The relationship between language production and verbal short-term memory: The role of stress grouping

MORGAN, Jane, EDWARDS, Stephanie and WHEELDON, Linda R.

Published version

MORGAN, Jane, EDWARDS, Stephanie and WHEELDON, Linda R. (2014). The relationship between language production and verbal short-term memory: The role of stress grouping. The Quarterly Journal of Experimental Psychology, 67 (2), 220-246.

The relationship between language production and verbal STM: The role of stress grouping

J. L. Morgan

Sheffield Hallam University, U.K.

S. Edwards and L. R. Wheeldon

University of Birmingham, U.K.

Short Title: Stress Grouping and Verbal Short Term Memory

Word count: 12, 809

Corresponding Author:

Jane Morgan,

Department of Psychology, Sociology and Politics,

Sheffield Hallam University,

Collegiate Crescent Campus,

Sheffield, S10 2LD, UK

e-mail: j.l.morgan@shu.ac.uk

tel: +44 (0) 114 225 5587

fax number: + 44 (0) 114 225 2430

Abstract

This study investigates the influence of stress grouping on verbal short-term memory (STM). English speakers show a preference to combine syllables into trochaic groups, both lexically and in continuous speech. In two serial recall experiments, auditory lists of nonsense syllables were presented with either trochaic (STRONG-weak) or iambic (weak-STRONG) stress patterns, or in monotone. The acoustic correlates that carry stress were also manipulated in order to examine the relationship between input and output processes during recall. In Experiment 1, stressed and unstressed syllables differed in intensity and pitch but were matched for spoken duration. Significantly more syllables were recalled in the trochaic stress pattern condition than in the iambic and monotone conditions which did not differ. In Experiment 2, spoken duration and pitch were manipulated but intensity was held constant. No effects of stress grouping were observed, suggesting that intensity is a critical acoustic factor for trochaic grouping. Acoustic analyses demonstrated that speech output was not identical to the auditory input, but that participants generated correct stress patterns by manipulating acoustic correlates in the same way in both experiments. These data challenge the idea of a language-independent STM store and support the notion of separable phonological input and output processes.

Key words: Verbal short-term memory, stress grouping, speech perception, speech production, acoustic correlates of stress

Intuitively it makes sense that there might be some degree of overlap between the processes that underlie speech processing and the temporary storage of linguistic information in verbal short-term memory (STM). For example, during normal conversation we often need to store sections of the speech code we have planned while waiting for a chance to articulate it (e.g., Levelt, 1989). Similarly, holding onto verbal material in STM, such as a phone number or a shopping list, requires the generation and maintenance of a sound-based representation (Baddeley & Hitch, 1974; Baddeley, Thomson & Buchanan, 1975; Burgess & Hitch, 1999). Such similarities have led to the proposal that the verbal STM and the language system might be more closely related than has been traditionally assumed and could even share architecture (Acheson & MacDonald, 2009a and b; Ellis, 1980; Howard & Nickels, 2005; Jacquemot & Scott, 2006; Martin, Lesch & Bartha, 1999; Page, Madge, Cumming & Norris, 2007). In addition, a link has been suggested between STM and long-term stored linguistic knowledge. A number of studies have demonstrated that properties of stored lexical representations can support STM performance (e.g., Gathercole, Pickering, Hall & Peaker, 2001; Roodenrys, Hulme, Alban, Ellis & Brown, 1994; Walker & Hulme, 1999). Furthermore, more general properties of language (e.g., phonotactic frequency) have also been shown to benefit short-term recall (Gathercole, Frankish, Pickering & Peaker, 1999; Thorn, Gathercole & Frankish, 2005).

The current study sought to investigate the relationship between the language system and verbal STM by focussing on the linguistic property of prosody. Prosodic structure has properties that make it an ideal tool to investigate STM processes and how they relate to language. Prosody is an inherent part of the phonological representation of language, which groups stressed and unstressed syllables and words into larger rhythmic units in a language-dependent way (e.g., Inkelas & Zec, 1990; Nespor & Vogel, 1986; Selkirk, 1984, 1995; Wheeldon & Lahiri, 1997, 2003; see Wheeldon 2000 for a review). Furthermore in relation to memory, stress grouping has been shown to improve serial recall in auditory STM tasks (Adams, 1915; Boucher, 2006; Frankish, 1989; Reeves, Schmauder, & Morris, 2000). We, therefore, investigated the effect of stored linguistic knowledge on STM by testing whether the stress grouping benefit to STM is influenced by language-specific preferences in prosodic structure.

In addition, prosodic structure provides us with a tool to investigate the relationship between input and output processes in verbal STM. Stress is a perceptual property of syllables that can be carried by variations in a number of acoustic features, namely intensity, duration and pitch. In the studies we report, we manipulated the acoustic correlates of stress in the auditory input and performed detailed acoustic analyses of participants' speech output during recall. Our aim was to test the extent to which speakers' output was determined by the acoustic features of the input or by their own production preferences. Such fine-grained analyses of participants' responses have typically been the concern of phoneticians but are not commonly undertaken in the memory literature and to our knowledge this study is the first to use such a method to investigate the relationship between memory performance and phonological encoding processes in speech production.

Language processes and verbal STM

Traditionally the systems which underlie memory and language have been modelled as distinct architectures. Influential models of verbal STM have conceptualised information as being represented in phonological form and maintained by means of sub-vocal rehearsal (Baddeley & Hitch, 1974; Burgess & Hitch, 1999; Baddeley, Thomson & Buchanan, 1975). Specifically, the working memory model proposes the phonological loop; a temporary store whose contents decay rapidly unless they are refreshed by rehearsal (a process termed articulatory control, Baddeley, 1986). According to this account the speech production architecture is used to output material stored in the phonological loop but a language-independent mechanism is responsible for its maintenance.

The need for a specialised, language-independent STM device has, however, been questioned. Instead, it has been suggested that there is a close, if not overlapping relationship between verbal STM and speech production processes. In particular separate phonological input and output buffers have been proposed that are closely related to language processing. Evidence to support this view comes predominantly from patient data, which have demonstrated dissociations between performance on tasks which involve phonological input and tasks which involve phonological output (e.g., Howard & Nickels, 2005; Martin, Lesch & Bartha, 1999). Furthermore, Jacquemot and Scott (2006) propose that the input buffer is located within the speech perception architecture and is responsible for the encoding of the incoming speech stream into the phonological representations which access the mental lexicon. In contrast, the output buffer is concerned with the construction of the phonological representations which are necessary for speech production. Within this framework verbal STM is the product of information being transferred between the two buffers during rehearsal and speech monitoring processes (e.g., Morgan & Wheeldon 2003; Wheeldon & Levelt 1995; Wheeldon & Morgan, 2002). Finally, it has also been claimed that the overlap between language and memory systems renders STM stores unnecessary. In particular it has been suggested that the phonological effects which have been typically ascribed to STM can instead be attributed to the activation of long-term lexical representations and processes within the language production architecture (see Acheson & Macdonald, 2009a).

Taken together the theoretical positions outlined above suggest that the language system can influence both the encoding and the output of the to-be-remembered verbal code in the following ways. First, stored long-term knowledge can be recruited to support the encoding of items in verbal STM. Second, generating output in STM tasks might not simply involve the reproduction of a stored representation but rather makes use of the encoding processes which are relied upon in spontaneous language production. We discuss the evidence for each of these claims in turn.

Stored linguistic knowledge and verbal STM

An increasing number of studies have demonstrated beneficial effects of stored lexical information on verbal STM performance. For example, it has been shown that the lexicality of the to-be-remembered items benefits retention in that immediate recall for words is better than for nonwords (Gathercole et al., 2001). Furthermore recall is improved for nonwords which have structures similar to words compared to less word-like nonwords (Gathercole, 1995). Other studies have demonstrated effects of lexical variables such as frequency, age of acquisition and concreteness on recall (e.g., Gathercole, Pickering, Hall & Peaker, 2001; Romani, McAlpine & Martin, 2008; Roodenrys, Hulme, Alban, Ellis & Brown, 1994; Walker & Hulme, 1999).

An important question concerns the nature of the stored information which can influence verbal STM performance. In particular, it has been suggested that verbal STM performance might not only be supported by features of lexical representations but also by generic properties of the language. Evidence which supports this view comes from studies which measure the short-term retention of nonwords. Similar to the findings with real words, the recall of nonwords has been shown to be influenced by specific lexical representations. For instance, superior recall has been found for nonwords with larger compared to smaller lexical neighbourhoods (Roodenrys & Hinton, 2002; Thorn & Frankish, 2005). Nonword recall has, however, also been shown to be influenced by more general emergent properties of lexicon, for example, the statistical frequency of certain phoneme combinations in the language (Gathercole, Frankish, Pickering & Peaker, 1999 with children; Thorn, Gathercole

& Frankish, 2005 with adults). Such findings have led Thorn, Gathercole and Frankish (2005) to conclude that long-term influences on STM performance could be mediated by more than one mechanism: one where specific lexical representations are activated; and another where more general language-specific preferences are drawn upon. One aim of our study, therefore, was to test this claim by investigating whether another general property of the language (i.e., language-specific preferences in prosodic structure) would similarly influence verbal STM.

Language production and verbal STM

In the speech production literature there is broad agreement that once a lexical item has been selected to convey the concept to be communicated, its phonological form is retrieved and encoded for output. This process of phonological encoding involves the generation of an abstract representation of sound form which is determined by prosodic structure (Levelt, 1989, 1992; Levelt, Roelofs & Meyer, 1999; Roelofs, 1997, 2002). A prosody generator is proposed which takes as its input information about the syllable structure and stress pattern of words and combines them into phonological words ($_{\omega}$) (i.e., a single unit with one main stress). For example, the sentence "give it to her" could be articulated as a single phonological word, [*giv*-i-ter]$_{\omega}$ in which several new syllables are constructed and stored until an entire phonological word is complete (Lahiri & Wheeldon, 2011; Wheeldon & Lahiri, 1997, 2003). The resulting ordered speech plans are further encoded according to their phonetic realisations for articulation (Cholin, Levelt & Schiller, 2006; Cholin, Schiller & Levelt, 2004; Levelt & Wheeldon, 1994). Prosodic structure also impacts on articulatory planning as metrical regularity has been shown to improve speech rate and accuracy in articulation (Tilsen, 2011a&b). The generation of prosodic structure is, therefore, central to current models of spoken language production.

The claim that speech production and verbal STM share architecture has been supported by data from the serial recall task. In this task participants hear lists of stimuli (e.g., letters, digits, words, nonsense words) and are required to repeat back the list immediately in the same order in which it was presented. A number of models have focused on the way in which serial order is maintained in both verbal STM (e.g., Brown, Preece & Hulme, 2000; Burgess & Hitch, 1999) and speech production (Vousden, Brown & Harley, 2000). It has been proposed that a timing/learning context signal is constructed from a set of internal oscillators which fluctuate to represent positional information both within a list and within a group (Brown et al., 2000; Burgess & Hitch, 1999). A similar framework has been applied to the maintenance of serial order during the process of phonological encoding with the formation of a phonological-context signal which represents the metrical frame which is generated for each fragment of speech (Vousden, Brown & Harley, 2000).

