Remembering With Your Tongue - Articulatory Embodiment in Memory and Speech

This thesis is presented in candidature for the degree of Doctor of Philosophy

2015

Alexander St. John

Declaration/Statements Page

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 fulfillment of the requirements for the degree of …………………………(insert MCh, MD, MPhil, PhD etc, as appropriate)

Signed ………………………………………… (candidate)

Date …………………………

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated.

Other sources are acknowledged by explicit references. The views expressed are my own.

Signed ………………………………………… (candidate)

 Date …………………………

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available 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 …………………………

Acknowledgments

I would like to thank my supervisor Bill Macken for his guidance, encouragement and understanding over the years. I am also grateful to Dominic Dwyer, Dylan Jones, Jacky Boivin, John Culling, Lewis Bott, Petroc Sumner and Tom Freeman for the mentoring and inspiration they have provided throughout my time in the department. Finally, I would like to acknowledge the patience and unwavering personal support of my partner, Nicola Horner.

Thesis Summary

Articulatory factors are typically relegated to a peripheral role in theoretical accounts of cognitive function. For example, verbal short-term memory functions are thought to be serviced by dedicated mechanisms that operate on abstract phonological (i.e., non-articulatory) items. An alternative tested here is that memory functions are supported by motor control processes that embody articulatory detail. To provide evidence for this viewpoint, this thesis focuses on the influence of articulatory effort-minimisation processes on memory and speech.

Chapter 1 demonstrated that verbal sequences involving fluent inter-item coarticulations are better remembered than disfluent counterparts. Because coarticulatory fluency was manipulated by reordering a single set of items, this effect cannot be explained by item-oriented mechanisms. Neither is it a consequence of misarticulation at output, because it persists in an order reconstruction task where participants are not required to articulate responses. This fluency effect also extends beyond memory contexts to constrain reading times in inner speech.

Chapter 2 investigated whether effort-minimisation processes can explain superior memory for words from dense phonological neighbourhoods. Analysis indicates these words tend to involve simple articulatory features – a pattern that may reflect a shaping

influence of lenition on the phonological distributions that underlie neighbourhood density effects.

Chapter 3 investigated whether superior memory for frequent words can be explained by their susceptibility to lenition. Because lenition and redintegration alike are influenced by frequency, a phonetic manipulation was devised to induce lenition in nonwords experimentally whilst controlling for frequency. However, this experimentally-induced lenition did not translate into memory improvements.

The findings indicate a central role for articulatory factors in memory and speech function, consistent with the view that verbal short-term memory function is supported by speech motor control processes. This resonates with embodied approaches to explaining cognition, whereby distributed action and perception-oriented processes are deployed to provide task-specific solutions to cognitive problems.

Contents

General Introduction and Thesis Overview

Articulatory factors relate to the production of speech gestures by the speech apparatus (e.g., the tongue, lips etc.). Conventionally, the role of articulatory factors and processes in cognitive function has been viewed as a peripheral one. That is, although articulatory factors may bear on the output of particular cognitive functions, they are not integral to these functions. By analogy, a computer monitor does not constrain computing processes but acts as a device to visually output the products of these processes. In this way, articulatory factors can sometimes lead to performance effects in memory and speech – for example, a complex verbal sequence may be misarticulated in a serial recall task. However, these performance effects are purely a matter of output, analogous to faults in a computer monitor rather than the computer.

Consistent with this view, cognitive functions have traditionally been thought to depend on mechanisms that operate on centralised, item-level phonological representations. This assumption is reflected in influential models of verbal short-term memory (vSTM) such as the standard model, where vSTM functions are serviced by a phonological loop system (e.g., Baddeley, 2012). This system comprises a phonological store, which passively stores phonological items, and an active articulatory rehearsal process, which revivifies items in the phonological store to offset the effects of trace decay. Information from long-term memory can also contribute to vSTM performance via a phonological redintegration process, which reconstitutes decayed memory traces at output by matching them to intact corresponding representations in long-term memory (e.g., Hulme, Roodenrys, Schweickert, Brown, Martin, & Stuart, 1997; Roodenrys, Hulme, Lethbridge, Hinton & Nimmo, 2002).

This thesis explores the alternative possibility that articulatory factors play a more integral role in memory and speech functions (particularly vSTM function) than has been traditionally assumed, and that this is made possible by the embodiment of articulatory detail in processes that support these functions. This position is difficult to reconcile with conventional notions that vSTM performance is supported by dedicated systems and mechanisms that operate on item-level phonological representations. Instead, it is argued that memory and speech functions are serviced by speech motor control processes that can be co-opted to retain ordered sequences of verbal information by recoding them into an articulatory form.

On what basis should we expect vSTM function to be serviced by speech motor control processes that embody articulatory detail? Firstly, vSTM and speech production show remarkable similarities in terms of performance patterns and error types (e.g., Ellis, 1980; Acheson & MacDonald, 2009). For example, phonological similarity often results in exchange errors in speech production (as in 'she sells sea shells by the she sore', where the onsets of the final two words are exchanged). Phonological similarity leads to comparable ordinal exchange errors in vSTM tasks (e.g., Fallon, Groves & Tehan, 1999). Exchange errors also exhibit clear patterns in terms of the positions between which they occur. In speech production, misplaced phonemes tend to appear within one or two words of the correct position (e.g., Shattuck-Hufnagel, 1979). Similarly, items recalled in incorrect serial positions in vSTM tasks are unlikely to appear more than one or two items from the correct position (e.g., Haberlandt, Thomas, Lawrence & Krohn, 2005). As another example, primacy and recency effects typically result in U-shaped distributions of correct performance in serial recall tasks (e.g., Murdock, 1962). A similar

pattern is observed in the accurate production of syllables in isolated multi-syllable nonwords such as 'keegainysannogeeray', where the onset and offset syllables are more likely to be articulated correctly (e.g., Gupta, Lipinski, Abbs & Lin, 2005).

These common patterns suggest the existence of common underlying mechanisms that support both vSTM and speech functions. Psycholinguistic theorists argue that these similarities reflect the dependency of vSTM functions on the language architecture, as opposed to memory-specific mechanisms (e.g., Martin & Saffran, 1997; MacDonald & Christiansen, 2002; Acheson & MacDonald, 2009). However, an alternative perspective is that cognitive functions are more generally supported by the opportunistic deployment of distributed perception and action-oriented processes to fulfil the demands of those cognitive tasks to which their capacities are well-suited. vSTM and language production happen to share common task demands (i.e., ordered behaviour, short-term retention) that can be fulfilled by the deployment of speech motor control processes.

Secondly, previous studies have demonstrated that vSTM performance can be improved by increasing the *coarticulatory fluency* of verbal sequences that must be remembered. Coarticulation refers to the accommodations made by the speech apparatus between articulatory gestures (particularly those that straddle the boundaries between words) in order to produce fluent, connected speech. Depending on the anatomical properties of the particular gestures that are coarticulated, some coarticulations can be implemented more fluently (i.e., more quickly and efficiently) than others. Hence, coarticulatory fluency refers to the ease with which the boundaries between words can be negotiated by the speech apparatus. In this case, verbal sequences that do not involve any changes

3

in place of articulation (i.e., a reconfiguration of the speech apparatus to form a speech constriction at a different site within the vocal tract) at word boundaries are better-remembered than sequences that involve a change in place of articulation at each word boundary.

Because previous research has typically examined articulatory effects in restricted contexts that involve single items or pairs of items (e.g., Caplan, Rochon & Waters, 1992; Baddeley & Andrade, 1994), measurements have failed to detect influential coarticulatory fluency effects that only emerge in longer sequences (e.g., Murray & Jones, 2002; Woodward, Macken & Jones 2008). It is difficult to account for these sequence-level articulatory constraints on vSTM performance in terms of the item-level phonological mechanisms (e.g., phonological storage; trace redintegration) that have conventionally been argued to service vSTM function. This evidence can be better accommodated by the position that vSTM function is serviced by speech motor control processes that embody articulatory detail.

Empirically, this thesis focuses on the effects of articulatory effort minimization processes in the context of vSTM tasks. This context allows for contrasts to be drawn between the conventional view that vSTM function is serviced by memory-specific phonological mechanisms and the alternative view that vSTM function is supported by speech motor control processes that embody articulatory detail. According to this alternative view, articulatory effects that constrain vSTM performance should not be considered as memory effects per se. This is because they originate in speech motor control processes that can be deployed to support other cognitive functions that involve similar task demands. That is, the same articulatory effects that constrain performance in vSTM tasks should also emerge outside of vSTM contexts.

For the purposes of the present investigation, effort minimization processes are split into two categories - *coarticulatory* processes that increase the fluency of articulatory transitions between words, and *lenition* processes that reduce the difficulty and complexity of articulatory features within words. If vSTM function depends on the action of speech motor control processes, reductions in articulatory complexity resulting from coarticulation and lenition should increase the efficacy of those speech motor control processes argued to support vSTM function, leading to improvements in performance.

Approach and aims

Chapter 1

Previous evidence for coarticulatory fluency effects in vSTM (Murray & Jones, 2002) has been reinterpreted as a consequence of redintegration. This reinterpretation was made possible by confounds between coarticulatory fluency and phonological neighbourhood density – a linguistic property known to facilitate vSTM performance (e.g., Roodenrys et al., 2002). Specifically, the words used to construct fluent sequences belonged to denser phonological neighbourhoods (Miller, 2010), possessing more phonologically similar neighbours (i.e., words that differ from a specified word by a single phoneme).

The first aim of the investigation undertaken in Chapter 1 is to show that coarticulatory constraints are neither peripheral to vSTM performance (i.e., explicable as a consequence of misarticulation at output) nor open to reinterpretation in terms of item-level phonological processes. The approach taken here is to devise a manipulation of coarticulatory fluency that can be applied by reordering the same set of verbal items. As well as controlling for variations in PND, this will control for variations in any item-

level properties. This manipulation can be combined with an order reconstruction task to eliminate the requirement for participants to articulate their responses, together with the possibility that any effects of the manipulation are due to misarticulation.

The second aim of this chapter is to show that coarticulatory fluency effects in vSTM performance cannot be explained by memory-specific mechanisms or even characterised as memory effects *per se*. The approach taken here is to look for evidence of coarticulatory effects outside of the context of vSTM tasks, in inner speech: Although vSTM tasks may place demands on inner speech processes, inner speech reading tasks will involve no memory demands. Measurements of reading times for fluent and disfluent sequences are compared between overt (i.e., vocalised) and inner speech (i.e., silent speech without lip movement) with the expectation that coarticulatory fluency factors will constrain both similarly.

Chapters 2 & 3

Chapter 1 aims to establish that coarticulatory fluency effects in vSTM cannot be reinterpreted as a consequence of item-level phonological mechanisms that are specific to memory. Chapters 2 and 3 take a different approach, exploring the possibility that effects of linguistic properties on vSTM performance, as are typically attributed to a phonological redintegration process, can instead be explained as a consequence of articulatory effort minimisation processes such as *lenition*, a language change process that reduces the articulatory complexity of affected words (e.g.,Bybee, 2010). Reductions in articulatory complexity should allow verbal materials to be more easily manipulated by the speech motor control processes argued to support vSTM function in Chapter 1.

Chapter 2

Words from dense phonological neighbourhoods (i.e., words with numerous similar-sounding 'neighbours' that differ by a single phoneme) are better remembered in vSTM tasks. This advantage is usually explained in terms of redintegrative mechanisms. However, this advantage depends more fundamentally on systematic patterns in the phonological distributions that underlie phonological neighbourhood density (PND) effects. Whereas some words belong to dense phonological neighbourhoods, others belong to more sparsely populated neighbourhoods; if words were distributed evenly across phonological space, there would be no basis for differential PND effects. Nevertheless, little consideration has been given to the reasons behind these systematic variations in phonological distributions. It is argued here that these systematic patterns in phonological distributions can be partly explained by pressures towards effort minimization. That is, densely-populated regions of phonological space will cluster around easier articulatory configurations. In order to test this hypothesis, a measure of articulatory difficulty is devised based on a combination of anatomical parameters. Chapter 2 investigates whether differences in PND, both in a sample of English words and the materials used in past experiments, are confounded with articulatory difficulty as quantified by this omnibus measure.

Chapter 3

Frequently-used words receive stronger support from redintegrative mechanisms in vSTM tasks. However, high-frequency words are also particularly susceptible to the language change process lenition, which reduces the articulatory complexity of affected words. Previous research (e.g., Murray & Jones, 2002) demonstrates that reductions in (co)articulatory complexity lead to better memory for verbal materials. Consequently, it is possible that better memory for frequently-occurring

words can be explained by the articulatory effects of lenition rather than a phonologically-oriented redintegration process. Chapter 3 explores methods for experimentally inducing lenition via contextual manipulations while holding frequency at a constant value. Ultimately, a successful contextual manipulation of lenition could be used to constrain vSTM performance.

Further discussion of relevant concepts and literature is provided in the appropriate chapters.

Chapter 1

Evidence for a sequence-level coarticulatory constraint in verbal short-term memory and inner speech

Conventional understanding of verbal short-term memory (vSTM) focuses on memory-specific mechanisms that operate on phonological items, to the neglect of articulatory factors and processes that operate on verbal sequences. Previous work has demonstrated superior memory for verbal sequences that involve fluent coarticulatory transitions between items. However, this evidence was left open to reinterpretation in terms of item-level redintegration processes due to confounds between the sequence-level manipulation of coarticulatory fluency and item-level properties known to improve vSTM performance. This problem is redressed here by using a novel manipulation of coarticulatory fluency that can be implemented across a single set of verbal items simply by reversing their order, thereby eliminating any variation in item-level properties. Superior memory for sequences with fluent coarticulations persists when item-level properties are controlled for in this manner. Performance advantages for sequences involving fluent coarticulations also extend beyond the context of vSTM tasks to inner speech, where sequences involving fluent coarticulations are read faster than disfluent counterparts. It is argued that these coarticulatory fluency effects reflect the dependency of vSTM function on speech motor control processes that operate in inner speech and embody articulatory detail.

Introduction

Past efforts to understand short-term memory for verbal materials (vSTM) have focused on effects that operate at the level of single items. These effects relate to the properties of particular words in to-be-remembered sequences, and how differences in the properties of these words can constrain vSTM performance. For example, frequently-encountered words tend to be better-remembered in vSTM tasks (e.g., Hulme et al.,1997). The same can be said for short words (e.g., Baddeley, Thomson & Buchanan, 1975) and words with concrete rather than abstract meanings (e.g., Walker & Hulme, 1999). By comparison, the role of sequence-level effects that operate across and between items has been largely overlooked (e.g., Woodward et al., 2008). Moreover, because documented item-level constraints on vSTM performance are numerous and ostensibly well-understood, new vSTM phenomena are increasingly likely to be interpreted and understood as item-level effects.

This chapter focuses on the effects of sequence-level coarticulatory fluency on vSTM. Coarticulation refers to the accommodations made by the speech apparatus (such as the tongue and lips) between speech gestures in order to produce fluent, connected speech. This includes accommodations between speech gestures that straddle word boundaries. Depending on the anatomical properties of the particular gestures that are coarticulated, some coarticulations can be implemented more fluently (i.e., more quickly and efficiently) than others. Hence, coarticulatory fluency refers to the ease with which the boundaries between words can be negotiated by the speech apparatus. As an example, the coarticulatory boundary between 'lap' and 'bat' is relatively easy to negotiate given that the offset of 'lap' and the onset of 'bat' are both articulated with the lips. By comparison, the boundary between 'lap' and 'get' is more difficult to

negotiate: While the offset of 'lap' is articulated with the lips, the onset of 'get' is implemented with the tongue body - a different articulator. Short-term memory for verbal sequences involving these difficult (i.e., complex or disfluent) coarticulatory transitions between items is worse than memory for sequences involving easy (i.e., simple or fluent) coarticulatory transitions (e.g., Murray & Jones, 2002): Henceforth, this phenomenon is referred to as a *coarticulatory fluency effect* in vSTM.

Coarticulatory fluency effects are one of a handful of sequence-level effects that have been implicated in vSTM performance, although they have come to light more recently than others, such as grouping effects (e.g., Harris & Burke, 1972). Because previous research has typically examined articulatory effects in restricted contexts that involve single items or pairs of items (e.g., Caplan, Rochon & Waters, 1992; Baddeley & Andrade, 1994), measurements have failed to detect coarticulatory effects that only emerge in longer sequences (e.g., Murray & Jones, 2002; Woodward et al., 2008). Although coarticulatory effects can be measured in terms of sequence duration (i.e., sequences with simpler coarticulations tend to have shorter articulatory durations), their influence on vSTM appears to be explained by differences in articulatory complexity/fluency rather than duration: Even when matched on articulatory duration, verbal materials with fewer syllables or fewer different phonemes are better-remembered in immediate recall tasks (e.g., Service, 1998).

Although experimental work offers evidence that vSTM is constrained by the sequence-level factor of coarticulatory fluency (e.g., Murray & Jones, 2002; Woodward et al., 2008), this evidence is viewed with some scepticism. This is because experimental manipulations of coarticulatory fluency are confounded by variations in item-level properties (such as phonological neighbourhood density - e.g., Miller, 2010) that prove

difficult to fully control for and allow for reinterpretations of coarticulatory fluency effects in terms of item-oriented phonological mechanisms (such as trace redintegration - e.g., Hulme et al., 1997; Roodenrys et al., 2002).

Evidence for a coarticulatory fluency constraint on vSTM comes from a study where the time taken for English-Welsh bilinguals to produce digit sequences was measured in each language (Murray & Jones, 2002). Spoken in isolation, the English digits 1 to 9 were produced more slowly (on average) than their Welsh counterparts (at 488ms vs. 456ms). However, when spoken in nine-digit sequences, the same English digits were produced more quickly (on average) than sequences of corresponding Welsh digits (at 255ms vs. 294ms). Therefore, the longer duration of Welsh digit sequences cannot be accounted for purely in terms of item-level properties. If this were the case, shorter Welsh digits should combine to form shorter Welsh digit sequences. Conversely, shorter Welsh digits combine to form longer Welsh sequences. Two implications can be drawn from this. Firstly, the time taken to articulate a verbal sequence is constrained by properties relating to the sequence as a whole, as well as properties relating to particular items. Secondly, these sequence-level properties are influential: English digits are subject to a sequence-level production advantage that does not merely offset item-level effects (i.e., the production advantage for isolated Welsh digits) but overturns them.

The faster production of English digit sequences is accounted for by a constraint on the fluency with which different speech gestures are coarticulated across word boundaries. On average, negotiating the coarticulatory boundaries between the Welsh digits 1 to 9 (i.e., from the offset of one digit to the onset of the next) necessitates more changes in *place of articulation* than the corresponding English digits. A change in place of articulation is a reconfiguration of the speech apparatus to form a

constriction at a different site within the vocal tract. For the purposes of this study, a coarticulatory transition entails a change in place of articulation when different active articulators (i.e., the lower lip, the front of the tongue and the body of the tongue) are recruited across a word boundary. Whereas eight of the coarticulatory transitions between the Welsh digits do not involve any changes in place of articulation, 18 of the transitions between the corresponding English digits do not involve any change in place of articulation. To illustrate, the coarticulatory boundary between the English digits 'seven' and 'two' involves two *medial,* coronal gestures, /n/ and /t/, both of which are articulated with the tongue tip. Therefore, no change in place of articulation is required. However, the coarticulation between the Welsh digits 'pump' and 'naw' requires a change in place of articulation from the *anterior* labial constriction /p/ (articulated with the lips) to the *medial* coronal constriction /n/ (articulated with the tongue tip). In summary, the fluency with which a given verbal sequence can be produced decreases with the number of changes in place of articulation; these fluency costs are reflected in sequence-level measures of articulatory duration.

Ostensibly, verbal sequences that can be more fluently produced should also be better-remembered in vSTM tasks due to more efficient articulatory rehearsal. The authors (Murray & Jones, 2002) tested this hypothesis experimentally: Coarticulatory fluency was manipulated by designing verbal sequences to include or exclude changes in place of articulation at word boundaries. For a disfluent condition, eight-item sequences of English consonant-vowel-consonant words were constructed such that all of the boundaries between words involved changes in place of articulation. For example, in the disfluent sequence 'tape, knife, turf...', the coarticulatory boundary between 'tape' and 'knife' involves a change from

the labial /p/ (articulated with the lower lip) to the coronal /n/ (articulated with the tongue tip). For a fluent condition, a second set of eight-item sequences was constructed such that the coarticulatory boundaries between words did not involve any changes in place of articulation. For example, in the fluent sequence 'rail, rice, nurse...', the boundary between 'rail' and 'rice' involves a transition from the coronal /l/ (articulated with the tongue tip) to another coronal, /r/ (also articulated with the tongue tip). Participants performed an order reconstruction task on these fluent and disfluent sequences: Sequences items were presented one at a time before reappearing together in a scrambled order. Participants were then required to select these scrambled items in their original order of presentation. Order reconstruction performance (i.e., the mean percentage of items correctly selected in their original order of presentation) was better for fluent sequences (at 61.1%) than for disfluent sequences (at 54.2%). This suggests that vSTM is constrained by the sequence-level property of inter-item coarticulatory fluency.

To isolate the sequence-level effect of their coarticulatory fluency manipulation, Murray and Jones matched their experimental materials on lexical frequency, an item-level property known to facilitate vSTM performance (e.g., Hulme et al., 1997). However, more recent examination (Miller, 2010) reveals that these materials were not matched on another item-level property known to facilitate vSTM performance - phonological neighbourhood density (PND - e.g., Roodenrys et al., 2002). This refers to the number of similar sounding neighbours a word possesses (i.e., those which differ by a single phoneme - therefore 'cat' and 'bat' are phonological neighbours of 'rat'). In fact, the facilitative item-level effects of PND were confounded with the sequence-level manipulation of coarticulatory fluency. Specifically, fluent sequences without changes in place of articulation at

word boundaries contained words from denser phonological neighbourhoods (mean PND = 31.38) than disfluent sequences involving changes in place of articulation (mean PND = 17.75; Miller, 2010). Because of this confound, it becomes unclear whether the observed memory effect was genuinely caused by the sequence-level manipulation of coarticulatory fluency; well-documented item-level mechanisms stand ready to offer competing interpretations.

One such hypothetical item-level mechanism is redintegration. This is a process by which short-term memory traces that have become degraded due to decay or interference are reconstructed from corresponding representations in long-term memory. The reconstruction process is more effective for high-frequency words that have highly available and accessible long-term representations (Hulme et al., 1997). It is also more effective for words from dense phonological neighbourhoods. This is because words from a given phonological neighbourhood form a network linked by mutual excitatory connections (e.g., Roodenrys et al., 2002). When any word from this network is presented, all its phonological neighbours are also activated to some degree. Via its mutual connections with these neighbours, the presented word receives additional supporting activation. Therefore, the more phonological neighbours a word has (i.e., the higher its PND value), the more supporting activation it receives during retrieval. Ultimately, the redintegrative process selects the word with the most activation as the basis for output. Therefore, via associative links with numerous phonological neighbours, high-PND words stand to receive superior redintegrative support in memory tasks.

Put in context, the PND confound in Murray and Jones is a symptom of a broader problem. PND is one of many item-level variables that can contribute to shaping vSTM performance via at least one item-level

mechanism (redintegration): Item-level constraints on vSTM are pervasive. Just as Murray and Jones controlled for lexical frequency but not PND, a subsequent experiment might control for PND but fail to anticipate the impact of another item-level variable – possibly one with an as-yet undocumented influence on vSTM. Even if these additional variables could be anticipated, it quickly becomes impractical to match sets of verbal materials on numerous item-level properties. It is a simple matter to find two sets of materials that match on a single criterion such as frequency, but significantly more difficult to find sets of materials that match on two criteria, such as frequency and PND. Given a finite pool of verbal materials, the precision with which sets of these materials can be matched suffers as more matching criteria are specified. On balance, matching is an imperfect control strategy. Yet, so long as potential item-level confounds remain imperfectly controlled, sequence-level interpretations for coarticulatory fluency effects on vSTM remain in doubt. Hence, an alternative solution to the problem of item-level control is called for.

For this solution, we turn to an alternative constraint on coarticulatory fluency. Given that coarticulation is a mechanically complex behaviour, coarticulatory fluency is constrained by anatomical characteristics besides the magnitude of changes in place of articulation. Findings from electropalatography research (where articulatory movement is measured via an electrode array attached to the tongue) reveal an asymmetry in the degree to which stop consonants (i.e., consonants involving a complete blockage of airflow, such as /b/, /d/ or /g/) are temporally overlapped across word boundaries. Overlap is a hallmark of efficiently coarticulated speech that involves the simultaneous production of adjacent speech gestures. That is, the production of a second gesture begins before the production of a preceding first gesture is complete.

Therefore, as overlap between the gestures increases, their combined production time (and complexity) is reduced, resulting in superior coarticulatory fluency.

To elaborate on this overlap asymmetry, inter-item coarticulations involving backward-moving changes in place of articulation (i.e., from a given articulator to a more posterior articulator) between stop consonants are more overlapped than corresponding forward-moving changes (Byrd, 1996). For example, a backwards-moving change from a /d/ articulated with the tongue tip to a /g/ articulated with the tongue body (as in 'bad-gab') affords more overlap than the reverse transition between /g/ and /d/ (as in 'bag-dab').

Superior overlap for backward-moving changes stems from anatomical constraints on coarticulated speech (Chitoran, Goldstein & Byrd, 2002). An overlapped forward-moving change in place of articulation (for example, from /g/ to /b/) requires that a secondary *anterior* constriction (such as /b/, formed with the lips) is formed just prior to the release of air trapped behind a primary *posterior* constriction (such as /g/, formed with the tongue body). However, the air expelled by the release of the *posterior* constriction /g/ cannot exit through the front of the mouth while it is blocked by the *anterior* constriction /b/. Therefore, overlap must be sacrificed to preserve the perceptual impact of the posterior gesture /g/. This sacrifice is unnecessary for an equivalent backward-moving change (for example, from /b/ to /g/). In this case, the primary constriction /b/ occupies the front of the vocal tract, where it can be released without interference from the simultaneous formation of a secondary *posterior* constriction /g/.

This coarticulatory fluency constraint differs from the constraint exploited in previous research (e.g., Murray & Jones, 2002; Woodward et

al., 2008) in that it is not based on the presence of a change in place of articulation, but on the direction of the change. This means it can be exploited to implement a manipulation of coarticulatory fluency across a set of identical items: A word with a *posterior* onset and offset (such as 'gig') can be followed by a word with a *medial* onset and offset (e.g., 'dad'), which in turn is followed by a word with an *anterior* onset and offset (e.g., 'bob'). The boundaries between the words involve changes in place of articulation that move incrementally forward through the vocal tract, resulting in a disfluent word sequence (e.g., 'gig-dad-bob'). By reversing the order of these items, the direction of the inter-item coarticulations can also be reversed to produce a fluent sequence that involves backward-moving inter-item coarticulations (e.g., 'bob-dad-gig'). In this fashion, inter-item coarticulatory fluency can be manipulated while eliminating variations in item-level properties entirely, since the same items are utilised in each case. This manipulation can therefore be used to measure the genuine influence of inter-item coarticulatory fluency on vSTM, free from the contamination of item-level confounds.

Experiment 1

Backwards-moving changes in place of articulation between stop consonants (e.g., /b/ - /g/) are more fluently overlapped than corresponding forward-moving changes (e.g., /g/ – /b/). However, unlike the magnitude constraint on coarticulatory fluency employed in previous work (major versus minor changes in place of articulation - e.g., Murray & Jones, 2002; Woodward et al., 2008) this directional constraint has not yet been tested in a vSTM context. Experiment 1 tests whether the directional constraint influences vSTM performance under similar conditions to those used in Murray & Jones (2002). That is, fluent and disfluent sequences are constructed from different sets of words that are matched on item-level

properties - in this case, both frequency and PND. In a repeated measures design, participants perform serial recall on fluent and disfluent sequences. These sequences are presented visually, and participant output is spoken.

The direction of inter-item coarticulations is manipulated across different six-item sequences of English words. Words with *posterior* onsets and *anterior* offsets (e.g., 'nap') are combined to generate fluent sequences with backwards-moving changes in place of articulation between each word (e.g., 'nap-doom-ripe-ship-jeep-loop'). To illustrate, the transition between 'nap' and 'doom' involves a backwards-moving change from the anterior labial offset 'p' (articulated with the lips) to the more posterior coronal onset 'd' (articulated with the tongue tip). Conversely, words with *anterior* onsets and *posterior* offsets (e.g., 'fan') are combined to generate disfluent sequences with exclusively forwards-moving changes in place of articulation between words (e.g., 'veil-boon-fan-peas-budge-vice'). Memory performance is assessed in a serial recall task with spoken output. Given that backward-moving inter-item coarticulations involve superior temporal overlap, fluent backward-moving sequences should be better remembered.

Method

Participants

Twenty-two participants completed the experiment in return for course credit or a payment of £3. These were Cardiff University undergraduate students (three male and nineteen female, between the ages of eighteen and twenty-five), all native English speakers reporting normal/corrected hearing and vision. Informed written consent was obtained from all participants.

Materials and Procedure

The consonants /b/, /f/, /m/, /p/, /v/, /k/, /tʃ/, /d/, /dʒ/, /n/, /l/, /r/, /s/, /ʃ/, /t/, /θ/, and /z/ were categorised according to place of articulation. In this case, the first category (anterior) contained consonants involving an *anterior*, labial place of articulation recruiting the lips. The second category (posterior) contained consonants involving both *medial* and *posterior* places of articulation recruiting the tongue tip, tongue body or glottis. The MRC Psycholinguistic database (Wilson, 1988) was searched exhaustively for English consonant-vowel-consonant (CVC) words corresponding to one of two formats – an *anterior* onset and *posterior* offset (e.g., 'pot'), or a *posterior* onset and *anterior* offset (e.g., 'tap'). From the results, two pools of CVCs were created according to onset location. These pools were reduced to a size of forty-eight items each by matching their contents on mean lexical frequency (specifically, CELEX frequency – e.g., Baayen, Pipenbrock & Van Rijn, 1995), at 20.72 (SD = 16.69) for the *posterior* onset pool and 20.96 (SD = 17.74) for the *anterior* onset pool; $t(48) = 0.07$, $p = 0.944$. The material pools were also matched on mean phonological neighbourhood density, at 16.58 (SD = 5.37) for the *posterior* onset pool and 16.74 (SD = 6.30) for the *anterior* onset pool; $t(48)=0.14$, $p = .89$. The linguistic statistics program N-watch (e.g., Davis, 2005) was used to facilitate this matching process.

Posterior-onset CVCs were used to construct six-word sequences involving fluent backwards-moving changes in place of articulation at word boundaries (e.g., 'nap-doom-ripe-ship-jeep-loop'), and anterior-onset CVCs were used to construct six-item sequences with disfluent forwards-moving changes (e.g., 'veil-boon-fan-peas-budge-vice'). Sequences were constructed for each experimental trial by randomly recruiting items from the appropriate item pool.

Stimuli were presented centrally on a computer monitor in black font, using Matlab software including Psychophysics Toolbox extensions (Brainard, 1997, Kleiner, Brainard & Pelli, 2007). Participants were tested individually while seated in a soundproof booth, where their responses were recorded for the duration of the experiment (approximately thirty minutes) using a microphone. Participants commenced each trial by pressing the space bar on a keyboard; on each trial, six words were presented one at time, each for 750ms, with a 750ms interstimulus interval. Presentation was followed by a ten-second retention interval during which participants subvocally rehearsed the sequence while fixating an onscreen cross. Participants were then prompted by an onscreen message ('Recall now') to speak the sequence aloud in its original order; if unsure of a word, participants guessed a response or said 'pass'. Before beginning the experiment, participants completed six practice trials (using three sequences from each condition, generated in the same manner as experimental sequences) under the experimenter's supervision, to check their understanding of the task. As part of a repeated-measures design, participants completed thirty-two experimental trials from each condition, distributed randomly across sixty-four total trials. These were divided into four sixteen-trial blocks, after each of which participants were prompted to take a short break.

