations in articulatory difficulty, experimental word pools containing CVC words that have been used to demonstrate an ND effect were selected (see Table 2). To establish whether those previously used word pools are confounded by articulatory difficulty an overall articulatory difficulty score was calculated for each word in each of the selected experimental word pools. This was achieved by totalling the manner of articulation and voicing scores for the onset and offset consonant of each word. This gave every word contained within each of the dense and sparse neighbourhood word pools a single overall difficulty score. For each of the selected experiments the overall difficulty scores for dense neighbourhood words and sparse neighbourhood words could then be used to allow a statistical comparison between the word pools (e.g., St. John, 2015).

##### **4.3.2.1 Results and Discussion**

For each dense and sparse neighbourhood CVC word pool used to demonstrate the ND effect, a one-tailed independent measures *t*-test was conducted to test the prediction that the words in the sparse neighbourhood word pools will have larger articulatory difficulty scores than the words in the dense neighbourhood word pools. Table 2 shows the mean articulatory difficulty scores and results of the *t*-tests.

Table 2

Mean articulatory difficulty score and *t*-test outcomes for pools of dense and sparse neighbourhood CVC words used in previous vSTM experiments.

Experiment	Mean Articulatory Difficulty Score (and SD)		<i>t</i> -tests (one-tailed) and effect sizes (Cohen's <i>d</i> )
	Dense Neighbourhood	Sparse Neighbourhood	
Roodenrys et al. (2002, Experiment 1) and Allen and Hulme (2006, Experiment 2)	3.22 (1.93)	4.75 (2.06)	$t(62) = 3.07, p = .002, d = 0.77$
Roodenrys et al. (2002, Experiment 3)	2.97 (1.77)	3.84 (2.08)	$t(62) = 1.81, p = .04, d = 0.45$
Jalbert et al. (2011a, Experiment 2) and Jalbert et al. (2011b, Experiment 1)	2.56 (1.71)	3.38 (2.22)	$t(30) = 1.16, p = .13, d = 0.41$
Clarkson et al. (2017, Experiment 2 and 3)	3.28 (1.61)	4.17 (2.11)	$t(92) = 2.31, p = .01, d = 0.48$

Across all four statistical tests there was a general trend for the sparse neighbourhood word pools to have higher articulatory difficulty scores than the dense neighbourhood word pools. In all cases the effect size (as measured by Cohen's *d*) was at least  $d > 0.4$  and the *t*-tests yielded a significant effect for Roodenrys et al. (2002, Experiment 1 and 3), Allen and Hulme (2006, Experiment 2) and Clarkson et al. (2017). Only the *t*-test for the word pool used by Jalbert et al. (2011a, b) failed to reach significance. However, considering these were the smallest word pools (16 items per pool) to be tested this could have been a consequence of statistical power (power analysis indicated that there would only have been a 53% chance of finding a significant effect given the effect size and sample size).

The results of Study 1b suggest that word pools that have been manipulated on neighbourhood density also vary on articulatory difficulty. The experiment also overcame the limitations of St. John (2015) whereby assigning difficulty scores to just 8 consonants meant that only a few items comprising the word pools were scored on articulatory difficulty. In combination with St. John the results of the current experiment add further weight to the possibility that these examples of the ND effect may be some consequence of the to-be-remembered material's articulatory difficulty,

rather than some consequence of dense neighbourhood words eliciting more supportive activation in LTM. For example, dense neighbourhood words, due to them having easier articulations, may afford more fluent segmental recoding which would likely lead to better vSTM task performance (e.g., Jones & Macken, 2018; Macken, Taylor & Jones, 2015). However, spoken communication does not usually involve a single word spoken in isolation and the words preceding and following can alter the ways in which words are articulated. This can occur through the process of lenition which refers to the reduction of vowel and consonant sounds in order to reduce articulatory effort (e.g., Kirchner, 1998; Honeybone, 2008). For example, the two words 'don't know' may not always be fully articulated in fluent speech with the /t/ being dropped and becoming 'dunno'. Therefore, the articulatory effort required to produce a word in isolation (e.g., 'don't') could be considered different to the articulatory effort required to produce that same word within a sequence. This means that a difficulty score for a word produced in isolation may no longer provide an accurate measure of that word's difficulty when it is produced within a sequence. This is a particularly important issue if suggesting that the ND effect may be some consequence of articulatory difficulty. This is because to-be-remembered items in a vSTM task are almost exclusively presented as a sequence of items and participants are often required to reproduce those items (e.g., via spoken production). Rehearsal strategies in vSTM tasks typically involve rehearsal of the entire sequence (often in a cumulative fashion) rather than rehearsing individual words in isolation (e.g., Tan & Ward, 2008).

A limitation of Study 1 is that it only investigated articulatory difficulty for words spoken in isolation. Any differences in articulatory difficulty at the lexical level could possibly attenuate if the articulations for sparse neighbourhood words become less effortful when spoken within sequences. Experiment 5 addresses this concern by providing time-based measures of the articulatory difficulty of dense and sparse neighbourhood CVC words spoken in isolation and within sequences.

#### **4.4 Experiment 5 - Articulatory difficulty as measured by spoken duration**

The scoring system in Study 1 provides a useful way to establish if certain CVC words are easy or difficult to articulate and whether there is a relationship between articulatory difficulty and neighbourhood density. However, it may not provide entirely accurate difficulty ratings for those same words spoken within a sequence. To return to the previous example of 'don't know', in Study 1b, the consonants /d/, /t/ and /n/ received a difficulty score of 0, 2 and 1 respectively (manner of articulation and voicing scores combined). If these scores were to be applied to the /d/ and /t/ in 'don't' and the onset /n/ in 'know' it would give the word 'don't' a difficulty score of 2 and the onset of 'know' a difficulty score of 1. However, as discussed, the word 'don't' is prone to lenition when spoken within the sequence 'don't know' and can become 'dunno'. If this lenition process were to occur, it means that the articulatory difficulty score for /t/ should no longer apply as it is no longer being fully realised in speech.

One way to overcome the limitations of the scoring system developed in Study 1 is to measure the spoken duration of items spoken in isolation and within sequences (e.g., Woodward et al., 2008). If a measurement of the time taken to vocalise 'don't know' and 'dunno' were taken it would likely reveal that 'dunno' can be vocalised faster than 'don't know'. In that respect, because of the faster spoken duration, 'dunno' could be considered more fluent, or easier, to articulate than 'don't know'. Woodward et al. (Experiment 1) used this logic to explore the lexicality effect. This had previously been suggested to be an effect independent of articulatory rehearsal because single words or word pairs are spoken at the same rate as single non-words or non-word pairs (e.g., Thorn & Gathercole, 2001). However, when the spoken duration of longer sequences (e.g., 6 items) of words and non-words was measured differences began to appear. The items contained within sequences of words were spoken faster than the items contained within sequences of non-words even when the individual items comprising both those sequences were spoken at the same rate in isolation. This is because, when vocalising sequences, not only are articulations for each item required, but the vocal apparatus must also prepare for articulation of the subsequent item (co-articulations). Measuring the length of time to produce a 6-item sequence can capture both the fluency of articulations required at the item-level and fluency of co-articulations at the sequence-level. Only measuring

the duration of individual items or item pairs can fail to capture differences that may only emerge at the sequence-level.

Woodward et al. (2008) also found that giving participants the opportunity to practice the articulations and co-articulations associated with words and non-words (by administering a serial recall test for those words) significantly reduced (by 33ms) the time taken to vocalise each item within a sequence of non-words. However, there was less benefit for the items contained in word sequences (reduction of only 14ms per item). If articulations/co-articulations are already being produced fluently (i.e., in the case of word sequences) then practice is not thought to provide much additional benefit. However, if the articulations/co-articulations are relatively difficult (i.e., in the case of non-word sequences) then there is more scope for improvement.

Interestingly, there was no benefit of practice upon the time taken to articulate words and non-words in isolation. Words took 511ms and non-words 512ms before familiarisation and post familiarisation the durations were similar at 508ms for both words and non-words. This suggests that practice only benefits articulatory fluency for words vs non-words at the sequence-level rather than the item-level. It also highlights how only measuring spoken durations for single items can fail to capture differences that exist in articulatory difficulty at the sequence-level.

Some indirect evidence for supposing that dense neighbourhood words will be vocalised more fluently than sparse neighbourhood words are that firstly, dense neighbourhood words are less prone to speech errors than sparse neighbourhood words (Vitevich, 2002). Secondly, in an analysis of conversational speech, Gahl et al. (2012) found that dense neighbourhood words are more prone to lenition than sparse neighbourhood words, with the vowel sounds comprising dense neighbourhood words tending to become more centralised. More direct evidence is the finding that dense neighbourhood words have a faster speech rate (in words per second) than sparse neighbourhood words (Roodenrys et al., 2002). However, the speech rate was calculated by asking participants to recite 8 pairs of dense and 8 pairs of sparse neighbourhood words. The mean of those eight times was then transformed into a measure of speech rate in words per second for dense and sparse words. While it may be suggestive that dense neighbourhood words are articulated more fluently it only provides an estimation of speech rate and because only the time taken to vocalise item-pairs was taken it does not establish whether any differences exist at the sequence-level (e.g., Woodward et al., 2008).

The only attempt to establish whether dense neighbourhood words are articulated more fluently than sparse neighbourhood words when spoken in isolation and within sequences was undertaken by Jalbert et al. (2011b). Participants vocalised ten sequences of six items drawn from each of the dense and sparse neighbourhood word pools used in their three experiments. The time taken to vocalise all the individual items was also measured. In these experiments there was no difference found in the time taken to vocalise dense and sparse neighbourhood items or sequences. This is suggestive that there is perhaps no difference in articulatory difficulty between dense and sparse neighbourhood word sequences. However, as highlighted in Chapter 3, their experimental word pools were confounded by onset letter frequency with all the word pools that sustained better vSTM also being comprised of more unique onset letters. If the onset letters were used as cues in which to inform the order in which to reconstruct the items, then the precise relationship between the spoken measures and subsequent memory for those particular items is unclear. Additionally, the word pools were fairly small (16 items) so may not be entirely representative of dense and sparse neighbourhood words more generally.

To test the possibility that dense neighbourhood words are easier to articulate than sparse neighbourhood words the current experiment measured the time taken to vocalise the 48 dense and 48 sparse neighbourhood words that were used in Experiments 1 and 2 (Chapter 2) in isolation and within 6-item sequences (e.g., Woodward et al., 2008). These words were chosen because they have previously revealed an ND effect in serial recall (Experiment 1) and serial reconstruction (Clarkson et al., 2017). As such, the ND effect in those instances is not likely to have been some consequence of the task and materials whereby a first letter strategy could possibly have produced the same outcome (e.g., see Chapter 3). Secondly, it goes some way in addressing the concern that the smaller word pools used by Jalbert et al. (2011b) were non-representative.

In the present experiment the spoken durations for items and sequences were taken before and after a familiarisation stage whereby participants practiced vocalising sequences of dense and sparse neighbourhood words. This meant that the experiment provided two measures of articulatory difficulty (e.g., Woodward et al., 2008). Firstly, at the item-level, if dense neighbourhood words contain easier articulations than sparse neighbourhood words, then the average spoken duration of



dense neighbourhood words is likely to be shorter than those for sparse neighbourhood words. Secondly, at the sequence-level, if the co-articulations required for dense neighbourhood word sequences are easier than those required for sparse neighbourhood word sequences then, within sequences, dense neighbourhood words are likely to be vocalised faster than sparse neighbourhood words. Additionally, if sparse neighbourhood words contain more difficult articulations/co-articulations than dense neighbourhood words then these may also benefit most from familiarisation. This might be because the articulations and co-articulations for dense neighbourhood sequences are already produced fairly fluently whereas sparse neighbourhood sequences may be more disfluent and benefit from practice. Post familiarisation, this would result in a bigger reduction in spoken durations for sparse neighbourhood words. However, if there is no difference in articulatory difficulty between dense and sparse neighbourhood words at the item or sequence-level then there will be no difference in duration measures for words spoken in isolation or in sequences. There is also unlikely to be any differential benefits of familiarisation because irrespective of neighbourhood density the items and sequences are all being produced fairly fluently.

As an additional measure of difficulty to those already outlined, speech onset latencies (the delay from initial presentation of a word/sequence to the spoken onset of the first word) were also collected in the current experiment. Fluent vocalisation of items and sequences is considered to be underpinned by the planning of the required articulations (e.g., Sternberg, Monsell, Knoll, & Wright, 1978). However, as sequence length increases, or sequences become more difficult (e.g., days of week in correct order vs random order) then speech latency times increase (Sternberg et al., 1978). If an increase in speech latency times are considered to reflect longer speech planning then it may be the case that sparse neighbourhood sequences, if they are comprised of more difficult articulations, will also take longer to plan and then vocalise than dense neighbourhood sequences.

#### **4.4.1 Method**

##### **4.4.1.1 Participants**

10 participants (mean age 24 years, 7 female and 3 male) were recruited from the Cardiff School of Psychology participation panel and awarded £10 for their participation. Participants were required to have normal/corrected vision and hearing. All stages of the experiment were conducted in accordance with the Cardiff School of Psychology ethics procedures.

##### **4.4.1.2 Materials**

The 48 dense and 48 sparse neighbourhood words used in Experiments 1 and 2 (Chapter 2) were also used in this experiment (See Appendix A). For each participant, dense and sparse neighbourhood sequences were constructed by selecting 6 items at random, without replacement, from the appropriate word pool until the pool was depleted. Once depleted, all items were returned to the pools and a further set of 6-item sequences were constructed. This process was repeated until a total of 24 sparse neighbourhood and 24 dense neighbourhood sequences were constructed.

##### **4.4.1.3 Procedure**

The experiment took place in a sound attenuated booth with all stimuli presented visually on a computer screen in Arial, 40-point font. At all stages of the experiment the items and sequences were presented in a random order without replacement. There were 6 distinct stages in the experimental procedure that are outlined below. At the beginning of each stage participants were presented with instructions explaining what they would be required to do in the upcoming stage. Participants were required to acknowledge that they had read and understood them (via a button press) before proceeding.

**Stage 1 - Pre-test.** All 96 words were presented individually in the centre of the screen and participants were required to read each item slowly and clearly so that the experimenter could check for correct pronunciation of each word. Any issues with pronunciation could be corrected at this stage. This stage was implemented to ensure that any later findings were not due to participants not knowing how certain

words should be pronounced. Once the word had been spoken, the participant pressed the spacebar to proceed to the next trial. This stage was completed once all 96 words had been presented and read out loud.

**Stage 2 - Baseline 1 (Items).** The procedure was similar in structure to the pre-test stage but once all items had been presented, they were returned to the word pools and then immediately re-presented. This meant that the participant was presented with each word twice during this stage, giving a total of 192 trials. Participants were instructed to read each presented word aloud as quickly and accurately as possible. The spoken responses were recorded at a sample rate of 44.1 kHz/16-bit using a condenser microphone. Participants indicated that they had finished vocalisation of an item, and initiated presentation of the next word, by pressing the spacebar.

**Stage 3 - Baseline 2 (Sequences).** Participants were presented with 24, 6-item sequences of dense neighbourhood words and 24, 6-item sequences of sparse neighbourhood words. All items in a sequence appeared simultaneously on screen in a horizontal presentation format. The first word appeared at the central left of the screen and each subsequent word was placed to the right of the preceding word. The final word in the sequence appeared at the central right of the screen. Upon presentation of the sequence, participants were instructed to read the sequence aloud from left to right as quickly and accurately as possible. The spoken responses were recorded via microphone. Participants indicated that they had finished vocalisation of a sequence, and initiated presentation of the next sequence, by pressing the spacebar.

**Stage 4 - Familiarisation Stage.** Participants were re-presented with each of the 48 sequences they had seen in Baseline 2. Participants were required to read each presented sequence aloud 3 times as quickly and accurately as possible. This stage ensured that participants were becoming equally practised with the articulations and co-articulations required for all the dense and sparse neighbourhood sequences. Participants indicated that they had finished 3 vocalisations of the sequence, and initiated presentation of the next sequence, by pressing the spacebar. To ensure that the instructions were followed accurately the spoken responses were recorded via microphone.

**Post Familiarisation Stage 1 (Sequences).** The procedure was identical to Baseline 2 and immediately followed the Familiarisation stage.

**Post Familiarisation Stage 2 (Items).** The procedure was identical to Baseline 1 and immediately followed Post Familiarisation Stage 1.

#### 4.4.2 Results.

##### 4.4.2.1 Spoken durations (pre- and post-familiarisation)

For each participant the item and sequence recordings were saved and labelled. The durations of spoken items and sequences were then hand-measured using Audacity® (version 2.0.1, 2012) software. The software allows visualisation of recorded waveforms and enables precise duration measurements by placing start and end points onto the waveform. The placed start and end points are then used by the software to provide a duration measurement in milliseconds. To further ensure the precision of duration measurements the waveform can also be replayed with the identified start and end points acting as the beginning and end of the replay. This allows a judgement of whether a vocalisation has been fully captured. A caveat of this methodology is that measurements were not taken while blind to the experimental conditions. However, great care was taken to ensure accuracy with each item/sequence replayed several times. For item recordings, a measure of each item's duration was taken. As each item was spoken twice by each participant in both the pre and post familiarisation stage a single measure for each item, in each stage, was calculated by averaging the duration of the two items spoken in that stage. For sequence recordings, the duration of the entire spoken sequence was measured and then divided by six (the number of items in the sequence) to calculate an average spoken duration per item in that sequence. Table 3 shows the mean spoken duration of dense and sparse neighbourhood words spoken in isolation and in sequences, before (Baseline) and after (Post) familiarisation.

Table 3

*Mean spoken duration of items spoken in isolation and within sequences before (Baseline) and after (Post) familiarisation.*

Item	Isolation (ms)		Sequences (ms)	
	Baseline	Post	Baseline	Post
	M (and SD)	M (and SD)	M (and SD)	M (and SD)
Dense	461 (110)	465 (113)	512 (110)	474 (116)
Sparse	492 (109)	496 (112)	525 (110)	487 (113)

A 2x2x2 within-subjects ANOVA with stage (baseline, post-familiarisation), material type (items, sequences) and neighbourhood density (dense, sparse) was performed to establish whether there was any effect of neighbourhood density upon the duration of items spoken in isolation and within sequences before and after familiarisation. The main effect of neighbourhood density was significant with words from a sparse neighbourhood having longer spoken durations than words from a dense neighbourhood,  $F(1, 9) = 71.02$ ,  $MSE = 139.1$ ,  $p < .001$ ,  $n_p^2 = .89$ . The main effect of stage was significant,  $F(1, 9) = 10.06$ ,  $MSE = 585.18$ ,  $p = .01$ ,  $n_p^2 = .53$ , but the main effect of material type was not significant  $F(1, 9) = 0.86$ ,  $MSE = 10067$ ,  $p = .38$ ,  $n_p^2 = .09$ . However, the interaction between stage and material type was significant,  $F(1, 9) = 4.96$ ,  $MSE = 1794$ ,  $p = .05$ ,  $n_p^2 = .36$ . The duration of items spoken in isolation remained similar before and after familiarisation whereas the duration of items contained within sequences was shorter post familiarisation. The interaction between neighbourhood density and material type was also significant,  $F(1, 9) = 15.86$ ,  $MSE = 96.47$ ,  $p < .01$ ,  $n_p^2 = .64$ . There was a tendency for dense neighbourhood words to be spoken faster in isolation, but this advantage reduced when spoken within sequences. Exploratory analysis of this interaction revealed that, while there was a clear reduction in the effect size, dense neighbourhood words were still spoken faster than sparse neighbourhood words when contained in sequences,  $t(19) = 4.54$ ,  $p < .001$ ,  $d = 1.01$ , as well as when spoken in isolation,  $t(19) = 11.7$ ,  $p < .001$ ,  $d = 2.61$ . The remaining interactions between stage and neighbourhood density,  $F(1, 9) = 0.01$ ,  $MSE = 54.58$ ,  $p = .93$ ,  $n_p^2 = .001$ , and three-way interaction, were not significant,  $F(1, 9) < 0.01$ ,  $MSE = 44.83$ ,  $p = .94$ ,  $n_p^2 = 0.001$ .

#### **4.4.2.2 Speech onset latencies**

For each participant, speech onset latencies (pre and post-familiarisation) for all items and sequences were hand-measured using Audacity® (version 2.0.1, 2012). The experimental program began recording immediately when the items/sequences appeared on screen. Therefore, for all trials the speech onset latency was measured as the time from the beginning of the recording until the onset of the first spoken item. Table 4 shows the mean speech latency times for dense and

sparse neighbourhood words spoken in isolation and in sequences, before (Baseline) and after (Post) familiarisation.

Table 4

*Mean speech latency times for items spoken in isolation and within sequences before (Baseline) and after (Post) familiarisation.*

Item	Isolation (ms)		Sequences (ms)	
	Baseline	Post	Baseline	Post
	M (and SD)	M (and SD)	M (and SD)	M (and SD)
Dense	662 (129)	607 (78)	832 (144)	878 (277)
Sparse	666 (130)	612 (81)	819 (133)	879 (231)

A 2x2x2 within-subjects ANOVA with stage (baseline, post-familiarisation), material type (items, sequences) and neighbourhood density (dense, sparse) was performed to establish whether there was any effect of neighbourhood density on the speech latency times for items and sequences before and after familiarisation. The main effect of neighbourhood density was not significant,  $F(1, 9) = .02$ ,  $MSE = 1089.7$ ,  $p = .89$ ,  $n_p^2 < .01$ , suggesting that irrespective of neighbourhood density speech latency times were similar. The main effect of material type was significant,  $F(1, 9) = 9.8$ ,  $MSE = 94,304$ ,  $p = .01$ ,  $n_p^2 = .52$ , with longer speech latency times for sequences than for items. The main effect of stage was not significant,  $F(1, 9) < 0.001$ ,  $MSE = 11,638$ ,  $p = .98$ ,  $n_p^2 < .001$ . However, the interaction between stage and material type was significant,  $F(1, 9) = 6.85$ ,  $MSE = 8432.3$ ,  $p = .03$ ,  $n_p^2 = .43$ . Compared to baseline, speech latency times tended to decrease for the items but increased for the sequences after familiarisation. The remaining interactions between stage and neighbourhood density,  $F(1, 9) = 0.27$ ,  $MSE = 1016.1$ ,  $p = .62$ ,  $n_p^2 = .03$ , neighbourhood density and material type,  $F(1, 9) = 0.49$ ,  $MSE = 1272$ ,  $p = .5$ ,  $n_p^2 = .05$ , and the three-way interaction,  $F(1, 9) = 0.24$ ,  $MSE = 935.4$ ,  $p = .64$ ,  $n_p^2 = 0.03$ , were not significant.

