Pergamon

Neuropsychologia, Vol. 33, No. 11, pp. 1433–1454, 1995

0028–3932/95 $9.50+0.00

0028–3932(95)00074–7

THE ROLE OF SUBVOCALIZATION IN AUDITORY IMAGERY

J. DAVID SMITH,* MARGARET WILSON† and DANIEL REISBERG‡

*Department of Psychology and Center for Cognitive Science, Park Hall, State University of New York at Buffalo, Amherst, NY 14260, U.S.A.; †The Salk Institute for Biological Studies, La Jolla, CA 92037, U.S.A.; and ‡Psychology Department, Reed College, Portland, OR 97202, U.S.A.

(*Received* 3 *June* 1994; *accepted* 16 *February* 1995)

Abstract—Five experiments explored the utility of subvocal rehearsal, and of an inner-ear/inner-voice partnership, in tasks of auditory imagery. In three tasks (reinterpreting ambiguous auditory images, parsing meaningful letter strings, scanning familiar melodies) subjects relied on a partnership between the inner ear and inner voice, one similar to the phonological loop system described in the short-term memory literature. Apparently subjects subvocally rehearsed the imagery material, which placed the material in a phonological store that allowed the imagery judgement. In a fourth task (distinguishing voiced and unvoiced consonants in imagery), subjects still subvocally rehearsed, but seemed to need no additional phonological store to respond correctly. In this case they may have consulted articulatory or kinesthetic cues instead. In a fifth experiment (making homophone judgements), subjects hardly even needed to subvocally rehearse, a result suggesting that homophone judgements rely on some direct route from print to phonology. We consider the breadth of the partnership between the inner ear and inner voice, the level that subvocal rehearsal occupies in the cognitive system, and the functional neuroanatomy of the phonological loop system.

Key Words: imagery; auditory imagery; subvocalization; inner speech.

INTRODUCTION

Research on mental imagery has typically focused on visual, not auditory, imagery. Fortunately, this gap is closing [50], for auditory imagery is important in its own right, occupying an intriguing position amidst diverse phenomena and research domains. For example, it may underlie the rehearsal processes of working memory [3, 7] and the phonological processes subserving some aspects of text comprehension [5, 10, 30]. Likewise, auditory imagery may play a role in music perception and cognition [19, 21, 36], the verbal processes of self-regulatory cognition and even the auditory hallucinations of schizophrenia [60].

In exploring auditory imagery, we cannot presume that insights about visual imagery will simply generalize, given the different sensory characteristics of sound and light, the different evolutionary histories of hearing and vision, or the different entanglements of hearing and vision with speech and language. For example, humans could well rely on processes of subvocal rehearsal to refresh or enliven auditory images, and the analogs to this in visual imagery remain unclear.

*Author to whom all correspondence should be addressed.

In fact, research on short-term memory confirms that covert rehearsal benefits some cognitive functions. Specifically, short-term memory for verbal material seems to rely on a phonological loop with two constituents—a short-lived store that represents material in a phonological form, and a process of rehearsal that re-enacts this material, re-presents it to the store, and thus refreshes and preserves its contents. Intuitively, this conception of working memory relies on a partnership between an 'inner ear' (the store) and an 'inner voice' (subvocal rehearsal). The interplay between these two resources in short-term memory is documented by experimental and neuropsychological evidence [3, 4, 9, 30, 59, 62, 63, 73, 74].

Our research focuses on other uses the phonological loop has for cognition. In particular, many imagery tasks require that subjects analyze or make judgements about auditory stimuli that are not currently on the scene. We consider the possibility that the inner-ear/inner-voice partnership provides a platform on which these imagery processes and judgements take place (see also [7]).

Evidence already exists that covert speech plays some role in auditory imagery. Reisberg *et al.* [53] examined the imagery analogue of the Verbal Transformation Effect [66–68], an effect which relies on the fact that certain words and phrases, if repeated over and over, yield a soundstream compatible with more than one segmentation. For example, rapid repetitions of the word 'life' produce a soundstream fully compatible with the perception that either 'life' or 'fly' is being repeated. These ambiguous soundstreams are usually perceived first in one way then the other, changing in phenomenal form just as the (visual) Necker cube and duck/rabbit figures do.

Reisberg *et al.* asked whether imagined repetitions produce verbal transformations, just as heard repetitions do. Subjects imagined the stimulus word's being repeated by a friend's voice, rather than actually hearing it, and often did report transformations of their image. However, these transformations by imagers seem to depend on subvocalization, for they are essentially eliminated when subarticulation is blocked by having subjects chew candy during the trial. The effects of chewing cannot be attributed to general distraction, because other concurrent activities, equally distracting but not involving the articulators, caused no disruption.

Moreover, Reisberg *et al.* found that the probability of success in this task gradually declines as subvocalization is undercut. Specifically, subjects in a later experiment were allowed differing degrees of vocalization: Some subjects actually spoke the ambiguous phrase; others whispered it; still others mouthed it with no movement of air; a fourth group imaged it without mouthing; and a fifth group imaged it with the articulators clamped still. Across these five groups, respectively, 80, 65, 55, 50 and 30% experienced a phenomenal shift in the target phrase (all nonadjacent percentages are reliably different from one another). As subvocalization waned, so waned the capacity to reinterpret the image. Those familiar with the short-term memory literature will see the close analogy between this result and that of Murray [42], who found that memory traces become more robust as their rehearsal becomes more explicit and overt (i.e. mouthing, whispering, saying).

These results on auditory imagery frame the issues for the present research. First, we ask if subvocal rehearsal provides support to auditory imagery in a broader range of tasks. Second, we ask about the precise *role* of subvocalization in auditory–imagery tasks. That is, we will ask whether the inner voice alone provides the critical information that allows imagery judgements and reconstruals, or whether the inner ear alone does, or whether the imagery judgements depend on a partnership between the two resources. Our attempt to specify this functional relationship recalls again the short-term memory literature, in which different effects (e.g. phonological similarity) are linked to the input or output components of the loop, and in which different interference manipulations (e.g. concurrent audition, concurrent articulation) and cerebral accidents selectively target one or the other [6, 62, 63, 71, 73].

In discussing the data, we will consider several further issues. First, we describe other instances of articulatory/phonological interactions in cognitive processing, to emphasize their importance. Second, we consider the special value of re-presentation by the inner voice when a task requires judgements about or analyses of imagery material. This idea of re-presentation explains more fully the restricted role the phonological loop plays in language processing. Third, we consider the level in the cognitive system at which the inner voice and the inner ear resources interact. Finally, we discuss the functional neuroanatomy of the inner ear and inner voice, drawing on recent brain-imaging studies.

EXPERIMENT 1: VERBAL TRANSFORMATIONS

Reisberg *et al.* [53] showed that the reinterpretation of auditory images depends on subvocal rehearsal processes. But what role does this rehearsal play? Is the kinesthetic support from articulation sufficient for performance, or is some auditory/phonological representation involved as well? Experiment 1 evaluated these two possibilities using the standard logic of selective interference. Subvocal rehearsal (the inner voice), is known to be blocked by concurrent articulation (e.g. tah-tah repeated aloud by the subject), and this interference manipulation thus allows one to ask how performance fares absent rehearsal. The phonological store (the inner ear) is known to be blocked by concurrent auditory input (e.g. tah-tah repeated through headphones), and this interference manipulation thus allows one to ask how performance fares absent the phonological store.

We already know that image reconstruals somehow depend on subvocal rehearsal. If blocking the inner ear disrupts performance, this will indicate that imagery reconstruals also depend on the phonological store, and that subvocal rehearsal and the phonological store work in partnership during performance. If blocking the phonological store does not disrupt performance, this will indicate that subvocalization provides mainly kinesthetic information for subjects, and will disconfirm a partnership pattern for that task.

Method

Subjects. Forty-five subjects were approached in various school and professional settings and asked to participate in a brief (5 min) procedure. Fifteen subjects were assigned at random to each of the three experimental conditions—no interference, articulatory suppression and irrelevant speech through headphones. Subjects received $1.00 for their participation.

Interference manipulations. Subjects in all conditions imagined a friend, of the same sex as themselves, repeating a target word. For the articulatory-suppression group, subjects were instructed not to say the repetitions at all. To enforce this injunction, subjects were told to press their lips together, clench their teeth, and to press their tongue to the roof of their mouth while imaging the repetitions. This suppression technique seemed unlikely to be generally distracting to subjects, and unlikely to generate a disturbing rhythm in conflict with the rhythm of the repetitions.

Other subjects imagined the repetitions while hearing tape-recorded prose, read by a speaker of the same sex as the subject. Distracting speech was chosen to increase the similarity between the interference and imagery materials, following studies in working memory that suggest that this similarity increases the interfering effects of the irrelevant auditory inputs on the performance of the focal task [3, 55].