Results and Discussion

Recorded participant responses were transcribed and marked against a log of presented items. Credit was only awarded for the recall of a correct item in its original serial position. This method was used to obtain an overall measure of performance for each participant under each condition. Although the effect was small (at $\eta_p^2 = .19$), mean correct performance for fluent sequences was significantly better (54.63%; SD =

15.08) than for disfluent sequences (51.21%; SD =15.51), F (1, 21) = 4.920, p = .038 (η_p^2 = .19). Figure 1 depicts mean correct performance as a function of sequence type and serial position. A two-way repeated measures ANOVA indicates there was no significant interaction between fluency and serial position: F (5, 105) = .742, p = .594 (η_p^2 = .034).
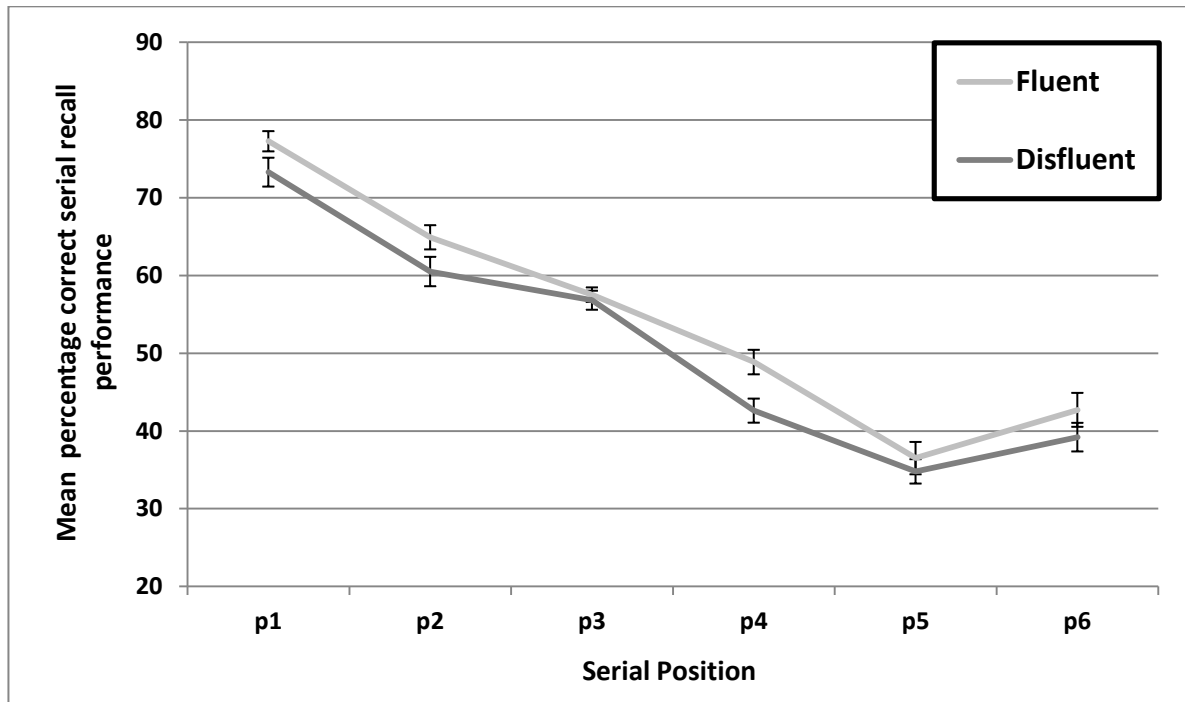
Figure 1. Mean percentage correct serial recall performance as a function of speech type and serial position. Error bars show Standard Error.

As expected, serial recall performance was significantly better for sequences with backward-moving inter-item coarticulations. This result corroborates the findings of previous research in which coarticulatory fluency influences vSTM performance (e.g., Murray & Jones, 2002). It also validates the use of the directional constraint (i.e., superior overlap for backward-moving changes in place of articulation at word boundaries) as an effective manipulation of coarticulatory fluency.

However, the experiment is not without caveats. For example, the directional coarticulatory fluency constraint applies selectively to stop consonants that involve a complete blockage of airflow through the oral cavity (i.e., /b/, /p/, /t/, /d/, /k/, /g/). These consonants can only be used to generate limited pools of English CVC words that cannot be sensitively matched on the item-level properties of both frequency and PND. In order to diversify the pool of available materials and allow for more sensitive matching, the stop-consonant restriction was lifted here. This means the experimental materials incorporated consonants that do not involve stoppage (including fricatives such as /f/ and approximants such as /l/), resulting in a dilution of the fluency manipulation. This dilution may have contributed to the small size of the fluency effect. Moreover, although this compromise allowed for the materials recruited in the fluent and disfluent conditions to be matched on frequency and PND, the potential remains for item-level confounds of unanticipated importance. For example, given the method used here for constructing fluent and disfluent sequences, fluent sequences will always involve words with posterior consonantal offsets and anterior offsets. A further issue is that the use of spoken output makes it unclear whether the observed coarticulatory fluency effect arises from memory/rehearsal processes or is merely an output effect (i.e., participants remember disfluent sequences correctly but misproduce them at output).

Experiment 2a

Experiment 2a employs a manipulation of coarticulatory fluency that can be applied to sets of identical items simply by reversing their order. This obviates the need to match materials on item-level properties by eliminating item-level variations and confounds entirely. This in turn allows a test of whether vSTM is genuinely constrained by inter-item coarticulatory fluency. Nonword materials were employed as an additional control in order

to limit the potential effects of pre-existing inter-item associations on vSTM (e.g., Stuart & Hulme, 2000). In a repeated measures design, participants performed serial recall on fluent and disfluent nonword sequences. As in Experiment 1, sequences were presented visually and participant output was spoken.

To implement the fluency manipulation, disfluent sequences were constructed from nonwords whose onsets and offsets move incrementally forward through the vocal tract. For example, the nonword 'kug' (which has a *posterior* onset and offset, articulated with the tongue body) is followed by 'dord' (with a *medial* onset and offset, articulated with the tongue tip), which is then followed by 'pobe' (which has an *anterior* onset and offset, articulated with the lips). Therefore, each inter-item coarticulation involves a change to a more anterior place of articulation. Fluent sequences were generated by reversing the order of disfluent sequences to produce a series of backward-moving coarticulations (e.g., 'pobe, dord, kug').

Six-item sequences were used in the experiment. This raises the limitation that incrementally forward-moving coarticulations cannot continue for more than three items. For example, the final offset in the forward-moving sequence 'kug, dord, pobe' (i.e., /b/) is articulated with the lips. Because no further anterior places of articulation remain, the speech apparatus must be returned to their original posterior configuration in order for forward movement to continue. This necessitates a discrepant backward-moving change, as can be observed in the centre of the otherwise forward-moving sequence 'kug, dord, pobe, geg, dat, bup'. The presence of this fluent change in a disfluent sequence threatens to dilute the directional effect of the manipulation. However, coarticulation can be eliminated by a prosodic boundary (i.e., a pause in speech - e.g., Cho & Keating, 2001). Therefore, in order to overcome this limitation, experimental

sequences were grouped into two sets of triplets at presentation: The first three items in each sequence were presented consecutively on the left side of the screen. These were followed after a 750ms pause by the latter three items, presented consecutively on the right side of the screen.

Method

Participants

Thirty-four participants were recruited from the same demographic as in Experiment 1, in return for course credit or a payment of £3. None of these had participated in the previous experiment.

Materials and Procedure

Stop consonants (i.e., consonants that involve a compete blockage of airflow) were combined with vowels to generate three pools of Consonant-Vowel-Consonant (CVC) nonwords, each containing twenty-two items. The first pool contained nonwords with *anterior* labial onsets and offsets involving the lips (e.g., 'bip'), the second contained nonwords with *medial* coronal onsets and offsets involving the tongue tip (e.g., 'tet') and the third contained nonwords with *posterior* velar onsets and offsets involving the tongue body (e.g., 'geg').

Disfluent six-item sequences were constructed by selecting a random nonword from the *posterior* pool followed by nonwords from the *medial* and *anterior* pools. This procedure generates sequences with inter-item coarticulations that move incrementally forward through the speech apparatus (e.g., 'kug (*posterior*), dord (*medial*), pobe (*anterior*), geg (*posterior*), dat (*medial*), bup (*anterior*)'), with the exception of the central coarticulation (see below). Fluent sequences were generated by reversing the order of disfluent sequences to produce a series of backward-moving

coarticulations (e.g., 'bup, dat, geg, pobe, dord, kug'). As part of a repeated-measures design, thirty-two pairs of fluent-disfluent sequences were generated for each participant. From these, sequences were selected randomly without replacement for each of the sixty-four experimental trials.

Participants were tested in a sound-attenuating booth, where sequences were presented on a computer monitor as in Experiment 1. Again, participant responses were recorded using a microphone for the duration of the experiment. Each trial commenced with a central fixation cross, displayed for 750ms. Six nonwords were then presented serially, each for 750ms with no interstimulus interval (in order to encourage coarticulation). The first three nonwords were presented on the left side of the screen. These were followed after a blank 750ms pause by the latter three nonwords on the right side of the screen. A ten-second retention interval ensued, during which participants were instructed to subvocally rehearse the nonword sequence while fixating a central onscreen cross. Finally, the impact of the fluency manipulation on vSTM was measured via spoken output in a serial recall task as in Experiment 1. Before commencing the experiment, participants completed six practice trials (three from each condition) under the supervision of the experimenter to check their understanding of the task.

Results and Discussion

Recorded responses were transcribed and scored as in Experiment 1 to obtain an overall measure of performance for each participant under each condition. Figure 2 depicts mean percentage correct performance (i.e., the proportion of nonwords recalled correctly and in their original serial position) as a function of sequence type and serial position. A two-way repeated measures ANOVA indicates that mean correct performance for

fluent sequences (28.43%; SD = 16.13) was significantly better than for disfluent sequences (26.07%; SD = 15.20): $F(1, 33) = 6.542$, $p = .015$ ($\eta_p^2 = .17$). There was also a significant interaction between fluency and serial position: $F(5, 165) = 4.538$, $p < .001$ ($\eta_p^2 = .12$).

This significant interaction between sequence type and serial position was unexpected, and is difficult to account for. The interaction centres on an anomaly at p2, where the otherwise consistent pattern of superior serial recall for fluent over disfluent sequences is reversed. Whereas performance for disfluent sequences conforms to a typical s-shaped function, performance for the fluent sequences does not, suggesting that the anomaly can be explained by some property that is unique to fluent sequences. One possibility is that an unforeseen difficulty in the transition between anterior labial consonants and medial coronal consonants accounts for the dip in performance at p2. However, given that a similar transition occurs at p5, in the absence of a similar anomaly, this seems unlikely.
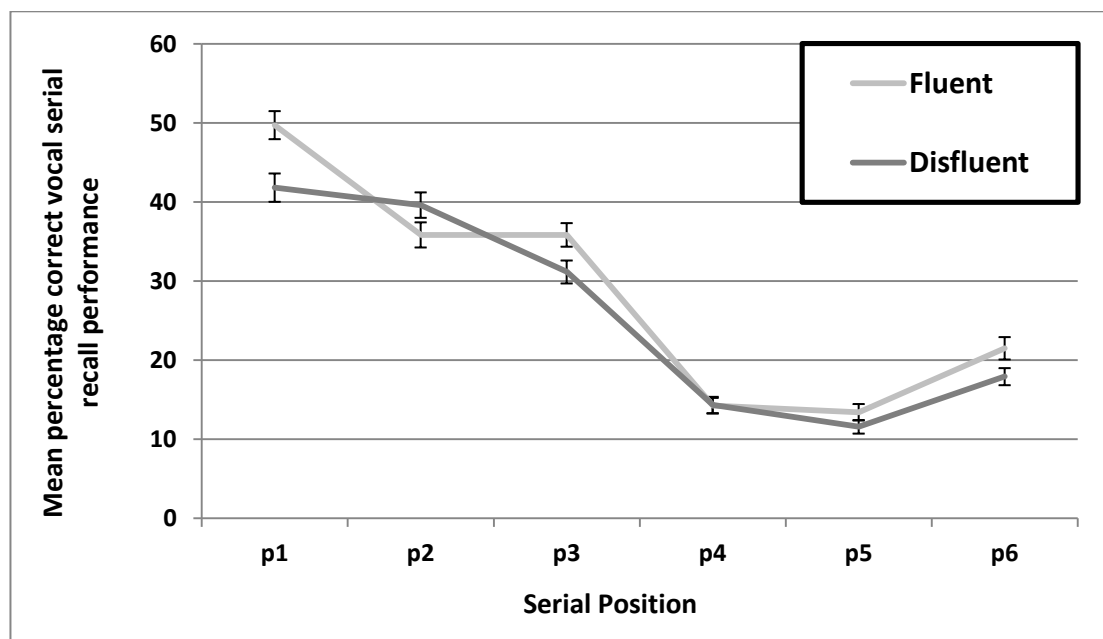

Figure 2. Mean percentage correct serial recall performance as a function of speech type and serial position. Error bars show Standard Error.

As hypothesised, fluent sequences were better remembered than their disfluent counterparts. Specifically, spoken serial recall performance was superior for nonword sequences involving more fluent backward-moving coarticulations between items. Not only does this experiment corroborate the findings of previous research (e.g., Murray & Jones, 2002), it shows that these findings hold when the potential for variation in item-level properties is eliminated. Removing this inlet for competing item-level interpretations is an important step in consolidating the argument that inter-item coarticulatory fluency exerts a genuine influence on vSTM.

As in previous work, coarticulatory fluency was manipulated by means of an anatomical constraint. However, where previous work exploits differences in the magnitude of changes in place of articulation, this manipulation exploits differences in the direction of these changes. Due to anatomical constraints on the overlapped production of adjacent stop consonants (Chitoran et al., 2002), backwards-moving coarticulations are more temporally overlapped than their forward-moving counterparts (Byrd, 1996). Therefore, sequences involving backwards-moving changes at word boundaries are more fluently implemented (and shorter in duration) than comparable sequences involving forwards-moving changes at word boundaries. This results in superior vSTM performance.

Experiment 2a provides evidence for a coarticulatory fluency effect in spoken serial recall performance. However, this is not necessarily the same thing as evidence for a fluency effect in memory. Because spoken serial recall involves a significant production component (i.e., responses must be overtly articulated), disfluent sequences might have been misproduced at output rather than misremembered. What appears to be a memory effect may instead be a production artefact.

28

Experiment 2b

Experiment 2b set out to eliminate the possibility that the effect observed in Experiment 2a is a production artefact caused by the requirement to overtly articulate responses. To this end, Experiment 2a was closely replicated using an order reconstruction task instead of spoken serial recall. In this task participants are not required to overtly articulate their responses at any point. Instead, the items from the original sequence are re-presented together in a randomly scrambled order following the retention interval. Participants must reconstruct the original sequence by clicking these scrambled items in their original order. The scrambled items are presented in black font and are recoloured red once clicked; each item can only be selected once.

Method

Participants

Forty-eight new participants were recruited from the same demographic as in previous experiments, for the same payment.

Materials and Procedure

Experiment 2b was identical to Experiment 2a in terms of both materials and procedure, with the sole exception that an order reconstruction task was employed in place of a spoken serial recall task. As in the serial recall task, six nonwords were presented serially, followed by a ten-second retention interval. However, rather than being recalled by participants, the six nonwords were re-presented together in a randomly scrambled order. Participants were required to reconstruct the original sequence by clicking the scrambled items in their original order of presentation. The scrambled nonwords were presented orthographically in

black font and were recoloured red once clicked; each nonword could only be selected once.

Results and Discussion

For each trial, the order in which participants reconstructed sequences was compared against the original order of presentation. A score was then assigned according to the number of items correctly selected in their original serial positions. This was done for each sequence to yield an overall measure of performance for each participant under each condition. Mean correct performance for fluent sequences (67.73%; SD = 15.92) was significantly better than for disfluent sequences (63.51%; SD = 16.04): $F (1, 47) = 13.96$, $p = 0.001$ ($\eta_p^2 = 0.23$). There was also a significant interaction between fluency and serial position: $F (5, 235) = 4.792$, $p < .001$ ($\eta_p^2 = .09$).

Once again, fluent sequences with backward moving inter-item coarticulations were better remembered than their disfluent forward-moving counterparts. In this case, because vSTM performance was measured using an order reconstruction task, participants were not required to overtly articulate their responses at any point. This provides an assurance that the observed vSTM performance effect was not merely a product of overtly misarticulated responses. If the fluency effect is not a product of overt articulation, it must instead originate from articulatory detail that is represented internally.

Experiment 3

Experiment 2b shows that nonword sequences with fluent coarticulations are better-remembered even in an order reconstruction task that does not involve any overt articulation of to-be-remembered sequences. Therefore, the coarticulatory fluency effect cannot be

characterised as a peripheral output effect explained by the misarticulation of verbal sequences. Rather, coarticulatory fluency must directly constrain whatever (ostensibly internal) process is deployed to support performance in vSTM tasks. One possibility is that the coarticulatory fluency effect is not a short-term memory effect *per se*, but originates from articulatory detail embodied in inner speech. By this token, although a fluency effect can be observed in vSTM performance, it is incidental: The fluency effect is not driven by the relative efficiency of vSTM-specific rehearsal processes so much as the difficulty inherent in implementing fluent and disfluent sequences in a medium that embodies articulatory detail (i.e., inner speech). In this case, coarticulatory fluency effects (and articulatory effects in general) should manifest in any task that involves inner speech, including tasks that place no demands on vSTM.

The question of whether inner speech embodies articulatory detail has been addressed in recent investigation outside the context of vSTM tasks. Such investigation has relied on introspective reports of the types of errors that occur in inner speech in order to inform our understanding of its representational nature (e.g., Oppenheim & Dell, 2008; Oppenheim, 2012; Corley, Brocklehurst & Moat, 2011). In particular, the search for articulatory detail has relied on reports of phonemic similarity errors. These are errors that involve the exchange of two similar speech sounds, as is often seen in tongue twisters. For example, 'reef leech' may be misproduced with the similar onsets /r/ and /l/ exchanging to 'leaf reach' (Oppenheim & Dell, 2008).

Because similar-sounding phonemes share articulatory details (for example, both /r/ and /l/ are articulated using the tongue tip), phonemic similarity errors are caused as much by articulatory similarity as by

phonemic similarity. As is more formally instantiated in a spreading activation model (e.g., Dell, 1986), representations of verbal items involve phonological and articulatory levels of detail that map onto one another. When the phoneme /r/ is presented, activation spreads to associated articulatory features such as the recruitment of the tongue tip. Activation then spreads from this articulatory feature to other phonemes that recruit the tongue tip, such as /l/. In some cases, activation for /l/ may exceed the activation for /r/, in which case a phonemic similarity error occurs as /l/ is incorrectly output. Therefore, phonemic similarity effects arise due to competing activation between phonemes with shared articulatory features. According to this model, the presence of a phonemic similarity effect in inner speech signals the presence of articulatory detail; conversely, the absence of a phonemic similarity effect signals an absence of articulatory detail (e.g., Oppenheim & Dell, 2008).

This logic has been applied in experiments where participants recite four-word sequences of similar-sounding words (such as 'lean, reed, reef, leech' – Oppenheim & Dell, 2008) in both overt and inner speech. As they do so, they are required to immediately report any errors they detect in their speech. If reported exchange errors involve more similar than dissimilar phonemes, a phonemic similarity effect is present. This paradigm has produced varied results, from a nonsignificant negative similarity effect (i.e., where exchange errors involve more dissimilar than similar phonemes) in inner speech (Oppenheim & Dell, 2008), to a nonsignificant positive similarity effect (Oppenheim & Dell, 2010), to a significant positive similarity effect (Brocklehurst & Corley, 2009; Corley, Brocklehurst & Moat, 2011). Only the last of these provides evidence for articulatory detail in inner speech, although on closer examination this evidence proves questionable.

If inner speech embodies articulatory detail in the same way as overt speech, phonemic similarity effects should manifest to a similar extent in both speech types. Statistically, this means speech error data should show no interaction between speech type (i.e., overt or inner speech) and phonemic similarity (i.e., phonetically similar or dissimilar phonemes). In a close replication of Oppenheim and Dell (2008), Corley et al. (2011) found a significant main effect of phonemic similarity, together with a non-significant interaction between speech type and phonemic similarity. However, the significance of this interaction, at p = .09, is not so weak as to be easily dismissed: Although a phonemic similarity effect was present in both overt and inner speech, the effect was appreciably weaker in inner speech (e.g., Oppenheim, 2012). Corley et al. suggest that phonemic similarity errors in inner speech may be underreported rather than absent because, unlike in overt speech, there is no sound to monitor (see also Postma, 2000). However, a counterargument is provided by the observation that a similar number of phonemic similarity errors are reported in overt and silently articulated (i.e., mouthed) speech (Oppenheim & Dell, 2010). That is, a reduction in the number of reported similarity errors cannot be accounted for simply by the absence of sound.

Nevertheless, there are other factors that might suppress the detection and reporting of phonemic similarity errors in inner speech. Firstly, it is unclear how acutely participants attend to their inner experience, particularly given the potentially distracting demands of the error report paradigm, where participants must recite sequences in time with a metronome. By comparison, even during silent articulation participants can focus on cues that may help monitor speech errors, such as the configuration of their speech apparatus. Therefore, some inner speech errors may go undetected in the absence of these cues. Secondly,

because participant reports of inner speech errors are necessarily subjective, some errors might be detected but go unreported. Such factors may account for some of the varied results that have been produced by experiments that employ the error report paradigm (e.g., Oppenheim & Dell, 2008; Oppenheim & Dell, 2010; Corley et al., 2011).

The dependency of inner speech investigation on potentially unreliable error reports can be circumvented by utilising a manipulation of coarticulatory fluency. Whereas phonemic similarity effects must be measured in terms of speech errors, the influence ofcoarticulatory fluency effects can be measured in terms of sequence durations (e.g., Murray & Jones, 2002; Woodward et al., 2008). The latter measure makes for a more straightforward paradigm: Participants begin reading a verbal sequence on cue and indicate when they have finished. Arguably, participants will report when they have finished reading sequences in inner speech more reliably than they will detect and report errors in inner speech.

Experiment 3 set out to determine whether temporal advantages afforded by coarticulatory fluency extend from overt to inner speech. To this end, a simple reading task was employed to test inner speech for articulatory detail. In a 2 x 2 repeated measures factorial design, the time taken for participants to read fluent and disfluent nonword sequences was measured in overt and inner speech. Longer sequences were used than in Experiment 2 (sequence length was increased to nine items) in order to increase the number of inter-item coarticulations. This in turn increases the likelihood of detecting differences in articulatory duration between the fluent and disfluent conditions (e.g., Woodward et al., 2008).

In previous research, manipulations of coarticulatory fluency result in shorter production times for fluent sequences in overt speech (Murray &

Jones, 2002). If inner speech embodies articulatory detail in the same way as overt speech, fluent sequences should be read faster than disfluent sequences to a similar degree irrespective of speech type.

Method

Participants

Thirty-two new participants were recruited from the same demographic as in previous experiments, for the same payment.

Materials and Procedure

Stimuli were nine-item CVC nonword sequences, presented visually on a computer monitor. Nonwords were generated by combining two stop consonants with a vowel, using the procedure from Experiment 1. To populate disfluent sequences, a pool of CVCs was generated by combining *medial* coronal onsets with *posterior* velar offsets (e.g., 'teg'; note – whereas medial and velar consonants were grouped into the same 'posterior' category in Experiment 1, they are differentiated here). When these are combined, the coarticulation from the offset of each nonword to the onset of the next involves a forward-moving change from a *posterior* to *medial* place of articulation (e.g., 'dak, deeg, dayg, teg, dook, teeg, durg, dag, darg'). For the fluent sequences, pools of CVCs were generated by combining *posterior* velar onsets with *medial* coronal offsets (e.g., 'gad'), such that the coarticulations between nonwords involve backward-moving changes from *medial* to *posterior* places of articulation (e.g., 'gid, keet, kood, gort, kade, gurt, gad, gat, kide').

Nonwords from these preliminary pools were matched on vowel types to control for potential durational factors. Therefore, if three nonwords from the fluent pool contained the vowel /ɑ/, three nonwords from the

disfluent pool would contain the same vowel. Further, consonant onsets were balanced across nonword pools to avoid potential phonological similarity confounds. Given that two stop consonants are available at each place of articulation, the *posterior* velar onsets /g/ and /k/ were divided evenly between nonwords from the fluent pool, and the *medial* coronal onsets /t/ and /d/ between nonwords from the disfluent pool. These procedures yielded two matched pools, each containing fourteen CVC nonwords that were used to populate the experimental sequences. As in Experiment 1, sequences were constructed on a trial-by-trial basis where nonwords are selected from the appropriate pool randomly without replacement.

The four experimental conditions were blocked into four sets of thirty-two trials, which were presented in a counterbalanced order for each participant as part of a repeated-measures design. Before each block commenced, participants were given on-screen instructions as to whether the following sequences should be read aloud or silently. They were instructed to read sequences as quickly as possible without making any errors. Moreover, sequences were to be read soundlessly and without lip movement in the inner speech conditions. To check for compliance with these instructions, participants were monitored using a microphone for the duration of the experiment; they were also monitored visually by the experimenter during practice trials.

Each trial began with a central fixation cross, displayed for 750ms, followed by the simultaneous presentation of a nine-item sequence. Participants began reading the sequence immediately upon its appearance and indicated completion of the task by pressing the spacebar. The time elapsed between these two points was measured, and participants were

then asked to report whether they made any errors while reading the sequence by pressing the Y (yes) or N (no) key; trials with reported errors were immediately repeated. Before commencing the experiment, participants completed four practice trials (one from each condition) under the supervision of the experimenter. They were instructed to use these practice trials to find a rapid yet error-free speech rate.

Results and Discussion

Figure 3 depicts sequence reading times as a function of coarticulatory fluency (fluent vs. disfluent) and speech type (inner speech vs. overt speech). A 2 x 2 repeated-measures ANOVA shows that reading times were significantly faster for fluent sequences than for disfluent sequences - $F(1, 31) = 5.53$, $p = .025$ ($\eta_p^2 = .15$). Reading times were also significantly faster for inner speech conditions than for overt speech conditions - $F(1, 31) = 29.90$, $p < .001$ ($\eta_p^2 = .49$). There was no significant interaction between fluency and speech type - $F(1, 31) = 0.02$, $p = .887$ ($\eta_p^2 = .00$).
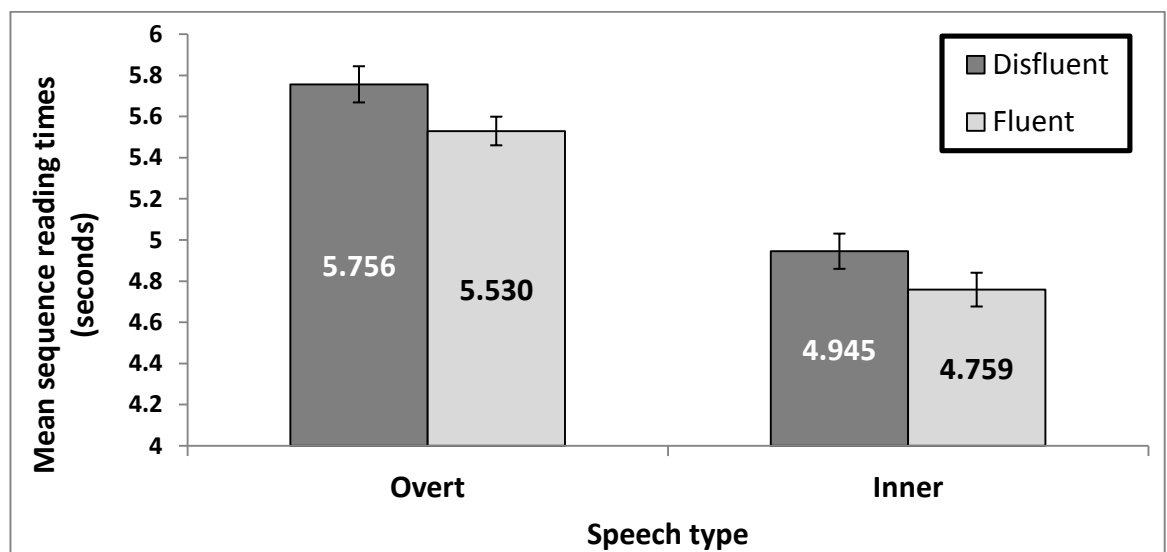

Figure 3. Mean production times (in seconds) as a function of speech type and sequence type. Error bars show Standard Error.

As might be expected, the coarticulatory fluency constraint applies to overtly articulated speech: Overt speech reading times were faster for fluent than for disfluent sequences. Critically, the coarticulatory fluency constraint also applies to inner speech. That is, reading times were also faster for fluent sequences in inner speech, despite the absence of any overt involvement of the articulators. Moreover, the magnitude of the fluency effect in inner speech is comparable to that of the fluency effect observed in overt speech. This is demonstrated by a significant main effect of fluency in the absence of any interaction between fluency and speech type (a negligible effect size of $\eta_p^2 = 0.00$ was observed): Irrespective of speech type, reading times are similarly constrained by coarticulatory fluency. Finally, reading times were (on average) 16.27% faster in inner speech than in overt speech. This corroborates previous reports of a 15-25% advantage (Coltheart, 1999), which likely indicates a fixed temporal cost of overt execution. In summary, the results of this experiment suggest that inner speech embodies sequence-level articulatory detail in the same way as overt speech.

The results of Experiment 3 are not entirely consistent with findings from recent work that employs an error report paradigm (e.g., Oppenheim & Dell, 2008, 2010; Corley et al., 2011). The results are most consistent with the findings of Corley et al. (2011), who also offer evidence for an articulatory effect in inner speech (in the form of a phonemic similarity effect). However, the phonemic similarity effect identified in Corley et al. is smaller in inner than in overt speech. By contrast, the present results indicate coarticulatory fluency effects of comparable magnitude in inner and overt speech.

One explanation for the inconsistency between the present results and recent findings from error report paradigms is that subjective error

reports provide an unreliable account of inner speech errors. Indeed, it is difficult to compare the present results to findings from previous speech error investigation due to lack of consensus in the latter. The application of the error report paradigm has variously produced evidence for nonsignificant positive, nonsignificant reverse, and significant positive similarity effects in inner speech (e.g., Oppenheim & Dell, 2010).

However, there are other explanations for this inconsistency. One is that inner speech need not be conceptualised as a fixed phenomenon that is necessarily phonological or articulatory in nature. Rather, the activation of articulatory detail in inner speech can take intermediate values. According to this flexible abstraction hypothesis (Oppenheim & Dell, 2010), certain conditions can elicit stronger activations of articulatory detail in inner speech. One such condition is silent articulation: Although phonemic similarity effects are not reported in inner speech, they are reported in silently articulated (i.e., mouthed) speech. If similar conditions are present in Experiment 3, its results may be reconciled with previous findings from speech error investigation. For example, participants were instructed to read inner speech sequences silently and without lip movement, and were visually monitored during practice trials to check for compliance. However, this does not preclude the possibility that participants began to engage in silent articulation at some later point during the experiment, resulting in the activation of articulatory detail.

Moreover, the flexible abstraction hypothesis is not explicit in specifying the conditions under which articulatory detail will be activated in inner speech. Conceivably, other conditions besides silent articulation might elicit a similar effect. One such condition could be the absence of lexical support. In Dell's (1986) model of word production, verbal materials are represented at multiple, hierarchical levels. At the top of the hierarchy,

words are represented at a lexical-semantic level. These words contain phonemes (a middle phonemic level) which correspond to specific articulatory features (a bottom feature level). If verbal materials do not correspond to existing lexical representations or are only weakly lexicalised, they must be represented exclusively at lower, more detailed (i.e., articulatory feature) levels. In this way, the absence of lexical support can result in compensatory activation of articulatory features. This compensatory effect might be compared to hyperarticulation - the exaggerated form of articulation that is often observed in infant or foreigner-directed speech (e.g., Uther, Knoll & Burnham, 2007). In the case of frequently-occurring words with strong lexical representations, the opposite (i.e., an attenuation of articulatory detail) can be observed. For example, the frequently-occurring word 'every' often reduces from its prescribed three-syllable form 'ev-e-ry' to the two-syllable 'ev-ry', via the omission of its central vowel (e.g., Hooper, 1976). That is, /e/ is so predictable within the context of the lexeme 'every' that it becomes redundant and need not be specified (e.g., Jurafsky, Bell, Gregory & Raymond, 2001).

For two reasons, Experiment 3 may elicit compensatory activation of articulatory detail in inner speech. Firstly, whereas recent speech error investigation (e.g., Oppenheim & Dell, 2008) has employed sequences of common English words, Experiment 3 employs nonword materials. These may elicit stronger activation of articulatory detail due to the absence of lexical support. Moreover, Experiment 3 employs nine-word sequences, whereas recent speech error investigation has employed four-word sequences. In the past, it has been observed that some articulatory effects only emerge in long sequences (e.g., Woodward et al., 2008). For example, coarticulatory effects that are absent in four-item sequences (e.g., Cowan et al., 1988) manifest in longer nine-item sequences that involve more

numerous inter-item coarticulations (e.g., Murray & Jones, 2002). The use of longer sequences magnifies the importance of inter-item coarticulations, and like nonwords, inter-item coarticulations are categorically non-lexical. Therefore, placing a greater emphasis on coarticulation may also magnify the importance of articulatory detail.