Importantly, for the purpose of the current study, all models of serial recall propose an explicit production component in the form of the generation and maintenance of a phonological representation to enable speech. A further aim of our study was, therefore, to establish whether the output generated in a serial recall task resembles the phonological representation which is constructed during normal speech production processes.

There is evidence to suggest that this might be the case. Ellis (1980) investigated the relationship between the speech errors which participants made in serial recall tasks and those observed in spontaneous speech production. Spontaneous speech errors have been the focus of a great deal of language production research (e.g., Dell, 1986; Fromkin, 1980; Garrett, 1980; see Meyer, 1992 for a review) and Ellis demonstrated that their main characteristics are mirrored in STM tasks. For example, exchanges which involve phonemes are more frequent than those which involve whole syllables. Exchanges between phonemes are more likely to occur when they possess similar distinctive features (feature similarity effect) and when the syllables in the lists contain the same vowel (context similarity effect). Evidence of syllable position constraints were also observed especially in the case of syllable onsets. Ellis explained this “error equivalence” by attributing them to the same locus, namely the response buffer which is used in both the recall of verbal material and in the storage of speech programmes prior to articulation.

A similar conclusion was reached by Page, Madge, Cumming and Norris (2007). They focused on the phonological similarity effect which typifies immediate serial recall such that errors are more likely to occur in lists in which the items share phonology compared to those that are dissimilar. Performance was compared on a standard serial recall task to that on a speeded reading task (which was used to eliminate the memory component of the task). On both tasks sub-lexical phonological similarity errors were observed which mirror those observed in spontaneous speech production. Based on this finding the authors concluded that a common mechanism must underlie both memory and speech production and that the phonological store is in fact the ordered speech plan that is used to generate fluent, connected speech.

Finally, Acheson and MacDonald (2009b) performed a detailed analysis of the type and distribution of speech errors produced using a serial recall task compared again to a speeded reading task. Three serial recall tasks were run which involved tongue twisters constructed using nonwords and each requiring a different output response (i.e., a spoken or typed response or a recognition task). Irrespective of the mode of response the pattern of errors observed mirrored the types of errors found in spontaneous production (e.g., sub-lexical errors, contextual substitutions and syllable position constraints). Based on these findings the authors concluded that such errors occur not because of properties of verbal STM but rather as a result of the recruitment of the representations and processes of normal language production. Further recent support for this proposition comes from neuropsychological evidence which suggests that the same brain structures are responsible for both phonological encoding and verbal STM processes (Acheson, Hamidi, Binder & Postle, 2011).

Prosodic structure and verbal STM

The current study focuses on the linguistic property of prosody, which refers to the grouping of syllables and words based on rhythmic principles (Selkirk, 1984, 1995; Nespor & Vogel, 1986). There is a great deal of evidence that English, like the other Germanic languages, has a trochaic grouping preference (STRONG-weak, e.g., Hayes, 1982). Cutler & Carter (1987) examined the MRC psycholinguistic database (Coltheart, 1981) and found that 73% of words had primary or secondary stress on the initial syllable. When a natural speech sample was analysed this increased to as many as 90% of lexical items having a strong initial syllable (Svartvik & Quirk, 1980). The trochaic stress grouping is also prevalent in phrasal groupings in Germanic languages, where there is evidence that a strong stress attracts upcoming unstressed elements. This preference for leftward attachment can be seen in the development of grammaticalised suffixes (e.g., Lahiri & Wheeldon 2011). Trochaic grouping has also been shown to determine the time it takes to produce spoken phrases in both Dutch and English. Wheeldon and Lahiri (1997) demonstrated that during the generation of phonological words, unstressed determiners cliticise leftwards in defiance of syntactic phrase boundaries (e.g., $[\text{Drink –de}]_{\omega}$ $[\text{wijn}]_{\omega}$ / $[\text{Drink –the}]_{\omega}$ $[\text{wine}]_{\omega}$). They showed that these phonological word groupings determined sentence onset latencies in both prepared and on-line sentence production tasks (see also, Lahiri & Wheeldon, 2011; Malpass, Lahiri, Wheeldon, 2010).

The predominance of the trochaic stress pattern in English has been shown to influence both lexical access and language acquisition. For example, it has been shown that when segmenting speech listeners exploit the most typical stress pattern of words and phrases in their language (e.g., Cutler, 1994; Cutler & Norris, 1988; Tyler & Cutler, 2009). Moreover, there is developmental evidence to suggest that infants are sensitive to the prosodic structure of their native language from a very early age. By 9 months English infants are able to make use of the trochaic stress pattern to segment an artificial language (Thiessen & Saffran, 2003). In addition based on syllable omission patterns it has been suggested that children as young as two years use trochaic metrical templates when planning their speech (Gerken, 1994).

Prosody is, therefore, an inherent part of the phonological representation of language. Moreover, evidence exists which suggests that prosodic grouping can also improve serial recall performance in auditory STM tasks. For example, recall is improved when pauses are inserted between groups of items in a to-be-remembered list (Frankish, 1989; Ryan, 1969a & b; Wickelgren, 1967). Moreover, boundary effects for grouped stimuli are observed in auditory serial recall, which are not seen when items are presented visually (Cowan, Saults, Elliot & Moreno, 2002; Frankish, 1985; 1989; Frick, 1989). Such findings suggest that auditory memory can exploit structural features of the list which may reflect prosodic units of the language (Frankish, 1989). Prosodic structures in language, however, are not normally delimited by pauses. Rather, stress patterns predominantly contribute to rhythmic grouping in speech. Again there is evidence to suggest that verbal STM can be influenced by stress grouping. Boucher (2006) found that stress groupings of 2-4 syllables in size (i.e., groups which reflect prosodic units) improved serial recall whereas groupings which were any larger in size were detrimental to performance. Enhanced recall has also been shown when items are grouped using other prosodic features such as intonation and stress (Frankish, 1995). Finally, Reeves, Schmauder and Morris (2000) found that the addition of a stress pattern (either anapest or dactyl) to the auditory presentation of the lists improved recall compared to the monotone lists.

The current study

The current study was designed to test two hypotheses. First that stored linguistic knowledge (i.e., the language's prosody) can be recruited to support the encoding of items in STM. Based on the evidence described above we would expect stress grouping to benefit verbal STM performance. What has not been demonstrated before, however, is whether the English language's preference for trochaic groupings increases retention compared to iambic groupings. Our research aims to directly test this hypothesis in a pair of serial recall experiments by manipulating the perceptual salience of the stress pattern of the to-be-remembered items to favour either trochaic or iambic groupings.

Second, that the output generated in serial recall tasks reflects the encoding processes which are relied upon in spontaneous language production rather that the acoustic properties of the auditory input. In order to explore the relationship between verbal input and output, the acoustic correlates of stress were manipulated across the two experiments to allow us to directly compare participants' production of stressed and unstressed syllables with the manipulated acoustic properties of the auditory lists which were heard.

In the two serial recall experiments which we report below, we presented lists of to-be-remembered items which were made up of alternating stressed and unstressed syllables (following a trochaic or an iambic pattern) or monotone syllables (all stressed or all unstressed). Imposing a stress pattern on the syllable lists contributes to the likelihood that the lists will be grouped given that stress is a perceptual characteristic of syllables that contributes both to the perception and production of rhythm in speech. For example, perceptual centres or *p-centres* (Morton, Marcus & Frankish, 1976) are located in stressed syllables and, although not signalled by any one acoustic feature, seem to be aligned to stressed vowel onsets. In particular it has been demonstrated that p-centres play an important role in the perceived timing of syllable sequences, such that the perception and production of an even rhythm is based on p-centre alignment (e.g., Barbosa, Arantes, Meireles & Vieira, 2005; Fowler, 1979; Marcus, 1981). Moreover timing regularities (such as inter-stimulus intervals) have been shown to improve speech perception as such regularities allow listeners to predict where important segmentation information is likely to occur (Quené & Port, 2005). However, the relationship between perceptual stress and acoustic factors is not simple. Syllable stress can be realised by variations in a number of acoustic factors, either alone or in combination: these are pitch (hz), intensity (dBs) and duration (ms). There is evidence that these factors differ in their effects on the perception of grouping. According to the iambic/trochaic law (Hayes, 1995; but see Revithiadou, 2004) items which differ in terms of *intensity* form groups with initial prominence (i.e., a trochaic stress pattern), whereas items which differ in terms of *duration* form groups with final prominence (i.e., an iambic stress pattern). There is behavioural evidence to support this assertion. In particular it has been demonstrated that both speech and non-speech sequences alternating in intensity are perceived as trochaic where as sequences alternating in duration are perceived as iambic (Hay & Diehl, 2007). We, therefore, manipulated these factors independently across the two experiments in order to test their effect on both grouping and on the participants' output. In Experiment 1, intensity and pitch were manipulated while spoken duration was held constant. In Experiment 2, spoken duration and pitch were varied while intensity was held constant. The aim of this manipulation was to test whether any observed grouping benefits were language-specific (as evidenced by better performance in the trochaic stress pattern condition in both experiments) or are attributable to a universal bias (as evidenced by better performance in the trochaic stress pattern condition in Experiment 1 and better performance in the iambic stress pattern condition in Experiment 2).

Finally, we also wished to test whether participants' output mirrored the acoustic nature of the input lists or whether speakers stressed their output using all the acoustic features they would in their normal speech production. Specifically we looked for evidence of characteristics which typify normal, error-free phonological encoding processes using speech measurements rather than the speech error analyses which have typically been used in previous research.

Experiment 1

Experiment 1 was designed to investigate the influence of prosodic grouping on the immediate serial recall of nonsense syllables. We compared, the retention of lists presented with either trochaic (STRONG-weak), iambic (weak-STRONG) or monotone stress patterns. In particular we were interested in whether the stress grouping of lists would benefit memory performance and whether any benefit observed would be dependent on the general preference in English for trochaic stress groups. We also manipulated the acoustic factors carrying stress such that intensity and pitch was varied across stressed and unstressed syllables but duration was held constant. This allowed us to ascertain whether participants' output matched the acoustic properties of the auditory input or was characterised by production preferences of spontaneous speech.