Procedure. The experimenter, who was blind to the experiment's hypotheses, explained that some words, when repeated, begin to sound like something else. Then she repeated the word 'life' for 1 minute, at two repetitions per second, and the subject listened and reported any transformations. Forty of 45 subjects heard the transformation to 'fly' and only these subjects' data were analyzed.

Next subjects were given a new word—'stress'—printed on an index card. They were to imagine a friend's voice repeating it with no gaps at the rate the experimenter had demonstrated. They were instructed not to say the word out loud at all, but just to imagine silently the friend's repetitions. Subjects imagined the repetitions for 1 min, pausing only to report transformations, but then resuming the imagery.

The articulatory-suppression subjects were told not to "say the word out loud at all, don't whisper, don't even move your teeth, tongue, or lips". To help them comply, they were told to "put your teeth together, your lips firmly together, and put your tongue firmly on the roof of your mouth. This will make sure you use pure imagination for your repetitions." Irrelevant-speech subjects were told that the speech heard through headphones would be only background speech, that they could completely ignore it, and that they would not be tested on it in any way.

Results

Following Reisberg *et al.* [53], we focused our data analysis on the transformation of 'stress' to 'dress', the transformation that was heard by 100% of subjects in their perceptual conditions, but was not easily guessed in their guessing condition. It is thus the transformation that most clearly signals a *bona fide* perceptual discovery.

In the no-interference condition, 77% of subjects reported a transformation of 'stress' to 'dress' during the 1 min of imagined repetitions. In the suppression and irrelevant-speech conditions, respectively, 25 and 13% heard a transformation to 'dress'. These percentages did not differ from each other, but both were reliably lower than the no-interference group's rate of transformation, $t\,(26)=3.04$, $P<0.01$, for control vs suppression; $t\,(28)=4.40$, $P<0.01$, for control vs irrelevant speech. Both percentages were similar to those found by Reisberg *et al.* for subjects who did not engage in imagery at all, but only tried to guess the transformations that might occur (about 25%). These subjects were instructed to guess what the new word could change into if it were repeated over the way the experimenter had repeated 'life'. They were cautioned not to say the word at all, and any visible vocalization was immediately corrected. These subjects tended to offer, as their responses, various anagrams of the stimulus (e.g. tress, rest, and so forth), but rarely gave the 'dress' response. This provides further assurance that this response does signal a genuine perceptual reversal.

Discussion

The large effect of articulatory suppression observed here replicates the earlier finding by Reisberg *et al.*: In order to discern these transformations, subjects need to rehearse subvocally the imagery material. But in addition, subjects also depend on the inner ear (the phonological store). If this access is blocked (by irrelevant speech), performance is also devastated. Apparently, then, subjects find reconstruals of ambiguous auditory images using both components of the phonological loop system.

Of course this interpretation rests on the presumption that articulatory suppression and irrelevant speech have their effects by specifically blocking subvocal rehearsal and the phonological store, respectively. But one could argue that these manipulations disrupt performance merely because they are generally distracting. Fortunately, this issue has been addressed in some detail in the working-memory literature. This evidence indicates that not all concurrent activities disrupt performance in the way that concurrent articulation does [6], and not all irrelevant inputs have the impact of task-irrelevant speech [55, 56]. These results are peculiar if these manipulations are simply generally distracting, and the most natural interpretation remains that these manipulations specifically disrupt phonological processing and the phonological loop [3, 4, 30].

This specificity applies to our imagery tasks, too. In the stress/dress task, for example, performance is disrupted if subjects must clamp their articulators in place. However, there is no interference with this task if subjects hum while imaging the repetitions [53]. In contrast, subjects in a further experiment were instructed in the difference between voiced and unvoiced consonants, and then received a printed list of 36 one-syllable words. Subjects were asked to imagine how each word would sound if pronounced aloud, and to judge whether the initial consonant in each word would be voiced or unvoiced. Subjects were 76% correct if this was their only task. Subjects were also successful if their articulators were clamped tightly shut (68%, not reliably different from 76%). However, performance dropped almost to chance levels if subjects were asked to hum aloud while making these judgements (56%, reliably different from both 76 and 68%).

This pattern makes little sense if clamping or humming are merely distracting in some general way. However, these results have a sensible selective-interference interpretation. For the voiced–unvoiced task, subjects need information about their own vocal chords (or the planning mechanisms for the vocal chords); access to this information is disrupted by humming but not by clamping. Verbal transformations, in contrast, seem to depend more on articulatory information, and access to this information is blocked by clamping but not by humming. This implies that stress/dress subjects are relying more on phonological representations engendered by the movement and positions of teeth, tongue and lips (or their planning or control mechanisms). The key point, though, is that these interference manipulations do what they are meant to do. They specifically preempt particular enactment processes, rather than causing general distraction.

Moreover, the voiced/unvoiced result draws attention to a further point: Apparently, the inner voice is not a monolithic entity; instead, covert speech has multiple aspects (just as overt speech does), and a particular task might require support from some, but not all, of these aspects. In addition, it seems possible to disrupt some aspects of the inner voice while sparing others—e.g. to disrupt voicing but not articulation. This fractionation of the inner voice suggests that covert rehearsal serves to create a quite specific stimulus to be judged or interpreted. Consequently, further research may be able to determine, for any given auditory–imagery task, not only whether subvocal rehearsal supports it, but also what features of subvocalization are critical to performance.

Two methodological points follow from this discussion. First, for some purposes one will prefer interference material that is quite similar to the imagery material for the focal task, thus increasing the power of the interference manipulation. Second, for other purposes, one may prefer a broad-band interference manipulation, like the commonly-used repetitions of 'tah-tah', that involve both the articulators and voicing. This will allow one to produce reliable interference effects without having to know in advance which specific aspects of subvocalization the focal task requires. Both of these considerations have guided our choice of procedures in the studies reported here.

In contrast, a further methodological consideration carries little weight. That is, it appears to make little difference whether one chooses a silent mode of articulatory suppression (as we did in Experiment 1), or an overt (aloud) mode of articulatory suppression (as nearly all other researchers have done). In either case, suppression overruns covert rehearsal processes; in neither case is the effective locus of interference the auditory input from the concurrent task. This is evident in our finding here that silent suppression has the same effect as overt suppression (see also Reisberg *et al.* [53], Experiments 1 and 4). Likewise, in memory research, Wilding and White [74] found in three experiments that mouthed and spoken articulatory suppression were equivalent in their disruption of rhyme judgements. Irrelevant auditory input had nearly no effect. Indeed, the literature shows in general that the effects of irrelevant auditory input are pale compared to the effects of overt articulatory suppression (Gathercole, personal communication, January 1995). This is true even when concurrent audition and concurrent articulation have the identical content [16].

This point is echoed by Gupta and MacWhinney [33], who compared memory performance when subjects articulated the word THE, with performance when subjects tapped their fingers as a motoric distractor while a tape loop played THE as concurrent audition. Articulatory suppression had stronger effects on performance than did the motor-hearing distraction. This obviously fits well with the claim that the effects of concurrent articulation are truly articulatory in nature; the effects seem not to arise from the acoustic consequences of this articulation.*

EXPERIMENT 2: MEANINGFUL LETTER STRINGS

Experiment 1 indicates that the imagined verbal-transformation effect relies on a partnership between the inner voice and inner ear. Under articulatory suppression, covert rehearsal processes are undermined, and so the output function that repeats the string operates poorly. Under concurrent audition, the phonological store is otherwise engaged, and the input provided by covert rehearsal goes 'unheard'. Either kind of interference makes image reconstrual difficult, because both kinds of processing underlie the successful reconstrual of an auditory image.

Do other auditory imagery tasks also involve this partnership? In our second Experiment, subjects imagined hearing a string of letters or numbers being named aloud, and had to write down the English word each string would sound like. For instance, the string N–M–E would sound like 'enemy' if spoken aloud. Like imaged verbal transformations, this task, on the face of it, requires auditory imagery; will it also require the support of inner speech, and perhaps the inner ear in partnership?

Method

Subjects. Twenty-four members of a university community were paid for their participation; six subjects were assigned at random to each of four experimental conditions.

Interference manipulations. This experiment adopted a standard 2 × 2 selective-interference design (between subjects). Articulatory suppression was provided by having subjects repeat 'tah-tah' aloud—the most frequent block to subvocalization in the literature. Auditory interference was again provided by irrelevant speech (in this case the poem *Jabberwocky*) heard through headphones. In all conditions, interference (when present) began before each stimulus sheet was presented, and continued until after the trial had ended. No strict pace was set for repetitions in the articulatory suppression condition, but subjects who paused were reminded to continue.