The results of Experiment 3 also leave us with the question of why inner speech embodies articulatory detail, given that it does not entail any direct movement of the articulators. Motor control theory offers one explanation. Articulatory control is supported by a predictive internal model: When a motor command is issued to an articulator such as the tongue, an efference copy of the command is fed into a predictive model. In parallel with the implementation of the motor command by the tongue, its outcome is estimated by the predictive model. This estimated outcome is then compared with the action's actual outcome, as determined by proprioception and other sensory feedback. Any discrepancy between estimated and actual outcomes produces an error signal which is fed back into the control system. The feedback signal is used to correct subsequent motor commands and to recalibrate the predictive model (e.g., Grush 2004).

As a concrete example, suppose a heavy weight is attached to the tongue. In response to the intention to articulate a /t/ gesture, a motor command is issued. Due to the attached weight, the tongue undershoots its intended, estimated position at the alveolar ridge and an error signal is generated. This signal is fed back into the control system with two consequences. Firstly, the next motor command will be corrected with additional, compensatory force. Secondly, the expectations of the predictive model are recalibrated: In response to the original motor command, the tongue would now be estimated to undershoot the alveolar ridge. In a

similar way, existing predictive models will have evolved to embody stable anatomical constraints such as the mass of the tongue. This allows them to produce realistic estimates of the outcomes of motor commands. These in turn are necessary to generate useful corrective feedback that is used to support normal articulatory control.

Although typically employed as part of a control system, predictive models can also be run autonomously to emulate the actions the system usually controls (Grush, 2004). In the case of articulatory control, predictive models can be co-opted to produce mental imagery that incorporates real anatomical constraints on the articulators. Inner speech has been described as a form of mental imagery (e.g., Dell & Oppenheim, 2008), and the results of Experiment 3 suggest that inner speech embodies detail that originates in the anatomy of the articulators. Therefore, one possibility is that inner speech embodies articulatory detail because it is an emulation of overt articulatory behaviour generated by a predictive model (e.g., Tian & Poeppel, 2010; Scott, 2013).

General Discussion

Previous evidence for sequence-level articulatory effects in vSTM (e.g., Murray & Jones, 2002) has been undermined by the confounding influence of item-level properties such as lexical frequency and PND. Variations in these properties are difficult to fully control for and provide an inlet for reinterpretations of sequence-level coarticulatory effects in terms of item-level phonological processes such as redintegration (e.g., Hulme et al., 1997; Roodenrys et al., 2002). However, the experiments presented here show that vSTM performance is constrained by inter-item coarticulatory fluency in a manner than cannot be explained by item-oriented or memory-specific mechanisms.

Previous research manipulates coarticulatory fluency via the inclusion or exclusion of changes in place of articulation at word boundaries (e.g., Murray & Jones, 2002; Woodward et al., 2008). Experiment 1 showed that coarticulatory fluency can be manipulated via an alternative anatomical constraint that relates to the *direction* of a change in place between stop consonants (i.e., whether a change in place of articulation involves a change to a more anterior or posterior speech constriction – e.g., Byrd, 1996). Specifically, performance in a serial recall task with spoken output was better for nonword sequences involving backward-moving coarticulatory transitions.

Experiment 2a exploited the directional nature of the coarticulatory constraint tested in Experiment 1 to devise a novel manipulation of coarticulatory fluency that can be implemented simply by reversing the order of a given sequence of items. Consequently, this manipulation can be utilised to influence coarticulatory fluency while controlling for variations in item-level properties (both anticipated and unanticipated) entirely. Superior serial recall for fluent nonword sequences (i.e., sequences with backward-moving coarticulatory transitions) persisted despite this stringent item-level control.

Controlling for variations in item-level properties may not eliminate the possibility for redintegrative contributions to vSTM performance entirely. Some evidence suggests that redintegration can operate based on the strength of inter-item associations (e.g., Stuart & Hulme, 2000). Nevertheless, recent work shows that coarticulatory fluency effects persist when these inter-item associations are controlled for (e.g., Woodward, 2006; Woodward et al., 2008). Specifically, coarticulatory fluency benefits (i.e., reductions in sequence duration) resulting from familiarisation with a given sequence of items will generalise to new sets of items, so long as

these share the same inter-item coarticulations. In any case, the use of nonword materials in the experiments presented here will restrict any effects that depend on pre-existing associations between items.

Experiment 2b replicated the result of Experiment 2a using an order reconstruction task. Because this alternative short-term memory task does not require participants to overtly articulate their responses, the observed fluency effect cannot be explained in terms of peripheral productive effects (such as the misarticulation of responses during serial recall). Rather, coarticulatory fluency must directly constrain whatever process supports performance in vSTM tasks.

Experiment 3 investigated whether the coarticulatory fluency effect identified in Experiments 1 and 2 extends beyond the context of vSTM. One possibility is that the fluency effect observed in memory tasks is a manifestation of articulatory detail embodied in inner speech, which is co-opted to support rehearsal processes used in vSTM tasks. If so, fluency effects should also be apparent in a task that recruits inner speech but does not place any demands on memory. Experiment 3 demonstrates that nonword sequences with fluent inter-item coarticulations are read faster in inner speech than sequences with disfluent coarticulations. This implies that coarticulatory fluency effects in vSTM cannot be explained by memory-specific mechanisms, given that the same effects extend beyond the context of vSTM tasks. An alternative possibility is that the fluency effects observed in vSTM tasks originate from motor control processes that embody articulatory detail and operate in inner speech.

It is assumed that superior memory for fluent sequences is explained by reductions in coarticulatory complexity rather than articulatory duration (although duration is taken as a generally reliable indicator of

complexity). This assumption is based on previous research that demonstrates poorer memory for complex verbal materials (e.g., materials that involve more syllables, or a more complex format such as CVCVC rather than CVCV) when articulatory duration is controlled for (e.g., Service, 1998). Based on the findings presented in this chapter, it is argued that vSTM function is supported by speech motor control processes that embody articulatory detail. These speech motor control processes enable the retention of an ordered series of verbal items by recoding these sequences into a sequential articulatory form. Sequences with less complex articulatory representations can be more easily recoded, and are therefore better remembered.

Chapter 2

Articulatory difficulty as an explanation for phonological neighbourhood density distributions

Heterogeneous patterns in the distribution of words across phonological space are well-described but poorly understood. A potential explanation for these patterns is the systematic influence of pressures to minimise articulatory effort. An analysis is employed here to test the hypothesis that densely populated phonological regions tend to incorporate more easily articulated speech sounds. An omnibus measure is devised to quantify articulatory difficulty based on three anatomical parameters – articulatory precision, muscular tension, and the efficiency of jaw movements. In a sample of English words, phonological neighbourhood density is found to differ significantly according to articulatory difficulty. By implication, phonological neighbourhood density distributions can ultimately be explained, at least partly, by articulatory pressures. The same can be said of effects that depend on these density distributions, such as the facilitatory effects observed in verbal short-term memory tasks.

Introduction

Phonological neighbourhood density (PND) refers to the number of words that inhabit a given region in phonological space. The density of a given word's phonological neighbourhood can be quantified according to the number of phonological neighbours it possesses (e.g., Roodenrys et al., 2002). These are words that differ from the specified word by a single phoneme. For example, 'tab and 'cub' are both phonological neighbours of 'cab', whereas 'tub' is not. Differences in PND have been shown to influence language comprehension and production (e.g., Garlock, Walley & Metsala, 2001; Munson & Solomon, 2004), and can even constrain performance in verbal short-term memory (vSTM) tasks (e.g., Roodenrys et al., 2002; Allen & Hulme, 2006). For example, memory span for words with a high mean PND (of 28.8) can be as much as nine percent higher than for words with a low mean PND (of 8.8 - e.g., Roodenrys et al., 2002). This advantage is usually explained in terms of redintegration, a phonological process by which decayed short-term memory traces are reconstructed from stable long-term representations (e.g., Hulme et al., 1997). More precisely, degraded short-term traces are matched with the closest corresponding representation in long-term memory; the latter then serves as a basis for output (e.g., Burgess & Hitch, 1992; Page & Norris, 1998). In the case of PND, words from a given phonological neighbourhood form a network linked by mutual excitatory connections (e.g., Roodenrys et al., 2002). When any word from this network is presented (e.g., 'rat'), all of its phonological neighbours (including 'bat' and 'cat') are also activated to some degree, based on shared phonemic features (i.e., 'at'). Via its connections with these neighbours, the presented word receives additional supporting activation. The more phonological neighbours a word has, the more supporting activation it receives, and the more likely it is to be output.

Therefore, high-PND words tend to be better-remembered because they receive superior redintegrative support from associations with numerous phonological neighbours.

Although the effects of PND on vSTM performance have been documented and explained (e.g., Roodenrys et al., 2002; Hulme & Roodenrys, 2009), somewhat less consideration has been given to origin of the PND distributions on which these putative redintegrative memory effects depend. In the absence of systematic shaping forces, words should be evenly distributed across phonological space. In this case, no PND effects would be observed in vSTM performance because there would be no variations in density. Instead, we observe that the distribution of words across phonological space is heterogeneous: Some words occupy very sparse neighbourhoods whereas others belong to dense phonological clusters. By analogy, the distribution of words across phonological space might resemble the distribution of a population across a geographical region. If we were to consult a detailed map, an inspection would reveal a correspondence between population density and fundamental geographical features. For example, population centres tend to cluster around rivers but avoid mountains. Arguably, our knowledge of PND distribution is not unlike a map whose only feature is a representation of neighbourhood density across phonological space – a map with which we can describe the distribution of PND but not explain it. In this spirit, it is suggested that as-yet uncharted features may underlie heterogeneous PND distributions. In particular, these distributions might be partly explained by articulatory pressures. Such an association between articulatory factors and PND would also, therefore, pose questions about the precise mechanisms underpinning the effects associated with variations in density (such as the facilitation of vSTM performance).

Moreover, previous evidence that implicates articulatory effects in constraining vSTM performance (Murray & Jones, 2002) is undermined by the presence of a PND confound. Specifically, sequences that are difficult to articulate also involve low-PND materials that receive poor redintegrative support (Miller, 2010). Yet if PND distributions are shaped by articulatory pressures as suggested above, this confound constitutes a natural part of a larger pattern. As part of this pattern, it would also be reasonable to expect to find articulatory difficulty confounds in previous experiments that manipulate PND to influence vSTM performance (e.g., Roodenrys et al., 2002).

The notion that articulatory pressures contribute to shaping PND distributions is not without precedent. A given region in phonological space encompasses a number of potential speech sounds (i.e., sounds that can be produced by the articulatory apparatus). However, only a portion of these will be realised as speech sounds and incorporated into in the linguistic inventory of any given language. For example, the click consonants found in the Khoisan languages of Africa (such as / Õ/ - e.g., Sands & Güldemann, 2009) are absent from English. For a given language, the population of a region in phonological space (and hence, its phonological density) is determined by the subset of speech sounds that are represented in the linguistic inventory of that language.

What, then, determines which sounds are or are not incorporated into linguistic inventories? The theory of natural phonology (or linguistic naturalness) asserts that the most common speech sounds in the languages of the world are those that are the most easily articulated or the most perceptible (e.g., Hooper, 1976; Lindblom, 1983; Ohala, 1983). More precisely, sound systems may be shaped by pressures to reduce articulatory effort while preserving perceptibility, often resulting in tradeoffs

between the two (e.g., Lindblom & Maddieson, 1988; Lindblom, 1990). Therefore, articulatory difficulty (in conjunction with perceptual factors) may constrain the extent to which particular phonemes and phoneme combinations are represented in linguistic inventories. Indeed, difficult speech sounds such as voiced sibilant affricates (for example, /dz/, which does not appear in English at all) are highly underrepresented in numerous languages (e.g., Zygis, Fuchs & Koening, 2012).

If PND distributions are determined by the contents of linguistic inventories, which in turn are shaped by articulatory pressures, these articulatory pressures should be reflected in PND distributions. Specifically, densely populated phonological neighbourhoods should tend to incorporate easily articulated speech sounds, and sparsely populated phonological regions should tend to incorporate speech sounds that are more difficult to articulate.

The aim of this investigation is to test for an influence of articulatory difficulty on PND. In past literature, PND has been quantified in a straightforward manner: A word's PND corresponds precisely to the number of phonological neighbours it possesses (i.e., those which differ from the given word by a single phoneme - e.g., Roodenrys et al., 2002). However, the same cannot be said for articulatory difficulty. Although articulatory difficulty is an intuitive concept, it has proved difficult to precisely define and quantify (e.g., Westbury & Keating, 1986; Lindblom, 1990; Ann, 2005). Importantly, the influence of articulatory difficulty on PND distributions cannot be tested if articulatory difficulty cannot be effectively characterised and quantified. Therefore, what follows is a consideration of different ways in which the articulatory difficulty of a particular class of speech sounds – consonants – can be characterised and quantified.

The most direct way to characterise articulatory difficulty is in terms of specific anatomical parameters that relate to the behaviour of the articulators themselves. However, the precision of this approach can also be a caveat. Articulation is a complex behaviour that involves numerous anatomical parameters, many of which have the potential to interact with each other. Although chosen parameters can be measured directly, how do we determine which parameters (or combinations thereof) are the most relevant?

An alternative approach is to infer articulatory difficulty from usage patterns in 'low effort' speech contexts. The premise here is that easily articulated speech sounds will be overrepresented in such contexts, whereas difficult sounds will be underrepresented. These usage patterns will reflect the combined influence of relevant articulatory (and perceptual) factors and any interactions between them. However, the resulting definitions of articulatory difficulty will not incorporate genuine explanations for why particular consonants are easier or more difficult to articulate than others. Consequently, they may become circular (e.g., Ann, 1993). For example, easily articulated consonants could be defined as those that are acquired the earliest. But which consonants are acquired earliest - those that are the most easily articulated. The argument that some consonants are favoured over others because they are easier to articulate is only given true meaning if we can specify why this is the case (e.g., Westbury & Keating, 1986).

An optimal compromise between these direct and indirect approaches is to employ low-effort usage patterns to support and validate definitions of articulatory difficulty that are grounded in explicit anatomical premises. Below, some prevalent patterns in low-effort speech contexts are

discussed. These patterns are then related to a number of anatomical parameters that index articulatory difficulty more directly.

What speech contexts are likely to involve low articulatory effort? Arguably, children whose articulatory abilities are still in development will acquire low-effort consonants before difficult consonants. Moreover, they may continue to favour these low-effort consonants even after they have acquired more difficult consonants. Indeed, similar kinds of consonants, such as stops (for example, /p/ and /t/) and nasals (for example, /m/ and /n/), tend to be acquired earliest and used the most frequently in child speech (e.g., Robb & Bleile, 1994). Other types of consonants, such as fricatives (such as /f/ and /s/), tend to be acquired later (e.g., Templin, 1953). Similarly, anterior consonants (i.e., labial consonants such as /b/, /p/ and /f/) are generally acquired before more posterior consonants (i.e., alveolar such as /t/ and /s/, or velar consonants such as /g/ and /k/; Edwards & Shriberg, 1983; Stoel-Gammon, 1988). Similar patterns in terms of age of acquisition are displayed by American, Japanese and deaf children (e.g., Locke, 1980), implying a common underlying cause, such as a shared articulatory anatomy.

Even when children have acquired a range of consonants, they display preferences for some over others, as evidenced by patterns of phoneme substitution in their speech. For example, they tend to replace word-initial fricatives such as /s/ with stops such as /t/ (e.g., 'sad' becomes 'tad'), voiceless consonants such as /t/ with voiced consonants such as /d/ (e.g., 'tail' becomes 'dail'), approximants such as /r/ with glides such as /w/ (e.g., 'dragon' becomes 'dwagon'), and posterior consonants such as /g/ with anterior consonants such as /d/ (e.g., 'girl' becomes 'dirl'; Oller, Wieman, Doyle & Ross, 1975). Similarities also exist between the substitution patterns found in child speech and the speech of apraxic

patients (e.g., Klich, Ireland & Weidner, 1971; Grigos & Kolenda, 2010). Finally, many of the same substitution patterns appear when low-effort speech conditions such as intoxication (e.g., Kaplan, 2010) are induced experimentally (e.g., Lester & Skousen, 1974; Johnson, Pisoni & Bernacki, 1990).

Adults might also be expected to favour more easily articulated consonants within contexts that involve fast or informal speech. Some of these preferences may be captured in characterisations of the effort-minimising process of consonant lenition (e.g., Bauer, 1988; Bybee & Scheibman, 1999). This process typically entails one of two types of sound changes that follow given trajectories (e.g., Lass, 1984). One type is sonorization, where consonants become more vowel-like. For example, the voiceless stop /p/ may sonorize to the voiced stop /b/, which may sonorize further to the continuant /v/. A second type is opening, where consonants become less tense and less resistant to airflow. For example, the stop /p/ may open to the fricative /f/, which may open further to the approximant /h/. Given the characterisation of lenition as a reduction in articulatory effort (e.g., Kirchner, 1988), consonants that are found further along these lenition trajectories should be easier to articulate.

Before moving on to consider direct anatomical measures of articulatory difficulty, another approach warrants a brief review. Perhaps the most straightforward means of determining which consonants are easy or difficult to articulate is to have speakers provide articulatory difficulty ratings for various consonants (e.g., Locke, 1972). These subjective ratings have been shown to correlate with several of the potential indicators for ease of articulation discussed above, including age of acquisition, substitutions in child speech, and frequency of occurrence in conversation (Parnell & Amerman, 1977). Nevertheless, a clear caveat of this approach is that

53

speaker ratings might reflect subjective familiarity with given consonants as much as any difficulty intrinsic to these consonants (e.g., Kirchner, 1988). For example, English speakers would likely rate the click consonants found in the Khoisan languages of Africa (e.g., Sands & Güldemann, 2009) as extremely difficult or impossible to articulate, although these consonants are routinely articulated by native speakers.

Finally, articulatory difficulty can be defined directly according to a number of anatomical parameters. For example, the production of some speech sounds requires a high degree of articulatory precision. That is, the active articulator must be placed and held within a small target area in order to produce the intended speech sound. More precise articulations demand additional articulatory effort in the form of fine motor control (e.g., Kirchner, 1988). Such a difference in articulatory precision is captured in the contrast between fricative and stop consonants: Fricative consonants such as /s/ require more precise articulations than stop consonants such as /t/. To elaborate, stop consonants involve a complete blockage of airflow through the oral cavity. Because full closure is an easy articulatory target that cannot be overshot, stops can be articulated with coarse ballistic movements (e.g., Ladefoged & Maddieson, 1996). However, fricative consonants are characterised by turbulent airflow, the production of which requires that an active articulator (such as the tongue tip) is precisely placed and held within a small target area where it approaches but narrowly undershoots full closure (e.g., Stevens, 1971). In the case of the fricative /s/, not only must effort be invested to elevate the tongue tip (as with the stop /t/): Additional, antagonistic effort must be invested to restrain the tongue tip in order to prevent full closure. Simply put, stop consonants should be easier to articulate than fricatives because they require less precise articulations. This contrast is supported by evidence from age of

acquisition and usage patterns: Stop consonants tend to be acquired earlier than fricatives and are used more frequently (e.g., Templin, 1953; Young, 1981). Children also tend to replace word-initial fricatives with stops (e.g., Oller et al., 1975).

In previous research, the articulatory difficulty of different handshapes in sign language has been quantified according to a combination of three anatomical contrasts (e.g., Ann, 1993; Ann, 1996). As an example of one of these contrasts, it should be easier to articulate handshapes that involve the thumb, index finger or little finger, because these digits have independent extensors (i.e., they can be fully extended when the other digits are contracted). According to a second contrast, handshapes should be easier to articulate if they prescribe the same configuration (i.e., extended or retracted) for the middle, ring and little fingers, because these digits depend on a common muscle. These individual contrasts are combined to form a quantitative omnibus measure of articulatory difficulty for handshapes. A similar approach can be applied here by combining the contrast between fricative and stop consonants with additional phonetic contrasts.

Another anatomical parameter that may constrain articulatory difficulty is muscular tension, as is incorporated into the distinction between fortis and lenis (or tense and lax) consonants (e.g., Hardcastle, 1973; Jaeger, 1983; see also Butcher, 2004). Differences in muscular tension are embodied in the contrast between voiced and voiceless consonants: The glottis must be abducted (i.e., constricted) to occlude airflow for the articulation of voiceless consonants such as /s/ (e.g., Kirchner, 1988), whereas voiced consonants such as /z/ can be articulated with an open glottis and a relative reduction in transglottal tension (Lass & Anderson, 1975). Accordingly, voiced consonants should be easier to articulate than

voiceless consonants. In support of this argument, voiceless consonants (such as /t/) are replaced by their voiced counterparts (such as /d/ - e.g., Lass, 1984) as part of a typical change along a sonorizing lenition trajectory (i.e., a reduction in articulatory effort as consonants become more vowel-like). The same substitution is observed in child speech, where voiceless consonants tend to be replaced with voiced consonants (e.g.,'tail' becomes 'dail' - Oller et al., 1975). It is interesting to note, however, that voiced consonants are normally acquired later than voiceless consonants (e.g., Edwards & Shirberg, 1983).

A final anatomical parameter that can be related to articulatory difficulty is the rotational efficiency of the jaw at different places of articulation. The active articulators (i.e., lower lip, tongue tip and tongue body) are attached to the lower jaw, which can be rotated to assist with articulations by moving the active articulators towards or away from the passive articulators (i.e., the upper lip, alveolar ridge and velum). Because the jaw rotates around a posterior pivot, it can assist with anterior articulations more efficiently than posterior articulations (e.g., Edwards, 1985; Mooshammer, Poole & Geumann, 2007). That is, less rotational movement (and therefore less articulatory effort) is required for the jaw to open or close a 5mm gap between the lips than is required to open or close the same 5mm gap between the tongue tip and the hard palate. Consequently, anterior labial consonants such as /p/ should be easier to articulate than medial, coronal consonants such as /t/.

This distinction in ease of articulation between labial and coronal consonants is also supported by evidence from age of acquisition and substitution patterns. Anterior consonants (i.e., labials) are generally acquired earlier than more posterior consonants (such as alveolar and velar consonants – e.g., Edwards & Shriberg, 1983; Stoel-Gammon, 1988), and

children tend to make substitutions in which they replace alveolar and velar consonants with labials (Oller et al., 1975).

Three phonetic contrasts have now been identified, each of which embodies an anatomical parameter that relates to articulatory difficulty: Fricatives should be more difficult to articulate than stops because they require greater articulatory precision. Voiceless consonants should be more difficult than voiced consonants because they involve additional muscular tension. Finally, coronal consonants should be more difficult to articulate than labial consonants because they receive less efficient assistance from jaw movements. These contrasts can be combined into a single omnibus measure to provide a quantitative measure of articulatory difficulty. That is, consonants can be assigned ordinal values between zero and three to quantify their articulatory difficulty according to the three phonetic contrasts detailed above. For example, the voiceless labial stop /p/ has one difficult articulatory feature according to the omnibus - it is voiceless. By comparison, the voiced alveolar coronal (alveolar coronals are articulated with the tongue tip against the alveolar ridge) fricative /z/ has two difficult features according to the omnibus - it is both coronal and fricative. Therefore, according to the omnibus measure, /z/ has a more difficult articulation than /p/.

The omnibus measure is used here to quantify the articulatory difficulty of the four labial obstruents (i.e., consonants that involve some degree of blocked airflow through the oral cavity) /b/, /p/, /v/ and /f/, and the four alveolar coronal obstruents, /t/, /d/, /s/, and /z/. These consonants represent the possible combinations of the three binary phonetic contrasts that form the omnibus. For example, half are fricatives (/v/, /f/, /s/ and /z/) and half are stops (/b/, /p/, /t/ and /d/). Similarly, half are voiced (/b/, /v/, /d/ and /z/) and half voiceless (/p/, /f/, /t/ and /s/). Finally, half have a labial

place of articulation (/b/, /p/, /f/ and /v/), whereas half have a coronal place of articulation (/t/, /d/, /s/ and /z/).

The purpose of the omnibus measure is to provide a metric for ease of articulation that can be related to PND values. However, articulatory difficulty is measured in terms of the articulatory configurations that are required to produce a single phoneme, whereas PND is measured at a word level. Therefore, the strategy used here is to restrict analysis to simple three-phoneme words. PND values for these words are then paired with the articulatory difficulty scores of their onset consonants. Relative to other positions such as offset, onset is a lexically and phonetically important position. For example, onsets are thought to play an important role in lexical access (e.g., Marslen-Wilson & Zwitserlood, 1989; Gow, Melvold & Manuel, 1996) and tend to resist phonetic weakening and assimilation (e.g., Cho & Keating, 2001). Therefore, the articulatory difficulty of onset consonants should provide a fair approximation of the articulatory difficulty of simple, three-phoneme host words.

Three-phoneme English words beginning with each of the eight consonants specified above were retrieved from the MRC psycholinguistic database (Coltheart, 1981). PND values for these words were then compared based on the articulatory difficulty scores allocated to their consonantal onsets. If ease of articulation contributes to shaping PND distributions, higher-difficulty consonantal onsets should tend to correspond to words from sparsely populated phonological neighbourhoods (i.e., words with low PND values).

The labial and coronal obstruents /b/, /p/, /v/, /f/, /t/, /d/, /s/, and /z/ were included in the analysis. Each consonant was assigned a cumulative articulatory difficulty score between zero and three based on the presence of frication, devoicing or a coronal place of articulation. Therefore, /b/ was assigned a score of zero, /p/, /v/, and /d/ were assigned a score of one, /f/, /t/, and /z/ were assigned a score of two and /s/ was assigned a score of three.

Consonant	Stop	Fricative (+)	Voiced	Voiceless (+)	Labial	Coronal (+)	Difficulty
b	o		o		o		0
d	o		o			+	1
f		+		+	o		2
p	o			+	o		1
s		+		+		+	3
t	o			+		+	2
v		+	o		o		1
z		+	o			+	2

Figure 4. Labial and coronal obstruents included in the analysis with articulatory features and difficulty scores; '+' indicates a 'difficult' feature.

The MRC psycholinguistic database was searched exhaustively for words beginning with each of the specified consonants followed by two phonemes of any identity. 655 words were obtained using this procedure. PND values for all but four of these were obtained using N-watch analysis software (Davis, 2005); the four words without corresponding PND values were excluded from further analysis. N-Watch calculates PND values via an orthographic comparison of a target word to all the other lexical entries

within a 30,605-word vocabulary (this excludes words with very low CELEX frequencies of less than 0.34 per million). Any word within this vocabulary that can be formed by the substitution of a single letter from the target word is counted as a neighbour. PND values for the sampled words were categorised according to the difficulty scores of their onset consonants. This yielded 118 words with an articulatory difficulty score of zero, 213 with one, 190 with two and 130 with three.

Results and Discussion

Figure 5 illustrates a tendency towards lower PND values in words with higher difficulty onsets. A one-way between-subjects ANOVA confirms that PND values differ significantly as a function of articulatory difficulty scores: $F_{(3, 647)} = 9.016$, $p < .001$ ($\eta_p^2 = .040$). A simple linear regression was also used to test the efficacy of the articulatory difficulty omnibus in predicting PND values. Omnibus difficulty scores significantly predicted PND values ($\beta = -.172$, ($t_{(649)} = -4.458$, $p < .001$), and the difficulty scores explained a small but significant proportion of variance in PND values: $R^2 = .030$; $F_{(1, 649)} = 19.87$, $p < .001$.
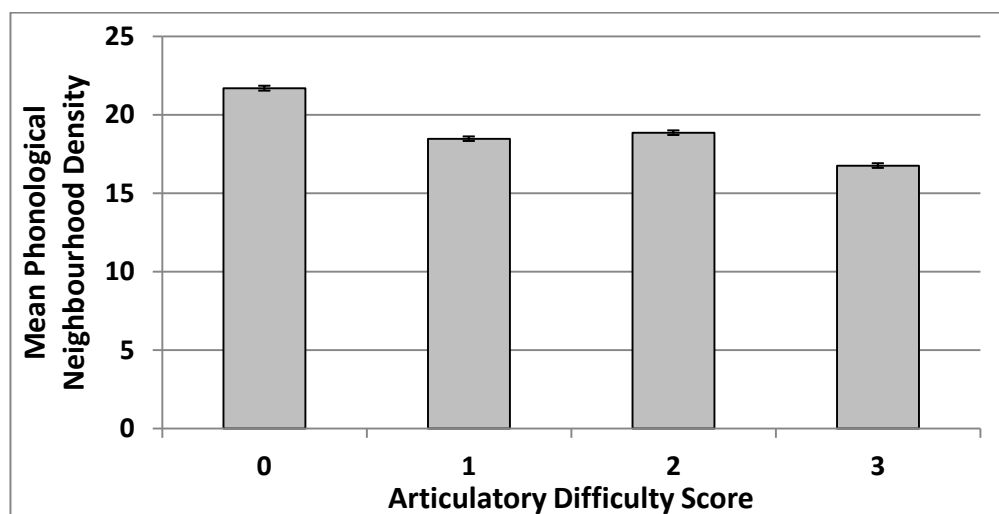

Figure 5. Mean phonological neighbourhood density as a function of articulatory difficulty scores for onset consonants. Error bars show standard error.

Figure 6 shows a breakdown of mean PND values according to each of the three phonetic contrasts that constitute the omnibus measure of articulatory difficulty. In a more detailed analysis, the influence of each of these contrasts on PND values was assessed independently in a one-way between subjects ANOVA. PND values were significantly lower for words with difficult fricative onsets (mean = 16.50; SD = 7.79) rather than stop consonant onsets (mean = 20.33; SD = 7.29): $F (1, 649) = 40.44$, $p < .001$ ($\eta_p^2 = .059$). PND values were not significantly lower for words with difficult voiceless onsets (mean = 18.71; SD = 7.66) rather than voiced onsets (mean = 19.04; SD = 7.82): $F (1, 649) = .279$, $p = .598$ ($\eta_p^2 = .000$). Finally, PND values were significantly lower for words with difficult alveolar onsets (mean = 18.09; SD = 7.08) rather than labial onsets (mean = 19.49; SD = 8.19): $F (1, 649) = 5.374$, $p = .021$ ($\eta_p^2 = .008$).
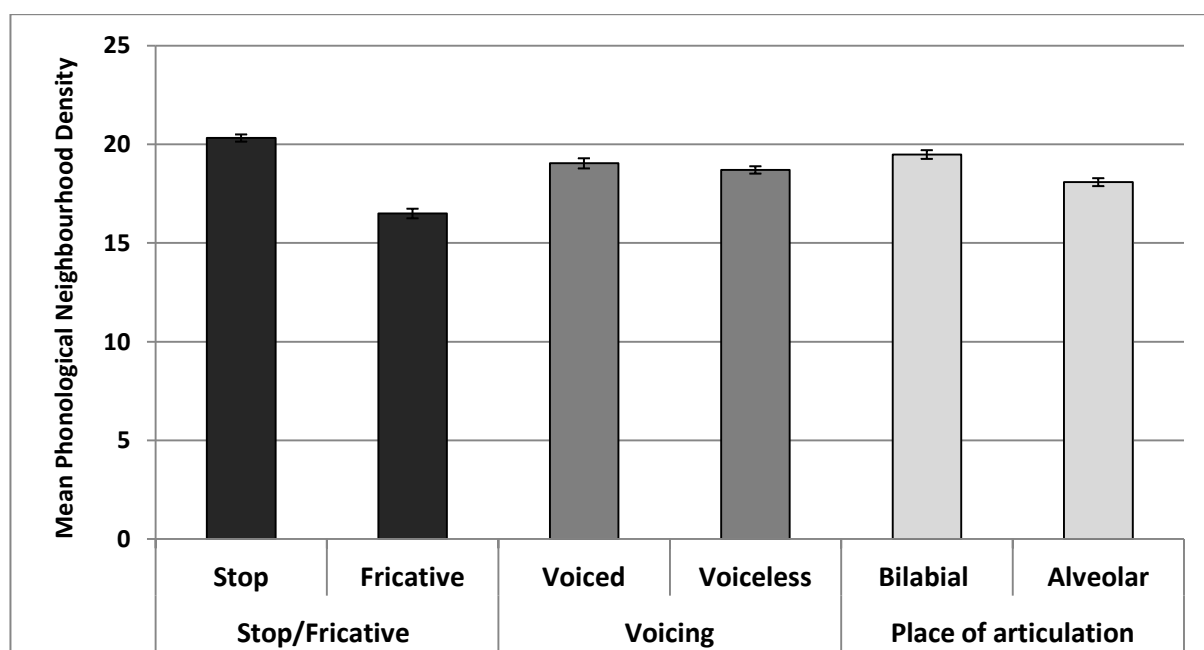

Figure 6. Mean phonological neighbourhood density as a function of the phonetic contrasts included in the omnibus measure of articulatory difficulty. For each contrast, the more difficult articulation is presented on the right. Error bars show standard error.

As hypothesised, words with high-difficulty articulatory onsets tended significantly towards lower PND values. This result extends the theory of natural phonology (e.g., Hooper, 1976; Lindblom, 1983; Ohala, 1983) to support the argument that heterogeneous phonological distributions can be explained (at least partly) by articulatory difficulty. A further implication is that PND-related effects, such as the facilitation of vSTM performance, may be driven by articulatory (as well as perceptual) factors, rather than by neighbourhood density per se.

