

Habituation to irrelevant speech: Effects on a visual short-term memory task

NEIL MORRIS and DYLAN M. JONES

University of Wales College of Cardiff, Cardiff, Wales

The Baddeley and Hitch (1974) formulation of short-term or working memory embodied a phonological store within the articulatory loop component of the model. Later formulations specifically postulated an acoustic filter that endowed only speech-like stimuli with obligatory access to this phonological store. This paper presents evidence that this phonological store may have two filters, one of which is subject to habituation and can therefore attenuate the entry of irrelevant speech, thus undermining the obligatory access assumption of the model. An experiment is reported in which subjects were presented with a habituation period comprising 20 min of irrelevant speech—speech to be ignored by the subject—before a test phase in which a visually presented serial recall task with concurrent irrelevant speech was performed. The effect of irrelevant speech, which impairs performance on the primary task when there is no habituation phase, is reduced markedly in those conditions where the speech used in the habituation phase is the same as that used in the test phase, if the irrelevant speech is in a language different from that presented during the subsequent trials or if the habituator is a hummed version of the irrelevant speech passage. When a nonspeech sound (pink noise) is used in the habituation phase, a large irrelevant speech effect is found in the test phase. Morris, Quayle, and Jones (1989) found that humming did not produce an irrelevant speech effect, which suggests that the first filter is permeable to humming but that the second filter is not. The results of the habituation study indicate that the first filter is permeable to all speech sounds but not to other acoustic input and that it has some attenuating device. A second filter appears to extract more complex speech features and thus excludes humming. It is concluded that irrelevant speech does not have obligatory access to this phonological store when exposure is prolonged, and that it has unvarying spectral features.

There has recently been renewed interest in the effects of noise on short-term or working memory. Such research has had a checkered history when white noise has been employed as a source of interference, because there is some evidence to support the contention that the disruptive effect of white noise is confined to strategic processes and does not involve competition for memory resources (Smith, 1989). Thus, whether or not effects of white noise are found seems to depend largely on the demand characteristics of the primary task (Jones, *in press*). When human speech is used as a source of interference, however, a much clearer pattern of results is forthcoming.

Salamé and Baddeley (1982) used human speech as a source of distraction to examine the structure of working memory. In particular, they proposed that such unattended or irrelevant speech—that is, speech that subjects should simply ignore—has *obligatory* access to a phonological short-term memory store that is a component of the articulatory loop. The articulatory loop is a speech-based rehearsal loop with a long history as a concept to explain

the use of articulatory rehearsal in short-term memory (see, e.g., Glanzer, 1964; Levy, 1971; Wickelgren, 1969). In its latest incarnation (Baddeley & Hitch, 1974), it is composed of two components: a phonological store holding speech-like representations, and an articulatory control process, which may use the vocal apparatus to create articulatory representations that can be refreshed by repeated subvocalization (Reisberg, Rappaport, & O'Shaughnessy, 1984; Sokolov, 1972).

In the Baddeley and Hitch (1974) formulation, visually presented verbal material undergoes a process of grapheme-to-phoneme conversion, and this process is mediated by subvocal rehearsal. However, auditorily presented information enters the articulatory loop by a more direct means. Salamé and Baddeley (1982) propose that auditorily presented verbal material can bypass this grapheme-to-phoneme conversion process and gain direct entry to the articulatory loop's phonological store, because the material already consists of phonemes. Furthermore, as noted above, they propose that such input has obligatory access—that is, subjects cannot exclude this material from the phonological store, even when it is irrelevant to the performance of the primary task.

Salamé and Baddeley (1982, 1983) have also shown that the irrelevant auditory input must be speech-like in nature to produce disruption when the primary task requires verbal working memory; white and pink noise do not im-

The authors would like to thank Paul Lucas for running this experiment and Jason Shute for providing the vocals. In addition, we would like to acknowledge the helpful comments of two anonymous referees. Correspondence may be addressed to Neil Morris, Wolverhampton Polytechnic, School of Health Sciences, 62-68 Lichfield Street, Wolverhampton WV1 1DJ, England.

pair performance. In addition, Colle and Welsh (1976) have demonstrated that the intensity of the speech is not an important variable in determining the magnitude of the effect. Indeed the effect is not even dependent on whether or not the subject can understand the language in which the speech is presented (Jones, Miles, and Page, *in press*; Morris, Quayle, and Jones, 1989; Salamé & Baddeley, 1982, 1989). In fact, the only variable that has so far been shown to affect the magnitude of the effect when a short-term memory task (for example, serial recall of digits, consonants, or words) is used is phonemic similarity. When the irrelevant speech source is phonemically very similar to the phonemic representations of visually presented items in the memory task, the magnitude of the effect is greater than when the two streams of information are phonemically dissimilar (Salamé & Baddeley, 1982). Thus, neither the intensity of the speech nor its semantic properties seem to be important when the primary task requires little semantic processing (but see Jones *et al.*, *in press*, for an examination of the effects of the semantic properties of irrelevant speech when the primary task does involve extensive semantic processing).