Method

Materials

A set of 24 syllables was constructed which comprised 12 CVC syllables and 12 VC syllables (see Appendix 1). In order to isolate the effects of acoustic cues to stress grouping, we explicitly designed our stimuli to reduce, as far as possible, other linguistic cues to stress. We chose either illegal or low frequency syllables of English in order to limit any possible effects of stored metrical structures on stress assignment. As far as possible, nonsense syllables or low frequency English syllables were used. All syllables had a frequency of 0 (calculated using CELEX counts per million; Baayen, Piepenbrock & Gulikers, 1995) with the exception of /rɛl/ (6), /ɛs/ (554), /ɔm/ (158), /æf/ (180), /ɛp/ (23) and /eɪn/ (2168). Moreover none of the syllables could be combined with others in the set to form an English word. Vowel reduction to shwa [ə] usually occurs in unstressed syllables in English (e.g., Clark & Yallop, 1995), and has been shown to be an important cue to stress (e.g., Cutler & Butterfield, 1992). Our interest, however, was in the role of prosodic rather than segmental cues to stress grouping. We, therefore, ensured that all syllables contained full vowels. Finally CVC and VC syllables sets were also matched for vowel weight with equal numbers of vowels containing two mora (i.e., diphthongs and long vowels - henceforth “long vowels”) and vowels containing one mora (henceforth “short vowels”) as long vowels are more likely than short vowels to attract stress.

Stressed and unstressed versions of each syllable were recorded by a female native speaker of British English with no obvious regional accent. All syllables were produced in isolation to prevent any co-articulation effects. Stressed and unstressed versions were produced consecutively to produce a clear contrast and several tokens of each syllable were recorded in different orders to allow the best exemplars to be identified. The spoken duration, intensity and pitch of the digitised recording of each syllable was measured using Praat software (Boersma & Weenink, 2009). The intensity and pitch measures were the averages computed across the entire syllable. As we were concerned with duration differences between tokens of syllables with the same segmental content to ensure consistency, spoken syllable duration was measured according to objective, segment-dependent criteria. For example all syllables beginning with voiceless stops were measured from the onset of aspiration.

A stressed and unstressed token of each nonsense syllable was then chosen to construct the serial recall lists according to the stress pattern condition. Stressed and unstressed syllables were selected such that the stressed version of each syllable was louder in intensity and higher in pitch than the unstressed version but that they were matched on spoken duration. The means of these measures for each syllable type are shown in Table 1. Across both syllable types the stressed and unstressed syllables differed significantly on average by 50.78Hz ($t(23) = 11.0$, $p < .001$) and 13.21dBs ($t(23) = 20.1$, $p < .001$). Stressed and unstressed syllables did not, however, differ in terms of spoken duration either of the whole syllable (mean duration difference of -11ms, $t < 1$, $p = .37$) or of the syllable constituents (onsets, nuclei and codas all ps>.05).

Table 1 about here

Serial recall lists. For the purposes of the design only the 24 nonsense syllables were arranged to make 72 CVC-VC pairs. Such structures were chosen to maximise the potential for resyllabification to occur during the production of the two stress pattern lists (i.e., trochaic and iambic). For example, if during recall, speakers grouped syllables into lexical or phonological word-like units according to the maximization of onset principle (e.g., Selkirk, 1980; Wheeldon & Lahiri, 1997) presenting a closed syllable followed by an onset-less syllable (e.g., CVC/VC) means that if resyllabification occurs the coda consonant of the first syllable may become the onset of the second syllable (e.g., CV-CVC) or become ambisyllabic (e.g., CV[C]VC). Such resyllabification in the participants' output would provide us with further evidence concerning the effects of stress grouping on output processes. Critically we were interested in whether our speakers would actively group and resyllabify temporally separated syllables based on stress patterns. Therefore, the presentation of each syllable in the input list was separated by a 100ms pause.

The 72 pairings constituted all possible combinations of the 24 syllables within these structural restrictions in order to reduce the possibility that speakers would be aware of the syllable pairings. Arranging the syllables in this way also meant that the alignment of p-centres remained constant over trochaic and iambic lists as the distance between each stressed syllable in a list was of equal length. This was not, however, the case for the monotone lists in which all syllables were either stressed or unstressed. In these lists, the alternation of closed and open syllables might also lead to a general tendency in our participants to group the six syllables into three syllable pairs due to closer p-centres within compared to across syllable pairs. We were, however, interested in whether stress grouping affected memory performance over and above any potential effects of grouping due to syllable structure.

Finally each syllable pair contained one long and one short vowel, which occurred equally often in each order across the 72 pairs. This was done to ensure that the stimuli were not biased towards either stress pattern. Stress tends to be assigned to long vowels, therefore, the pairs with initial long vowels would be more naturally produced with a trochaic stress pattern whereas the pairs with initial short vowels would be more naturally produced with a iambic stress pattern.

Three syllable pairs were used to make up each serial recall list and each of the 72 syllables pairs was used twice resulting in the formation of 48 lists. It is important to note that the syllable pairings were made for design purposes only and were not apparent to participants, as within each list all syllables were separated by 100ms of silence. Four versions of each of the 48 serial recall lists were constructed by concatenating each of the discrete digitised recordings of the stressed and unstressed syllable tokens. Each version corresponded to one of the four stress pattern conditions as follows (bold italics denote stress):

(1) Trochaic stress pattern e.g., ***hif*** abe ***sut*** eed ***rame*** af

(2) Iambic stress pattern e.g., hif ***abe*** sut ***eed*** rame ***af***

(3) Stressed monotone e.g., ***hif*** ***abe*** ***sut*** ***eed*** ***rame*** ***af***

(4) Unstressed monotone e.g., hif abe sut eed rame af

Experimental Design

The 48 serial recall lists were arranged into four blocks of 12 lists of the same stress pattern. Within a block each nonsense syllable occurred three times but each time within a different pairing and in a different list position. Each experimental block was preceded by three practice trials which conformed to the stress pattern of the subsequent block but which were constructed from a different pool of nonsense syllables. Across an experimental session every participant was presented with four blocks of serial recall lists, the order of which was rotated across participants, alternating between the stress pattern and monotone blocks.

Apparatus

The experiment was built and run using the experiment generator software, E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA). Participants were tested individually in a sound-attenuated booth. A tascam recorder was used to make a digitised recording of each participant’s speech output.

Procedure

In order to familiarise participants with the nonsense syllables prior to the experimental trials, they heard and were asked to repeat all of the syllables in isolation. Both unstressed and stressed versions of all 24 experimental syllables plus the 18 practice syllables were presented one at a time. Each syllable appeared twice, in a different random order for every participant. Participants were given two seconds from the onset of each syllable to repeat it back before the next syllable was presented.

The experimental session then began with the presentation of the four blocks of serial recall lists. Every block began with three practice lists. Each list was presented auditorily and commenced with a fixation cross that appeared on the screen in front of the participant for 250ms. The first list was presented 500ms after the fixation cross was displayed. At the offset of each list presentation there was a one second delay followed by a beep signal. Participants were instructed to listen to the list of syllables and to repeat back each list when they heard the beep. They were instructed to try to be as accurate as possible and to assign the same stress pattern to that which they had heard when the list had been presented. The experimenter controlled the presentation of the next list when the participant had finished giving their response. After each block of lists there was a participant timed break after which the instructions were repeated. All spoken responses were recorded for analysis. The whole procedure for each participant was completed within 45 minutes.

Participants

20 undergraduate students at the University of Birmingham acted as participants. They were native speakers of English and had no known hearing impairment. Participation was voluntary and individuals received course credit in return for their time.

Results

The data for each participant were transcribed by two listeners who were blind to the experimental conditions. Any differences between transcriptions were listened to and transcribed by a third listener. The number of syllables which participants attempted to produce for each list regardless of accuracy was recorded (range 0-6). Every trial was also coded for accuracy of stress production. Trochaic and iambic lists were coded as correct when they were produced with the correct stress pattern throughout the list. Stressed and unstressed lists were coded as correct when they were produced with an unvarying stress pattern. Although stress is relative, speakers did produce the stressed and unstressed lists with clear differences in overall prominence and it was, therefore, also possible to judge the unvarying list production on overall stress level. Participants failed to produce the correct stress pattern on only 4% of trials. These errors were almost entirely attributable to speakers' continuing use of the stress pattern required in a previous block, rather than to a failure to maintain the correct stress within a list. As we were concerned with the effect of correctly generated stress patterns on recall the syllable accuracy measures were taken from the correctly stressed lists only. For each condition we calculated the number of syllables each participant produced correctly. To count as correct, a syllable had to have all segments correctly produced in the right order, occur in the correct serial position, and be produced with the correct stress[1]. The error score was broken down into the overall number of correct syllables (range 0-6), the number of correct stressed syllables (range 0-3) and the number of correct unstressed syllables (range 0-3) for each list. Three participants were replaced due to high error rates either in terms of number of correct syllables produced or correct stress pattern applied to the lists. The mean percentage accuracy for the rated performance measures is shown in Table 2.

Table 2 about here

Analyses were conducted with both subjects (F_1) and lists (F_2) as random factors and on percentage accuracy arcsine transformed. As all participants responded in each grouping condition and all lists were generated in each grouping condition across participants, the effect of Stress Pattern was both within-subjects and within-items. As can be seen in Table 2, the percentage of syllables participants attempted to recall for each list was high. On most trials participants produced six syllables after each list presentation. No difference was observed between the percentage of syllables attempted in each of the Stress Pattern conditions, F_1 & $F_2 < 1$.

The percentage of trials on which participants produced the correct stress pattern was also high. A main effect of Stress Pattern was observed, F_1 (3, 57) = 4.4, *MSE*= 0.1, $p = .008$, F_2 (3, 141) = 11.3, *MSE* = 0.2, $p < .001$. Pairwise comparisons were performed and revealed no significant differences according to the subject analysis but the items analysis revealed differences between the unstressed condition and the trochaic condition ($p <.001$), the unstressed condition and the iambic condition ($p < .001$) and between the stressed condition and the iambic condition ($p = .005$). Therefore, in general, speakers were more likely to produce the ungrouped lists more accurately, particularly the unstressed lists.

We were critically interested, however, in how stress pattern would affect participants' memory for the nonsense syllables. On the whole, recall was quite poor with participants producing, on average, about two syllables (30%) correctly per list. This was probably due to the syllables used being unfamiliar to speakers of English. An ANOVA conducted on the percentage of correct syllables (arcsine transformed) produced in each of the four conditions yielded a significant effect, $F_1(3, 57) = 5.7$, MSE=.00, p=.002, $F_2(3, 141) = 5.8$, MSE=.01, p <.001. Bonferroni pairwise comparisons revealed that performance in the trochaic stress pattern condition was significantly better than in the iambic condition according to both the subject (p=.004) and items analyses (p<.001). Performance in the trochaic condition was also significantly better than in the stressed condition (p=.021 by subjects and p=.01 by items) and the unstressed condition (p=.017 by subjects and p=.04 by items). None of the other conditions differed significantly from one another. Therefore, recall was found to be the most accurate for the lists presented with the trochaic stress pattern.