Instructions to subjects. Subjects were told that they would be given strings of letters and numbers that, when pronounced out loud, sound like English words. For instance, the string M-T would be 'empty'. Subjects were told to imagine hearing each string spoken aloud, and to write down the English word it formed. Subjects were then given a single sheet of paper showing 10 strings. (Thus all strings were simultaneously visible to the subject.)

*We note that Gupta and MacWhinney [34] have rethought their 1993 study [33], arguing that flaws in the previous study block the conclusions they had earlier endorsed—namely that the effects of concurrent articulation are truly articulatory in nature. However, and despite their arguments, their current data do seem still to emphasize those articulatory effects. Even so, it remains an important goal to specify exactly how articulatory suppression does reduce performance, and the research of Gupta and MacWhinney is an innovative effort in this regard.

Subjects were instructed that they could skip items if they wished, and return to them later if they wished. Subjects had 90 sec to complete the task.

Results

With no interference, subjects interpreted 72% of the strings correctly. But, under articulatory suppression or irrelevant auditory input, performance declined to 21 and 40%, respectively, and to 19% when subjects were under both kinds of interference simultaneously (see Fig. 1). These data were analyzed using a two-way ANOVA with Suppression (absent–present) and Irrelevant Speech (absent–present) as between-subjects factors. There was a main effect for both kinds of interference—for Articulatory Suppression [$F\,(1, 28) = 27.4$, $P < 0.05$, $MSe = 3.83$]; for Irrelevant Speech [$F\,(1, 28) = 6.40$, $P < 0.05$, $MSe = 3.83$]. Importantly, there was also a reliable interaction between Suppression and Irrelevant Speech [$F\,(1, 28) = 4.70$, $P < 0.05$, $MSe = 3.83$]. *Post-hoc* tests revealed the character of the interaction—performance with neither kind of interference on the scene was better than performance in any other condition; there was no difference among the other conditions (articulatory suppression, irrelevant speech, or both combined). (Note: all *post-hoc* tests in this paper relied on the LSmeans procedure from the SAS package. To preserve overall protection levels for these tests across the series of experiments, the required level of significance for a contrast between means was set at 0.02.)

Discussion

Again subjects clearly needed to subvocalize in order to perform this task—with subvocalization blocked they interpreted many fewer strings correctly. But subvocal rehearsal alone was not sufficient to support interpretation. Subjects were not just using the fact that articulating the string N–M–E feels like articulating 'enemy'. Given this kinesthetic strategy, the inner ear would not be needed and irrelevant speech would not impair performance. Yet it clearly does, indicating that subjects use a strategy that requires both the inner voice and the inner ear. Apparently, subjects produce the repetitions with the inner voice, 'listen' with the inner ear, and then interpret the auditory or phonological stream. These results generalize the pattern observed in Experiment 1—this task also relies on a partnership between subvocalization and the auditory buffer.

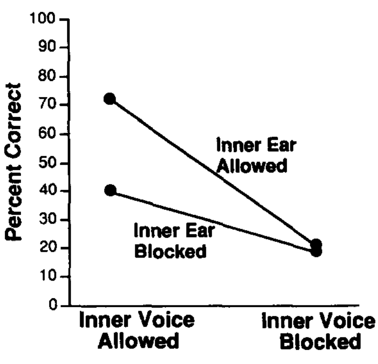

Fig. 1. Performance (percent correct) when subjects tried to interpret letter strings pronounceable as English words. Subjects performed with the inner voice available or blocked (by articulatory suppression) and with the inner ear available or blocked (by irrelevant auditory input).

EXPERIMENT 3: FAMILIAR TUNES

Both of the tasks we have examined so far have relied on subvocalization, but both have also featured judgements about language. Is the use of subvocalization restricted to imagery for linguistic materials, or is it also used for a more general range of representations? To address this question, we turn to a task that requires a judgement about *pitch*. Will subjects still use subvocalization for this task?

Method

Subjects. Twenty members of a university community were paid for their participation.

Interference manipulations. This experiment again adopted a standard 2 × 2 selective-interference design (within-subjects), with the order of presentation of the conditions counterbalanced across subjects. Subjects were verbally instructed about the interference at the beginning of each condition. Articulatory suppression was again provided by having subjects repeat 'tah-tah' aloud, and auditory interference was provided through headphones, with subjects hearing "Take Me Out To The Ball Game" being sung. Melodic material was used as the interference to maximize the similarity between the interference material and the task material. (Moreover, in this case a pilot study had shown no interference effect from listening to prose.) The interference activity, when required for a condition, began before the stimulus sheet was presented to the subject, and continued until the last item was marked. No strict pace was set for repeating, but subjects who stopped repeating altogether were reminded by the experimenter's tapping on the table.

Stimuli. Stimuli were 32 familiar tunes (children's songs, showtunes, etc.), eight per condition. Stimuli always appeared in the same order for each subject, but, because the sequence of conditions was counterbalanced, the assignment of stimuli to the various conditions was as well.

Procedure. Subjects first heard the opening bars of each tune being played on a piano (tape-recorded), and were asked to identify it. This ensured that all stimuli had been heard recently by the subject, and also allowed us to exclude from analysis any stimuli unfamiliar to the subject. After subjects had heard all the tunes once, they were then given the following instructions: "You will be shown the names of familiar songs, and asked to decide whether the melody rises or falls from the *second* to the *third* note. For example, 'Pop Goes The Weasel' should be marked 'rises', while 'Take Me Out To The Ball Game' should be marked 'falls'. Do not actually sing or hum the tunes. Instead, *imagine* hearing the tunes to help you decide." The song titles were then presented, with the songs for that condition listed on a single sheet. All of the songs listed did in fact change in pitch between the second and third notes—i.e. the correct response was never 'stays the same'. The appropriate interference for each condition (if any) was begun before the list of titles was presented.

Results

With no interference, subjects judged pitch contours successfully 83% of the time (chance performance would be 50%). But, under suppression or irrelevant auditory input, performance declined to 65 and 68%, respectively, and stayed at 68% when subjects were under both kinds of interference simultaneously. (Performance in all conditions was reliably above chance-levels.) This pattern suggests that all interference conditions hurt performance about equivalently. This pattern was confirmed by analyzing the data using a two-way ANOVA with Suppression (absent–present) and Irrelevant Speech (absent–present) as within-subject factors. There was a significant interaction between Suppression and Irrelevant Speech [$F(1, 19) = 4.70$, $P < 0.05$, $MSe = 301.1$]. Moreover, *post-hoc* tests confirmed that all three interference conditions (suppression, irrelevant speech, or both) produced equivalent performance, and all produced worse performance than when no interference was present.

Discussion

This result reveals the partnership pattern once again. We know that subjects must subvocalize to perform well in this task because blocking subvocalization impairs performance. But they also need access to the inner ear, given the impact of irrelevant auditory inputs on performance. We therefore conclude that subjects sing these tunes and listen to themselves with the inner ear, judging the pitch changes from these auditory representations.

While this result plainly does extend the partnership pattern to a musical judgement, further research would be useful for judging how lyrics figure in the present results. Our task, requiring subjects to judge a pitch change from the second to the third note, can surely be done without reference to lyrics, but, at the same time, we know that song lyrics are a powerful retrieval cue for melodies [22, 35, 36]. Thus, subjects might rely on the lyrics to help scan through the target melodies. This would introduce a linguistic element into our task, and this could encourage use of the phonological loop. Thus, a valuable experiment in this area would be one that taught subjects wordless tunes, and then studied the effects of interference.

EXPERIMENT 4: PLURAL AND PAST-TENSE MORPHEMES

Our first three experiments have documented a consistent pattern. We have chosen tasks that seemed likely to draw on auditory imagery and, in each case, imagery performance has relied on an inner-ear/inner-voice partnership. Experiment 4, however, demonstrates that there are exceptions to this partnership pattern. In its focal task, subjects made judgements about the pronunciation of word-final phonemes. In particular, we used the inflectional morphemes for past tense and for plurals, taking advantage of the fact that the pronunciation of these morphemes is altered by the voicing of the preceding segment. For example, 'cats' ends in an unvoiced /s/ sound but 'dogs' ends in a voiced /z/ sound. Similarly, 'walked' ends in an unvoiced /t/ sound but 'jogged' ends in a voiced /d/ sound.

Method

Subjects. Twenty-four members of a university community were paid participants in the brief experiment. (These same subjects had also served in Experiment 2.) Subjects were assigned at random to one of four experimental conditions—no interference, articulatory suppression alone, irrelevant speech alone, or both.

