Interference by process, not content, determines semantic auditory distraction

John E. Marsh *, Robert W. Hughes, Dylan M. Jones

School of Psychology, Cardiff University, Cardiff, Wales CF10 3AT, UK

Abstract

Distraction by irrelevant background sound of visually-based cognitive tasks illustrates the vulnerability of attentional selectivity across modalities. Four experiments centred on auditory distraction during tests of memory for visually-presented semantic information. Meaningful irrelevant speech disrupted the free recall of semantic category-exemplars more than meaningless irrelevant sound (Experiment 1). This effect was exacerbated when the irrelevant speech was semantically related to the to-be-remembered material (Experiment 2). Importantly, however, these effects of meaningfulness and semantic relatedness were shown to arise only when instructions emphasized recall by category rather than by serial order (Experiments 3 and 4). The results favor a process-oriented, rather than a structural, approach to the breakdown of attentional selectivity and forgetting: performance is impaired by the similarity of process brought to bear on the relevant and irrelevant material, not the similarity in item content.

Keywords:
Auditory distraction
Semantic interference
Selective attention
Interference-by-process
Semantic-category clustering

1. Introduction

One of the most influential constructs in memory research is interference: the ease with which items are retrieved from memory is dictated, at least in part, by other stimuli or events that are similar in some way to the target (see, e.g., Anderson, 2003; Baddeley, 1986; McGeoch, 1942; Nairen, 1990; Nairne, 2002; Neath, 2000). The classical, structuralist, view has been that such interference-by-similarity-of-content directly causes forgetting, that is, forgetting is a passive side-effect of structural changes that result from the storing of new, similar, events in memory (Anderson, 1983; Cowan, 1999; McGeoch, 1942; Mensink & Raaijmakers, 1988; Oberauer & Lange, 2008; Oberauer, Lange, & Engle, 2004; Salamé & Baddeley, 1982). However, an alternative, more functional, view is that ‘forgetting’ (or the impairment of retrieval) reflects the legacy of dynamic and adaptive selective attention processes (such as inhibition; e.g., Houghton & Tipper, 1994) that are designed to resolve conflict during the selection of candidates at retrieval (e.g., Anderson, 2003). Set within this quintessentially attentional approach to forgetting, the present article explores the nature of phenomena relating to impaired retrieval from memory due to distraction from irrelevant auditory events using the structuralist, interference-by-similarity-of-content, approach as a theoretical counterpoint.

One line of research in which a dynamic selective attention framework has been used to reconstrue putatively mnemonic phenomena is that concerned with the disruptive effects of to-be-ignored sound on visual-verbal serial recall whereby a list of around 6–8 verbal items (e.g., letters or digits) is to be recalled in strict serial order (the irrelevant sound effect—hereafter ISE—e.g., Colle & Welsh, 1976; Jones, Madden, & Miles, 1992; Jones & Tremblay, 2000; Salamé & Baddeley, 1982). The mere presence of background sound depresses serial recall appreciably, the weight of evidence favoring the view that the effect results from interference-by-process, and is not a passive side-effect of having similar items to remember and to ignore.

* Corresponding author. Tel.: +44 29 20 876788; fax: +44 029 20 874858.
E-mail address: marshje@cardiff.ac.uk (J.E. Marsh).
(Hughes & Jones, 2005; Jones & Tremblay, 2000). Specifically, this ‘classical’ ISE is thought to result from the obligatory, preattentive, seriation (or ordering) of sound sequences producing competition for the deliberate process of seriating the to-be-remembered items. Here we examine whether the principle of interference-by-process can be extended to a setting in which the focal memory task involves not serial processing but semantic retrieval strategies: Does the concurrence of similar semantic processing (rather than serial processing) applied to relevant and irrelevant material now dictate the form and degree of distraction? What little evidence there is seems to suggest that the structural accounts seem perfectly adequate and irrelevant material now dictate the form and degree of distraction. Does the concurrence of similar semantic processes producing competition for the deliberate process (or ordering) of sound sequences producing competition for the deliberate process of seriating the to-be-remembered items. Here we examine whether the principle of interference-by-process account rather paradigm-bound to serial short-term memory. The goals of the present series were to revisit the empirical signature of auditory distraction in the context of episodic short-term memory tasks that tap semantic memory processes (particularly given the paucity of studies on the issue) to establish the degree to which it is distinct from that found in serial recall, and to examine thereafter how such distinct phenomena might be reconciled with a dynamic process-oriented approach to interference.

1.1. Irrelevant sound effect in serial recall

The debate between the structuralist and process-based viewpoints can be observed in microcosm in a body of research showing that the presence of irrelevant, to-be-ignored sounds markedly increases forgetting in a (usually visually-presented) serial recall task (e.g., Colle & Welsh, 1976; Jones et al., 1992; Salamé & Baddeley, 1982). The conventional viewpoint, that forgetting can occur as a direct and passive consequence of the structural similarity between to-be-remembered and irrelevant episodes or stimuli (e.g., McGeoch, 1942), is evident in several theoretical accounts of the ISE that view it as a mere consequence of auditory stimuli gaining access to the same representational space as the to-be-remembered items (e.g., phonological store, Burgess & Hitch, 1992; Gathercole & Baddeley, 1993; Salamé & Baddeley, 1982; primary memory, Neath, 2000). Although these accounts differ in their detail of how interference arises, the important point for present purposes is that they are all examples of an interference-by-content approach: recall is impaired as a result of the similarity in identity (i.e., content) between to-be-remembered and to-be-ignored items.

Several strands of evidence converge to weaken the interference-by-content approach. First, non-speech sounds such as tones—which bear little or no resemblance to the to-be-remembered items—produce disruption similar in degree and kind to that from irrelevant speech (e.g., Jones & Macken, 1993; Neath & Surprenant, 2001). Second, the magnitude of disruption is unrelated to the degree of phonological similarity between to-be-remembered and to-be-ignored items (Jones & Macken, 1995; LeCompte & Shaibe, 1997; but see Hughes & Jones, 2005) thereby disconfirming the predictions of an early account based on the concept of a phonological store (Gathercole & Baddeley, 1993; Salamé & Baddeley, 1982). As a result of these findings, the phonological store account of the ISE has been modified and expressed computationally such that irrelevant speech disrupts a representation of order within the passive store rather than interfering with item representations (Norris, Baddeley, & Page, 2004; Page & Norris, 2003). However, problematic for any account that views irrelevant speech as disrupting the phonological store is recent evidence showing that rehearsal is a precondition for its expression (Jones, Macken, & Nicholls, 2004).

Third, the interference-by-content approach fails to acknowledge adequately the critical importance of the nature of focal task processing, the impairment of recall being chiefly determined by the co-existence of similar to-be-recalled and to-be-ignored items within a store. That is, they cannot account for why the ISE is only found if the focal task necessitates or tends to encourage a seriation process (e.g., serial rehearsal) and why the mere presence of similar content between the memory material and the sound is not sufficient (or necessary) for the effect (Beaman & Jones, 1997; Farley, Neath, Albbrinton, & Surprenant, 2007; Henson, Hartley, Burgess, Hitch, & Flude, 2003; Hughes, Vachon, & Jones, 2007; Perham, Banbury, & Jones, 2007).

Whilst the preoccupation of the interference-by-content approach is with item identity, on the interference-by-process account, the key determinant of the disruption in serial recall is the extent to which both the irrelevant sound and the focal memory task share similar seriation (or ordering) processes (Jones, 1993; Jones & Tremblay, 2000). A key observation underpinning this account is the changing-state effect (e.g., Jones et al., 1992) whereby a sound sequence—regardless of whether it comprises speech or non-speech—that exhibits abrupt changes in acoustic properties (e.g., “k v h g . . .”, or a sequence of tones changing in frequency) is invariably more disruptive than a continuous or repeating stimulus (e.g., “k k k k . . .”, or a repeated tone). On the interference-by-process account it is assumed that the preattentive perception of acoustic changes between segmentable elements in the sound yields cues as to the order of those elements as a by-product of primitive, acoustic-based, perceptual organization processes (cf. Bregman, 1990). These irrelevant order cues compete for—and hence impair—the deliberate seriation process (serial rehearsal) supporting ordered recall of the to-be-remembered items (Hughes & Jones, 2005; Jones, 1993). In support of this view, the ability to encode the order of stimuli in an attended changing-state auditory sequence predicts the degree to which that sequence is disruptive when presented as irrelevant sound during serial recall (Macken, Phelps, & Jones, in press).

