Semantic categorisation of a word supports its phonological integrity in verbal short-term memory

Nicola Savill, Tim Metcalfe, Andrew W. Ellis, Elizabeth Jefferies

Department of Psychology, University of York, UK

To appear in ***Journal of Memory and Language***

doi:10.1016/j.jml.2015.06.003

Corresponding Author:

Nicola Savill

Department of Psychology,

University of York,

Heslington,

York

YO10 5DD

UK

Email: nicola.savill@york.ac.uk

Tel: +44 (0)1904 322937

Fax: +44(0)1904 323181

Acknowledgements:

This work was funded by a European Research Council grant (283530 – SEMBIND)

Abstract

In three immediate serial recall (ISR) experiments we tested the hypothesis that interactive processing between semantics and phonology supports phonological coherence in verbal short-term memory (STM). Participants categorised spoken words in six-item lists as they were presented, according to their semantic or phonological properties, then repeated the items in presentation order (Experiment 1). Despite matched categorisation performance between conditions, semantically-categorised words were correctly recalled more often than phonologically-categorised words. This accuracy advantage in the semantic condition was accompanied by fewer phoneme recombination errors. Comparisons with a no-categorisation ISR baseline (Experiment 2) indicated that, although categorisations were disruptive overall, recombination errors were specifically rarer following semantic categorisation. Experiment 3 replicated the key findings from Experiment 1 and also revealed fewer phonologically-related errors following semantic categorisation compared to a perceptual categorisation of high or low pitch. Therefore, augmented activation of semantic representations stabilises the phonological traces of words within verbal short-term memory, in line with the "semantic binding" hypothesis.

Introduction

A wealth of evidence demonstrates that knowledge of the sounds and meanings of words supports their maintenance within verbal short-term memory (STM). Measures of immediate serial recall (ISR) – where a sequence of verbal material is immediately repeated back in order – consistently show higher accuracy for familiar words compared to unfamiliar nonwords (e.g., Hoffman, Jefferies, Ehsan, Jones, & Lambon Ralph, 2009; Hulme, Maughan, & Brown, 1991; Hulme, Roodenrys, & Brown, 1995; Jefferies, Frankish, & Lambon Ralph, 2006; Saint-Aubin & Poirier, 1999, 2000; Thorn, Gathercole, & Frankish, 2005). The independent contribution of phonological knowledge to this effect is demonstrated by recall/repetition advantages for phonologically-familiarised nonwords (or otherwise unfamiliar) stimuli compared to untrained items (Majerus, Linden, Mulder, Meulemans, & Peters, 2004; Melby-Lervåg & Hulme, 2010; Savill et al., 2015) and effects of phonotactic frequency on ISR accuracy (Thorn & Frankish, 2005). Meanwhile, independent influences of semantic knowledge are revealed by the impact of semantic manipulations such as word imageability/concreteness (Bourassa & Besner, 1994; Caza & Belleville, 1999; Walker & Hulme, 1999; Majerus & van der Linden, 2003; Jefferies et al., 2006a; Romani, McAlpine, & Martin, 2008) and by neuropsychological studies of word recall deficits in patients whose semantic knowledge is impaired (Patterson, Graham, & Hodges, 1994; N. Martin & Saffran, 1997; Jefferies, Jones, Bateman, & Lambon Ralph, 2004, 2005; Majerus, Van der Linden, Poncelet, & Metz-Lutz, 2004).

The explanations offered for these phenomena tend to attribute them to processes either (a) at the point of recall, where the accessibility of the lexical forms of words in long-term memory (LTM) influences the likelihood of correctly restoring the degraded phonological trace (‘redintegration’; Schweickert, 1993; see also Hulme et al., 1997, 1991,

1995; Saint-Aubin & Poirier, 2000), or (b) prior to recall, with temporary activation of long-term linguistic representations directly supporting STM (hereon referred to as 'language-based' accounts'; Patterson et al., 1994; Martin & Saffran, 1997; Acheson & MacDonald, 2009; Majerus, 2013). In redintegration accounts, long-term linguistic knowledge facilitates item reconstruction, with little provision for improved order memory (beyond recall of existing inter-item associations; Stuart & Hulme, 2000), while in language-based accounts it influences phonological encoding and maintenance of the sequence: phonological-lexical and semantic knowledge is thought to contribute to phoneme order memory (Patterson et al., 1994; Jefferies et al., 2004, 2006a; Jefferies, Frankish, & Noble, 2009; Hoffman et al., 2009) while syntactic knowledge supports word order (Acheson & MacDonald, 2009). These broad perspectives offer different predictions regarding how and when semantic representations influence recall. According to the redintegration account, LTM representations accessed at recall would most likely influence the rate of items recalled in *any* position (i.e., an increase in targets reconstructed successfully both in and out of position. In contrast, language-based accounts make specific predictions about the effect of semantic knowledge on phoneme ordering in STM (cf. "semantic binding hypothesis").

Most studies of verbal STM have examined item and order recall at a whole-item level (i.e., whether items are recalled in the correct serial position, or out of position, or not recalled at all) but have not examined recall at the phoneme level. The current study analyses phoneme-level errors to examine the predictions of the semantic binding hypothesis, which holds that both phonological-lexical and semantic-level representations support the coherence of phonological representations in STM. This hypothesis (Patterson et al., 1994), inspired by parallel-distributed-processing (PDP) models of language, holds that prior exposure to the sequence of speech sounds that comprises a known word influences the likelihood of those

speech sounds emerging together at recall, while the semantic activation that co-occurs with a word's phonological form over time provides a second source of constraint. A loosening of lexical/semantic constraints – when lexical/semantic activation for target items is relatively weak – should therefore particularly increase the likelihood of phonemes breaking away from list items and migrating between them. This pattern is evident in recall errors made by semantic dementia patients to words with degraded semantic representations whose phonological task performance is otherwise normal ('mint, rug' → 'rint, mug') (Patterson et al., 1994; Jefferies et al., 2005; Majerus, Norris, & Patterson, 2007) and in errors to words and nonwords when they are mixed together in a list (Jefferies et al., 2006a). Yet there is difficulty in establishing whether such effects are semantic or largely lexical (since the contribution of phonological-lexical and semantic information from known words is often confounded) (Jefferies et al., 2006a; Papagno, Vernice, & Cecchetto, 2013).