Stimuli. Two stimulus lists were prepared, with 10 stimuli on each list. One of the lists contained words all ending in the letter 's'; some of the words were plural nouns; others were present-tense verbs; others were possessives. Each of the words on the list had a different penultimate phoneme (i.e. the phoneme determining whether the final 's' will be voiced or not); in other words, each penultimate phoneme appeared only once in the series. Given the variety of penultimate phonemes, half of the words on the list would be pronounced with the word-final phoneme voiced, and other half unvoiced. The second list contained past-tense verbs, each spelled with the regular 'ed' ending. Again, each word on the list had a unique penultimate phoneme, such that half of the words would be pronounced with a terminal /ed/, and half with a terminal /t/. No words were included in either list for which the inflection would add a syllable—i.e. we excluded words like 'bat' (past tense 'batted') and 'house' (plural 'houses').

Presumably, subjects do not know explicitly which word-final phonemes must be voiced and which are unvoiced. But subjects surely can discriminate upon hearing which are which. The question in Experiment 4 is whether subjects can make this judgement from imaged sounds.

Interference manipulations. This experiment again employed a 2×2 selective-interference design (between subjects). Articulatory suppression was provided by having subjects repeat 'Suzie-Suzie' aloud while making all the S vs Z judgements, and 'teddy-teddy' for all of the T vs D judgements. In this way we provided subarticulated material that was phonologically similar to the to-be-judged material. Irrelevant speech input took the form of repetitions of either 'Suzie' or 'teddy' through headphones, corresponding to the to-be-judged material, once again to provide disruptive phonological material. Interference began before the stimulus sheet was presented, and continued until after the list was completed. Again, no strict pace was set for repeating, but subjects who paused were reminded to continue.

Procedure. S–Z subjects were told that they would see words ending in S that would either sound as though they ended in 's' or ended in 'z'. They were to imagine hearing each word spoken aloud, and mark the ones that sound like they end in 'z'. T–D subjects were told that some past-tense verbs that end in 'ed' can be pronounced as if they end with 't' instead. They were told to imagine each word on their list spoken aloud, and to mark the ones that sound like they end with a 't' sound. The wording for the T–D instruction was more careful than for the S–Z instruction because some pilot subjects refused to believe that any of the stimuli could end in a 't' sound. Following these instructions, subjects received instructions appropriate to their interference condition.

Results

Table 1 shows the levels of performance in all eight conditions of the experiment. When the plural data were analyzed using a two-way ANOVA with Suppression (absent–present) and Irrelevant Speech (absent–present) as between-subject factors, there was a main effect for Suppression [$F(1, 28) = 5.76$, $P < 0.05$, $MSe = 1.96$]; but not for Irrelevant Speech [$F(1, 28) = 0.78$, $MSe = 1.96$], and there was no interaction between Suppression and Irrelevant Speech [$F(1, 28) = 0.02$, $MSe = 1.96$]. In all conditions, performance was reliably above chance-levels.

The past-tense data showed the same pattern: a main effect for Suppression [$F(1, 28) = 7.15$, $P < 0.05$, $MSe = 1.26$]; but not for Irrelevant Speech [$F(1, 28) = 1.21$, $MSe = 1.26$], and no interaction [$F(1, 28) = 0.62$, $MSe = 1.26$]. Again, performance in all conditions was above chance levels.

Discussion

Like the tasks of Experiments 1–3, subjects performed best when they were able to subvocalize. Articulatory suppression hurt their capacity to judge the sound of plural and past-tense morphemes. However, there the resemblance with previous results ends: In Experiment 4, this covert articulation alone was sufficient to support optimal performance. Subjects may have been just articulating 'dogs' and 'walked' and feeling the kinesthetic cues provided by that enactment. There is no evidence, as there was in previous experiments, that the inner ear was needed for this task, or that this task reveals any partnership between the inner ear and inner voice.

One might worry, of course, that the morpheme task is just too easy to be affected by interference in the inner ear. Or perhaps the inner-ear interference used was too weak. But the morpheme task was not at ceiling, and was subject to interference (in the inner voice). And the inner-ear manipulation was very similar to those that impaired performance in Experiments 1 and 2, and moreover relied on phonological material highly similar to the material in the focal task, making it potentially especially disruptive in this particular

Table 1. Performance under selective-interference conditions when the inner voice was available or blocked (by articulatory suppression) and when the inner ear was available or blocked (by irrelevant auditory input). (A) Percent correct when subjects judged whether words spelled with 'S' are pronounced with 'S' or 'Z'. (B) Percent correct when subjects judged whether words spelled with 'ED' are pronounced with 'D' or 'T'

	Inner voice	
	Available	Blocked
(A) S/Z judgements		
Inner ear		
Available	90	78
Blocked	94	85
(B) T/D judgements		
Inner ear		
Available	84	70
Blocked	85	77

instance. Yet there was no disruption. There seems to be something different about the morpheme task that needs explaining. In particular, it is intriguing that the inner voice remains relevant for this task even when the inner ear drops out entirely. This suggests that the inner voice can provide different sorts of cues and signals to the cognitive system. Possibly subvocal rehearsal is used more kinesthetically in this case, so that the rest of the phonological loop system does not come into play for the s/z or t/d tasks. Indeed, given this result, one might argue that the s/z and t/d tasks are 'kinesthetic judgements,' and not tasks of auditory imagery proper. We would suggest, however, that little hangs on this classification, and, on any account, these tasks are clearly pertinent to our understanding of the function of the inner-ear/inner-voice partnership.

EXPERIMENT 5: HOMOPHONES

Thus far all four auditory–imagery tasks have benefited from subvocal rehearsal, and three of them have revealed the inner ear and inner voice in partnership. It appears that judgements about imagined sounds often rely on the components of, or even have the precise architecture of, the phonological loop system proposed by Baddeley and others.

However, this reliance may not extend to all tasks of auditory imagery, and judgements about pseudo-homophones seem a candidate for such a task. Would 'phyte' sound like an English word if pronounced aloud? Would 'cayoss'? Answering these questions seems to require a judgement about sound, yet these judgements show substantial independence from the phonological store and from subarticulation. For example, Baddeley and Salame [8] found that irrelevant speech input leaves homophone judgements unimpaired, suggesting their independence from the inner ear. Similarly, articulatory suppression leaves homophone judgements essentially intact [5, 12].

However, note an important limitation of these *separate* interference studies. Perhaps subjects can use *either* a kinesthetic *or* a phonological strategy in making homophone judgements. In this case, the representations used to perform the task might 'move around', as subjects simply switch modes depending on which variety of interference is on the scene. On this view, then, homophone tasks would be immune to disruption from either the inner voice or inner ear *alone*. Obviously the crucial condition for detecting this switching strategy is to run subjects with both kinds of interference present simultaneously. To our knowledge, subjects had never been run under this double phonological jeopardy, and so Experiment 5 attempted to clarify this situation.

Method

Subjects. Twelve members of a university community were paid for their participation in the experiment. Each subject made homophone judgements under all four experimental conditions—no interference, articulatory suppression alone, irrelevant speech alone, or both, and were given these experimental treatments in counterbalanced fashion.

Interference manipulations. This experiment's 2 × 2 selective-interference design was carried out within subjects. Articulatory suppression was provided by having subjects count from one to six repetitively, a suppression manipulation adopted by Besner *et al.* [12]. Subjects were coached towards the repetition rate used in Besner *et al.*'s Experiment 6 (metronome 170). As previously, auditory interference was provided by irrelevant speech heard through headphones (the reading of prose material).

Stimuli. The stimuli were nonsense words, both pseudo-homophones and otherwise, which were constructed following Besner *et al.* [12]. The stimuli were drawn from a set of 48 carefully matched pairs. One member of each pair was a pseudo-homophone (aynjel or raynbo); the other a nonsense word (ayntel or raunbo). Using these pairs, four 24-word lists were constructed. Each list contained 12 pseudo-homophones and 12 nonsense words, but never contained both members of a pair.

Procedure. Subjects were told that they would see strings of letters arranged in lists, and that some of the strings would sound like actual English words. They were to put a checkmark by the ones that do. It was stressed that they would be timed, and that they should work as fast as they could without making errors. Subjects first worked through a practice list of 12 words structured as described above. When an interference manipulation was applied, it was initiated first, and then at the experimenter's "Go" the subject turned the page and began the list.

In preparation for articulatory suppression, subjects were told that they would also perform a repetitive counting task, counting from one to six at a rate that they then practiced. In preparation for irrelevant speech, subjects were told they would perform while hearing someone reading over headphones. They were told that they would not be quizzed on this auditory material and could just ignore it as they worked through the list. In preparation for receiving both interference manipulations simultaneously, subjects received both sets of instructions.

Results

Table 2 shows decision time (latency to complete the 24-item lists) and percent accuracy under the four conditions. When the latency data were analyzed using a two-way ANOVA with Suppression (absent–present) and Irrelevant Speech (absent–present) as within-subject factors, there were no main effects and no interaction.