In sum, results based on research using the serial recall paradigm favor a dynamic process-based approach (Jones & Tremblay, 2000). However, the phenomenon of interference-by-process seems highly specific to a particular process (seriation) and little evidence is available with respect to whether such conflict occurs between other types of processes. In the present study, therefore, we addressed whether the phenomenon extends to auditory distraction in the context of a focal task that is likely to be
dominated by semantic-based, rather than seriation, processes. This is a particularly pertinent issue in light of the fact that the little evidence there is on semantic auditory distraction effects suggests that the interference-by-content approach offers a perfectly adequate explanation.

1.2. Semantic auditory distraction: interference-by-content?

On the interference-by-process account of auditory distraction in serial recall, it is the processing of the precategorical, acoustic, attributes of the sound that is key to the disruption (e.g., Jones & Macken, 1993). Consistent with this view, neither the lexical-semantic content of the irrelevant sound (when speech is used) nor the similarity in terms of semantic content between the speech and the to-be-remembered list has any bearing on the magnitude of disruption (Buchner, Irmenn, & Erdfelder, 1996; Jones, Miles, & Page, 1990; LeCompte, Neely, & Wilson, 1997; Surprenant, Neath, & Bireta, 2007; but see Buchner, Rothermund, Wentura, & Mehl, 2004). Such findings are easily explicable on the interference-by-process account: The key process supporting serial recall is an articulatory-based seriation process, not a semantically-based one (e.g., Jones et al., 2004). Thus, the lexical-semantic attributes of irrelevant speech would not be expected to conflict with the focal seriation process. However, the absence of such lexical-semantic effects in the context of serial recall can also be accommodated within the interference-by-content approach. In the typical serial recall task, the to-be-recalled items (e.g., digits, letters) are relatively impoverished in terms of semantic content. Thus, the representations of to-be-recalled items—devoid of rich semantic content—may not be susceptible to degradation or retrieval-confusion as a result of activated semantic representations of the irrelevant speech items.

In line with the interference-by-content approach, the results of a small number of studies suggest that when the items are semantically rich, recall is indeed impaired by the semantic attributes of the irrelevant sound (Beaman, 2004; Jones et al., 1990; Martin, Wogalter, & Forlano, 1988; Neely & LeCompte, 1999; Oswald, Tremblay, & Surprenant, 2003). For example, in a category-exemplar recall task, in which a list of, say, 16 semantically-rich items (nouns) taken from a single semantic category are presented for free recall, the semantic similarity between the relevant and irrelevant items impairs performance (Beaman, 2004; Neely & LeCompte, 1999). The free recall of relatively low-dominance category-exemplars (e.g., “avocado”) is disrupted (as reflected in omission errors) more by related, high-dominance, irrelevant category-items (that are not included in the to-be-remembered list; e.g., “apple”) than by high-dominance, categorically-unrelated, irrelevant items (e.g., “hammer”). Such results seem to be readily explained within an interference-by-content approach: The semantic representations of the to-be-recalled items may be degraded or otherwise made less accessible as a function of their semantic similarity to the irrelevant items (e.g., Anderson, 1983; Oberauer & Lange, 2008; Oberauer et al., 2004; Rundus, 1973).

However, our central contention in this paper is that these semantic auditory distraction effects may also be amenable to, and indeed be better explained by, an interference-by-process analysis (cf. Marsh, Hughes, & Jones, submitted for publication, 2008). The starting point for this analysis is that, compared to serial recall, the semantic richness of the items in category-exemplar recall, as well as the longer list-length, demotes the likelihood of a seriation strategy and instead promotes the use of semantic-based organization processes. It is this shift in the nature of the dominant process strategy used to support performance in the focal task, not simply the semantic richness of the to-be-recalled items, that renders such tasks vulnerable to competition from the processing of the semantic content of the irrelevant sound. Thus, in contrast to the interference-by-content approach, it is not the mere co-activation or co-registration of relevant and irrelevant semantic representations that impairs performance. Rather, it is the integrity of dynamic, semantic-based, organizational processes engaged in support of retrieval that is compromised by the semantic processing of the irrelevant sound. In this way, semantic auditory distraction may be regarded as an extension of the general case of interference-by-process: It may reflect a difficulty in selecting amongst two sets of semantic representations both of which represent plausible candidates for populating the semantic-organizational skill used in the focal task, just as with two sets of serial representations in the classical irrelevant sound effect.

In the studies that follow, we scrutinize further the nature of semantic auditory distraction and attempt to clarify the extent to which semantic-organizational processing in the focal task—as opposed to the mere semantic richness of the to-be-recalled items—is responsible for disruption by the lexical-semantic attributes of irrelevant speech. We use a setting in which a relatively long list of 32 exemplars (e.g., “strawberry”, “pigeon”, etc.) drawn from a smaller set of 4 semantic categories (e.g., “Fruit”, “Birds”, etc.) are presented for recall. It is well established that under such conditions participants tend at test to cluster the randomly presented exemplars according to their category at a greater-than-chance level even without instruction to do so (Bousfield, 1953; Smith, Jones, & Broadbent, 1981). This semantic category-clustering (henceforth termed “semantic-categorization”) implies secondary organization whereby participants bring to bear pre-existing conceptual relationships or semantic associations to guide encoding and retrieval of episodic information which is distinct from primary organization whereby the organization corresponds to the serial order of the list (Tulving, 1968).

In another study (Marsh et al., submitted for publication, 2008), we have used lists of words drawn from a single semantic category to investigate the role that source-monitoring processes play in governing the false recall of items presented as irrelevant sound. Single category lists being ‘blocked’ by semantic category were ideally suited to this purpose because this method of list presentation increases false recall (Brainerd, Payne, Wright, & Reyna, 2003). The research reported here, though conceptually similar, follows a distinct empirical line. In the following experiments, we are not interested in false recall (indeed random presentation of exemplars drawn from several semantic categories attenuates false recall), but rather in
how meaningful irrelevant sound may impair the recruitment of semantic category knowledge as an organizing principle. Recall of semantically categorizable lists, as compared with lists of words drawn from single semantic categories, yields measures of semantic processing in the form of the semantic organization of responses and the probability of producing each category at test (Burns & Brown, 2000). The purpose of the present study was thus to examine whether the meaningfulness of irrelevant speech interferes with semantic organization during free recall of categorizable word-lists.

Assuming quite distinct, semantic-based, processing in the category-exemplar recall task as compared with the seriation-based processing in serial recall, what predictions does the interference-by-process approach make? The first is that the empirical signature of auditory distraction in semantic task settings will be qualitatively distinct from that found in serial recall: the semantic properties of the sound should be endowed with disruptive power but the acoustic, changing-state, properties of the sound should prove relatively impotent. Whilst this first prediction also flows from the interference-by-content approach, the second is unique to the interference-by-process approach: the impairment should be a product not only of the processing of the lexical-semantic attributes of the sound but also of the deployment of semantic-based processes as a means of supporting retrieval. That is, the mere presence of semantic content within the to-be-recalled and irrelevant material should not be sufficient to produce disruption. Thus, if the task does not necessitate or encourage the use of semantic processes but instead seriation processes, the semantic attributes of the sound should not have disruptive potency even when the to-be-recalled items are rich in semantic content. The same prediction holds for between-sequence semantic similarity effects: the greater disruption found with greater semantic similarity between to-be-recalled and irrelevant items is not a passive by-product of their greater overlap within some semantic-psychological space (e.g., Oberauer et al., 2004). Rather, such similarity exacerbates the difficulty of coupling the correct (i.e., task-relevant) set of semantic representations to the semantic-organization processes being engaged to support retrieval.

The present series of experiments begins by seeking to establish whether or not the action of irrelevant sound in the context of semantically-driven episodic tasks—using a category-exemplar task—is indeed distinct from that in the standard serial recall setting. Later experiments in the series investigate the nature of between-sequence semantic similarity effects in this setting and explore the key hypothesis—based on the interference-by-process framework—that when the same type of lists are subject to primary organization (seriation) their recall should become immune to semantic auditory distraction.