One way to test for purely semantic effects is to examine the phonological coherence of lists while manipulating the degree of semantic activation during encoding. Acheson, MacDonald, & Postle (2011) disrupted semantic processing for list items using irrelevant category judgments to pictures presented concurrently. They found increased item order errors in ISR for concrete words but not for nonwords, relative to non-semantic orientation judgments. While that study supports the view that semantic activation influences serial ordering in STM, compatible with language-based accounts, the mechanism underpinning the effect remains unclear for two reasons: (i) dual-task testing continued during recall itself, which could disrupt semantically-driven redintegrative processes; and (ii) the increase in item order errors during semantic categorisation may have reflected an increase in phoneme movement at the sub-item level (in line with the semantic binding account). Given the nature

of the stimuli and strategic editing of responses to produce real words, such phoneme movement could have produced whole-item order errors.

The present study

The following experiments took a similar approach to Acheson et al. (2011) but addressed the question of whether influencing semantic activation of items *at encoding* would impact upon their phonological coherence in subsequent serial recall (i.e., ordering at the sub-item level), as predicted by the semantic binding hypothesis. Rather than manipulate ISR stimuli or disrupt semantic activation with an unrelated task, we biased the activation of language representations with encoding tasks that directed attention to different aspects of the stimuli. Word lists were carefully constructed to enable tracking of phoneme migrations between words without their potential categorisation as whole word movement and to match for linguistic properties between list sets. Since, in each experiment, the word properties and recall task were matched between categorisation conditions, any difference in ISR between conditions would be attributable to the encoding state (differences which would not be expected in the case of a redintegration mechanism operating in isolation). In two experiments (Experiments 1 and 3), participants categorised spoken words according to a semantic or phonological property ('natural or man-made' or 'long or short vowel' respectively) – or, in Experiment 3, also a perceptual property ('high or low pitch') – and, after categorisation of the sixth word, verbally recalled the stimuli in sequence. In each case, participants were told which categorisation decision to make prior to the first list item to minimise interference with the phonological trace to be recalled. Following the semantic binding hypothesis, we predicted more phonologically coherent item recall – measurable in terms of *fewer phonologically-related errors* (i.e., where phonemes have broken away from the target item, and may have recombined with phonemes from another item) alongside more

accurate recall overall – for lists where semantic processing was enhanced during encoding. Experiment 2 involved ISR without any categorisation tasks during encoding to provide a baseline for Experiment 1. This established the direction of categorisation effects on recall (i.e., increased phonological coherence following semantic encoding or decreased coherence following phonological encoding).

Experiment 1: Semantic vs. Phonological encoding

Experiment 1 tested whether augmenting semantic activation of the list items at the point of encoding, with the use of a semantic categorisation task, would lead to more phonologically coherent recall than following a categorisation task directing attention to the words' phonological properties.

Method

Participants: Participants were 24 native British English undergraduates with normal hearing from the University of York, aged 18-21 years, who took part in exchange for course credit.

Stimuli: Stimuli were 60 lists of six auditorily-presented monosyllabic English nouns with a CVC/CCVC structure, selected from the MRC Psycholinguistic Database (http://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa_mrc.htm) on the basis of their suitability for categorisation, and being of similar imageability and lexical frequency (see list properties detailed below). Words with homophones of a higher lexical frequency were excluded. Word stimuli were recorded in a sound-attenuated booth, spoken by a male British English speaker with flat intonation. Lists were constructed such that no phoneme was

repeated in the same syllabic position[1] and to ensure a similar ratio of yes:no responses for categorisation. 180 words were used to create the first 30 lists: The next 30 were formed by recombining the same 180 words into new lists (imageability M = 565.43, SD = 46.14 and M = 571.48, SD = 40.70 and frequency M = 43.78, SD = 75.61 and M = 40.98, SD = 71.50, according to the MRC Psycholinguistic Database norms (Coltheart, 1981), for the two sets of lists assigned to semantic and phonological conditions, counterbalanced across participants). Each word was thus presented twice over the experiment. An additional 30 nouns not used in the main experiment formed five practice lists.

Procedure: The 60 test lists, following five practice lists, were presented in E-Prime. Lists were delivered in blocks of 15, separated by rest breaks. Before the practice task, short vowel sounds were identified to the participant as æ e ɪ ɒ ʌ and illustrative yes/no examples for both categorisations were provided (e.g., no to "sheep" and yes to "pan" for the short vowel decision and yes to "sheep" and no to "pan" for the natural decision). Participants were asked to categorise each word as quickly and accurately as possible and, at the end of the list, attempt to verbally recall in order all six items. Instructions were to respond with anything they felt they might have heard, even if unsure, and to skip items which they could not recall at all (to avoid interference from target-unrelated responses).

Participants wore a headset with a microphone connected to a digital recorder to listen to and recall the words. At the beginning of a trial, a screen informing participants of the categorisation to be made for the following six words was displayed. Trials alternated between semantic categorisation ('Instruction: Is it a natural thing?') and phonological categorisation ('Instruction: Does it have a short vowel sound?'). The instruction remained on screen until the participant pressed a key to continue. A fixation cross was displayed from 1 s

[1] Our list construction strategy was to avoid repetition of phonemes in the same syllable position based on the tendency of phonemes to retain their syllabic position (Ellis, 1980)

before first word onset until the end of the list. Participants made their respective yes/no decision after each word with the keyboard, which cued the next list item. After the sixth item response, a screen displaying 'now recall the list in order' prompted verbal recall. List order was fixed but categorisation order was swapped for half of the participants (i.e., half began with the phonological rather than semantic condition).

ISR analysis: Verbal responses were transcribed phonemically. When fewer than six responses were given on a trial, whole item omissions were positioned within the transcript in a way that minimised the error score. For example, if five responses were produced and these largely corresponded with the second through to the sixth target word respectively, responses would be transcribed as attempts at the second through to the sixth targets. In the few instances where participants gave seven rather than six responses, the seventh response was omitted from analyses.

A random subset of four data sets was independently transcribed by a second rater. Transcriptions were over 99% in agreement. ISR responses were automatically classified from the transcription according to the categories shown in Table 1[2]. The choice to classify errors item-by-item was encouraged by our use of simple monosyllabic words. Participants tend to strategically edit out potential nonword responses in pure word lists (Baars, Motley, & MacKay, 1975; Jefferies et al., 2009) and thus phoneme migration responses were likely to be rare and moderated by the number of opportunities for phoneme migrations to create real words within a given list. By employing a metric that classified each item-level response according to whether phoneme migration(s) had occurred at all (e.g., whether comprised of one migrated phoneme combined with correct-in-position phoneme(s) or three migrated phonemes), sensitivity to changes in phonological integrity at the word level was maximised.