The properties of the articulatory loop seem in general to have been mapped out fairly well (see Baddeley, 1986, and Morris, 1986, for extensive reviews of the evidence), and the phonological store in particular has now become the focus of attention of much further research. Indeed, the "irrelevant speech paradigm" is proving to be a useful tool in the examination of speech processing in working memory. This is probably because (1) the properties of the irrelevant speech can be easily manipulated, (2) it is not necessary to monitor the subjects' compliance in performing the secondary activity beyond ensuring that headphones are not removed, and (3) irrelevant speech, in general, produces effects of large magnitude, and the effects are confined specifically to verbal memory tasks (spatial memory, for example, is not disrupted by irrelevant speech; see Morris, Quayle, & Jones, 1988).

Overall, the recent literature largely supports the existence of a phonological store with a mechanism preventing material not composed of phonemes (or, indirectly, graphemes by conversion) from entering this store. Such a mechanism may be an auditory filter, or, alternatively, it may involve production-system-type elements (Newell & Simon, 1972) that are activated only when "speech" arrives. These possible mechanisms are functionally isomorphic with respect to the published literature. However, it is clear that simple speech sounds are not the unit of currency of the phonological store. Recent research by Morris *et al.* (1989) has shown that both speech and singing produce disruption, but that, critically, humming produces no irrelevant speech effect. Thus it appears that the "currency" of the phonological store has considerable linguistic complexity. In their study, humming was linguistically impoverished but was similar in prosody to the sung material. The study reported in this paper is relevant to this endeavor, but its primary purpose is to ex-

amine further the notion of obligatory access to the phonological store.

The contention that access is obligatory—that is, unsuppressible—is based on parsimony. Everyday experience, however, suggests that the entry of irrelevant speech to the phonological store is not obligatory at all times. Most cognitive processes requiring the articulatory loop (see Baddeley, 1986, and Morris, 1986, for recent reviews of the range of activities known to require the resources of the articulatory loop) are not seriously impaired in workplaces where there is a large volume of background conversation, and the "cocktail party effect" (Broadbent, 1958; Cherry, 1953) suggests that some filtering of speech occurs. How, then, can these findings be reconciled with a model of the phonological store that allows direct access to speech at all times?

The most obvious solution seems to be to postulate a filter that habituates to incoming speech unless the "portal" is maintained by some active process coupled with dishabituation occurring during periods of quiet. Hillyard, Squires, Bauer, and Lindsay (1971), for example, recorded auditory event-related potentials in subjects who listened to a series of tones in one ear and ignored a series of tones in the other. They found that there was a large positive component (P300) approximately 300 msec after infrequent stimulus changes in the attended ear but no comparable component for the unattended ear. They compared this to the "cocktail party" effect, and, by extrapolation, they suggested that there is a stimulus set for a particular speaker's voice and a response set to recognize the contents of the speaker's speech. This suggests that the filter can be tuned. However, in studies of irrelevant speech, only one voice is presented, so tuning may indeed be obligatory in these circumstances because such speech is the only candidate for a stimulus set. Nevertheless, a filter with the capabilities proposed by Hillyard *et al.* would have the capacity to exclude human speech, and extensive exposure to a single human voice may result in exclusion of that voice from the phonological store. Such filtration might require central, attentional resources, in which case one might expect habituation to any monotonous input, regardless of modality of entry or stimulus properties. This remains to be demonstrated, but what is clear is that disinhibition must occur in quiet, as a logical necessity; otherwise, prior exposure to a particular voice would render it useless as a source of irrelevant speech on another occasion. Our own studies, in which subjects have, of practical necessity, participated in a number of irrelevant speech experiments in which the same voice has been used repeatedly, continue to show large performance deficits with irrelevant speech.

The experiment reported in this paper was designed specifically to address the issue of habituation to irrelevant speech, in an attempt to isolate some of the features of speech that produce habituation. This was achieved by taking baseline measurements of memory performance in quiet, exposing four groups of subjects for 20 min to

noise during the interval. On Block 2 trials, all subjects again undertook a memory task, this time with Italian irrelevant speech (as used in Condition B). Finally, a posttest baseline measurement was undertaken in quiet. The experimental design is shown schematically in Figure 1.