2. Experiment 1

The first experiment examined whether category-exemplar recall is disrupted more by meaningful than by meaningless speech: we contrasted the effect of an English narrative with that of the same narrative read in a language that the participants neither spoke nor comprehended (Welsh). Given that the meaningfulness of speech does not influence the ISE in the context of serial recall (e.g., Jones et al., 1990), an effect of meaningfulness in the present setting would imply that semantic-based tasks are peculiarly susceptible to impairment by the lexical-semantic attributes of irrelevant speech. We also examined whether any such impairment could be identified with a reduction in the efficacy of a semantic-organization strategy, as indexed by a diminution in the propensity to cluster by category (Smith et al., 1981). Furthermore, in order to assess whether meaningless irrelevant speech was capable of producing disruption that is attributable to its changing-state properties—as is the case in the context of serial recall—we also included a pink noise condition which conveys no changing-state information: Aperiodic broadband noise such as pink noise fails to disrupt serial recall performance (Ellermeier & Zimmer, 1997) but sometimes disrupts tasks that involve semantic processing (see Smith & Jones, 1992).

2.1. Method

2.1.1. Participants

Thirty-six Cardiff University students took part in return for course credit. Each participant reported normal or corrected-to-normal vision and normal hearing and was a monolingual native English speaker.

2.1.2. Apparatus and materials

2.1.2.1. To-be-remembered material. Eight instances were chosen from each of 64 categories in the Yoon et al. (2004) norms in order to construct 16 lists of 32 words, each list drawn from 4 semantic categories. Categories chosen had minimal category-exemplar overlap, and exemplars and categories were not repeated between or within lists in order to reduce the influence of proactive interference (e.g., Shuell, 1968). The exemplars chosen were sampled outside of the 10 most frequently produced instances so as diminish the likelihood that items could be recalled by simple free association or guessing (e.g., Shuell, 1969).

Categories were randomized assigned to each list but with the constraint that associated categories (e.g., “Musical Instruments” and “Type of Music”) did not appear together. Category-exemplars within each list were arranged pseudo-randomly, so that no two members of the same category were presented adjacent and that each category was represented equally in each quarter of the list.

2.1.2.2. Irrelevant sound. The meaningful speech was English narrative taken from a horticultural essay, recorded in a female voice and sampled with a 16-bit resolution, at a sampling rate of 44.1 kHz using Sound Forge 5 software (Sonic Inc., Madison, WI; 2000). The meaningless speech was a Welsh translation of this narrative. The speech in each of the irrelevant sound conditions was played to participants at 65–70 dB(A) via stereo headphones that were worn throughout the experiment. Third-octave pink noise was generated using a C program
that generates digital waveforms (the program is available from: http://www.moshier.net/pink.html). The irrelevant sounds were presented throughout the study and test phases of the task.

2.1.3. Design

The experiment had a within-participants design with one factor: ‘Sound Condition’ which had four levels: English speech, Welsh speech, pink noise, and quiet. The 16 to-be-remembered lists were randomized but presented in a fixed order for each participant. The sound conditions were randomized as follows: The 16 lists were divided into four blocks. In each block the four lists were randomly assigned to one of the four speech conditions. To control for potential order effects, the order of the sound conditions within each block was counterbalanced across participants.

2.1.4. Procedure

Participants were seated at a viewing distance of approximately 60 cm from a PC monitor on which category-exemplars were displayed in a central position. Lists of category-exemplars appeared in lower case black 72-point Times New Roman font one word at a time against a white background. Each word appeared for 2 s with an inter-stimulus interval of 1 s. Retrieval was immediate with the end of the list being notified by the appearance of a red ‘RECALL’ cue.

Participants were tested in small groups of six participants in a sound-attenuated room. Participants were seated in individual cubicles each equipped with a Samsung Syncmaster 171S PC and display. Participants were informed that they would be presented with sixteen 32-word lists, and that each list would be presented one word at a time on the computer monitor from which they were asked to recall as many words as possible and write the words they remembered down in the order which they recalled them on recall sheets when a ‘RECALL’ cue appeared on the screen. Recall sheets contained 17 columns of 32 rows each. One practice trial was presented before the experimental trials. Participants were not explicitly told that the lists were categorizable. Participants were informed that they would have 2 min to retrieve as much as they could of the list and that after this time a tone would sound to signal the beginning of the next list (some 5 s later). Participants were instructed to ignore any sound that they heard through the headphones and were told that they would not be tested on its content at any point in the experiment. The experiment lasted approximately 50 min.

2.2. Results and discussion

2.2.1. Recall measures

Recall measures came in three forms: the overall mean probability of correctly recalling category-exemplars across all categories, the mean probability of recalling exemplars from within each category recalled, and the probability of recalling each category (based on recalling at least one word from a category). The data from each recall measure were analyzed using a one-way analysis of variance (ANOVA) with Sound Condition as the within-participant variable. Other types of response (e.g., intrusions) were so low as to defy statistical analysis.

Table 1 shows the results of the various recall measures in the four sound conditions. Section A shows the mean scores for the overall probability of correctly recalling exemplars in each condition. These indicate generally that performance was numerically better in quiet than in both the pink noise and Welsh speech conditions (performance in these latter two conditions appeared comparable), which, in turn, produced better performance than the English speech condition. An ANOVA confirmed a main effect of Sound Condition on the overall probability of recalling category-exemplars, $F(3, 105) = 15.50$, $MSE = .006$, $p < .001$, with post hoc tests (Fisher’s PLSD) revealing significant differences between quiet and pink noise ($p < .01$), quiet and Welsh speech ($p < .005$), quiet and English speech ($p < .001$), pink noise and English speech ($p < .001$), and between Welsh and English speech ($p < .001$). The same pattern of means was also evident when considering the probability of recalling exemplars within each category recalled (Section B of Table 1): Performance in quiet was better than in all the sound conditions and performance was better in the pink noise and Welsh speech conditions than in the English speech condition. An ANOVA revealed a main effect of Sound Condition on the probability of recalling exemplars within each category recalled, $F(3, 105) = 18.31$, $MSE = .004$, $p < .001$, with post hoc tests (Fisher’s PLSD) revealing the same significant differences as for the analysis of the overall probability of correctly recalled items. However, the pattern of results was different for the probability of recalling each category (Section C of Table 1). Here, the means for performance in English speech were lower than those for the quiet, pink noise and Welsh speech conditions. An ANOVA revealed a main effect of Sound Condition on the probability of recalling categories.

| (A) Overall probability of recalling category-exemplars |
|-----------------|------------------|
| Quiet            | .51 (.019)        |
| Pink noise       | .46 (.016)        |
| Welsh speech     | .45 (.021)        |
| English speech   | .39 (.022)        |

| (B) Probability of recalling exemplars from within each category recalled |
|-----------------|------------------|
| Quiet            | .53 (.016)        |
| Pink noise       | .48 (.014)        |
| Welsh speech     | .48 (.019)        |
| English speech   | .42 (.018)        |

| (C) Probability of recalling each category |
|-----------------|------------------|
| Quiet            | .96 (.011)        |
| Pink noise       | .96 (.009)        |
| Welsh speech     | .94 (.012)        |
| English speech   | .90 (.019)        |

| (D) Mean Z scores |
|-----------------|------------------|
| Quiet            | 3.17 (.25)        |
| Pink noise       | 3.04 (.25)        |
| Welsh speech     | 2.88 (.25)        |
| English speech   | 2.45 (.24)        |
each category, $F(3, 105) = 5.18$, $MSE = .006$, $p < .005$. However, in contrast to the other recall measures, post hoc testing (Fisher’s PLSD) revealed significant differences only between the English speech condition and all other conditions ($p < .05$, all comparisons).

2.2.2. Clustering measure

Whilst there are several potential ways of measuring semantic-categorization (for a review, see Murphy, 1979), we restricted our analyses to the Z score (Frankel & Cole, 1971). Z scores range from negative to positive values the maximum of which is determined by the number of categories and category-exemplars represented by the to-be-remembered list. The Z score is higher for outputted sequences that demonstrate use of the semantic structure inherent in the to-be-remembered list. For example, the outputted sequence: “brass, lead, giraffe, pig, cherry, lemon, pliers, chisel”, that demonstrates strong use of semantic organization (as is evident from the number of adjacent repetitions of same-category items) will receive a higher score than a more randomly organized sequence: “brass, lead, cherry, pliers, lemon, chisel, giraffe, pig”, in which exactly the same number of categories and category-exemplars are recalled but in which fewer same-category adjacent repetitions are produced. Severely fragmented recall patterns such as “brass, pig, cherry, pliers, giraffe, lemon, chisel, lead” will receive negative Z scores because they represent below-chance clustering (in fact none of the responses in this example are clustered by semantic-category as indicated by the absence of any same-category adjacent repetitions).