[2] Manual coding was only employed to identify semantic errors: Responses automatically identified as phonologically unrelated were coded as semantic errors if the response was closely semantically related to one of the target list words.

TABLE 1

We analysed ISR response categories that constituted at least 1% of possible ISR responses and compared the rates of each error type across the categorisation tasks using paired sample *t*-tests.

Results and discussion

One participant was excluded on the basis of unusually poor ISR performance with, on average, only one item out of six correct in each list (two standard deviations below average group correct-in-position performance). A further five participants with the greatest differences in categorisation performance were also excluded (whose phonological categorisation accuracy differed by more than 14% from the semantic task; in each case their phonological decision accuracy was below 70%), allowing categorisation accuracy, reaction time (RT) and weighted reaction time (WRT) (reaction time divided by proportion correct: Townsend & Ashby, 1978) measures to be closely matched. This ensured that any differences in ISR performance according to categorisation condition could not be attributed to overall differences in the difficulty or time allocated to the semantic and phonological judgments. Accordingly, there was no difference in accuracy [Semantic $M = 89\%$; $SD = 5.68$; Phonological $M = 88\%$; $SD = 7.15$], mean reaction times [Semantic $M = 1797$ ms; $SD = 301.20$; Phonological $M = 1774$ ms; $SD = 255.61$] or weighted RT [Semantic $M = 2026.52$ ms; $SD = 291.30$; Phonological $M = 2026.64$ ms; $SD = 370.79$] for the semantic and phonological decisions for participants taken into the analyses [paired *t*-tests: all $p > .33$].

Analyses were performed on ISR data from these 18 participants[3]. Table 2 shows the percentage of ISR responses of each type at the item level, for each categorisation task. Semantic errors, extra-list intrusions and phonologically-unrelated response errors each

[3] The pattern of statistical outcomes for the ISR results with all participants from Experiment 1 included was the same as the reported data.

accounted for less than 1% of possible responses and thus were not analysed: their descriptive data is provided for completeness only.

TABLE 2

As predicted, participants recalled significantly more words correct in position (CIP) following semantic than phonological categorisation (Table 2). In terms of ISR errors, this difference in recall accuracy reflected the production of significantly fewer recombination errors following semantic categorisation compared to phonological categorisation (Table 2). No other analysed error type was significantly influenced by categorisation condition (all $p >$ .15).

This pattern of data indicates that semantic analysis of the words had a protective effect on their phonological integrity in subsequent recall, in terms of a greater percentage of items recalled and fewer recombination error responses, compared to analysis of their phonological properties. While item integrity was differentially affected by the semantic condition, item order was not – unlike Acheson et al.'s data but congruent with lexical/semantic knowledge primarily supporting item coherence/identity (Patterson et al., 1994; Poirier & Saint-Aubin, 1995; Hulme et al., 1997; Saint-Aubin & Poirier, 1999, 2000; Gathercole, Pickering, Hall, & Peaker, 2001; Jefferies et al., 2006a; Jefferies, Frankish, & Lambon Ralph, 2006b). These categorisation effects cannot be readily explained in terms of differences in attention or encoding time, since performance on the two categorisation tasks was equivalent in terms of average accuracy, RT and weighted RT. This pattern of results is compatible with a semantic binding account. However, since only phonological and semantic judgments were tested, it is possible phonological categorisation had a disruptive effect on maintenance in ISR that impacted phonological integrity, as opposed to semantic categorisations having a protective effect (e.g., the vowel judgments may have reduced

encoding of the item consonants, which are more vulnerable to migration; Jefferies et al., 2006a). We therefore ran a second experiment that did not involve concurrent categorisation during encoding, providing 'baseline' ISR data to compare with the semantic and phonological categorisation conditions from Experiment 1.

Experiment 2: Supplementary Baseline ISR comparison

We tested baseline ISR performance for the same set of stimuli without categorisation in a new set of participants to confirm the direction of the semantic and phonological effects in Experiment 1. Specifically, the difference between categorisation conditions might have reflected improved phonological binding following semantic categorisation or weakened phonological binding following phonological categorisation (or both). While we expected that the overall capacity for recall would be superior without an encoding task diverting attention away from ongoing rehearsal of previously-presented items, the semantic binding account would predict a higher percentage of phoneme recombination errors in standard ISR compared with ISR following semantic categorisation, since this manipulation should strengthen semantic support for phoneme ordering. In contrast, phoneme recombinations might occur at an equivalent rate in baseline ISR and following phonological categorisation.

Method

Participants: Eighteen British English students, aged 18-22 years, took part in the experiment. These participants did not take part in Experiment 1, allowing a between-subjects comparison of ISR performance with and without concurrent categorisation. The data in Experiment 2 are therefore not influenced by participants' experience of categorisation at encoding.

Stimuli: Since the word stimuli had been repeated in the second half of Experiment 1 in recombined ISR lists, this experiment used the first 30 lists containing non-repeated items that had been categorised once, *either* semantically or phonologically.

Procedure: Trial procedure was identical to Experiment 1 with the exception that the onset of word presentation was fixed to a rate of 1770 ms and, instead of a categorisation instruction, each trial began with the displayed instruction "Press SPACE BAR as quickly as possible after you hear each list item". Presentation rate was fixed in order to remove a potential ISR advantage in the baseline condition attributable to a faster self-paced presentation rate (in the absence of item decisions). The chosen inter-stimulus interval matched the average word onset intervals (i.e., RTs) in Experiment 1, as an alternative to using a faster presentation rate of one second, which is the typical rate tested in ISR tasks. Participants completed two practice trials to familiarise them with the ISR procedure before commencing the experiment.

ISR analysis: ISR transcription and coding procedures were identical to Experiment 1.

We compared the percentage of responses in each of the response categories for baseline ISR in Experiment 2 with equivalent data from Experiment 1 in two parallel analyses, using independent t-tests. First, in Analysis 1, we compared the 30 lists in the baseline ISR task with the 30 lists presented in each of the semantic and phonological conditions in Experiment 1 (i.e., all 60 lists that were presented in Experiment 1 were included in this analysis). Secondly, in Analysis 2, we controlled for task learning and fatigue effects by comparing the 30 lists in the baseline ISR task with the same first 30 lists in Experiment 1 (incorporating 15 lists tested with each categorisation type).