#### **4.4.3 Discussion.**

Words from a sparse neighbourhood had longer spoken durations, both in isolation and within sequences, than words from a dense neighbourhood. This suggests that there are differences in spoken duration times at both the item and the sequence-level. However, the advantage for the dense neighbourhood words compared to sparse neighbourhood words spoken in isolation reduced when spoken in sequences. If the dense neighbourhood words were comprised of more fluent co-articulations than the sparse neighbourhood words then, similarly to words being vocalised quicker than non-words in sequences but not in isolation (Woodward et al. 2008), then an increase, rather than reduction, in the difference between spoken durations for dense and sparse neighbourhood words would have been expected. The finding in the current experiment suggests that the relative fluency at the sequence-level (i.e., co-articulations) is similar between dense and sparse neighbourhood words. However, differences in articulatory difficulty at the item-level still result in there being longer spoken duration times for sparse neighbourhood words contained in sequences. As a second measure of articulatory difficulty speech latency times were also collected. As sequence lengths increase, speech latency times increase (e.g., Sternberg et al., 1978) and this was also the case in the current experiment with longer speech latency times for the sequences compared to items. This suggests that the speech latency times captured motor planning with the time taken to plan vocalisations increasing as the number of items that require planning increases. However, as the difficulty of sequences increases (e.g., days of week in correct vs random order) so do speech latency times (Sternberg et al., 1978). The speech latency times in the current experiment did not differ between dense and sparse neighbourhood sequences pre or post-familiarisation. This is suggestive that there is no difference in the difficulty, as measured by speech latency times, between the articulatory planning of dense and sparse neighbourhood sequences.

The implications of the current findings are that the ND effect using these materials (Clarkson et al., 2017; Experiment 1 in the current thesis) may have been some consequence of differences in spoken durations. If overt speech rates are considered to reflect rehearsal rates within vSTM (e.g., Hulme et al., 1984) then, because the sparse neighbourhood words took longer to vocalise, this would also likely equate to slower rehearsal rates for those sparse neighbourhood items. This presents a few possible re-interpretations of the ND effect in those experiments.

One possibility for the present findings is that the ND effect may have been due to time-based decay, essentially making them an example of a time-based word length effect (better memory for words that can be spoken fastest even when equated on length; e.g., Baddeley et al. 1975). Faster rehearsal theoretically helps offset the detrimental effects of decay within STM. However, there are very few examples of the time-based word length effect and it tends to be far less robust than the syllable-based word length effect (e.g., Neath, Bireta & Surprenant, 2003). This has led researchers to question whether the time-based word length effect really exists or whether it was some unusual consequence of the items originally used by Baddeley et al. (e.g., Jalbert et al., 2011a; Neath et al., 2003). Alternatively, slower rehearsal of items could impact the fluency of rehearsal and introduce more errors into a speech motor plan (e.g., Jones & Macken, 2018; Macken et al., 2015). For example, items in a vSTM task are typically rehearsed in a cumulative fashion (e.g., Tan & Ward, 2008). To-be-remembered items are also very often presented at a rate of 1-item per second. The longer that each to-be-remembered item takes to vocalise then the earlier that cumulative rehearsal of those items is likely to be disrupted by presentation of the next to-be-remembered item. If two sparse neighbourhood items have a combined spoken duration of 1050ms and two dense neighbourhood items have a combined duration of 990ms then it is possible to cumulatively rehearse both dense items before presentation of the third item. However, rehearsal of the second sparse neighbourhood item will be interrupted by presentation of the third sparse neighbourhood item. Once there is no longer enough time to cumulatively rehearse the items participants will adopt a fixed rehearsal strategy instead (e.g., Page & Norris, 1998; Tan & Ward, 2008). However, this is a less effective strategy than cumulative rehearsal and leads to more errors at recall (e.g., Tan & Ward, 2008). It may be the case that sparse neighbourhood words, due to longer spoken durations, force participants to adopt a less effective fixed rehearsal strategy earlier than it is necessary to do so for dense neighbourhood words.

A third possibility for the present findings is that, rather than switching to a fixed rehearsal strategy, participants may still attempt to rehearse the items cumulatively but at a faster rate to ensure a full cumulative rehearsal cycle before presentation of the next item. Faster speech rates can result in more speech errors though (e.g., Dell, 1986; Dell, Burger & Svec, 1997; Fossett, McNeil & Pratt, 2016) and due to the longer spoken durations, it would be the sparse neighbourhood words



needing the greatest increase in rehearsal rate. This could result in more errors during rehearsal of the sparse neighbourhood items and therefore more errors at recall. Finally, it may be the case that rehearsal has no impact upon vSTM and the ND effect is a consequence of supportive activation being elicited within LTM (e.g., Allen & Hulme, 2006; Clarkson et al., 2017; Derragh et al., 2017; Jalbert et al., 2011a, b). However, the current data at least demonstrates that rehearsal cannot yet be entirely ruled out as an explanation for the ND effect.

#### **4.5 General Discussion.**

Study 1 and Experiment 5 investigated whether there is a relationship between neighbourhood density and articulatory difficulty. Study 1 treated articulatory difficulty as the relative effort required by the vocal apparatus to articulate consonant sounds. Three dimensions - their manner of articulation, their place of articulation and their voicing - were used to help quantify difficulty. The results of a regression analysis revealed a tendency for CVC words from denser neighbourhoods to contain easier articulations than CVC words from sparser neighbourhoods. Additionally, in Study 1b, there was evidence that word pools containing CVC words that have been used to investigate the ND effect are confounded by articulatory difficulty. The sparse neighbourhood word pools were found to have higher articulatory difficulty scores than the dense neighbourhood word pools.

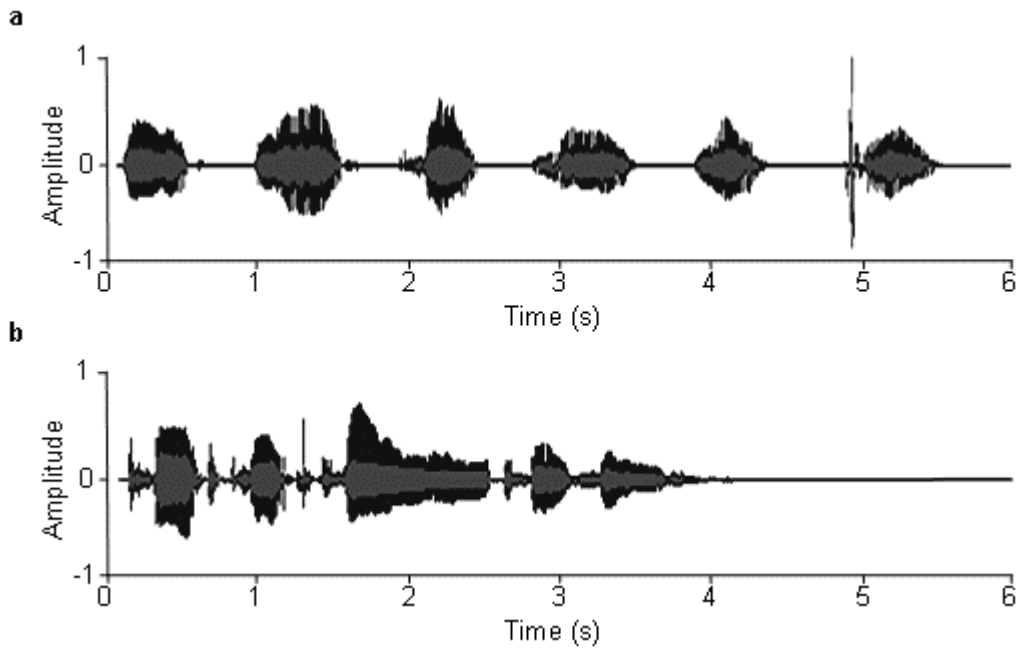
Experiment 5 treated articulatory difficulty as the time taken to plan and then vocalise individual words and word sequences. Sparse neighbourhood words took longer to vocalise both within a sequence and in isolation. However, the differences in spoken durations between dense and sparse neighbourhood words did not increase when the items were spoken within sequences. This suggests that the relative fluency in which sequences of dense and sparse neighbourhood words can be vocalised is similar. However, the differences in spoken durations at the item-level, that still remain to some extent within sequences, indicate that sparse neighbourhood words are perhaps more difficult to articulate. Both sets of data raise doubts over ruling out a possible causal role for articulatory rehearsal when explaining the ND effect (e.g., Jalbert et al., 2011a, b).

The scoring system in Study 1 was an improvement upon the original version devised by St. John (2015) and managed to capture 18% of the total variance. This does still leave a large proportion of the variance left unexplained though and until a difficulty scoring system manages to capture more of the variance, it would perhaps be best to treat any conclusions with some caution. The failure to capture more variance may have been a consequence of only scoring consonants. However, an attempt was also made to develop a vowel difficulty scoring system (see Appendix C) but this failed to predict any variance in phonological neighbourhood density distributions. As such, it seems unlikely that only including consonants in the current scoring system prevented it from capturing more variance. A second possibility is that the fairly broad dimensions used meant that a lot of important articulatory detail was still missed. For example, the fricative consonants /θ/ and /ð/ could perhaps be considered more difficult to articulate than other fricatives. This is because they are usually the final consonant sounds to be acquired by infants (e.g., McLeod & Crowe, 2018) as they require a high degree of motor control and precision. The consonant sound /ð/ is acquired at around 5 years of age and /θ/ at 6. While the age that different consonant sounds are acquired was taken into some consideration when quantifying their relative difficulty (e.g., affricates and fricatives are some of the later consonant sounds to emerge and were therefore categorised as more difficult than stops, nasals and approximants) the age that different consonants within those categories are acquired was not taken into consideration. So, for example, the fricative /θ/ was rated the same difficulty as the fricative /f/ despite /f/ typically being acquired around three years earlier than /θ/ (e.g., McLeod & Crowe, 2018). Future modelling could perhaps incorporate more fine-grained information regarding developmental age.

Another way in which the current scoring system could possibly be improved would be to include additional scoring parameters. For example, as discussed in the introduction, voiced sibilant affricates (e.g., /dʒ/ /dʒ/) tend to be avoided in languages due to the high air pressure required for their vocalisation making them effortful to produce (e.g., Zygis et al., 2012). It may be possible to incorporate air pressure into future scoring systems. However, it is also important to consider that articulatory pressures are just one possible influence upon language development. Lindblom (1990) suggests a trade-off between perceptibility and articulatory pressures whereas other authors caution against a strong focus upon articulatory pressure and

highlight a more important role for perceptual factors (e.g., Simpson, 2002; Ohala, 1990). Therefore, while articulatory difficulty may have had some influence upon neighbourhood density distributions, other factors may have exerted stronger influences. In that respect, it may not be possible for any scoring system only incorporating articulatory difficulty to capture much more of the variance. Future scoring may not only need to include a more fine-grained analysis of articulatory difficulty but also include perceptual pressures to form a more accurate picture. Despite some of these limitations the scoring system still captured 18% of the variance though and adds weight to previous evidence (e.g., St. John, 2015) that neighbourhood density is closely related to articulatory difficulty.

The finding that item lengths spoken within sequences were generally longer than item lengths spoken in isolation is surprising given that items are typically spoken faster within sequences (e.g., Woodward et al. 2008). An inspection of individual participant data revealed that four out of the ten participants did show this pattern of performance though. This raises questions over how accurately the remaining participants followed the task instructions. They were instructed to vocalise the sequences as quickly and accurately as possible but many of the sequence recordings contained waveforms with very distinct items (see Figure 11 for an example). This suggests that in some instances there was very little, if any, co-articulatory overlap and may have meant that any possible differences in co-articulatory fluency between dense and sparse neighbourhood words were not able to be captured. It is possible that the lack of co-articulatory overlap was some consequence of the long baseline measures. Participants spent a large proportion of the experiment (288 trials) vocalising items in isolation before any sequences were presented. Participants may have been focussed on ensuring that they still clearly and accurately vocalised each item rather than providing more natural vocalisation of the sequences. It may be useful for future experiments to counterbalance the order of item and sequence vocalisation although correct pronunciations would still need to be checked beforehand which may lead to similar issues. Alternatively, although it would reduce the generalisability of any findings, it may be better to use materials drawn from far smaller word pools to reduce the length of each stage.



*Figure 11. Waveforms from two 6-item sequences of dense neighbourhood words vocalised by different participants. a|Participant vocalising six items with no co-articulation. b|Participant vocalising six items with co-articulation.*

A further limitation with the spoken recording measures used in Experiment 5 is that they did not require participants to replicate cumulative rehearsal patterns that are commonly used in vSTM tasks (e.g., Tan & Ward, 2008). This may explain why the dense neighbourhood advantage for words spoken within sequences was smaller than the advantage when those same words were spoken in isolation. For example, if participants were required to read the sequences in a cumulative fashion (e.g., item 1, item 1 and 2, item 1, 2 and 3 etc.) there could potentially have been larger differences in the time taken to cumulatively read the dense neighbourhood sequences compared to the sparse neighbourhood sequences. Future experiments could perhaps incorporate this manipulation to ensure a higher degree of similarity between overt articulation and the rehearsal methods commonly used in vSTM tasks. Additionally, to extend the generalisability of the current findings a similar methodology could be applied to the words used in other examples of the ND effect. Along with the results obtained using the articulatory difficulty scoring system, finding that other sparse neighbourhood word pools also take longer to vocalise would provide further evidence for the ND effect perhaps being some consequence of rehearsal.

Despite some limitations the investigations in the present chapter do highlight the importance of giving more consideration to what exactly neighbourhood density distributions reveal. Study 1 and Experiment 5 both highlight how articulatory difficulty may be a causal predictor of neighbourhood density distributions and, as such, could also be responsible for the advantage typically afforded in vSTM tasks to dense neighbourhood words. There is also a broader problem however in respect to neighbourhood density calculations and how they are used in the vSTM literature. Our ability to represent words in written or phonetic form is not necessarily the same thing as that which is being represented (e.g., see Macken, Taylor & Jones, 2015; Wray, 2015). Suggesting that phonological or orthographic neighbours of words have an impact upon vSTM requires the assumption that the tools that have enabled us to visualise language (e.g., letters, phonemes, linguistic networks) are also some inherent property of language itself that can impact vSTM. For example, suggesting that phonological neighbourhood density distributions can impact vSTM requires that linguistic inventories exist in LTM and that these inventories are comprised of the same properties (e.g., phonemes) used to help visualise language (e.g., see Port, 2007). If linguistic inventories in LTM are not comprised of phonemes, then the number of phonological neighbours a word has is arguably then irrelevant when attempting to find an explanation of why that word is, or is not, remembered. Neighbourhood density distributions are clearly useful in helping to visualise and describe patterns of sounds within a given language but the claim that those same measures are a determinant of vSTM should perhaps be given far more consideration than it currently receives.

In conclusion, Study 1 and Experiment 5 have provided evidence suggesting that neighbourhood density distributions are confounded by articulatory difficulty. Not only do sparse neighbourhood words score more highly on articulatory difficulty than dense neighbourhood words but they also take longer to produce in speech. Jalbert et al. (2011b) discounted articulatory factors as a potential explanation for the ND effect. However, the current investigation raises the possibility that the ND effect could be some consequence of the articulatory difficulty of the to-be-remembered items.

## Chapter 5

### Can manipulating articulatory difficulty produce an analogous effect to the ND effect?

#### 5.1 Abstract

The ND effect is typically explained by suggesting that dense neighbourhood words elicit more supportive activation within linguistic networks in LTM than sparse neighbourhood words. However, the results of Chapter 4 indicated that neighbourhood density is confounded with articulatory difficulty. Words from denser neighbourhoods tend to consist of easier articulations and can be vocalised quicker in isolation and within sequences. As such, variations in articulatory difficulty, rather than activation within linguistic networks, may be responsible for some previous examples of the ND effect. However, to sustain this line of argument it is important to demonstrate that effects analogous to the ND effect can be elicited when neighbourhood density and other item-level variables are held constant but where articulatory difficulty is free to vary. One way in which articulatory difficulty can be manipulated is by varying co-articulatory fluency between items that are otherwise matched on item-level properties. Co-articulatory fluency refers to the relative ease with which co-articulations – the transitions required by the vocal apparatus to navigate from the offset of one word to the onset of the next - contained within a to-be-remembered sequence can be navigated by the speech apparatus. Even when item-level variables are held constant, the more fluently that the co-articulations can be navigated by the vocal apparatus then the better vSTM is (e.g., Murray & Jones, 2002; St. John, 2015; Woodward et al., 2008). Experiment 6 explored whether differences in serial recognition accuracy would emerge when using sequences controlled for neighbourhood density but varying in co-articulatory fluency. However, serial recognition performance was similar for the fluent and disfluent sequences. Experiments 7a and 7b tested whether co-articulatory fluency could impact performance in serial reconstruction tasks. The experiments also explored two possible explanations (length of retention interval and presentation rate of items) for the absence of a co-articulatory fluency effect in Experiment 6. However, both experiments again failed to elicit a vSTM advantage for the sequences containing more fluent co-articulations. In Experiment 7a there was an advantage for the

sequences consisting of disfluent co-articulations. Because the experiments failed to demonstrate that variations in articulatory difficulty can produce effects that are analogous to the ND effect it raises some questions about the viability of suggesting that the ND effect might be some consequence of differences in articulatory difficulty between dense and sparse neighbourhood words.

## **5.2 Introduction to Experiments 6 and 7**

Neighbourhood density is considered a powerful predictor of vSTM task performance (e.g., Allen & Hulme, 2006; Clarkson et al., 2017; Derragh, et al., 2017; Jalbert et al., 2011a, b; Roodenrys et al., 2002). As already dealt with in Chapter 3, the work by Jalbert et al. (2011b) suggested that the ND effect is reliant upon the level of activation elicited within LTM and independent of articulatory rehearsal. This raised doubts whether accounts of vSTM that incorporate some role for rehearsal, such as an object-oriented account (e.g., Jones & Macken, 2018; Macken et al., 2015), are adequately able to explain the ND effect. However, via a series of simulations, Chapter 3 demonstrated that almost identical results to Jalbert et al. (2011b) would have been obtained had participants only remembered the first letter of each to-be-remembered item and used them as cues to complete the serial reconstruction tasks. As such, Jalbert and colleagues may have incorrectly ruled out a possible role for rehearsal in explaining the ND effect. Chapter 4 went onto provide some evidence that phonological neighbourhood density is correlated with articulatory difficulty. Words from dense neighbourhoods typically require less articulatory effort and are quicker to vocalise than sparse neighbourhood words.

The findings of Chapters 3 and 4 enable the questioning of whether the levels of supportive activation within LTM (e.g., Derragh et al., 2017; Jalbert et al., 2011a, b) or perhaps the more difficult articulations associated with sparse neighbourhood words are responsible for some previous examples of the ND effect. However, a problem is that the relationship identified in Chapter 4 between neighbourhood density and articulatory difficulty is purely correlational. Any impact that articulatory difficulty may have upon the ND effect is therefore unclear. As such, in order to demonstrate that vSTM effects such as the ND effect may be driven by variations in articulatory difficulty it is important to demonstrate an equivalent effect when neighbourhood density is controlled for, but where articulatory difficulty is free to

vary. A demonstration that vSTM can be impacted solely by articulatory difficulty, when any item-level variations are held constant, would provide some further justification for considering that the ND effect may be a consequence of articulatory difficulty rather than varying levels of supportive activation being elicited within LTM.

To establish the impact that articulatory difficulty might upon vSTM the current chapter focusses upon the manipulation of co-articulatory fluency (e.g., Murray & Jones, 2002; St. John, 2015; Woodward et al., 2008). Co-articulatory fluency refers to the relative ease with which the co-articulations required to vocalise the offset of one word and the onset of a subsequent word can be achieved by the vocal apparatus. Fluent vocalisation of items is thought to be underpinned by a comprehensive forward planning process (e.g., Sternberg et al., 1978; Sternberg, Wright, Knoll & Monsell, 1980). When required to vocalise a sequence of items (e.g., item 1, item 2, item 3 etc.) the plan will not only include the articulations required for a current item but also the articulations required for future items and the co-articulations required to navigate the vocal apparatus to those future items. Because of this forward planning, the articulatory movements needed for adjacent gestures can anticipate and accommodate each other. So, for example, the articulation required for the offset of item 1 can assimilate some of the features required for the articulation for the onset of item 2. The articulations required for the offset of item 2 can assimilate some of the features required for the articulation of item 3 and so on. Assimilation of features allows for the production of fluid and connected speech. However, despite this comprehensive planning process, the co-articulatory fluency between items can still vary. For example, the co-articulation between the words 'tap' and 'bath' involves navigating the vocal apparatus from the offset consonant /p/ to the onset consonant /b/. This could be considered a fluent co-articulation as both consonants have bilabial places of articulation requiring very little reconfiguration of the vocal apparatus between them. However, co-articulation between 'tap' and 'gas' could be considered a disfluent co-articulation as the onset consonant /g/ has a velar place of articulation. The navigation from /p/ to /g/ is further than the fluent navigation from /p/ to /b/ resulting in a more complex and time-consuming transition. Murray and Jones (2002) used a serial reconstruction task to test participant's ability to reconstruct the order of 8-item sequences comprised solely of either fluent or disfluent co-articulations. Reconstruction was better for fluent sequences (M = 61.1%) whereby all co-articulations contained small changes in place of articulation



(e.g., 'rail-rice-nurse-wren-sill-sun-lean-ran') compared to disfluent sequences (M = 54.2%) all containing larger changes in place of articulation (e.g., 'tape-knife-turf-deaf-nib-deep-cup-cap'). It was suggested that the fluent sequences allowed for more efficient and less error prone (i.e., fewer naturally occurring speech errors) assembly of the items into a motor plan and a more fluent rehearsal process. Rather than rehearsal reviving degraded traces within STM (e.g., Baddeley, 1986; Baddeley, 2000) rehearsal is considered a process whereby the assembled speech motor plan is maintained via cyclical pre-output execution (e.g., Woodward et al., 2008). This task advantage for fluent sequences will be referred to as the *co-articulatory fluency effect*.