Z scores were calculated with all repeat and intrusion errors removed. Section D of Table 1 shows the mean clustering measure for each sound condition. The Z score means are lower in all the sound conditions than in the quiet condition, and lower in the English speech condition compared to the other two conditions. An ANOVA confirmed a main effect of sound on Z scores, $F(3, 105) = 4.22$, $MSE = .84$, $p < .001$. Follow-up post hoc tests (Fisher’s PLSD) revealed that the only significant differences were between the English speech condition and each of the other sound conditions ($p < .05$). Thus, English speech reduced the level of semantic-categorization as measured by the Z score.

The results of the recall measures, and the Z score clustering measure, suggest that there is an impairment to semantic free recall tasks that is attributable to the meaningfulness of irrelevant speech, given that semantic content would, arguably, have been the most salient difference between the Welsh and English speech conditions. There was, however, some degree of disruption produced by pink noise and Welsh speech. This is not easily explained if the free recall task engaged a purely semantic-based retrieval strategy. The pink noise and Welsh speech, being devoid of semantic content, should not have produced disruption. However, because there were no statistically significant differences between the pink noise and Welsh speech conditions it is reasonable to conclude that they both exerted the same general effect of sound that is quite often reported in the context of tasks involving semantic memory (Martin et al., 1988; Smith et al., 1981) rather than one attributable to acoustic variability which produces the classical irrelevant sound effect in serial recall. Further evidence to buttress this conclusion was found in a supplementary experiment—one not reported in full here for economy of exposition—which involved a standard visually-presented serial recall task (for a typical procedure, see Hughes & Jones, 2005) with the irrelevant sound stimuli used in Experiment 1. In this experiment—in which the participants were 40 non-Welsh speakers sampled from the same population as in Experiment 1—a within-participant design was used. This supplementary experiment revealed an irrelevant sound effect, $F(3, 117) = 17.76$, $MSE = .033$, $p < .001$—that is, an effect of sound compared to quiet—but no effect of meaning thus replicating Jones et al. (1990) with the particular materials used in Experiment 1. The experiment also revealed no effect of pink noise thus replicating Ellermeier and Zimmer (1997). The mean probability of correct recall, marked with the strict serial recall criterion and collapsed across serial position, did not differ between the Welsh speech condition, $M = .60$; $SE = .03$, and English speech condition, $M = .61$; $SE = .03$ ($p > .7$), but both these means were significantly lower than those in the quiet condition, $M = .73$, $SE = .02$, and the pink noise condition, $M = .74$, $SE = .02$ ($p < .001$ for both comparisons). This result replicates prior research showing that the meaningfulness of irrelevant speech does not have a disruptive effect in the serial recall setting but also shows that the Welsh speech used in Experiment 1 was indeed capable of producing an ISE in the serial recall context. Its failure to produce any more disruption than pink noise in Experiment 1 suggests that it was impotent in terms of producing a classical ISE in the context of category-exemplar recall.

The effect of meaningfulness in Experiment 1 appears to be explicable as a conflict of semantic processing that emerges when the primary task promotes a degree of such processing. The best indication of an impairment of semantic processing produced by the meaningful irrelevant speech is that meaningful speech reduces the use of secondary organization, that is, the semantic-categorization of the to-be-recalled material, and also reduces the probability with which a category is recalled. Both these recall measures are thought to reflect semantic or “relational” processing (Burns & Brown, 2000; Hunt & McDaniel, 1993). The reduction in category recall reflects, possibly, a failure to establish adequately at encoding, or use at retrieval, higher-order semantic encodings that can be used as a retrieval plan for enabling inter-category transitions (e.g., Bower, Clark, Lesgold, & Winzenz, 1969). One of these semantic encoding strategies might involve forging some kind of semantic association between category-exemplars (e.g., “pigeon”, “chisel”) or categories (e.g., “Birds”, “Tools”) where pre-experimentally there is none (cf. Wingfield, Lindfield, & Kahana, 1998). There is also the possibility that the meaningful speech impairs generation of category-names or exemplars (e.g., for use as retrieval cues) during encoding or retrieval: It is well accepted that tasks that involve semantic memory are underpinned by generation processes (Gronlund & Shiffrin, 1986).
3. Experiment 2

In Experiment 2, the process-based account of semantic auditory distraction is once again scrutinized, this time not by manipulating the mere meaningfulness of the speech but rather its semantic similarity to the to-be-recalled exemplars (see also Marsh et al., submitted for publication, 2008). The interference-by-process and interference-by-content accounts both assume that greater disruption will be found when irrelevant and to-be-remembered category-exemplars are drawn from the same semantic category but the two approaches differ with regard to the causal mechanism underlying such a between-sequence semantic similarity effect.

The interference-by-process approach supposes that between-sequence semantic similarity will impair recall performance not because of the mere similarity in content between the to-be-recalled and irrelevant category-exemplars but because it disrupts the strategy or process underpinning the semantic focal task. In this way the interference-by-process approach is similar to the strategy-disruption interpretation which holds currency in other domains of forgetting research (e.g., Basden & Basden, 1995; Basden, Basden, & Stephens, 2002). In the context of semantic auditory distraction the interference-by-process account holds that the irrelevant information produces disruption because it suggests an organization of semantic information that is incongruent with the participant’s encoding or retrieval strategy (cf. Greitzer, 1976). For example, suppose that a given participant’s initial retrieval plan involves recall of category A first, then D, then C, then B. If irrelevant sound related to category B coincides with the point in the participant’s plan when B exemplars are to be retrieved, the sound will specify information that cues a category (category B) that, if responded to, would impair the orderly execution of the plan. Irrelevant exemplars that are drawn from a category represented on the list (e.g., “sparrow” and “robin” for the category “Birds”) are particularly likely to provide a misleading cue in this way than irrelevant exemplars not represented on the list because they convey information that is “congruent-yet-incongruent” with the task demand; “robin” is a “Bird” and hence context-appropriate but was not presented on the to-be-recalled list and thus is response-inappropriate.

A unique prediction of this process-based view that between-sequence semantic similarity disrupts the participant’s capacity to adhere to their original retrieval plan is that the irrelevant exemplars need only be semantically related to a subset (e.g., one out of four) of the categories represented on the to-be-remembered list. For example, it would matter not whether there are irrelevant items that are semantically related to category A to disrupt recall of category A exemplars: so long as there are irrelevant items on that trial that are semantically similar to at least one of the other subsets of exemplars (say, those comprising category B) this may still corrupt the overall integrity of the retrieval plan. In contrast, whilst the interference-by-content approach predicts that greater disruption will be produced with a greater similarity between to-be-recalled and irrelevant exemplars this effect should be confined solely, or largely to the set of to-be-recalled category-items matching the irrelevant category-items (e.g., Anderson, 1983; Oberauer et al., 2004; Rundus, 1973). Thus, in order to adjudicate between the interference-by-process and interference-by-content accounts we adopted the approach of presenting just eight irrelevant items randomly throughout the presentation and recall periods that, in some conditions, were semantically related to the to-be-recalled items belonging to just one of the four categories represented on the to-be-remembered list.

3.1. Method

3.1.1. Participants

Thirty participants from Cardiff University took part in return for course credit. Each participant reported normal or corrected-to-normal vision and normal hearing and was a native English speaker.

3.1.2. Apparatus and materials

Sixty categories were chosen from the Van Overschelde, Rawson, and Dunlosky (2004) norms. From each of these categories sixteen exemplars were chosen. The 8 highest-dominant exemplars were used for the irrelevant exemplars, whilst the to-be-remembered exemplars were chosen from the 11th to 18th ranked positions. The 60 categories were sorted into 12 pools of 5 categories between which there was no obvious semantic relation. For the related trials, the category presented as sound matched one of the 4 categories represented in the to-be-remembered list. For the unrelated trials, the sound consisted of the category in each pool that was not represented on the to-be-remembered list. The 8 irrelevant items for the unrelated and related conditions were randomly presented throughout the study and test phases of each trial at a rate of two per second, thus each irrelevant item was presented 54 times throughout a given trial. The presence of any given category as part of the to-be-remembered list and related and unrelated sound was counterbalanced between participants. This procedure resulted in the construction of 12 categorized lists comprising 32 exemplars, 4 to be used for each of the related, unrelated and quiet conditions.