Results and discussion

TABLE 3

Both baseline comparisons revealed better recall without a concurrent categorisation task, with significantly more words recalled in position, fewer order errors and fewer omissions than those in Experiment 1 (Table 3). Importantly, however, despite being an objectively easier ISR task with no distraction from active maintenance and accordingly producing fewer errors overall, the percentage of recombination errors was similar to the phonological condition in Experiment 1, and significantly *higher* than for the semantic encoding condition (Table 3). Phonologically related non-recombination errors, however, were similar to both categorisation conditions.

The lower percentage of recombinations in the semantic than the no-categorisation condition (in the absence of meaningful shifts in other incorrect response categories) supports the suggestion that active recruitment of items' semantic representations in the semantic encoding task had a relatively stabilising effect on the targets' phonological trace and allows us to reject an alternative explanation by which the phoneme judgments in Experiment 1 weakened phoneme position encoding. In addition, the finding that the percentage of phonologically-related incorrect responses was not smaller in the baseline control discourages an interpretation that the greater phonological coherence of ISR in the semantic condition stemmed from that somehow being the easier task.

Experiment 3: Experiment 1 replication with perceptual task extension

To test the reliability of a semantic encoding enhancement for phonological stability in STM seen in Experiment 1, and to assuage any concerns about conclusions relying upon the comparison of data in Experiments 1 and 2, we ran one more experiment in a new set of

participants to test whether (i) the semantic encoding condition's recall advantage over the phonological encoding condition replicated and (ii) whether the semantic advantage remained in comparison to a different matched decision.

In addition to testing ISR following the same semantic and phonological categorisations at encoding as participants did in Experiment 1, this experiment tested the impact of categorising spoken items according to a perceptual property, specifically of high or low pitch. Categorising words according to this suprasegmental psychoacoustic property should encourage attention to the whole auditory signal without promoting attention to its linguistic features. We reasoned that if semantic categorisation again yielded more phonologically coherent recall than phonological categorisation and it also yielded more robust recall than a matched non-linguistic pitch categorisation condition, this would be good corroborating evidence for the semantic encoding benefit.

Method

Participants: A new set of 24 native British English participants (aged 19-34) took part in the experiment.

Stimuli: To accommodate the pitch condition, the same 180 words from the previous experiments were recombined to add a third set of 30 lists, producing 90 trial lists in total (i.e., 30 lists per encoding condition). To create the pitch manipulation, the pitch contours of sound files used in the previous experiments were altered by ±10 Hz in Praat, and the low and high pitch files were distributed so that they had 50:50 occurrence within a categorisation set (and so that no ISR list contained all high or all low pitch files). The size for these pitch changes was determined by pilot testing of categorisation performance alone (n = 16) which found that semantic, phonological and pitch categorisations of 10 Hz (but not 20 Hz) pitch-shifted files had similar accuracy.

Procedure: Prior to the experimental task, participants practiced each of the categorisations with six words not used in the main experiment followed by practice ISR trials for each categorisation condition. The 90 test lists were presented in six blocks of 15 lists via E-Prime. Since there were now three categorisation conditions, lists were grouped into categorisation mini-blocks of five lists, rather than cycled between conditions every trial (as in Experiment 1), to reduce cognitive load from switching between three task sets. Pitch decisions were prompted with "Is it the low or high pitch?" on screen at the start of the trial. Trial procedure was otherwise identical to Experiment 1. While ISR list order was fixed, the categorisations performed on a given list were counterbalanced across participants.

ISR Analysis: Transcription and ISR response coding was performed identically to the previous experiments. As per the previous experiments, response categories that were sufficiently frequent to permit statistical analysis (at least 1% of ISR responses) were submitted to repeated-measures ANOVAs with a single within-subject factor of categorisation condition (semantic, phonological, pitch). Greenhouse-Geisser correction of degrees of freedom was applied where relevant.

Results and discussion

Since we were primarily interested in the new comparison of semantic encoding with the perceptual encoding pitch condition, we included participants on the basis of similar semantic and pitch weighted RTs, to leave 18 participants (in practice this involved excluding those with accuracy weighted RT differences of over 500 ms between the pitch and semantic conditions; the pitch categorisation accuracy in these excluded participants differed from the semantic task by more than 20%)[4]. In these participants, semantic categorisation accuracy was significantly higher than the other two conditions [F (2, 34) = 14.13, p < .001; Accuracy:

[4] The pattern of statistical outcomes for the ISR results with all participants from Experiment 3 included was the same as the reported data.

Pitch M = 83%, SD = 2.40; Semantic M = 94%, SD = 0.57; Phonological M = 84%, SD = 2.57], while pitch categorisations were faster than both semantic and phonological conditions [F (2, 34) = 36.86, p < .001; RTs: Pitch M = 1478 ms, SD = 96.37; Semantic M = 1677, SD = 89.09; Phonological M = 1731.34, SD = 95.30]. These categorisation times were quicker than those in Experiment 1, probably related to the categorisation conditions being tested in mini-blocks. This structure may have contributed to accuracy and reaction times trading off differently between conditions, with pitch categorisation faster but less accurate than semantic categorisation; for this reason, accuracy-weighted RT was used as the measure to match performance across the categorisation tasks. Importantly for this experiment, however, while phonological categorisation was overall less efficient [F (2, 34) = 13.03, p < .001; Weighted RTs: Pitch M = 1804 ms, SD = 159.00; Semantic M = 1778 ms, SD = 92.90; Phonological M = 2050 ms, SD = 172.15], semantic and pitch weighted RTs were matched [p = .62].

Table 4 shows the percentage of responses of each type in ISR at the item level, for each categorisation task. Semantic errors, extra-list intrusions and phonologically-unrelated response errors each accounted for less than 1% of possible responses and thus were not analysed: Their descriptive data is provided for completeness only.