If this is the case, could it be that past manipulations of PND that result in vSTM effects are confounded with differences in articulatory difficulty? One way to investigate this possibility is to compare the articulatory difficulty scores of the high and low-PND materials employed in past vSTM experiments. However, only a handful of previous experiments (Roodenrys et al., 2002; Clarkson, 2013) implement 'pure' manipulations of PND. For example, in one experiment (Goh & Pisoni, 2003), manipulations of PND are (deliberately) confounded with neighbourhood frequency (i.e., the frequency of a word's phonological neighbours) such that low-PND materials have high neighbourhood frequencies and vice versa. In another experiment (Thomson, Richardson & Goswami, 2005), *rime neighbourhood density* is manipulated to influence vSTM. This relates to how many words share a vowel and coda with the target word (e.g., 'hat', 'bat', 'cat'). Because these neighbours differ in onset, it would be inappropriate to relate the omnibus measure of articulatory difficulty (which is based on word onsets) to rime neighbourhood density (i.e., the articulatory ease of /b/ in 'bap' does little to explain the presence of the rime neighbours 'tap' and 'sap').

Two further experiments (Allen & Hulme, 2006 and Jalbert, Neath, Bireta & Suprenant, 2011) were excluded from the analysis because they

re-use materials from Roodenrys et al. (2002; Experiments 1 and 3, respectively), which are already being included in the analysis. A final experiment (Roodenrys & Hinton, 2002) was excluded due to the use of nonwords materials. Given the argument that more easily articulated speech sounds are more likely to be incorporated into linguistic inventories (i.e., attested as real words in real languages), scoring the onset difficulty of nonword onsets (i.e., words that are not attested) would be inappropriate.

Experiment Excluded	Reason for Exclusion
Roodenrys and Hinton (2002; Experiment 2)	Nonword materials used
Goh and Pisoni (2003; Experiment 1)	Manipulation of PND confounded with neighbourhood frequency
Thomson, Richardson and Goswami (2005: Experiment 1)	Rime PND manipulated; does not apply to word onsets
Allen and Hulme (2006; Experiment 2)	Materials from Roodenrys et al. (2002; Experiment 1) re-used; redundant.
Jalbert, Neath, Bireta and Suprenant (2011; Experiment 2)	Materials from Roodenrys et al. (2002; Experiment 3) re-used; redundant.

Figure 7: Experiments excluded from difficulty score analysis

A caveat of this approach is that the omnibus measure of articulatory difficulty developed here only applies to a subset of consonants (i.e., labial and alveolar coronal obstruents). Therefore, it can only be used to quantify the difficulty of a portion of the materials used in those previous experiments that implement pure manipulations of PND (between 33% and 43%). This effectively reduces sample size and the power of any statistical comparisons. Nevertheless, such analysis proves suggestive. For example, in Roodenrys et al. (2002, Experiment 1), low-PND materials (words with 14 neighbours or fewer) have a higher mean difficulty score (mean difficulty = 2.27; SD = 0.79) than materials from a high-PND category (with 18 neighbours or more; mean difficulty = 1.36; SD = 1.21). Despite the small sample size (11 words from each category) and heterogeneous variance between groups (Levene's test indicates unequal variances at $F = 3.56$, $p = .043$), the difference in mean articulatory difficulty is very close to

significance at $t_{(17.2)}$ = -2.094, p = .051, (η_p^2 = .180; degrees of freedom adjusted from 20 to 17.2).

Experiment	Articulatory difficulty of sampled High-PND Materials	Articulatory difficulty of sampled Low-PND Materials
Roodenrys et al. (2002; Experiment 1)	1.36	2.27
Roodenrys et al. (2002; Experiment 3)	1.18	1.3
Clarkson (2013; Experiment 11)	1.36	1.5

Figure 8: Mean articulatory difficulty scores of high and low-PND materials employed in previous experiments that manipulate PND to influence vSTM performance.

The power of this comparison can be improved by employing a larger sample of similar materials together with the same selection criteria used by Roodenrys et al. (2002): Taking the 651-word sample employed in the earlier analysis, a low-PND category (words with 14 neighbours or fewer) and high-PND category (words with 18 neighbours or more) are defined. The words in this low-PND category (mean PND = 10.11; SD = 3.08) have a significantly higher mean articulatory difficulty rating (mean difficulty = 1.76, SD = .98) than words from the high-PND (mean PND = 24.74; SD = 4.45) category (mean difficulty = 1.37, SD = 1.00); $F_{(1, 568)}$ = 20.58, p < .001 (η_p^2 = .035).

These analyses further substantiate the possibility that vSTM effects attributed to manipulations of PND in past experiments can alternatively be explained by differences in articulatory difficulty. This is an interesting possibility in light of the reciprocal PND confound identified (Miller, 2010) in previous work that manipulates articulatory factors to influence vSTM

performance (Murray & Jones, 2002): The confound cuts both ways. If PND distributions are shaped by articulatory pressures, both of these instances may be part of a larger pattern by which PND and articulatory difficulty are naturally confounded.

The validity of the conclusions drawn here rests on the assumption that the omnibus measure used in the analyses provides a valid characterisation of articulatory difficulty. There are good reasons to suppose this is the case. The phonetic contrasts on which the omnibus measure is based are each underpinned by explicit anatomical premises. For example, the articulation of stop consonants demands less articulatory precision than the articulation of fricative consonants (e.g., Stevens, 1971; Ladefoged & Maddieson, 1996). Voiced consonants involve less transglottal muscular tension than voiced consonants (e.g., Lass & Anderson, 1975; Kirchner, 1998 – although the voicing contrast did not significantly influence PND). Finally, the jaw can assist in the articulation of labial consonants more efficiently than it can assist with articulations of alveolar consonants (e.g., Mooshammer et al., 2007).

These anatomical premises are further backed by evidence from age of acquisition and usage patterns. Stop consonants are generally acquired earlier than fricatives and are used more frequently (e.g., Templin, 1953; Young, 1981). Children also tend to replace word-initial fricatives with stops (e.g., Oller et al., 1975). The voicing contrast draws similar support from substitutions in child speech, where voiceless consonants tend to be replaced with voiced consonants (e.g., Oller et al., 1975). In addition, the voicing of a voiceless consonant corresponds to a typical change along an opening lenition trajectory (e.g., Lass, 1984). Support for the voicing contrast is strong but not unanimous: One discrepancy is that voiced consonants are normally acquired later than voiceless consonants (e.g.,

Edwards & Shirberg, 1983). Finally, anterior consonants (i.e., labials) are generally acquired earlier than more posterior consonants (such as alveolar and velar consonants; e.g., Edwards & Shrirberg, 1983; Stoel-Gammon, 1988). Again, children tend to make substitutions in which they replace alveolar and velar consonants with labials (e.g., Oller et al., 1975). Despite the consistency of the ANOVA results with these contrastive indicators of articulatory difficulty, a negative correlation was found between omnibus scores of articulatory difficulty and the subjective ratings of articulatory difficulty provided in Locke (1972): r_s (649) = -.161, $p < .001$.

PND values vary significantly according to two of the three phonetic contrasts that comprise the omnibus measure of articulatory difficulty: Fricative consonants are associated with significantly lower PND values than stops, and coronals are associated with significantly lower values than labials. However, voiceless consonants are not associated with significantly lower PND values than voiceless consonants. A potential explanation for this lack of an effect is that the voicing feature produces weak perceptual contrasts. That is, voiced and voiceless articulations may sometimes be difficult to differentiate perceptually.

To expand on this, discriminations between voiced and voiceless consonants rely on temporal (i.e., continuous) acoustic cues. One such cue is voice onset time (or VOT - e.g., Abramson & Lisker, 1964). This refers to the time elapsed between the release of an airflow-obstructing constriction and the onset of voicing (i.e., when the vocal folds begin to vibrate). For a fully-voiced fricative consonant such as /v/, voicing precedes release. For a part-voiced stop consonant such as /b/, voicing coincides with release. Finally, for a voiceless stop consonant such as /p/, voicing follows release after an intervening period of aspiration (an intense burst of airflow). Although VOT is a continuous acoustic cue, it leads to overwhelmingly

categorical percepts (e.g., Eimas & Corbit, 1973). Nevertheless, a narrow area of uncertainty surrounds the perceptual boundary between voiced and voiceless consonants (i.e., the area that corresponds to intermediate VOT values). By comparison, the acoustic cue for the contrast between stop and fricative consonants is itself categorical: Stops involve an abrupt burst of airflow, whereas fricatives involve sustained turbulence.

Moreover, voiced and voiceless consonants in word-final position can be discriminated on the basis of an additional acoustic cue – the ratio of a final consonant's duration relative to a preceding vowel (e.g., Denes, 1955; Port & Dalby, 1982). Specifically, the shorter the final consonant relative to the preceding vowel, the more likely it is to be perceived as voiced rather than voiceless. Because the analyses presented here relate exclusively to the properties of onset consonants, they will not account for any contributions of this additional cue to voicing discriminations.

One possibility here is that voicing acts as a supporting phonetic feature whose perceptual value is defined by its interactions with other features. These interactions can be missed when only single-feature contrasts are considered. In support of this argument, some of the most striking patterns in terms of consonantal prevalence and usage involve combinations of voicing with other features. For example, whereas voiced fricatives (e.g., /z/) are underrepresented in the linguistic inventories of many languages, voiceless fricatives (e.g., /s/) are relatively common (e.g., Ohala, 1983). This difference can be explained by an interaction between aerodynamics and perception: Because the perceptual impact of both voicing and frication relies on a steady stream of airflow, the articulation of voiced fricatives requires that subglottal pressure is split between the glottis and a fricative oral constriction (e.g., Stevens, 1971). Consequently, voiced

fricatives are produced with weaker frication than voiceless fricatives, rendering them less perceptually distinctive.

In summary, it may be that voicing contrasts fail to influence PND distributions because they have limited intrinsic value in terms of perceptual distinctiveness. However, when the voicing contrast is combined with other features – notably frication – it becomes more influential. This is because changes in voicing also affect the quality of frication due to the reliance of both features on the same stream of airflow. The importance of this perceptual interaction between voicing and frication is reflected in PND distributions. As reported earlier, voicing does not have a significant influence on PND values when a large sample of 651 consonants is analysed. However, when this analysis is restricted to the 255 words with fricative onsets (i.e., /s/, /f/, /z/ and /v/), those with voiced fricative onsets (i.e., /z/ and /v/) are associated with significantly lower PND values (mean PND = 11.74; SD = 5.94) than words with voiceless fricative onsets (i.e., /s/ and /f/; mean PND = 17.24; SD = 7.79): t (52.23) = -4.8, $p < .001$ (η_p^2 = .058; Levene's test indicates unequal variances at F = 7.790, p = .006, therefore degrees of freedom were adjusted from 253 to 52.2).

The above also serves as a reminder that linguistic inventories are shaped by a combination of articulatory and perceptual pressures. Moreover, these pressures can interact in different ways. Generally, the most common sounds in a linguistic inventory will be determined by trade-offs between articulatory ease and perceptibility (e.g., Lindblom, 1990). The importance of other phonetic contrasts will be determined by predominantly articulatory or perceptual pressures (as may be the case with voiced and voiceless fricatives). Other contrasts again may involve an alignment of articulatory and perceptual pressures, rather than competition between them. For example, not only are stops easier to articulate than fricatives

68

due to a lower demand for articulatory precision, they also feature clear cues as to place of articulation (such as aspiration – an abrupt burst of airflow) which can make them more perceptible (e.g., Wright, 2004). Because this analysis only investigates the influence of articulatory difficulty on PND distributions, it may underestimate or fail to account for distribution patterns that are motivated by perceptual pressures. Accordingly, future analyses might benefit from combining measures that incorporate articulatory difficulty as well as perceptibility.

Nevertheless, it would be useful to optimise measures of articulatory difficulty before combining them with measures of perceptibility. One way to do this would be to incorporate additional anatomical parameters into the omnibus measure of articulatory difficulty. For example, an alternative characterisation of articulatory difficulty is offered by articulatory phonology, which describes speech sounds in terms of articulatory events unfolding across 'tract variables' that correspond to the configuration of different parts of the vocal tract (for example, lip aperture or glottal aperture - e.g., Browman & Goldstein, 1992). Some speech sounds involve more articulatory events (i.e., changes to different tract variables) than others. For example, relative to simple stops (such as /p/), ejective stops (such as /p'/, which is found in Zulu – e.g., Ladefoged, 1971) involve additional glottal events. These are arguably more difficult to articulate, not only because they necessitate quantitatively more action within the vocal tract, but because they will require the increasingly difficult coordination of events across different parts of the vocal tract (e.g., Willerman, 1991; Ann, 2005).

Although the validity of the articulatory difficulty omnibus is supported by a range of evidence, it is important to bear in mind other limitations of the analysis. For example, the results are qualified by the use of a potentially unrepresentative sample of three-phoneme words. The

development of more comprehensive word-level measures of articulatory difficulty would facilitate investigation appreciably. However, such measures would need to account for numerous factors including vowel difficulty, positional interactions, and coarticulatory effects. Given that articulatory difficulty remains difficult to characterise even at a segmental level (e.g., Ann, 2005), this constitutes a significant challenge for future research.

Similarly, the analysis is restricted to English words. Given the universality of articulatory anatomy, articulatory pressures might be expected to hold across larger cross-language samples (e.g., Shariatmadari, 2006). However, linguistic inventories are shaped by both articulatory and perceptual pressures (e.g., Lindblom & Maddieson, 1988). Even if articulatory pressures are predetermined by a universal articulatory physiology, perceptual pressures are demonstrably malleable. Language-specific perceptual contrasts can develop in response to early linguistic experience and persist throughout adult life. For example, Japanese adults have difficulty distinguishing /r/ and /l/ (e.g., Goto, 1971), although this is an easy perceptual contrast for English listeners (and indeed, 4-month old Japanese infants). At the same time, Hindi listeners are able to make additional perceptual contrasts relative to English adults – for example, they are able to reliably discriminate between the retroflex /Da/ and dental /da/ (e.g., Werker & Lalonde, 1988). Consequently, a given phonological region might support more perceptually contrastive speech sounds (and denser phonological clusters) for a Hindi speaker than for an English speaker. Effectively, these differences mean that phonological space cannot be treated as a fixed quantity. This complicates the search for universal associations between articulatory pressures and PND distributions: English and Hindi may be subject to similar articulatory pressures. However,

relationships between articulatory difficulty and PND might differ across these languages because PND distributions in Hindi are determined by a different linguistic inventory that is shaped by different perceptual pressures.

As discussed earlier, the finding that dense phonological neighbourhoods tend to incorporate easily articulated consonants can be explained by the theory of natural phonology (e.g., Hooper, 1976; Lindblom, 1983; Ohala, 1983). Some phonological regions encompass sounds that are particularly amenable to production by the articulatory apparatus. A disproportionately large number of the possible sounds within these regions will be realised as speech sounds and incorporated into linguistic inventories. As a result, dense clusters of these easily articulated speech sounds will form in phonological space. However, this does not account for how, of the possible sounds that exist within phonological space, those that are more easily articulated are incorporated into linguistic inventories as speech sounds to begin with.

One mechanistic explanation for this pattern is offered by the language change process of diachronic lenition: Over time and with repeated use, word forms can become permanently phonetically weaker (e.g., Bybee, 2002). This is illustrated by historical changes. For example, the Latin 'mittere' becomes the Spanish 'meter' (Bauer, 2008), and the Old English 'swerd' becomes the Middle English 'zuord' (Honeybone, 2012). As a more recent example (although not yet reflected orthographically), the three-syllable word 'ev-e-ry' is often reduced in speech to the two-syllable 'ev-ry'. Lenition can be characterised as a persistent and systematic bias towards effort minimization (e.g., Pierrehumbert, 2001). Ultimately, this bias could lead linguistic inventories to converge on easily articulated sounds,

thereby explaining the formation of dense clusters of speech sounds in phonological space.

If this is the case, how can we account for the incorporation of sounds that are difficult to articulate into linguistic inventories? When a phonological region becomes heavily saturated with speech sounds, the perceptual contrastiveness of these speech sounds will begin to suffer. Eventually, this region will become sufficiently saturated that the articulatory benefits of any additional speech sounds is outweighed by prohibitive losses in terms of perceptual contrastiveness. When this threshold is reached, speech sounds will spill out into more sparsely populated regions in order to preserve perceptual contrastiveness, even if these regions offer speech sounds that are sub-optimal in terms of articulatory difficulty. Therefore, the existence of speech sounds with difficult articulations is explained by a trade-off between articulatory and perceptual factors (e.g., Lindblom, 1990): Speech sounds that involve high articulatory costs are adopted because the alternative is an even greater cost to perceptual contrastiveness.

Such costs to perceptual contrastiveness are language-dependent, and can be related to the concept of *functional load*. This refers to the number of language-specific contrasts in which a given phoneme is involved (e.g., Hockett, 1967; Shariatmadari, 2006). For example, together with /p/, the consonant /b/ forms an opposition that contrasts numerous pairs of words such as 'pig' and 'big' or 'peer' and 'beer'. If /b/ is lost, these words become homophones. Therefore, speech sounds with a high functional load may prove especially resistant to reductions in articulatory difficulty via phonetic change because they play a disproportionately important role in maintaining perceptual contrasts. As an example, the difficult consonant cluster /kt/ is routinely assimilated into the easier /tt/ in

languages such as Italian (e.g., from the Latin 'doctor' to the Italian 'dottore'). However, the same cluster tends to actively emerge in Moroccan Arabic (e.g., from the Classic Arabic 'kataba' to the Moroccan Arabic 'ktib' – Ploch, 2003). This might be explained by the importance of the triconsonantal root in Moroccan Arabic. The triconsonantal root is a collection of three consonants that denotes a particular meaning across an entire class of words (including both nouns and verbs). For example, the root /k-t-b/ is used to form words that relate to writing, and often results in the formation of /kt/ clusters, as in 'yaktubu' (he writes) or 'maktûb' (letter). Therefore, although /kt/ may be more difficult to articulate than /tt/, the /k/ within /kt/ bears a high functional. Should the /kt/ cluster assimilate to /tt/, the reduction in contrastiveness that would accompany the loss of the /k/ would also lead to the loss of numerous words belonging to the class formed by the /k-t-b/ root. Therefore, in this particular case the cost to contrastiveness likely outweighs any savings from reductions in articulatory difficulty.

In conclusion, the analysis presented here provides evidence for an influence of consonantal articulatory difficulty on PND. Specifically, dense phonological neighbourhoods tend to incorporate easily articulated consonants, whereas sparse neighbourhoods tend to incorporate consonants that are more difficult to articulate. This suggests that heterogeneous PND distributions can be explained at least partly by differences in articulatory difficulty. Ostensibly, these differences contribute to systematically shaping the linguistic inventories that in turn determine PND distributions.

A further implication relates to the facilitatory effects of PND on vSTM performance, as are usually explained in terms of a phonological redintegration process (e.g., Hulme et al., 2002). On the basis of recent

evidence, it has been argued that articulatory factors (notably inter-item coarticulatory difficulty) play a previously overlooked role in constraining verbal short-term memory performance (e.g., Murray & Jones, 2002; Woodward et al., 2008). The results of this analysis imply the existence of another articulatory constraint on memory that has been similarly overlooked: The very differences in PND on which the redintegration process acts may ultimately be constrained by articulatory factors.

This is not to say that redintegrative effects should be discounted: In practice, vSTM performance is likely influenced by both articulatory and redintegrative effects. However, discounting articulatory explanations for vSTM effects in favour of redintegrative explanations would be equally implausible for several reasons. Firstly, manipulations of PND in previous vSTM experiments (e.g., Roodenrys et al., 2002) appear to suffer from articulatory confounds in much the same way that manipulations of articulatory factors are confounded with differences in PND (Murray & Jones, 2002; Miller, 2010). Secondly, articulatory effects persist when redintegration is controlled. Thirdly, when we consider the origin of the heterogeneous PND distributions on which related vSTM effects depend, it transpires that existing redintegrative effects may be partly underpinned by hidden articulatory pressures in any case.

Chapter 3

Lenition as an explanation for frequency effects in vSTM

Frequent words are better remembered in short-term memory tasks, as is typically explained by a redintegration process whereby partially-decayed phonological traces are reconstructed at retrieval based on matches with corresponding representations in long-term memory. An alternative possibility investigated here is that frequent words are better remembered because they are more affected by lenition, a language change process that reduces the articulatory complexity of affected words. Both previous work and Chapter 1 show that reductions in articulatory complexity lead to improvements in short-term memory. Lenition depends on the frequent recurrence of words *within reducing contexts*. Therefore, to isolate lenition effects from redintegrative frequency effects, this chapter aims to experimentally induce lenition in nonwords via contextual manipulations, while controlling for frequency. The repeated articulation of nonwords within carrier phrases did not result in more lenition (i.e., reductions in articulatory duration) than isolated production. Neither did nonwords lenit more following exposure in which they were semantically associated with familiar objects (e.g., a banana). However, more lenition was measured in nonwords designed to have a high phonetic potential for reduction – specifically, nonwords designed with strong (as opposed to weak) consonant gestures in phonetically-weakening (intervocalic and word-final) positions. Ultimately, the lenition induced via this manipulation did not translate into memory improvements in an order reconstruction task. Nevertheless, this investigation demonstrates the principle that lenition effects dependent on contextual frequency can be experimentally disentangled from redintegration effects that depend on simple frequency.

Introduction

The more often a word is used or encountered in speech, the higher its lexical frequency. For investigative purposes, lexical frequency is calculated as a statistic based on how often given words occur in large corpuses, typically taken from written texts. Two popular measures of lexical frequency calculated in this manner are the Kucera-Francis and CELEX measures of (e.g., Kucera & Francis, 1967; Baayen, Piepenbrock & Van Rijn, 1993).

High-frequency words tend to be better-remembered in verbal short-term memory tasks than low-frequency words (e.g., e.g., Hulme et al., 1997; et al., 2002). This facilitative effect of frequency on vSTM performance is usually explained by appeal to redintegration, a phonological process by which decayed short-term memory traces are reconstructed from stable long-term representations (e.g., Hulme et al., 1997). More precisely, degraded short-term traces are matched with the closest corresponding representation in long-term memory, with the latter then serving as a basis for output (e.g., Burgess & Hitch, 1992; Page & Norris, 1998). The redintegration process is argued to be more effective for high-frequency words because they have better-specified and more accessible phonological representations in long-term memory (e.g., Hulme et al., 1997). This argument can be clarified by invoking the concept of entrenchment: Through repetition, words accumulate lexical strength that results in faster and easier recognition, access and retrieval (e.g., Bybee, 2010).

This chapter explores the possibility that an alternative articulatory mechanism - a language change process known as lenition - can explain the facilitative effects of frequency on vSTM. Lenition refers to the tendency

for the articulations of some words to reduce in complexity following extensive exposure, or when they are used in particular contexts. More precisely, lenition is characterised by reductions in the duration and magnitude of articulatory gestures within the affected word (Browman & Goldstein, 1992); in some cases, these gestures are lost entirely. As an example, consider the reduction of the high frequency 'ev-e-ry' from its prescribed three-syllable form to the two-syllable 'ev-ry', via the omission of the central vowel. Because lenition is practically synonymous with phonetic weakening or *reduction* (e.g., Honeybone, 2008), these terms are often used interchangeably. One of the most widely-accepted formal definitions of lenition is that a given gesture (*x*) is more lenited (i.e., weaker) than another (*y*) if *y* passes through an *x* stage on its way to zero (Hyman, 1975). For example, the 'strong' geminated stop consonant /tt/ lenits to zero / / (deletion) through a series of incrementally weaker stages: /tt/>/t/>/ts/>/s/>/h/>/ / (e.g., Honeybone, 2012).

Previous approaches to investigating lenition typically involve static analyses of pre-existing patterns within natural language environments. These analyses broadly focus on diachronic sound changes (i.e., historical changes between two points in time) such as those found between Latin and western romance languages (e.g., Bauer, 2008; Hualde, Simonet & Nadeu, 2011) or the synchronic differences that are often found between dialects (such as the 'flapping' of /t/ and /d/ to /r/ following stressed vowels in Irish and American-English - e.g., Carr & Honeybone, 2007; Marotta, 2008; Honeybone, 2012).

What motivates the suggestion that lenition can explain the facilitative effects of frequency on vSTM performance? Firstly, high-frequency words are particularly susceptible to lenition, tending to lenit at faster rates and to greater extents than low-frequency words (e.g., Hooper,

1976; Bybee, 2010). For example, the central vowel in the relatively high-frequency word 'memory' is often omitted in speech, resulting in a two-syllable form 'mem-ry' rather than the three-syllable 'mem-o-ry'. By comparison, the central vowel in the similar but lower-frequency 'mammary' is preserved. Secondly, recent experimental work demonstrates that experimentally-induced reductions in articulatory complexity facilitate vSTM performance (see Chapter 1; Murray & Jones, 2002; Woodward, Macken & Jones, 2008). Specifically, verbal sequences that involve more fluent coarticulations between items are better-remembered. Because lenition also involves reductions in articulatory complexity, it follows that advantaged vSTM performance for high-frequency words could be explained by lenition-related reductions in articulatory complexity rather than phonologically-oriented redintegration effects. In the case of 'mem-(o)-ry' and 'mamm-a-ry', it is assumed that the former (whose lenited form involves two syllables rather than three) can be more can be more easily manipulated by speech motor control processes deployed to support vSTM function.

Importantly, although frequency has an important influence on lenition, it is not necessarily a causal one (e.g., Deese, 1960). Specifically, a high frequency in itself is not sufficient to cause lenition unless it is combined with a suitable *reducing context* (i.e., a context that permits or fosters lenition – e.g., Bybee, 2010; Raymond & Brown, 2012). This contextual dependency will prove critical in demonstrating that lenition, rather than redintegration, can explain superior vSTM performance for frequent words. In principle, it should be possible to influence vSTM performance via contextual manipulations of lenition while frequency is held at a constant value. This will exclude the possibility for frequency-related contributions from redintegration, which benefits from frequency directly in

that representations of higher frequency words are better-specified and more accessible in long-term memory (e.g., Hulme et al., 1997). In order to understand and exploit lenition's contextual dependencies, it is necessary to consider how various stylistic, semantic and phonetic factors contribute to establishing a reducing context. These factors are discussed in more detail below.

Stylistic factors that shape reducing contexts

Stylistic factors such as dialect, register and speech rate can contribute to the reducing contexts that foster lenition. Often, variations in these stylistic factors will be dictated by social constraints such as the relationship of the speaker to the listener or the formality of the social context. Speech rate is of particular interest here as a stylistic factor that is particularly amenable to experimental manipulation. Lenition is more likely to occur in contexts that involve rapid speech (e.g., Donegan & Stampe, 1972; Lindblom, 1990; Kohler, 1991; Byrd, 1994), which necessitates reductions in the duration of articulatory gestures. One way to achieve these reductions is by executing full speech gestures at an accelerated rate. In practice, however, rapid speech is more commonly achieved via articulatory undershoot - a phenomenon by which speech gestures are not fully executed. Instead, the active articulators (such as the tongue tip or the lower lip) fall short of their intended targets during speech production (such as the alveolar ridge or upper lip - e.g., Lindblom, 1990). For example, when the word 'don't' is articulated in rapid speech, the tongue tip may fail to reach the alveolar ridge to produce the burst of airflow that usually characterises the final /t/.

Semantic factors that shape reducing contexts

Semantic factors also play an important part in shaping the reducing contexts that foster lenition. For example, words tend to lenit more when they are spoken within the context of a discourse topic to which they are semantically related (e.g., Gregory Raymond, Bell, Fosler-Lussier & Jurafsky, 1999). More specifically, the probabilistic reduction hypothesis states that words tend to be more subject to lenition if they can be easily predicted from the encompassing context (e.g., Bell, Brenier, Gregory, Girand, & Jurafsky, 2009). For example, the articulatory duration of the word 'grand' is shorter when it appears in the context of the familiar and predictable pairing 'grand canyon' than when it occurs in less predictable pairings such as 'grand river' (e.g., Gregory et al., 1999).

The importance of contextual predictability is further illustrated by the phonemic restoration effect: When a phoneme or cluster or phonemes in speech is masked by a non-speech sound such as a cough, listeners perceive the identity of the masked phoneme according to the context of its host word. For example, in the sentences 'the *eel was on the axle' and 'the *eel was on the orange', listeners hear 'wheel' and 'peel', respectively (e.g., Warren, 1970). If listeners can recover entirely masked phonemes from predictable contexts, speakers can afford to heavily lenit their speech in these contexts despite any perceptual costs: So long as a speaker knows that a listener can easily access a word within a given context (i.e., the listener can predict the word), the speaker has license to produce the word with reduced articulatory effort and detail (e.g., Bybee, 2010).

Some strings of words form phrases that convey a particular meaning (e.g., 'I do not know'). These phrases, sometimes referred to as 'lexical bundles,' (e.g., Arnon & Snider, 2010) are subject to frequency effects that cannot be accounted for by the frequency properties of their word-level constituents. For example, participants are faster to recognise

higher-frequency phrases (e.g., 'all over the place') as possible English sequences than lower-frequency phrases (e.g., 'all over the city'). This suggests that such phrases can develop holistic, chunk-like holistic representations in memory (e.g., Bybee & Scheibman, 1999; Arnon & Snider, 2010; Tremblay, Derwing, Libben, & Westbury, 2011). As phrases recur more frequently, they become more chunk-like. That is, their word-level constituents become increasingly redundant and their production increasingly automatized (e.g., Bybee, 2010; Kapatsinski, 2010). This leads to word-level lenition effects (through the loss of redundant phonetic detail; e.g., Bybee, 2010) as well as increased coarticulation between items (e.g., Woodward, 2008). These lenition effects are sometimes quite drastic, as in the case of the frequently-occurring phrase 'I do not know', which is often produced 'dunno'. Critically, a word need not have a high frequency to benefit from lenition effects, so long as is produced as part of a chunk-like (i.e., frequently-occurring) phrase. Moreover, it is not necessary for phrases to recur with extremely high frequencies in order to take on chunk-like properties (e.g., Bybee, 2010).

Phonetic factors that shape reducing contexts

　　Phonetic factors also contribute to reducing contexts at a sublexical level. For example, lenition in consonant gestures such as /t/ is subject to positional constraints. Consonants are less likely to lenit when they occur in a strong phonetic position – for example, at word onset (as in 'tell'). However, they are more likely to lenit when they occur in a phonetically weak position – for example, between two vowels (as in 'iota') or in a word-final position (as in 'don't'; e.g., Lass & Anderson, 1975; Segeral & Scheer, 2008). Interactions can also occur between adjacent positions, depending on the identities of the speech gestures that occupy those positions. For example, a word-final /t/ gesture tends to be preserved before a following

vowel (e.g., 'don't argue') but deleted before a consonant (e.g., 'don' go'; Bybee, 2002).

Online versus offline lenition.

If a reducing context is sufficient to elicit lenition effects, what is the importance of a word's frequency in a reducing context? A reducing context alone will elicit a situational lenition effect (i.e., an effect that persists so long as a word is used in a reducing context), henceforth referred to as *online lenition*. Online lenition can be differentiated from a more permanent and transferable effect that develops when a word is repeatedly encountered in a reducing context (i.e., when a reducing context is combined with frequency). In this latter case, words become more likely to lenit even when they occur outside of the originally reducing context (e.g., Bybee, 2002). This effect, henceforth referred to as *offline lenition*, is thought to be explained by incremental changes to a word's long-term linguistic representation as a result of its repeated usage within reducing contexts (e.g., Pierrehumbert, 2001; this mechanism is explained in greater detail following Experiment 4). It is illustrated by historical sound changes such as the transformation of the Latin 'mittere' into the Spanish 'meter' (Bauer, 2008).

The ultimate aim of this chapter is to demonstrate that superior memory performance for high-frequency words can be explained by lenition, as opposed to redintegrative frequency effects. However, the context of a vSTM task is not necessarily a reducing one - for example, words are unlikely to appear as part of a familiar sequence. Therefore, the experiments presented here are chiefly concerned with inducing offline lenition effects that will transfer from reductive training contexts to vSTM tasks.

Experimental approach

This investigation adopts an exploratory approach whose aim is to experimentally induce lenition in nonword materials via a contextual manipulation. A successful manipulation can later be used in the context of a vSTM task to determine if lenition effects, independent of variations in frequency (and associated redintegrative effects), translate into memory improvements. The advantages of this experimental approach are that it allows strict control over the frequency with which participants are exposed to the materials, the context of this exposure, and the properties of the materials themselves. Nevertheless, this experimental approach diverges from previous methods that investigate lenition in a natural language environment. Removing this language change process from its natural language environment will marginalise factors that might otherwise contribute to its development - notably, semantic factors and social/communicative factors that relate to speaker-listener interactions. Therefore, although the ultimate aim of the investigation is to utilise a frequency-matched manipulation of offline lenition to influence vSTM performance, the greater challenge lies in establishing a protocol and a manipulation that can be used to experimentally induce offline lenition in an artificial environment.