3.1.3. Design

A repeated-measures design was used with a single factor—Sound Condition—which had three levels: (1) Speech categorically-unrelated to the to-be-remembered material; (2) categorically-related speech; and (3) quiet. The procedure was the same as in Experiment 1.

3.2. Results and discussion

3.2.1. Recall measures

Table 2 provides the results of the various recall measures in the three sound conditions (again, intrusion and repetition rates were too low for statistical analysis). Section A of Table 2 shows the mean scores for the overall probability of correctly recalling category-exemplars in each condition. The means show generally that performance was better in quiet than unrelated speech, which in turn was better than related speech. An ANOVA con-
firmed a main effect of Sound Condition on the overall probability of recalling category-exemplars, $F(2, 58) = 65.06, MSE = .002, p < .001$, with post hoc tests (Fisher’s PLSD) revealing significant differences between quiet and unrelated speech ($p < .001$) and between unrelated and related speech ($p < .001$). The same pattern of means was also evident for probability of recalling exemplars within each category recalled (Section B of Table 2): Performance in quiet was better than for unrelated speech and performance in unrelated speech condition was better than for related speech. An ANOVA revealed a main effect of Sound Condition on the probability of recalling exemplar within each category recalled, $F(2, 58) = 22.61, MSE = .002, p < .001$, with post hoc tests (Fisher’s PLSD) revealing the same significant differences as for the analysis of the overall probability of correctly recalling category-exemplars ($ps < .05$). However, the pattern of results was different for the probability of recalling each category (Section C of Table 2). Here, the means for performance in the sound conditions were lower than in the quiet condition but there appeared to be no difference between the unrelated and related sound conditions. An ANOVA revealed a main effect of Sound Condition on the probability of recalling each category, $F(2, 58) = 19.12, MSE = .006, p < .001$. However, in contrast to the other two measures, post hoc testing (Fisher’s PLSD) revealed significant differences only between the quiet and unrelated speech and related speech conditions (both $p < .001$). Thus, despite having an effect on the other two recall measures, semantic relatedness had no effect on the probability of producing each category.

To examine the prediction of the interference-by-process account—that the additional disruption produced by related speech will not be specific to the category that matches the irrelevant exemplars but will be generalized to all categories on the to-be-remembered list—we averaged the probability of recalling category-exemplars within each category recalled across the three categories that did not match the irrelevant exemplars and compared these with the mean probability of correctly recalling exemplars from the category that did. The resulting means were $.45 (SE = .016)$ for the non-matching categories and $.48 (SE = .023)$ for the matching category which was not a significant difference, $t(29) = 1.72, p > .05$. Thus, the disruption was not confined to the retrieval of to-be-remembered category-exemplars that matched the category represented by the irrelevant exemplars.

### 3.2.2. Clustering measure

Section D of Table 2 shows the results of the mean clustering measure for each sound condition. The $Z$ score means are lower in all the sound conditions than in the quiet condition, and lower in the related-speech condition compared to the unrelated-speech condition. An ANOVA confirmed a main effect of Sound Condition on $Z$ scores, $F(2, 58) = 13.64, MSE = .996, p < .001$. Follow-up post hoc tests (Fisher’s PLSD) revealed significant differences between quiet and unrelated speech ($p < .001$), and between unrelated and related speech ($p < .001$). Thus the semantic relationship between the to-be-recalled and irrelevant items impairs semantic-categorization above and beyond the effect of mere meaningfulness.

Whilst replicating the key features of Experiment 1, Experiment 2 adds to the few studies that have demonstrated between-sequence semantic similarity effects in semantic free recall tasks (Beaman, 2004; Marsh et al., submitted for publication, 2008; Neely & LeCompte, 1999). Moreover, the between-sequence semantic similarity effect obtained in Experiment 2 appears to fit better within a process-oriented rather than an interference-by-content approach. The finding that appears particularly supportive of the interference-by-process account but at the same time problematic for the interference-by-content view is the lack of category-specific impairment within the between-sequence semantic similarity effect. That is, the disruption produced by irrelevant exemplars that matched one of the to-be-remembered categories was not confined to that matching category (for category-specific impairment in a different context, see Mueller & Watkins, 1977). If disruption is produced as a passive side-effect of the structural similarity of irrelevant and to-be-remembered items within a representational space (e.g., Gathercole & Baddeley, 1993; Oberauer & Lange, 2008; Oberauer et al., 2004) or changes in the nature of the hypothetical associative strengths between to-be-remembered items and retrieval cues within memory (Anderson, 1983; Rundus, 1973), then category-specific impairment would be expected. This finding appears consistent, however, with the notion of process or strategy-based disruption. The disruption, for example, could be attributed to the impairment of a retrieval strategy rather than a clash between the content of the to-be-remembered and irrelevant items (Basden & Basden, 1995; Basden et al., 2002). It seems reasonable to suggest that the related irrelevant items can impair the construction, or use of, a retrieval plan (e.g., Bower et al., 1969) as sometimes the category the sound items activate will be in conflict with the desired (unrelated) list-category being encoded or retrieved. Impairment in recalling exemplars from the categories that do not match the irrelevant speech could be a result of the related irrelevant information suggesting an output order that is incongruent with the participant’s original retrieval plan.

### Table 2

Mean recall and clustering measures as a function of irrelevant sound condition in Experiment 2. Standard errors of the means are presented in parenthesis.

<table>
<thead>
<tr>
<th>(A) Overall probability of recalling category-exemplars</th>
<th>Quiet</th>
<th>Unrelated speech</th>
<th>Related speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>.53 (0.016)</td>
<td>.44 (0.019)</td>
<td>.40 (0.019)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(B) Probability of recalling exemplars from within each category recalled</th>
<th>Quiet</th>
<th>Unrelated speech</th>
<th>Related speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability of recalling within each category</td>
<td>.54 (0.015)</td>
<td>.50 (0.019)</td>
<td>.45 (0.016)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(C) Probability of recalling each category</th>
<th>Quiet</th>
<th>Unrelated speech</th>
<th>Related speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability of recalling each category</td>
<td>.99 (0.005)</td>
<td>.89 (0.018)</td>
<td>.88 (0.017)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(D) Mean Z scores</th>
<th>Quiet</th>
<th>Unrelated speech</th>
<th>Related speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z score</td>
<td>3.74 (0.27)</td>
<td>2.70 (0.27)</td>
<td>2.48 (0.26)</td>
</tr>
</tbody>
</table>
(or perhaps reflect the overhead of inhibitory processes designed to minimize the ensuing conflict; see Section 6).

We turn now to assess the interference-by-process account further by testing more directly its tenet that the semantic effects of irrelevant sound are jointly determined by the nature of the focal task processes and the nature of the sound. Thus, the hypothesis tested in Experiments 3 and 4 is that the semantic effects of irrelevant sound should only arise when (a) the irrelevant sound conveys semantic information and (b) the focal task involves semantic processing.

4. Experiment 3

Experiments 1 and 2 demonstrated that semantic auditory distraction disrupts the semantic-categorization process. This represents an entirely distinct effect from that found in the serial recall setting in which the acoustic, not semantic, properties of the sound disrupts the strategy of serial rehearsal (or seriation; Beaman & Jones, 1998). Experiment 3 sought further, arguably more direct, evidence that the two findings are distinct by making use of a task instruction manipulation (Marsh et al., submitted for publication, 2008; Weist & Crawford, 1973): By instructing one group of participants to recall in serial order, and another to recall the same kind of list by category, we can assess the degree to which the dominant focal process modulates which attributes of the sound dictate disruption. If the effects of the meaningfulness of irrelevant speech in Experiments 1 and 2 are indeed due to semantic processing of the to-be-remembered material, then we would expect semantic effects of irrelevant speech to be found when participants are instructed to retrieve the to-be-remembered material according to semantic category but not when encouraged to use, instead, a seriation process to recall the very same content by instructing recall in serial order. In contrast, the interference-by-content accounts might be taken to predict an effect of the meaningfulness of irrelevant speech regardless of task instruction. There is good evidence that semantic activation of to-be-remembered items occurs regardless of task instruction (e.g., Rouibah, Tiberghien, & Lupker, 1999), thus the propensity for passive interference between semantically similar representations should not differ according to the particular focal task strategy employed. In Experiment 3, we use shorter lists (16-exemplars) than in Experiments 1 and 2 to facilitate the instructed use of a serial recall strategy.