TABLE 4

Consistent with Experiment 1, participants correctly recalled a significantly higher percentage of words in position (CIP) following semantic than phonological categorisation [t(17) = 3.15, p < .001; see Table 4]. However, overall accuracy differences between the semantic and pitch conditions did not reach significance (p = .17). In terms of ISR errors, on the other hand, the semantic condition elicited significantly fewer phonologically-related incorrect productions than both the phonological and pitch conditions. Specifically,

recombination errors were, as we previously found, significantly reduced in the semantic condition compared to phonological and pitch conditions [semantic vs. phonological, t (17) = -2.95, $p < .01$; semantic vs. pitch, t (17) = -3.87, $p < .01$; phonological vs. pitch, t (17) = -1.41, *ns*]. In addition, phonologically-related non-recombination errors were also significantly reduced in the semantic condition compared to phonological and pitch conditions [semantic vs. phonological, t (17) = -2.69, $p < .05$; semantic vs. pitch, t (17) = -2.37, $p < .05$; phonological vs. pitch, t (17) = 1.16, *ns*]. Meanwhile, encoding condition did not modulate the percentage of item order errors or omissions (all $p > .14$).

This pattern of data largely mirrors that found in Experiment 1; the exception being that phonologically-related non-recombinations were also reduced following semantic encoding. A shift in phonologically-related non-recombination errors (i.e., a change in partially-correct responses) alongside recombination errors is also consistent with a semantic binding account (phoneme recombinations and phonologically-related non-recombination errors tend to pattern together: see Jefferies et al., 2006a, 2009; Patterson et al., 1994).

General Discussion

This series of experiments shows that orienting attention to the semantic features of words benefits their phonological coherence in verbal STM. Relative to matched phonological decisions, perceptual decisions and baseline ISR without a concurrent categorisation task, semantic categorisation of words presented for serial recall made the words less vulnerable to breaking apart and their phonemes recombining. The accuracy of ISR was greater overall following semantic categorisation and this effect was explained by a reduction in phoneme recombination errors in all experiments (and in Experiment 3, also by fewer phonologically-related non-recombination errors) compatible with predictions of the

semantic binding hypothesis. We can be confident that the semantic categorisation task was responsible for producing more phonologically coherent recall since the targets did not differ between conditions and categorisation performance (accuracy and time on task) was matched between the phonological and semantic tasks. Moreover, the comparison of semantic and phonological categorisation conditions (Experiment 1) with baseline ISR in the absence of concurrent categorisation (Experiment 2) suggests that semantic encoding specifically benefitted phonological coherence. The rate of recombination errors was equivalent in the phonological categorisation and baseline ISR conditions: these errors were specifically reduced by semantic categorisation. This challenges alternative explanations for the difference between semantic and phonological conditions, by which phoneme recombination rates reflected accuracy-related opportunities for migration errors or a relative *increase* in recombination errors in the phonological encoding condition following attention to sub-item fragments.

Our findings are broadly compatible with language-based accounts of verbal STM (Patterson et al., 1994; N. Martin & Saffran, 1997; Romani et al., 2008; Acheson & MacDonald, 2009). In subtly different ways, these propose that verbal STM arises via an interaction of temporary phonological activations with existing semantic representations. More specifically, the semantic binding hypothesis predicts that semantic activation reduces phoneme migration rates. This study provides the first clear demonstration of this effect in healthy adults since previous studies have either manipulated semantic variables but focused on whole item recall (Bourassa & Besner, 1994; Poirier & Saint-Aubin, 1995; Walker & Hulme, 1999; Tse & Altarriba, 2007; Romani et al., 2008) or have revealed effects of lexicality and/or frequency on phoneme ordering errors without disentangling the contributions of phonological-lexical and semantic knowledge (Patterson et al., 1994;

Jefferies et al., 2006a; Hoffman et al., 2009). We propose that familiar sequences of phonemes constituting phonological-lexical representations for the word targets were maintained within the phonological system during the ISR task (for example, through interactive-activation of acoustic and articulatory codes; Jacquemot & Scott, 2006; Plaut & Kello, 1999). The veracity of this phoneme sequence benefitted from interaction with the semantic system particularly after semantic categorisation during encoding – i.e., stronger semantic activation for these lists increased pattern completion effects within the phonological system. Within the context of maintaining these representations in a novel sequence beyond typical span length, the phonological system was sufficiently challenged such that phonemes were vulnerable to breaking away from their respective item (i.e., in the case of recombination and phonologically-related non-recombination errors) and intruding into other word positions (i.e., in the case of recombination errors) in the absence of strong semantic support.

We consider that such language-perspectives on STM provide a fuller explanation of the present data than recall-based accounts. Redintegration is thought to have its key effects on serial recall at the whole-item lexical level via reconstruction at the recall stage (Gathercole et al., 2001) while semantic binding has its key effects on serial order at the phoneme level via activation at encoding (Jefferies et al., 2006a). Semantic binding specifically predicts semantic activation should help to protect word items from their phonemes breaking away and recombining, while redintegration-based accounts do not. There is no strong *a priori* reason to predict that redintegration alone – that is, item-based reconstruction of degraded trace output from the phonological loop – would specifically modulate phoneme recombination errors. Nevertheless, this mechanism could produce the pattern we observed with several additional assumptions: (i) that as the phonological trace

decays, the position of phonemes within the sequence becomes less certain; (ii) that redintegration, through a comparison of this degraded phonological trace with long-term representations of lexical-phonological forms of words, reduces this uncertainty and (iii) that rich semantic processing at the point of encoding makes it easier for redintegration mechanisms to select the appropriate lexical-phonological forms to use during reconstruction. However, previous studies using mixed lists of words and nonwords have found that not only does the presence of 'unbound' nonword phonemes increase phoneme recombination errors for words but, in addition, the presence of words reduces phoneme migrations for *nonwords*. Since redintegration is item-based, and it should not be possible to engage this mechanism for nonwords, it is less clear how the redintegration framework could be modified to account for this finding (see Jefferies et al., 2006a, 2009). We propose that strong semantic constraints improve the stability of the phonological trace directly during maintenance while acknowledging that a semantic redintegration mechanism may also improve STM performance at the point of recall.