The research by Murray and Jones (2002) has not been without criticisms though. Miller (2010) noted that while Murray and Jones matched the fluent/disfluent sequences on two item-level properties known to impact vSTM task performance (frequency and length) the items had not been matched on phonological neighbourhood density. Miller calculated the mean phonological neighbourhood density in Murray and Jones as 31.38 for fluent sequences and 17.75 for disfluent sequences. This neighbourhood density confound is of importance as item-based accounts of vSTM (e.g., Jalbert et al., 2011a, b; Derragh et al., 2017) can explain the results of Murray and Jones without any need to acknowledge a possible role for the fluency of motor planning and rehearsal. Instead, the effect found by Murray and Jones can be considered another example of the ND effect. It is therefore difficult to establish whether the results were due to the manipulation of co-articulatory fluency or the neighbourhood density confound.

St. John (2015, Experiment 1) attempted to address the neighbourhood density confound highlighted by Miller (2010) by matching fluent/disfluent word pools on item-level properties including phonological neighbourhood density. To help facilitate the matching process an alternative fluency manipulation to that used by Murray and Jones (2002) was employed. Instead of small/large changes in place of articulation, co-articulatory fluency was manipulated by varying the direction of the co-articulations between items. Due to anatomical constraints (e.g., see Chitoran, Goldstein & Byrd, 2002) a backwards moving change (e.g., from /b/ to /g/) has a higher degree of co-articulatory overlap than the equivalent forwards moving change (e.g., from /g/ to /b/). In the examples, this is because the primary constriction of /b/ is formed at the front of the vocal tract by blocking airflow using the lips. This

constriction can be released without sacrificing the perceptual impact of /b/. This then makes it possible to simultaneously continue the vocalisation of /b/ while preparing for the primary constriction of /g/ which has a posterior constriction formed with the tongue body. However, because /b/ is formed by blocking airflow with the lips it is not possible to prepare for the vocalisation of /b/ while simultaneously vocalising /g/. This is because /g/ requires continuous air expulsion. As soon as air is blocked /g/ can no longer be vocalised. As such, compared to the backwards moving co-articulation between /b/ and /g/, there is very little co-articulatory overlap between the forwards moving co-articulation between /g/ and /b/. As the amount of co-articulatory overlap increases the production time and articulatory complexity reduces which results in a more fluent co-articulation. Using a serial recall task St. John (2015) presented to-be-remembered sequences that consisted entirely of fluent backwards moving co-articulations (e.g., '*nap-doom-ripe-ship-jeep-loop*') or entirely of disfluent forwards moving co-articulations (e.g., '*veil-boon-fan-peas-budge-vice*'). Serial recall was more accurate for the fluent sequences ( $M = 54.63\%$ ) compared to the disfluent sequences ( $M = 51.21\%$ ). However, an issue with using serial recall is that any errors at recall may be a consequence of spoken errors rather than memory errors. For example, the participant may have correctly remembered a word but when attempting to vocalise the word may then have produced a speech error. This would be scored as incorrect despite the participant having remembered the correct word.

To address the concern that spoken errors may be responsible for the co-articulatory fluency effect St. John (2015, Experiments 2a/2b) compared performance between serial recall and serial reconstruction. Unlike serial recall, serial reconstruction requires no spoken output. This is sometimes considered to minimise the impact that possible differences in spoken production may have upon the task (e.g., Jalbert et al., 2011a, b). However, rather than testing memory for the same sets of items as used in Experiment 1, St. John also utilised a different fluency manipulation whereby the same items were used for both the fluent and disfluent sequences. This was achieved by presenting sequences of non-words consisting predominantly of fluent backwards moving co-articulations (e.g., '*bup-dat-geg-pobe-dord-kug*') and then reversing the order of those same non-words to produce a sequence with predominantly disfluent forwards moving co-articulations (e.g., '*kug-dord-pobe-geg-dat-bup*'). In all instances the co-articulation between items 3 and 4

was in the opposite direction to the rest of the sequence but an increased ISI (750ms rather than 0ms) between those two items was considered to reduce the likelihood of co-articulation occurring between those two items (cf. Cho & Keating, 2001). This presentation rate adjustment kept all the co-articulations within a particular sequence fluent or disfluent. Serial recall (Experiment 2a) and serial reconstruction (Experiment 2b) was best for the sequences containing the fluent backwards moving co-articulations. This suggests that the fluency of co-articulations within to-be-remembered sequences have measurable impacts upon vSTM. Because the effect was also found in serial reconstruction it suggests that the effect when using non-words was not some consequence of spoken errors.

The work by St. John (2015) controlled for neighbourhood density and only manipulated co-articulatory fluency. As such, it went some way in addressing concerns that the previous example of the co-articulatory fluency effect (Murray & Jones, 2002) was just another example of the ND effect. The work also demonstrates that vSTM effects can be elicited by manipulating sequence-level, rather than item-level, properties. This raises the possibility that vSTM effects typically ascribed to item-level manipulations, such as neighbourhood density, should perhaps be considered as effects that are driven by variations in sequence-level properties such as articulatory difficulty. However, before making any generalisations there are two key issues in the work by St. John that first need addressing. These relate to the use of serial recall when testing vSTM for words and the use of non-words in St. John's later experiments. Firstly, serial recall requires participants to reproduce the items (e.g., spoken, written, typed) in their correct order. This means that any effects may be a result of errors during production of the items (e.g., spoken errors, written errors) rather than memory errors. While St. John went some way in addressing this concern by using serial reconstruction the task was only used when testing memory for non-words. This means that the co-articulatory fluency effect for words (St. John, Experiment 1) was only demonstrated using serial recall and could therefore still have been a result of spoken errors. To address this concern, it is important to provide further tests of the co-articulatory fluency effect using vSTM tasks that do not require spoken output. Serial recognition and serial reconstruction are two such vSTM tasks meeting this requirement.

A second issue with the work by St. John (2015) is that a demonstration of a co-articulatory fluency effect using non-words is not necessarily evidence of a co-

articulatory fluency effect that will generalise to vSTM for words. For example, redintegration is used as an explanation for a variety of vSTM effects such as lexicality (e.g., Gathercole et al., 2001), frequency (e.g., Hulme et al., 1997) and neighbourhood density (e.g., Jalbert et al., 2011a, b; Derragh et al., 2017) and could therefore be considered an essential component of successful vSTM. However, non-words are not considered to have pre-existing representations in LTM (e.g., Gathercole et al., 2001) and without any pre-existing representations it should not be possible for LTM to provide redintegrative support to any degraded non-words in STM (although it has been argued that non-words can still elicit activation within their orthographic or phonological neighbourhoods in LTM; see Jalbert et al., 2011b). When a co-articulatory fluency effect is demonstrated using non-words (St. John, Experiments 2a & 2b) it could therefore be argued that there is no opportunity for LTM to provide support to STM via redintegration. In the absence of usual vSTM processes, other processes such as the learning and practicing of the novel articulations required to rehearse non-words may have unusually strong impacts upon vSTM task performance that they otherwise would not have had. It is therefore important to demonstrate that co-articulatory fluency effects can still be elicited when using words.

Although it has been argued that it is important to demonstrate a co-articulatory effect using words rather than non-words a problem with using words is that despite best efforts to match them on item-level variables there remains the risk that they will unintentionally vary on some other item-level variable. To address this problem the first experiment in the current series utilised the experimental paradigm of Macken et al. (2014). As discussed in Chapter 1 (Section 1.7.1), the lexicality effect is usually robust in serial recall but attenuates in serial recognition (e.g., Gathercole et al., 2001). This is typically considered to be a consequence of the re-presented items at test repairing degraded items within STM. Because the repair process for all items is equally supported by the re-presented items there is no longer any requirement to rely upon LTM whereby the level of redintegrative support varies as a function of any item-level differences. However, Macken et al. (2014) noted that previous tests of the lexicality effect had almost exclusively used auditorily presented serial recognition. Despite the lexicality effect supposedly being an item-level effect whereby using serial recognition would mean that re-presented items repair all degraded traces in STM, Macken et al. successfully elicited the lexicality

effect in visual serial recognition. The lexicality effect emerged in visual serial recognition because the task is thought to first require that the standard sequence is segmentally recoded, via speech motor processes, into articulatory form before it can later be compared with the test sequence. The relative fluency of the segmental recoding process determines subsequent recognition accuracy. The articulations required to segmentally recode words are considered to be more fluent than those required for non-words because participants will have had more prior experience articulating words. However, auditory serial recognition provides a setting whereby the auditory input can promote auditory object formation (e.g., Bregman, 1990). An order judgement between two auditory objects can be made without needing to have any knowledge of the specific items that comprise that object (e.g., Warren et al., 1969). In auditorily presented serial recognition an accurate order judgement can be achieved by pattern matching whereby the overall global sound of the standard sequence is compared with the overall global sound of the test sequence. This is particularly the case if the presentation rates are optimised in such a way to promote auditory object formation. Macken et al. used very fast presentation rates (250ms with 100ms ISI) in order to promote auditory object formation and reduce the likelihood of participants segmentally recoding the auditory items. As such, the lexicality effect was eliminated in auditory serial recognition.

The differences between the accounts of the lexicality effect in serial recognition tasks offered by Gathercole et al. (2001) and Macken et al. (2014) provide an ideal setting in which to test the co-articulatory fluency effect. Firstly, according to Gathercole et al., serial recognition (irrespective of presentation modality) reduces the requirement for LTM with re-presented items repairing any degraded traces in STM. As such, any overlooked item-level variables should not impact the outcome of the task. Additionally, because the co-articulatory fluency effect is a sequence-level, not an item-level effect (i.e., item-level properties are held constant), then it should not be possible to elicit a co-articulatory fluency effect in visual or auditory serial recognition. Irrespective of any variations in fluency the to-be-remembered items in STM will receive equivalent levels of support from the re-presentation of the items. However, if visual serial recognition first requires segmental recoding whereas auditory serial recognition can be completed via pattern matching (e.g., Macken et al.) then the more fluently the visually presented to-be-remembered sequences can be segmentally recoded then the better serial

recognition performance will be. As such, an effect of co-articulatory fluency should be elicited in visual serial recognition. However, irrespective of any variations in fluency, auditory serial recognition can be completed via pattern matching which will attenuate or eliminate the effect.

To summarise, the current experiments extend upon previous work into the co-articulatory fluency effect (e.g., Murray & Jones, 2002; St. John, 2015) and establish whether the effect persists when neighbourhood density is controlled for and when spoken output is not required. This was achieved by comparing memory for fluent and disfluent word sequences in vSTM memory tasks that require no spoken output. Experiment 6 used serial recognition and Experiment 7 used serial reconstruction. Secondly, by using serial recognition in Experiment 6 it was also possible to use the Macken et al. (2014) paradigm to further test whether the co-articulatory fluency effect is a consequence of segmental recoding. Such a finding would highlight that variations in articulatory difficulty, rather than just item-level variables, have measurable impacts upon vSTM performance. This would lend weight to the suggestion that vSTM effects such as the ND effect can be explained by drawing reference to variations in articulatory difficulty rather than the differing levels of supportive activation dense and sparse neighbourhood words are considered to elicit within linguistic networks in LTM.

### **5.3 Experiment 6**

Experiment 6 assessed the effects of modality on the co-articulatory fluency effect by testing serial recognition of 5-item sequences presented both visually and auditorily at a fast (250ms per item, 100ms ISI) presentation rate. Sequences were constructed with words drawn from 40-item word pools containing either fluent backwards moving or disfluent forwards moving co-articulations.

#### **5.3.1 Method**

##### **5.3.1.2 Participants**

30 participants (mean age 20 years, 23 female and 7 male) were recruited from the Cardiff School of Psychology participation pool and awarded course credits for participating. Participants were required to have normal/corrected vision and

hearing. All stages of the experiment were conducted in accordance with the Cardiff School of Psychology ethics procedures.

### **5.3.1.2 Materials**

Stimuli were constructed by employing the directional fluency manipulation used by St. John (2015, Experiment 1). The consonants /b/, /f/, /m/, /p/, /v/ were categorised as ‘anterior’ as articulation involves recruitment of the lips. The consonants /k/, /g/, /d/, /dʒ/, /n/, /l/, /r/, /s/, /t/, /z/ were categorised as ‘posterior’ as articulation is further back in the vocal apparatus and involves the recruitment of the tongue tip, tongue body or glottis. A *disfluent* word pool was created containing words with anterior onsets and posterior offsets (e.g., ‘pin’) and a *fluent* word pool was created containing words with posterior onsets and anterior offsets (e.g., ‘nap’). The words were all obtained from The MRC Psycholinguistic Database (Wilson, 1988). Each pool contained 40 items (See Appendix D) and *t*-tests revealed that the pools did not significantly differ on word frequency,  $t(78) = 0.06$ ,  $p = .96$ , phonological neighbourhood density,  $t(78) = 0.03$ ,  $p = .98$ , or orthographic neighbourhood density,  $t(78) = 1.38$ ,  $p = .17$ .

Standard sequences were constructed by selecting 5 items at random, without replacement, from the appropriate word pool. Sequences containing fluent backwards moving co-articulations sampled items from the fluent word pool (e.g., ‘nap-rub-doom-cape-dip’) and sequences containing disfluent forwards moving co-articulations sampled items from the disfluent word pool (e.g., ‘fan-bid-bite-pin-bout’). Once the word pool was depleted all items were returned and a further 8 sequences were constructed. A total of 32 fluent and 32 disfluent standard sequences were constructed. *Different* test sequences were then constructed by randomly selecting 16 fluent and 16 disfluent test sequences and transposing two adjacent items (excluding the first and last item). Each possible transposition occurred an equal number of times. To familiarise participants with the experimental procedure an additional 8 (4 fluent/4 disfluent) practice sequences were constructed using the same method used to construct the experimental sequences.

For the auditory stimuli, each item was recorded in a monotone male voice at a sample rate 44.1 kHz/16-bit using a condenser microphone. Each item was digitized using Audacity® (version 2.0.1, 2012) software and edited to a duration of

250ms using the 'adjust tempo' option which preserves the pitch of edited items. Items were then assembled into 5-item sequences (100ms ISI) and exported as 16-bit WAV files.

### **5.3.1.3 Design and procedure**

The experiment used a 2x2 within-subject, repeated measures design, with modality (auditory, visual) and fluency (fluent, disfluent) as factors. The experimental procedure took place in a sound attenuated booth with visual sequences presented on a computer monitor and auditory sequences presented via Sennheisser HD280 headphones. For both modalities, the onset of all trials was initiated by the participant pressing the spacebar. A fixation cross then appeared on screen for 1000ms followed by presentation of the first item. Each item in the standard sequence was presented for 250ms separated by a 100ms interstimulus interval which consisted of a blank screen (visual presentation) or silence (auditory presentation). After presentation of the final item in the standard sequence there was a 1500ms delay (blank screen in both modalities). This delay was incorporated to allow some additional time for participants to rehearse and increase the likelihood of any co-articulatory fluency effects emerging. The fixation cross then re-appeared for 1000ms before serial presentation (same parameters as the standard sequence) of the test sequence. Immediately after presentation of the test sequence a centrally located question mark prompted participants to provide a same/different response via the keyboard ('z' for same and 'm' for different). To ensure that participants understood which key corresponded to which response, 'Same' appeared at the bottom left of the computer screen and 'Different' at the bottom right (locations directly above the corresponding keys). The two physical keys were also labelled with a "Same" or "Different" sticker.

The visual and auditory stimuli were presented separately in two blocks of 64 trials. The blocks were counterbalanced across participants and for each modality the 32 5-item fluent sequences and 32 5-item disfluent sequences were presented randomly without replacement.



### 5.3.2 Results and Discussion

Data were collapsed across the *same/different* trials and a percentage correct was calculated for each fluency manipulation in each modality (see Figure 12).

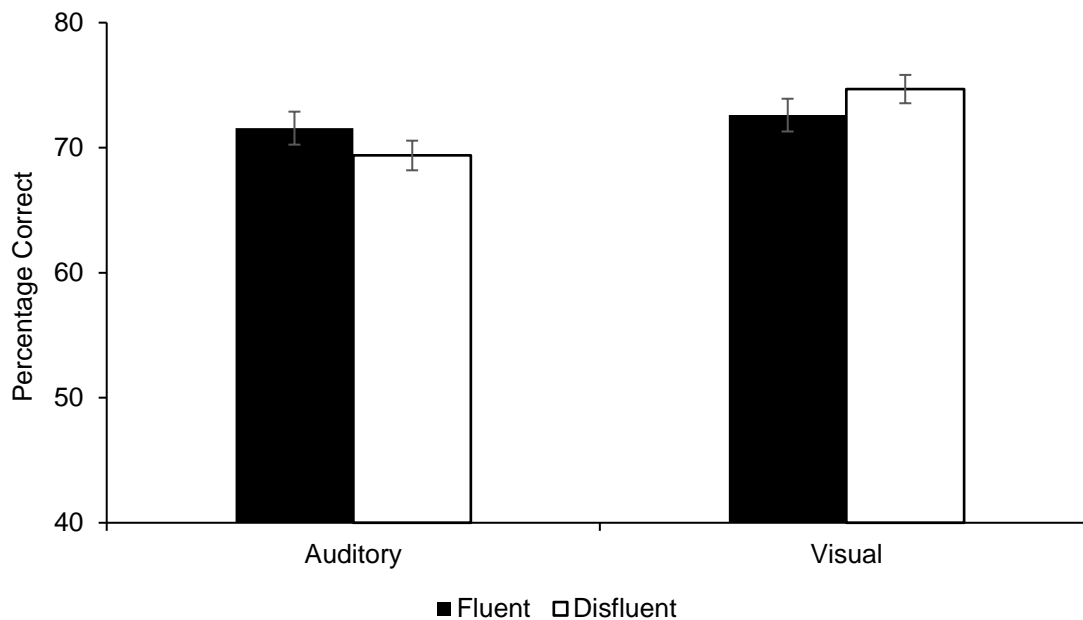


Figure 12. Mean percentage correct in each condition. Error bars denote within-subject standard error (cf. Cousineau, 2005).

A 2x2 repeated measures ANOVA was conducted to assess the effects of modality (auditory, visual) and fluency (dense, sparse) on serial recognition performance. The main effects of fluency,  $F(1, 29) < 0.01$ ,  $MSE = 43.35$ ,  $p = .97$ ,  $n_p^2 < .01$ , and modality were not significant,  $F(1, 29) = 2.92$ ,  $MSE = 103.63$ ,  $p = .1$ ,  $n_p^2 = .09$ . The key interaction of interest between fluency and modality was significant,  $F(1, 29) = 4.38$ ,  $MSE = 31.23$ ,  $p = 0.045$ ,  $n_p^2 = .13$ . However, pairwise comparisons (two-tailed) suggest that this interaction was caused by an improvement in performance for visually presented disfluent sequences compared to auditorily presented disfluent sequences,  $t(29) = 2.74$ ,  $p = .01$ , whereas performance was similar for the fluent sequences irrespective of modality,  $t(29) = 0.46$ ,  $p = .65$ . There was no indication that serial recognition accuracy differed between the fluent and disfluent sequences in either the auditory,  $t(29) = 1.35$ ,  $p = .19$ , or visual,  $t(29) = 1.36$ ,  $p = .18$ , modality.

The results of Experiment 6 demonstrate that serial recognition performance is similar for fluent and disfluent word sequences and that this is irrespective of what modality the to-be-remembered sequences are presented in. If auditory presentation of to-be-remembered materials at fast rates affords participants with the opportunity to complete a serial recognition task via pattern matching (e.g., Macken et al., 2014) then the absence of a significant co-articulatory fluency effect in auditory serial recognition is not unexpected. However, a failure to elicit a co-articulatory fluency effect in visual serial recognition is unexpected. Visual presentation is considered to require that each item is segmentally recoded before an order judgement can be made. The relative fluency of this segmental recoding process is thought to impact vSTM task accuracy (e.g., Macken et al., 2014). Because of this it would be expected that recognition accuracy be better for the fluent visual sequences compared to the disfluent visual sequences. No evidence for a co-articulatory fluency effect in serial recognition can be easily accommodated if serial recognition is considered a task whereby the re-presentation of items obviates the need for LTM though (e.g., Gathercole et al., 2002). Properties of the items themselves, not variations in co-articulatory fluency, are considered to impact vSTM. Besides from all items being matched on item-level properties they were also all re-presented at test. This would have served to repair any degraded items in STM irrespective of whether they were from fluent or disfluent word pools. Such an account of vSTM would make the finding that serial recognition accuracy is similar between fluent and disfluent sequences, irrespective of modality, unsurprising.