In this experiment meaningfulness was manipulated by contrasting forward with reversed speech. Reversing speech removes phonetic properties that allow lexical access, and thus semantic processing (Sheffert, Pisoni, Fellowes, & Remez, 2002) but preserves the overall spectral features of forward speech. Given that the meaningfulness of speech (as indicated by the comparison between forward and reversed speech) does not influence the classical ISE in the context of serial recall (e.g., Jones et al., 1990), an effect of meaningfulness in the current experiment would further support the prediction that focal semantic processing is peculiarly susceptible to meaningful irrelevant speech, and thus lend support to the interference-by-process account. Moreover, the interference-by-process account would receive further support should the meaningfulness of the irrelevant speech be endowed with disruptive potency only when the to-be-recalled exemplars are recalled by their meaning (e.g., by semantic-category) rather than serial order.

4.1. Method

4.1.1. Participants

Forty students from Cardiff University (none of whom took part in Experiments 1 and 2) participated for course credit. Each reported normal hearing and normal or corrected-to-normal vision and was a native English speaker. Participants were randomly assigned to one of two groups: semantic-categorization or seriation instructions.

4.1.2. Apparatus and materials

These were similar to Experiment 1 with the following exception: Four exemplars were chosen, namely, the 11th to 14th most frequently produced responses for each of the 72 categories chosen from the Yoon et al. (2004) norms. These were combined as in Experiment 1 to create 18 lists of 16 words each, each list having 4 categories.

4.1.3. Design

A mixed design was used with one between- and one within-participant factor. The between-participants factor was ‘Task Instruction’ of which there were two levels: semantic-categorization and seriation. The within-participants factor was ‘Sound Condition’ as in Experiments 1 and 2.

4.1.4. Procedure

The procedure was similar to Experiments 1 and 2 apart from the following: Participants were informed that they would be presented with a total of 18 lists of words that each contained 16 exemplars, 4 from each of 4 different semantic categories. The category-exemplars were drawn from positions 11–14 of each of 72 categories chosen from the Yoon et al. (2004) category norms. The irrelevant speech was English narrative. Meaningless speech was the same narrative reversed using Sound Forge 5 software (Sonic Inc., Madison, WI; 2000). Participants given semantic-categorization instructions were asked to try to remember as many words as possible by semantically-categorizing them and writing them down according to their categories when the RECALL cue appeared: Participants were told to write down the exemplars as they came to mind, and to attempt to recall all the exemplars they could remember from one category, exhausting that category, before moving on, and doing the same with the next category and so on. They were also told that if they could remember any individual words after the semantic-category clusters they should write them at the end of the clusters.

Participants given seriation instructions were instructed to try and remember the words in their original order of presentation and to recall each exemplar by assigning it to its original serial position when the RECALL
cue appeared. To maximise the level of exemplar recall, participants in the seriation group were instructed that they could leave gaps if necessary but were also told that if they had a list-exemplar available to them for recall, but could not remember the position, that they should guess the original position.

Recall sheets contained 18 columns of 16 rows each. Participants given seriation instructions were given specially-prepared recall sheets with serial positions marked on them, whilst the group instructed to categorize had the same recall sheets but without the serial positions marked. Participants were explicitly instructed to ignore any sound that they might hear during the task. Sounds were presented throughout the presentation and test phases. Because the list length was halved for this experiment the retrieval time allotted for each list was reduced to 1 min.

4.2. Results and discussion

4.2.1. Recall measures

The recall measures are distinguished as in Experiments 1 and 2. For the seriation group, items were scored as correct regardless of whether they occupied their correct positions in the output protocols. Section A of Table 3 shows the overall probability of correctly recalling category-exemplars for both Task Instruction groups. Performance in quiet is clearly superior to performance in the speech conditions for both groups. Of greater interest, and consistent with the interference-by-process approach, performance with forward speech was poorer than that with reversed speech for the semantic-categorization group but not the seriation group. A 3 (Sound Condition) × 2 (Task Instructions) ANOVA confirmed a main effect of Sound Condition on the overall probability of correctly recalling category-exemplars, F(2, 76) = 58.06, MSE = .004, p < .001. There was no main effect of Task Instruction, F(1, 38) = 3.8, MSE = .032, p = .059). However, there was a significant interaction between Sound Condition and Task Instruction, F(2, 76) = 16.03, MSE = .004, p < .001, whereby the disruptive effects of meaningfulness arose when the retrieval strategy was semantic-categorization but not when it was seriation. Simple effects analyses (LSD) revealed significant differences between quiet and the reversed condition and between quiet and forward speech (p < .001), and the difference between reversed and forward speech was significant (p < .005) for the semantic-categorization group. In contrast, for the seriation group the only significant differences were those between the quiet and reversed conditions and between the quiet and forward speech conditions (p < .005). The simple effects analyses also revealed that recall performance for the semantic-categorization group exceeded that of the seriation group in all three conditions (all p < .05).

Section C of Table 3 shows the mean probability of correctly recalling each category. In general, category recall was better in the seriation than semantic-categorization group. For the semantic-categorization group, the probability of recalling each category was lower in the forward speech compared to reversed speech and quiet conditions. Interpretation of this particular subset of the results is complicated by the fact that the probability of recalling each category was at ceiling in the seriation group. The reason for this is that recalling the first four presented exemplars in their original presentation positions (as is required under seriation instructions) would guarantee category recall for all categories represented on the to-be-recalled lists. Despite this complication, the pattern of results for the semantic-categorization group appears to be consistent with Experiment 1: Meaningful speech, as compared with meaningless speech, disrupted the probability of correctly recalling each category. Confirming this, an ANOVA restricted to the semantic-categorization group revealed a main effect of sound on the probability of recalling each category, F(2, 38) = 75.38, MSE = .003, p < .001, with post hoc tests (Fisher’s PLSD) revealing significant differences between quiet and forward speech and between reversed and forward speech (p < .001).

4.2.2. Clustering measure

Z scores were negative for the seriation group consistent with the assumption that participants in this group would not have clustered the to-be-recalled exemplars

<table>
<thead>
<tr>
<th>Sound condition</th>
<th>Categorization</th>
<th>Seriation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SE)</td>
<td>Mean (SE)</td>
</tr>
<tr>
<td>(A) Overall probability of recalling category-exemplars</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quiet</td>
<td>.65 (.027)</td>
<td>.56 (.027)</td>
</tr>
<tr>
<td>Reversed speech</td>
<td>.60 (.027)</td>
<td>.48 (.024)</td>
</tr>
<tr>
<td>Forward speech</td>
<td>.45 (.024)</td>
<td>.48 (.025)</td>
</tr>
<tr>
<td>(B) Probability of recalling exemplars from within each category</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quiet</td>
<td>.70 (.020)</td>
<td>.57 (.027)</td>
</tr>
<tr>
<td>Reversed speech</td>
<td>.64 (.017)</td>
<td>.48 (.024)</td>
</tr>
<tr>
<td>Forward speech</td>
<td>.61 (.023)</td>
<td>.48 (.025)</td>
</tr>
<tr>
<td>(C) Probability of recalling each category</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quiet</td>
<td>.94 (.024)</td>
<td>1.0 (.002)</td>
</tr>
<tr>
<td>Reversed speech</td>
<td>.93 (.030)</td>
<td>.99 (.006)</td>
</tr>
<tr>
<td>Forward speech</td>
<td>.75 (.025)</td>
<td>.99 (.006)</td>
</tr>
</tbody>
</table>

F(1, 38) = 25.04, MSE = .02, p < .001, but no interaction between these variables, F(2, 76) = 1.06, MSE = .004, p > .05. Simple effects (LSD) revealed significant differences between quiet and reversed speech (p < .05), and quiet and forward speech (p < .001), and the difference between reversed and forward speech was significant (p < .005) for the semantic-categorization group. In contrast, for the seriation group the only significant differences were those between the quiet and reversed conditions and between the quiet and forward speech conditions (p < .005). The simple effects analyses also revealed that recall performance for the semantic-categorization group exceeded that of the seriation group in all three conditions (all p < .05).
by semantic category. As such, only the scores for the categorization group were included in the following analysis. The mean Z scores were lower in the reversed speech ($M = 2.86$, $SE = 0.19$) and forward speech ($M = 1.79$, $SE = 0.16$) conditions compared to the quiet condition ($M = 3.14$, $SE = 0.17$) and were lower in the forward compared to reversed speech conditions. An ANOVA revealed a main effect of sound on Z scores, $F(2, 38) = 47.15$, $MSE = .213$, $p < .001$. Post hoc testing (Fisher’s PLSD) revealed significant differences between quiet and forward speech, and between reversed and forward speech (both $p < .001$). Thus, like Experiments 1 and 2, the degree of semantic-organization was impaired by meaningful speech as compared with meaningless speech.