The present work builds on studies which have observed recall advantages when semantic processing is emphasised at encoding, both in STM and in LTM. Importantly, however, these effects of semantic encoding have been examined at the whole-item level, and not the phoneme level. For example, Campoy and Baddeley (2008) found better serial order memory for closed sets of stimuli when participants were explicitly instructed to focus on the meaning of visually-presented words and attempt to link them, compared to participants instructed to adopt a purely phonological encoding strategy or those who had no instruction (see also Hanley & Bakopoulou, 2003). In their study, semantic encoding yielded recall that was relatively resistant to effects of phonological similarity and word-length, and this was attributed to access to a *separate* semantic store (either in long-term memory, consciously

accessed via an episodic buffer, or via a separate semantic short-term store, cf. Campoy & Baddeley, 2008). From this perspective, durable semantic traces (cf. Baddeley & Ecob, 1970), may have informed redintegrative processes by providing an additional cue to lexical candidates from long-term memory at recall (following Saint-Aubin & Poirier, 1999, 2000). Their study, however, was not designed to examine the phonological coherence of recall, which is how and where we observe the present encoding-based effects. The serial order reconstruction task they used emphasised retention of the order of whole items and did not permit phoneme migrations.

Better recall of stimuli following semantic encoding is also associated with the levels of processing (LOP) framework in studies of LTM (Craik & Lockhart, 1972; Craik & Tulving, 1975). Our experimental manipulations have some similarity with LOP experiments, since participants made semantic or phonological categorisations of each word item during encoding. In the LOP literature, using recall measures at the level of individual items, semantic encoding is strongly associated with an advantage for retrieval of whole items from LTM, while effects on short-term maintenance are rare (e.g., Mazuryk & Lockhart, 1974; Rose & Craik, 2012). However, we have shown that when recall measures are sensitive to phoneme level accuracy, encouraging attention to the semantic properties of words at encoding benefits STM performance (relative to phonological or perceptual properties or no emphasis).

It appears that there are at least two independent mechanisms which underpin superior memory performance following semantic encoding, which differentially impact STM and LTM tasks: In the short-term, recall is strongly influenced by the still-active phonological trace, and semantic encoding effects improve the coherence of this phonological sequence; however, ISR is primarily influenced by phonological manipulations. In the longer term,

when the phonological trace is no longer available, whole item retrieval is strongly influenced by the distinctiveness of LTM traces and the ease with which targets can be differentiated from competing long-term representations (Jacoby & Craik, 1976). Semantic encoding is thought to have a strong effect on the *distinctiveness and durability* of LTM, according to LOP theory (Craik & Lockhart, 1972; Jacoby & Craik, 1976). For example, semantic encoding may promote visual imagery of the concept that the word denotes (Durso & Johnson, 1980), which will help to differentiate a representation of the target word in episodic LTM from other possible targets. These processes will also potentially influence STM when the remnant phonological trace is incomplete, since distinctive items will be more available for redintegration. In summary, we suggest that the semantic categorisation advantage in short-term recall is the combined result of a more resilient phonological trace (resulting from semantic binding) supplemented by the availability of stronger/more distinctive episodic memory for the targets at the point of recall.

Finally, we consider why Acheson et al., (2011) found that a semantic secondary task produced serial order errors at the whole item level (rather than at the sub-item phoneme level), by noting relevant design differences. First, this study used six item lists, which is at the limits of or just beyond typical word span, whereas Acheson et al.'s participants were tested on up to five items. We know that non-target responses (including phoneme migrations) proliferate when challenged beyond span (Unsworth & Engle, 2007), while shorter lists are more likely to produce whole item order errors. Secondly, our lists were constructed to avoid target phoneme repetition within a list, to optimise tracking of recombinations. Thirdly, the nature of our concurrent task differed: Acheson et al.'s semantic task was designed to be disruptive, was potentially more difficult than their comparison task (Acheson et al., 2011, p. 56), and placed attentional priority on picture categorisation rather

than serial encoding and recall, while our categorisation tasks drew attention to different aspects of the stimuli-to-be-encoded. Since item order accuracy is partially reliant on executive mechanisms distinct from the language system, while phoneme order accuracy within items is strongly supported by linguistic knowledge (Majerus, 2009; Hoffman, Jefferies, Ehsan, Jones, & Lambon Ralph, 2012), attentional differences could account for i) the preponderance of item order errors in Acheson et al.'s study and ii) the smaller number of item order errors in our baseline condition compared to the categorisation conditions, given that there was no concurrent task to distract from serial order encoding. Lastly, the manipulation in the present study focused on the encoding stage, while the concurrent tasks used by Acheson et al. extended throughout encoding and retrieval. Therefore, Acheson et al. may have disrupted semantic redintegration at recall (potentially as well as semantic phoneme binding).

In summary, although language-based accounts of verbal STM have been gaining ground and form the basis of several recent and influential models of verbal STM (N. Martin & Saffran, 1997; R. C. Martin, Lesch, & Bartha, 1999; Buchsbaum & Esposito, 2008; Acheson & MacDonald, 2009; Majerus, 2013), the influence of linguistic representations on the stability of the phonological trace has not been widely studied. The current study provides unique evidence showing that semantic processing during encoding influences the stability of the phonological trace, even in healthy participants (providing evidence that is complementary to previous studies of patients with semantic dementia).

References

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

Acheson, D. J., MacDonald, M. C., & Postle, B. R. (2011). The effect of concurrent semantic categorization on delayed serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 44–59.

Baars, B. J., Motley, M. T., & MacKay, D. G. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, *14*, 382–391.

Baddeley, A. D., & Ecob, J. R. (1970). Simultaneous acoustic and semantic coding in short-term memory. *Nature*, *227*, 287–289.

Bourassa, D. C., & Besner, D. (1994). Beyond the articulatory loop: A semantic contribution to serial order recall of subspan lists. *Psychonomic Bulletin & Review*, *1*, 122–125.

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

Campoy, G., & Baddeley, A. D. (2008). Phonological and semantic strategies in immediate serial recall. *Memory*, *16*, 329–340.

Caza, N., & Belleville, S. (1999). Semantic contribution to immediate serial recall using an unlimited set of items: evidence for a multi-level capacity view of short-term memory. *International Journal of Psychology*, *34*, 334–338.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology*, *33A*, 497–505.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*, 268–294.

Durso, F. T., & Johnson, M. K. (1980). The effects of orienting tasks on recognition, recall, and modality confusion of pictures and words. *Journal of Verbal Learning and Verbal Behavior*, *19*, 416–429.

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

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

Hanley, J. R., & Bakopoulou, E. (2003). Irrelevant speech, articulatory suppression, and phonological similarity: a test of the phonological loop model and the feature model. *Psychonomic Bulletin & Review*, *10*, 435–444.