Separate analyses were performed on the number of correct stressed and unstressed syllables recalled in the trochaic and iambic stress pattern conditions. In the trochaic lists stressed syllables fell in list positions 1, 3 and 5, whereas in the iambic lists stressed syllables fell in list positions 2, 4 and 6. Analyses revealed that more stressed syllables were recalled correctly in the trochaic condition compared to the iambic condition ($F_1(1, 19) = 28.1$, MSE = .01, p<.001, $F_2(1, 47) = 27.7$, MSE = .03, p<.001). Whereas when the number of correct unstressed syllables was compared no difference was found between the conditions, F_1& F_2< 1. Such a pattern of results suggests stressed syllables are better remembered when they form the onset of a grouping.

However, it must be noted that in the trochaic condition stressed syllables were not only the onset of a grouping but were also the onset of a list. It is possible, therefore, that the superior memory performance observed for stressed syllables in the trochaic condition is due to a primacy effect. If this were the case any trochaic benefit should be limited to the initial position in the list. In order to determine the effect of stress pattern across each list position we selected those trials on which participants produced a six-syllable response. A further three participants were removed from the positional analysis due to empty cells. The resulting percentage error rates across each of the six list positions in each condition are plotted in Figure 1.

Figure 1 about here

As can be seen in Figure 1, for all list positions (with the exception of the final list position) memory performance is the most accurate in the trochaic stress pattern condition suggesting that the benefit observed in this condition is not restricted to the first syllable in the list. An ANOVA was performed which compared recall accuracy in each condition across all six list positions. A main effect of List Position was observed, $F_1(5, 80) = 51.2$, $MSE = .08$, $p<.001$, $F_2(5, 235) = 43.4$, $MSE = .29$, $p<.001$. Bonferroni pairwise comparisons revealed that recall accuracy declined incrementally across the list until the final position where performance was found to improve. Specifically, performance on the first position in the list was more accurate than all the other list positions (all subjects and items *p*s<.001). Performance in second list position was more accurate than in third position by subjects only ($p=.022$ by subjects, $p=.21$ by items), and fourth and fifth positions (subjects and items *ps*< .001). Performance in third position was better than fourth position by subjects only ($p=.043$ by subjects, $p=.13$ by items) and the fifth position ($p<.001$ by subjects, $p=.02$ by items). Performance in positions four and five did not significantly differ ($p = 1.0$), however, performance in these list positions was significantly worse than in the final list position (position four, $p=.035$ by subjects and $p=.028$ by items; position five, $p=.005$ by subjects and $p=.007$ by items). Therefore, when recall accuracy is compared across all six positions we find evidence of the primacy and recency effects usually observed in serial recall tasks.

There was also a significant interaction between List Position and Stress Pattern by items that approached significance by subjects, $F_1(15, 240) = 1.7$, $MSE = .05$, $p = .056$, F_2(15,

705) = 2.0, *MSE* = .12, p =.012. Post hoc ANOVAs were conducted which explored the effect of Stress Pattern on recall in each of the list positions separately. For the first list position a main effect of Stress Pattern was observed, $F_1(3, 48) = 3.8$, *MSE* = .05, p=.017, $F_2(3, 141) = 3.1$, *MSE* = .13, p=.029. Subsequent Bonferroni pairwise comparisons failed to yield any significant differences between the four conditions, although the difference between the trochaic and iambic conditions approached significance (p=.075 by subjects, p =.061 by items). For positions 2 and 6 no main effects of Stress Pattern were observed (position 2 $F_1(3, 48) = 1.3$, *MSE*= .08, p=.28, $F_2(3, 141) < 1$; position 6, $F_1(3, 48) < 1$, $F_2(3, 141) = 1.8$, *MSE*= .13, p=.16). However, main effects of Stress Pattern were observed for position 3, $F_1(3, 48) = 5.5$, *MSE* = .07, p=.002, $F_2(3, 141) = 6.2$, *MSE* = .12, p <.001; position 4, $F_1(3, 48) = 3.1$, *MSE* =.05, p=.035, $F_2(3, 141) = 3.1$, *MSE* = .12, p=.03; and position 5, in the items analysis only, $F_1(3, 48) = 2.1$, *MSE* =.05, p=.11, $F_2(3, 141) = 4.3$, *MSE* = .09, p=.006. These effects of Stress Pattern were further explored using Bonferroni pairwise comparisons which revealed that performance was better in the trochaic condition compared to the iambic condition in list position 3 (p=.013 by subjects, p<.001 by items) and by items in list position 5 (p=.38 by subjects, p=.002 by items) and better than in the unstressed condition in list position 4 (p=.045 by subjects, p=.013 by items). This analysis performed on a subset of the data does suggest that the trochaic benefit is observed across the list (in particular list positions 1, 3, 4 and 5) and is not just a feature of the initial stressed syllable.

Discussion

The results of Experiment 1 show that participants had little difficulty in perceiving and reproducing the intended stress patterns. They produced six syllable responses on most trials and the vast majority of their responses were perceived by independent listeners to be correctly stressed. Our acoustic measures of participants' spoken responses (reported below) confirm that they did indeed reproduce the stress patterns they heard in the input.

Our main interest was in the effect of stress grouping on recall accuracy. Recall was more accurate for lists presented with the trochaic stress pattern than for those presented with an iambic stress pattern, which did not differ in accuracy from ungrouped lists. In addition, stressed syllables were better recalled in the trochaic than in the iambic stress pattern condition suggesting that stressed syllables are more salient when they form the onset of a grouping. There was no difference in recall for the unstressed syllables in the trochaic and iambic lists. Finally, an analysis of the influence of list position on recall confirmed that the observed benefit for the trochaic stress pattern was not restricted to the first position in the list. Taken together these findings suggest that trochaic but not iambic grouping benefits verbal STM performance compared to monotone lists.

As discussed in the Method section above, the to-be-remembered syllables were arranged into CVC-VC pairs to allow resyllabification to occur. However, we observed no evidence of resyllabification. Speakers did not run the syllables together to generate fluent resyllabified phonological words. Instead, speakers reproduced the syllables as discrete units. This may be due to the difficulty of the task, as syllable accuracy was low and resyllabification is an optional process which usually occurs in fluent connected speech. Nevertheless we observed a clear benefit for trochaic but not iambic stress on recall accuracy, suggesting that stress grouping did occur.

It is possible, however, that the significant difference between trochaic and iambic lists may have been due to perceptual effects of the acoustic cues we manipulated rather than a language-specific preference for trochaic grouping. In this experiment our stressed and unstressed syllables differed in terms of pitch and intensity but were matched in duration. As outlined in the Introduction, according to the iambic/trochaic law (Hayes, 1995), the specific acoustic properties of an auditory signal universally bias listeners towards the perception of certain groupings. That is, items which differ in terms of *intensity* form groups with initial prominence (i.e., a trochaic stress pattern). Items which differ in terms of *duration* form groups with final prominence (i.e., an iambic stress pattern). Therefore, the acoustic profile of our stimuli favoured the trochaic over the iambic grouping. As a consequence our participants may have grouped both stress pattern lists into trochees from the first stressed syllable leading to three groups in the trochaic lists (e.g., [σ́σ][σ́σ][σ́σ]) compared to four groups in the iambic lists (e.g., [σ][σ́σ][σ́σ][σ́]). In this case better performance on the trochaic lists would be due to the number of groupings based on a universal perceptual grouping rather than a linguistic grouping preference. However, rhythmic judgements have been shown to be strongly influenced by the first syllable heard in the sequence (see Hay & Diehl, 2007; Woodrow, 1909). In particular in Hay & Diehl's (2007) perceptual grouping experiments, the syllable sequences were presented with a long fade-in and fade-out precisely because list onset is an extremely strong cue to group onset. Nevertheless it is still possible that performance in the trochaic lists was better than in the iambic lists because the acoustic stress cues were better correlated with the trochaic grouping.

Experiment 2 was, therefore, designed to replicate Experiment 1 however the nonsense syllables were now matched on intensity but varied in both pitch and duration. If variation in duration creates a universal bias towards an iambic grouping then we should now see a benefit for the iambically grouped syllable lists in STM performance. However if our original finding was due to a linguistic preference for the trochaic stress pattern then we should again observe a trochaic but no iambic benefit in this experiment.

Finally as with Experiment 1, we are also interested in the degree to which participants' speech output is determined by the acoustic information in the input. A comparison of the speech output in Experiments 1 and 2 will also provide us with a test of the relationship between acoustic input and speech output during serial recall.

Experiment 2

Method

Stress manipulation

Experiment 2 was identical in design to Experiment 1, except that syllable intensity was held constant across stressed and unstressed syllables while pitch and duration varied. A range of stressed and unstressed tokens were again produced by a native female speaker of British English and the digitised recordings were measured for the acoustic correlates of stress: spoken duration, intensity and pitch. Stressed and unstressed syllables were selected such that they differed in spoken duration and pitch characteristics but were matched on intensity. It was ensured that differences were in the right direction for all pairs (i.e., the stressed version of the syllable was longer in duration and higher in pitch than the unstressed version. The means of these measures for each syllable type are shown in Table 3.

Table 3 about here

Across both syllable types the stressed and unstressed syllables differed significantly on average by 68.7 Hz ($t(23) = 17.9$, $p<.001$) and 182 ms in duration ($t(23) = 20.7$, $p<.001$). The duration increase in stressed syllables was mainly carried by the nuclei (143 ms, $t(23) =$

12.2, $p<.001$) although the codas also showed a small but significant increase (32ms, $t(23) = 2.8$, $p<.001$). The duration of onsets in stressed and unstressed syllables did not differ (8ms, $t(11) = 1.1$, $p=.30$). In terms of intensity, stressed and unstressed syllables differed by only 1.9dB.

The block construction was the same as in Experiment 1. The procedure and experimental instructions were also identical. A further 20 participants took part, from the same subject pool.

Results

Participants' responses were transcribed and coded in the same way as for Experiment 1. Two participants were replaced due to high error rates (loss of more than 25% of data). Subject means for the rated performance measures (percent correct) are given in Table 4.

Table 4 about here

As with Experiment 1, all analyses were performed on percent correct measures arcsine transformed. Participants again produced on average 5 to 6 syllables on each trial and there was once again no significant difference between the mean number of syllables attempted in each of the Stress Pattern conditions (F_1 & $F_2<1$). The percentage of trials on which participants produced the correct stress pattern was again very high across all conditions and an ANOVA yielded a non-significant main effect of Stress Pattern, $F_1(3, 57) = 1.3$, $MSE=.02$, $p=.28$, $F_2(3, 141) = 2.2$, $MSE=.02$, $p=.091$.

In contrast to Experiment 1, however, the number of nonsense syllables correctly recalled did not differ across the conditions and the effect of Stress Pattern on correct recall was not significant, $F_1(3, 57) = 1.4$, $MSE=.01$, $p=.26$, $F_2(3, 141) = 2.2$, $MSE=.01$, $p=.09$. Furthermore, the number of correct stressed and unstressed syllables recalled in the trochaic and iambic stress pattern conditions did not differ greatly. An analysis of the number of correct stressed syllables recalled in the trochaic and iambic stress pattern conditions showed no significant effect (F_1 & $F_2 < 1$). Similarly, Stress Pattern did not significantly affect the number of unstressed syllables correctly recalled, $F_1(1, 19) = 2.2$, $MSE = .01$, $p=.16$, $F_2(1, 47) = 2.2$, $MSE = .03$, $p=.15$.