Another possibility is that the serial recognition task failed to successfully capture the errors elicited by the co-articulatory fluency manipulation. For example, a participant could make an error when segmentally recoding items 4 and 5 (e.g., phonemes being substituted or a transposition of the items) from a disfluent to-be-remembered standard sequence. However, if a transposition in the test sequence occurs between positions 2 and 3, then the participant could still correctly respond *different* despite having remembered the sequence incorrectly. As such, the serial recognition task may not have captured enough errors in which to demonstrate any effect of co-articulatory fluency upon the task. Macken et al. (2014) did successfully demonstrate a lexicality effect in visual serial recognition though. This suggests that serial recognition is sensitive enough to detect some effects. One possibility is that non-word sequences may involve a more disfluent segmental recoding process than

the disfluent sequences in the current experiment did. The more disfluent segmental recoding is, then the more likely that there will be errors at multiple serial positions and therefore, the more successful a serial recognition task is likely to be in capturing those errors.

Asides from the serial recognition task possibly preventing a co-articulatory effect from being detected, it is also important to remember that Macken et al. (2014) found a robust lexicality effect in visual serial recognition but not in auditory serial recognition. Previous research (e.g., Gathercole et al., 2001) had failed to consider the presentation modality of to-be-remembered items meaning that explanations of the lexicality effect in serial recognition were based almost exclusively upon the results of auditory serial recognition tasks. This oversight highlights just how important a wide variety of other task parameters, often not considered, are for theory development. A final possibility for the failure to elicit a co-articulatory fluency effect in the current experiment is that the particular task parameters somehow prevented the co-articulatory fluency effect from emerging in serial recognition. Two task parameters that may have prevented the co-articulatory fluency effect from emerging, retention interval and presentation rate, are returned to and addressed in the upcoming section.

#### **5.4 Experiment 7**

Experiment 7 used serial reconstruction as an alternative to serial recognition. This provided a vSTM task whereby there is still no requirement for participants to vocalise items. This not only enabled a further test of whether the directional co-articulatory fluency manipulation can impact vSTM but also whether the effect found originally by St. John (2015, Experiment 1) was possibly a consequence of spoken errors during output rather than the fluency of segmental recoding. Using serial reconstruction also addresses the concern highlighted in the previous section that a serial recognition task is perhaps not sensitive enough to detect a co-articulatory fluency effect. Unlike serial recognition, serial reconstruction can capture errors at any of the serial positions during every trial. In the earlier example it was highlighted how serial recognition could miss a transposition error between items 4 and 5 if the standard sequence switches the position of items 2 and 3 (the participant would still correctly respond *different*). However, in serial reconstruction, this error would be

captured when the participant incorrectly reconstructs the sequence by clicking items 4 and 5 in the wrong order.

Experiment 7 also enabled a further test of whether the task parameters used in Experiment 6 prevented elicitation of a co-articulatory fluency effect. Two task parameters, retention interval and the presentation rate of items, were more closely examined. Previous demonstrations of the co-articulatory fluency effect using serial recall/reconstruction have all had slower item presentation rates or incorporated a longer retention interval. For example, presentation of the to-be-remembered items in St. John (2015, Experiments 1 and 2) was 750ms per item with a 750ms ISI. After presentation of the final item there was a 10s retention interval. Item presentation in Murray and Jones (2002) was 750ms per item with a 250ms ISI before immediate serial reconstruction after presentation of the final item. This means that, in those experiments, participants were provided with more overall time in which they could engage in rehearsal. If rehearsal consists of a speech output plan being cyclically executed (e.g., Woodward et al., 2008) then longer rehearsal periods (>4000ms) may be necessary for errors to become incorporated into the speech motor plan and for the co-articulatory fluency effect to be elicited. To use the example of a 'tongue twister', (e.g., '*leap-note-nap-lute*') repeated reiteration is known to induce speech errors (e.g., Wilshire, 1999). The first reiteration (i.e., analogous to immediate recall after very little opportunity for rehearsal) will likely contain fewer errors than the sixth reiteration (i.e., analogous to delayed recall after several cumulative rehearsal cycles). In Experiment 6 the short retention interval (2500ms) may not have provided adequate time for the entire to-be-remembered sequence to be cumulatively rehearsed enough times to successfully elicit a co-articulatory fluency effect.

Secondly, not only do slower presentation rates provide more overall time for rehearsal but they can also alter the way in which participants rehearse the items. For example, Tan and Ward (2008) suggest that the most commonly used and beneficial rehearsal style for vSTM tasks is a cumulative rehearsal strategy that begins as soon as the first item is presented (e.g., item 1, item 1 and item 2, item 1 and item 2 and item 3, and so forth). However, this strategy is not possible with fast presentation rates because there is not enough time for a participant to rehearse multiple items before presentation of the next item. Instead, at faster rates it is only possible to rehearse the currently presented item (a fixed rehearsal strategy). In Experiment 6, a failure to provide participants with adequate time to cumulatively

rehearse while the to-be-remembered items were presented may have led to a situation whereby a co-articulatory fluency effect could not be elicited. A fixed rehearsal method would mean that each of the disfluent co-articulations would only have occurred once, rather than multiple times, during presentation of the to-be-remembered sequence.

As well as testing whether a co-articulatory fluency effect for words can be elicited in serial reconstruction when spoken output is not required, Experiment 7a also tested whether a longer retention interval is necessary to elicit the co-articulatory fluency effect by comparing serial reconstruction of sequences of fluent or disfluent words after a short (2500ms) or long (10s) retention interval. Experiment 7b tested whether slower presentation rates are required to elicit a co-articulatory fluency effect by decreasing the presentation rate from 250ms per item (100ms ISI) to 750ms per item (750ms ISI).

#### **5.4.1 Experiment 7a Method**

##### **5.4.1.1 Participants**

30 participants (mean age 19 years, 25 female and 5 male) who had not participated in Experiment 6 were recruited from the same demographic as the previous experiment and awarded course credits for participating.

##### **5.4.1.2 Materials, Design and Procedure**

A 2x2x5 within subject, repeated-measures design was employed with fluency (fluent, disfluent), retention interval (2500ms, 10s) and serial position (1-5) as factors. The to-be-remembered word sequences and presentation parameters were identical to the visual presentation of the standard sequences in Experiment 6. However, following visual presentation of the to-be-remembered sequence there was a retention interval (blank screen) of either 2500ms or 10s (blocked and counterbalanced across participants). After the retention interval all five to-be-remembered words were re-presented on screen as labels around a centrally located question mark. Participants were instructed to reconstruct the to-be-remembered sequence by clicking on each word in the same order that they had originally been presented. Once clicked, each label changed colour to white and the

word disappeared with no opportunity to correct errors. If unsure of the order of the to-be-remembered sequence participants were encouraged to guess the order of the items. After all the words had been selected, participants were prompted to initiate the next trial by pressing the spacebar.

### 5.4.2 Results and Discussion

Responses were scored as correct if the item was selected in the correct serial position. Serial position curves for the four combinations of retention interval and fluency are shown in Figure 13.

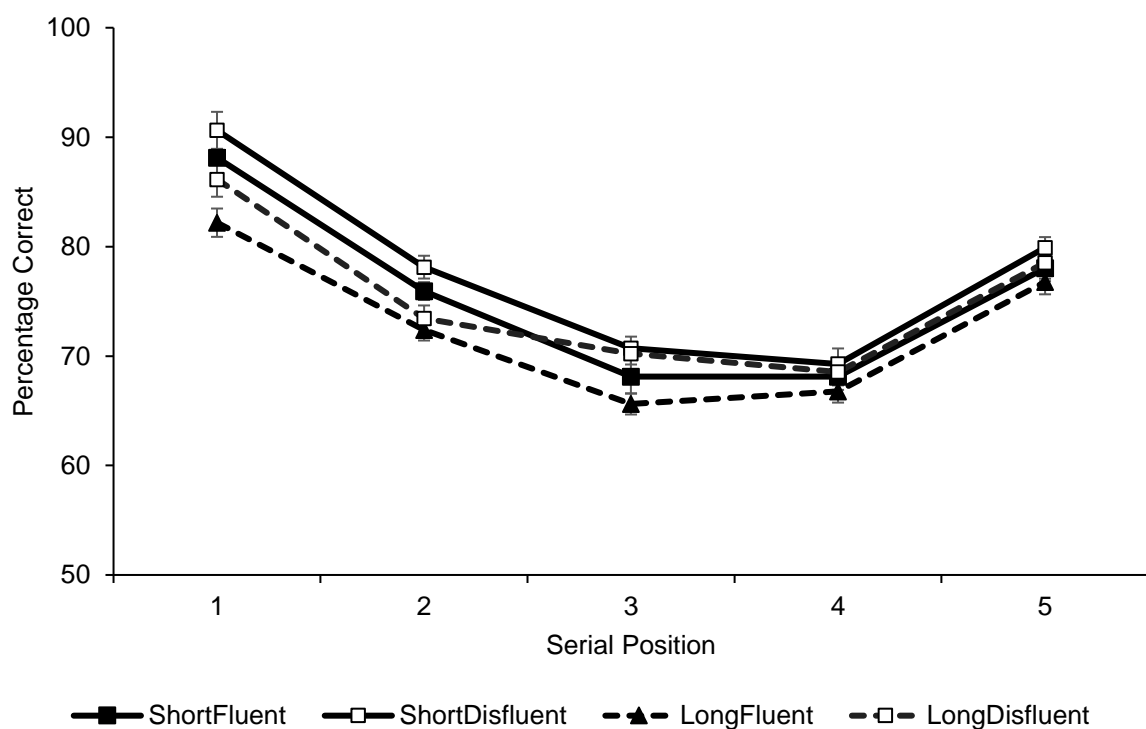
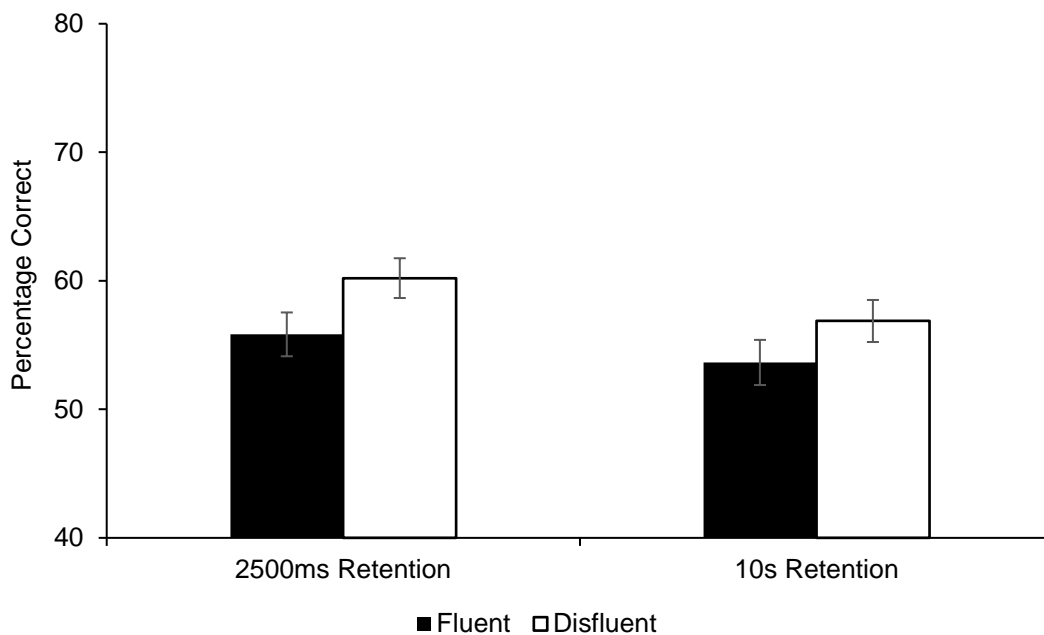


Figure 13. Mean serial position curves for recall of 5-item sequences of fluent and disfluent words with either a short (2500ms) or long (10s) retention interval. Error bars denote within-subject standard error (cf. Cousineau, 2005).

To assess the effects of fluency (fluent, disfluent), retention interval (1500ms, 10s) and serial position (1-5) on the serial reconstruction task a 2x2x5 repeated measures ANOVA was conducted. There was a significant main effect of fluency,  $F(1, 29) = 8.16$ ,  $MSE = 101.02$ ,  $p = .01$ ,  $n_p^2 = .22$ , with more accurate reconstruction of the disfluent rather than the fluent sequences. The main effect of serial position,  $F(4, 116) = 50.91$ ,  $MSE = 138.69$ ,  $p < .001$ ,  $n_p^2 = .64$ , was also significant reflecting a

general decline in accuracy across serial positions until serial position 5 where accuracy increased irrespective of retention interval. The main effect of retention interval,  $F(1, 29) = 2.01$ ,  $MSE = 518.81$ ,  $p = .17$ ,  $n_p^2 = .07$ , was not significant. However, the interaction between serial position and retention interval,  $F(4, 116) = 2.66$ ,  $MSE = 40.51$ ,  $p = .04$ ,  $n_p^2 = .08$ , was significant. This most likely reflects the less accurate reconstruction performance at the first two serial positions with a 10s retention interval compared to the 2500ms retention interval. Performance was then similar at the remaining serial positions irrespective of the retention interval length. The interaction of interest between fluency and retention interval was not significant,  $F(1, 29) = 0.17$ ,  $MSE = 71.97$ ,  $p = .69$ ,  $n_p^2 = .01$ . The effect of fluency in each of the retention intervals, collapsed across serial position, is shown in Figure 14. This suggests that, irrespective of retention interval, any differences in reconstruction accuracy between the fluent and disfluent sequences were similar in size. The remaining interactions were not significant ( $p > .05$ ).



*Figure 14.* Mean percentage correct in each condition collapsed across serial position. Error bars denote within-subject standard error (cf. Cousineau, 2005).

Experiment 7a used serial reconstruction to test whether a directional co-articulatory fluency effect (e.g., St. John, 2015, Experiment 1) could be elicited in a task not requiring spoken output. This would ensure that any emergence of a co-

articulatory fluency effect was not a consequence of spoken errors. However, similarly to Experiment 6, there was once again no support for the hypothesis that vSTM task performance would be better for fluent rather than disfluent sequences (e.g., Murray & Jones, 2002; Woodward et al., 2008; St. John, 2015). Instead, reconstruction accuracy was better for the disfluent sequences. The implications of this advantage for the disfluent sequences are considered in light of the further findings of Experiment 7b. A second aim of the experiment was to establish whether a longer retention interval than 2500ms is required to elicit a co-articulatory fluency effect. Despite better serial reconstruction for the disfluent sequences this effect was similar irrespective of whether the retention interval was 2500ms or 10s. This suggests that the short (2500ms) retention interval in Experiment 6 was not the reason for failing to elicit a co-articulatory fluency effect in serial recognition.

Experiment 7b provided a further test of whether the co-articulatory fluency effect found by St. John (2015, Experiment 1) was possibly a consequence of spoken errors at output. This was achieved by once again using a serial reconstruction task whereby spoken reproduction of the items is not required. The experiment also provided a test of whether the item presentation rate in Experiment 6 was responsible for the lack of co-articulatory fluency effect. The directional co-articulatory fluency effect was originally found by St. John when using serial recall and a presentation rate of 750ms per item with a 750ms ISI. This means that the effect may have been a consequence of participants cumulatively rehearsing the items while they were being presented (e.g., Tan & Ward, 2008). To test this possibility the serial reconstruction task used in Experiment 7b matched the presentation parameters originally used by St. John. For the sake of correspondence between the current experiment and St. John this also meant using 6-item, rather than 5-item, to be remembered sequences. Additionally, even though the long retention interval in Experiment 7a led to slightly lower overall reconstruction accuracy than the short retention interval there was no suggestion that the retention intervals interacted with fluency. It was therefore decided that Experiment 7b solely employ a 10 second retention interval. This was not only the same as the retention interval used by St. John but was also considered the best option to maximise the opportunities for participants to engage in rehearsal and for a co-articulatory fluency effect to emerge.



### **5.4.3 Experiment 7b Method**

#### **5.4.3.1 Participants**

30 participants (mean age 21 years, 26 female and 4 male) who had not participated in Experiment 6 or Experiment 7a were recruited from the same demographic as Experiment 6 and awarded course credits or £6 for participating.

#### **5.4.3.2 Materials, Design and Procedure**

To construct 6-item word sequences while ensuring that each item was presented an equal number of times two additional items were added to each word pool bringing the total items in each pool to 42. 'Lap' and 'tub' were added to the fluent word pool and 'fade' and 'vice' were added to the disfluent word pool (See Appendix E). *t*-tests revealed that the two word pools did not significantly differ on word frequency,  $t(82) = 0.03$ ,  $p = .98$ , phonological neighbourhood density,  $t(82) = 0.03$ ,  $p = .98$ , or orthographic neighbourhood density,  $t(82) = 1.61$ ,  $p = .11$ .

6-item to-be-remembered word sequences were constructed using the same methodology as Experiment 6. Presentation rate of items was 750ms per item with a 750ms ISI (identical presentation rates to St. John, 2015). Following the visual presentation of 6 to-be-remembered items there was a 10s delay (blank screen) before all words were re-presented on screen as labels around a centrally located question mark. Participants reconstructed the sequence by clicking on each word in the same order that they were originally presented.

#### **5.4.4 Results and Discussion**

Serial position curves for the two combinations of fluency are shown in Figure 15. Responses were scored as correct if the item was selected in the correct serial position. To assess the effects of fluency (fluent, disfluent) and serial position (1-6) on the serial reconstruction task a 2 x 6 repeated measures ANOVA was conducted. Similarly to Experiment 6, the main effect of fluency,  $F(1, 29) = 0.28$ ,  $MSE = 111.88$ ,  $p = .6$ , was not significant. This reflected similar reconstruction accuracy irrespective of fluency. Similarly to Experiment 7a, the main effect of serial position was significant,  $F(5, 145) = 15.22$ ,  $MSE = 133.76$ ,  $p < .001$ ,  $\eta_p^2 = .34$ , reflecting a general decline in accuracy across serial positions until the final serial position where there

was an improvement in accuracy. The interaction between fluency and serial position,  $F(5, 145) = 1.07$ ,  $MSE = 37.48$ ,  $p = .38$ ,  $\eta_p^2 = .04$ , was not significant suggesting that the shape of the serial position curve was similar irrespective of fluency.

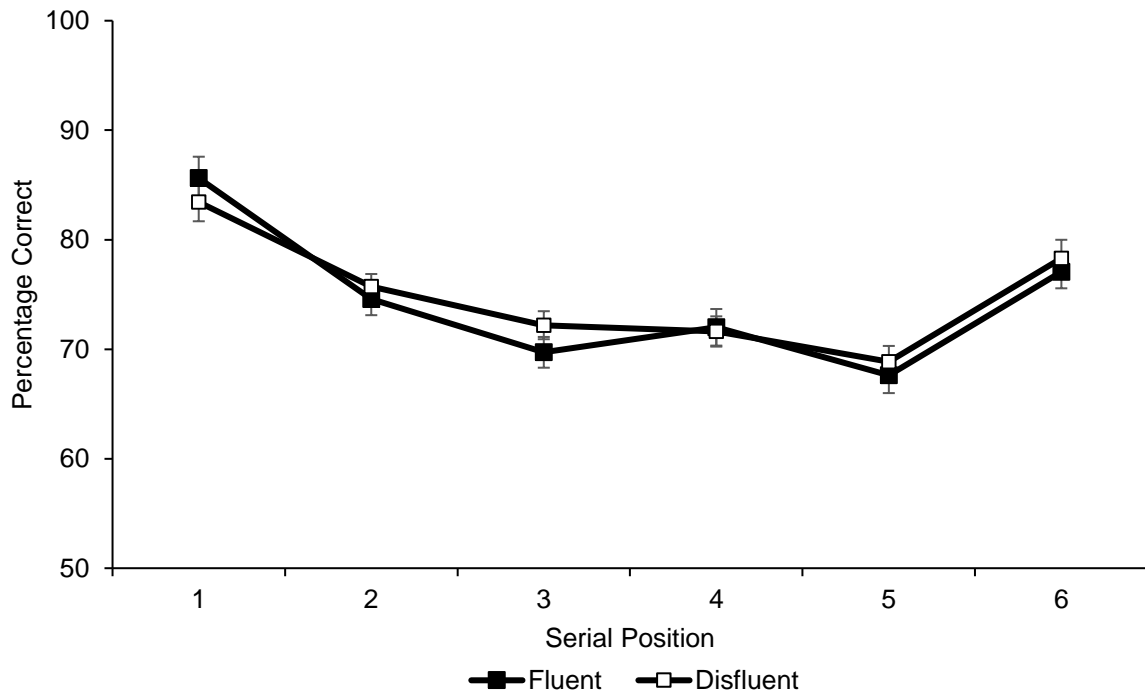


Figure 15. Mean serial position curves for recall of 6-item sequences of fluent and disfluent words. Error bars denote within-subject standard error.

Similarly to Experiment 6 and Experiment 7a, there was no support for the hypothesis that sequences containing more fluent co-articulations would afford more accurate serial reconstruction. Additionally, the slowing of presentation rate timings from 250ms (100ms ISI) per item to 750ms (750ms ISI) per item had minimal impact upon task accuracy. Therefore, it is unlikely that the faster presentation rates in Experiment 6 were the reason for the serial recognition task failing to elicit a co-articulatory fluency effect.

Returning to the significant advantage for the disfluent sequences in Experiment 7a it is possible that it was caused by an overlooked, item-based, confound. Any item-level variations in the word pools leave the results open to interpretations based upon LTM providing differential levels of support for the fluent and disfluent words. However, a similar effect was not elicited in Experiment 7b. If there were an overlooked item-based confound it would also be expected to impact

serial reconstruction for 6-item sequences when item presentation rates are slowed. There are two reasons for suggesting this. Firstly, this is because other vSTM effects such as lexicality (e.g., Macken et al., 2014), word length and the phonological similarity effect (e.g., Coltheart & Langdon, 1998) have all been found when presentation rates are at 250ms per item or even faster (as fast as 114ms per item in Coltheart & Langdon). As presentation rates slow, the sizes of those effects typically increase (e.g., Coltheart & Langdon). Secondly, while increasing the sequence length from 5 to 6 to-be-remembered items could be considered a fairly small parametric difference it does provide a slight increase in task difficulty because there are more opportunities for errors in each trial. Redintegrative accounts (e.g., Neale & Tehan, 2007) suggest that as task difficulty increases (e.g., increasing retention interval or number of to-be-remembered items) there is an increased likelihood of item representations within STM becoming degraded and therefore an increased likelihood of vSTM being reliant upon a successful redintegrative process. If, as Neale and Tehan suggest, an additional to-be-remembered item increases task difficulty then, relative to the difficulty of remembering 5 items, the increased difficulty of remembering 6 items in Experiment 7b should have meant that task performance was even more dependent on redintegration. Therefore, any overlooked item-based confounds should have had at least a similar, if not bigger, effect upon reconstruction accuracy compared to Experiment 7a. However, this was not the case.