When the error data were similarly analyzed, there was a main effect for suppression [$F(1, 11)=5.32$, $P<0.05$, $MSe=4.79$]; but no main effect for irrelevant speech [$F(1, 11)<1.0$, $MSe=1.43$]; and no interaction [$F(1, 11)<1.0$, $MSe=4.19$].

Discussion

The latency result replicates others showing that homophone judgements are slowed by neither articulatory suppression nor irrelevant auditory input. The error result shows that homophone judgements are made only slightly less accurate by articulatory suppression, and are not affected by auditory input. These two results reflect accurately the combined findings from several other experiments [5, 8, 12]. The key result, though, lies in the condition with both forms of interference simultaneously present, and both components of the phonological loop simultaneously blocked. Subjects make homophone judgements fluently even then, when they have nowhere to hide.

Thus, the pseudo-homophone task represents a second case in which judgements of imagined sounds seem not to depend on the partnership between the inner ear and inner voice. We do find a reliable effect of concurrent articulation. As in Experiment 4, this suggests that if either component of the phonological loop is the more essential element, it is the inner voice. However, Experiment 5 joins others in confirming that homophone

Table 2. Performance under selective-interference conditions when the inner voice was available or blocked (by articulatory suppression) and when the inner ear was available or blocked (by irrelevant auditory input). (A) Average time (sec) to complete 24 homophone judgements. (B) Percent correct in making 24 homophone judgements

	Inner voice	
	Available	Blocked
(A) Time to completion		
Inner ear		
Available	30.5	31.7
Blocked	30.8	29.8
(B) Percent correct		
Inner ear		
Available	89	82
Blocked	89	83

performance is substantially spared even with the inner voice denied. Several factors probably contribute to this pattern: Our pseudo-homophone task requires a judgement about a single representation, rather than a comparison among multiple representations; the task relies on an enormously well-practiced translation (from print to phonology); the task also requires no analysis or dissection of the represented sound. Which of these factors are crucial, though, remains a topic for future research. Clearly, though, pseudo-homophone performance raises interesting questions about when a task will show the partnership pattern, and when not. We consider this issue below.

GENERAL DISCUSSION

In three tasks we observed a partnership between the inner ear and inner voice. That is, subjects subvocally rehearsed the required imagery material, thereby representing it in a phonological store (the inner ear), and this enabled them to make a judgement about the image's sound properties. Evidence from other domains emphasizes both the breadth and the theoretical significance of this partnership pattern.

First, as we have mentioned, short-term memory rehearsal relies on a phonological loop system that also features an interaction between subvocal rehearsal and a phonological store [3, 4, 30]. In fact, the similarities that led us to borrow heavily from the working-memory model in researching auditory imagery led Baddeley and Logie [7] to ask if the phonological memory system might even be the seat of auditory imagery.

As another example of the partnership, silent mouthing of visually-presented lists creates memory effects that were long attributed to acoustic coding (e.g. the suffix effect), and some of these effects do not occur if lip movements are discouraged [43] (further discussion of auditory–modality effects can be found in [18, 20, 23, 44]). The partnership model explains this pattern: subvocalized (mouthed) speech images could well load the inner ear, and produce phonological ('auditory') effects in a way that images denied subvocalization do not.

There is also evidence that whispered and mouthed repetitions of syllables cause VOT boundary shifts in the identification of speech stimuli [17]. This result is problematic if habituation effects are viewed as purely auditory; but not if we assume that subvocal rehearsal loads the inner ear, and that some habituation effects reside within that quasi-perceptual medium.

Finally, the inner-ear/inner-voice partnership also holds promise for explaining the hallucinations of schizophrenia [60]. Sensory theories of hallucinations have failed to explain why auditory hallucinations (i.e. voices) so dominate the schizophrenic's experience, and why the voices seem poorly externalized (i.e. they seem to reside in the patient's throat and head). As one schizophrenic said, "I'm hearing the voices again—they're coming right through the voice box." Another said, "somebody is attacking me and my lips move" (Gould [31], p. 424; see also Gould [32]). These observations are easily explained, though, on the view that schizophrenics subvocalize and then "hear" the hallucinatory speech images that result. Consistent with this, electromyography sometimes reveals articulatory activity correlated with hallucinations. Moreover, some therapeutic interventions (and some controlled studies) have reduced hallucinations by having schizophrenics clamp their articulators, much as we do in our experiments to block use of the inner voice. Thus, schizophrenic voices may represent yet another (albeit bizarre) instance of the inner-ear/inner-voice partnership. (For detailed discussion, see [13, 29, 60].)

Thus speech perception and speech production interact in a variety of tasks and domains. In other contexts, similar interactions have been explained by the Motor Theory of Speech Perception [38], or by linking speech perception closely to the perception of articulatory gestures [28]. Without endorsing either account, we do endorse a concerted effort to describe more fully the processing interactions linking auditory perception and production. Tracing the connections joining speech, imagery, and memory research may encourage a closer analysis of this interactive utility.

Of course, questions do remain about the nature of the inner-ear/inner-voice partnership, and we consider three now. First, we consider the role of the partnership in cognitive processing, with an emphasis on when it is used, and when it is not. Second, we consider the level of abstraction at which the partnership operates. Third, we consider the functional neuroanatomy of the inner ear and inner voice.

The function of the partnership in cognitive processing

According to an emerging consensus, the phonological loop system appears not to be needed for a variety of language functions. It is not needed, for example, as a buffer or assembly platform for normal speech production. Nor is it necessary for normal language comprehension, or for reading (either single-word tasks, or even everyday reading comprehension).

Regarding speech production, even patients with profoundly impaired short-term memory systems can have reasonably normal conversational speech. Vallar and Baddeley [62, 63], for example, studied a patient (P.V.) who was impaired on a variety of standard working-memory measures. Nonetheless, this patient showed rapid voluntary articulation and normal conversational speech (see also [47, 58]). Gathercole and Baddeley ([30], p. 89) conclude that "the weight of neuropsychological evidence favours the view that normal phonological working memory skills are not necessary for the planning and production of spontaneous speech". Regarding speech perception, severe deficits of the phonological loop system can still spare language comprehension. All three of Warrington *et al.*'s [69] short-term memory patients had reasonable language comprehension, as did P.V. (see also [65]).

Reading also proceeds without using the phonological loop as a print-to-sound buffer. The phonology of single words can be fluently derived from print even while articulation is suppressed (e.g. the speed and accuracy of pseudo-homophone verification is preserved; see [5, 12]), and even while both components of the loop are blocked simultaneously (Experiment 5). These pseudo-homophones also cause frequent errors by experimental 'proofreaders' (e.g. they overlook the error in "idle and board" [24]); and pseudo-words also lodge more securely in memory than do other nonwords [11]. Both of these effects persist even when articulation is suppressed.

Articulatory suppression does not even affect the speed or accuracy with which subjects verify simple sentences [2] or subjects' memory for prose passages [72]. Apparently, then, the phonological loop plays only a minor role in normal reading, either at the single-word level or in basic reading comprehension.

Thus, three separate streams of on-line language comprehension—speaking, listening, reading—flow fluently even when the inner-ear/inner-voice partnership is denied to cognitive processing. So what is the partnership good for, then?

The present results show that subvocal rehearsal and the inner-ear/inner-voice partnership is valuable in finding new construals for imagery material (Experiment 1 in [53]); assembling visually presented phonological material into a possible language stream (Experiment 2); analyzing a familiar melody to compare the pitches of different notes (Experiment 3); and so forth. Thus, even though the partnership is unnecessary for 'on-line' language processing, it does appear necessary whenever a task requires subjects to re-present auditory material or re-scan it, in order to analyze the material, make comparisons, or form interpretations of it.

Several other results are consistent with this general view. For example, rhyme judgements with visual stimuli (e.g. do 'phraigm' and 'stame' rhyme?) require the inspection of mentally-represented phonological material, its segmental analysis, and then the comparison of two separate representations. Rhyme judgements should therefore use the phonological loop, and should be impaired by articulatory suppression. They are [12, 37, 74]. Judging whether two words have the same stress pattern also requires replaying the words comparatively, and also borrows the phonological loop [16]; see also [15].

One might also expect that processing of particularly complex language would over-tax the on-line parsing routines, creating a need to rely on some sort of 'back-up' system, in which inputs are preserved and re-inspected, off-line. In such cases listeners might switch into 'loop mode', using the obvious buffering utility of the rehearsal loop in order to get by. Gathercole and Baddeley [30] discuss the evidence supporting this view.

Though music remains thinly researched in relevant ways, we would expect the same pattern. Receptive music (listening) and productive music (singing) might well not borrow the phonological loop system, and might survive disruptions of this loop (e.g. the relevant neuropsychological disorders). However, when the task requires the comparison of novel melodic fragments [40], or when analytic judgements of musical material are required (Experiment 3 in [36]), we would predict a reliance on the phonological loop, and performance deficits under articulatory suppression.