As in Experiment 1, recall accuracy across the six list positions was examined. Figure 2 shows the mean percentage error rate plotted across the list positions in each of the four Stress Pattern conditions. As can be seen performance across the list is very similar regardless of the condition. An analysis was performed which explored the effect of Stress Pattern on recall accuracy in each list position. Trials where six syllables were attempted were selected and this resulted in the loss of no participants' data. A main effect of List Position was observed, $F_1(5, 95) = 55.1$, $MSE = .02$, $p<.001$, $F_2(5, 235) = 49.5$, $MSE = .06$, $p<.001$. Bonferroni pairwise comparisons demonstrated that performance in the first position was significantly more accurate than all the other list positions (all *p*s< .001) suggesting a primacy effect. Performance in position 2 was significantly better to that in positions 3, 4, and 5 (all *p*s<.001) but not to the final list position (all *ps*=.10). Performance in position 3 was only significantly better than position 5 in the subjects analysis ($p=.023$ by subjects, $p=.14$ by items). Finally, performance on positions 4 and 5 did not differ from one another (all *ps*=.10) but performance on both positions was significantly worse than that observed in position 6 (position 4, $p<.001$ by subjects and $p=.003$ by items; position 5, $p<.001$ by subjects and $p<.001$ by items) suggesting a recency effect. Crucially, however, no interaction of List

Position and Stress Pattern was found, $F_1(15, 285) = 1.1$, MSE=.06, p=.40, $F_2(15, 705) = 1.5$, MSE=.11, p=.12, demonstrating that stress grouping did not affect recall performance in any list position.

Figure 2 about here

Extended Participant Testing

We were concerned that an iambic grouping effect might emerge with larger participant numbers. We, therefore, tested a further 12 participants in this experiment and repeated the analyses with data from all 32 participants. There was very little change in the percentage accuracy observed in each stress pattern (Trochaic, 29%; Iambic, 31%; Stressed, 29%; Unstressed, 27%). The effect of Stress Pattern on accuracy (arcsine transformed) was once again non-significant, $F_1(3, 93) = 1.3$, MSE=.01, p=.27, $F_2(3, 141) = 1.1$, MSE=.01, p=.37. **N**o other differences in the pattern of results were observed.

Comparison of Experiments 1 and 2

Finally we conducted an analysis of syllable accuracy, including the factor Experiment, to compare the results from Experiments 1 and 2. This analysis included only the initial 20 participants tested in Experiment 2 to match the power of this experiment to that of Experiment1. The ANOVA yielded a main effect of Experiment in the items analysis, $F_1(1, 38) = 1.3$, MSE=.04, p=.26, $F_2(1, 94) = 4.4$, MSE=.03, p=.04, due to the higher percentage accuracy in Experiment 1 than in Experiment 2 (33% and 30% respectively). In addition, there was a significant interaction of Experiment and Stress Pattern, $F_1(3, 114) = 3.3$, *MSE*=.01, *p*=.02, F_2(3, 282) = 3.6, *MSE*=.01, *p*=.01. Planned comparisons showed that this interaction was due to the more accurate performance on the trochaic lists in Experiment 1 compared to Experiment 2, F_1(1, 38) = 4.2, *MSE*=.01, *p*=.048, F_2(1, 94) = 7.4, *MSE*=.02, *p*=.008. Performance on the iambic lists did not differ between experiments, F_1&F_2<1, neither did performance in the stressed lists, F_1 <1, F_2(1, 94) = 2.7, *MSE*=.01, *p*=.10. Performance on the unstressed lists, differed significantly only in the items analysis, F_1(1, 38) = 2.4, *MSE*=.01, *p*=.13, F_2(1, 94) = 4.7, *MSE*=.02, *p*=.03.

Discussion

In contrast to Experiment 1, we observed no benefit to STM performance for either the trochaic or iambic lists compared to the monotone lists. In particular the manipulation of spoken duration did not produce evidence of the predicted benefits of iambic grouping. Instead the benefit of trochaic grouping disappeared.

Why do we observe no benefit to recall of our stress manipulation in this experiment? One possibility is that the manipulation of duration rather than intensity disrupted the perception of relative stress compared to Experiment 1. In other words participants in this experiment might have failed to hear the difference in prominence between stressed and unstressed syllables and as a consequence no stress grouping occurred. Despite this however, as with Experiment 1, there is strong evidence to suggest that participants both successfully perceived and reproduced the intended relative stress patterns. Participants' responses to stressed lists in Experiments 1 and 2 were perceived by independent listeners to be correctly stressed to the same extent (92% and 95% respectively). Likewise the monotone list output in both experiments was also clearly perceived as such (99% and 96% correct respectively). Furthermore, as will be shown below, speech measurements of the acoustic correlates of stress demonstrated that participants appropriately stressed their output in both stressed and monotone lists.

Nevertheless, the manipulation of the acoustic cues to stress in Experiment 2 has changed the effect of stress grouping on STM performance compared to Experiment 1. One possibility is that our acoustic manipulations affected the way in which lists were grouped. In most grouping studies duration has been shown to be a very strong and universal cue to right-edge boundaries (e.g., Bion, Benavides-Varela & Nespor, 2011; Tyler & Cutler, 2009). Following the logic we used in Experiment 1, this grouping strategy would result in three groups in the iambic lists (e.g., [σσ́][σσ́][σσ́]) but four groups in the trochaic lists (e.g., [σ́][σσ́][σσ́][σ]). However, this grouping would predict better performance on the iambic lists than the trochaic lists, which is not what we find.

Another possibility is that we observe no STM benefits of either stress pattern in Experiment 2 because *no* form of grouping has occurred. For the trochaic lists this may be due to the lack of an intensity difference to mark group onset. In the case of the iambic lists, it is possible that the grouping effect of duration was weakened by the concurrent manipulation of pitch. There is some evidence that (when other acoustic cues are held constant) pitch can function as a left-edge marker for syllable grouping in English. For example, Tyler & Cutler (2009) employed an Artificial Language Learning (ALL) paradigm where participants listen to a continuous speech stream which is made up of novel 'words' (the familiarisation phase). Participants are subsequently tested on their recognition of these component words of the language or on words which have been constructed from groups of syllables which had spanned word boundaries in the familiarisation phase. Using such a paradigm it was found that lengthening was a strong and universal cue to word-final boundaries in English, French and Dutch. So this acoustic cue is used in the same way across languages. In contrast pitch was shown to be a better left-edge cue for English and Dutch (trochaic languages) and a better right-edge cue for French (iambic language) suggesting that pitch is a language-specific cue. However, in English the difference between left and right-edge effects was not large and other studies have shown evidence for both left and right-edge pitch marking (e.g., Toro, Sebastián-Gallés & Mattys, 2009). Moreover, pitch and duration can also vary together to mark prominence (Fry, 1958). It seems unlikely, therefore, that the pitch manipulation would compete with the strong duration cue to iambic grouping and our acoustic analysis of speakers' outputs (reported below) also provide no evidence for such a conflict.

A final possibility is that the stress manipulations were successfully perceived and did result in the expected grouping of syllables, but nevertheless failed to benefit recall. This would suggest that factors other than a language-specific trochaic grouping bias are playing a role in the observed improvement to STM performance. In particular, the advantage to memory of trochaic groupings, which was demonstrated in Experiment 1, may be dependent on the acoustic naturalness of the groupings and that in particular intensity is necessary for the benefit to be observed.

The effects of perceptual grouping cues and language-dependent grouping preferences on memory are difficult to untangle. However, relevant evidence could come from the way in which the lists were produced. We, therefore, performed acoustic analyses on our participants' speech output in order to find evidence of grouping in their use of acoustic cues.

Acoustic Speech Analyses

In this study we were not only interested in the effect of our stress manipulations on the participants' STM performance, we were also interested in its effect on their speech output. Examining the participants' spoken production of the nonsense lists can provide critical information about the processes underlying their performance. Initially we wished to be able to confirm the perceptual judgments made about the speakers' production of stress using acoustic speech measurements. Furthermore we wanted to see whether participants did indeed alter their stress production to match the auditory grouping of the stimuli in the trochaic and iambic lists. Our experimental stimuli differed in the properties we manipulated in order to vary stress. In Experiment 1, intensity and pitch were manipulated while spoken duration was held constant. In Experiment 2, spoken duration and pitch were varied while intensity was held constant. We, therefore, also wished to test whether speakers' output mirrored the acoustic nature of the input lists or whether speakers stressed their output using all the acoustic features they would in their normal speech production.

Experiment 1: speech analysis

We included in the measurement analysis syllable pairs that were correctly pronounced and judged to have the correct stress pattern. As syllable pairs differed considerably in their phonetic characteristics we searched for pairs that were correctly produced an equal number of times in each stressed condition and at least once in the unstressed condition. We took the unstressed condition as the baseline because speakers' output in this condition was judged to be closest to normal output. Furthermore, the inclusion of both ungrouped conditions would have resulted in too a large reduction in the data set.

The acoustic properties of syllable pairs would also be influenced by their overall position in a list, therefore, each syllable pair was matched for list position across conditions. In total, 23 of the 48 experiment lists yielded at least one such triplet of syllable pairs and the data set included a total of 83 individual pairs.

The digitised recording of each of the syllables was then analysed using Praat software (Boersma & Weenick, 2009). For each syllable, measurements were taken of its spoken duration (ms), mean intensity (dbl) and mean pitch (Hz). These measurements were performed in the same way as for the acoustic stimuli presented to participants: mean intensity and pitch across the whole syllable and whole spoken syllable duration measured using the same segment dependent criteria. Two independent measurers analysed all speech files. The two measurement sets for each syllable correlated very highly with each other (all rs >0.8, $p<.001$) and none of the analyses we conducted differed between the two sets of measures. We, therefore, report the data from one measurement set only. The mean acoustic measures for Experiment 1 are shown in Table 5.

Table 5 about here

The absolute differences between the first and second syllables in a pair are influenced by the particular segmental content of the syllable, as different speech sounds have intrinsic differences in their spoken duration, pitch and intensity. Moreover, individual differences in the articulation of speakers will affect any comparison of the absolute measures. We, therefore, computed the relative difference between the first and second syllables for each syllable pair. The mean differences are shown in Figure 3a-c. A positive difference shows that the syllable 1 measure was greater than the syllable 2 measure and a negative different shows the reverse. As can be seen, all measures are positive for trochaic and unstressed syllable pairs and negative for iambic syllable pairs. Therefore in both the trochaic and iambic conditions intensity, duration and pitch were varied together to mark relative prominence.