An alternative possibility is that, despite the disfluent sequences being better remembered in Experiment 7a this was not a 'real' effect and was possibly a Type 1 error. If this is the case then the findings of all three experiments can be accommodated fairly easily by item-based accounts of vSTM (e.g., Derragh et al., 2017; Gathercole et al., 2001; Jalbert et al., 2011a, b). The word pools were matched on item-level properties and therefore no differences would be expected in serial recognition or serial reconstruction.

## **5.5 General Discussion**

The aim of Experiments 6 and 7 was to demonstrate that effects analogous to the ND effect can be elicited even when item-level variables are held constant but

where sequence-level properties such as articulatory fluency vary. Such a finding would lend weight to the possibility raised in Chapter 4 that the ND effect is possibly driven by variations in articulatory difficulty rather than varying levels of supportive activation being elicited within LTM. However, Experiment 6 failed to elicit a co-articulatory fluency effect in serial recognition irrespective of whether the to-be-remembered sequences were presented visually or auditorily. Experiments 7a and 7b also failed to elicit a co-articulatory fluency effect in visually presented serial reconstruction tasks. Some evidence was found in Experiment 7a that serial reconstruction is more accurate for disfluent sequences, but no similar effect was found in Experiment 7b. Experiment 7b, similarly to Experiment 6, found no significant difference in task accuracy between the fluent and disfluent sequences. The failure to elicit a co-articulatory fluency effect in the expected direction means that the current series of experiments provide very little support for the hypothesis that the fluency of segmental recoding has an effect upon vSTM task performance (e.g., Macken et al., 2014; Murray & Jones, 2002; Woodward et al., 2008; St. John, 2015). Instead, if Experiment 7a is considered a Type 1 error, the results can perhaps be best accommodated by item-based accounts of vSTM (e.g., Derragh et al., 2017; Jalbert et al., 2011a, b). Because the pools were matched on item-level properties then no differences in serial recognition or serial reconstruction accuracy would be expected.

The results of the current experiments raise the possibility that the co-articulatory fluency effect found by St. John (2015, Experiment 1) was simply a consequence of spoken errors during output. It may not reflect an effect based upon fluent/disfluent segmental recoding. Additionally, as discussed in the introduction to the present experiments, the co-articulatory fluency effect found in serial reconstruction (St. John, Experiment 2b) may reflect processes involved specifically with memory for non-words. There was no evidence in the current set of experiments to support the notion that accurate vSTM performance is a consequence of the fluency of segmental recoding. As such, whatever process is responsible for the ND effect has far more robust effects upon vSTM task performance than the directional co-articulatory fluency manipulation used in the present experiments does. Therefore, the usefulness of suggesting that the ND effect may be some consequence of variations in articulatory difficulty between dense and sparse neighbourhood items is called into some question. However, the directional co-

articulatory fluency manipulation used in the current experiments is just one way in which co-articulatory fluency can be manipulated. The distance the vocal apparatus must travel (e.g., Murray & Jones, 2002) or manipulating prior experience with co-articulations (e.g., Woodward et al., 2008) are just a couple of other examples. While there is relatively little evidence that the current manipulation (fluent backwards moving vs disfluent forwards moving co-articulations) can elicit a co-articulatory fluency effect and that the previous example (St. John, Experiment 1) was not just a consequence of spoken errors the same cannot yet be said for the other fluency manipulations. As such, it is important to provide a similar series of tests to alternative fluency manipulations. Systematic demonstrations that fluency is, or perhaps is not, a key determinant of vSTM task performance may go some way in furthering our understanding of vSTM. It will also allow us to question whether effects proposed to be reliant upon activation elicited within LTM (e.g., the ND effect) may be better accounted for by considering the impact that fluency has upon vSTM. It is possible, that compared to differences in articulatory difficulty between dense and sparse neighbourhood words, the directional co-articulatory fluency manipulation is simply far too subtle to demonstrate analogous effects to the ND effect whereas other manipulations may yield more robust effects of a similar magnitude to the ND effect.

A possible reason already discussed for the failure to elicit a co-articulatory fluency effect in serial recognition is that the task was not sensitive enough to detect errors at all serial positions. This was addressed in Experiment 7 by using serial reconstruction. However, both tasks are most likely to capture a mistake at the word level (e.g., intrusion, order or omission errors). An issue discussed in Chapter 3 is that serial reconstruction can be completed with only partial knowledge of the to-be-remembered word. This issue also applies to serial recognition though. Any error whereby some of the word is maintained (e.g., substituting phonemes between items) could very easily be overcome. For example, if a to-be-remembered standard sequence contains the items *beak* and *mead* and during segmental recoding there is a speech error whereby items break down at, for example, their CV/C boundary (e.g., Treiman & Danis, 1988) then the participant could end up comparing the incorrectly remembered words *bead* and *meeek* with the re-presented items. This would still allow for a successful judgement of order though. For example, the test sequence may present the items again in the same order (e.g., *beak* followed by

*mead*) and while the participant may not remember *beak* and *mead* being presented, they do remember *bead* and *meek*. These incorrect words can be used to inform a correct judgement about the order of the re-presented items (e.g., participant realises that they must have remembered *bead* instead of *beak* and then correctly judge the test sequence to be in the same order as the standard sequence). It is possible that Macken et al. (2014) were able to elicit a lexicality effect in serial recognition because the segmental recoding of words is far more fluent than the segmental recoding of non-words. As such there may have been more errors at the lexical level. The directional co-articulatory fluency effect may be far more subtle however with errors more likely to occur at the sublexical level such as the substitution of phonemes between items (e.g., Ellis, 1980; Treiman & Danis, 1988). As such, it may be far more difficult to capture those errors when using serial recognition. One way to overcome this limitation in future research would be to use a more sensitive serial recognition task (e.g., Jefferies, Frankish & Lambon-Ralph, 2006). Rather than the test sequence being the same or different to the standard sequence via the transposition of two adjacent items a more sensitive serial recognition task transposes phonemes in adjacent words instead (e.g., '*bag, rock, sun, half*' becomes '*bag, sock, run, half*'). Such a task may more successfully detect typical speech errors and be more successful in detecting co-articulatory fluency effects.

A second possible reason that the current experiments failed to elicit a co-articulatory fluency effect is that while the disfluent word pools varied in the direction of co-articulations there may have been some other overlooked difference in fluency at the sequence-level. For example, familiarising participants with the co-articulatory transitions comprising to-be-remembered items can help improve vSTM for those items (e.g., Woodward et al., 2008). It is possible that the effect in Experiment 7a was a consequence of the participants becoming overly familiarised with the co-articulations required to navigate to-be-remembered sequences drawn from the disfluent word pool. The onsets of words within the disfluent word pool began with five different anterior onset consonants whereas the onsets of words within the fluent word pool began with eight different posterior onset consonants. During the experimental procedure participants will therefore have had more experience segmentally recoding the five onset consonants and will have had more experience with the forward moving co-articulations required to navigate to those onset

consonants. The advantage for disfluent sequences in Experiment 7a (and very slight tendency for the disfluent visual items in Experiments 6 and 7b to be better remembered) could possibly reflect that increased familiarity with the disfluent co-articulations. For example, familiarisation could possibly have improved the fluency of segmental recoding for the disfluent word sequences. There may only have been a significant effect in Experiment 7a because, as already discussed, the serial recognition task used in Experiment 6 was perhaps not sensitive enough to detect an effect. Secondly, the fast presentation rates in Experiment 7a could possibly have increased the likelihood of there being segmental recoding errors compared to the slower presentation rates in Experiment 7b. Faster speech rates typically result in more speech errors (e.g., Dell, 1986; Dell et al., 1997; Fossett et al., 2016) and any unintended variations in sequence-level fluency may therefore have been more likely to emerge. St. John (2015) did find a directional co-articulatory fluency effect with the same dispersion of onset consonants though and the word pools were constructed in a similar way to the current experiments. However, despite closely following St. John's methodology for word-pool construction it is unlikely that the word pools in the current experiment were identical to St. John's word pools (St. John's materials are not publicly available). Any small differences between the word pools, such as the number of unique co-articulations, may have contributed to the advantage for disfluent sequences in Experiment 7a.

In conclusion, irrespective of the task (serial recognition or serial reconstruction), the retention interval (2500ms or 10s) or the presentation rate of to-be-remembered items (350ms per item vs 1500ms per item) the current experiments in this final empirical chapter failed to elicit a co-articulatory fluency effect. As such, the experiments were unable to demonstrate that vSTM task performance can be impacted by variations in articulatory difficulty when item-level variables are held constant. The benefit of attempting to interpret the ND effect and other vSTM effects as resulting from fluent/disfluent segmental recoding is therefore called into some question.

## Chapter 6

### General overview of the thesis, theoretical implications and discussion

#### *6.1 Aims of the thesis and general overview*

The thesis has been concerned with the item-level variable neighbourhood density and the impact that manipulating it has upon vSTM. vSTM, as measured by span tasks, serial recall and serial reconstruction has been claimed to be better when to-be-remembered words are from a dense rather than a sparse (either phonological or orthographic) neighbourhood. Typically, this advantage for dense neighbourhood words (the ND effect) is used as evidence of LTM providing support, via the activation of lexico-phonological representations, to volatile representations of target words in STM. It has been suggested that LTM supports STM via a redintegration process whereby the level of supportive activation in LTM determines the likelihood of degraded items being repaired at recall (e.g., Derragh et al., 2017; Jalbert et al., 2011a, b; Roodenrys et al., 2002). Alternatively, it has also been suggested that higher levels of activation in LTM support more robust order encoding which increases the likelihood of items being remembered in their correct serial position (e.g., Clarkson et al., 2017). The primary aim of the thesis was to gauge our current understanding of the ND effect and to consider what it reveals about vSTM.

The thesis investigated neighbourhood density by adopting several approaches. Firstly, the thesis began by exploring the parameters under which the ND effect manifests and establishing what models of vSTM can best accommodate the findings. Secondly, the thesis more closely examined the parameters used by some previous experiments that have investigated neighbourhood density. It was considered whether a first letter confound combined with alternative task completion strategies (e.g., only remembering part of the word), rather than neighbourhood density, could have been determining vSTM task accuracy in some instances. Thirdly, with exception to St. John (2015), little consideration has been given for the cause of neighbourhood density distributions and the impact this could have upon vSTM. The thesis expanded upon the work by St. John to establish whether neighbourhood density is confounded by articulatory difficulty and that, rather than neighbourhood density per se, variations in articulatory difficulty may be responsible



for the ND effect. Finally, by varying the articulatory difficulty of to-be-remembered sequences an attempt was made to demonstrate that effects analogous to the ND effect can still be elicited even when neighbourhood density is controlled. The following section provides a brief overview of each empirical chapter and the key findings.

## **6.2 Summary of empirical chapters and key findings**

### **6.2.1 Chapter 2 – Establishing the parameters under which the ND effect manifests**

Presentation modality, word pool size and memory task type were identified as factors requiring further exploration before more robust claims about the precise way in which neighbourhood density impacts vSTM can be made. Across 4 experiments the impact of neighbourhood density on vSTM for 6-item sequences as a function of task type (serial recall vs serial recognition), modality of item presentation (auditory vs visual) and the size of the pool from which the sequences were drawn (48 vs 12) was tested. In Experiment 1 there was a robust advantage for dense neighbourhood words compared to sparse neighbourhood words in both auditorily and visually presented serial recall. The key finding is that the size of the dense neighbourhood advantage was similar irrespective of presentation modality. This supports claims (e.g., Roodenrys et al., 2002; Allen & Hulme, 2006) that whatever memory process drives the ND effect in vSTM is most likely distinct from processes occurring during encoding whereby the processing speed of dense and sparse neighbourhood words varies as a function of modality (e.g., Luce & Pisoni, 1998; Yates et al., 2004).

In Experiment 2, the dense neighbourhood advantage was eliminated in visually and auditorily presented serial recognition. The finding mirrors those of other linguistic effects (e.g., the lexicality effect, word length effect) whereby there is typically a robust effect in serial recall but a smaller or eliminated effect in serial recognition (e.g., Baddeley et al., 2002; Gathercole et al., 2001). The attenuation or absence of such linguistic effects in serial recognition is typically used to highlight the distinct but interactive nature of LTM and STM. The availability of pre-existing representations in LTM is considered important for serial recall because they can assist with reintegration. When items are re-presented in serial recognition the

requirement for LTM is obviated as the re-presented items can repair any degraded traces in STM (e.g., Gathercole et al., 2001). However, in Experiment 3, when the size of the dense and sparse neighbourhood word pools was reduced, serial recall was better for the sparse rather than the dense neighbourhood words. A non-significant interaction between modality and neighbourhood density suggested that once again the effect was not differentially affected by modality. Additionally, this same advantage for sparse neighbourhood words was also found in Experiment 4 when using visually and auditorily presented serial recognition.

Experiments 3 and 4 raise doubts over interpreting the ND effect as some consequence of supportive activation being elicited within linguistic networks in LTM. In these experiments, the dense neighbourhood words should still arguably have elicited more activation in LTM than the sparse neighbourhood words. The dense neighbourhood words should therefore have provided more supportive activation to either assist with a redintegrative process (e.g., Derragh et al., 2017; Jalbert et al., 2011a, b; Roodenrys et al., 2002) or assist with the encoding of order information (e.g., Clarkson et al., 2017). The results of the 4 experiments in Chapter 2 raise questions about the viability of typical accounts of the ND effect that invoke support from LTM whereby the level of supportive activation within lexical neighbourhoods determines the likelihood of successful vSTM. Instead, the results indicate a critical role for a variety of task factors such as word pool size and task type in modulating the ND effect.

### **6.2.2 Chapter 3 – The ND effect in Serial Reconstruction tasks: A consequence of Neighbourhood Density, Orthographic Frequency or a First Letter Confound?**

Chapter 2 revealed the ND effect to vary as a function of task type and word pool size. This prompted a closer inspection of some previous experiments that have investigated the ND effect. A commonly used vSTM task in the ND effect literature is serial reconstruction whereby all to-be-remembered items are re-presented at test and participants are required to click on the items in their originally presented order. However, a potential problem with the conclusions drawn from serial reconstruction is that it can be accurately completed with only partial knowledge of the to-be-remembered words. For example, by only remembering the first letter of each to-be-remembered word (a first letter strategy) that first letter can still successfully be used

as a cue in which to decide the correct order that the items should be clicked. A closer look was taken at the word pools used by Jalbert et al. (2011b) and Guitard, Saint-Aubin, et al. (2018) because their dense and sparse word pools were not matched on the number of onset letters. This was achieved by running a series of simulations that modelled serial reconstruction performance across their experiments with the assumption that participants solely adopted a first letter strategy.

The key finding in Chapter 3 was that the simulation results very closely resembled the pattern of results found originally by Jalbert et al. (2011b) and Guitard, Saint-Aubin, et al. (2018). The simulation results are important because the effects found in the two papers have been used as a basis to make strong claims about the nature of vSTM. For example, Jalbert et al. (2011b) found that dense neighbourhood non-words were better remembered even when the dense neighbourhood items were long non-words and the sparse neighbourhood items were short non-words. This demonstrated that neighbourhood density could override the word length effect and is possibly the cause of previous examples of the word length effect (long words typically come from sparser neighbourhoods than short words). Their interpretation of this finding was that articulatory rehearsal cannot be a causal mechanism of vSTM because long words will take longer to articulate and therefore be more prone to decay than short words. Guitard, Saint-Aubin, et al. reviewed the materials used by Jalbert et al. (2011b) and suggested that their findings were likely a consequence of the short non-words from a sparse neighbourhood having unusual orthographic structures. Across a series of 8 experiments a task advantage for long words compared to short words was only found when the long words were comprised of more typical orthographic structures than the short words. In all other instances, vSTM was better for the short words. However, whether caused by neighbourhood density or orthographic structure any situation where there is an advantage for long words over short words still raises doubts over articulatory rehearsal being a causal predictor of vSTM. The long words will take longer to articulate and therefore be more prone to decay (e.g., Baddeley, 1975). However, more familiar orthographic structures could possibly afford more fluent segmental recoding (e.g., Jones & Macken, 2018; Macken et al., 2015).

While it is impossible to know exactly what strategies participants adopted in the studies conducted by Jalbert et al. (2011b) and Guitard, Saint-Aubin, et al. (2018) a first letter strategy is one that is acknowledged in the literature and where

steps, such as using phonologically similar onset consonants (e.g., Baddeley et al., 2002), are sometimes taken to reduce the likelihood of participants using such a strategy. Morrison et al. (2016) also highlight that a single vSTM task is usually completed in a variety of different ways by different participants (e.g., in serial recall tasks participants report using mental imagery, grouping and rehearsal strategies). As such, there is a possibility that some participants may have adopted a first letter strategy rather than attempting to remember all the words in their entirety. Additionally, even if participants attempt to remember the words in their entirety, naturally occurring speech errors (e.g., Ellis, 1980; Treiman & Danis, 1988) may lead to participants remembering an incorrect word beginning with the correct onset letter. This onset letter could then be used as a cue in which to decide the item to click. Because no steps were taken by Jalbert et al. (2011b) or Guitard, Saint-Aubin, et al. to reduce the likelihood of participants using a first letter strategy and because very similar robust effects would have emerged had participants used the first letters to complete the serial reconstruction tasks this raises some doubts over what their data reveals. Some participants relying upon the first letter to complete the vSTM tasks provides an alternative explanation for their findings that is not reliant upon theoretical processes such as reintegration or decay. Until similar effects are found in tasks whereby alternative strategies will not yield equivalent effects it would seem best to remain cautious before ruling out some causal role for rehearsal upon the ND effect.

### ***6.2.3 Chapter 4 - Exploring the Relationship Between Neighbourhood Density and Articulatory Difficulty: Effort-Based and Duration-Based Measures of Articulatory Difficulty***

The results of Chapter 2 raised doubts over interpretations of the ND effect that invoke some level of supportive activation in LTM and Chapter 3 highlighted that, despite the claims made by Jalbert et al. (2011b), a causal role for rehearsal in explaining the ND effect might not yet have been ruled out. As such, a possible alternative explanation for the ND effect was sought. With exception to St. John (2015), very little consideration has been given to the cause of phonological neighbourhood distributions and the possibility that whatever has shaped the distributions may also play some role in vSTM performance. Chapter 4 investigated whether phonological neighbourhood distributions are possibly a consequence of

pressures to minimise articulatory effort during the natural development and evolution of language (e.g., Lindblom, 1990). Sparse neighbourhood words may have fewer neighbours because generally they consist of more effortful articulations than dense neighbourhood words and so the speech sounds comprising those sparse neighbourhood words have tended to be avoided. Study 1 tested this prediction by quantifying the articulatory effort required for the 24 English consonant sounds by scoring them on their relative difficulty based upon three dimensions – their manner of articulation, their place of articulation and their voicing. These scores were then applied to the onset and offset of English CVC words. A regression analysis revealed that CVC words from denser neighbourhoods tend to consist of easier articulations than CVC words from sparser neighbourhoods. This raises the possibility that previous examples of the ND effect are due to variations in articulatory difficulty across the dense and sparse neighbourhood word pools. Lending additional weight to this possibility was the finding that several dense and sparse neighbourhood CVC word pools used by researchers to test the ND effect also differ in articulatory difficulty. Dense neighbourhood word pools have tended to consist of words with less effortful articulations than the words contained in the sparse neighbourhood word pools.

Experiment 5 measured the time taken to vocalise dense and sparse neighbourhood words in isolation and within sequences. This not only provided a measure of articulatory difficulty at the item-level but also at the sequence-level (e.g., Woodward et al., 2008). Dense neighbourhood words were spoken faster both within sequences and in isolation. This is the first demonstration that items contained within sequences of dense neighbourhood words are vocalised quicker than items contained within sequences of sparse neighbourhood words. Overall, the possibility raised by Chapter 4 is that the ND effect may be some consequence of differences in articulatory difficulty between dense and sparse neighbourhood words. Dense neighbourhood words not only tend to consist of easier articulations but also tend to be vocalised more quickly than sparse neighbourhood words.

#### **6.2.4 Chapter 5 – Can manipulating articulatory difficulty produce an analogous effect to the ND effect?**

Chapter 4 established that neighbourhood density is confounded by articulatory difficulty (dense neighbourhood words typically require less articulatory effort and are quicker to vocalise than sparse neighbourhood words) but the relationship between neighbourhood density and articulatory difficulty is purely correlational. As such, the relative impact that articulatory difficulty has upon the ND effect is unclear. In order to demonstrate that vSTM effects such as the ND effect may be driven by variations in articulatory difficulty it is important to demonstrate an equivalent effect when neighbourhood density is controlled for, but where articulatory difficulty is free to vary. In Chapter 5, across three experiments, articulatory difficulty was varied by manipulating the direction of co-articulations between items (e.g., St. John, 2015) in to-be-remembered sequences. Backwards moving co-articulations involve an increased degree of co-articulatory overlap and can be considered more fluent, requiring less articulatory effort, than forwards moving co-articulations (e.g., Chitoran et al., 2002). It was tested whether vSTM would be better for to-be-remembered sequences consisting of fluent backwards moving rather than disfluent forwards moving co-articulations in serial recognition (Experiment 6) and serial reconstruction (Experiments 7a & 7b). However, no difference in task accuracy was found in Experiments 6 or 7b and in Experiment 7a serial reconstruction was better for the sequences containing disfluent co-articulations. A failure to demonstrate that sequences of words comprised of more fluent co-articulations yield an effect equivalent to the ND effect raises some doubt over the usefulness of suggesting that articulatory difficulty, rather than neighbourhood density, is a possible cause of the ND effect.