Experiments 4 and 5 combine some of the stylistic and semantic factors discussed earlier into a simple contextual manipulation that should encourage or discourage lenition. Specifically, speakers articulate nonwords either in isolation, or embedded within carrier phrases. Nonwords that recur within the same carrier phrase will be subject to several reducing pressures: They will become predictable within the context of their carrier phrase, could benefit from chunking effects, and should be articulated quite rapidly (partly due to the potential for coarticulation) - previous work shows

that words embedded in verbal sequences are articulated more rapidly than isolated words (e.g., Murray & Jones, 2002). This difference in speech rate can be exaggerated further by explicitly instructing speakers to articulate phrases rapidly and isolated nonwords slowly. By comparison, nonwords articulated in isolation cannot benefit from contextual factors related to predictability and chunking because they do not belong to a phrasal context. Similarly, they will be unable to benefit from coarticulatory effects that increase speech rate.

By definition, nonwords lack lexical and semantic properties, although they are specified in the same phonetic detail as real words. As part of a factorial design, Experiment 6 employs two manipulations that target nonwords' existing phonetic properties on the one hand, and their absent lexical-semantic properties on the other. The first manipulation focuses on nonwords' phonetic potential for lenition. To maximise this potential, nonwords are designed such that strong consonant gestures (i.e., effortful consonant gestures that occur early in lenition trajectories, such as 'tt') occupy phonetically-weakening positions (i.e., intervocalic or word-final positions within which they are predisposed to lenit) and weak consonant gestures occupy strong phonetic positions (i.e., onset position, within which they are unlikely to lenit). Conversely, to minimise this potential, nonwords are designed based on the reverse configuration – weak consonant gestures in phonetically-weakening positions and strong consonant gestures in strong positions. The second manipulation is intended to endow nonwords with semantic properties under the premise that they will benefit more effectively from semantic factors that contribute to the development of lenition effects. Throughout the experiment, targeted nonwords are cued via pictorial representations of a familiar fruit item (e.g., a banana). In order to articulate these nonwords, participants must consult a translation sheet

where the pictorial cues are matched to orthographic representations for corresponding nonwords.

The premise here is that repeated exposure to nonwords within reducing contexts should result in the induction of lenition effects that will persist outside of these contexts. Critically, these lenition effects should persist when nonwords are recombined into new sequences that must be remembered for a vSTM task. Experiment 7 tests whether lenition effects induced in Experiment 6 translate into performance improvements in an order reconstruction task (i.e., the same task used in Experiment 2b, where a verbal sequence is presented serially then re-presented simultaneously in a scrambled order; participants must select items in this scrambled sequence in their original order of presentation). Although this method does not involve a direct manipulation of frequency, it achieves a functionally equivalent effect by manipulating a factor (i.e., whether or not a context is reducing) that determines whether lenition can benefit from a given exposure frequency. In this way, the result will bear on the hypothesis that lenition, rather than redintegration, explains frequency effects in vSTM.

Experiment 4

Experiment 4 investigated whether offline lenition can be selectively induced in nonwords by manipulating exposure context while holding exposure frequency constant. In a training phase, participants repeatedly articulated nonwords in one of two conditions. In an *isolated production* condition, participants carefully articulated isolated nonwords (e.g., 'atta'). In a *phrasal production* condition, participants rapidly articulated carrier phrases with embedded nonwords (e.g., 'try atta next'). Baseline and post-training measurements of nonword articulatory durations were compared to

calculate how much these nonwords lenited as a consequence of training in isolated or phrasal contexts.

The context in which nonwords were produced (i.e., embedded within a carrier phrase or in isolation) was also manipulated during the baseline and post-training test phases in which measurements of articulatory duration were taken. This was done under the premise that offline lenition effects can be latent in nature – that is, they may only emerge in reducing contexts similar to those within which they developed (e.g., Kohler, 1991; Kirchner, 2001). Consequently, the measurement of offline lenition effects may require that these effects are not only contextually induced, but contextually elicited. In this case, offline lenition effects may manifest preferentially within phrasal contexts.

A related issue is that nonwords recur in the same carrier phrases throughout training, which should subject them to reducing online pressures related to predictability and chunking. However, these reducing pressures will be lost when nonwords occur in different phrases during test stages, even if other aspects of the phrasal context (e.g., sequential production) remain. In order to quantify the importance of these pressures, materials trained in phrasal contexts are tested under two subconditions: In a *different phrase* subcondition, nonwords appear in mismatched carrier phrases at test and training (for example, 'try atta next' during training, but 'all atta long' at test). In a *same phrase* subcondition, nonwords appear in the same carrier phrases at test and training (for example, 'try atta next' in both cases).

Two-syllable vowel-consonant-vowel (VCV) nonwords (e.g. 'atta') were generated for use as experimental materials, subject to a few constraints. The strongest available intervocalic consonants (such as the

geminated consonant /tt/) were employed. 'Strong' consonants are specified according to the working definition of lenition as deletion (e.g., Lavoie, 2001; Kaplan, 2010), by which lenition is considered any step along a phonetic trajectory that ends in deletion (/ /). For example, the strong geminated stop /tt/ lenits to zero / / via incremental reductive stages (/tt/>/t/>/ts/>/s/>/h/>/ / - e.g., Honeybone, 2012). Consonants from the beginning of this trajectory were employed in order to maximise the distance from / /, thereby allowing a high potential for lenition. Combined with the attested importance of intervocalic position as a weakening phonetic context (e.g., Lass & Anderson, 1975), this should ensure that potential lenition effects are realised.

Nonwords were embedded in carrier phrases selected for their schematic nature. Schematic constructions are popularly described as pairings between form and meaning (e.g., Bybee, 2010). They take the form of phrases with open slots that can be occupied by various words, so long as these words belong to an appropriate abstract category. For example, 'what (*x*) is it?' is a schematic construction in which *x* is an open slot. The contents of the slot are not fixed, though in this example they must take the form of a noun such as 'time', or 'date'. The premise behind using schematic constructions as carrier phrases is that they readily integrate novel words, and words that integrate well with their carriers should reduce more effectively. Nonwords have the potential to acquire abstract categories appropriate to whichever carrier they are embedded in (e.g., a nonword used in 'what (*x*) is it?' should acquire an implicit noun-phrase status). This affords a tight integration between nonwords and their carriers that could lead to reductive benefits.

It is expected that nonwords trained in phrasal contexts will be subject to more offline lenition (i.e., they will reduce more in duration

between baseline and a post-training test) than nonwords trained in isolated contexts. It is also expected that offline lenition effects will be elicited more effectively in nonwords *tested* in phrasal contexts relative to nonwords tested in isolated contexts. Finally, nonwords that are trained and tested in the same phrases are expected to lenit more than nonwords in the different phrase subcondition due to additional chunking effects (e.g., Bybee, 2010).

Method

Participants and Design

Twelve Cardiff University undergraduate students participated in return for course credit or a payment of £3. These were all female native English speakers between the ages of eighteen and twenty-five. Each participant completed all experimental conditions in a 2x2 repeated measures design.

Materials and Procedure

Eight vowel-consonant-vowel format nonwords (VCVs - e.g. 'atta') were generated for use in the experiment. These were constructed using the geminated consonants /tt/, /pp/ and /kk/ in intervocalic position, combined with the flanking vowels /a/, /ɪ/, /o/, and /ɛ/. Two groups of nonwords were generated: 'akka', 'ikko', 'oppa' and 'itta', (Group 1), and 'atti', 'eppi, 'otti' and 'appo' (Group 2). Each group of nonwords was allocated to one of the two training conditions (isolated or phrasal production), and the nonword group allocated to each condition was alternated across participants. Nonwords were presented to participants in black font in the centre of a computer monitor. This was done using Matlab software including Psychophysics Toolbox extensions (Brainard, 1997, Kleiner et al., 2007).

Nonwords were presented either individually (*isolated production*) or embedded within a carrier phrase (*phrasal production*), depending on the experimental condition. For phrasal production conditions, nonwords were embedded in one of eight carrier phrases. These consisted of two short English words or phrases flanking a target nonword. As with the nonword materials, carrier phrases were allocated to one of two groups of four: 'Try x next', 'An x deal', 'Say x again', 'One x please' (group 1) and 'All x long', 'How's x going', 'Have a x day', 'It's x time' (group 2).

For phrasal training conditions and the *same phrase* subcondition of the test phases, nonwords and carrier phrases were paired by group. Therefore, for example, group 1 nonwords (e.g., 'akka') were paired with group 1 carrier phrases (e.g., 'try x next'). Conversely, during the *different phrase* subcondition of the test phases, nonwords and carrier phrases were paired across groups in untrained combinations. For example, group 1 nonwords (e.g., 'akka') would be paired with group 2 phrases (e.g., 'all x long').

The experiment consisted of a baseline test followed by a training phase and a post-training test, each separated by one-minute breaks. Before commencing the experiment, participants completed several practice trials in which they articulated the nonword 'ukka' in isolation and embedded within the carrier phrase 'please (x) now' (these materials were not used in the experiment itself). Practice trials were completed under the supervision of the experimenter until it was clear that participants understood the task. Throughout the experiment participants commenced each trial with a keypress. This prompted the brief appearance of a fixation cross, followed by the presentation of a single nonword or phrase.

For the baseline and post-training test phases, participants were instructed via an onscreen message to read presented nonwords and phrases aloud at a natural pace. Participants' productions for these trials were recorded for later measurement. For the medial training phase of the experiment, participants were instructed via an onscreen message to articulate isolated nonwords slowly and carefully, and phrases rapidly but clearly. Each nonword was articulated twenty-four times during training, forming a one-hundred and ninety-two-trial training block.

At test, those nonwords that were trained in isolation (e.g., group 1 nonwords) were produced in two contexts – both in isolation and in untrained phrases. Those nonwords that were trained in phrases (e.g., group 2 nonwords) were produced in three contexts – in isolation, in untrained (i.e., group 1) phrases and in trained (i.e., group 2) phrases. Each of the eight nonwords was produced and recorded three times per condition, culminating in two sixty-trial test blocks. At every stage of the experiment, the presentation order of trials was randomised by condition.

Results and Discussion

Spectrograms of participant recordings were obtained using Praat acoustic analysis software (e.g., Boersema, 2001). From these, nonword durations were measured manually. Measurements from the three productions of each nonword were averaged to provide mean baseline and post-training durations. The mean post-training durations for each nonword were then divided by their baseline durations to yield the proportion of each nonword's final duration relative to its baseline duration (i.e., the amount of *reduction* or offline lenition measured in each condition). For example, a nonword with a baseline duration of 3000ms and a post-training duration of

2700ms has a lenition proportion of 0.9; lower values indicate more lenition. These lenition proportions are displayed in Figure 9.

Because online lenition is a prerequisite for the development of persistent offline lenition effects, the first important question to be addressed is whether online lenition manifests preferentially in phrasal testing contexts. That is, irrespective of any persistent reductions as a consequence of training, were nonwords articulated more rapidly in phrasal contexts in general? An analysis of the mean durations of nonwords (collapsed across conditions and test stages) shows that they were articulated significantly faster in phrasal contexts (mean duration = 2897ms; SD = .020) than in isolated contexts (mean = 3695ms; SD = .051): t (11) = 7.448, $p < .001$ (η_p^2 = .835).
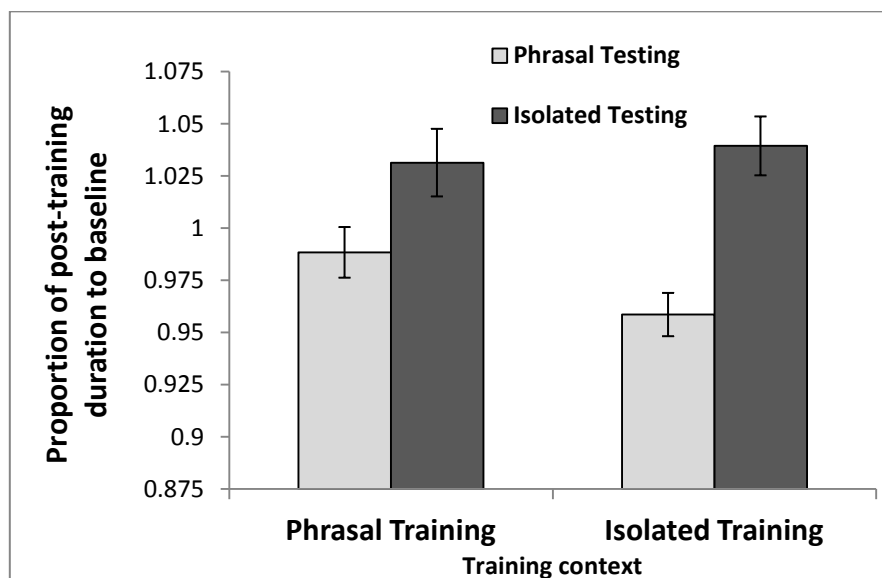

Figure 9. Mean proportions of post-training nonword durations relative to baseline (lower values indicate greater reduction). Error bars display Standard Error.

The next question posed is whether the manipulation of training context selectively induced offline lenition. Specifically, was the amount of offline lenition induced between baseline and post-training tests (i.e., as a consequence of training exposure) higher for nonwords trained in phrasal

contexts? A two-way repeated-measures ANOVA indicates that it was not. Lenition measured in nonwords trained in phrasal contexts (mean = 1.011; SD = .099) was no greater than for nonwords trained in isolation (mean = .999; SD = .085): F (1, 11) = .740, *p* =.408, (η_p^2 = .063).

It is possible that offline lenition effects may only emerge within favourable contexts (i.e., the same kind of contexts that encourage online lenition – e.g., Kirchner, 2001). Therefore, offline lenition might manifest preferentially in phrasal testing contexts. This suggestion is borne out by the results, which indicate more offline lenition for nonwords tested in phrasal contexts (mean = .974; SD = .078) than for nonwords tested in isolated contexts (mean = 1.035; SD = .103): F (1, 11) = 7.755, *p* = .018 (η_p^2 = .413).

This raises the possibility that offline lenition effects might be strongest in a specific set of circumstances where they have been both effectively induced (i.e., nonwords were trained in a phrasal context) and elicited (i.e., nonwords were tested in a phrasal context). In other words, we might expect an interaction between training and testing context. However, analysis does not indicate a significant interaction between testing and training context: F (1, 11) = 1.128, *p* = .311, (η_p^2 = .093).

A related question, in the case that nonwords are both trained and tested in phrasal contexts, is whether more lenition occurs when these contexts match. Specifically, does more lenition occur when nonwords are trained and tested in the very same carrier phrases between training and test, as opposed to different phrases? A paired samples t-test indicates that this was not the case: Nonwords tested in the same phrases as they were trained in did not reduce significantly more (mean reduction to .944; SD =

.096) than nonwords tested in untrained phrases (mean = .988; SD = .084): t (11) = -1.602, p = .137 (η_p^2 = .189).

To review, Experiment 4 employed a contextual manipulation to investigate whether persistent offline lenition can be selectively induced in one of two subsets of frequency-matched nonwords. During a training phase, nonwords were repeatedly articulated in either phrasal or isolated contexts. Production context was also manipulated for the baseline and post-training phases during which nonword durations were measured. This was done in order to investigate whether any induced lenition manifests preferentially in favourable environments.

Contrary to expectations, there was no selective effect of training context on offline lenition (operationalised here as a reduction in articulatory duration between a baseline and post-training test). Nonwords trained in phrasal contexts did not lenit any more between baseline and post-training tests than those nonwords trained in isolation. However, there was a significant effect of testing context: More offline lenition was measured in nonwords produced in phrasal testing contexts. This implies that lenition effects induced during training are latent in nature, but will manifest preferentially in favourable phrasal production contexts (e.g., Kirchner, 2001). By extension, measurements taken in isolated production contexts are likely to underestimate the extent of offline lenition induced during training. Finally, although a small increase in offline lenition was measured when nonwords were produced in pre-trained phrases rather than untrained phrases, the increase was non-significant. It is possible that the reductive effects of chunking were unable to reach their full potential under the conditions of the experiment. Importantly, this finding suggests that offline lenition effects are not limited to the specific phrasal context (i.e., carrie

phrase) within which they were induced, but can transfer to other phrasal contexts.

It is useful to know that induced offline lenition manifests preferentially in favourable (i.e., phrasal) testing contexts. Indeed, unfavourable (i.e., isolated) testing contexts appear not only to suppress lenition but to elicit the reverse effect of fortition (i.e., an increase in articulatory duration between baseline and test). This finding will simplify the task of detecting any offline lenition induced by future manipulations. However, the experiment did not succeed in selectively inducing offline lenition: Nonwords reduced to a similar extent regardless of whether they were subjected to isolated or phrasal training. How might the absence of a selective reduction effect be explained?

One possibility is that the apparent fortition measured in isolated testing contexts was not unique to the isolated production context. Instead, this effect could have been caused by global factors that also apply to the phrasal testing contexts. In particular, repeatedly articulating phonetically similar nonwords is a demanding and repetitive task that will likely result in an accumulation of boredom and fatigue effects. These in turn could result in slower speech rates. Given the format of the experimental paradigm (two measurement phases preceding and following a lengthy exposure phase) and the nature of the dependent variable (the proportion of post-training articulatory durations to baseline durations), these order effects could result in an under-measurement of offline lenition effects. This will reduce the chances of detecting significant lenition effects of any sort.

Might any other factors have limited the amount of lenition induced in the experiment? Offline lenition is conceptualised as a consequence of repeated exposure to reducing online pressures (e.g., Bybee & Scheibman,

1999). Consequently, another possible explanation for the paucity of offline lenition is that there was an insufficient amount of online lenition to consolidate into offline effects. If this were the case, nonwords would not have been articulated any faster in phrasal contexts than in isolated contexts. However, a quick examination reveals that this is not the case. Even in the different phrase subcondition, nonwords were articulated significantly faster in phrasal contexts (mean duration = 2897ms; SD = .020) than in isolated contexts (mean = 3695ms; SD = .051): t (11) = 7.448, p < .001 (η_p^2 = .835). Therefore, a lack of online reduction cannot be responsible for curtailing offline lenition effects.

Another unexpected result was that the production of nonwords in the very same phrases as they were trained in did not elicit significantly more reduction than their production in different carrier phrases. This suggests that participants did not form or deploy chunk-like constructions of the kind that are responsible for substantial lenition effects in natural language contexts (e.g., 'dunno'). A possible explanation for this is that a speaker's license to use lenited chunk-like forms is implicit in a natural language environment - at least in informal communicative contexts. However, speakers are unlikely to infer similar license in a formal and artificial laboratory environment.

An additional barrier to the development of substantial offline lenition is created by presenting nonwords and their carrier phrases in a prescriptive orthographic medium. For example, the utterance 'I do not know' regularly lenits to the chunk-like 'dunno' in informal speech. However, this outcome is clearly implausible when participants are asked to rapidly read the onscreen phrase 'I do not know' aloud.

Lenition effects might also have been inhibited by the nature of the experimental materials. The nonwords used in the experiment were designed to exploit the strong predisposition towards lenition in consonants that occur between two vowels (e.g., Lass & Anderson, 1975; Segeral & Scheer, 2008). However, their potential for reduction is limited in other ways. Words with complex articulations stand to reduce more than words with simpler articulations, given that they have more detail to lose. One important correlate of articulatory complexity is the number of syllables in a word (e.g., Caplan, Rochon & Waters, 1992; Baldo, Wilkins, Ogar, Willock & Dronkers, 2011), of which the experimental materials only have two. Therefore, the use of more complex nonwords with additional syllables and articulatory features might lead to substantially more lenition.

On a similar note, the eight nonwords used in the experiment were constructed from a limited pool of consonants and vowels. Consequently, the materials are both structurally and phonetically similar to each other. This may lead participants to confuse the identities of nonwords assigned to different experimental conditions, effectively undermining any selective effects of the training manipulation. This potential for confusion between the materials also relates to another, deeper issue. The offline lenition measured in the experiment is assumed to be caused by *exemplar shift* (e.g., Pierrehumbert, 2001). This mechanism combines the effects of reducing pressures and frequency to explain the consolidation of online lenition into a persistent offline effect. Owing to its importance in explaining how the experimental effects develop, this exemplar-based mechanism warrants deeper consideration here.

Exemplar models assume that long-term linguistic representations are shaped by numerous tokens of experience. In practice, every token of word usage is mapped onto an exemplar, strengthening it, whereas

disused exemplars weaken and are ultimately lost (e.g., Pierrehumbert, 2003; Wedel, 2007). If a token is too dissimilar to an existing exemplar to map onto it, a new exemplar is created close in phonetic space to similar existing exemplars. Therefore, a word's various phonetic realisations are captured in a cluster of exemplars that contribute to a central linguistic representation.

Speech production involves persistent and systematic biases towards effort minimization (e.g., Lindblom, 1990). When such a bias is introduced into the exemplar model, every time a speech token is produced, its form is very slightly more lenited than that of the exemplar that served as a target for production. Given a sufficient number of iterations (i.e., if the word is used frequently), this will result in the formation of new, more lenited exemplars. Plausibly, the same principle of effort minimization will lead speakers to select lenited exemplars as production targets more often than not. With continued usage, this should lead lenited exemplars to strengthen. Conversely, unlenited exemplars will be reinforced less and less often, leading them to weaken and be lost. In this way, the introduction of a small but systematic production bias will cause a word's exemplar cluster to incrementally shift in favour of a more lenited long-term representation. The role of frequency in this exemplar shift process is that of a catalyst: Because higher-frequency words are subjected to reducing biases in speech production more often, they take on lenited representations more quickly.

Usually, when a speech token cannot be mapped onto an existing exemplar, a new exemplar is added to the cluster. However, if a token cannot be matched to a central representation at all (usually because the token is too ambiguous) it is ignored rather than stored (e.g., Pierrehumbert, 2001). The use of nonword materials that lack pre-existing

lexical representations may create a similar situation. In this case the token is ambiguous because no reasonably corresponding central lexical representation has been established yet. Therefore, it is possible that nonwords will be unable to fully benefit from lenition via exemplar shift until they have lexicalised sufficiently. Similarly, confusion caused by the similarities between the eight nonwords used in the experiment may have hindered the formation of discrete lexical representations. This is likely exacerbated by the absence of typical natural language properties (such as semantic attributes) that could normally be used to help differentiate the materials.

One way to address this problem could be by using pseudoword materials rather than pure nonwords. These pseudowords, although they can technically be classified as nonwords, are based on existing lexical entries (for example, the pseudoword 'cathedruke' is based on 'cathedral') and will likely share some semantic attributes with real words. Previous research shows that pseudoword materials show evidence of lexicalisation (specifically, they engage in lexical competition) in response to moderate phonological exposure (e.g., Dumay & Gaskell, 2007). However, the use of pseudowords will reduce the experimental control afforded by pure nonwords – for example, pseudowords could be subject to partial redintegrative effects based on the frequencies of those real words on which they are based.

Experiment 5

The quantities of offline lenition measured in Experiment 4 were quite small. If these can be improved upon, it may become easier to detect any effects of the experimental manipulations. The pilot experiments presented here test two interventions intended to boost the amount of

offline lenition induced or measured: The first focuses on using nonword materials with more complex articulations, and the second focuses on the use of a delayed test to offset boredom and fatigue effects and allow an incubation period for any offline lenition induced during training.

Experiment 5a

Experiment 5a investigated whether the amount of offline lenition induced in Experiment 4 could be boosted by employing a smaller set of more complex and distinctive nonword materials - specifically, six three-syllable nonwords in place of eight two-syllable nonwords. This was done with the expectation of less confusion between nonwords and a higher potential for reduction. In addition, the number of training trials was reduced from 24 exposures per word to 15 to offset potential fatigue and boredom effects. Finally, additional recordings of nonword productions were taken during baseline and post-training test phases (five instead of three) to provide a more reliable measure of lenition. Having quantified the online lenition benefits associated with nonwords recurring within pre-trained phrases in Experiment 1, all nonwords were tested in different phrases to those in which they were trained. This better simulates the context of a vSTM task, in which nonwords are unlikely to occur within familiar pre-trained phrases.

Method

Participants

Ten naive participants were recruited from the same demographic as in Experiment 4, for the same payment. Experiment 5a utilized the same repeated measures design and experimental format as Experiment 4, with the notable exception that the *same phrase* subcondition (used to measure the reducing consequences of nonwords appearing in pre-trained chunks)

was omitted. Nonwords trained and tested in phrasal contexts were embedded in different carrier phrases between training and test.

Materials and Procedure

Six three-syllable nonwords were generated for use in the experiment: 'takkody', 'mikkoda', 'deppiry' (Group 1), 'tattina', 'mittala' and 'bappolo' (Group 2). Of those nonwords used in Experiment 1, the six that reduced the most were used as base forms to which prefixes and affixes were added to form three-syllable nonwords. For example, the two-syllable 'ikko' from Experiment 1 was prefixed with 'm' and affixed with 'da' to form the three-syllable 'mikkoda'. As in Experiment 1, the group of nonwords used for each training condition was alternated across participants. The single carrier phrase 'try (x) next' was used for phrasal conditions in the training phase and a different carrier phrase, 'say (x) again', was used for phrasal conditions in the baseline and post-training testing stages.

Procedure

The procedure for Experiment 5 deviated from Experiment 4 in two ways: Five recordings were taken for each nonword at each test phase (compared to three in Experiment 1), culminating in two 60-trial test blocks. Additionally, each nonword was articulated a total of fifteen times during the training phase (rather than twenty-four as in Experiment 1), culminating in a 90-trial training block.

Results and Discussion

Experiment 5a investigated whether the amount of offline lenition measured in Experiment 4 could be boosted by utilising fewer and more complex nonword materials in conjunction with a shorter training stage (intended to counteract fatigue and boredom effects). These changes

resulted in a non-significant (4.62%) boost to lenition in phrasal testing

contexts compared to that observed in Experiment 4. A mean reduction

proportion of .931 (SD = .103) was measured in phrasal testing contexts,

compared to .951 (SD = .083) in Experiment 4: F (1, 20) = 1.485, p = .237

(η_p^2 = .069).

This small boost to lenition did not alter the nature of the effects

reported in Experiment 4: A two-way repeated-measures ANOVA indicates

that once again, nonwords tested in phrase reduced significantly more

(mean = .931; SD = .103) than nonwords tested in isolation (mean = 1.039;

SD = .097): (F (1, 9) = 24.618, p = .001 (η_p^2 = .732). Similarly, nonwords

trained in phrase did not reduce any more (at .982; SD = .116) than

nonwords trained in isolation (at .987; SD = .112): F (1, 11) = 0.196, p =

.668 (η_p^2 = .021). Finally, there was no significant interaction between

training and testing context: (F (1, 9) = 1.411, p = .265 (η_p^2 = .136).
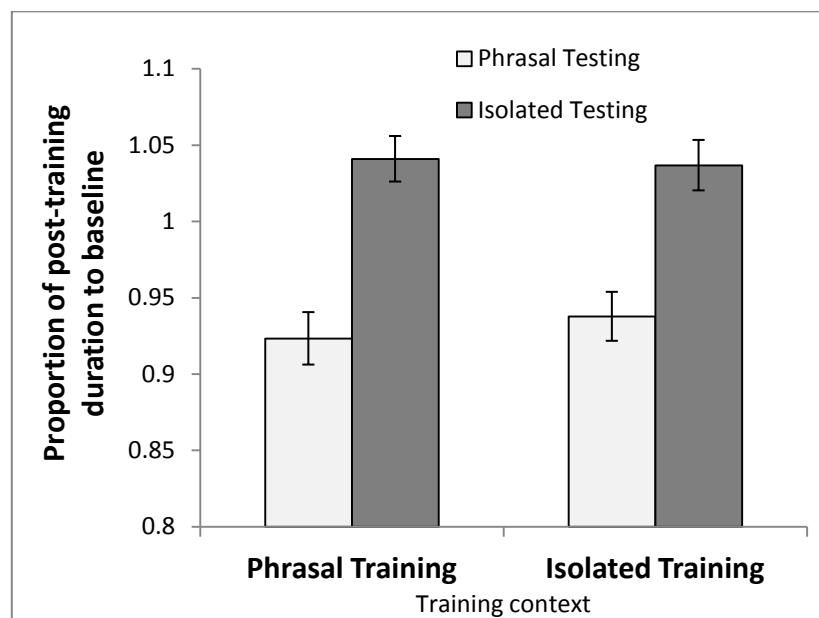

Figure 10. Mean proportions of post-training nonword durations relative to

baseline. Lower values indicate greater lenition. Error bars display Standard Error.

Experiment 5b

Experiment 5b tests a stronger intervention against fatigue and boredom effects that may have lessened the amount of lenition measured in Experiment 4. Such effects are assumed to accumulate throughout the course of the experiment, and may therefore result in slower speech rates during the post-training test than the baseline test. Because the measure of offline lenition used here is based on a comparison between baseline and post-training durations, boredom and fatigue effects that accumulate between the baseline and post-training test could offset any lenition effects that develop during the exposure phase. Experiment 5a was replicated using a small sample of four participants with the addition of a second post-training test (i.e., in addition to the post-training test immediately after the training phase) after a delay period of 42 to 72 hours. A delayed test should allow for the dissipation of any accumulated fatigue or boredom effects. The impact of this approach can be measured by comparing reduction between immediate and delayed post-training tests.

The use of a delay period between the training phase and post-training test may offer additional benefits besides the mitigation of boredom and fatigue effects that could lead to an under-measurement of the offline lenition. Experiment 4 raised the possibility that nonwords fail to benefit from lenition via exemplar shift because they lack lexical representations. Evidence suggests that in order for lexicalization processes to proceed effectively, repeated exposure must be followed by a sufficient consolidation period (e.g., Dumay & Gaskell, 2007). Specifically, a 12-hour interval including nocturnal sleep was required in order for nonwords to begin engaging in lexical competition - an important indicator of their integration into long-term lexical memory (e.g., Jusczyk & Luce, 2002).

Results and Discussion

The use of a delayed post-training test did not result in any significant improvements in reduction proportions relative to an immediate test. Mean reduction in phrasal testing contexts was no greater between baseline and a delayed test (mean reduction = 1.029; SD = .066) than between baseline and an immediate test (mean = .0954; SD = .029), as indicated by a one-way repeated-measures ANOVA: $F(1, 3) = 9.856$, $p > .05$ ($\eta_p^2 = .767$). This suggests that fatigue and boredom effects are not accountable for an under-measurement of induced lenition. Equally, the lexical consolidation benefits associated with a delayed test did not result in increased lenition. However, it is possible that a consolidation period, although necessary for lexicalisation, must be combined with other factors that contribute to this process - such as semantic attributes, for example.

As noted earlier, previous research finds evidence of nonword lexicalisation after as little as thirty exposures (Dumay & Gaskell, 2007), and the twenty-four exposures utilised in Experiment 4 do not fall very short of this number. Nevertheless, the nature of the verbal materials used in this previous research may have led to an underestimation of the requirements (including exposure frequency) for nonword lexicalisation. Specifically, this research utilised pseudo-nonwords that were based closely on real words. For example, the pseudo-nonword 'cathedruke' is derived from 'cathedral'; indeed, the phonological differences between the two are quite superficial. Consequently, unlike the semantically-impoverished nonwords utilised in the present series of experiments, pseudo-nonwords such as 'cathedruke' share semantic (and phonetic) properties with pre-existing lexical entries. Given the importance attributed to the development of meaning (i.e., semantics) in the lexicalisation process (e.g., Brinton & Traugott, 2005), these semantic associations likely facilitate the integration of pseudo-nonwords into the lexicon. Moreover, by influencing the lexicalisation

process, semantic differences stand to constrain lenition effects. According to exemplar theory (Pierrehumbert, 2001), articulatory reductions correspond to changes in lexical representation. In order for verbal materials to fully benefit from the articulatory reductions that accompany representational change, it follows that they must first develop lexical representations.

This suggests a manipulation to test whether semantics plays an important role in enabling lenition via exemplar shift: Select nonwords can be enriched with semantic associations in order to facilitate lexicalisation, and as a consequence, lenition. Notably, the use of a similar strategy in previous experimental work led to the conclusion that semantic associations do not facilitate lexicalisation (Dumay, Gaskell & Feng, 2004). In this previous experiment, participants were familiarised with the nonword materials via one of two tasks. The first involved semantic exposure via a categorisation task, where participants were required to verify whether nonwords belonged to pre-assigned conceptual categories (e.g., 'a cathedruke is a variety of vegetable'). The second involved phonological exposure via a phoneme-monitoring task, where participants were required to indicate whether a specified phoneme was present in a given nonword. Those nonwords familiarised via semantic exposure did not lexicalise significantly more (specifically, they did not engage in greater lexical competition) than nonwords familiarised via simple phonological exposure.