Thus the phonological loop seems to be required whenever the cognitive system pauses, and steps back from its fluid input and output modes, neither of which borrows the loop. The loop provides a distinctive intermeshing of the input and output utilities for the purpose of re-presentation, inspection, off-line judgements and interpretations.

The level of the partnership in the cognitive system

Can we be more precise, though, about how this re-presentation occurs, or about how the 'inner voice' is realized? The neuropsychological evidence has critically shaped ideas on this point. It is clear, for example, that subvocal rehearsal does not depend on peripheral articulatory mechanisms. Demonstrating this point, Baddeley and Wilson [9] studied a group of dysarthrics, whose language was (centrally) intact, but who had lost the ability to control the articulatory musculature and speak. Subvocal rehearsal was clearly intact in these patients, for they showed normal memory spans; and showed the usual word-length and phonological similarity effects for visually presented memory material. Baddeley and Wilson concluded that phonological coding and subvocal rehearsal can operate centrally without the actual realization of speech (see also [45, 64]).

Baddeley and Wilson suggest, therefore, that subvocal rehearsal involves the central mechanisms and motor sequencing routines of speech *planning*. These control mechanisms were presumably intact in their patients who had long spoken normally, before suddenly forfeiting the surface and outward realization of speech acts. Logie *et al.* [39] endorse this speech-planning view of rehearsal, but also add a cautionary note: Having overt speech, and not just covert, may increase span somewhat, just as Murray [42] showed that overt rehearsal by normal individuals increases span. Possibly, too, some judgements about speech sounds may really need some more kinesthetic signal provided by the articulators (e.g. the s–z and t–d judgements of Experiment 4). In cases like this, more surface realizations for speech could be required, and we would expect dysarthric individuals to fail in making these judgements.

Under this speech-control proposal, apraxic patients—specifically those with a disturbance of speech planning—should show deficits in the functioning of the phonological loop, and they do. These patients, operating without articulatory suppression, perform a variety of tasks just as normals do when operating with articulatory suppression [73].

We would urge one caution, though, in interpreting this 'speech-planning' view of subvocalization. As we have seen, subvocalization is disrupted by concurrent chewing [53], and also by the requirement that subjects 'clamp' their articulators in a frozen position (Experiment 1, or [53]). These results are easily accommodated if we assume that 'speech-planning' includes both the selection of sounds to be uttered, and also some early steps toward planning the actual articulator movements, for producing these sounds. It would then be this latter step that is disrupted by 'clamping' the articulators, or by chewing.

However, Bishop and Robson [14] provide a dissenting voice. They showed that congenitally speechless individuals, who had never spoken and who had arguably never learned the motor plans for producing and controlling speech, nevertheless showed normal phonological loop systems in several different tasks. There are several ways to read this finding; as one possibility, one could argue that the connections between the inner voice and inner ear develop automatically, i.e. with no need for overt practice in speaking. In this case, these connections would be available to these anarthric individuals. A different possibility (and the conception preferred by Bishop and Robson) is that the representations underlying subvocal rehearsal are not articulatory gestures (either explicit or at some abstract, planning stage), but rather are more abstract phonological representations, of a kind available to individuals who have never spoken. Bishop and Robson suggest that, for normal individuals, these representations might be the ones that enter the system that derives a speech motor program, and thus there is an obvious link between these representations and speech-planning. Nonetheless, these representations would not in themselves contain articulatory specifications. (For related discussion, see [34].)

In summary, then, the phonological loop system does not depend on inner speech in any literal sense, with covert muscle movements and the like. Whether its proper locus lies one level deeper (in motor plans for speech) or two levels deeper (in a more abstract phonological code—see [30]) remains undecided.

The functional neuroanatomy of the auditory imagery partnership

Several lines of research, in memory, auditory imagery, and so forth, reveal a phonological loop with two constituents, one more closely linked to audition, one more closely linked to speech production. We now consider the neural substrate of these input and output resources.

The inner ear. The candidates for a phonological coding and storage center lie in the left temporoparietal cortex, close to the left supramarginal gyrus, the angular gyrus, and including the left posterior superior temporal gyrus (Wernicke's area). For relevant findings see [26, 27, 46, 57, 64]. More recently, Zatorre *et al.* [77] found these areas active in a task involving imagery for familiar songs (with lyrics), and concluded that, in the absence of any real auditory input, those activation sites may reflect "endogenous auditory processing" (i.e. imagery). Paulesu *et al.* [48] imaged the brain during an immediate memory span task and a rhyme task, both involving visually presented materials. Their results led them to implicate the superior temporal gyrus in phoneme perception and phonological processing, and they noted (p. 344) that this area "can be activated even in the absence of external auditory phonological stimulation".

The inner voice. As we have already noted, lesions that produce dysarthria (e.g. in the bulbar motor neurons, the basal ganglia, the cerebellum, the deep white-matter tracts) spare the short-term memory system and the rehearsal process that serves it. Thus candidate loci for subvocal rehearsal instead probably lie in the cortical regions that serve the planning and sequencing of articulatory events. These regions are generally considered to be in the anterior cortico-subcortical regions. Damage in these areas does cause apraxic symptoms (difficulty in initiating speech, poorer articulation as word length increases, simplification of consonant clusters, phoneme substitutions, and better articulation for highly automated speech—counting—than voluntary expressive speech). Moreover, these apraxic individuals often do have disorders of short-term memory. They act as though they are operating under articulatory suppression (the inner voice blocked) even when they are not ([73]; see also [47, 70]).

In particular, Broca's area is often implicated in language output (e.g. [49]) and in the pre-motor organization of articulation. It seems involved, for example, in the initiation of speech [1], in the organization of articulatory sequences and in the covert formulation of language (inner speech [61]). Thus even small focal lesions to Broca's area cause stuttering and oral apraxia (though not full-blown Broca's aphasia—see [41]).

The inner-ear/inner-voice partnership. Zatorre *et al.* [76] used PET scans to examine brain activity while subjects heard words passively and while they performed a phonetic monitoring task (e.g. subjects were asked whether 'big' and 'leg', auditorily presented, end in the same segment). These analyses and comparisons of phonetic segments might well engage subvocal rehearsal and the phonological store. As expected, the segment-monitoring task created significant activity in the superior temporal gyri. In addition, it created a large increase of activity in left Brodmann's areas 44 and 6, Broca's area especially near the junction with the premotor cortex. Zatorre *et al.* (p. 848) point out that Broca's area seems not to be involved in the normal reception of phonetic sequences (i.e. word perception). However, for off-line phonetic comparisons, "the auditory features of the stimulus, extracted by temporal-lobe mechanisms, must be related to articulation." This is precisely the argument this paper makes based on the relevant experimental evidence.

Demonet *et al.* [25] compared a tone monitoring task (press response button to ascending tone triplets); a phonetic monitoring task (press to D-syllables followed by B-syllables in auditorily presented nonsense words—e.g. respond to "reDozaBu"), and a semantic monitoring task (find positive adjectives modifying small animals in auditorily presented noun phrases—e.g. respond to "kind mouse"). Note that the phonetic monitoring task requires just the kind of secondary analyses of phonetic sequences that we have argued should invoke the phonological loop. This task strongly activated the anterior part of Wernicke's area, spreading towards the more anterior portions of the superior temporal gyrus, and also activated left Brodmann areas 44 and 45 (Broca's area).

Demonet *et al.* also singled out one brain area left *inactive* by this task, namely, the Supplementary Motor Area. This leads Demonet *et al.* to be skeptical of the claim that phonological monitoring truly requires the inner voice proper. Rather, they suggest that the activation in Broca's area reflects a phonemic, not articulatory, component of inner speech. This argument, played out in the neuropsychological domain, obviously parallels the debate mentioned earlier, concerned with whether subvocal rehearsal is more articulatory in nature, or more based in an abstract phonological code [14, 30, 34].

Finally, both the span task and the rhyming task of Paulesu *et al.* [48] strongly activated Broca's area, in addition to the temporo-parietal areas already described. The activation of other areas, such as the supplementary motor area, also indicated to these authors (p. 344) the recruitment of "a more general neuronal network serving language planning and execution." Their data, they argued, "support a multicomponent model for the articulatory loop, in which the subvocal rehearsal system and the phonological store are anatomically distinct."

Thus the evidence suggests that the functional neuroanatomy of the phonological loop does have the duality of structure expected by the phonological loop model. It apparently includes both a set of loci allied to the receptive phonological processing of speech, and a set of loci allied to productive phonological processing (i.e. the pre-motor planning and sequencing of speech). Moreover, these two areas seem to interact under two circumstances: when phonological maintenance is required [48] and when off-line judgements about phonology are required [25, 48, 76].