#### **6.3 Overall Implications - How reliable is the ND effect?**

Across the thesis the results have raised questions over the interpretations of several previous experiments demonstrating the ND effect. Table 5 shows each of the studies that have reported the ND effect and highlights whether the issues raised in the present thesis apply to those studies.

Table 5

*A summary of experiments that have demonstrated the ND effect and whether they are susceptible to the issues identified in this thesis.*

Experiment	vSTM task and confound		
	Task	Modality	Confound?
Roodenrys et al. (2002, Experiment 1)	Span Task	Auditory	Articulatory Effort
Roodenrys et al. (2002, Experiment 3)	Span Task	Auditory	Articulatory Effort
Allen and Hulme (2006, Experiment 2)	Serial Recall	Auditory	Articulatory Effort
Jalbert, et al. (2011a, Experiment 2)	Serial Reconstruction	Visual	Articulatory Effort
Jalbert et al. (2011b, Experiment 1, 2 & 3)	Serial Reconstruction	Visual	First letter
Clarkson et al. (2017, Experiment 2 and 3)	Serial Reconstruction	Visual	Articulatory Effort
Derraugh et al. (2017, Experiment 1 and 2)	Serial Reconstruction	Visual	? (Multiple word formats used)
Guitard, Gabel et al. (2018, Experiment 7)	Serial Reconstruction	Visual	z-transformed constrained trigram count measure

As can be seen from Table 5 there are only two studies that have found the ND effect and not susceptible to the criticisms developed in this thesis. Chapter 4 highlighted a tendency for sparse neighbourhood CVC words to require more effortful articulations and take longer to vocalise than dense neighbourhood CVC words. It is possible that a similar pattern would emerge in experiments (i.e., Derraugh et al., 2017; Guitard, Gabel, et al., 2018) using dense and sparse neighbourhood words with different formats (e.g., VCC words or those containing consonant blends). Until the articulatory difficulty scoring system developed in Chapter 4 is expanded to include different word formats though that is purely speculative. However, given the size of the word pools (282 items) used by Derraugh et al. (2017) it is possible that they may also vary on other item or sequence-level properties that they did not control for. The dense and sparse neighbourhood word pools used by Guitard, Gabel, et al. (2018) could also vary on articulatory difficulty

but more definitively they report that the materials vary on z-transformed constrained trigram counts. This variable is calculated by first taking trigram counts of each word (the frequency at which groups of three consecutive letters within a word appear in other words within a given language). A z score is then calculated for each word (the difference between the trigram value of a target word and the overall mean divided by the standard deviation) and then the overall average of the z scores is calculated to produce a mean z-transformed constrained trigram count for a particular word pool. Whether this variable can elicit an effect of comparable size to the ND effect is not known but it does mean that it cannot currently be claimed with any certainty that neighbourhood density, rather than z-transformed constrained trigram counts, was driving vSTM performance in their experiment.

While the ND effect is currently considered as being robust, evidenced by it being used to make strong claims about the nature of vSTM, the interpretation of the effect is brought into some question by the issues raised within the present thesis. Until, at the very least, the ND effect is demonstrated in a variety of vSTM tasks when onset letter distributions within word pools, articulatory difficulty and item-level variables are all controlled for, it would seem best to remain more cautious before using the ND effect to draw conclusions about the nature of vSTM.

#### **6.4 Theoretical Implications: Models of memory**

The key findings of the thesis are difficult for any of the models of memory introduced in the thesis to accommodate. In the following section these difficulties are discussed.

##### **6.4.1 Redintegrative accounts of vSTM**

Current redintegrative accounts of the ND effect propose that words from a dense neighbourhood elicit more supportive activation within linguistic networks in LTM. This increases the likelihood of any degraded traces in STM being repaired (e.g., Derragh et al., 2017; Jalbert et al., 2011a, b; Roodenrys et al., 2002). However, in Experiments 3 and 4 there was a vSTM advantage for words from a sparse neighbourhood. According to redintegrative accounts of the ND effect, sparse neighbourhood words should elicit less supportive activation than dense



neighbourhood words and therefore be less likely than dense neighbourhood words to undergo a successful redintegration process.

The findings of the present thesis are particularly problematic for models of vSTM incorporating both trace decay and redintegration (e.g., Burgess & Hitch, 1999, 2006; Page & Norris, 1998). According to Study 1 and Experiment 5 sparse neighbourhood words, as well as supposedly eliciting less activation in LTM, also consist of more effortful articulations and are slower to vocalise than dense neighbourhood words. This means that sparse neighbourhood words are likely to take longer to rehearse and therefore be subject to higher levels of decay than dense neighbourhood words. It is difficult to explain why sparse neighbourhood words would then be better remembered if, not only are they likely to decay more than dense neighbourhood words, but also likely to elicit less supportive activation in LTM. There are other examples whereby vSTM accuracy is not dependent on rehearsal time though. For example, Jefferies, Frankish, and Noble (2011) demonstrated a reversal of the word length effect when the short items were comprised of non-words but the long items comprised of words. Within a trace decay framework it is possible that, while the long words were more susceptible to decay than the short non-words during rehearsal, it was only the long words that were able to benefit from LTM providing redintegrative support (see, e.g., Lewandowsky & Farrell, 2000 for similar suggestion). This would have left the short non-words decayed to some extent but unable to be repaired and the long words decayed to a further extent but able to be repaired via redintegration. A similar suggestion could be made for the results of Experiments 3 and 4. The word pools may have varied on some overlooked property that was able to override any decay of the sparse neighbourhood words but less able to repair any decayed dense neighbourhood words. However, when words vary on length but are equated on neighbourhood density the word length effect disappears (e.g., Jalbert al., 2011a). Within a trace decay framework, the long words in that experiment should have been more susceptible to decay than the short words while receiving equivalent levels of redintegrative support at recall. If both trace decay and redintegration determine whether items are recalled correctly from vSTM, then the word length effect should not have attenuated in that instance. Such findings, along with the results of Chapter 2, raise doubts over the usefulness of suggesting that both redintegration and decay offset by rehearsal impact vSTM.

If there were some overlooked item-level variable in Experiments 3 and 4 then the findings are perhaps easier to accommodate within redintegrative frameworks that eschew the notion of decay offset by rehearsal (e.g., Nairne, 1990). Repeatedly drawing words from smaller word pools may have activated the lexical representations of all to-be-remembered words in LTM to a similar level (e.g., Goh & Pisoni, 2003). As such, the level of redintegrative support provided would be similar for both the dense and sparse neighbourhood words. This may have attenuated or eliminated the ND effect and provided a situation whereby some other activation within LTM was able to provide more support to the sparse neighbourhood words. However, the ND effect has also been demonstrated elsewhere when using relatively small word pools (16 items; e.g., Jalbert et al., 2011a, b) so such an account seems unlikely or would at least require identification of some other item-level variable that provided more support to the sparse neighbourhood words in the present experiments but more support to the dense neighbourhood words in Jalbert et al. (2011a, b). However, this approach would then mean that neighbourhood density was possibly not causing the effects found by Jalbert et al. (2011a, b). Alternatively, the activation in LTM elicited by dense and sparse neighbourhood words may still have varied but some other, even more powerful, overlooked variable may then have been able to override that activation. While possible, the problem with this type of speculation is that if variables exist that are powerful enough to reverse the ND effect (imageability has also reversed the ND effect; e.g., Goh & Pisoni, 2003) then it raises questions over the relative usefulness of using neighbourhood density as a variable to help to explain vSTM.

Experiment 7a demonstrated an advantage for disfluent sequences even when the to-be-remembered word pools were controlled for item-level variables. Again, similarly to the results of Experiments 3 and 4, one way to accommodate the findings within a redintegrative framework is to suggest that the materials varied on some other overlooked item-level property. However, despite increasing the number of to-be-remembered words from 5 to 6 there was no significant effect found in Experiment 7b. As task difficulty increases (e.g., by increasing the number of to-be-remembered items) it is suggested that there is an increased likelihood of item representations becoming degraded and therefore an increased reliance upon redintegration (e.g., Neale & Tehan, 2007). Additionally, other vSTM effects such as word length and phonological similarity typically get bigger, not smaller, when

presentation rates are slowed (e.g., Coltheart & Langdon, 1998). As such, if there were some item-level confound in Experiment 7a then a similar, if not bigger, effect should also have been elicited in Experiment 7b when the presentation rates were slowed. A possible way to accommodate the findings within a redintegrative framework is to suggest that Experiment 7a may have been a Type 1 error and across Experiments 6, 7a and 7b there was no true effect because the word pools did not vary on any item-level properties.

Overall, the results are difficult for any models of memory that incorporate a redintegrative stage reliant upon the level of activation in LTM. However, while speculative, the results could possibly be accommodated if some overlooked item-level confound in Experiments 3 and 4 was able to override the activation elicited within lexical neighbourhoods. The results are particularly problematic for models incorporating trace decay offset by rehearsal though and call into question the value of incorporating it as an explanatory mechanism for vSTM. Even if the word pools used in Experiments 3 and 4 varied on some other item-level confound the sparse neighbourhood words are likely to have taken longer to rehearse and therefore have been more prone to decay.

#### **6.4.2 Order based account of the ND effect**

The ND effect was robust in serial recall (Experiment 1) but attenuated in serial recognition (Experiment 2). Clarkson et al. (2017) suggested that the ND effect may be some consequence of LTM supporting more robust order encoding for the dense neighbourhood words. However, because this account suggests that the ND effect is one upon order memory it would predict an ND effect in a serial recognition task that is considered to be a test of order memory (e.g., Gathercole et al., 2001). Models of order memory (e.g., Burgess & Hitch, 1999, 2006; Page & Norris, 1998) incorporating a late redintegrative stage would predict no ND effect in tasks that are a pure test of order memory. This is because items (regardless of whether they are partially degraded) are retrieved in order from STM prior to LTM having any influence upon the items. Re-presented items can repair all degraded traces in a serial recognition task but LTM is thought to be required to repair degraded traces in serial recall. In the second instance the level of supportive activation would determine the success of the redintegrative process. However, both models incorporate a decay

process offset by rehearsal and suffer from the issues discussed in Section 6.4.1. Additionally, if activation in LTM helps to support order encoding then there should not be a reversal of the ND effect when the word pool size is reduced. This is because activation in LTM should still be greater for the dense neighbourhood items than it is for the sparse neighbourhood items.

Similarly to the accounts discussed in Section 6.4.1, the reversal of the ND effect could possibly be considered some consequence of an overlooked item-level confound but it would first need to be explained why the ND effect did not appear in Experiment 2 when using serial recognition. The advantage for disfluent sequences in Experiment 7a could also once again be considered a Type 1 error because the word pools did not differ on item-level variables.

### **6.4.3 Object-oriented account of the ND effect**

vSTM task performance is best for sequences affording the most fluent segmental recoding (e.g., Macken et al., 2014; Murray & Jones, 2002; Woodward et al., 2008). Therefore, words from dense neighbourhoods are likely to be better remembered because, according to Chapter 4, they generally consist of less effortful articulations and can be vocalised faster (i.e., more fluently) than words from sparse neighbourhoods. The segmental recoding and later reproduction of words from a dense neighbourhood should therefore be more fluent than it would be for sparse neighbourhood words. However, in Experiments 3 and 4, vSTM was better for the words from a sparse neighbourhood. This is difficult to reconcile under an object-oriented account because the dense neighbourhood words should still have afforded more fluent segmental recoding than the sparse neighbourhood words.

One possibility discussed in Section 2.7 was that, while the dense and sparse neighbourhood word pools in Experiments 3 and 4 were matched on the number of unique onset consonants that appeared in each word pool, they were not matched on the number of phoneme onsets. The fewer unique phoneme onsets in the dense neighbourhood word pool may have meant that the words within the dense neighbourhood to-be-remembered sequences were more likely to act as lexical competitors with each other (e.g., Sevald & Dell, 1994) and this could possibly have led to the advantage for the sparse neighbourhood words. It is also important to consider that differences in articulatory difficulty and fluency apply at both the item

and sequence-level. While the dense neighbourhood words used in Experiments 3 and 4 may have afforded more fluent segmental recoding in isolation (as suggested in Experiment 5 with dense neighbourhood words having faster vocalisation times) there may have been some particular combination of the transitions between items that was more fluent among the sparse items. For example, co-articulatory fluency can vary as a function of the distance travelled by the vocal apparatus between the onset and offset of to-be-remembered items (e.g., Murray & Jones, 2002) the direction of the co-articulations (e.g., St. John., 2015) or as a consequence of familiarising participants with the co-articulatory transitions comprising the to-be-remembered items (e.g., Woodward et al., 2008). Co-articulatory fluency was manipulated in Chapter 4 by altering the direction of co-articulations between items. Ease of fluency did not predict performance even when item-level variables were matched so it seems unlikely that directional differences in co-articulation would be able to reverse the ND effect. However, while speculative, it is possible that other variations in co-articulatory fluency, such as the distance travelled by the vocal apparatus or familiarisation with co-articulations, were able to exert a stronger influence than neighbourhood density in Experiments 3 and 4. Some other variation in fluency may also have led to the serial reconstruction advantage for the disfluent sequences in Experiment 7a. Although no effect emerged in visually presented serial recognition (a task arguably requiring segmental recoding; e.g., Macken et al., 2014), or in Experiment 7b, possibilities already discussed in Section 5.5 are that the serial recognition task was perhaps not sensitive enough to detect a co-articulatory fluency effect in Experiment 6. Additionally, if there were any other differences in fluency then the fast presentation rates in Experiment 7a may have increased the likelihood of there being speech errors during segmental recoding, and therefore increased the likelihood of an effect emerging, compared to the slower presentation rates of Experiment 7b.

Asides from speculative explanations of performance, what is of particular importance is that, while the results of the present thesis do not readily map onto an object-oriented account of vSTM, there is perhaps more flexibility than item-based accounts in that a number of factors beyond the item-level variables can also be considered and explored in future research. Performance in a vSTM task is considered to not only be a consequence of the to-be-remembered materials but also a consequence of skills possessed by the participant (e.g., perceptual-motor

skills, linguistic ability) and the specific task (e.g., Jones & Macken, 2018; Macken et al., 2015). Performance is a dynamic outcome of these factors and modifications to one factor can modify the influence of each of the others. Any modification (e.g., word pool size, presentation rates, materials) could have had an unexpected influence upon the other factors and yielded what may initially seem like incompatible results for an object-oriented account.

#### **6.4.4 Psycholinguistic models of vSTM**

As an alternative to the models of memory discussed so far in the thesis there are also psycholinguistic models of vSTM (e.g., Martin and Saffran, 1997; Martin and Gupta, 2004). These identify vSTM performance as being based upon the temporary activation of long-term verbal codes that underpin general language processing. For example, Martin and Gupta (2004) suggest that, rather than to-be-remembered items accessing some specialised store, order of to-be-remembered items is maintained via *sequence memory* which temporarily stores the pattern of activation elicited within the linguistic system. Rather than activation just being elicited at the lexical level (e.g., only among phonological/orthographic neighbours) each to-be-remembered item (e.g., in a serial recall task) elicits temporary activation at the lexical level which then also spreads into corresponding semantic layers (e.g., ‘dog’ can activate its lexical neighbours as well as activating semantic features such as ‘pet’, ‘labrador’, ‘animal’, ‘fur’ etc.). Activation in the semantic layer feeds back down to the lexical layer and helps to sustain activation at the lexical level. Temporary learning of the connection weights between each word at the lexical level (i.e., the serial order) is maintained in sequence memory. Successful retrieval of the items is dependent upon the accuracy in which sequence memory *replays* the original sequence of activation. Assuming the connection weights have not decayed then the sequence will be correctly replayed, and recall will be correct. However, decayed connection weights will lead to an increased possibility of an incorrect replay and incorrect recall. Under this account of vSTM the ND effect could perhaps be considered a consequence of the connection weights being better specified for dense neighbourhood words (because of increased activation in the lexical layers) compared to sparse neighbourhood words and therefore being less likely to decay. A similar process can also account for various other vSTM effects. For example,

lexicality effects might emerge because non-words will have no pre-existing lexical representations in which to develop a well specified sequence of activation at the lexical level. Similarly, the less frequently a word is used then the weaker their lexical representations will be and again the less specified the sequence of activation will be at the lexical level compared to more frequently used words.

A psycholinguistic account may initially seem to be unable to accommodate the findings of Experiments 3 and 4 whereby there was a reversal of the ND effect. However, in much the same way that successful redintegration is proposed to be reliant upon the level of activation in LTM so is the specification of connection weights. The word pools were controlled for imageability (a variable proposed to modulate the semantic system; e.g., Sabsevitz, Medler, Seidenberg & Binder, 2005). However, other semantic features not considered or controlled for, such as word pleasantness, have also been shown to modulate the semantic system and impact vSTM (e.g., Monnier & Syssau, 2008). Any overlooked semantic features of the to-be-remembered materials could possibly have fed back enough supportive activation into the lexical layers to override the higher levels of activation that dense neighbourhood words usually benefit from.

## ***6.5 Further directions and considerations***

The difficulty for models of vSTM to accommodate all the findings in the present thesis means that future research into neighbourhood density will need to examine several factors far more closely. This not only relates to factors highlighted in the thesis (e.g., type of task, word pool size, articulatory difficulty) but also factors that were beyond the scope of the present thesis. Some suggestions and directions for future research are outlined below.

### ***6.5.1 Expanding the articulatory difficulty scoring system and using it to create word pools***

While Chapter 4 demonstrated that dense neighbourhood CVC words tend to consist of more difficult articulations the scoring system did not provide difficulty ratings for dense and sparse neighbourhood words with alternative formats (e.g., VCC, CCV or longer). It is possible that, while sparse neighbourhood CVC words tend to consist of more difficult articulations, sparse neighbourhood VCC or CCV words may not. Such a finding would undermine the possibility that the ND effect

may be driven by differences in articulatory difficulty. A future version of the scoring system should attempt to include these word formats and establish whether all sparse neighbourhood words tend to consist of more effortful articulations. Additionally, future research may also be able to use a version of the scoring system to factorially manipulate neighbourhood density and articulatory difficulty. If articulatory difficulty, rather than neighbourhood density, is responsible for the ND effect then it should be possible to elicit a vSTM advantage for sparse neighbourhood words over dense neighbourhood words if those sparse neighbourhood words contain easier articulations. Similarly, it should also be possible to eliminate any effect if the dense and sparse neighbourhood words are comprised of similar articulatory difficulty.

### ***6.5.2 Assessing whether different manipulations of co-articulatory fluency can override neighbourhood density***

While Chapter 5 failed to demonstrate a vSTM advantage for fluent backwards moving co-articulations it is only one manipulation of co-articulatory fluency and it may have been too subtle to elicit an analogous effect to the ND effect. Another manipulation is the distance travelled by the articulatory apparatus between the offset of one word and onset of the next (e.g., Murray & Jones, 2002). The less distance travelled then the more co-articulatory overlap there is and the more fluent the co-articulations are. Another possible way to test whether articulatory difficulty has some role in the ND effect is to manipulate co-articulatory fluency while also manipulating neighbourhood density. Dense neighbourhood words seem to be comprised of less effortful articulations (Study 1a) and word pools manipulating neighbourhood density have also manipulated articulatory difficulty (Study 1b). As such, less effortful articulations rather than neighbourhood density per se may be responsible for the ND effect. However, to further sustain this suggestion, it may be possible to attenuate, or even reverse the ND effect, if the to-be-remembered dense neighbourhood words are arranged in such a way that the co-articulations are more effortful (e.g., greater distance between offsets and onsets) than the co-articulations required for to-be-remembered sequences of sparse neighbourhood words.



### ***6.5.3 Separating the potentially independent contributions of phonological and orthographic neighbourhood density***

One possible weakness in the thesis, and in particular to the experiments conducted in Chapter 2, is that neighbourhood density (both orthographic and phonological) was manipulated. Word pools from a dense phonological neighbourhood were also from a dense orthographic neighbourhood and vice versa. Currently, because the two variables are highly correlated (e.g., Mulatti et al., 2003), this is an issue in all the vSTM literature investigating the ND effect though. As such, even if the ND effect is a consequence of supportive activation being elicited within linguistic networks in LTM, it is not clear whether that supportive activation is elicited in orthographic networks, phonological networks or perhaps both. In a lexical decision task, Grainger, Muneaux, Farioli, and Ziegler (2005) found that when visually presented words were from a sparse orthographic neighbourhood participants were fastest to respond when the words were from a dense rather than a sparse phonological neighbourhood. However, when words were from a dense orthographic neighbourhood participants were fastest to respond to words from a sparse rather than a dense phonological neighbourhood. Such a factorial design has not yet been used when testing vSTM but it may also be the case that similar interactions occur in vSTM tasks. This would rule out a simple interpretation of the ND effect solely being based upon a word's phonological or orthographic neighbourhood. Instead, it might need to be considered that both networks play some supportive role.

### ***6.5.4 Considerations beyond orthographic and phonological neighbourhoods***

A further consideration is that the ways in which words are suggested to interact within linguistic networks in LTM is constantly undergoing development. A more recent conception to phonological neighbourhood density measures is the clustering coefficient (e.g., Chan & Vitevich, 2009). This measure not only includes a count of how many phonological neighbours a word has but also includes a count of how many of the neighbours are phonological neighbours with each other. Unlike phonological neighbourhood density whereby auditory recognition is inhibited (e.g., Luce & Pisoni, 1998) but visual recognition is facilitated (e.g., Yates et al., 2004) words with a low clustering co-efficient are recognised faster than words with a high clustering co-efficient when presented visually (Yates, 2013) and auditorily (Chan &

Vitevitch, 2009). Because the clustering co-efficient does not indicate the existence of opposing effects during the perception of auditory and visual material it may provide a better measure in which to understand vSTM. There has been some use of the clustering co-efficient in vSTM research with Vitevitch, Chan and Roodenrys (2012) finding better serial recall for words with a high compared to a low clustering co-efficient. Further use of the clustering co-efficient in a variety of vSTM tasks may provide further understanding of how lexical representations of words interact and what effect such interactions have upon vSTM.