Importantly, this experiment is subject to the same caveat as Dumay and Gaskell (2007): Again, the verbal materials employed were pseudo-nonwords based closely on pre-existing lexical entries. In this case, the use of pseudo-nonwords with strong semantic associations renders the semantic exposure condition somewhat redundant. That is, semantic exposure may have failed to convey any special lexicalisation benefits (i.e.,

beyond simple phonological exposure) because any such benefits were already assured by the nature of the materials. Consequently, the importance of semantics for lexicalization cannot be conclusively ruled out.

Moreover, manipulating semantics could be a more pragmatic strategy for enabling lexicalization-dependent lenition effects than increasing exposure frequency, given that it is unclear what constitutes a sufficient exposure frequency. Available estimates (30 exposures in Dumay & Feng, 2007) from previous work may be unreliable due to the use of pseudo-nonword materials possessing semantic properties that could facilitate lexicalisation.

Experiment 6

Experiments 4 and 5 simulate lenition under controlled conditions that marginalise the contributions of naturalistic pressures that may contribute to this language change process. For example, nonwords are semantically impoverished. Experiment 6 addresses this absence of semantic attributes with a manipulation that enriches select nonwords with semantic associations. Equally, because nonwords lack lexical and semantic properties that might usually contribute to the development of offline lenition effects, phonetic factors could play an exaggerated (or compensatory) role in nonword lenition. Therefore, Experiment 6 also involves a manipulation of nonwords' phonetic potential for reduction.

Semantics could make an important contribution to lenition via several mechanisms. For example, as discussed earlier, exemplar shifts act on lexical representations whose acquisition may be facilitated by semantics. Semantic development is acknowledged as an important part of the lexicalisation process (e.g., Brinton & Traugott, 2005). Moreover, the same experiments that dismiss the importance of semantics in contributing

105

to lexicalisation (e.g., Dumay et al., 2004; Dumay & Gaskell, 2007) involve the rapid lexicalisation of pseudo-nonwords that are endowed with semantic attributes because of their similarity to real words. Because semantics was not isolated effectively in these experiments, its role in lexicalization (and by extension, lenition) remains unclear.

Similarly, lenition can result from automatization – the acquisition of lasting neuromotor production efficiency through repeated practice (e.g., Bybee & Hopper, 2001). However, automatization proceeds only when a speaker knows that a given word is easily accessible within its context (e.g., Bybee, 2002). Semantic factors can play an important part in determining this accessibility. For example, reaction times in a lexical decision task (a measure of lexical accessibility) are improved when a target word is preceded by a semantically-related prime (e.g., when 'butter' is preceded by 'bread'; Schvaneveldt & Meyer, 1973). In addition, words that are semantically related to discourse are articulated more rapidly (e.g., Gregory et al., 1999). Because nonwords lack semantic attributes that could improve their accessibility within particular contexts, they may fail to fully benefit from automatization processes.

As well as facilitating reductive mechanisms, a semantic intervention can be used to address a potential limit on reduction. Sound changes, particularly those relating to unfamiliar words, can be constrained by orthography (e.g., Derwing, 1992). Repeatedly presenting a nonword in written form may serve to anchor that nonword's pronunciation, counteracting the articulatory target drift that is typical of an exemplar shift. However, if a nonword's phonetic form (e.g., 'bappolo') can be enriched with a semantic association (e.g., a 'bappolo' is a banana), it becomes unnecessary to use orthographic prompts. In this example, 'bappolo' can

be replaced with a picture of a banana, thereby reducing any constraining influence of orthography on the nonword's pronunciation.

Experiment 6 employs a manipulation that enriches select nonwords with semantic associations. Participants are be provided with a cartoon picture of either an apple or banana in the centre of the phrase 'one ( ) please', and are required to ask for the object using a corresponding nonword. This nonword is provided on a translation sheet that includes two object-nonword pairings. For example, the apple may be paired with the nonword 'baput'. Therefore, when participants encounter 'one (apple) please', they consult the translation sheet and speak 'one baput please'. Participants are encouraged to produce phrases without consulting the translation sheet as they become more familiar with the object-nonword pairings. The intention here is that participants will transition to producing nonwords from an internal representation rather than an external orthographic one.

The contextual manipulations utilised in Experiment 4 and 5 may have proved ineffective in selectively inducing offline lenition because the use of a controlled laboratory environment (and nonword materials in particular) eliminates many naturalistic pressures towards reduction. The semantic manipulation outlined above addresses the impoverished nature of nonword materials. However, although nonwords are semantically impoverished, they are fully phonetically specified. Phonetic factors are an important determinant of whether lenition does or does not proceed in real words. If anything, the contributions of these phonetic factors towards lenition should be exaggerated in nonwords due to the absence of other contributing factors. Therefore, in addition to manipulating nonwords' semantic attributes, Experiment 6 manipulates nonwords' phonetic

attributes such that they are predisposed to lenit to greater or lesser extents. The basis of this phonetic manipulation is outlined below.

The likelihood of reduction is strongly influenced by an interaction between *gestural* and *positional* strength. Stronger gestures possess a higher potential for reduction, given that they must proceed through more reductive stages than weaker gestures before reaching a common endpoint. For example, to reach the given endpoint /s/, the gesture /d/ proceeds through two reductive stages: /**d**/ > /ts/ > /**s**/. However, the gesture /t/ begins at a stronger starting point and must proceed through three reductive stages: /**t**/ > /d/ > /ts/ > /**s**/. Given the definition of lenition as a reduction in duration and/or magnitude (Browman & Goldstein, 1992), each of these stages should typically be accompanied by measurable reductions in duration. Therefore, given an opportunity to weaken, a strong gesture should reduce in duration more than a weak gesture. Whether or not a gesture realises its potential for lenition depends partly on the phonetic environment that it gesture occupies (i.e., its positional strength). Words contain phonetically 'strong' and 'weak' positions: Word-initial and word-final positions are strong, and intervocalic positions are weak (e.g., Segeral and Scheer, 2008). Gestures that occupy strong positions will resist reduction, whereas gestures that occupy weak positions are highly disposed to reduce.

Through a manipulation of gestural-positional strength, it should be possible to design nonwords that are phonetically predisposed to reduce to lesser or greater extents. Indeed, gestural-positional strength was exploited in Experiments 4 and 5 to maximise nonwords' potential for reduction in response to manipulations of word context. Specifically, strong consonants such as /tt/ were placed in weak intervocalic positions (between two vowels - e.g., 'atta') where their high potential for reduction could be realised.

However, the influence of gestural-positional strength on lenition outcomes was not systematically investigated in its own right. It should also be possible to manipulate gestural-positional strength so as to inhibit reduction rather than facilitate it.

In order to implement this manipulation, Consonant-Vowel-Consonant-Vowel-Consonant (CVCVC)-format nonwords (e.g., 'tadid') were utilised. These contain one strong consonantal position (onset) and two weak positions (the central intervocalic position and the terminal offset position). These positions were filled with strong or weak consonants to satisfy one of two conditions. In the first condition, positional and consonantal strength are *matched* such that strong positions are filled with strong consonants and weak positions with weak consonants (e.g., 'tadid'). Consequently, strong consonants are disinclined to realise their high potential for reduction, whereas the reducing effect of weak positions is counteracted by using consonants with a low potential for reduction. In the second condition, positional and consonantal strength are *mismatched* such that weak positions are filled with strong consonants and strong positions with weak consonants (e.g., 'datit'). Consequently, strong consonants will be inclined to realise their high reduction potential, whereas weak consonants with low reduction potential are placed in positions where reduction was unlikely to occur in any case.

In some circumstances, phonetically strong positions will not just protect occupying consonants from weakening but can cause them to strengthen (the reverse of lenition - a phenomenon known as 'fortition'; e.g., Segeral & Scheer, 2008). However, a further advantage of the CVCVC format is that the ratio of two weak consonantal positions (i.e., intervocalic and offset) to one strong (onset) should overcome any potential strengthening effects of the word-initial position.

109

The net effect of these manipulations is to create nonwords that are phonetically predisposed to reduce to a greater or lesser extent in response to exposure. Given sustained exposure, nonwords with mismatched gestural-positional strength should continue to reduce after nonwords with mismatched gestural-positional strength have exhausted their potential.

Experiment 6 utilises a factorial design to investigate the contribution of semantics and phonetic reduction potential to lenition. nNonwords are trained and tested in identical carrier phrases. Again, a delayed test will be employed in order to allow for the consolidation of learned phonetic forms (e.g., Dumay & Feng, 2007), and to mitigate any fatigue or boredom effects that accumulate during training. Based on the implicit role of semantics in lexicalisation, and of lexicalisation in lenition, it is expected that nonwords enriched with semantic associations will reduce in duration more than those without. Further, it is expected that nonwords with a high phonetic potential for reduction (i.e., nonwords in which strong consonant gestures are placed in phonetically weak positions) will reduce in duration more than their low-potential counterparts.

Method

Participants and Design

Twenty-four naive participants were recruited from the same demographic as in the previous experiments for the same payment. Each participant completed all experimental conditions in a 2x2 repeated measures design.

As a manipulation of phonetic reduction potential, participants were required to articulate nonwords with mismatched gestural-positional strength (i.e., high reduction potential) and nonwords with matched gestural-positional strength (i.e., low reduction potential). As part of a

semantic enrichment manipulation, nonwords were presented either orthographically (semantically impoverished condition), or via a cartoon picture of a corresponding fruit object (semantically enriched condition). On a separate translation sheet, the two cartoon pictures participants could encounter were paired with orthographic representations of corresponding nonwords. As in previous experiments, nonword durations were compared between a baseline and post-training test.

Materials

Nonwords were constructed in a Consonant-Vowel-Consonant-Vowel-Consonant (CVCVC) format. These were formed by combining the vowels 'a', 'e', 'i', and 'u' with strong (voiceless) and weak (voiced) varieties of labial stop consonant (e.g., 'p' and 'b'), alveolar stop consonant (e.g., 't' and 'd') and velar stop consonant (e.g., 'k' and 'g').

For the high reduction potential condition, nonwords were constructed with mismatched gestural-positional strength. That is, a strong consonant (such as 't') was allocated to the weak intervocalic and word-final positions, whereas a weak stop consonant (such as 'd') was allocated to the strong onset position, to produce a nonword such as 'datit'. Conversely, for the low reduction potential condition, nonwords were constructed with matched gestural-positional strength. In this case, a weak consonant (such as 'd') was allocated to the weak intervocalic and word-final positions, whereas a strong stop consonant (such as 't') was allocated to the strong onset position, to produce a nonword such as 'tadid'.

Nonwords with high potential for reduction (i.e., mismatched gestural-positional strength)	Nonwords with low potential for reduction (i.e., matched gestural-positional strength)
datit	*tadid*
betak	*pedag*
baput	*pabud*
gekip	*kegib*

Figure 11. The nonword pairs used in Experiment 6, arranged by high and low phonetic potential for reduction.

Four pairs of nonwords were constructed in this fashion, each pair including a nonword form with mismatched gestural-positional strength (e.g., 'datit') and a counterpart form with matched strength (e.g., 'tadid). Each participant was assigned four nonwords - one from each of the four pairs, including two matched forms and two mismatched. This ensured that no participant ever encountered both nonwords from a pair (e.g., 'tadid' and 'datit'), which could lead to phonemic confusions across conditions. High and low potential nonwords were split between the two semantic conditions such that each of the four nonwords corresponded to a unique factorial condition. The particular nonwords allocated to each condition were rotated across participants.

Throughout the experiment, nonwords or pictorial substitutes were embedded within the carrier phrase 'one (x) please', which was presented orthographically. Participants were provided with a translation sheet on which cartoon pictures of an apple and banana were labelled (orthographically) with corresponding nonwords.

Procedure

An initial experimental session involved a baseline test phase followed by a training phase. As in Experiment 5b, a delayed post-training test phase took place during a second session scheduled between 42 and 72 hours after training. Exposure frequency for each nonword in the training phase was increased to thirty-six (to match the frequency used in Dumay & Gaskell, 2004).

For semantically-impoverished trials, participants were orthographically presented with short carrier phrases including centrally-embedded nonwords. Participants read these phrases aloud. For semantically enriched trials, nonwords were replaced with one of two pictorial substitutes - a cartoon picture of an apple or a banana. Participants consulted a translation sheet to identify the nonword corresponding to the presented picture before reading the entire phrase aloud, including the nonword corresponding to the picture cue. Written instructions issued prior to the experiment encouraged participants to refrain from referring to the translation sheet as they became more confident of the picture-nonword associations.

Results and Discussion

The first question addressed in this experiment was whether manipulating nonwords' phonetic properties so as to maximise their potential for lenition will cause them to lenit more in response to moderate articulatory exposure. As expected, nonwords with a high phonetic potential for reduction (i.e., mismatched gestural-positional strength) reduced significantly more between baseline and a delayed test (at .801; SD = .123) than counterpart nonwords with a low potential for reduction (i.e., with matched gestural-positional strength, at .863; SD = .114), as indicated in a two-way repeated-measures ANOVA: $F_{(1, 23)} = 9.479$, $p = .005$ ($\eta_p^2 =$

.292). This is a valuable result, indicating that the manipulation of phonetic reduction potential fulfils the demand for an effective frequency-independent manipulation of offline lenition.

The second question addressed here was whether nonwords will lenit more if they are endowed with semantic properties - specifically, if they are associated with and cued via pictorial representations of familiar fruit items throughout the experiment. It is suggested that nonwords with semantic properties will lexicalise more effectively and become more susceptible to exemplar shift processes. However, semantically enriched nonwords (i.e., words whose phonetic forms were cued pictorially rather than presented orthographically) did not reduce significantly more (at .848; SD = .110) than semantically impoverished nonwords (at .816; SD = .131): $F (1, 23) = 1.447$, $p > .05$ ($\eta_p^2 = .059$).
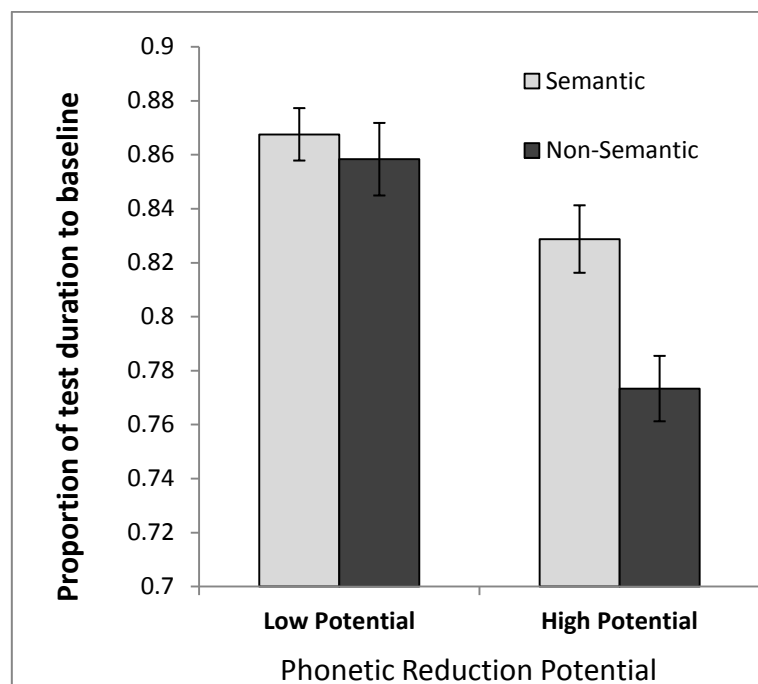

Figure 12: Mean durational reduction according to phonetic reduction potential and semantic properties (lower values indicate greater reduction). Error bars show Standard error.

A final possibility is that increasing nonwords' susceptibility to mechanisms through which offline lenition develops (i.e., exemplar shift) via semantic manipulations will only have a substantial effect when these nonwords already possess a sizeable potential for articulatory reduction. That is, the manipulations of nonwords' semantic and phonetic properties may interact such that nonwords endowed with semantic properties lenit more effectively, but only if they possess a high phonetic potential for lenition. However, no significant interaction was found between the phonetic and semantic factors manipulated in the experiment: $F(1, 23) = .867$, $p > .05$ ($\eta_p^2 = .036$).

How exactly did the phonetic manipulation contribute to superior lenition for high-potential nonwords? An examination of the articulatory duration data (see Figure 13, which displays mean nonword durations at baseline and test according to phonetic reduction potential) indicates that superior reduction for high over low-potential nonwords was accounted for by longer articulatory durations at baseline (mean = .472 seconds; SD = .079, versus .452; SD = .082) and shorter terminal durations (mean = .380 seconds; SD = .081 versus .393; SD = .082). This pattern is reflected in a significant interaction between reduction potential and testing phase in a two-way repeated-measures ANOVA: $F(1, 23) = 12.251$, $p = .002$ ($\eta_p^2 = .348$).
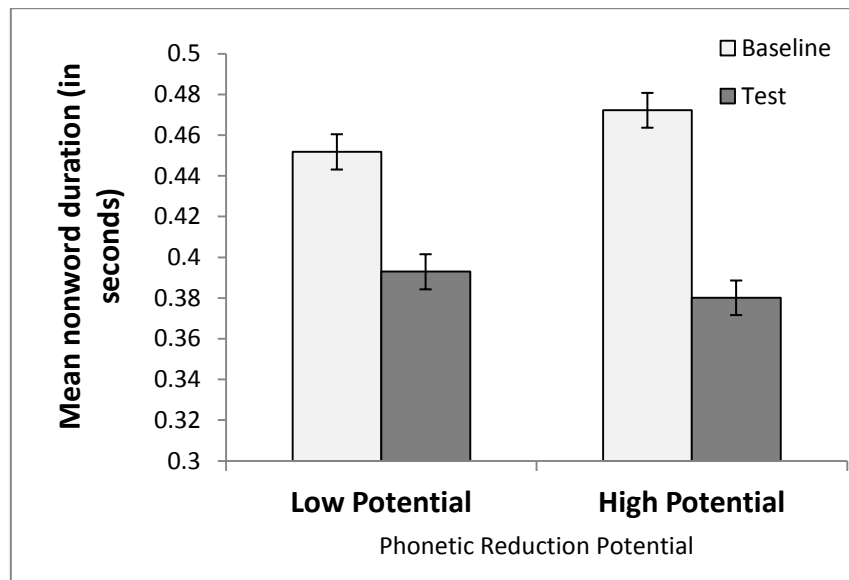
Figure 13: Mean durations of nonwords with high and low phonetic potential at baseline and post-training test. Error bars show Standard error.

Longer baseline durations for high-potential nonwords can be explained in terms of their phonetic composition: Whereas low potential nonwords comprise two weak consonant gestures and one strong, high potential nonwords comprise two strong gestures and one weak. This additional strong gesture will result in longer baseline durations for high-potential nonwords. High potential nonwords also reached lower terminal durations than low-potential nonwords, which can again be explained in phonetic terms. Both of the strong consonants in a high potential nonword occupy phonetically weakening positions (intervocalic and word-final - e.g., Segeral and Scheer, 2008). Therefore, all of the strong gestures in high potential nonwords are susceptible to lenition. By comparison, the one strong gesture in a low potential nonword is protected by its strong onset position. Ultimately, high-potential nonwords stand to lose all of their strong gestures, resulting in shorter terminal durations than for low potential nonwords whose one strong gesture is preserved. Again, shorter terminal durations will contribute to superior reduction for high potential nonwords. These explanations are supported by patterns in the durational data.

Superior reduction for high-potential nonwords was accounted for by a 4.425% higher mean baseline duration and a 3.421% lower mean terminal duration.

The semantic enrichment manipulation did not have a significant impact on nonword reduction. One explanation for this result relates to the experimental procedure. Participants were instructed to reduce their reliance on the provided translation sheets as they become more familiar with the two picture-nonword associations in the experiment. Because familiarity with these associations develops largely through training, participants are likely to produce nonwords from the translation sheet during the baseline phase, but from memory during the test phase. However, although participants may be sufficiently confident to produce nonwords from memory during the test phase, they may not yet be entirely certain of the correct forms. This is important because uncertainty in speech tends to be accompanied by symptoms of disfluency including hesitations, increases in syllable stress and reductions in speech rate (e.g., Starkweather, 2014). These often manifest in unfamiliar second language contexts, where they are actively deployed by speakers as strategies to compensate for communicative uncertainty (e.g., Poulisse, 1990). Such strategies can include explicit markers (e.g., the speaker request 'how do you say x'), repetitions, and again, reductions in speech rate. Crucially, any uncertainty present in the test phase but not the baseline phase may lead participants to selectively reduce their speech rate. This will effectively offset any durational reduction that might otherwise apply.

There is an alternative interpretation for the ineffectiveness of the semantic manipulation that incorporates the subtle yet unexpected tendency towards inferior reduction in semantically-enriched nonwords. Although unexpected, this outcome closely corroborates findings from

previous research on lexicalisation. Specifically, pseudowords subjected to semantic exposure via a semantic categorisation task did not show any more evidence of lexicalisation (as assessed by whether these pseudowords engaged in lexical competition) than pseudowords subjected to simple phonological exposure. If anything, the semantic treatment mildly inhibited normal lexicalisation (Dumay et al., 2004). The authors concluded that simple exposure to phonological form is sufficient to advance lexicalization processes, whereas semantic exposure is not.

It was originally unclear whether these assertions should be taken at face value. Arguably, the use of pseudowords with pre-existing semantic attributes (e.g., 'cathedruke') prevents a clean manipulation of semantics. Nevertheless, a very similar pattern of results was produced in Experiment 6 when pure nonwords were used to assess the contribution of semantic properties to the development of offline lenition. These similarities lend weight to the credibility of Dumay et al.'s findings. Moreover, if there is a link between lenition and lexicalization (i.e., lexicalisation is a prerequisite for lenition via exemplar shift) and the latter is encouraged by phonological but not semantic exposure, increasing the frequency of nonword exposure might be a more effective strategy for developing offline lenition than semantic treatments. Indeed, if offline lenition proceeds via an exemplar shift process that only affects lexical materials, initial exposure (which is required to establish a lexical representation) may not contribute to lenition.

Experiment 7

The aim of the previous experiments presented in this chapter was to identify a manipulation that can be used to induce lenition in nonwords under frequency-matched conditions. As discussed earlier, a manipulation such as this can be used to support the hypothesis that lenition, rather than

frequency-related redintegration effects, explains frequency effects in vSTM. Rather than directly manipulating frequency to influence lenition, this method achieves a functionally equivalent outcome by manipulating factors that determine whether or not lenition can benefit from a given exposure frequency. Experiment 6 identified a manipulation suitable for this purpose, whereby nonwords' phonetic properties are designed such that they are predisposed to reduce more or less in response to frequency-matched exposure. Therefore, the specific aim of Experiment 7 was to test whether a phonetic predisposition towards greater offline lenition in nonwords will translate into vSTM improvements, independently of any direct effects of frequency. To this end, order reconstruction tasks were used in place of the baseline and test-phase measurements of articulatory duration from previous experiments. Order reconstruction involves the serial presentation of a verbal sequence which is subsequently re-presented in a scrambled order. Participants must then select these scrambled items in their original order of presentation. As explained in Chapter 1, performance in this task cannot be explained in terms of misarticulation because participants are not required to overtly articulate their responses.

One issue here is that if the pool of nonwords from Experiment 6 were to be reused and allocated to participants in the same way, only two nonwords per condition could be provided for each participant. This was not sufficient to populate sequences for an order reconstruction task. Therefore, the four pairs of nonwords from Experiment 6 were supplemented with four new pairs of nonwords to double the number of available materials. This was sufficient to allow each participant four nonwords per condition – enough to construct sequences for an order reconstruction task. However, a consequence of increasing the total number of nonwords is an increase in the risk that they will be phonemically

confused. This is dangerous because it could result in phonemically-similar nonwords mapping onto the same lexical representation. In this case, lenition effects intended for nonwords with a high phonetic potential for reduction could generalise to low-potential nonwords, diluting the effect of the experimental manipulation. Due to this risk, care was taken to reduce the phonemic similarity between nonword pairs by minimising the number of phonemes shared between them. As an additional precaution, nonword pairs were split into two phonemically dissimilar groups, each containing four nonword pairs. Each participant was assigned high-potential nonwords from one group and low-potential words from another. This ensured that no participant ever encountered both nonwords from a pair (e.g., 'tadid' and 'datit').

Method

Participants and Design

Fifteen participants were recruited from the same demographic as in the previous experiments for the same payment. Each participant completed both experimental conditions in a repeated measures design. Participants performed an order reconstruction task on four-item sequences of nonwords with high phonetic reduction potential (i.e., mismatched gestural-positional strength) or low reduction potential (i.e., matched gestural-positional strength): Nonwords sequences were presented serially (i.e., one at a time) before reappearing together in a scrambled order. Participants were then required to select these scrambled items in their original order of presentation. As in previous experiments, performance was compared between a baseline and post-training test. The dependent variable, improvement in order reconstruction performance between baseline and post-training test, is expressed as the proportion of post-

training test performance to baseline performance. Higher values therefore indicate greater improvements - for example, 50% correct performance at baseline and 75% correct performance at post-training test will yield an improvement value of 1.5.

Materials

The four nonword pairs from Experiment 6 (tadid, datit; kegib, gekip; pabud, baput; pedag, betak) were supplemented with four additional nonword pairs (kudeg, gutek; togeb, dokep; tibad, dipat; kibug, gipuk). Care was taken to reduce the phonemic similarity between these nonwords by minimising the number of phonemes shared between each.

In order to avoid confusion between nonwords from the same pair (e.g., 'tadid' and 'datit'), nonword pairs were split into two phonemically dissimilar groups, each containing four nonword pairs - Group A (tadid, datit; kegib, gekip; pabud, baput; pedag, betak) and group B (kudeg, gutek; togeb, dokep; tibad, dipat; kibug, gipuk). Each participant was assigned high-potential nonwords from one group and low-potential words from another (e.g., 'datit' from Group A and 'kudeg' from Group B). The groups from which materials were drawn were rotated across participants.

Procedure

Participants completed 20 order reconstruction trials during a baseline phase and a further 20 during a post-training test which took place after a delay of 42-72 hours. For the order reconstruction trials, four nonwords were presented orthographically in the centre of the screen. These were presented serially, with an interstimulus interval of 750ms. After a retention period of 10 seconds, the nonwords reappeared together simultaneously, presented from left to right in a randomly determined order. Participants were required to click the nonwords in the same order that they

were originally presented in. Each nonword could only be selected once and changed colour when selected.

During the training phase participants were required to articulate orthographically-presented nonwords embedded within the carrier phrase 'one (x) please'. Participants articulated each nonword 10 times throughout the training phase. Before commencing the experiment, participants completed sample practice trials from the training and testing phases of the experiment under the supervision of the experimenter.

Results and Discussion

The purpose of Experiment 7 was to test whether superior offline lenition for nonwords with a high phonetic potential for reduction (see Experiment 6) translates into comparable improvements in vSTM performance. In other words, will memory for high-potential nonwords improve more as a result of articulatory exposure than memory for low-potential nonwords? Improvements in order reconstruction performance were calculated by comparing performance between a baseline and post-training test. Contrary to expectations, there was no significant difference in memory improvements between nonwords with a low phonetic potential for reduction (mean = 1.338; SD = .538) and nonwords with a high potential (mean = 1.208; SD = .384), as indicated in a paired-samples t-test: t (14) = .923, p > .05 (η_p^2 = .057).
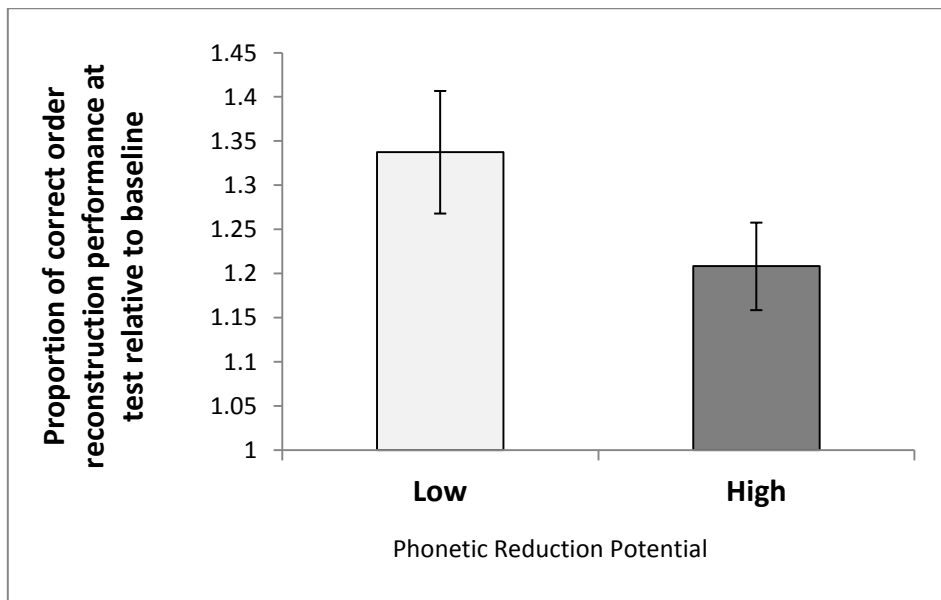
Figure 14. Mean proportion of improvement in order reconstruction performance between baseline and test phases, according to phonetic reduction potential. Error bars show standard error.

It is possible that the offline lenition effects measured in Experiment 6 do not translate linearly into comparable improvements in vSTM performance (i.e., a 10% reduction in articulatory duration may correspond to a smaller 5% improvement in vSTM performance), although this does not account for (non-significantly) greater memory improvements for nonwords with low phonetic reduction potential. A potential explanation for this unexpected pattern is considered below.

Due to the limited number of available nonwords and the nature of the order reconstruction task, participants encountered the same four nonwords in each condition throughout the experiment. Because these nonwords have phonologically distinct onsets (e.g., tadid, kegib, pabud, pedag) and need not be reproduced for an order reconstruction task, participants could complete the task simply by memorising the onset syllables of each nonword. For example, participants can effectively reconstruct the sequence 'tadid, kegib, pabud, pedag' by memorising the

four-syllable amalgam 'ta-ke-pa-pe'. This is problematic because the

manipulation of phonetic reduction potential centres on the medial and final

consonants in each nonword. By only memorising the onset syllables of

each nonword, participants could circumvent or even reverse the effect of

the manipulation. If the onset syllable strategy is utilised, order

reconstruction performance is determined by how well participants

memorise four-syllable strings of onsets (e.g., 'ta-ke-pa-pe'). If this strategy

is applied to high-potential nonwords, participants must memorise a series

of weak consonants (e.g., 'da-ge-ba-be'), but if the strategy is applied to

low-potential nonwords, participants must memorise a series of strong

consonants (e.g., 'ta-ke-pa-pe'). The manipulation of reduction potential is

based on the premise that equal exposure results in greater reduction for

strong consonants than for weak ones. Therefore, when only the onsets of

the nonwords are considered, the effect of the manipulation reverses:

Participants stand to improve more at memorising a series of strong

consonants (i.e., onsets in the low-potential condition) than memorising a

series of weak consonants (i.e., onsets in the high-potential condition).

General Discussion

The ultimate aim of this investigation was to influence vSTM

performance by means of a frequency-matched manipulation of offline

lenition. The facilitative effects of frequency on vSTM performance are

typically explained in terms of redintegration, a phonological process by

which decayed short-term memory traces are reconstructed from stable

long-term representations (e.g., Hulme et al., 1997). This redintegration

process is argued to be more effective for high-frequency words because

they have better-specified and more accessible phonological representations in long-term memory (e.g., Hulme et al., 1997). However, a demonstration that lenition (a process that reduces the articulatory complexity of affected items) influences vSTM performance independently of any variations in frequency will support the case for an alternative articulatory explanation. One way to demonstrate the link between frequency and lenition processes (while excluding redintegration) is to manipulate contextual factors that modulate the frequency's contribution to lenition effects while holding frequency at a constant value. Critically, it was unclear how - and indeed whether - a frequency-matched manipulation of offline lenition could be implemented. Therefore, the main objective of the investigation was to devise a means of experimentally inducing lenition in nonwords under frequency-matched conditions. This involved testing the contribution of various frequency-independent factors (including phrasal speech contexts, as well as semantic and phonetic properties) to the development of offline lenition in nonwords.

The first step was to establish a protocol that could be used to induce and elicit substantial amounts of offline lenition. One of the main issues to be addressed here was that speakers do not always use lenited word forms, even when they are available (e.g., Kohler, 1991; Kirchner, 2001). Experiment 4 established that offline lenition effects (i.e., persistent lenition effects that generalise beyond the reducing contexts in which they originally developed) manifest preferentially in nonwords that are articulated as part of carrier phrases rather than in isolation. However, implementing the same manipulation during nonword acquisition did not significantly influence the amount of offline lenition induced. Experiment 5 showed that lenition effects are magnified in longer and more complex nonwords. It also discounts the possibility that the experimental paradigm (which involves

comparisons between durational measurements taken before and after an exposure phase) generates fatigue or boredom-related order effects that lead to an under-measurement of lenition.