Experiment 3 revealed that semantic effects of irrelevant speech—in terms of its meaningfulness—appear to be process—rather than content-driven: The mere presence of similar semantic content within the relevant and irrelevant material is insufficient for disruption; the semantic processing of irrelevant speech disrupts only when dynamic semantic-organisation based processes are deployed as a recall strategy. Consistent with Experiments 1 and 2, meaningful speech, as compared with meaningless speech, disrupted the overall probability of recalling category-exemplars and it also produced semantic interference that was specific to the probability of correctly recalling each category and the degree of semantic-categorization evident in the recall protocols. These results, therefore, provide further support for the interpretation of auditory distraction as process-based (e.g., Jones & Tremblay, 2000) as opposed to structural or content-based (e.g., Gathercole & Baddeley, 1993; Neath, 2000).

5. Experiment 4

The results of the current series are thus far consistent generally with the notion that the semantic processing required by the primary task can be disrupted by semantic processing of irrelevant speech. Extending the rationale of Experiment 3, this suggests that between-sequence semantic similarity should also be disruptive when the primary task requires semantic processing, but not when it relies on serial rehearsal. Experiment 4 sought additional support for the process-based interpretation of semantic auditory distraction by manipulating between-sequence semantic similarity (as in Experiment 3) and also task instructions (as in Experiment 3). Finding that between-sequence semantic similarity impairs category-exemplar recall only under semantic-categorization instructions and not under seriation instructions would serve to bolster the interference-by-process approach and undermine further the interference-by-content approach.

5.1. Method

5.1.1. Participants

Sixty students from Cardiff University took part in Experiment 4. None had taken part in Experiments 1–3. Each participant reported normal hearing and normal or corrected-to-normal vision and was a native English speaker. Participants were randomly assigned to one of the between-participants groups: semantic-categorization or seriation instructions.

5.1.2. Apparatus and materials

These aspects of the method were similar to Experiment 2 with the following differences: The 4 highest-dominant exemplars from each category were used for the irrelevant exemplars, whilst the to-be-remembered exemplars were chosen from the 11th to 14th positions.

5.1.3. Design

A mixed design was used with one between- and one within-participant factor. The between-participants factor was Task instruction as in Experiment 3. The within-participants factor was, like Experiment 2, Sound Condition of which there were three levels: (1) Speech categorically-unrelated to the to-be-remembered material; (2) categorically-related speech; and (3) quiet.

5.1.4. Procedure

The procedure was the same as Experiment 3 with the exception of the following: Participants were informed that they would be presented with a total of 12 lists of words that each contained a total of 16 exemplars. 4 from each of 4 different semantic categories, and response sheets contained 12 columns of 16 rows each. Like the foregoing experiments, sounds were presented throughout the presentation and test phases of the tasks.

5.2. Results and discussion

5.2.1. Recall measures

The recall measures were the same as in Experiments 1–3. Section A of Table 4 shows the overall probability of correctly recalling category-exemplars. It is evident that performance in both speech conditions was poorer than quiet regardless of task instruction. Moreover, performance in related speech was poorer than performance in unrelated speech for the semantic-categorization, but not the seriation, group. A 3 (Sound Condition) × 2 (Task

<p>| Table 4 |
| Mean recall measures as a function of irrelevant sound condition and task instruction in Experiment 4. Standard errors of the means are presented in parenthesis. |</p>
<table>
<thead>
<tr>
<th>Sound condition</th>
<th>Categorization</th>
<th>Seriation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) Overall probability of recalling category-exemplars</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quiet</td>
<td>.62 (.013)</td>
<td>.53 (.014)</td>
</tr>
<tr>
<td>Unrelated speech</td>
<td>.54 (.019)</td>
<td>.48 (.018)</td>
</tr>
<tr>
<td>Related speech</td>
<td>.51 (.022)</td>
<td>.48 (.015)</td>
</tr>
<tr>
<td>(B) Probability of recalling exemplars from within each category recalled</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quiet</td>
<td>.63 (.013)</td>
<td>.54 (.017)</td>
</tr>
<tr>
<td>Unrelated speech</td>
<td>.63 (.014)</td>
<td>.48 (.018)</td>
</tr>
<tr>
<td>Related speech</td>
<td>.58 (.018)</td>
<td>.48 (.015)</td>
</tr>
<tr>
<td>(C) Probability of recalling each category</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quiet</td>
<td>.98 (.006)</td>
<td>.99 (.006)</td>
</tr>
<tr>
<td>Unrelated speech</td>
<td>.87 (.016)</td>
<td>.99 (.004)</td>
</tr>
<tr>
<td>Related speech</td>
<td>.86 (.021)</td>
<td>.99 (.003)</td>
</tr>
</tbody>
</table>
the category-exemplars within each category recalled were .58 (SE = 0.021) for the non-matching categories and .57 (SE = 0.024) for the matching category. A paired t-test revealed the impairment was not specific to the recall of that category, t(29) = 0.44, p > .05. Thus, like Experiment 2, the disruption was not confined to the retrieval of to-be-remembered category-exemplars that matched the category from which the irrelevant exemplars were drawn.

5.2.2. Clustering measure

The mean Z scores for the semantic-categorization group were lower in the unrelated speech (M = 2.61, SE = 0.14) and related speech (M = 2.33, SE = 0.17) conditions than in the quiet condition (M = 3.15, SE = 0.12) and also appeared to be lower in the related compared to unrelated speech condition. An ANOVA revealed a main effect of Sound Condition on Z scores, F(2, 58) = 19.15, MSE = .274, p < .001. Post hoc testing (Fisher’s PLSD) revealed significant differences between quiet and unrelated speech (p < .001), quiet and related speech (p < .001), and between unrelated and related speech (p < .05). Thus the degree of semantic-categorization was impaired by both unrelated and related speech but the impairment was greater in the related condition.

The results thus confirm that between-sequence semantic similarity effects occur only when semantic processing is part of the retrieval strategy (see also Marsh, Hughes, & Jones, 2006, 2008). Such a finding harmonizes with the notion that such effects are better explained in terms of a process-oriented approach than by a content-based approach to auditory distraction.

6. General discussion

The results of the current series can be summarized as follows: Experiment 1 demonstrated that the meaningfulness of irrelevant speech produces greater disruption to the free recall of categorizable word lists than meaningless speech or pink noise. This experiment also revealed that the pattern of semantic interference as compared with the disruption produced by noise (pink noise and meaningless speech) shows a unique characteristic: it affects the recall of categories as well as the degree of semantic-categorization demonstrated at test. Experiment 2 revealed an effect of between-sequence semantic similarity whereby the semantic relatedness between the to-be-recalled and irrelevant items produces additional disruption to the overall probability of recalling exemplars, the probability of recalling exemplars within each category and the degree of semantic-categorization observed although it has no effect on the probability of recalling each category. Experiments 3 and 4 revealed that effects of meaningfulness and between-sequence semantic similarity are found only when semantic-categorization is adopted by the participant and not when serial order is used as an organizing strategy.

The results of all the experiments appear to be well accounted for within a process-oriented approach to disruption from task-extraneous material (Hughes & Jones, 2005; Jones & Tremblay, 2000; Neumann, 1996). The interfe-
ence-by-process view holds that in the case of the classical ISE, the processing of serial order in the sound is in conflict with the processing of serial order in the primary task; the interference may be construed as the residual cost of mechanisms (perhaps inhibition; see Hughes & Jones, 2003) designed to resolve a ‘competition-for-action’ conferred by the order cues generated from irrelevant and relevant sources of information. This approach explains why neither the meaningfulness of irrelevant speech nor between-sequence semantic similarity plays a role in the disruption when serial recall is instructed (Experiments 3 and 4; see also Buchner et al., 1996 and Jones et al., 1999): In this case it is the information that the irrelevant sound yields about serial order, not its meaning, that is broadly compatible with the action (or process) of serial rehearsal. The interference-by-process approach also explains why the meaning of speech becomes disruptive to the performance of free recall tasks only when semantic-categorization is an obvious or instructed strategy (Experiments 1-4). When the primary task involves dynamic semantic encoding and retrieval processes—unlike the case with serial recall—the irrelevant semantic information extracted from the speech produces competition for these processes. Impairment can thus be understood in terms of a relative difficulty in selecting the correct source of semantic information as they both compete for the category of action being called for in the semantic recall task. That between-sequence semantic similarity produces more interference than mere meaningfulness (Experiment 2 and 4) is particularly supportive of a process-oriented account: The irrelevant speech in this case specifies highly context-compatible, but ultimately response-inappropriate, information in the context of the semantic recall task.