#### **6.5.5 Completely forget the Neighbourhood: Modelling the degree of overlap between experimental materials and participants' prior knowledge**

Throughout the thesis there have been examples that highlight the importance of giving greater consideration to the memory task and the particular materials being used. One emerging line of research that addresses some of these issues, and that may help with our understanding of vSTM, is the CLASSIC model of language acquisition and development. CLASSIC is an associative learning model whereby learning is based upon the chunking of incrementally larger amounts of verbal information (e.g., Jones & Macken, 2015, 2017). Modelling begins at the phoneme level and the more exposure the model has to repeated combinations of phonemes then the more likely those phoneme combinations will be chunked into a larger unit. Frequently presented phoneme combinations, words and phrases will eventually be combined into single chunks. So, for example, the word sequence 'what time is dinner?' would initially be comprised of all the individual phonemes making up that sequence and because the model has no prior knowledge of the sequence it could be considered difficult to produce in speech. Repeated exposure to sets of phonemes comprising that sequence, across different phrases and sentences ('*what time does the match start?*', 'I don't know *what time* it is', 'it *is dinner*', 'when *is dinner?*' etc.), cause the phonemes to chunk beyond the word-level (e.g., '*what time*' and '*is dinner*'). The word sequence 'what time is dinner?' could now be considered to comprise of only two chunks and therefore easier to produce in speech because, despite the model still not having knowledge of the entire sequence, it does have knowledge of the two chunks comprising the sequence. Finally, if the model is repeatedly exposed to '*what time is dinner?*' then this results in one single sequence-level chunk that subsumes all phonemes and words comprising that sequence. This

is now a fully learnt sequence and therefore could be considered easier to produce in speech compared to sequences comprised of more chunks.

Regarding vSTM task performance, rather than considering word pools as varying on item-level properties, they could instead be considered as varying on a continuum that relates to the participants' prior linguistic experience (e.g., Jones & Macken, 2015, 2017). According to this account, the more experience someone has with particular chunks of phonemes, words and phrases then the more readily they will map onto their speech motor planning processes and therefore sustain better vSTM. The word sequence 'What time is dinner?' would be best remembered in a vSTM task when it is comprised of as few chunks as possible. While item-level frequency and other item-level properties such as phonotactic probability or n-gram measures are usually controlled for in the neighbourhood density and wider vSTM literature, properties such as the frequency of phoneme combinations at co-articulatory boundaries and across sets of items are not controlled for. What may appear to be a seemingly random sequence of to-be-remembered words could contain some very common chunks. In that respect the reversal of the ND effect in Experiments 3 and 4, and advantage for disfluent sequences in Experiment 7a, could possibly have resulted from the sparse neighbourhood word pool and disfluent word pool containing more familiar phoneme chunks. While it is very speculative in relation to the results of the current thesis, it does potentially provide a fruitful line of research in which to further understand why neighbourhood density impacts vSTM.

CLASSIC has proved successful elsewhere in modelling widely used measures of STM capacity such as digit span, non-word repetition and sentence recall (Jones & Macken, 2017). Additionally, by manipulating sequences of to-be-remembered digits so that they are either comprised of digit pairs that appear frequently or infrequently in the British National Corpus (e.g., *two* regularly appears in digit sequences whereas *nine* occurs less frequently) digit-span has been shown to be superior for sequences containing more frequent digit pairs (Jones & Macken, 2015). The more familiar a digit combination is for a participant then the more readily it will map onto speech planning processes and the better vSTM will be. By modelling the materials used to demonstrate the ND effect (and other linguistic effects) it may be possible to establish whether the word pools vary in the frequency of exposure at not only the item-level but also at the sequence-level.

## **6.6. From theory development to beyond: Greater consideration must be given to the vSTM task**

In an attempt to explore what neighbourhood density reveals about vSTM some important issues that have been highlighted in this thesis are that task variables such as the memory task and word pool size seemingly modulate the ND effect in ways not predicted by current models of vSTM. Additionally, it is possible that a vSTM task may not always be completed in the way intended by the experimenter but still yield results that seemingly progress (although may be misleading) theory development. These issues are of importance because, to understand vSTM, it seems necessary to question in far more depth what exactly is being tested by any particular task. For example, within the ND effect literature some researchers consider serial reconstruction to reduce the impact that any differences in spoken production times between to-be-remembered items could have upon the task (e.g., Jalbert et al., 2011 a, b). Others have considered it a pure test of order memory (e.g., Clarkson et al., 2017). However, besides from theoretical disagreements regarding serial reconstruction, it is also a task whereby participants could successfully complete it by choosing to adopt a first letter strategy. In that respect, depending upon the strategy the participant adopts, the task may not always be testing item or order memory. Instead it may be also be testing the impact that alternative strategies have upon the likelihood of effects appearing/attenuating.

A failure to consider that alternative strategies are available for task completion may lead to incorrect interpretations of the data. It is not just the chosen memory task that could possibly mislead researchers though but also task variables (e.g., presentation modality). For example, some researchers consider that auditorily presented material has direct access to the phonological store (e.g., Baddeley, 2000). However, it may instead be providing a situation whereby obligatory perceptual processes such as auditory object formation can impact the outcome of the task. As already discussed in Section 1.7.1 the lexicality effect is robust in visually presented serial recognition but disappears in auditorily presented serial recognition (Macken et al., 2014). An almost exclusive use of auditorily presented serial recognition had previously led researchers to assume that the lexicality effect attenuates in serial recognition (e.g., Gathercole et al., 2001) and those findings were used to make claims about the distinct but interactive nature of LTM and STM.

Another example of presentation modality being crucial to the presence or absence of effects is the survival of the phonological similarity effect under conditions of articulatory suppression when material is presented auditorily but not visually (e.g., see Section 1.2.1). This has been used as key evidence supporting the notion that auditory, but not visual material, benefits from obligatory access to the phonological store (e.g., Baddeley, 2000). However, closer analysis of serial position data reveals that the survival of phonological similarity effect when using auditory presentation is almost exclusively a consequence of the final auditory item having a recall advantage compared to the visually presented items (i.e., a strong recency effect for auditory but not visual items; Jones, Macken & Nicholls, 2004). Key here is that because auditory presentation affords auditory object formation (e.g., Bregman, 1990) information contained at the boundaries of an auditory object is more accessible than information contained within the object (e.g., Bregman & Rudnicky, 1975). As such, survival of the phonological similarity effect when using auditory presentation can be explained because of the final item advantage afforded to auditory but not visually presented sequences. To support this interpretation, the phonological similarity effect in the auditory modality can be eliminated under conditions of articulatory suppression by adding a redundant word (e.g., 'zero') at the end of a to-be-remembered sequence in the same voice and presentation rate as the to-be-remembered materials (Jones et al., 2004). In these instances, the auditory object is now extended and comprises the additional word. Because of this, the final to-be-remembered item no longer benefits from being at the boundary of the auditory object. In that respect, rather than providing alternative routes to the phonological store (e.g., Baddeley, 2000), the manipulation of modality tests the ways in which presented items are bound (or not) into auditory objects and, depending upon the demands of the memory task, how successful participants are at extracting information contained within an auditory object. The examples discussed so far in this section highlight how greater consideration and diversification of tasks and materials are essential in enriching our understanding of memory.

The issue of what exactly a vSTM task is testing does not just apply to theory development though. Perhaps even more importantly is that the theories developed from vSTM tasks are applied in more general settings. The working memory model (Baddeley & Hitch, 1974; Baddeley, 1986; Baddeley, 2000) is used extensively as a framework across multiple psychology disciplines (e.g., neuroscientific,

developmental) to understand a wide variety of phenomena. For example, Dupont-Boime and Thevenot (2018) found that children with larger working memory capacity (as assessed by the total number of digits that could be recited in reverse order) also tended to use finger counting strategies when engaging in arithmetic tasks. The authors concluded that the ability to use finger counting must therefore require high working memory capacity. This finding has implications in assessing children's developmental rates and should a child not be using finger counting, when their peers are, could raise concerns about the development of their working memory capacity. However, an alternative is to consider that children who spontaneously employ alternative strategies to a task (e.g., finger counting to complete a maths task) may also be able to employ such alternative strategies in a vSTM task. Rather than providing a measure of working memory capacity the digit span task used by Dupont-Boime and Thevenot (2018) may instead have been testing what skills their participants had that they could opportunistically use to help them later reproduce a sequence of digits (e.g., Macken et al., 2015; Jones & Macken, 2018). Finger counting is a particularly useful skill in this setting as, similar to using finger counting when completing maths questions, it provides an additional way to help remember numbers (e.g., Reisberg, Rappaport, & O'Shaughnessy, 1984). If finger counting was used to assist with the digit-span task then the task ceases being a test of vSTM capacity but a test of how well the participants can employ previously learnt skills to help optimise their ability to re-produce a sequence of digits.

Anderson, Wagovich & Brown (2019) found that children who stutter have lower memory spans (again measured by a digit span task) than children who do not. The authors concluded that children who stutter therefore show weaknesses in vSTM. However, the results were obtained by using a span task requiring spoken output. As such, the successful completion of the task could be considered to require both motor planning and successful production of the motor plan in the form of a spoken response. Of relevance here is that individuals who stutter report more errors in inner speech as well as producing more overt speech errors than participants who do not stutter (Brocklehurst & Corley, 2011). If inner speech is considered analogous to motor planning (e.g., Oppenheim & Dell, 2010) then any task that requires the preparation and output of a motor plan is likely to be performed worse by someone who stutters, not because of a deficit in vSTM, but because the task does not readily map onto the skills possessed by the participants (i.e., the ability to plan and then

produce fluent speech). It may be the case that vSTM tasks such as auditorily presented serial recognition would not reveal 'deficits' in vSTM relative to non-stutterers because the task no longer requires assembly and output of a speech motor plan. The task can be completed via global pattern matching (e.g., Macken et al., 2014).

While the work by Dupont-Boime and Thevenot (2018) and Anderson et al. (2019) are just a couple of examples they highlight that, without far more consideration of what exactly is being tested by a particular task, the interpretations of the data may be incorrect and possibly unhelpful in improving the future outcomes for those being tested. For example, the results of Anderson et al. could be interpreted as there being a benefit of helping to improve the vSTM capacity of those who stutter (perhaps by having them complete a variety of vSTM tasks or complete working memory training; e.g., Schwarb, Nail & Schumacher, 2016). However, if reduced vSTM span (as measured by relatively poor performance in vSTM tasks) is the outcome, rather than the cause of stuttering, then attempts to improve vSTM capacity are likely to have no measurable benefits upon stuttering (aside from providing some additional practice in preparing, and then producing, strings of speech comprised of random and unrelated items).

Returning to the results of the current thesis, Chapter 4 highlighted how the manipulation of neighbourhood density may also be a manipulation of articulatory difficulty. In that sense it raises questions over what exactly a vSTM task manipulating neighbourhood density is testing. Is the manipulation a test of the relative level of support each item in STM receives from LTM, or is it a test of the relative ease in which a sequence of items can be assembled into a speech motor plan and then produced? In Chapter 2, the word pool size was reduced in Experiments 3 and 4. This is a manipulation suggested to reduce the burden upon item memory (e.g., Goh & Pisoni, 2003) whereby any effects that impact item rather than order memory attenuate. However, there was a reversal of the ND effect which raises the question of whether the reduction in word pool size really reduced the burden upon item memory. Instead, it may have been testing whether some strategy was in fact more useful in remembering the sparse neighbourhood items under the particular combination of word pool size, materials and the memory task. To understand the ND effect and what, if anything, it really reveals about memory it is essential that, moving forwards, even greater consideration is given to all aspects of

the vSTM tasks used in the research. A failure to do so may only serve to obscure, rather than help, our understanding of memory.

### **6.7 Summary and Conclusions**

In conclusion, the present thesis has demonstrated that, given the current evidence available, interpreting the ND effect as being some consequence of the activation elicited within linguistic networks in LTM is problematic and possibly misleading. The vSTM advantage for sparse neighbourhood words elicited in Experiments 3 and 4 cannot easily be accommodated by vSTM accounts that incorporate redintegration, or explanations based upon order encoding, that are dependent upon the level of activation elicited in LTM. Additionally, the finding that sparse neighbourhood words not only tend to consist of more difficult articulations (Study 1) but also take longer to vocalise (Experiment 5) means that generally the ND effect could be re-interpreted as some consequence of segmental recoding being more fluent for dense rather than sparse neighbourhood items (e.g., Jones & Macken, 2018; Macken et al., 2015). However, the thesis failed to demonstrate an equivalent effect when just articulatory difficulty was manipulated. Despite this, there is an increasing line of evidence (e.g., Macken et al., 2014; Murray & Jones, 2002; St. John, 2015; Woodward et al., 2008) that articulatory difficulty at both the item and sequence-level can impact the fluency in which a to-be-remembered sequence can be segmentally recoded and output. Until the neighbourhood density literature addresses the number of concerns highlighted in the present thesis then it would be best to remain cautious before drawing any firm conclusions regarding the ND effect and what exactly it reveals about vSTM.



## References

- Allen, R., & Hulme, C. (2006). Speech and language processing mechanisms in verbal serial recall. *Journal of Memory and Language, 55*, 64-88.
- Anderson, J. D., Wagovich, S. A., & Brown, B. T. (2019). Phonological and semantic contributions to verbal short-term memory in young children with developmental stuttering. *Journal of Speech, Language, and Hearing Research, 62*, 644-667.
- Audacity Team (2012). Audacity (Version 2.0.1-beta) [Computer software]. Retrieved from <http://audacityteam.org/>.
- Baayen, R. H., Popenbrock, R., & van Rijn, H. (1993). The CELEX lexical database [Computer software]. *Philadelphia, PA: University of Pennsylvania, Linguistic Data Consortium.*
- Baddeley, A. D. (1986). *Working memory*. Oxford, England: Oxford University Press.
- Baddeley, A. D. (2000). Short-term and working memory. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 77– 92). Oxford, England: Oxford University Press.
- Baddeley, A. D. (2002). Is working memory still working? *European Psychologist, 7*, 85-97.
- Baddeley, A. D., Chincotta, D., Stafford, L., & Turk, D. (2002). Is the word length effect in STM entirely attributable to output delay? Evidence from serial recognition. *Quarterly Journal of Experimental Psychology, 55*, 353-369.
- Baddeley, A. D., and Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation*, (Vol. 8. Pp. 47-90). New York: Academic Press.

- Baddeley, A. D., Lewis, V. J., & Vallar, G. (1984). Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology*, *36*, 233-252.
- Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*, 575-589.
- Baddeley, A. D., & Wilson, B. A. (1985). Phonological coding and short-term memory in patients without speech. *Journal of Memory and Language*, *24*, 490-502.
- Bailey, H., Dunlosky, J., & Hertzog, C. (2009). Does differential strategy use account for age-related deficits in working-memory performance? *Psychology and Aging*, *24*, 82-92.
- Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments, & Computers*, *33*, 73-79.
- Bireta, T. J., Neath, I., & Surprenant, A. M. (2006). The syllable-based word length effect and stimulus set specificity. *Psychonomic Bulletin & Review*, *13*, 434-438.
- Bregman, A. S. (1990). *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*, 263-267.
- Brocklehurst, P. H., & Corley, M. (2011). Investigating the inner speech of people who stutter: Evidence for (and against) the Covert Repair Hypothesis. *Journal of Communication Disorders*, *44*, 246-260.

- Burgess, N., and Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, 106, 551–581.
- Burgess, N., and Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language*, 55, 627–652.
- Caplan, D., Rochon, E., & Waters, G. S. (1992). Articulatory and phonological determinants of word length effects in span tasks. *Quarterly Journal of Experimental Psychology*, 45, 177-192.
- Chan, K. Y., and Vitevitch, M. S. (2009). The influence of the phonological neighborhood clustering coefficient on spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 1934–1949.
- Chitoran, I., Goldstein, L., & Byrd, D. (2002). Gestural overlap and recoverability: Articulatory evidence from Georgian. In C. Gussenhoven, & N. Warner (Eds.), *Laboratory phonology VII* (pp. 419–448). Berlin: Walter de Gruyter.
- Cho, T. & Keating, P. A. (2001). Articulatory and acoustic studies of domain-initial strengthening in Korean. *Journal of Phonetics*, 29, 155–190.
- Clarkson, L., Roodenrys, S., Miller, L. M., & Hulme, C. (2017). The phonological neighbourhood effect on short-term memory for order. *Memory*, 25, 391-402.
- Coe R. (2002). *It's the effect size, stupid: What effect size is and why it is important*. Paper presented at the Annual Conference of the British Educational Research Association, University of Exeter, Exeter, England.
- Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology*, 33, 497-505.

- Coltheart, M., Davelaar, E., Jonasson, J. F., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535-555). Hillsdale, NJ: Erlbaum.
- Coltheart, V., & Langdon, R. (1998). Recall of short word lists presented visually at fast rates: Effects of phonological similarity and word length. *Memory & Cognition, 26*, 330-342.
- Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology, 55*, 429-432.
- Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology, 1*, 42-45.
- Crannell, C. W., & Parrish, J. M. (1957). A comparison of immediate memory span for digits, letters, and words. *Journal of Psychology, 44*, 319-327.
- Davis, C. J. (2005). N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behaviour Research Methods, 37*, 65-70.
- Dell, G. S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review, 93*, 283-321.
- Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review, 104*, 123-147.
- DeLosh, E. L., & McDaniel, M. A. (1996). The role of order information in free recall: Application to the word-frequency effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1136-1146.

- Derrough, L. S., Neath, I., Surprenant, A. M., Beaudry, O., & Saint-Aubin, J. (2017). The effect of lexical factors on recall from working memory: Generalizing the neighborhood size effect. *Canadian Journal of Experimental Psychology, 71*, 23-31.
- Deterding, D. H. (1997). The formants of monophthong vowels in standard southern British English pronunciation. *Journal of the International Phonetic Association, 27*, 47-55.
- Dupont-Boime, J., & Thevenot, C. (2018). High working memory capacity favours the use of finger counting in six-year-old children. *Journal of Cognitive Psychology, 30*, 35-42.
- Durand, J. (2005). Tense/lax, the vowel system of English and phonological theory. In Philip Carr, Jacques Durand & Colin J. Ewen (Eds.), *Headhood, elements, specification and contrastivity*, 77–98. Amsterdam: John Benjamins.
- Edwards, M. L., & Shriberg, L. D. (1983). *Phonology: Applications in communicative disorders*. San Diego, California: College Hill Press.
- Ellis, A. W. (1980). Errors in speech and short-term memory: The effects of phonological similarity and syllable position. *Journal of Verbal Learning and Verbal Behavior, 19*, 624-634.
- Epstein, M., Hacopian, N., Ladefoged, P. (Eds.). (2002). Dissection of the Speech Production Mechanism. *UCLA Working Papers in Phonetics, 102*.
- Fant, G. (1973). *Speech sounds and features*. Cambridge, MA: MIT Press.
- Fossett, T. R. D., McNeil, M. R., Pratt, S. R., Tompkins, C. A., & Shuster, L. I. (2016). The effect of speaking rate on serial-order sound-level errors in normal healthy controls and persons with aphasia. *Aphasiology, 30*, 74-95.

- Fowler, C. A., Treiman, R., & Gross, J. (1993). The structure of English syllables and polysyllables. *Journal of Memory and Language*, *32*, 115-140.
- Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, *66*, 789-806.
- Gathercole, S. E., Pickering, S. J., Hall, M., & Peaker, S. M. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *Quarterly Journal of Experimental Psychology*, *54*, 1-30.
- Goh, W. D., & Pisoni, D. B. (2003). Effects of lexical competition on immediate memory span for spoken words. *Quarterly Journal of Experimental Psychology*, *56*, 929-954.
- Grainger, J., Muneaux, M., Farioli, F., & Ziegler, J. C. (2005). Effects of phonological and orthographic neighbourhood density interact in visual word recognition. *Quarterly Journal of Experimental Psychology*, *58*, 981-998.
- Grimm, P. (2010). Social desirability bias. In J. Sheth & N. Malhotra (Eds.), *Wiley International Encyclopedia of Marketing*. Chichester, UK: John Wiley & Sons.
- Guitard, D., Gabel, A. J., Saint-Aubin, J., Surprenant, A. M., & Neath, I. (2018). Word length, set size, and lexical factors: Re-examining what causes the word length effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 1824-1844.
- Guitard, D., Saint-Aubin, J., Tehan, G., & Tolan, A. (2018). Does neighborhood size really cause the word length effect? *Memory & Cognition*, *46*, 244-260.
- Honeybone, P. (2008). Lenition, weakening and consonantal strength: tracing concepts through the history of phonology. In J. de Carvalho, T. Scheer, & P. Ségéral (Eds.), *Lenition and Fortition*. Berlin: Mouton de Gruyter. 9-93.