Experiment 6 demonstrated that enriching nonwords with semantic attributes does not predispose them to lenit more effectively, and may even inhibit the development of offline lenition. The consistency of this pattern with results from previous research into lexicalisation (e.g., Dumay et al., 2004) suggests the development of offline lenition might depend on lexicalisation rather than semantics. A second manipulation of nonwords' phonetic potential for reduction significantly influenced the degree to which they lenited. Specifically, nonwords with strong consonant gestures in phonetically-weakening positions and weak gestures in phonetically-strengthening positions lenited more than nonwords with the reverse configuration (i.e., weak consonant gestures in phonetically-weakening positions and strong gestures in phonetically-strengthening positions). As such, this phonetic manipulation satisfied the requirement for an effective frequency-matched manipulation of offline lenition.

Experiment 7 combined the manipulation of phonetic reduction potential with an order reconstruction task to determine if the resulting differences in lenition translate into vSTM performance. That is, would nonwords with a high potential for reduction also be subject to greater improvements in vSTM performance between baseline and test? Although this was not the case, there was some evidence that participants exploited mnemonic strategies to circumvent the effects of the manipulation. An alternative serial recall task was utilised to block this strategy, resulting in an improved but non-significant effect.

The exploratory experimental approach used here differs from conventional approaches to investigating lenition. Previous investigations into lenition have favoured static approaches that examine the language change process in its natural environment between two points. Diachronic analyses focus on historical sound changes such as those between Latin and Western Romance languages (i.e., French, Spanish, North-Italian; e.g., Bauer, 2008; Hualde, Simonet & Nadeu, 2011). Synchronic analyses typically focus on sound changes across dialects – for example, the 'flapping' of /t/ and /d/ to /r/ following stressed vowels in Irish and American-English (e.g., Carr & Honeybone, 2007; Marotta, 2008; Honeybone, 2012).

By comparison, the artificial induction of lenition in the laboratory is a challenging and largely unprecedented exploratory approach that involves numerous unknowns. This approach presents some unique methodological challenges - for example, the development of offline lenition cannot even be measured unless it can first be elicited. Nevertheless, this approach also offers some unique opportunities to understand the contribution of factors that are usually taken for granted. For example, the development of offline lenition may proceed via an exemplar shift process (e.g., Pierrehumbert, 2001) that only affects lexicalised words (and therefore has a limited effect on nonwords). Such a contribution of lexicalisation to the development of offline lenition could easily be overlooked in a natural language environment, where it is a given that most words will possess robust lexical representations. This investigation also adds to the findings of previous work on the importance of reducing environments for lenition (see Bybee, 2010) by demonstrating experimentally that these environments play an important part in eliciting as well as inducing persistent offline lenition effects.

The artificial induction of lenition in the laboratory is an approach that allows great control over variables such as exposure context and frequency. However, this level of control comes at a cost: Many of the factors that contribute to lenition in a rich natural language context - even if they can be anticipated effectively – prove difficult to reproduce in the laboratory. An obvious example is the difficulty of reproducing the scale involved in natural lenition, both in terms of time and exposure: Historical sound changes (such as the Latin 'mittere' to the Spanish 'meter' – e.g., Bauer, 2008) can require extended periods of time to fully unfold, and exemplar-based models of lenition (e.g., Pierrehumbert, 2001) incorporate tens of thousands of exposure events to simulate substantial lenition outcomes.

Similarly, the absence of pressures native to a genuine communicative context could prove problematic. Sound changes often result from compromises between pressures on a speaker to minimise articulatory effort while accommodating a listener's perceptual requirements (e.g., Lindblom, 1990). Usually, a speaker's purpose is to communicate a given message to a listener. The form of this message can be sacrificed subject to the listener's perceptual and comprehensive capabilities, so long as the message itself is received and understood. However, the paradigm utilised in these experiments does not involve speakers communicating a message to a listener. Instead, speakers articulate an utterance to a microphone. This context may not provide speakers with sufficient license to allow reductions into their speech.

Lenition is also subject to social and stylistic factors such as nature of the speaker-listener relationship and speech context. Even when lenited forms are available, whether or not they are used is at least partly subject

to speaker discretion (e.g., Kohler, 1991; Kirchner, 2001). Therefore, participants' awareness that their utterances are being recorded for examination, coupled with the laboratory setting, may lead them to suppress lenition and adopt a more careful speech style. Another notable social factor is the tendency for female speakers to use lenited forms less often than males (e.g., Byrd,1994). Given the predominantly female samples recruited throughout the investigation, this may have led to an under-elicitation of lenition.

Experiment 6 aimed to encourage lenition by introducing semantic factors into the laboratory setting. Specifically, nonwords were associated with one of two familiar visual objects (an apple or banana) in a bid to enrich them with semantic attributes. However, this is a limited and simplistic manipulation. Although it is possible to artificially introduce natural language factors into the laboratory, it is difficult to reproduce the depth, complexity and influence that these factors will possess in a natural language environment.

The strategy used here to investigate the development of offline lenition focuses narrowly on a handful of promising factors. However, this strategy suffers from a few blind spots. For example, the development of offline lenition could be heavily influenced by interactions between different factors. Examining such factors in isolation could lead to the mistaken conclusion that each is unimportant. Alternatively, lenition could be precipitated by an accumulation of highly numerous factors whose individual contributions are only weakly influential. Again, this state of affairs would prove difficult to detect with the present strategy. Further enquiry could benefit from a broader approach that prioritises the identification of additional factors that contribute to lenition and the

investigation of their combined effects. Particularly for the purposes of shaping vSTM performance via manipulations of lenition, it may be necessary to incorporate multiple factors, such that the size of the lenition effect exceeds the desired size of the vSTM effect.

One additional factor that could be investigated is the manner in which participants are exposed to novel words. For example, owing to automatization processes (by which words benefit from improved neuromotor efficiency with repeated practice; e.g., Bybee, 2002; Kapatsinski, 2010), nonwords could lenit more effectively if participants are exposed to them actively (i.e., by articulating them) rather than passively (i.e., by hearing them). Similarly, orthographically-presented nonwords may automatize more effectively than auditorily-presented nonwords. It has been argued that there is no direct access from orthography to the lexicon. Therefore, in order to allow lexical processing, verbal materials encountered in an orthographic form are automatically recoded into a phonological form (e.g., Luo, Johnson & Gallo, 1998; Peng, Ding, Perry, Xu, Jin & Luo, 2004). This conversion is not direct, but is mediated by articulatory recoding (e.g., Allport, 1979). Auditorily-presented nonwords, on the other hand, have direct access to the lexicon and do not require recoding. Consequently, it is possible that orthographic exposure to verbal materials also involves implicit but automatic articulatory exposure, whereas auditory exposure does not.

Conclusion

The ultimate aim of the investigation - to influence vSTM performance via a frequency-matched manipulation of offline lenition - was not realised. However, the results establish that this is possible in principle

by demonstrating that frequency is not the only factor capable of significantly influencing the development of offline lenition. For example, the distribution of strong and weak gestures throughout different consonantal positions can be manipulated such that nonwords are predisposed to lenit significantly more or less in response to moderate exposure. Studying offline lenition under artificial conditions (the use of nonword materials in particular) also provides insight into the importance of factors that might easily be overlooked in a more naturalistic context: In combination with results from previous research, the present findings suggest a role for lexicalisation processes in shaping nonword susceptibility to offline lenition.

General Thesis Discussion

The aim of the thesis was to provide evidence for the embodiment of articulatory detail in motoric processes that support and constrain both vSTM and speech function. However, vSTM function has conventionally been explained in terms of memory-specific mechanisms (e.g., phonological storage and trace redintegration) that prescribe a peripheral role for articulatory details and processes, acting instead on item-level phonological representations (e.g., Baddeley, 2012; Hulme et al., 1997). In order to provide plausible evidence for articulatory embodiment, it was necessary to experimentally control for these mechanisms. The experiments in Chapter 1 demonstrated that sequence-level coarticulatory fluency effects in vSTM cannot be classified as peripheral effects relating to misarticulated output because they persist in an order reconstruction task that does not directly involve the articulators. Neither can these coarticulatory effects be reinterpreted in terms of processes that operate on phonological items, because they persist when coarticulatory fluency is manipulated by reordering sequences of identical items. Moreover, the coarticulatory effects appear to originate in inner speech, outside of the context of vSTM tasks and the influence of memory-specific mechanisms.

Whereas Chapter 1 establishes that coarticulatory effects cannot be reinterpreted in terms of phonological processes, Chapters 2 and 3 examine whether supposedly phonological effects in vSTM can be reinterpreted as a consequence of articulatory processes. In particular, these chapters investigate whether superior vSTM performance for high-frequency and high-PND verbal materials (as is usually explained in terms of a phonological redintegration process) can be accounted for by articulatory effort minimization processes. Chapter 2 demonstrates that ease of articulation and PND are confounded, both in a sample of English

132

words, and in verbal materials utilised in previous experiments where PND is manipulated to influence vSTM performance (e.g., Roodenrys et al., 2002; Clarkson, 2013). Based on this pattern, it is argued that the language change process lenition (which reduces the articulatory complexity of affected words – e.g., Kirchner, 1998; Bauer, 2008) constrains PND effects in memory and language by shaping the phonological distributions that underlie these effects. Similarly, high-frequency words are particularly susceptible to lenition (e.g., Hooper, 1976; Bybee, 2010), meaning that frequency effects in vSTM may be partly mediated by reductions in articulatory complexity (e.g., Macken et al., 2014). However, it is difficult to evaluate this possibility because both redintegration and lenition are influenced by frequency. Chapter 3 explores methods for experimentally inducing lenition without recourse to manipulations of frequency, in order to disentangle lenition-based contributions to vSTM performance from those of redintegration. The contributions of these empirical chapters are reviewed in greater detail below.

The articulatory effects reported in this thesis, both in memory and in speech, are interpreted as follows: Serial recall requires the recoding of a presented verbal sequence into an articulatory form for output - a requirement that is shared by articulatory rehearsal processes more generally (e.g., Burton, Small & Blumstein, 2000). Consequently, performance in various memory tasks (as well as inner speech) depends on the efficacy with which speech motor control processes can recode verbal sequences into an articulatory form, as is determined by factors such as articulatory complexity and coarticulatory fluency. Moreover, factors such as articulatory complexity and coarticulatory fluency are not fixed but can be influenced by the language change process lenition.

Review of empirical contribution

Chapter 1 – Coarticulatory fluency in vSTM and Inner Speech

The experiments in Chapter 1 were concerned with the effects of coarticulatory fluency in vSTM tasks - an articulatory factor that relates to the sequence-level properties of verbal information. Previous experimental work shows that verbal sequences involving more fluent coarticulatory transitions between items are better-remembered in memory tasks (Murray & Jones, 2002). However, the manipulation of coarticulatory fluency used in this previous work was systematically confounded with variations in PND, an item-level variable known to facilitate vSTM performance (e.g., Roodenrys et al., 2002; Allen & Hulme, 2006). Specifically, verbal sequences involving fluent coarticulatory transitions between items also involved items with more phonological neighbours (see Miller, 2010). Consequently, the effect of the manipulation on memory performance was left open to reinterpretation in terms of a phonological redintegration process; the articulatory and sequence-level aspects of the manipulation could have been inconsequential.

The first aim of the experiments in Chapter 1 was to reassess the contribution of coarticulatory fluency to vSTM performance whilst controlling more effectively for variations in item-level properties. Previous work has manipulated coarticulatory fluency by including or excluding changes in place of articulation from the coarticulatory transitions between items (e.g., Murray & Jones, 2002; Woodward et al., 2008). However, Experiment 1 tested an alternative means of manipulating coarticulatory fluency by specifying the direction of changes in place of articulation. Specifically, transitions between stop consonants that involve a change from a given place of articulation to a more posterior place of articulation (e.g., from /p/

to /t/; a backward-moving change in anatomical terms) are more fluent than corresponding forward-moving changes (i.e., from /t/ to /p/; Byrd, 1996). In Experiment 1, this difference in fluency was reflected in better serial recall performance for word sequences involving exclusively backward rather than forward-moving coarticulatory transitions (words were also matched on lexical frequency and PND across conditions).

The value of this directional constraint on coarticulatory fluency is that it affords a manipulation of coarticulatory fluency across a set of identical items. For example, the nonword sequence 'pobe, dord, kug' involves exclusively fluent, backward-moving coarticulatory transitions. However, if the same items are presented in reverse order (i.e., 'kug, dord, pobe'), the sequence now involves exclusively disfluent, forward-moving transitions: Reversing the order of the sequence also reverses the direction (and hence the fluency) of the coarticulatory transitions within that sequence. Consequently, this directional constraint on articulatory fluency can be exploited to devise a manipulation that controls for variations in item-level properties entirely. Experiment 2a showed that nonword sequences with fluent, backward-moving coarticulatory transitions were better remembered in a serial recall task than reversed sequences with disfluent, forward-moving transitions. Because the manipulation controls for item-level factors entirely, this coarticulatory fluency effect cannot be reinterpreted in terms of item-oriented redintegration processes that are specific to memory (e.g., Hulme et al., 1997; Roodenrys et al., 2002). Nor can the effect be fully accounted for by psycholinguistic explanations that share a focus on item-level processes (e.g., Martin & Saffran, 1997; MacDonald & Christiansen, 2002; Acheson & MacDonald, 2009; these are considered in greater detail later). Experiment 2b replicated the coarticulatory effect found in Experiment 2a using an order reconstruction

task. Because order reconstruction does not involve the articulators directly (participants must reconstruct a scrambled version of the presented sequence into its original order rather than reproducing the original sequence vocally), the coarticulatory effect cannot be characterised as a peripheral effect that relates to the misarticulation of otherwise correctly-remembered sequences. Taken together, the results of Experiments 2a and 2b suggest that coarticulatory fluency effects in vSTM can be explained in terms of processes that support vSTM function directly but are not specific to this purpose.

The second aim of Chapter 1 was to show that coarticulatory fluency effects transcend the context of vSTM tasks. Experiment 3 investigated whether the coarticulatory fluency effect identified in Experiment 2 also constrains the timing of inner speech. Although not related to vSTM per se, inner speech could be co-opted to support vSTM functions via rehearsal. Whether or not the timing of inner speech is constrained by a manipulation of coarticulatory fluency also bears on a pre-existing debate as to whether inner speech specifies articulatory detail more generally (e.g., Oppenheim & Dell, 2008). The directional constraint from Experiments 1 and 2 was used to manipulate the coarticulatory fluency of nonword sequences. This manipulation had a similar impact on the time taken to read nonword sequences in overtly-articulated speech and inner speech (i.e., not involving any sound or articulatory movement). This suggests that inner speech embodies articulatory detail. Consequently, coarticulatory fluency effects in vSTM (e.g., Experiments 1 and 2; Murray & Jones, 2002) might best be characterised not as memory effects per se, but as inner speech effects that manifest in vSTM tasks.

Chapter 2 – PND effects and ease of articulation

Phonological neighbourhood density is a linguistic property that has been demonstrated to influence both vSTM performance and language production. Words with more phonologically-similar neighbours (a phonological neighbour being any word that differs from a given word by a single phoneme) are better-remembered in memory tasks, and are also more rapidly articulated (e.g., Roodenrys et al., 2002). Critically, PND effects in memory and language are underpinned by the heterogeneous distribution of speech sounds across phonological space. That is, some regions of phonological space are densely populated with attested speech sounds and words whereas others are more sparsely populated; if not for this uneven distribution, there would be no basis for differential PND effects in speech and memory. However, despite the importance of these distributions, little consideration has been given to their origin. Chapter 2 investigated the possibility that differences in the phonological distributions that underlie PND effects can be explained by the systematic influence of pressures to minimise articulatory effort. If articulatory factors are embodied in the phonological distributions that underlie PND effects on memory and language, they can indirectly account for these effects.

An analysis was employed to test the hypothesis that densely populated phonological regions tend to incorporate more easily articulated speech sounds. In order to quantify articulatory difficulty, an omnibus measure was devised based on three anatomical parameters – articulatory precision, muscular tension, and the efficiency of jaw movements. In a sample of English words, phonological neighbourhood density was found to differ significantly according to this omnibus measure of articulatory difficulty. By implication, phonological neighbourhood density distributions can ultimately be explained, at least partly, by articulatory pressures. The

same can be said of effects that depend on these density distributions, such as the facilitatory effects observed in vSTM tasks. Further analysis showed that previous manipulations of PND to influence vSTM performance are confounded with ease of articulation.

Chapter 3 – Frequency effects and lenition

Chapter 3 explored the possibility that the facilitative effects of lexical frequency on vSTM are mediated by the reductive articulatory consequences of the language change process known as lenition. Lenition is an effort-minimization process whereby commonly-used words tend to reduce in articulatory complexity (for example, the three-syllable 'ev-e-ry' often reduces to 'ev-ry' in speech). In order to establish that lenition can effectively mediate the facilitative effects of frequency on vSTM, it is necessary to demonstrate that the reductive articulatory consequences of lenition influence vSTM when frequency is controlled; it has already been established that lenition is associated with frequency (e.g., Hooper, 1976; Bybee, 2010), and that reductions in articulatory complexity are associated with superior vSTM performance (e.g., Experiments 1 & 2; Murray & Jones, 2002). Because frequency and lenition are naturally confounded, this investigation adopted an exploratory approach that aims to experimentally induce lenition in nonword materials by manipulating frequency-independent variables. The effects of these manipulations on lenition were measured by comparing measurements of nonword articulatory duration taken before and after an exposure phase where participants repeatedly articulated nonwords.

Experiment 4 established that lasting lenition effects (i.e., reductions in articulatory duration) can be induced in nonword materials through repeated articulatory exposure. Specifically, the same nonwords are

articulated more rapidly following an exposure phase than before this exposure phase. However, lenition effects induced through articulatory exposure only manifest when nonwords are subsequently articulated within phrasal speech contexts (i.e., when they are centrally embedded within carrier phrases). Experiment 5 shows that modest gains can be made to the amount of lenition induced in Experiment 4 by utilising longer and more complex nonword materials (i.e., trisyllabic CVCVC-format nonwords such as 'takkody' rather than disyllabic VCV-format nonwords such as 'akko'). Experiment 6 utilised a factorial design to test the influence of semantic and phonetic factors on lenition. Associating nonwords with pictorial cues for familiar fruit items (an apple or banana) in a bid to enrich them with semantic associations did not result in any more lenition than simple phonetic exposure. However, manipulating nonwords so as maximise their phonetic potential for reduction resulted in more lenition than a reverse treatment designed to minimise their phonetic potential for reduction. Specifically, more lenition was measured in nonwords where strong consonant gestures (e.g., such as 't') were placed in weakening lexical positions (i.e., positions where they were likely to reduce, such as intervocalic and offset position) and weak gestures (e.g., such as 'd') were placed in strong positions (i.e., positions where they were likely to be preserved, such as onset). This procedure formed nonwords with a high phonetic potential for reduction (such as 'datit'), which were contrasted against low-potential nonwords with the reverse configuration (i.e., strong gestures in strong positions and weak gestures in weak positions, such as 'tadid'). Experiment 7 tested whether lenition effects resulting from the manipulation of phonetic reduction potential in Experiment 6 would translate into vSTM improvements. To this end, pre-and post-exposure measures of articulatory duration were replaced with order reconstruction

tasks. However, improvements in order reconstruction performance were no greater for high-potential nonwords than for low-potential nonwords.

Theoretical and methodological implications

Memory-specific approaches to vSTM function

The evidence for articulatory effects in vSTM and speech presented in this thesis can be counterposed against accounts that invoke dedicated mechanisms to explain vSTM performance. The most influential of these is the standard model of short-term memory (see Baddeley, 2012), which proposes that vSTM is supported by a phonological loop system comprising two subcomponents. The first component is a phonological store whose sole purpose is to passively store phonological items. In this specialised capacity, the phonological store is distinct from both long-term memory and language systems and processes. Phonological traces held within this store decay over time but can be revivified by a separate active articulatory rehearsal process (the second subcomponent of the phonological loop). This model also allows for the influence of long-term linguistic properties (such as lexicality, semantic properties, frequency and neighbourhood density) on vSTM function. Degraded short-term traces can be reconstituted at output via a phonologically-oriented redintegration mechanism, which matches these degraded traces with (intact) corresponding items in long-term memory which can serve as an alternate basis for output.

For several reasons, it is difficult for memory-specific accounts to accommodate the findings presented within this thesis. Chapter 1 demonstrates a sequence-level (co)articulatory effect in vSTM that cannot be interpreted as a consequence of an item-level redintegration process due to careful controls for variation in item-level properties. The same

articulatory effect persists outside of the context of vSTM tasks in inner speech, where there are no grounds for the involvement of vSTM-specific mechanisms. Moreover, Chapters 2 and 3 show how articulatory processes relating to effort minimization can provide explanations for PND and frequency effects in vSTM without recourse to redintegrative processes that act on decayed phonological traces.

Psycholinguistic accounts of vSTM function

An alternative class of *psycholinguistic* accounts argues that vSTM phenomena can be explained in terms of linguistic systems and processes rather than memory-specific ones. This argument draws support from positional, lexical and phonological similarity constraints that are shared between the language production architecture and vSTM performance. According to one instantiation of this view, vSTM function depends on the temporary activation of long-term verbal (i.e., phonological, lexical and semantic) representations that underlie linguistic processing. Specifically, activation of phonological features spreads upwards to corresponding representations in a hierarchy of lexical and semantic layers. These representations are linked by mutual excitatory connections. For example, the phonological features /k/, / æ/, and /t/ correspond to the lexeme 'kæt' (cat), which corresponds to various semantic features such as 'animal', 'pet', and 'feline'. Activation of these phonological features spreads upwards to lexical and semantic representations, and then feeds back down to the originally activated phonological features, sustaining their activation. Therefore, the temporary storage of verbal material need not be accounted for by a dedicated storage component, but is fulfilled by the same system that processes verbal material (e.g., Martin & Saffran; 1997). This process can account for various memory effects – for example, it can explain why linguistic familiarity/lexicality effects emerge in vSTM. Because

unfamiliar verbal materials will have weak or absent lexical and semantic representations, they will benefit less (if at all, in the case of nonwords) from feedback activation from lexical and semantic representations.

Although the psycholinguistic approach does not appeal to memory-specific mechanisms to explain vSTM function, it shares with the memory-specific approach a focus on item-level properties and explanations (e.g., MacDonald & Christiansen, 2002). Consequently, it is difficult to reconcile the psycholinguistic approach with the results of Chapter 1, where sequence-level articulatory factors were shown to constrain memory and speech even when variations in item-level properties were controlled for.

Embodied accounts of vSTM function

A third approach within which evidence for articulatory effects in vSTM and speech can be situated is an *embodied* approach to understanding cognition (e.g., Wilson, 2002; Postle, 2006; Barsalou, 2008; Shapiro, 2011; Wilson & Golonka, 2013). According to this approach, neither vSTM nor language are special cognitive functions supported by dedicated systems. Instead, cognitive functions (e.g., memory, language, attention) are not only situated in the brain, but within bodily interactions with the outside environment (e.g., Wilson, 2002). These interactions are mediated by distributed perception and action-oriented processes that can be exploited as resources to support cognitive function. By this token, perceptual and productive factors that have conventionally been thought to play a peripheral role in constraining cognitive performance instead play a central role in supporting cognitive function.

The importance of task demands

In embodied cognition, cognitive problems are solved by improvising task-specific solutions from available resources (e.g., Wilson & Golonka, 2013). This can be counterposed against the more conventional notion that general cognitive problems (e.g., vSTM tasks) are solved by dedicated supporting systems (e.g., the phonological loop). Consequently, the embodied approach places a special importance on identifying the specific demands of particular cognitive tasks and on recognizing existing resources (i.e., perception and action-oriented skills; e.g., Glenberg, 1987) that can be deployed to satisfy these demands.

The importance of identifying the specific demands of cognitive tasks can be illustrated in relation to short-term memory. Although various cognitive tasks are commonly classified as 'verbal short-term memory tasks' (e.g., free recall, serial recall, order reconstruction, serial recognition), the particular demands of these tasks are diverse: Some require that materials are retained for later reproduction (e.g., recall tasks) whereas in others the materials are provided (e.g., reconstruction/recognition tasks). In some of these tasks it is necessary to retain the order of the materials (e.g., serial recall and order reconstruction tasks), but in others, only their identities must be retained (e.g., free recall). Moreover, participants could be required to remember materials that are presented either visually or auditorily in any of these tasks.

The importance of existing resources

It is also important to recognize existing resources (i.e., perception and action-oriented skills) that can be deployed to satisfy the demands of a given cognitive task. An accumulation of evidence now demonstrates how

various perception and action-oriented skills can be co-opted to complete cognitive tasks in an opportunistic fashion, obviating the need to invoke memory-specific processes (e.g., Jones & Nicholls, 2002; Macken & Jones, 2003; Jones, Macken & Nicholls, 2004; Jones, Hughes & Macken, 2006; Macken, Taylor & Jones, 2014).

For example, a common demand in short-term memory tasks (as well as speech and behaviour more generally) is the retention and production of ordered sequences of materials (e.g., Lashley, 1951). One way in which order can be imposed on the materials in vSTM tasks is by deploying speech motor control processes to encode these materials into an articulatory sequence (see Chapter 1; Woodward et al., 2008). However, motor control processes will not necessarily be deployed in all vSTM tasks. Because perceptual and gestural skills are deployed in an opportunistic fashion, motor control processes will not be deployed if a vSTM task can be more efficiently solved with an alternative skill. This opportunistic deployment has been demonstrated in recent experimental work (Macken et al., 2014), which is considered in more detail below.

Various linguistic familiarity effects are observed in serial recall tasks. For example, real words are better remembered than nonwords (the lexicality effect – e.g., Hulme, Maughan, & Brown, 1991). According to the embodied view, a lexicality effect is observed in serial recall because familiar linguistic materials (i.e., real words as opposed to nonwords) are more efficiently processed by speech motor control processes that are co-opted to solve the task. However, although robust lexicality effects occur in serial recall, they are much smaller in serial recognition tasks (where participants must judge whether a sequence of verbal materials is the same as a previously presented sequence). Lexicality effects in vSTM are typically explained in terms of a redintegration process (e.g., Hulme et al,

1997; Roodenrys et al., 2002), and smaller lexicality effects in serial recognition can be explained by the smaller part that reconstructive redintegration processes play when participants are provided with intact items.

An alternative embodied explanation for the marginal lexicality effects in serial recognition relates to the almost exclusive use of auditory presentation in these tasks. Critically, presenting materials auditorily rather than visually allows the recognition task (which in effect involves matching two extended auditory 'objects') to be solved by deploying perceptual (acoustic) pattern-matching processes. This efficient solution obviates the need to deploy speech motor control processes as in serial recall tasks. Whereas speech motor control processes are constrained by properties that relate to linguistic familiarity, acoustic pattern-matching processes are not – hence the attenuation of the lexicality effects usually observed in serial recognition. In support of this embodied interpretation, robust lexicality effects emerge in serial recognition tasks when visual presentation is utilised instead of the more traditional auditory presentation (e.g., Macken et al., 2014).

An important principle of the embodied approach is that there is no need to stipulate theoretical components dedicated to servicing particular cognitive functions if existing perception and action-oriented processes can be co-opted to service the same functions (e.g., Crowder, 1993; Postle, 2006). It is therefore important to recognize perception and action-oriented resources that could be used to solve cognitive tasks. This principle is reflected in the argument that memory and speech functions conventionally fulfilled by a bespoke phonological store (e.g., Baddeley, 2012) can instead be fulfilled by speech motor processes.

The phonological store is a specialised subcomponent of the phonological loop system postulated by the standard model of short-term memory. This store, whose sole purpose is to passively store phonological items, works in conjunction with an active articulatory rehearsal process that refreshes stored items as their traces decay over time (e.g., Baddeley, Lewis & Vallar 1984). This phonological loop system arose from the reconceptualisation of an originally articulatory store. The articulatory store was subdivided into an articulatory process and a phonological store to accommodate the dissociation between speech production and phonological storage capacity implied by patients with vSTM-specific impairments (despite apparently preserved language function). The complex of the vSTM patient was assumed to reflect a specific deficit in phonological storage, as might be accounted for by damage to a distinct phonological storage component. By the same token, normal vSTM function was argued to be supported by an intact phonological store.

An alternative proposal is that vSTM function is supported by the action of an auditory-motor interface whose purpose is to translate between auditory and articulatory codes (e.g., Buchsbaum & D'Espositio, 2008). This interface can be thought of as an auditory counterpart to visuomotor integration circuits previously discovered in the posterior parietal cortex (e.g., Andersen, Snyder, Bradley & Xing, 1997). These visuomotor circuits translate between visual and motoric representations, allowing actions to be guided by sensory feedback. For example, the visual representation of a cup can be translated into a motoric representation that bears on how the cup should be grasped (e.g., by curling the fingers around a visible handle); visual feedback can also guide a grasping action as it unfolds. The auditory-motor interface is argued to play an analogous role in guiding articulatory behaviour with auditory information (e.g., Buchsbaum, Baldo,

146

Okada, Berman, Dronkers, D'Esposito & Hickok, 2011). As regards vSTM function, articulatory rehearsal processes that support vSTM depend on this translation of auditory input (e.g., the consonant /t/) into instructions for articulatory output (e.g., form a constriction between the tongue tip and alveolar ridge; allow a brief buildup of air pressure behind this constriction before releasing it).

However, the role of this auditory-motor interface is not restricted to supporting vSTM; it also supports speech production and comprehension functions. Notably, patients with supposedly vSTM-specific impairments (which have been taken as evidence for a phonological store that is distinct from language production processes – e.g., Vallar & Baddeley, 1984) also display subtle language deficits that relate to the production of nonlexical materials (i.e., nonwords and single syllables – e.g., Allport, 1984). These additional deficits can be accounted for by damage to an auditory-motor interface: The inability to directly translate between auditory and acoustic representations may be masked in the case of lexical materials, because their articulatory representations can be accessed via an alternative semantic route (i.e., acoustic representations can be translated into semantic representations, which can be translated into articulatory representations). However, production deficits resulting from a damaged auditory-motor interface become clear when this alternative is eliminated, as in the case of nonword materials (which do not possess mediating semantic representations).

Methodological implications and considerations

The findings in Chapter 1 and 3 reinforce previous claims that some articulatory phenomena only emerge in sequence (e.g., Wheeldon, 2000). Not only is this the case for coarticulatory effects that relate specifically to

the transitions between words (see Chapter 1; Murray & Jones, 2002; Woodward et al., 2008), it applies to the emergence of lenition effects in single words (see Chapter 3). One implication is that where articulatory factors are concerned, measurements of sequence-level duration should be taken so as to avoid overlooking effects that only emerge in sequence.

At first, it seems that sequence-level articulatory duration plays an important part in accounting for vSTM performance. However, previous evidence shows that vSTM performance is not constrained by articulatory duration so much as articulatory complexity. Specifically, factors such as the number of phonemes or syllables in a word constrain vSTM performance even when articulatory duration is controlled for (e.g., Service, 1998). This is not to say that articulatory complexity and articulatory duration are naturally unrelated; however, it is articulatory complexity that makes the crucial contribution to vSTM performance. Based on this evidence, the position taken in this thesis (and in related previous work) is that differences in vSTM performance can be accounted for by articulatory (or coarticulatory) complexity rather than duration. This position can be contrasted with the notion that vSTM performance is determined by temporally-constrained processes of trace decay and articulatory refreshment (e.g., Baddeley, 2012).

Nevertheless, both in the present investigation and in related previous work that argues for the importance of articulatory complexity in vSTM (e.g., Murray & Jones, 2002; Woodward et al., 2008), measurements of articulatory duration were utilised. These durational measurements are assumed to provide a sufficient, if imperfect, approximation of complexity. Therefore, one way to strengthen the case for the importance of complexity in vSTM would be to incorporate more direct measures of articulatory complexity. However, whereas word-level complexity could be accounted

for in terms of syllable or phoneme count (for example, as in Service, 1998), it might prove more challenging to account for coarticulatory complexity in the same way (except in cases where coarticulation results in phoneme or syllable deletion). A potential way to quantify coarticulatory complexity would be in terms of the gestural scores used in articulatory phonology (e.g., Browman & Goldstein, 1992). Specifically, more complex coarticulations could be accounted for by more activity across different tract variables (e.g., lip aperture; tongue tip constriction degree). For example, it should be possible to account for a coarticulation between the alveolar consonant /t/ and the alveolar consonant /d/ in terms of tract variables relating to the tongue tip (e.g., tongue tip constriction degree and location). However, accounting for a change in place of articulation between the alveolar consonant /t/ and the labial consonant /p/ would require a combination of activity across tract variables relating the tongue tip (e.g., tongue tip constriction degree and location) as well as the lips (e.g., lip aperture and protrusion).