Hoffman, P., Jefferies, E., Ehsan, S., Jones, R. W., & Lambon Ralph, M. A. (2009). Semantic memory is key to binding phonology: converging evidence from immediate serial recall in semantic dementia and healthy participants. *Neuropsychologia*, *47*, 747–760.

Hoffman, P., Jefferies, E., Ehsan, S., Jones, R. W., & Lambon Ralph, M. A. (2012). How does linguistic knowledge contribute to short-term memory? Contrasting effects of impaired semantic knowledge and executive control. *Aphasiology*, *26*, 383–403.

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

Hulme, C., Roodenrys, S., & Brown, G. (1995). The role of long term memory mechanisms in memory span. *British Journal of Psychology*, *86*, 527–536.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, Gordon, 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*, 1217–1232.

Jacoby, L. L., & Craik, F. I. M. (1976). Effects of Elaboration of Processing at Encoding and Retrieval : Trace Distinctiveness and Recovery of Initial Context. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing and human memory* (pp. 1–22). Hillsdale, NJ, USA: Erlbaum.

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

Jefferies, E., Frankish, C. R., & Lambon Ralph, M. A. (2006a). Lexical and semantic binding in verbal short-term memory. *Journal of Memory and Language*, *54*, 81–98.

Jefferies, E., Frankish, C. R., & Lambon Ralph, M. A. (2006b). Lexical and semantic influences on item and order memory in immediate serial recognition: evidence from a novel task. *Quarterly Journal of Experimental Psychology*, *59*, 949–964.

Jefferies, E., Frankish, C. R., & Noble, K. (2009). Lexical coherence in short-term memory: strategic reconstruction or "semantic glue"? *Quarterly Journal of Experimental Psychology*, *62*, 1967–1982.

Jefferies, E., Jones, R., Bateman, D., & Lambon Ralph, M. A. (2004). When does word meaning affect immediate serial recall in semantic dementia? *Cognitive, Affective & Behavioral Neuroscience*, *4*, 20–42.

Jefferies, E., Jones, R. W., Bateman, D., & Lambon Ralph, M. A. (2005). A semantic contribution to nonword recall? Evidence for intact phonological processes in semantic dementia. *Cognitive Neuropsychology*, *22*, 183–212.

Majerus, S. (2009). Verbal short-term memory and temporary activation of language representations: The importance of distinguishing item and order information. In A. Thorn & M. P. A. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 244–276). London, UK: Psychological Press.

Majerus, S. (2013). Language repetition and short-term memory: an integrative framework. *Frontiers in Human Neuroscience*, *7*, 357.

Majerus, S., Linden, M. Van Der, Mulder, L., Meulemans, T., & Peters, F. (2004). Verbal short-term memory reflects the sublexical organization of the phonological language network: Evidence from an incidental phonotactic learning paradigm. *Journal of Memory and Language*, *51*, 297–306.

Majerus, S., Norris, D. G., & Patterson, K. E. (2007). What does a patient with semantic dementia remember in verbal short-term memory? Order and sound but not words. *Cognitive Neuropsychology*, *24*, 131–151.

Majerus, S., & van der Linden, M. (2003). Long-term memory effects on verbal short-term memory: A replication study. *British Journal of Developmental Psychology*, *21*, 303–310.

Majerus, S., Van der Linden, M., Poncelet, M., & Metz-Lutz, M.-N. (2004). Can phonological and semantic short-term memory be dissociated? Further evidence from landau-kleffner syndrome. *Cognitive Neuropsychology*, *21*, 491–512.

Martin, N., & Saffran, E. M. (1997). Language and auditory-verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology*, *14*, 641–682.

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.

Mazuryk, G. F., & Lockhart, R. S. (1974). Negative recency and levels of processing in free recall. *Canadian Journal of Psychology*, *28*, 114–123.

Melby-Lervåg, M., & Hulme, C. (2010). Serial and free recall in children can be improved by training: evidence for the importance of phonological and semantic representations in immediate memory tasks. *Psychological Science*, *21*, 1694–700.

Papagno, C., Vernice, M., & Cecchetto, C. (2013). Phonology without semantics? Good enough for verbal short-term memory. Evidence from a patient with semantic dementia. *Cortex*, *49*, 626–636.

Patterson, K. E., Graham, N., & Hodges, J. R. (1994). The impact of semantic memory loss on phonological representations. *Journal of Cognitive Neuroscience*, *6*, 57–69.

Plaut, D. C., & Kello, C. T. (1999). The emergence of phonology from the interplay of speech comprehension and production: a distributed connectionist approach. In B. MacWhinney (Ed.), *The emergence of language* (pp. 381–415). Mahweh, NJ: Erlbaum.

Poirier, M., & Saint-Aubin, J. (1995). Memory for related and unrelated words: further evidence on the influence of semantic factors in immediate serial recall. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *48*, 384–404.

Romani, C., McAlpine, S., & Martin, R. C. (2008). Concreteness effects in different tasks: implications for models of short-term memory. *Quarterly Journal of Experimental Psychology*, *61*, 292–323.

Rose, N. S., & Craik, F. I. M. (2012). A processing approach to the working memory/long-term memory distinction: evidence from the levels-of-processing span task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1019–1029.

Saint-Aubin, J., & Poirier, M. (1999). The influence of long-term memory factors on immediate serial recall: an item and order analysis. *International Journal of Psychology*, *34*, 347–352.

Saint-Aubin, J., & Poirier, M. (2000). Immediate serial recall of words and nonwords: Tests of the retrieval-based hypothesis. *Psychonomic Bulletin & Review*, *7*, 332–340.

Savill, N., Ashton, J., Gugliuzza, J., Poole, C., Sim, Z., Ellis, A. W., & Jefferies, E. (2015). tDCS to temporoparietal cortex during familiarisation enhances the subsequent phonological coherence of nonwords in immediate serial recall. *Cortex*, *63*, 132–144.

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration. *Memory & Cognition*, *21*, 168–175.

Stuart, G., & Hulme, C. (2000). The effects of word co-occurance on short-term memory: Associative links in long-term memory affect short-term memory performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 796–802.

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*, 729–735.

Thorn, A. S. C., Gathercole, S. E., & Frankish, C. R. (2005). Redintegration 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*, 133–158.