- Hulme, C., Roodenrys, S., Brown, G., & Mercer, R. (1995). The role of long-term memory mechanisms in memory span. *British Journal of Psychology*, *86*, 527-536.
- Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D., Martin, S., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a reintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 1217-1232.
- Hulme, C., Thomson, N., Muir, C., & Lawrence, A. (1984). Speech rate and the development of short-term memory span. *Journal of Experimental Child Psychology*, *38*, 241-253.
- Jalbert, A., Neath, I., Bireta, T. J., & Surprenant, A. M. (2011a). When does length cause the word length effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 338-353.
- Jalbert, A., Neath, I., & Surprenant, A. M. (2011b). Does length or neighborhood size cause the word length effect? *Memory & Cognition*, *39*, 1198-1210.
- Jefferies, E., Frankish, C. R., & Lambon-Ralph, M. A. (2006). Lexical and semantic influences of item and order memory in immediate serial recognition: Evidence from a novel task. *Quarterly Journal of Experimental Psychology*, *59*, 949-964.
- Jefferies, E., Frankish, C. R., & Noble, K. (2011). Strong and long: Effects of word length on phonological binding in verbal short-term memory. *Quarterly Journal of Experimental Psychology*, *64*, 241-260.
- Jones, D. M., & Macken, B. (2018). In the beginning was the deed: Verbal short-term memory as object-oriented action. *Current Directions in Psychological Science*, *27*, 351-356.

- 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, 656-674.
- Jones, G., & Macken, B. (2015). Questioning short-term memory and its measurement: Why digit span measures long-term associative learning. *Cognition*, 144, 1-13.
- Jones, G., & Macken, B. (2018). Long-term associative learning predicts verbal short-term memory performance. *Memory & Cognition*, 46, 216-229.
- Kern, S., Davis, B., MacNeilage, P., Koçbas, D., Kuntay, A., & Zink, I. (2014). Crosslinguistic similarities and differences in babbling: phylogenetic implications. In J.M. Hombert (Ed.), *Towards the origins of language and languages*.
- Kirchner, R. (1998). *An effort-based approach to consonant lenition* (Doctoral Dissertation). Los Angeles: University of California.
- Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford, England: Blackwell.
- Lewandowsky, S., & Farrell, S. (2000). A reintegration account of the effects of speech rate, lexicality, and word frequency in immediate serial recall. *Psychological Research*, 63, 163-173.
- Lewandowsky, S., & Oberauer, K. (2015). Rehearsal in serial recall: An unworkable solution to the nonexistent problem of decay. *Psychological Review*, 122, 674-699.
- Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 403-439). Dordrecht: Kluwer.



- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*, 1-36.
- Lucidi, A., Langerock, N., Hoareau, V., Lemaire, B., Camos, V., & Barrouillet, P. (2016). Working memory still needs verbal rehearsal. *Memory & Cognition, 44*, 197-206.
- Macken, B., Taylor, J. C., & Jones, D. M. (2014). Language and short-term memory: The role of perceptual-motor affordance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 1257-1270.
- Macken, B., Taylor, J. C., & Jones, D. M. (2015). Limitless capacity: a dynamic object-oriented approach to short-term memory. *Frontiers in Psychology, 6*, 293.
- Macken, B., Taylor, J. C., Kozlov, M. D., Hughes, R. W., & Jones, D. M. (2016). Memory as embodiment: The case of modality and serial short-term memory. *Cognition, 155*, 113-124.
- Martin, N., & Gupta, P. (2004). Exploring the relationship between word processing and verbal short-term memory: Evidence from associations and dissociations. *Cognitive Neuropsychology, 21*, 213-228.
- Martin, N., & Saffran, E. M. (1997). Language and auditory-verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology, 24*, 641-682.
- McLeod, S., & Crowe, K. (2018). Children's consonant acquisition in 27 languages: A cross-linguistic review. *American Journal of Speech-Language Pathology, 27*, 1546-1571.
- Miller, L. M. (2010). *The word frequency effect in short-term serial recall* (Doctoral Dissertation). School of Psychology-Faculty of Health & Behavioural Sciences, University of Wollongong.

- Monnier, C., & Syssau, A. (2008). Semantic contribution to verbal short-term memory: Are pleasant words easier to remember than neutral words in serial recall and serial recognition? *Memory & Cognition*, *36*, 35-42.
- Mooshammer, C., Hoole, P. & Geumann, A. (2007). Jaw and order. *Language and Speech* *50*, 145-176.
- Morrison, A. B., Rosenbaum, G. M., Fair, D., & Chein, J. M. (2016). Variation in strategy use across measures of verbal working memory. *Memory & Cognition*, *44*, 922-936.
- Mulatti, C., Besner, D., & Job, R. (2003). *Reading aloud: Evidence for a phonologically based neighborhood effect*. Paper presented at the 44th Annual Meeting of the Psychonomic Society, Vancouver, BC, Canada.
- Murray, A., & Jones, D. M. (2002). Articulatory complexity at item boundaries in serial recall: The case of Welsh and English digit span. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 594-598.
- Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*, 251–269.
- Naveh-Benjamin, M., & Ayres, T. J. (1986). Digit span, reading rate, and linguistic relativity. *Quarterly Journal of Experimental Psychology*, *38*, 739-751.
- Neale, K., & Tehan, G. (2007). Age and redintegration in immediate memory and their relationship to task difficulty. *Memory & Cognition*, *35*, 1940-1953.
- Neath, I., & Nairne, J. S. (1995). Word-length effects in immediate memory: Overwriting trace decay theory. *Psychonomic Bulletin & Review*, *2*, 429-441.
- Neath, I., Bireta, T. J., & Surprenant, A. M. (2003). The time-based word length effect and stimulus set specificity. *Psychonomic Bulletin & Review*, *10*, 430-434.

- Ohala, J. J. (1990). The phonetics and phonology of aspects of assimilation. In J. Kingston, & M. E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech* (pp. 258-275). Cambridge: Cambridge University Press.
- Oller, D. K., Wieman, L. A., Doyle, W. J., & Ross, C. (1976). Infant babbling and speech. *Journal of Child Language*, 3, 1-11.
- Oppenheim, G. M., & Dell, G. S. (2010). Motor movement matters: The flexible abstractness of inner speech. *Memory & Cognition*, 38, 1147-1160.
- Page, M. P. A., and Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, 105, 761–781.
- Peirce, J. W. (2007). PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods*, 162, 8-13.
- Port, R. (2007). How are words stored in memory? Beyond phones and phonemes. *New Ideas in Psychology*, 25, 143-170.
- Reisberg, D., Rappaport, I., & O'Shaughnessy, M. (1984). Limits of working memory: The digit-digit span. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 203-221.
- Roodenrys, S., Hulme, C., Lethbridge, A., Hinton, M., & Nimmo, L. M. (2002). Word-frequency and phonological-neighborhood effects on verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 1019-1034.
- Sabsevitz, D. S., Medler, D. A., Seidenberg, M., & Binder, J. R. (2005). Modulation of the semantic system by word imageability. *NeuroImage*, 27, 188-200.

- Saint-Aubin, J., & Poirier, M. (1999). The influence of long-term memory factors on immediate serial recall: An item and order analysis. *International Journal of Psychology, 34*, 347-352.
- Sandoval, S., Berisha, V., Utianski, R. L., Liss, J. M., & Spanias, A. (2013). Automatic assessment of vowel space area. *Journal of the Acoustical Society of America, 134*, 477-483.
- Schwarb, H., Nail, J., & Schumacher, E. H. (2016). Working memory training improves visual short-term memory capacity. *Psychological Research, 80*, 128-148.
- Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition, 21*, 168-175.
- Schweickert, R., & Boruff, B. (1986). Short-term memory capacity: Magic number or magic spell? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 419-425.
- Schweickert, R., Guentert, L., & Hersberger, L. (1990). Phonological similarity, pronunciation rate, and memory span. *Psychological Science, 1*, 74-77.
- Service, E. (1998). The effect of word length on immediate serial recall depends on phonological complexity, not articulatory duration. *Quarterly Journal of Experimental Psychology, 51*, 283-304.
- Sevold, C. A., & Dell, G. S. (1994). The sequential cuing effect in speech production. *Cognition, 58*, 91-127.
- Shattuck-Hufnagel, S. (1992). The role of word structure in segmental serial ordering. *Cognition, 42*, 213-259.
- Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences, 12*, 182-186.

- Simpson, A. P. (2001). Does articulatory reduction miss more patterns than it accounts for? *Journal of the International Phonetic Association*, 31, 29-39.
- Souza, A. S., & Oberauer, K. (2018). Does articulatory rehearsal help immediate serial recall? *Cognitive Psychology*, 107, 1-21.
- St. John, A. (2015). *Remembering with your tongue: Articulatory embodiment in memory and speech* (Doctoral Dissertation). School of Psychology, Cardiff University.
- Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. E. Stelmach (Ed.), *Information processing in motor control and learning* (pp. 118-152). New York: Academic Press.
- Sternberg, S., Wright, C. E., Knoll, R. L., & Monsell, S. (1980). Motor programs in rapid speech: Additional evidence. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 507-534). Hillsdale, N.J: Erlbaum.
- Stevens, K. N. (1971). Airflow and turbulence noise for fricative and stop consonants: Static considerations. *Journal of the Acoustical Society of America*, 50, 1180-1192.
- Storkel, H. L. (2004). Methods for minimizing the confounding effects of word length in the analysis of phonotactic probability and neighborhood density. *Journal of Speech, Language, & Hearing Research*, 47, 1454-1468.
- Stuart, G., & Hulme, C. (2000). The effects of word co-occurrence on short-term memory: Associative links in long-term memory affect short-term memory performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 796-802.
- Tan, L., & Ward, G. (2008). Rehearsal in immediate serial recall. *Psychonomic Bulletin & Review*, 15, 535-542.

- Thorn, A. S., & Gathercole, S. E. (2001). Language differences in verbal short-term memory do not exclusively originate in the process of subvocal rehearsal. *Psychonomic Bulletin & Review*, 8, 357-364.
- Treiman, R., & Danis, C. (1988). Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 145-152.
- Treiman, R., Fowler, C. A., Gross, J., Berch, D., & Weatherston, S. (1995). Syllable structure or word structure? Evidence for onset and rime units with disyllabic and trisyllabic stimuli. *Journal of Memory and Language*, 34, 132-155.
- Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 735-747.
- Vitevitch, M. S., Chan, K. Y., & Roodenrys, S. (2012). Complex network structure influences processing in long-term and short-term memory. *Journal of Memory and Language*, 67, 30-44.
- Vitevitch, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and Language*, 68, 306-311.
- Ward, G., Tan, L., & Grenfell-Essam, R. (2010). Examining the relationship between free recall and immediate serial recall: The effects of list length and output order. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 1207-1241.
- Warren, R. M., Obusek, C. J., Farmer, R. M., & Warren, R. P. (1969). Auditory sequence: Confusion of patterns other than speech or music. *Science*, 164, 586-587.

- Warren, R.M. (1974a). Auditory temporal discrimination by trained listeners. *Cognitive Psychology*, 6, 237-256.
- Warren, R.M. (1974b). Auditory pattern discrimination by untrained listeners. *Perception & Psychophysics*, 15, 495-500.
- Wilshire, C. E. (1999). The “tongue twister” paradigm as a technique for studying phonological encoding. *Language and Speech*, 42, 57-82.
- Wilson, M. (1988). MRC Psycholinguistic Database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, and Computers*, 20, 6-10.
- Woodward, A. J., Macken, W. J., & Jones, D. M. (2008). Linguistic familiarity in short-term memory: A role for (co-) articulatory fluency? *Journal of Memory and Language*, 58, 48-65.
- Wray, A. (2015). Why are we so sure we know what a word is? In J. R. Taylor (Ed.), *Oxford Handbook of the Word* (pp. 726-750). Oxford: Oxford University Press.
- Yates, M. (2013). How the clustering of phonological neighbors affects visual word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39, 1649-1656.
- Yates, M., Locker, L., & Simpson, G. B. (2004). The influence of phonological neighborhood on visual word perception. *Psychonomic Bulletin & Review*, 11, 452-457.
- Yavas, M. S. (2006). *Applied English phonology*. Malden, MA: Blackwell Publishers.
- Zemlin, W. R. (1998). *Speech and hearing science, Anatomy and physiology* (4<sup>th</sup> ed.). Boston: Allyn and Bacon.

Zygis, M., Fuchs, S., & Koenig, L. (2012). Phonetic explanations for the infrequency of voiced sibilant affricates across languages. *Laboratory Phonology*, 3, 299–336.



## Appendix A

### Word pools used in Experiments 1, 2 and 5

Dense Neighbourhood Words		Sparse Neighbourhood Words	
bark	node	badge	morgue
bead	nut	beige	moth
bin	peach	carve	noose
boot	pearl	chase	notch
buzz	pod	chef	nudge
cone	poise	chime	peg
cop	poke	churn	pierce
cork	pun	couch	rib
cull	rake	dab	shove
duck	rim	dodge	sieve
fade	ripe	fang	soothe
fizz	robe	fog	thatch
ham	sane	forge	thief
haze	sap	geese	thud
hike	shack	germ	torch
hood	siege	gig	turf
kite	sock	gown	verse
lag	tan	gush	vogue
lard	tart	herb	void
lice	thorn	ledge	web
lick	tile	loaf	wharf
maim	vine	lurch	yarn
meek	weep	merge	zip
mole	weird	mesh	zoom

	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Phonological neighbourhood Size	22.81	4.97	9.73	2.78
Orthographic neighbourhood Size	10.9	5.77	3.75	3.53
Word frequency	5.08	3.35	4.25	3.16
Imageability	509.7	98.42	511.1	88.72

## Appendix B

### Word pools used in Experiments 3 and 4

Dense Neighbourhood Words		Sparse Neighbourhood Words	
bark	pearl	badge	peg
boot	pod	beige	pierce
cone	tan	chase	thatch
cop	tart	chef	thief
cork	thorn	chime	torch
cull	tile	couch	turf
<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Phonological neighbourhood size			
24.5	3.78	8.75	1.82
Orthographic neighbourhood size			
12.42	6.39	3	3.59
Word frequency			
5.6	3.02	5.65	3.97
Imageability			
537.59	85.34	532.2	58.17
Percentage Correct in Experiment 1			
60.69	16.16	53.1	13.55

A possible confound when selecting a subset of words based upon the percentage correct achieved in Experiment 1 is that because the selected words appeared in different serial positions then any differences in the descriptive statistics may have resulted from which serial position the selected words most often appeared in. For example, if one word pool had more words that predominantly appeared in earlier serial positions then this pool would likely have achieved a higher percentage correct in Experiment 1 simply because of primacy effects. To check that this was not the case with the selected materials the average percentage correct for each serial position in Experiment 1 was calculated (collapsed across modality and neighbourhood density). Counts were then taken for the number of times that each word in the 12-item dense and sparse word pools appeared in each serial position and an estimated percentage correct, based upon serial position appearance, assigned to each word. For a hypothetical example, serial position 1 might have had

an overall average percentage correct of 100% and serial position 4 an overall average percentage correct of 50%. If a word was presented a total of 200 times over the course of all participants and it appeared in each of those serial positions 100 times then, out of those 200 instances, 150 instances would be correct (100 at serial position 1 and 50 at serial position 4). An average for each word was then calculated (e.g., in the earlier example the word would be assigned a percentage correct of 75%) and finally an overall average percentage correct for all the words in each word pool was calculated. The subset of dense neighbourhood words had an overall average of 59.66% and the subset of sparse neighbourhood words had an overall average of 58.06%. This means that, at most, the serial position that words appeared in could account for 1.6% of the 7.59% difference between the dense and sparse neighbourhood word pools. 5.99% must therefore have been due to some other factor i.e., the ND effect.

## Appendix C

### Vowel Difficulty Scoring

In addition to the consonant scoring system outlined in Chapter 5 an attempt was also made to model articulatory difficulty by assigning difficulty scores to vowels. Along with a consonant score this would allow entire CVC words to be scored. Quantifying difficulty for vowels was more ambiguous than it was for consonants though because they are more difficult to describe phonetically. This is because vowel sounds are essentially a continuum whereby different sounds fall in areas within a continuous space in the mouth. However, this space has been modelled as the vowel space area (e.g., Durand, 2005; Fant, 1973) which places vowels on a two-dimensional chart plotting the first (F1) and second (F2) formant frequency of the vowels. The frequencies are related to the size and the shape of the space that is created in the mouth as a function of the jaw opening (F1) and tongue position (F2) (e.g., Sandoval, Berisha, Utianski, Liss, & Spanias, 2013). In the very centre of the vowel space area is the schwa vowel sound (e.g., '/ə/' in 'əbaut' – 'about') and all other vowel sounds are placed at varying distances away from the schwa dependent upon the degree of movement required by the tongue and jaw. The logic of the current vowel difficulty scoring system is based upon findings that when vowel sounds undergo lenition (thereby reducing articulatory effort; e.g., Kirchner, 1998) vowel sounds tend to become more centralised (e.g., Deterding, 1997). As such, any vowel sounds moving away from the central positions could be considered more effortful and therefore more difficult to articulate. A similar rationale to the consonant scoring system was then used whereby the more muscular effort that is required to move away from the central vowel position then the more difficult the required articulation for that vowel sound is considered to be. Vowel sounds are also a consequence of lip position (e.g., Epstein, Hacopian & Ladefoged, 2002) and tenseness (e.g., Yavas, 2006). To ensure a comprehensive measure of vowel difficulty these were also included as dimensions. This meant that vowels were scored on features that fell under four distinct dimensions which will be described in detail in the upcoming section; horizontal tongue position, vertical tongue position, lip position and tenseness. Each score obtained within the three dimensions was totalled to provide a single overall difficulty score for vowels (see Table 6).

### ***Dimension 1 - Tongue position (Vertical and Horizontal)***

In order to move away from centralised vowel sounds the extrinsic muscles of the tongue must be utilised. The styloglossus moves the tongue back into the mouth and upwards, the hyoglossus moves the tongue back and down and the genioglossus pulls the tongue forwards (Epstein et al., 2002). The scoring system considered each single movement (up or down, or forwards or back) to be more effortful than a centralised tongue position and therefore contributed 1 to a vowel's overall difficulty score. If a vowel requires a combination of both movements (e.g., an upwards and forwards movement) then it would score 2 (1 for the upward movement and 1 for the forward movement). Centralised vowels contributed 0 because they require no tongue movement.

### ***Dimension 2 - Lip roundedness***

In order to move away from centralised vowel sounds the lips must move from a neutral position and be placed in either a rounded or spread position. This requires utilisation of the orbicularis oris muscles (rounded) or the risorius muscles (spread) (Epstein et al., 2002). The scoring system considered movement of these muscles to be more effortful than keeping the lips in a neutral position. As such, vowels requiring either the rounding or the spreading of the lips contributed 1 to a vowel's overall difficulty score. Vowels requiring a neutral lip position contributed 0.

### ***Dimension 3 - Tenseness***

Vowels can also either be tense or lax. Tense vowels have longer durations and require a greater degree of muscular tension than lax vowels (e.g., Yavas, 2006). Because tense vowels require more muscular tension the scoring system considered tense vowels to be more effortful than lax vowels with tense vowels contributing 1 to a vowel's overall difficulty score. Lax vowels contributed 0.

Table 6

Articulatory difficulty scores assigned to each vowel (V) in each dimension and the overall difficulty score (ODS)

V	Vertical Tongue Position			Horizontal Tongue Position			Lip Position			Tenseness		ODS
	High	Mid	Low	Front	Central	Back	Spread	Neutral	Rounded	Tense	Lax	
i:	1			1			1			1		4
i	1			1			1				0	3
ɪ	1			1			1				0	3
ɛ		0		1				0			0	1
æ			1	1				0			0	2
ʌ			1		0			0			0	1
ɜ:		0			0			0		1		1
ə		0			0			0			0	0
ɑ:			1			1		0		1		3
ɒ			1			1			1		0	3
ɔ:		0				1			1	1		3
ɔ			1			1			1		0	3
u:			1			1			1	1		4

Note. '0' indicates a least difficult feature and '1' indicates a difficult feature.

From the pool of 1218 CVC words used in Study 1a a vowel difficulty score was assigned to all words containing a monophthong. This yielded a final pool of 930 words. Results of a simple linear regression suggest that the vowel difficulty scores failed to predict a significant amount of variance in the phonological neighbourhood density distributions,  $F(1, 929) = 0.55$ ,  $p = .46$ ,  $R^2 < .01$ .

## Appendix D

### Word pools used in Experiments 6 and 7a

Posterior Onset Pool (Disfluent)		Anterior Onset Pool (Fluent)		
beak	mice	cam	nap	
bid	mid	cape	nip	
bike	moan	cob	ram	
bite	moat	cop	reap	
boot	mud	cub	rib	
bout	muse	dame	rim	
budge	pan	dim	roam	
fad	peas	dime	rob	
fake	peat	dip	robe	
fan	pen	dome	rub	
fen	pike	doom	rum	
fin	pin	kip	sap	
foal	pine	leap	seep	
foul	poke	lime	sip	
fuse	pout	lip	soap	
mace	pun	loam	sob	
maul	pus	lob	tame	
maze	veer	loop	tap	
mead	vet	lop	tip	
meek	vine	lope	tome	

	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Phonological neighbourhood Size	21.73	3.84	21.7	3.27
Orthographic neighbourhood Size	11.38	4.8	12.8	4.44
Word frequency	7.14	6.99	7.06	6.18

## Appendix E

### Word pools used in Experiment 7b

Posterior Onset Pool (Disfluent)		Anterior Onset Pool (Fluent)		
beak	mice	cam		nap
bid	mid	cape		nip
bike	moan	cob		ram
bite	moat	cop		reap
boot	mud	cub		rib
bout	muse	dame		rim
budge	pan	dim		roam
fad	peas	dime		rob
fade*	peat	dip		robe
fake	pen	dome		rub
fan	pike	doom		rum
fen	pin	kip		sap
fin	pine	lap		seep
foal	poke	leap		sip
foul	pout	lime		soap
fuse	pun	lip		sob
mace	pus	loam		tame
maul	veer	lob		tap
maze	vet	loop		tip
mead	vice*	lop		tome
meek	vine	lope		tub

(\* indicates additional words to those used in Experiment 6 and 8a)

	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Phonological neighbourhood Size				
	21.71	4.29	21.74	3.5
Orthographic neighbourhood Size				
	11.26	4.72	12.88	4.51
Word frequency				
	7.31	6.91	7.35	6.29