Figure 3 about here

An analysis of variance was performed on the difference in intensity between the first and second syllables across conditions with syllable pair as the random factor. There was a significant effect of Stress Pattern condition, $F_2(2, 44) = 5.3$, $MSE = 107$, $p = .009$. Pairwise comparisons showed that the trochaic condition had a significantly higher intensity difference than both the iambic ($p = .038$) and the unstressed conditions ($p = .001$). In contrast, the iambic and unstressed conditions did not significantly differ.

A similar analysis of syllable durations also yielded a significant effect of Stress Pattern, $F_2(2, 44) = 15.3$, $MSE = 66786$, $p < .001$. Pairwise comparisons showed that the trochaic condition differed significantly from the iambic condition ($p < .001$) and that the iambic condition also differed from the unstressed condition ($p = .004$). The trochaic and unstressed conditions did not significantly differ.

A similar pattern was observed with the pitch measurements. The main effect of Stress Pattern was again significant, $F_2(2, 44) = 13.8$, $MSE = 17145$, $p < .001$. The trochaic condition differed significantly from the iambic condition ($p < .001$) and that the iambic condition also differed from the unstressed condition ($p = .019$). The trochaic and unstressed conditions did not significantly differ.

For all measures, therefore, we see significant differences between the two stress pattern conditions, demonstrating that participants did indeed reproduce the stress patterns they heard in their output. Interestingly however, the way participants generated their stress patterns was not identical to the input. Stress in the trochaic lists was marked by a significant difference in intensity but not duration or pitch, compared to unstressed lists. In contrast, stress in iambic lists was marked by significant differences in duration and pitch but not intensity, compared to unstressed lists. The participants' productions did not, therefore, match the input as iambic pairs significantly differed in syllable duration despite the fact that syllable duration was held constant in the auditory stimuli.

Experiment 2: speech analysis

Similar data inclusion criteria were applied to Experiment 2. In this Experiment 19 lists yielded at least one matched set of syllable pairs and the data set included 64 individual pairs. The same independent measurers analysed all speech files. Once again the two sets of measures for each syllable correlated very highly with each other (all rs >.9, *p*s<.001) and none of the analyses we conducted differed between the two sets of measures. We, therefore, report the data from the same measurer as we did for Experiment 1. The mean acoustic measures for Experiment 2 are shown in Table 6.

Table 6 about here

We again computed the relative difference between the first and second syllables for each syllable pair. The mean differences are shown in Figure 3d-f. As can be seen the overall pattern is very similar to Experiment 1, with stressed syllables in both the trochaic and iambic conditions tending to be relatively louder, longer and higher pitch than unstressed syllables.

The intensity measures again showed a significant effect of Stress Pattern, $F_2(2, 34) = 13.3$, $MSE = 88$, $p < .001$. Pairwise comparisons showed that the trochaic condition differed significantly from both the iambic (p<.001) and the unstressed conditions (p=.021) but that the iambic and unstressed conditions did not significantly differ.

Durations also yielded a significant effect of Stress Pattern, $F_2(2, 34) = 11.9$, MSE = 216079, p<.001. The trochaic condition differed significantly from the iambic condition (p<.001) and the iambic condition also differed from the unstressed condition (p=.02). The trochaic and unstressed conditions did not significantly differ.

The pitch measurements yielded a significant main effect of Stress Pattern, $F_2(2, 34) = 23.3$, MSE=10239, p<.001. The trochaic condition differed significantly from the iambic condition (p<.001) and unstressed condition (p<.001). The iambic and unstressed condition did not significantly differ.

Once again, therefore, the stressed grouped conditions differed significantly from each other on all acoustic measures, including intensity, which in Experiment 2, was held constant in the acoustic stimuli. Compared to the unstressed lists, stress in the trochaic lists was marked by a significant difference in intensity and pitch but not duration, whereas stress in the iambic lists was marked by significant differences in duration but not intensity and pitch.

Comparison of Experiments 1 and 2

In order to test for any significant differences in the speech characteristics of Experiments 1 and 2, we ran analyses including the measurements from both experiments. Only eight of the experimental lists yielded at least one triplet of syllable pairs in both experiments and the joint data set included 64 individual pairs. Nevertheless the pattern of results across this subset of the data was very similar to the larger set. ANOVAs were performed on all acoustic measures including the factors Experiment (1 and 2) and Stress Pattern condition (trochaic, iambic, unstressed).

The pattern of results for intensity was very similar in both experiments. The analysis showed a main effect of Stress Pattern, $F_2(2, 14) = 11.9$, MSE=100, p<.001. However, the main effect of Experiment and the interaction of Experiment and Stress Pattern were not significant.

In contrast the duration differences were almost twice as large in Experiment 2, than in Experiment 1, although the pattern of difference remained similar. This was confirmed in the analysis of durations, which showed main effects of Stress Pattern, $F_2(2, 14) = 12.9$, MSE=114620, p=.001 and Experiment, $F_2(1, 7) = 11.0$, MSE=130292, p=.013 but no interaction, $F_2 < 1$.

Finally, the pitch analysis yielded no main effect of Experiment, $F_2 < 1$. However, there was a significant effect of Stress Pattern, $F_2(2, 14) = 11.8$, MSE =13780, p=.001 and an interaction of Stress Pattern and Experiment, $F_2(2, 14) = 4.2$, MSE =2675, p=.038. This was due to a surprisingly large decrease in the pitch difference in the unstressed condition from Experiment 1 to Experiment 2. The pitch differences for the grouped conditions did not change.

Discussion

Taken together the acoustic analysis of participants' spoken output has provided data relevant to a number of issues. Firstly, they confirm that speakers did indeed reproduce the intended stress patterns which they heard in the input for both Experiment 1 and 2 as, stressed syllables tended to be louder, longer and of higher pitch than unstressed syllables. In addition, our acoustic data are consistent with participants generating the trochaic and iambic groupings we expected. In both Experiments 1 and 2, stress in the trochaic lists was marked by a significant difference in intensity but not duration compared to unstressed lists. In contrast, in the iambic lists, stress was marked by significant differences in duration but not intensity compared to unstressed lists. This pattern is consistent with the trochaic/iambic law (Hayes, 1995) and suggests that our participants were producing the correct acoustic patterns for trochaic and iambic groupings irrespective of the input they had received. Importantly these differing effects of intensity and duration would not be predicted if all stressed syllables were functioning as trochaic onsets in Experiment 1 and as iambic offsets in Experiment 2. The pattern of pitch effects in our production data is less consistent, with significant differences observed on stressed syllables in iambs but not trochees in Experiment 1 and the reverse pattern observed in Experiment 2. Importantly, however, there is no clear evidence in the production data that speakers mark only left-edge boundaries with pitch.

It is very clear that speakers did not simply reproduce the acoustic properties of the speech that they heard. The finding that participants' speech output remains largely the same despite changes in the auditory input suggests that they are encoding the lists in the same way as they would for internally driven speech. Nevertheless, some aspects of the speech input did affect participants’ speech output as the difference in syllable duration was significantly less in Experiment 1 than in Experiment 2 suggesting that participants showed a general sensitivity to the duration manipulation in the second experiment. This duration difference, however, was of the same magnitude across all conditions.

The acoustic analyses, therefore, provide evidence that participants have grouped the lists in both experiments. However the production of stress groups in the output is unrelated to the grouping benefit for STM performance, which was only observed for trochaic lists in Experiment 1. This pattern of results is, therefore, more consistent with a perceptual rather than a production locus for the effect of stress grouping on STM performance.

General Discussion

We have presented a study which investigated the effect of stress grouping on verbal STM performance in order to further our understanding of the relationship between the language system and verbal STM. Our data contribute to this knowledge in the following ways. When syllable intensity was manipulated but syllable duration was held constant (Experiment 1), recall of nonsense syllables was enhanced by the trochaic stress patterns but not by iambic stress patterns. When syllable duration was manipulated but syllable intensity was held constant (Experiment 2), we observed no recall benefit for either trochaic or iambic stress patterns. Such a pattern of findings provides some support for the claim that stored long-term knowledge (i.e., language-specific prosodic preferences) can be recruited to support the encoding of items in STM. Our analysis of participants' speech output showed that recalled lists were appropriately stressed in both experiments. However the way in which stress differences were generated was largely independent both of the speech input participants heard and of the recall benefits observed. This finding supports the notion that the output generated in STM tasks does not simply involve the reproduction of a stored representation but rather makes use of the encoding processes which are relied upon in spontaneous language production.

Our findings are consistent with previous research which suggests that stress grouping of auditory input benefits verbal serial recall performance (Boucher, 2006; Reeves, Schmauder & Morris, 2000). However, ours is the first study to compare the effects of trochaic and iambic patterns and a number of aspects of our data suggest that this benefit is modulated by language-specific preferences. First, we only observed an increase in recall when the stress pattern was trochaic, which is the preferred prosodic grouping in English, and not when the stress pattern was iambic. Second, for the trochaic grouping benefit to be observed the input must be grouped using appropriate acoustic properties; the trochaic grouping only led to STM performance benefits when it was marked by differences in intensity. The importance of intensity in determining language structure in English has also been shown by studies investigating the use of stress to distinguish lexical and phrasal structures (e.g., a GREENhouse/a green HOUSE). These studies have demonstrated that the contrasting stress patterns were predominantly marked by differences in intensity (e.g., Morrill, 2011; Shilling, Wheeldon & Krott, in prep). The explanation of our grouping effect cannot, therefore, refer only to a universal perceptual bias (i.e., iambic/trochaic law) but must incorporate an additional language-dependent prosodic preference.

We argue, therefore, that our data are in line with the claim that verbal STM can be supported by general properties of the language system (e.g., Gathercole et al., 1999; Thorn, Gathercole & Frankish, 2005) as well as by specific characteristics of lexical representations (e.g., Gathercole, Pickering, Hall & Peaker, 2001; Romani, McAlpine & Martin, 2008; Roodenrys, Hulme, Alban, Ellis & Brown, 1994; Walker & Hulme, 1999). Such an influence of stored knowledge on verbal STM has been accounted for in a number of ways (as part of the working memory framework, Burgess & Hitch, 2005, 2006; in terms of redintegration: Hulme et al., 1997; Hulme, Newton, Cowan, Stuart & Brown, 1999; Schweickert, 1993; and within the multiple-code hypothesis, Martin et al, 1999; Romani et al., 2008). In particular strong versions of the redintegration hypothesis suggest that *only* stored information which is specific to a lexical representation can be used to rebuild decaying memory traces. However, in order to maintain and output the material in our task, participants could not rely upon such specific lexical information because the stimuli were nonsense syllables. Nevertheless language-specific prosodic preferences affected participants' serial recall performance suggesting that general properties of the language can also be recruited to support verbal STM.