Townsend, J. T., & Ashby, F. G. (1978). Methods of modeling capacity in simple processing systems. In J. Castellan & F. Restle (Eds.), *Cognitive Theory Vol. 3* (pp. 200–239). Hillsdale, NJ, USA: Erlbaum.

Tse, C.-S., & Altarriba, J. (2007). Testing the associative-link hypothesis in immediate serial recall: Evidence from word frequency and word imageability effects. *Memory*, *15*, 675–90.

Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: an examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, *133*, 1038–1066.

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

Table 1. Coding scheme for ISR responses

Response Category	**Criteria**	**Examples (where the category applies to both responses)**
Correct in position (CIP)	Target word recalled in correct serial position.	CAT, SHOP → 'cat, shop'
Item order error (ORD)	Target word recalled but in an incorrect list position.	CAT, SHOP → 'shop, cat'
Recombination error (RECOMB)	Incorrect response contained phonemes from more than one target word (maintaining the CVC position of the target phonemes).	CAT, SHOP → 'cap, shot/cot' (but not 'top/cash')
Non-recombination phonological error (NON)	Incorrect response contained 1 or 2 phonemes from one target word (maintaining the CVC position of the target phonemes).	CAT, SHOP → 'cab, shell' (but not 'tick/bush')
Semantic error (SEM)	Response was semantically but not phonologically related to one of the list items.	CAT, SHOP → 'mouse, till'
Extra-list intrusion (INT)	Response appeared in one of the previous six target lists, and did not contain phonemes from the current target list.	[any words from the previous six lists, that were not classified above]
Omission (OM)	Missing response.	CAT, SHOP → '… …'
Unrelated (UNR)	Response had no phonemes in common with any target (in relative CVC position), and was not present in one of the previous six lists.	CAT, SHOP → 'head, leaf'

Table 2. Percentage of ISR responses in Experiment 1 for each categorisation condition and their paired comparisons

	Phonological Categorisation		**Semantic Categorisation**				
Response Type	***M***	***SD***	***M***	***SD***	***t* (df=17)**	***p***	***d***
CIP	48.52	13.11	52.28	14.75	-2.351	**.031**	-0.278
ITEM ORDER	17.13	9.39	17.96	9.68	-0.758	*ns*	-0.090
OMISSION	15.49	13.29	13.74	12.86	1.496	*ns*	0.138
RECOMBINATION	9.97	4.62	7.59	4.35	3.264	**.005**	0.545
NON-RECOMB. PHON	7.53	4.20	6.82	4.67	1.025	*ns*	0.165
INTRUSION	0.49	0.78	0.86	0.61	-1.584	*ns*	-0.543
SEMANTIC	0.19	0.38	0.25	0.34	-0.511	*ns*	-0.175
UNRELATED	0.68	0.98	0.49	0.68	0.656	*ns*	0.225

Note. Mean and SD values relate to percentage of total ISR items, given to 2 d.p. *t* values related to paired sample *t*-tests, and Cohen's *d* effect sizes. Significant comparisons are highlighted in bold. CIP = correct in position. Non-recomb. phon = non-recombination phonological error. A full explanation of each response type is provided in Table 1.

Table 3. Percentage of baseline ISR responses in Experiment 2 and their comparisons with the categorisation conditions in Experiment 1

	Baseline ISR (%)		**Baseline vs PHON (df = 34)**				**Baseline vs SEM (df = 34)**			
			Analysis 1		**Analysis 2**		**Analysis 1**		**Analysis 2**	
	M	***SD***	***t***	***p***	***t***	***p***	***t***	***p***	***t***	***p***
CIP	63.43	10.66	3.743	**.001**	4.517	**.000**	2.598	**.014**	2.841	**.008**
ITEM ORDER	9.60	4.87	-3.336	**.006**	-3.330	**.003**	-3.814	**.001**	-3.036	**.005**
OMISSION	6.48	5.90	-2.630	**.015**	-2.948	**.007**	-2.176	**.040**	-2.575	**.017**
RECOMBINATION	11.42	5.08	0.896	*ns*	0.909	*ns*	2.427	**.021**	2.965	**.006**
NON-RECOMB. PHON	8.46	3.97	0.680	*ns*	0.686	*ns*	1.133	*ns*	0.663	*ns*
INTRUSION	0.22	0.28	-1.418	*ns*	-0.238	*ns*	-4.10	**<.001**	-3.198	**.003**
SEMANTIC	0.19	0.38	0	*ns*	0	*ns*	-0.403	*ns*	-0.393	*ns*
UNRELATED	0.22	0.58	-1.725	*ns*	-1.725	*ns*	-1.318	*ns*	-1.318	*ns*

Note. Mean and SD values relate to percentage of total ISR items in the baseline task. *t* and *p* values related to independent sample *t*-test comparisons with the baseline task with significant comparisons highlighted in bold. Analysis 1 compares the baseline with the full semantic and phonological conditions in Experiment 1; Analysis 2 is the comparison of the same 30 ISR lists (i.e., the first half of each categorisation condition).

Table 4. Percentage of ISR responses for each categorisation condition in Experiment 3

	Phonological Categorisation		**Semantic Categorisation**		**Pitch Categorisation**			
Response Type	***M***	***SD***	***M***	***SD***	***M***	***SD***	***p***	η_p^2
CIP	58.52	14.31	66.05	15.58	62.01	15.87	**.020**	.24
ITEM ORDER	13.09	6.48	11.30	5.98	11.51	5.37	*ns*	.07
OMISSION	12.10	14.07	10	12.64	9.63	13.09	*ns*	.12
RECOMBINATION	7.99	5.65	6.23	4.93	9.07	6.07	**.002**	.33
NON-RECOMB. PHON	6.98	4.66	5.43	4.29	6.45	3.50	**.015**	.23
INTRUSION	0.43	0.56	0.49	0.63	0.62	0.63	*ns*	.04
SEMANTIC	0.43	0.52	0.15	0.32	0.25	0.39	*ns*	.14
UNRELATED	0.46	0.81	0.34	0.74	0.46	0.81	*ns*	.13

Note. Mean and SD values relate to percentage of total ISR items, given to 2 d.p. *p* values relate to repeated measures ANOVA, and η_p^2 denotes partial Eta-squared effect sizes. Significant comparisons are highlighted in bold. CIP = correct in position. Non-recomb. phon = non-recombination phonological error. A full explanation of each response type is provided in Table 1.