The subjects were given no other task than to wear the headphones during the habituation phase. The experiment was carried out in a soundproofed room, and there was no written material present for the subjects to read. All the subjects commenced work on the trials in the test phase less than 1 min after the habituation phase had finished.

All subjects were fully briefed during the practice period, but they were not told what the predicted results were until all 48 subjects had been run. When all the data had been collected, the subjects were debriefed.

RESULTS

Since the hypotheses in these studies were postulated a priori, the appropriate, and most powerful, statistical tests were planned comparisons (Kirk, 1968). The predictions were that greater interference from irrelevant speech would be found in Conditions A and D (with humming and pink noise as habituators, respectively) and that less interference and thus more habituation would be found in Conditions B and C (Italian and English speech as habituators, respectively). Planned comparison analyses of variance were therefore performed on the mean number of consonants correctly recalled for each condition, using a design with three trial blocks and nine serial positions. Analysis of Condition A, with humming during the habituation phase, showed, contrary to predictions, no reliable differences between trial blocks [$F(2,22) = 2.42, p > .10$], a large effect of serial position [$F(8,88) = 23.75, p < .001$], and no interaction [$F(16,176) < 1$]. With Italian speech as a habituator (Condition B), a similar pattern of results appeared, with no reliable differences between blocks [$F(2,22) = 2.94, p > .10$], a main effect of serial position [$F(8,88) = 15.04, p < .001$], and no interaction [$F(16,176) = 1.19, p > .10$]. When English speech was used as a habituator (Condition C), there was no main effect of trial block [$F(2,22) = 1.62, p > .10$], a large effect for serial position [$F(8,88) = 14.05, p < .001$], and a significant interaction of trial blocks and serial position [$F(16,176) = 2.10, p < .01$]. The mean numbers correct for all blocks \times serial position are shown in Figure 2. Inspection of the panel for the English speech habituation condition suggests that the interaction occurred because performance at Serial Positions 5 and 6 was superior on Trial Block 3 as compared with the two earlier blocks. This was examined by repeating the analysis of variance for this condition with Block 3 trials removed. This analysis revealed no differences between Blocks 1 and 2 [$F(1,11) < 1$], a large effect of serial position [$F(8,88) = 12.72, p < .001$], and a loss of the interaction [$F(8,88) < 1$]. In the fourth condition (D), with pink noise as a habituator, there was a main effect of trial block, as predicted [$F(2,22) = 5.44, p < .025$], a main effect of serial position [$F(8,88) = 25.12, p < .001$], and no interaction [$F(16,176) < 1$]. It is clear

from Figure 2 that performance on Block 2, the block in which irrelevant speech was presented, was considerably poorer than on the other two blocks. The critical comparison, however, is between Block 1 and Block 2, so a further analysis of variance was performed on these two blocks. Even with reduced statistical power, there was still a main effect of block [$F(1,11) = 4.82, p < .05$] and an effect of serial position [$F(8,88) = 16.46, p < .001$], but no interaction [$F(8,88) = 1.14, p > .10$]. Tukey tests revealed that performance was poorer on Block 2 than on Block 1 ($p < .05$) or Block 3 ($p < .01$), and that performance was better on Block 3 than on Block 1 ($p < .05$).

Finally, a further analysis compared the first block of trials between conditions to see if the groups differed in baseline performance. This proved not to be the case. There was no main effect of condition [$F(3,44) < 1$] or interaction [$F(24,352) = 1.42, p > .05$]. The serial position effect was, of course, significant [$F(8,352) = 36.54, p < .001$].