As such, our data provide no support for the notion of a separate, language-independent memory store. Rather, our data are consistent with the claim that it is not necessary to distinguish between the phonological encoding processes necessary for language production and those responsible for verbal STM output (Acheson and Macdonald, 2009a and b; Ellis, 1980; Page et al, 2007). In our speech measurement analysis we have provided clear evidence for similarities between normal phonological encoding processes and the output processes for serial recall. Despite significant difference in the acoustic features we manipulated to generate syllable prominence in Experiments 1 and 2, our speech measurements demonstrate that our participants varied *all* relevant acoustic features (i.e., intensity, duration and pitch) in their list output in the same way as would be expected for speech driven by an internally generated conceptual representations. Our speech measurement data, therefore, provide unique evidence that participants recruit their normal speech production processes to maintain and output information in verbal STM tasks.

However, it is clear that the nature of the auditory input did influence the participants' output to some extent as the overall size of the duration difference produced by participants was greater when syllable duration was varied in the input (Experiment 2). Nevertheless, several aspects of our data clearly support the distinction between input and output processes for verbal STM (Howard & Nickels, 2005; Jacquemot & Scott, 2006; Martin et al., 1999). First, the grouping benefit for STM recall was limited to trochaic lists despite the speech measurement evidence that both kinds of stress patterns were generated by our speakers. Second, the way participants generated their stress patterns was not identical to the input they heard. These findings, therefore, require that a clear distinction is made between perceptual and production processes during short-term recall.

Such a conceptualisation of verbal STM has received strong support from cognitive neuropsychological studies which have described patients who are characterised by dissociations in their ability to perform memory tasks which involve production (i.e., serial recall tasks) and those which involve perception (i.e., matching or probe tasks) (e.g., Howard and Nickels, 2005; Martin & Freedman, 2001; Martin, Lesch & Bartha,1999). Of particular relevance to our findings is the anomic patient MS reported by Martin et al (1999) who showed good performance on STM tasks which did not require response output. Moreover MS was not impaired in all STM tasks which involved production. Rather MS performed worse on recall tasks which involved words compared to nonwords suggesting that the nature of his impairments related to the accessing of stored lexical knowledge. In line with this finding our data are consistent with the existence of separable phonological input and output buffers which have access to long-term language-specific knowledge (Martin et al., 1999) or with the activation of long-term knowledge or processes associated with both the perception and production of phonological representations (Acheson & MacDonald, 2009a).

Finally, our data point to a perceptual rather than a production locus for our STM trochaic grouping benefit. This is suggested by the mismatch which was observed between the auditory input and participants' speech output. The importance of perceptual factors in the effect of stress grouping in recall is also shown by the dependency of the trochaic grouping effect on the intensity manipulation, as evidenced by the null effect we observed when intensity was held constant (Experiment2). These findings are consistent with STM research which attributes grouping effects to the presence of salient perceptual cues in the auditory input (e.g., timing differences and physical/acoustic attributes of lists). Such perceptual cues impose list-internal structural features on encoded representations which act as superior retrieval cues for items in grouped compared to ungrouped lists (Frankish, 1989; Frankish & Turner, 1984). In a similar vein our findings suggest that the grouping benefit lies in the early process of determining the units to-be-remembered rather than in the process of building their structure for output.

In conclusion we have presented data which contribute to our understanding of the relationship between language processing and memory in a number of ways. In terms of the way memory is conceptualised we have demonstrated that short-term and long-term memory representations are closely linked. In particular verbal STM performance can be influenced by stored linguistic knowledge which speaks against a language-independent short-term store. Furthermore our data support the notion that within verbal STM phonological input and output processes are separable. Finally, in terms of which aspects of verbal STM and speech production share architecture we have demonstrated that participants' output in STM tasks relies on the same phonological encoding processes that are recruited during spontaneous speech production. Taken together our findings have not only added important detail to our theoretical knowledge of how memory and language processing interact but also our speech measurements analysis has provided us with a novel methodological approach to address these issues.

Acknowledgements

Authors one and three are equal first authors on the manuscript. The experiments were designed by these authors and run at the University of Birmingham by the second author who also contributed to the design and analysis. Thanks are also due to three research assistants, Elizabeth Cartwright, Grace Jenkins and Luke Condley who assisted with data collection and speech measurements of Experiment 2.

Footnote

1. The vast majority of the errors contributing were segmental in nature, therefore, the accuracy scores reported largely reflect segmental accuracy. However, this experiment was not designed to examine segmental error patterns and a detailed analysis of the segmental data is beyond the scope of this article.

Appendix 1

The twenty-four nonsense syllables (and IPA transcriptions in slashes) presented in Experiments 1 and 2

CVC				VC			
long vowels		short vowels		long vowels		short vowels	
sape	/seɪp/	pag	/pæg/	eed	/i:d/	es	/ɛs/
heek	/hi:k/	jek	/ʤɛk/	ipe	/aɪp/	ib	/ɪb/
nige	/naɪʤ/	hif	/hɪf/	oge	/əʊg/	om	/ɔm/
boop	/bu:p/	nol	/nɔl/	abe	/e:b/	af	/æf/
fote	/fəʊt/	sut	/sʌt/	ane	/eɪn/	ep	/ɛp/
rame	/reɪm/	rel	/rɛl/	oot	/u:t/	ut	/ʌt/

REFERENCES

Acheson, D. J., & MacDonald, M. C. (2009a). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin, 135(1),* 50-68. doi: 10.1037/a0014411

Acheson, D. J., & MacDonald, M. C. (2009b). Twisting tongues and memories: Explorations of the relationship between language production and verbal working memory. *Journal of Memory and Language, 60*, 329-350. doi: 10.1016/j.jml.2008.12.002

Acheson, D. J., Hamidi, M., Binder, J. R., & Postle, B. R. (2011). A common neural substrate for language production and verbal working memory. *Journal of Cognitive Neuroscience. Cognitive Neuroscience, 23(6)*, 1358-1367. doi: 10.1162/jocn.2010.21519

Adams, H. F. (1915). A note on the effect of rhythm on memory. *Psychological Review, 22 (4)*, 289-298.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* (Release 2) [CD-ROM]. Linguistic Data Consortium, University of Pennsylvania, Philadelphia,

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

Baddeley, A. D., & Hitch, G. J. (1974). Working Memory. In G. H. Bower (Ed.), *The psychology of learning and motivation (Vol. 8)*. London: Academic Press.

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*, 575-589. doi: 10.1016/S0022-5371(75)80045-4

Barbosa, P. A., Arantes, P., Meireles, A. R., & Vieira, J. M. (2005). Abstractness in Speech-

Metronome Synchronisation: P-Centres as Cyclic Attractors. *Proceedings of the Ninth European Conference on Speech Communication and Technology* (Interspeech 2005) Lisbon, Portugal. pp.1441-1444.

Bion, R. A. H., Benavides-Varela, S., & Nespor, M. (2011). Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Language and Speech, 54,* 123-140. doi: 10.1177/0023830910388018

Boersma, P., & Weenink, D. (2009). *Praat: doing phonetics by computer (Version 5.1.05),* from http://www.praat.org/.

Boucher, V. J. (2006). On the function of stress rhythms in speech: Evidence of a link with grouping effects on serial memory. *Language & Speech, 49(4),* 495-519. doi: 10.1177/00238309060490040301

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review, 107(1),* 127-181. doi: 10.1037/0033-295X.107.1.127

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review, 106(3),* 551-581. doi: 10.1037/0033-295X.106.3.551

Burgess, N., & Hitch, G. J. (2005). Computational models of working memory: Putting long term memory into context. *TRENDS in Cognitive Sciences, 9(11),* 535-541. doi: 10.1016/j.tics.2005.09.011

Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language, 55(4)*, 627-652. doi: 10.1016/j.jml.2006.08.005

Cholin, J., Levelt, W. J. M., & Schiller, N. O. (2006). Effects of syllable frequency in speech production. *Cognition, 99,* 205-235. doi: 10.1016/j.cognition.2005.01.009

Cholin, J., Schiller, N. O., & Levelt, W. J. M. (2004). The preparation of syllables in speech production. *Journal of Memory and Language, 50*, 47-61. doi: 10.1016/j.jml.2003.08.003

Clark, J., & Yallop, C. (1995). *An introduction to phonetics and phonology.* Oxford: Blackwell.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology, 33A*, 497-505. doi:10.1080/14640748108400805

Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech, 45(3),* 207-228. doi: 10.1177/00238309020450030101

Cowan, N., Saults, J. S., Elliot, E. M., & Moreno, M. V. (2002). Deconfounding serial recall. *Journal of Memory and Language, 46(1)*, 153-177. doi: 10.1006/jmla.2001.2805

Cutler, A. (1994). Segmentation problems, rhythmic solutions. *Lingua, 92,* 81-104. doi: 10.1016/0024-3841(94)90338-7

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language, 31(2)*, 218-236. doi: 10.1016/0749-596X(92)90012-M

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2*, 133-142. doi: 10.1016/0885-2308(87)90004-0

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 113-121. doi: 10.1037/0096-1523.14.1.113

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*, 283-321. doi: 10.1037//0033-295X.93.3.283

Ellis, A. W. (1980). Errors in speech and short-term memory: The effects of phonemic similarity and syllable position. *Journal of Verbal Learning and Verbal Behaviour, 19,* 624-634. doi: 10.1016/S0022-5371(80)90672-6

Fowler, C. A. (1979). “Perceptual centres” in speech production and perception. *Perception & Psychophysics, 25(5),* 375-388. doi:10.3758/BF03199846

Frankish, C. (1985). Modality-specific grouping effects in short-term memory. *Journal of Memory and Language, 24(2),* 200-209. doi: 10.1016/0749-596X(85)90024-5

Frankish, C. (1989). Perceptual organization and precategorical acoustic storage. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15(3),* 469-470. doi: 10.1037/0278-7393.15.3.469

Frankish, C., & Turner, J. (1984). Delayed suffix effect at very short delays. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10(4),* 767-777. doi: 10.1037//0278-7393.10.4.767

Frick, R. W. (1989). Explanations of grouping in immediate ordered recall. *Memory and Cognition, 17(5),* 551-562. doi: 10.3758/BF03197078

Fromkin, V. (Ed.) (1980). *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand*. San Francisco: Academic Press.

Fry, D. B., (1958). Experiments in the perception of stress. *Language and Speech, 1,* 126-152.

Garrett, M. F. (1980). The limits of accommodation: Arguments for independent processing levels in sentence production. In V. A. Fromkin (Ed.), *Errors in linguistic performance: slips of the tongue, ear, pen and hand.* New York: Academic Press.