However, note again that the existing research heavily favors language. Analogous music studies are only beginning to emerge. For example, Zatorre *et al.* [77] presented subjects with two words from a familiar song, and asked them to use auditory imagery to judge whether the pitch of the first word was higher or lower than the pitch of the second. Here too regions of the superior and middle temporal gyri indicated to the authors that subjects were engaging in endogenous auditory processing (for no real auditory input was arriving). Other regions of activation (e.g. the Supplementary Motor Area) suggested that subjects were covertly singing the songs to themselves. Clearly, this experiment (and our Experiment 3 also) included both a tonal and a language component, and this may increase or change the brain areas serving the interaction of the output and input resources. Thus it seems certain that a more balanced picture of the loci serving covert sound enactment will be possible only when other species of sound have been explored more fully (and more 'purely'—e.g. tunes that have no words).

As a related point, we note that relatively little research has examined imagery for materials not-easily-articulated—the sound of squealing brakes, or breaking glass, and so on. It seems plausible that inner-speech could not be involved in the imagery of these noises (cf. [19, 21]), but some authors have raised questions about this intuition (e.g. [7, 54]). Hence this too remains a question for future research.

Imagery, re-presentation, and mental life

As we have seen, neither the inner ear nor the inner voice is required for ordinary on-line language processing. Instead, these resources are called into play whenever the cognitive system must pause to consider a phonological representation. This may occur when the ordinary channels of processing are over-taxed, in which case the rehearsal loop serves as an input or output buffer. This also occurs when a judgement is required about a stimulus not currently on the scene. In this case the system uses a production resource to re-present the stimulus, so that it can be scanned, analyzed, and interpreted. This is reflected in the neuropsychological data, which show that off-line phonological tasks involve areas of both auditory-input and articulatory-output processing, in the absence of actual input or output.

Other output–input partnerships probably exist within cognition. Indeed, Experiment 4 suggests one possibility. In that task we found a role for the inner voice, but not the inner ear. Here the inner voice may actually be feeding a kinesthetic cue to an articulatory monitor, creating a partnership that depends on kinesthesis, not audition. Likewise, when language uses the visuo-spatial modality, as it does for people fluent in sign language, results suggest striking parallels to the articulatory loop [75], again emphasizing the breadth of these output–input partnerships. More broadly, we believe that this strategy of 'self-provided stimuli' may be prevalent in everyday mental life, as when we write down a word to see its correct spelling, or use a deliberate motor act to feel which way our house key turns [51].

The idea of partnerships between production and monitoring resources may also prove useful in considering other modalities of mental imagery. For example, does visual imagery draw on rehearsal mechanisms, and on an output–input partnership analogous to that between the inner voice and inner ear (i.e. a partnership between an 'inner scribe' and 'inner eye' [52])? If so, are these constituents as cleanly separable as they seem to be in auditory imagery? If not, what consequences does this have for the functioning of visual imagery? Such questions encourage greater cross-talk among researchers of imagery in different modalities.

REFERENCES

1. Alexander, M. P., Naesser, M. A. and Palumbo, C. Broca's area aphasias: Aphasia after lesions including the frontal operculum. *Neurology* **40,** 353–362, 1990.
2. Baddeley, A. D. Working memory and reading. In *Processing of Visible Language*, P. A. Kolers, M. E. Wrolstad and H. Bouma (Editors), pp. 355–370. Plenum Press, New York, 1978.
3. Baddeley, A. D. *Working Memory*. Clarendon Press, Oxford, 1986.
4. Baddeley, A. D. *Human Memory*. Allyn & Bacon, London, 1990.
5. Baddeley, A. D. and Lewis, V. Inner active processes in reading: The inner voice, the inner ear and the inner eye. In *Interactive Processes in Reading*, A. M. Lesgold and C. A. Perfetti (Editors), pp. 107–129, Lawrence Erlbaum, Hillsdale, New Jersey, 1981.
6. Baddeley, A. D., Lewis, V. and Vallar, G. Exploring the articulatory loop. *Q. J. exp. Psychol.* **36,** 233–252, 1984.
7. Baddeley, A. D. and Logie, R. Auditory imagery and working memory. In *Auditory Imagery*, D. Reisberg (Editor), pp. 179–197. Lawrence Erlbaum, Hillsdale, New Jersey, 1992.
8. Baddeley, A. D. and Salame, P. The unattended speech effect: Perception or memory? *J. exp. Psychol.: Learn. Mem. Cognit.* **12,** 525–529, 1986.
9. Baddeley, A. D. and Wilson, B. Phonological coding and short-term memory in patients without speech. *J. Mem. Lang.* **24,** 490–502, 1985.
10. Besner, D. Phonology, lexical access in reading, and articulatory suppression: A critical review. *Q. J. exp. Psychol.* **39,** 467–478, 1987.
11. Besner, D. and Davelaar, E. Basic processes in reading: Two phonological codes. *Can. J. Psychol.* **36,** 701–711, 1982.
12. Besner, D., Davies, J. and Daniels, S. Reading for meaning: The effects of concurrent articulation. *Q. J. exp. Psychol.* **33,** 415–437, 1981.
13. Bick, P. and Kinsbourne, M. Auditory hallucinations and subvocal speech in schizophrenic patients. *Am. J. Psychiat.* **144,** 222–225, 1986.

14. Bishop, D. V. M. and Robson, J. Unimpaired short-term memory and rhyme judgement in congenitally speechless individuals: Implications for the notion of "Articulatory Coding". *Q. J. exp. Psychol.* **41A**, 123–140, 1989.
15. Burani, C., Vallar, G. and Bottni, G. Articulatory coding and phonological judgements on written words and pictures: The role of the phonological output buffer. *Eur. J. cognit. Psychol.* **3**, 379–398, 1991.
16. Campbell, R., Rosen, S., Solis-Macias, V. and White, T. Stress in silent reading: Effects of concurrent articulation on the detection of syllabic stress patterns in written words in English speakers. *Lang. Cognit. Process.* **6**, 29-47, 1991.
17. Cooper, W. E., Blumstein, S. E. and Nigro, G. Articulatory effects on speech perception: A preliminary report. *J. Phon.* **3**, 87–98, 1975.
18. Crowder, R. G. Prefix effects in immediate memory. *Can. J. Psychol.* **21**, 450–461, 1967.
19. Crowder, R. G. Imagery for musical timbre. *J. exp. Psychol.: Hum. Percept. Perform.* **15**, 472–478, 1989.
20. Crowder, R. G. and Morton, J. Precategorical acoustic storage (PAS). *Percept. Psychophys.* **5**, 365–373, 1969.
21. Crowder, R. G. and Pitt, M. A. Research on memory/imagery for musical timbre. In *Auditory Imagery*, D. Reisberg (Editor), pp. 29-44. Lawrence Erlbaum, Hillsdale, New Jersey, 1992.
22. Crowder, R., Serafine, M. L. and Repp, B. Physical interaction and association by contiguity in memory for the words and melodies of songs. *Mem. Cognit.* **18**, 469–476, 1990.
23. Dallett, K. "Primary memory": The effects of redundancy upon digit repetition. *Psychon. Sci.* **3**, 237–238, 1965.
24. Daneman, M. and Stainton, M. Phonological recoding in silent reading. *J. exp. Psychol.: Learn. Mem. Cognit.* **17**, 618–632, 1991.
25. Demonet, J., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J., Wise, R., Rascol, A. and Frackowiak, R. The anatomy of phonological and semantic processing in normal subjects. *Brain* **115**, 1753–1768, 1992.
26. Duhamel, J. R. and Poncet, M. Deep dysphasia in a case of phonemic deafness: Role of the right hemisphere in auditory language comprehension. *Neuropsychologia* **24**, 769–779, 1986.
27. Dwyer, J., Blumstein, S. E. and Ryalls, J. The role of duration and rapid temporal processing on the lateral perception of consonants and vowels. *Brain Lang.* **17**, 272–286, 1982.
28. Fowler, C. and Rosenblum, L. D. The perception of phonetic gestures. In *Modularity and the Motor Theory of Speech Perception*, I. G. Mattingly and M. Studdert-Kennedy (Editors), pp. 33-59. Lawrence Erlbaum, Hillsdale, New Jersey, 1991.
29. Frith, C. D. *The Cognitive Neuropsychology of Schizophrenia.* Lawrence Erlbaum, Hillsdale, New Jersey, 1992.
30. Gathercole, S. E. and Baddeley, A. D. *Working Memory and Language.* Lawrence Erlbaum, Hillsdale, New Jersey, 1993.
31. Gould, L. N. Auditory hallucinations and subvocal speech: Objective study in a case of schizophrenia. *J. Ment. Dis.* **109**, 418–427, 1949.
32. Gould, L. N. Verbal hallucinations as automatic speech: the reactivation of dormant speech habit. *Am. J. Psychiat.* **107**, 110–119, 1950.
33. Gupta, P. and MacWhinney, B. Is the phonological loop articulatory or auditory? *Proceedings of the Cognitive Science Meetings,* Boulder Colorado, pp. 510–515. Lawrence Erlbaum, Hillsdale, New Jersey, 1993.
34. Gupta, P. and MacWhinney, B. Is the articulatory loop articulatory or auditory? Reexamining the effects of concurrent articulation on immediate serial recall. *J. Mem. Lang.* **33**, 1–26, 1995.
35. Halpern, A. Mental scanning in auditory imagery for tunes. *J. exp. Psychol.: Learn. Mem. Cognit.* **14**, 434–443, 1988.
36. Halpern, A. Musical aspects of auditory imagery. In *Auditory Imagery*, D. Reisberg (Editor), pp. 1–27. Lawrence Erlbaum, Hillsdale, New Jersey, 1992.
37. Johnston, R. S. and McDermott, E. A. Suppression effects in rhyme judgement tasks. *Q. J. exp. Psychol.* **38A**, 111–124, 1986.
38. Liberman, A. and Mattingly, I. The motor theory of speech perception revised. *Cognition* **2l**, 1–36, 1985.
39. Logie, R. H., Cubelli, R., Della Salla, S., Alberoni, M. and Michelli, P. Anarthria and verbal short-term memory. In *Developments in Clinical and Experimental Neuropsychology*, J. Crawford and D. Parker (Editors), pp. 203–211. Plenum Press, New York, 1989.
40. Logie, R. H. and Edworthy, J. Shared mechanisms in the processing of verbal and musical material. In *Imagery 2*, D. G. Russell, D. Marks and J. Richardson (Editors), pp. 33–37. Human Performance Associates, Dunedin, New Zealand, 1986.
41. Mohr, J. P., Pessin, M. S., Finkelstien, S., Funkenstein, H. H., Duncan, G. W. and Davis, K. R. Broca aphasia: Pathologic and clinical. *Neurology* **28**, 311–324, 1978.
42. Murray, D. The role of speech responses in short-term memory. *Can. J. Psychol.* **21**, 263–276, 1967.
43. Nairne, J. S. and Walters, V. L. Silent mouthing produces modality- and suffix-like effects. *J. Verb. Learn. Verb. Behav.* **22**, 475–483, 1983.
44. Nairne, J. S. and Crowder, R. G. On the locus of the stimulus suffix effect. *Mem. Cognit.* **10**, 350–357, 1982.
45. Nebes, R. N. The nature of internal speech in a patient with aphemia. *Brain Lang.* **2**, 489–497, 1975.