The importance of sequence-level articulatory effects has been overlooked in the past due to the use of restrictive item-level measurements (e.g., Woodward et al., 2008), but these item-level measurements are also poorly suited to detecting potential sublexical effects. In order to avoid a fixation on item-level phenomena, it could be useful to incorporate both sequence-level and sub-item measures (such as sub-item speech error analysis; Acheson & MacDonald, 2009) into future experimental work.

Just as it would be useful to make use of alternative measures in memory tasks, it would be beneficial to employ a broader range of memory tasks. As discussed earlier, various vSTM tasks entail quite different demands that can lead to different effects. Just as the use of traditional

item-level measures has obscured sequence-level articulatory effects (e.g., Woodward et al., 2008), the traditional use of auditory (but not visual) serial recognition tasks has obscured lexicality effects previously thought to be absent from serial recognition (e.g., Macken et al., 2014). These examples show that diversifying experimental tasks and measures reveals additional phenomena that can be used to enrich and re-evaluate relevant theory.

Another methodological issue relates to the use of nonword materials in the present investigation (see Chapters 1 and 3) as well as in previous research into articulatory effects in memory (e.g., Woodward, 2006). Relative to regular verbal materials, nonwords can be used to more effectively isolate and manipulate articulatory properties and effects. However, it is possible that the exclusive use of nonword materials magnifies the importance of these articulatory properties and effects. Specifically, articulatory processes may compensate for the absence of other processes that usually contribute to memory and speech functions (for example, processes related to semantics or the context and frequency of prior usage). Although the importance of articulatory factors in memory and speech has been underestimated in the past due to the use of restrictive item-level measurements (e.g., Woodward et al, 2008), it is also important not to overestimate the importance of these articulatory factors due to restrictions in the nature of the experimental materials.

Are embodied effects extensive enough to account for cognition more generally?

One question raised by demonstrations of embodied effects in cognitive performance is that of how extensive these effects are. At first, this appears to be a question of how influential embodied mechanisms (e.g., speech motor control processes in inner speech) are relative to

'disembodied' mechanisms (e.g., trace redintegration) that support and constrain cognition. However, a strong version of embodied cognition maintains that embodied and disembodied mechanisms cannot coexist. Instead, the supposed functions of disembodied mechanisms must ultimately be accounted for in purely embodied terms (the 'replacement hypothesis' – e.g., Shapiro, 2011) The basis for this argument is that acknowledging embodied effects at all requires a redefinition of our understanding of the foundations on which cognition operates. By definition, embodied effects must act through an embodied system. Therefore, embodied effects cannot be considered as additional factors that act on otherwise disembodied processes (e.g., Wilson & Golonka, 2013).

In light of this argument, is it reasonable to expect that embodied effects could account for cognition more generally? How do the present findings bear on this question? The articulatory effects revealed in the previous experimental chapters (e.g., coarticulatory fluency effects in Chapter 1; articulatory difficulty effects in Chapter 2) were modest in size, which may at first seem at odds with the notion of a replacement hypothesis. However, consider that the priority of those experiments reported in Chapter 1 was not to demonstrate the size of coarticulatory effects but to control for the contributions of item-level mechanisms to vSTM performance: A subtle constraint on coarticulatory fluency was utilised (i.e., the direction of changes in place of articulation) specifically because it afforded a manipulation that controlled for variations in item-level properties. Alternative constraints on coarticulatory fluency exist, some of which are coarser and likely to elicit larger effects (e.g., the presence or absence of changes in place of articulation - Murray & Jones, 2002; Woodward et al., 2008). Moreover, coarticulatory fluency constraints represent just one class of articulatory effects in memory and speech. For

example, articulatory effects can also operate at the level of single items, and the effects of articulatory suppression on memory are well-documented (e.g., Besner, 1987).

In the case of Chapter 2, the difficulty of word onsets was used to approximate the articulatory difficulty of simple (single-syllable) English words. It is possible that fuller characterisations of articulatory difficulty (i.e., word-level characterisations that account for difficulty across a short sequence of gestures) would lead to stronger links with word-level PND properties. However, characterisations of articulatory difficulty are admittedly still lacking, even at the level of single gestures (e.g., Ann, 2005). At the least, it would be premature to dismiss the performance impact of articulatory difficulty factors that cannot yet be satisfactorily measured.

Moreover, production-oriented constraints on cognitive performance are not limited to the verbal domain. Analogous effects can be found in visuospatial order reconstruction tasks. In these tasks, a number of dots are serially presented across various spatial locations. These are then simultaneously re-presented in their correct locations, and participants are required to select the dots in their original order of presentation. In these tasks, memory performance is constrained by the characteristics of the visual path participants must take between the dots to correctly reconstruct the presentation order of the original sequence (e.g., Parmentier, Elford & Maybery, 2005). Three factors in particular impair performance on this task - the total length of the correct visual path between the items, the acuteness of the changes in angle required at each item, and the number of occasions on which the visual path crosses itself. Therefore, as in the verbal domain, productive factors related to the transitions between sequence items constrain memory performance.

152

The importance of perception

Embodied cognition prescribes an important role for perceptual as well as productive factors in constraining cognitive performance. This is justified by the essential role perception plays in guiding action. Yet, much like production-related articulatory factors, perceptual factors have conventionally been viewed as a peripheral influence on cognition. Therefore, those processes essential to the support of vSTM function are typically assumed to be post-perceptual. However, more recent work shows that perceptual organisation processes can provide alternative explanations for supposedly memory-specific phenomena such as the phonological similarity effect.

The phonological similarity effect (e.g., Conrad, 1964) refers to impaired vSTM performance for sequences of similar-sounding items (for example, 'b, d, g, t, c' as opposed to 'f, q, r, h, y').The phonological similarity effect has been taken as evidence for the phonological loop model of vSTM because, under conditions of articulatory suppression (where participants must perform while they concurrently articulate an irrelevant sound), it is abolished in visually-presented sequences but persists in auditorily-presented sequences. Ostensibly, this is because auditory material has direct access to the phonological store (and is therefore subject to phonological similarity effects). Visual material, on the other hand, must be recoded into a phonological medium in order to gain access to the phonological store. This recoding is performed by articulatory processes that are otherwise engaged during articulatory suppression. Consequently, visually-presented sequences are exempt from the phonological similarity effect under conditions of articulatory suppression because they cannot enter the phonological store.

153

However, more recent experimental work (e.g., Jones, Macken & Nicholls, 2004) shows that the phonological similarity effect can be reinterpreted as a consequence of perceptual organisation processes rather than memory-specific phonological processes. An analysis of serial position data indicates that the survival of the phonological similarity effect in auditorily-presented sequences under articulatory suppression is accounted for by differences in serial recall performance for sequence-final items. This effect, being specific to the final items in auditory sequences, can be characterised as an auditory recency effect of a perceptual (i.e., acoustic) rather than phonological nature. The importance of auditory recency is demonstrated by suffix effects: Adding a redundant suffix to the end of an auditory sequence disrupts memory performance for that sequence (e.g., Crowder & Morton, 1969), particularly if the suffix is perceptually similar to other items in the sequence (e.g., it is presented in the same female voice, as opposed to a male voice). This suffix effect can be explained in terms of perceptual organization processes: An auditory sequence can be thought of as an extended auditory object with perceptually distinctive edges (hence primacy and recency effects - e.g., Bregman & Rudnicky, 1975; Botvinick & Plaut, 2006). If a redundant but perceptually similar suffix is added to the end of this sequence, it is treated as a part of the same auditory object (e.g., Jones et al 2004), and will disrupt the encoding of order within the target sequence (e.g., Nicholls & Jones, 2002).

If the phonological similarity effect survives in auditory sequences under articulatory suppression due to auditory recency, and auditory recency can be eliminated with a suffix, it should be possible to abolish the survival of the similarity under these conditions by appending an auditory sequence with a redundant suffix. Experimental work bears out this

prediction exactly (e.g., Jones et al, 2006). Under these conditions, the phonological similarity effect can be explained by auditory factors embodied in perceptual organization processes rather than amodal, memory-specific processes.

The embodied effects of perceptual processes add to the embodied effects of productive processes to provide a more extensive account of cognitive performance. However, these two types of processes also interact to account for further phenomena that exceed the scope of either perceptual or productive factors acting in isolation. For example, perceptual and productive processes interact to influence the effort minimization processes discussed in Chapters 2 and 3. Communication is subject to competing constraints that favour efficient production on the one hand and perceptual clarity on the other (e.g., Lindblom, 1990). Consequently, reductions in articulatory complexity cannot be understood purely as a consequence of production-oriented effort minimization processes. The same reductions could result from an easing of perceptual clarity pressures, as might accompany a change from a noisy environment to a quieter one. An opposite tendency towards increased articulatory effort might be observed when perceptual clarity pressures are magnified, as in foreigner or child-directed speech (e.g., Uther et al., 2007).

Conclusions

This thesis focuses on the role of articulatory effort minimization processes in memory and speech. It demonstrates that vSTM performance is constrained by coarticulatory fluency effects that operate on verbal sequences as opposed to items, and that these coarticulatory fluency effects extend beyond the context of vSTM tasks into inner speech, despite the absence of any direct involvement of the articulatory apparatus.

Moreover, vSTM advantages for high-frequency and high-PND words, as are conventionally accounted for by a redintegration process that operates on phonological items, can be alternatively explained by lenition – a language change process by which affected words reduce in articulatory complexity. The tendency for higher-PND words to involve less effortful articulatory features suggests that lenition plays a role in shaping the phonological distributions that underlie PND effects in memory and speech. Similarly, high-frequency words are particularly susceptible to lenition.

These findings, together with a handful of other studies, imply a previously underappreciated role for articulatory factors in memory and speech. Such findings are difficult to account for in terms of theoretical approaches that ascribe a peripheral role to articulatory factors in cognition but emphasize the importance of item-oriented phonological processes. Although particularly problematic for accounts that invoke memory-specific mechanisms to explain vSTM function, the present findings cannot be fully accounted for by alternative psycholinguistic accounts either, given that both share a focus on item-level processes. The findings can be accommodated within an increasingly influential embodied approach to explaining cognition, which argues that cognitive functions are embodied in the action of various perceptual and motoric processes. These processes are deployed to solve cognitive tasks based on the match between their inherent capacities and the particular demands of the tasks. The embodied approach accommodates the view that articulatory effects in vSTM reflect the dependency of some vSTM tasks on sequential motor control processes that operate in inner speech.

REFERENCES

Abramson, A.S., & Lisker, L. (1964) A cross-language study of voicing in initial stops: acoustical measurements. *Word. 20*, 384–422.

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

Allen, R., & Hulme, C. (2006). Speech and language processing mechanisms in verbal serial recall. *Journal of Memory and Language*, *55*(1), 64-88.

Allport, A. (1979). Word recognition in reading (Tutorial paper). In *Processing of visible language* (pp. 227-257). Springer: US.

Allport, D. A. (1984). Auditory–verbal short-term memory and conduction aphasia. In H. Bouma & D. G. Bouwhuis (Eds.), Attention and performance: X. Control of language processes (Vol. 10, pp. 313–325). Hillsdale, NJ: Erlbaum.

Andersen, R. A., Snyder, L. H., Bradley, D. C., & Xing, J. (1997). Multimodal representation of space in the posterior parietal cortex and its use in planning movements. Annual Review of Neuroscience, 20, 303–330.

Ann, J. (1993). *A linguistic investigation of the relationship between physiology and handshape*. Tuscon: University of Arizona dissertation.

Ann, J. (1996). On the relation between ease of articulation and frequency of occurrence of handshapes in two sign languages. *Lingua, 98(*1), 19-41.

Ann, J. (2005). A functional explanation of Taiwan Sign Language handshape frequency. *Language and linguistics. 6* (2), 217.

Anttila, R. (1972). *An introduction to historical and comparative linguistics*. New York: Macmillan.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*(1), 67-82.

Baayen, R.H., Piepenbrock, R., & Van Rijan, H. (1993). *The Celex Lexical Database*. Philadelphia: Linguistic data consortium, University of Pennsylvania.

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

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

Baddeley, A., & Andrade, J. (1994). Reversing the word-length effect: A comment on Caplan, Rochon, and Waters. *The Quarterly Journal of Experimental Psychology*, *47*(4), 1047-1054.

Baddeley, A., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology,* 36, 233–252.

157

Baldo, J. V., Wilkins, D. P., Ogar, J., Willock, S., & Dronkers, N. F. (2011). Role of the precentral gyrus of the insula in complex articulation. *Cortex*, *47*(7), 800-807.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology.*, *59*, 617-645.

Bauer, L. (2008). Lenition revisited. *Journal of Linguistics*, *44*(03), 605-624.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, *60*(1), 92-111.

Besner, D. (1987). Phonology, lexical access in reading, and articulatory suppression: A critical review. *The Quarterly Journal of Experimental Psychology*, *39*(3), 467-478.

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

Botvinick, M. M., & Plaut, D. C. (2006). Short-term memory for serial order: a recurrent neural network model. *Psychological review*, *113*(2), 201.

Brainard, D. H. (1997) The Psychophysics Toolbox, *Spatial Vision. 10*: 433-436.

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

Brinton, L. J., & Traugott, E. C. (2005). *Lexicalization and language change*. Cambridge University Press.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, *49*(3-4), 155-180.

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

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

Burgess, N., & Hitch, G. J. (1992). Toward a network model of the articulatory loop. *Journal of memory and language*, *31*(4), 429-460.

Burton, M. W., Small, S. L., & Blumstein, S. E. (2000). The role of segmentation in phonological processing: an fMRI investigation. *Cognitive Neuroscience, Journal of*, *12*(4), 679-690.

Butcher, A. (2004). 'Fortis/lenis' revisited one more time: The aerodynamics of some oral stop contrasts in three continents. *Clinical linguistics & phonetics*, *18*(6-8), 547-557.

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, *14*(03), 261-290.

Bybee, J. (2010). *Language, usage and cognition*. Cambridge, England: Cambridge University Press.

Bybee, J., & Hopper, P. (Eds.). (2001). Frequency and the emergence of linguistic structure. Amsterdam: Benjamins.

Bybee, J., & Scheibman, J. (1999). The effect of usage on degrees of constituency: the reduction of don't in English. *Linguistics*, *37*(4), 575-596.

Byrd, D. (1994). *Articulatory timing in English consonant sequences* (Vol. 86). Phonetics Laboratory, Deptartment of Linguistics, UCLA.

Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Commun*, *3*, 39–54.

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, *24*(2), 209-244.

Caplan, D., Rochon, E., & Waters, G. S. (1992). Articulatory and phonological determinants of word length effects in span tasks. *The Quarterly Journal of Experimental Psychology*, *45*(2), 177-192.

Carr, P., & Honeybone, P. (2007). English phonology and linguistic theory: an introduction to issues, and to 'Issues in English Phonology'. *Language Sciences*, *29*(2), 117-153.

Chitoran, I. Goldstein, L. & Byrd, D. (2002). Gestural overlap and recoverability: articulatory evidence from Georgian. In C. Gussenhoven, & N. Warner (Eds.), Laboratory phonology VII. Berlin: Walter de Gruyter.

Cho, T. & Keating, P. (2001). Articulatory and acoustic studies of domain-initial strengthening in Korean. *Journal of Phonetics, 29*, 155–190.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics*, *27*(2), 207-229.

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

Clarkson, L. (2013). *The phonological neighbourhood effect on short-term memory for order*. Doctoral thesis, University of Wollowong.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology*, *33*(4), 497-505.

Coltheart, V. (1999). Phonological codes in reading comprehension, short-term memory, and memory for rapid visual sequences. In V. Coltheart (Ed.), Fleeting Memories Cognition of Brief Visual Stimuli (pp. 181 – 225). Cambridge MA: The MIT Press.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British journal of Psychology*, *55*(1), 75-84.

Corley, M., Brocklehurst, P. H., & Moat, H. S. (2011). Error biases in inner and overt speech: Evidence from tongue twisters. *Journal of experimental psychology: Learning, memory, and cognition*, *37*(1), 162.

Crowder, R. G. (1993). Short term memory—Where do we stand? *Memory & Cognition, 21,* 142–145.

Davis, C. J. (2005). N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behaviour research methods*, *37*(1), 65-70.

Deese, J. (1960). Frequency of usage and number of words in free recall: The role of association. *Psychological Reports*, *7*(2), 337-344.

Delattre, P. (1965). *Comparing the phonetic features of English, German, Spanish, and French*. Heidelberg, Germany: Verlog.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological review*, *93*(3), 283.

Denes, P. (1955). Effect of duration on the perception of voicing. *The Journal of the Acoustical Society of America*, *27*(4), 761-764.

Derwing, B. L. (1992). Orthographic aspects of linguistic competence. *The linguistics of literacy*, 193-210.

Donegan, P. J., & Stampe, D. (1979). The study of natural phonology. In *Current approaches to phonological theory*, *pp. 126-174.*

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, *18*(1), 35-39.

Dumay, N., Gaskell, M.G., & Feng, X. (2004). A day in the life of a spoken word. In K. Forbus, D. Gentner, & T. Regier (Eds.), Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society (pp. 339–344). Mahwah, NJ: Erlbaum.

Edwards, M. L., & Shriberg, L. D. (1983). *Phonology: Applications in communicative disorders*. College Hill Press.

Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, *4*(1), 99-109.

Ellis, N. C., & Hennelly, R. A. (1980). A bilingual word-length effect: Implications for intelligence testing and the relative ease of mental calculation in Welsh and English. *British Journal of Psychology*, *71*(1), 43-51.

Fallon, A. B., Groves, K., & Tehan, G. (1999). Phonological similarity and trace degradation in the serial recall task: When CAT helps RAT, but not MAN. *International Journal of Psychology*, *34*(5-6), 301-307.

Garlock, V. M., Walley, A. C., & Metsala, J. L. (2001). Age-of-acquisition, word frequency, and neighborhood density effects on spoken word recognition by children and adults. *Journal of Memory and language*, *45*(3), 468-492.

Goh, W. D., & Pisoni, D. B. (2003). Effects of lexical competition on immediate memory span for spoken words. *The Quarterly journal of Experimental Psychology*, *56*(6), 929-954.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia*, *9*(3), 317-323.

Gow, D. W., Manuel, S., & Melvold, J. (1996). How word onsets drive lexical access and segmentation: Evidence from acoustics, phonology and processing. Paper presented at the Fourth International Conference on Spoken Language Processing, Philadelphia.

Gregory, M. L., Raymond, W. D., Bell, A., Fosler-Lussier, E., & Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production. In *Chicago Linguistic Society* (Vol. 35, pp. 151-166).

Grigos, M. I., & Kolenda, N. (2010). The relationship between articulatory control and improved phonemic accuracy in childhood apraxia of speech: A longitudinal case study. *Clinical linguistics & phonetics*, *24*(1), 17-40.

Grimme, B., Fuchs, S., Perrier, P., & Schöner, G. (2011). Limb versus speech motor control: A conceptual review. *Motor control*, *15*(1), 5-33.

Grush, R. (2004). The emulation theory of representation: motor control, imagery, and perception. *Behavioural and brain sciences*, *27*(03), 377-396.

Gupta, P., Lipinski, J., Abbs, B., & Lin, P. H. (2005). Serial position effects in nonword repetition. *Journal of Memory and Language*, *53*(1), 141-162.

Haberlandt, K., Lawrence, H., Krohn, L., Bower, K., & Thomas, J. G. (2005). Pauses and durations exhibit a serial position effect. *Psychonomic bulletin & review*, *12*(1), 152-158.

Hardcastle, W.J., (1973). Some observations on the tense-lax distinction in initial stops in Korean. *Journal of phonetics, 1*, 263-272.

Harris, G. J., & Burke, D. (1972). The effects of grouping on short-term serial recall of digits by children: Developmental trends. *Child Development*, 710-716.

Hockett, C. F. (1967). The quantification of functional load. *Word*, *23*(1-3), 300-320.

Honeybone, P. (2008). Lenition, weakening and consonantal strength: tracing concepts through the history of phonology. *Lenition and Fortition. Berlin: Mouton de Gruyter*, 9-93.

Honeybone, P. (2012). Lenition in English. In T. Nevalainen, & E. Traugott (Eds.), *The Oxford Handbook of the History of English*. New York: Oxford University Press.

Hooper, J. D. (1976). Word frequency in lexical diffusion and the source of morphophonological change (pp. 96–105). In W. Christie (Ed.), *Current progress in historical linguistics*. Amsterdam, the Netherlands: North Holland.

Hualde, J. I., Simonet, M., & Nadeu, M. (2011). Consonant lenition and phonological recategorization. *Laboratory Phonology*, *2*(2), 301-329.

Hulme, C., Maughan, S., & Brown, G. D. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. Journal of Memory and Language, 30, 685-701.

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 redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(5), 1217.

Hulme, C., Stuart, G., Brown, G. D., & Morin, C. (2003). High-and low-frequency words are recalled equally well in alternating lists: Evidence for associative effects in serial recall. *Journal of Memory and Language*, *49*(4), 500-518.

Ingham, R. J. (2006). Speech effort measurement and stuttering: Investigating the chorus reading effect. *Journal of Speech, Language, and Hearing Research*, *49*(3), 660-670.

Jaeger, J. J. (1983). The fortis/lenis question: evidence from Zapotec and Jawoñ. *Journal of phonetics*, *11*(2), 177-189.

Jalbert, A., Neath, I., Bireta, T. J. & Suprenant, A. M., (2011). When does length cause the word length effect? *Journal of Experimental Psychology: Learning, Memory and Cognition*, 37 (2), 338.

Johnson, K., Pisoni, D. B., & Bernacki, R. H. (1990). Do voice recordings reveal whether a person is intoxicated? A case study. *Phonetica*, *47*(3-4), 215-237.

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

Jusczyk, P.W., & Luce, P.A. (2002). Speech perception and spoken word recognition: Past and present. Ear and Hearing, 23, 2–40.

Kapatsinski, V. (2010). Frequency of use leads to automaticity of production: Evidence from repair in conversation. *Language and speech*, *53*(1), 71-105.

Kaplan, A. (2010). *Phonology Shaped by Phonetics: The Case of Intervocalic Lenition*. Ph.D. dissertation, University of California, Santa Cruz.

Keating, P. A., Lindblom, B., Lubker, J., & Kreiman, J. (1994). Variability in jaw height for segments in English and Swedish VCVs. *Journal of Phonetics*, *22*(4), 407-422.

Kirchner, R. (1998). *An effort-based approach to consonant lenition*. Los Angeles: University of California dissertation.

Kirchner, R. M. (2001). *An effort based approach to consonant lenition*. Psychology Press.

Klich, R. J., Ireland, J. V., & Weidner, W. E. (1979). Articulatory and phonological aspects of consonant substitutions in apraxia of speech. *Cortex*, *15*(3), 451-470.

Kohler, K. J. (1991). The phonetics/phonology issue in the study of articulatory reduction. *Phonetica*, *48*(2-4), 180-192.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, R.I: Brown University Press.

Kuehn, D. P., & Moll, K. L. (1976). A cineradiographic study of VC and CV articulatory velocities. *Journal of Phonetics*, *3*, 303-320.

Ladefoged, P. (1971). *Preliminaries to linguistic phonetics*. Chicago: Chicago University Press.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Cambridge, MA: Blackwell.

Lass, R. (1984). *Phonology: an introduction to basic concepts*. Cambridge: Cambridge University Press.

Lass, R., & Anderson, J. M. (1975). *Old English Phonology*. Cambridge: Cambridge University Press.

Lavoie, L. M. (2001). *Consonant strength: Phonological patterns and phonetic manifestations*. Psychology Press.

Lester, L. & Skousen, R. (1974). The Phonology of Drunkeness. In A. Bruck, R. A. Fox, & M. W. LaGaly (Eds.), *Papers from the Parasession on Natural Phonology*. Chicago: Chicago Linguistic Society.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In *Speech production and speech modelling* (pp. 403-439). Springer: Netherlands.

Lindblom, B. E., & Maddieson, I. (1988). Phonetic universals in consonant systems. In L. Hyman, & C. N. Li (Eds.), *Language, Speech, and Mind*. London: Routledge.

Lindblom, B., & Sundberg, J. E. (1971). Acoustical consequences of lip, tongue, jaw, and larynx movement. *The Journal of the Acoustical Society of America*, *50*(4B), 1166-1179.

Locke, J. L. (1972). Ease of articulation. *Journal of Speech, Language, and Hearing Research*, *15*(1), 194-200.

Locke, J.L. (1980). Mechanisms of phonological development in children: Maintenance, learning, and loss. In *Papers from the sixteenth regional meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society.

Luo, C. R., Johnson, R. A., & Gallo, D. A. (1998). Automatic activation of phonological information in reading: Evidence from the semantic relatedness decision task. *Memory & cognition*, *26*(4), 833-843.

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.

Maddieson, I. (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press.

Mannell, R. (2008). Coarticulation and Assimilation. Macquarie University, available here: http://clas. mq. edu. au/phonetics/phonetics/coarticulation/index.html.

Marotta, G. (2008). Lenition in Tuscan Italian (gorgia toscana). *Lenition and fortition*, 235-270.

Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human perception and performance*, *15*(3), 576.

Maybery, M. T., Parmentier, F. B., & Jones, D. M. (2002). Grouping of list items reflected in the timing of recall: Implications for models of serial verbal memory. *Journal of Memory and Language*, *47*(3), 360-385.

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.

Mooshammer, C.; Hoole, P. & Geumann, A. (2007). Jaw and order. *Language and Speech* 50, 145-176.

Munson, B., & Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of speech, language, and hearing research*, *47*(5), 1048-1058.

Murdock Jr, B. B. (1962). The serial position effect of free recall. *Journal of experimental psychology*, *64*(5), 482.

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*(3), 594.

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

Ohala, J. (1997) The relation between phonetics and phonology. In W. Hardcastle, & J. Laver (Eds.) *Handbook of Phonetic Sciences*. Oxford: Blackwell.

Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In *The production of speech* (pp. 189-216). Springer: New York.

Oller, D. K., Wieman, L. A., Doyle, W. J., & Ross, C. (1976). Infant babbling and speech. *Journal of Child Language*, *3*(01), 1-11.

Oppenheim, G. M. (2012). The case for subphonemic attenuation in inner speech: Comment on Corley, Brocklehurst, and Moat (2011). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38(3):* 502–512.

Oppenheim, G. M., & Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition*, *106*(1), 528-537.

Oppenheim, G. M., & Dell, G. S. (2010). Motor movement matters: The flexible abstractness of inner speech. *Memory & cognition*, *38*(8), 1147-1160.

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

Parmentier, F. B., Elford, G., & Maybery, M. (2005). Transitional information in spatial serial memory: path characteristics affect recall performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 412.

Parnell, M., & Amerman, J. D. (1977). Subjective evaluation of articulatory effort. *Journal of Speech, Language, and Hearing Research*, *20*(4), 644-652.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision*, *10*(4), 437-442.

Peng, D. L., Ding, G. S., Perry, C., Xu, D., Jin, Z., Luo, Q., & Deng, Y. (2004). fMRI evidence for the automatic phonological activation of briefly presented words. *Cognitive Brain Research*, *20*(2), 156-164.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee, & P. Hopper (Eds.), *Frequency effects and the emergence of linguistic structure*. Amsterdam: Benjamins.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and speech*, *46*(2-3), 115-154.

Ploch, S. (2003). Metatheoretical problems in phonology with Occam's Razor and non-ad-hoc-ness. *Living on the Edge*, *28*, 149-201.

Port, R. F., & Dalby, J. (1982). Consonant/vowel ratio as a cue for voicing in English. *Perception & Psychophysics*, *32*(2), 141-152.

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

Poulisse, N. (1990). The Use of Compensatory Strategies by Dutch Learners of English, Dordrecht: Foris Publications.

Raymond, W. D., & Brown, E. L. (2012). Are effects of word frequency effects of context of use? An analysis of initial fricative reduction in Spanish. *Frequency effects in language learning and processing*, *1*, 35-52.

Robb, M. P., & Bleile, K. M. (1994). Consonant inventories of young children from 8 to 25 months. *Clinical linguistics & phonetics*, *8*(4), 295-320.

Roodenrys, S. (2009). Explaining phonological neighbourhood effects in short-term memory. In A. Thorn, & M. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain*. UK: Psychology Press.

Roodenrys, S., & Hinton, M. (2002). Sublexical or lexical effects on serial recall of nonwords? *Journal of Experimental Psychology: Learning, Memory and Cognition*, 28 (1), 29.

Roodenrys, S., Hulme, C., Lethbridge, A., Hinton, M., & Nimmo, L. M. (2002). Word-frequency and phonological-neighbourhood effects on verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(6), 1019.

Sands, B., & Güldemann, T. (2009). 11 What click languages can and can't tell us about language origins. In R. Botha, & C. Knight (Eds.), *The cradle of language*. New York: Oxford University Press.

Schvaneveldt, R. W., & Meyer, D. E. (1973). Retrieval and comparison processes in semantic memory. *Attention and performance IV*, 395-409.

Scott, M. (2013). Corollary discharge provides the sensory content of inner speech. *Psychological science* (in press).

Ségéral, P., & Scheer, T. (2008). Positional factors in lenition and fortition. *Lenition and fortition*, 131-172.

Service, E. (1998). The effect of word length on immediate serial recall depends on phonological complexity, not articulatory duration. *The Quarterly Journal of Experimental Psychology: Section A*, *51*(2), 283-304.

Shapiro L. (2011). Embodied Cognition. New York: Routledge Press.

Shariatmadari, D. (2006). Sounds difficult? Why phonological theory needs' ease of articulation'. *School of Oriental and African Studies Working Papers in Linguistics*, *14*, 207-226.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial-ordering mechanism in sentence production. In *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 295-342).

Starkweather, C.W. (2014). Speech fluency and its development in children. In Norman Lass (ed) *Speech and language: Advances in basic research and practice*.

Stevens, K. N. (1971). Airflow and turbulence noise for fricative and stop consonants: Static considerations. *The Journal of the Acoustical Society of America*, *50*(4B), 1180-1192.

Stoel-Gammon, C. (1988). Prelinguistic Vocalizations of Hearing-Impaired and Normally Hearing Subjects. A Comparison of Consonantal Inventories. *Journal of Speech and Hearing Disorders*, *53*(3), 302-315.

Stuart, G., & Hulme, C. (2000). The effects of word co-occurance on short-term memory: Associative links in long-term memory affect short-term memory performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(3), 796.

Templin, M. C. (1953). Norms on a screening test of articulation for ages three through eight. *Journal of Speech and Hearing Disorders*, *18*(4), 323-331.

Thomson, J. M., Richardson, U., & Goswami, U. (2005). Phonological similarity neighborhoods and children's short-term memory: Typical development and dyslexia. *Memory & Cognition*, *33*(7), 1210-1219.

Tian, X., & Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. Frontiers in Psychology, 1(166).

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: evidence from self-paced reading and sentence recall tasks. Language Learning, 61(2), 569-613.

Uther, M., Knoll, M. A., & Burnham, D. (2007). Do you speak E-NG-LI-SH? A comparison of foreigner-and infant-directed speech. Speech Communication, 49(1), 2-7.

Vallar, G., & Baddeley, A. D. (1984). Phonological short-term store, phonological processing and sentence comprehension: A neuropsychological case study. Cognitive Neuropsychology, 1, 121–141.

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(5), 1256.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, *167*(3917), 392-393.

Wedel, A. B. (2007). Feedback and regularity in the lexicon. *Phonology*, *24*(01), 147-185.

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: initial capabilities and developmental change. *Developmental psychology*, *24*(5), 672.

Westbury, J. R., & Keating, P. A. (1986). On the naturalness of stop consonant voicing. *Journal of linguistics*, *22*(01), 145-166.

Wheeldon, L.R., (2000). *Aspects o f Language Production,* Psychology Press; Hove, UK.

Willerman, R. (1991). The phonetics of pronouns: A typological perspective. Manuscript: University of Texas.

Willerman, R., & Lindblom, B. (1991). The phonetics of pronouns. In *PERILUS XIII: Papers from the Fifth National Phonetics Conference* (Vol. 13, pp. 19-23).

Wilson, A. D., & Golonka, S. (2013). Embodied cognition is not what you think it is. *Frontiers in psychology*, *4,* 58.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic bulletin & review*, *9*(4), 625-636.

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*(1), 48-65.

Wright, R. (2004). A review of perceptual cues and cue robustness. *Phonetically based phonology*, 34-57.

Young, M. A. (1981). Articulation effort: Transitivity and observer agreement. *Journal of Speech and Hearing Research*, *24*, 224–232.

Zygis, M., Fuchs, S., & Koenig, L. (2012). Phonetic explanations for the infrequency of voiced sibilant affricates across languages. Laboratory Phonology, 3(2), 299–336.