Gathercole, S. E. (1995). Nonword repetition: More than just a phonological output task. *Cognitive Neuropsychology, 12(8),* 857-861. doi: 10.1080/02643299508251405

Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 25(1)*, 84-95. doi: 10.1037/0278-7393.25.3.562

Gathercole, S. E., Pickering, S. J., Hall, M., & Peaker, S. M. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *The Quarterly Journal of Experimental Psychology, 54A(1)*, 1-30. doi: 10.1080/02724980042000002

Gerken, L. A. (1994). Young children's representation of prosodic phonology: Evidence from English-speakers' weak syllable productions. *Journal of Memory and Language*, *33,* 19-38. doi: 10.1006/jmla.1994.1

Hay, J. S. F., & Diehl, R. L. (2007). Perception of rhythmic grouping: Testing the iambic/trochaic law. *Perception and Psychophysics, 69(1)*, 113-122. doi: 10.3758/BF03194458

Hayes, B. (1982). Extrametricality and English stress. *Linguistic Inquiry*, *13,* 227-276.

Hayes, B. (1995). The rhythmic basis of the foot inventory. In B. Hayes (Ed.), *Metrical stress theory: Principles and case studies* (pp.79-85). Chicago: University of Chicago Press.

Howard, D., & Nickels, L. (2005). Separating input and output phonology: Semantic, phonological, and orthographic effects in short-term memory impairment. *Cognitive Neuropsychology, 22(1)*, 42-77. doi: 10.1080/02643290342000582

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., 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)*, 1217-1232. doi: 10.1037/0278-7393.23.5.1217

Hulme, C., Newton, P., Cowan, N., Stuart, G., & Brown, G. (1999). Think before you speak: Pauses, memory search and trace redintegration processes in verbal memory span. *Journal of Experimental Psychology: Learning, Memory and Cognition, 25(2)*, 447-463. doi: 10.1037/0278-7393.25.2.447

Inkelas, S., & Zec, D. (1990).*The phonology-syntax connection*. Chicago: University of Chicago Press.

Jacquemot, C., & Scott, S. K. (2006). What is the relationship between phonological short-term memory and speech processing? *TRENDS in Cognitive Sciences, 10(11),* 480-486. doi: 10.1016/j.tics.2006.09.002

Lahiri, A., & Wheeldon, L. R. (2011). Phonological trochaic grouping in language planning and language change. In S. Frota (Ed.) *Prosodic categories: production, perception and comprehension*. Springer.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M. (1992). Accessing word in speech production: Stages, processes and representations. *Cognition, 42,* 1-22. doi: 10.1016/0010-0277(92)90038-J

Levelt, W. J. M. (1999). Models of word production. *TRENDS in Cognitive Sciences, 3,* 223-232. doi: 10.1016/S1364-6613(99)01319-4

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in language production. *Behavioural and Brain Sciences, 22,* 1-38. doi:10.1017/S0140525X99001776

Levelt, W. J. M., & Wheeldon, L. R. (1994). Do speakers have access to a mental syllabary? *Cognition, 50,* 239-269. doi: 10.1016/0010-0277(94)90030-2

Malpass, D., Lahiri, A., & Wheeldon, L. (2010). *Generating phonological words: evidence from English phrase production*. CUNY: Phonology Forum Conference on the Word. New York, Jan 14-15.

Marcus, S. M. (1981). Acoustic determinants of perceptual center (P-center) location. *Perception & Psychophysics, 30(3),* 247-256. doi: 10.3758/BF03214280

Martin, R. C., & Freedman, M. L. (2001). Short-term retention of lexical-semantic representations: Implications for speech production. *Memory, 9 (4/5/6),* 261-280. doi: 10.1080/09658210143000173

Martin, R. C., Lesch, M. F., & Bartha, M. C. (1999). Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language, 41*, 3-29. doi: 10.1006/jmla.1999.2637

Meyer, A. S. (1992). Investigation of phonological encoding through speech error analysis: Achievements, limitations, and alternatives. *Cognition, 42,* 181-211. doi:10.1016/0010-0277(92)90043-H

Morgan, J. L., & Wheeldon, L. R. (2003). Syllable monitoring in internally and externally generated English words. *Journal of Psycholinguistic Research, 32(3)*, 269-296. doi: 10.1023/A:1023591518131

Morrill, T. (2011). Acoustic Correlates of Stress in English Adjective–Noun Compounds. *Language and Speech,* 1-35. doi: 10.1177/0023830911417251

Morton, J., Marcus, S. M., & Frankish, C. R. (1976). Perceptual centers (P-centers). *Psychological Review, 83*, 405-408. doi: 10.1037//0033-295X.83.5.405

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht, The Netherlands: Foris.

Page, M. P. A., Madge, A., Cumming, N., & Norris, D. G. (2007). Speech errors and the phonological similarity effect in short-term memory: Evidence suggesting a common locus. *Journal of Memory and Language, 56,* 49-64. doi: 10.1016/j.jml.2006.09.002

Quené, H., & Port, R. F. (2005). Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica, 62(1)*, 1-13. doi: 10.1159/000087222

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review, 107*, 460-499. doi: 10.1037/0033-295X.107.3.460

Reeves, C., Schmauder, A. R., & Morris, R. K. (2000). Stress grouping improves performance on an immediate serial list recall task. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26(6)*, 1638-1654. doi: 10.1037//0278-7393.26.6.1638

Revithiadou, A. (2004). The iambic/trochaic law revisited: Lengthening and shortening in trochaic systems. In B. Arsenijevic, N. Elouazizi, M. Salzmann, & M. de Vos, (eds.) *Leiden Papers in Linguistics 1.1*, 37-62. Leiden Universiteit/ULCL, Leiden.

Roelofs, A. (1997). Syllabification in speech production: Evaluation of WEAVER. *Language and Cognitive Processes*, *12,* 657-693. doi: 10.1080/016909697386655

Roelofs, A. (2002). Spoken language planning and the initiation of articulation. *The Quarterly Journal of Experimental Psychology - Section A, 55*, 465-483. doi: 10.1080/02724980143000488

Romani, C., McAlpine, S., & Martin, R. C. (2008). Concreteness effects in different tasks: Implications for models of short-term memory. *The Quarterly Journal of Experimental Psychology, 61(2),* 292-323. doi: 10.1080/17470210601147747

Roodenrys, S., & Hinton, M. (2002). Sublexical or lexical effects on serial recall of nonwords? *Journal of Experimental Psychology: Learning, Memory and Cognition, 28(1),* 29-33. doi: 10.1037//0278-7393.28.1.29

Roodenrys, S., Hulme, C., Alban, J., Ellis, A. W., & Brown, G. D. A. (1994). Effects of word frequency and age of acquisition on short-term memory span. *Memory & Cognition, 22(6),* 695-701. doi: 10.3758/BF03209254

Ryan, J. (1969a). Grouping and short-term memory: Different means and patterns of grouping. *The Quarterly Journal of Experimental Psychology, 21*, 137-147.

Ryan, J. (1969b). Temporal grouping, rehearsal and short-term memory. *The Quarterly Journal of Experimental Psychology, 21*, 148-155.

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition, 21(2)*, 168-175. doi: 10.3758/BF03202729

Selkirk, E. O. (1984). *Phonology and syntax: the relation between sound and structure.* Cambridge, MA: MIT Press.

Selkirk, E. O. (1995). The prosodic structure of function words. In J. L. Beckman, D. Walsh, & S. Urbanczyk (eds.) *Papers in Optimality Theory, UMASS Occasional Papers in Phonology*, Amherst, GLSA.

Shilling, H., Wheeldon, L. R., & Krott, A. (in prep). Compound and phrasal Stress acquisition: When does Greenhouse Become Different to Green House?

Svartvik, J., & Quirk, R. (Eds.) (1980). *A corpus of English conversation*. Lund, Sweden: Gleerup.

Theissen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology, 39(4)*, 706-716. doi: 10.1037/0012-1649.39.4.706

Thorn, A. S. C., & Frankish, C. R. (2005). Long-term knowledge effects on serial recall of nonwords are not exclusively lexical. *Journal of Experimental Psychology: Learning, Memory and Cognition, 31(4)*, 729-735. doi: 10.1037/0278-7393.31.4.729

Thorn, A. S. C., Gathercole, S. E., & Frankish, C. R. (2005). Reintegration and the benefits of long-term knowledge in verbal short-term memory: An evaluation of Schweickert's (1993) multinomial processing tree model. *Cognitive Psychology, 50(2)*, 133-158. doi: 10.1016/j.cogpsych.2004.07.001

Tilsen, S. (2011a). Metrical regularity facilitates speech planning and production. *Laboratory Phonology, 2:1*, 185-218. doi: 10.1515/LABPHON.2011.006

Tilsen, S. (2011b). Effects of syllable stress on articulatory planning observed in a stop-signal experiment. *Journal of Phonetics, 39*, 642-659. doi: 10.1016/j.wocn.2011.04.002

Toro, J. M., Sebastián-Gallés, N., & Mattys, S. (2009). The role of perceptual salience during the segmentation of connected speech. *The European Journal of Cognitive Psychology, 21(5),* 786-800. doi: 10.1080/09541440802405584

Tyler, M. D., & Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *Journal of the Acoustical Society of America, 126(1)*, 367-376. doi: 10.1121/1.3129127

Vousden, J. I., Brown, G. D. A., & Harley, T. A. (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive Psychology, 41*, 101-175. doi: 10.1006/cogp.2000.0739

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term recall. *Journal of Experimental Psychology: Learning, Memory and Cognition, 25(5)*, 1256-1271. doi: 10.1037/0278-7393.25.5.1256

Wheeldon, L. R. (2000). Generating prosodic structure. In L. R. Wheeldon (Ed.), *Aspects of sentence production* (pp. 249-274). Hove: Psychology Press.

Wheeldon, L. R., & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language, 37*, 356-381. doi: 10.1006/jmla.1997.2517

Wheeldon, L. R., & Lahiri, A. (2003). Is the phonological word a unit of language production? *Proceeding of ICPhS*, 426-431.

Wheeldon, L. R., & Levelt, W. J. M. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language, 34*, 311-334. doi: 10.1006/jmla.1995.1014

Wheeldon, L. R., & Morgan, J. L. (2002). Phoneme monitoring in internal and external speech. *Language and Cognitive Processes, 17(5)*, 503-535. doi: 10.1080/01690960143000308

Wickelgren, W. (1967). Rehearsal grouping and hierarchical organization of serial position cues in short-term memory. *The Quarterly Journal of Experimental Psychology, 14(2)*, 97-102.

Woodrow, H. (1909). A quantitative study of rhythm: The effect of variations in intensity, rate and duration. *Archives in Psychology, 14,* 1-66.

Figure captions

Figure 1. Mean percentage accuracy rates for syllables in each of the 6 serial positions for the four Stress Pattern conditions of Experiment 1.

Figure 2. Mean percentage accuracy rates for syllables in each of the 6 serial positions for the four Stress Pattern conditions of Experiment 2.

Figure 3. Mean differences (syllable 1 – syllable 2) and standard errors, for the acoustic measures of intensity, duration and pitch for matched syllable pairs in the trochaic, iambic and unstressed conditions of Experiment 1 and 2.