One potentially problematic finding for the process-oriented approach is that meaningless irrelevant speech disrupted recall of categorizable lists when semantic-categorization was either spontaneously adopted (Experiments 1 and 2) or instructed (Experiments 3 and 4). However, that pink noise (which does not convey acoustic variation) in Experiment 1 produced comparable disruption to meaningless (Welsh) speech suggests that the effect of meaningless irrelevant speech in this paradigm is not the typical effect—attributable to acoustic variability—that disrupts serial recall. It seems, therefore, that this effect is a general effect of noise that often impairs performance on tasks that call upon semantic processing (see Smith & Jones, 1992).

That the disruption observed in the context of semantic tasks is determined by organizational processes that are brought to bear to meet the demands of the instructed retrieval strategy (Experiments 3 and 4) suggests it is the process, rather than content, that dictates the degree and type of disruption from irrelevant speech. This view of the impairment produced by irrelevant auditory stimuli is consistent with a functionalist approach to memory generally which advocates that the goals of the individual and the retrieval environment (instructions, cues, task demands) play a critical role in remembering and forgetting (Toth & Hunt, 1999), and according to which attempts to delineate the “structure(s)” of memory is an ill-rewarded endeavour.

The process-oriented approach also seems to provide a better interpretation of the results reported here than attentional resource-based accounts of disruption from irrelevant sound (Cowan, 1995; Neath, 2000; see also Lange, 2005). For example, the data seem particularly at odds with an attentional capture approach (Cowan, 1995; Elliott, 2002; see also Lange, 2005). On this account, the classical ISE is the result of acoustic changes-in-state from one irrelevant item to the next causing an orienting response away (or capturing attention) from the focal task. At first glance, the account might be readily extended to deal with the kind of semantic auditory distraction effects observed here by appeal to the notion that between-sequence semantic similarity would also give rise to orienting responses via the priming of semantic features of the sound based on the selectivity of the to-be-recalled items (cf. Cowan, 1995). However, as acknowledged by Saults and Cowan (2007), recent evidence indicates that the changing-state effect and attentional capture effects are functionally distinct (Hughes, Vachon, & Jones, 2005; Hughes et al., 2007). For example, attentional capture effects are evident in short-term memory tasks that are assumed not to engage a seriation process (missing-item task, cf. Beaman & Jones, 1997; Buschke, 1963) whereas the changing-state effect is contingent on seriation in the focal task (Hughes et al., 2007). In relation to the present data, there is no reason to suppose that semantic features of irrelevant sound should be primed—and hence endowed with attentional capturing power—any less by the to-be-recalled items under seriation instructions than under semantic-categorization instructions. Based on the notion that the priming of semantic features is underpinned by automatic semantic priming (e.g., Neely, 1976), attentional capture due to the meaningfulness of the sound or due to between-sequence semantic similarity would be expected even when serial recall is emphasized because automatic semantic priming occurs ‘full blown’ regardless of the focal task (see Neely & Kahan, 2001). Thus, that semantic auditory distraction was only produced under semantic-categorization instructions (Experiments 3 and 4) provides further evidence against this approach. The key difficulty for attentional resource-based accounts therefore is that they are too general; they cannot readily explain why the nature of auditory distraction is dictated by the prevailing mental activity, whilst this is an axiomatic tenet of the interference-by-process approach.

One challenge that flows from the view that the disruption reported in the present experiments is produced by a conflict between the semantic processing of the sound and semantic processing in the focal task is to identify more precisely the nature of that focal semantic processing. This is because it is likely that a number of diverse semantic processes contribute to performance on the semantic-categorization task, any of which could be potentially disrupted by the meaning of irrelevant speech. For example, semantic processing is required in the task for at least the following: (a) identifying the categorical structure of the list (Murphy, 1979); (b) reorganizing list exemplars to encode and rehearse same-category-exemplars together (Weist, 1972); (c) coupling semantic retrieval and rehearsal processes; same-category-exemplars (e.g., “dog”, "cat", "mouse"
“horse”) may be automatically (e.g., by spreading activation; Collins & Loftus, 1975), or deliberately, retrieved and rehearsed together upon presentation of a related category-exemplar (“bear”) after intervening unrelated category-exemplars (Weist & Crawford, 1973); and (d) cued search of long-term memory (Raaijmakers & Shiffrin, 1981; Rundus, 1973), which involves generating list, or candidate list, category names and exemplars for search of, and retrieval from, long-term lexical-semantic memory (Gronlund & Shiffrin, 1986). Although there is already some evidence that the latter generative process is susceptible to semantic-based disruption (Marsh, Hughes, & Jones, 2006, 2008), a more fine-grained analysis of which other semantic processes are impaired is needed.

The disruption produced by irrelevant auditory stimuli seems much better captured by a process-oriented approach (e.g., Hughes & Jones, 2005; Jones & Tremblay, 2000) than either a structural (or content-based) approach (e.g., Gathercole & Baddeley, 1993; Neath, 2000) or an attentional resource-based approach (e.g., Cowan, 1995; Lange, 2005; Neath, 2000). Latterly the interference-by-process view has been considered in terms of the selection-for-action account to attentional selectivity (cf. Allport, 1993; Hughes & Jones, 2005; Neumann, 1996). This approach holds that the planning and execution of a contextually-appropriate act—that demanded by the primary task—is compromised to the extent that irrelevant information is compatible with the act at a general level but incompatible with it at the level of the particular response required.

On the selection-for-action account, interference from the irrelevant information reflects an overhead of the action of mechanisms that prevent it from actually assuming the control of action. Thus, the nature and extent of interference is a joint product of the character of the primary task and the nature of the potentially-distracting information: When both irrelevant and relevant events constitute plausible candidates for the skill deployed to perform the focal task a selection-for-action problem is generated. For example, the classical ISE with serial recall may be conceived as a residual cost of preventing the competing irrelevant source of order information from assuming control of the skill of serial rehearsal deployed to retain the to-be-remembered sequence (Hughes & Jones, 2005).

We would argue that the strength of the process-oriented approach, embedded in the selection-for-action view, derives from its questioning of the more generally received wisdom that attentional selectivity is imposed by some shortfall of the cognitive system as is traditionally held by structuralist approaches (e.g., a limitation on processing or a limited attentional resource or set of resources; Cowan, 1995; Neath, 2000). Instead, the selection-for-action view is that selection (for action) is the problem to be solved (not the solution) and limitations in performance (including susceptibility to disruption from irrelevant information) are the consequence of adaptive mechanisms designed to resolve that selection problem (for extensive discussions, see Allport, 1993; Neumann, 1996; see also Anderson, 2003). Thus, on the selection-for-action approach, the relationship between selectivity and limited capacity is turned on its head: A human performer’s limited capacity (in an empirical sense) reflects the achievement of selective attention mechanisms designed to ensure that only task-relevant information assumes the control of goal-directed action, an achievement that gives the illusion of a limited capacity in the sense of a hypothetical property of the mind. The qualitatively different impairments that are produced by different aspects of irrelevant auditory stimuli (e.g., acoustic or semantic) in different settings (e.g., serial recall versus category-exemplar recall) are the manifestation of selective attention mechanisms operating to avoid ‘cross talk’ from information appropriate to actions or skills involved in, but inappropriate to the specific demands of, performing a given task. Two streams of information do not come into conflict for what they contain but for how they are processed.

Acknowledgement

The research reported in this article received financial support from an Economic and Social Research Council (UK) grant awarded to Dylan Jones, William Macken and Robert Hughes (RES-000-22-1526). We thank Steve Mosher for allowing us to use his C program to generate pink noise for Experiment 1.

References