46. Ojemann, G. A. Brain organization for language from the perspective of electrical stimulation mapping. *Behav. Brain Sci.* **6,** 189–230, 1983.
47. Ostergaard, A. L. and Meudell, P. R. Immediate memory span, recognition memory for subspan series of words, and serial position effects in recognition memory for supraspan series of verbal and nonverbal items in Broca's and Wernicke's aphasia. *Brain Lang.* **22,** 1–13, 1984.
48. Paulesu, E., Frith, C. D. and Frackowiak, R. S. J. The neural correlates of the verbal component of working memory. *Nature* **362,** 342–345, 1993.
49. Peterson, S. E., Fox, P. T., Posner, M. I., Mintum, M. and Raichle, M. E. Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature* **331,** 585–589, 1988.
50. Reisberg, D. (Editor). *Auditory Imagery.* Lawrence Erlbaum, Hillsdale, New Jersey, 1992.
51. Reisberg, D. External representations and the advantages of externalizing one's thought. In *The Ninth Annual Conference of the Cognitive Science Society,* E. Hunt (Editor). Lawrence Erlbaum, Hillsdale, New Jersey, 1987.
52. Reisberg, D. and Logie, R. The ins and outs of working memory. In *Imagery, Creativity and Discovery,* M. Intons-Peterson, B. Roskos-Ewoldsen, R. Blake and K. Clayton (Editors), pp. 39–76. Lawrence Erlbaum, Hillsdale, New Jersey, 1993.
53. Reisberg, D., Smith, J. D., Baxter, D. and Sonenshine, M. "Enacted" auditory images are ambiguous; "Pure" auditory images are not. *Q. J. Psychol.* **41,** 619–641, 1989.
54. Reisberg, D., Wilson, M. and Smith, J. D. Auditory imagery and inner speech. In *Mental Images in Human Cognition,* R. Logie and M. Denis (Editors), pp. 59–81. Elsevier, Amsterdam, 1991.
55. Salame, P. and Baddeley, A. D. Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *J. Verb. Learn. Verb. Behav.* **21,** 150–164, 1982.
56. Salame, P. and Baddeley, A. D. Phonological factors in STM: Similarity and the unattended speech effect. *Bull. Psychon. Soc.* **24,** 263–265, 1986.
57. Selnes, O. A., Knopman, D. S., Niccum, N. and Rubens, A. B. The critical role of Wernicke's area in sentence repetition. *Ann. Neurol.* **17,** 549–557, 1985.
58. Shallice, T. and Butterworth, B. Short-term memory impairment and spontaneous speech. *Neuropsychologia* **15,** 729–735, 1977.
59. Shallice, T. and Warrington, E. K. Auditory-verbal short-term memory impairment and conduction aphasia. *Brain Lang.* **4,** 479–491, 1977.
60. Smith, J. D. The auditory hallucinations of schizophrenia. In *Auditory Imagery,* D. Reisberg (Editor), pp. 151–178. Lawrence Erlbaum, Hillsdale, New Jersey, 1992.
61. Stuss, D. T. and Benson, D. F. *The Frontal Lobes.* Raven Press, New York, 1986.
62. Vallar, G. and Baddeley, A. D. Fractionation of working memory: Neuropsychological evidence for a short-term store. *J. Verb. Learn. Verb. Behav.* **23,** 121–141, 1984a.
63. Vallar, G. and Baddeley, A. D. Phonological short-term store, phonological processing, and sentence comprehension: A neuropsychological case study. *Cognit. Neuropsychol.* **4,** 417–438, 1984b.
64. Vallar, G. and Cappa, S. F. Articulation and verbal short-term memory: Evidence from anarthria. *Cognit. Neuropsychol.* **4,** 55-78, 1987.
65. Vallar, G. and Shallice, T. *Neuropsychological Impairments of Short-term Memory.* Cambridge University Press, Cambridge, 1990.
66. Warren, R. Illusory changes of distinct speech upon repetition—the verbal transformation effect. *Br. J. Psychol.* **52,** 249, 1961.
67. Warren, R. *Auditory Perception.* Pergamon Press, New York, 1982.
68. Warren, R. and Gregory, R. An auditory analogue of the visual reversible figure. *Am. J. Psychol.* **71,** 612–613, 1958.
69. Warrington, E. K., Logue, V. and Pratt, R. T. The anatomical localisation of selective impairment of auditory verbal short-term memory. *Neuropsychologia* **9,** 377–387, 1971.
70. Waters, G. S., Caplan, D. and Hildebrandt, N. Working memory and written sentence comprehension. In *Attention and Performance XII,* M. Coltheart (Editor), pp. 531–555. Lawrence Erlbaum, Hove, 1987.
71. Waters, G. S., Caplan, D. and Hildebrandt, N. On the structure of verbal short-term memory and its functional role in sentence comprehension: Evidence from neuropsychology. *Cognit. Neuropsychol.* **8,** 81–126, 1991.
72. Waters, G. S., Komoda, M. K. and Arbuckle, T. Y. The effects of concurrent tasks on reading: Implications for phonological recoding. *J. Mem. Lang.* **24,** 27–45, 1985.
73. Waters, G. S., Rochon, E. and Caplan, D. The role of high-level planning in rehearsal: Evidence from patients with apraxia of speech. *J. Mem. Lang.* **31,** 54–73, 1992.
74. Wilding, J. and White, W. Impairment of rhyme judgements by silent and overt articulatory suppression. *Q. J. exp. Psychol.* **37A,** 95–107, 1985.
75. Wilson, M. and Emmorey, K. (in preparation). Linguistic working memory in a visuo-spatial modality: Evidence from similarity and suppression in ASL.

76. Zatorre, R. J., Evans, A. C., Meyer, E. and Gjedde, A. Lateralization of phonetic and pitch discrimination in speech processing. *Science* **256**, 846–849, 1992.
77. Zatorre, R. J., Halpern, A., Perry, D., Meyer, E. and Evans, A. Hearing in the mind's ear: A PET investigation of musical imagery and perception. *J. Cognit. Neurosci.*, in press.