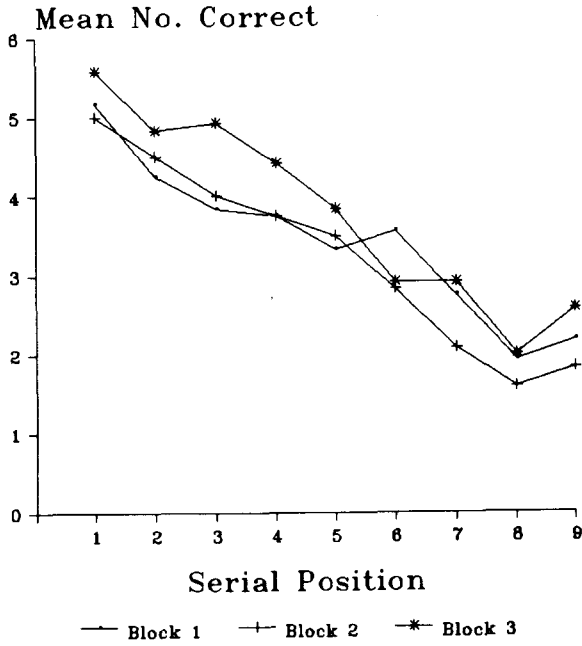
These results therefore provide a fairly clear picture. When subjects are exposed to 20 min of repetitive speech, they habituate to human speech sounds, and the unattended or irrelevant speech effect is lost. However, 20 min of pink noise produces little habituation, and an irrelevant speech effect of a magnitude comparable to that found by Morris et al. (1989) is demonstrated. These two studies are also of comparable statistical power. In addition, there is some evidence for a practice effect's having resulted in improved performance on Block 3 when pink noise was used as a habituator, and for a nonsignificant trend of improved performance on Block 3 trials when English and Italian speech were used as habituators, and this supports Poulton's (1973) caveat about repeated measures designs.

DISCUSSION

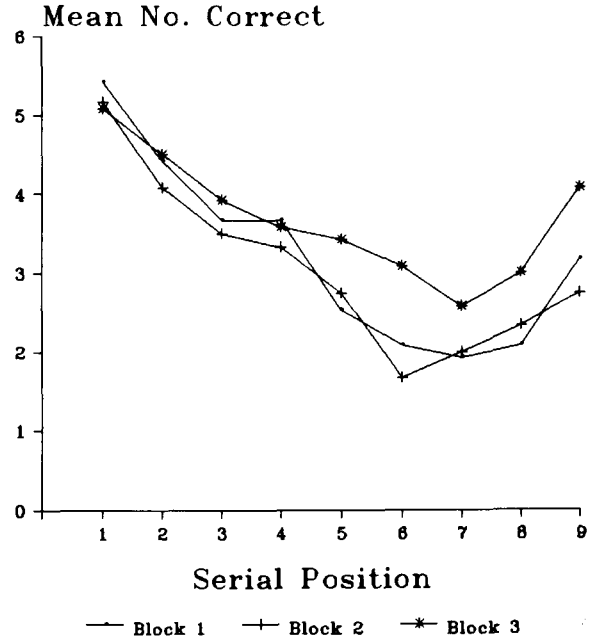
The most surprising result from this study is that humming, an irrelevant speech source that produces no irrelevant speech effect, produces a habituation effect. This result was not predicted, and it may shed some light on the nature of the acoustic filter. It suggests that there is indeed a filter that excludes extraneous acoustic sources from the phonological store, and that this filter is permeable to phonemes. However, it also suggests that in addition to this filter there is a second filter, located at some point beyond the acoustic filter, that allows access to the phonological store only when the input is phonemic in nature and when some other as yet unspecified properties are present in the input. These results also suggest that it is the first filter that habituates to the input. An alternative, but functionally equivalent, explanation is that input is parsed at least twice with more complex feature extraction occurring on the second parse.

Thus, there is strong evidence that irrelevant speech does not always have obligatory access to the phonological store. Rather, the entry of such speech to the store is dependent on the permeability, or attenuation, of the

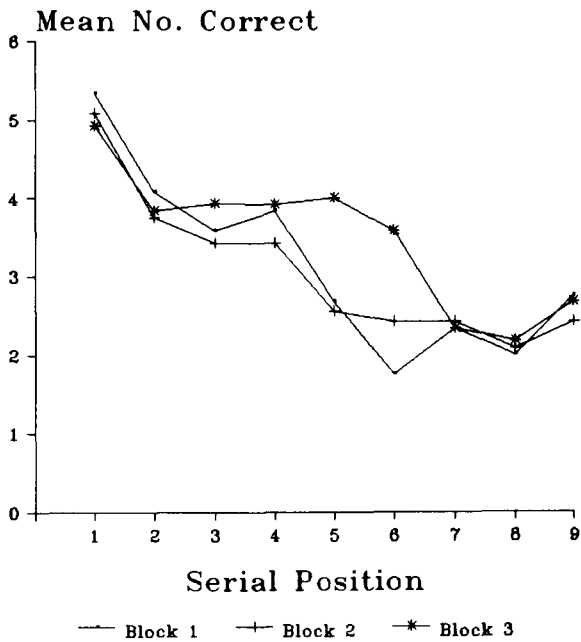
a. Humming



b. Italian



c. English



d. Pink Noise

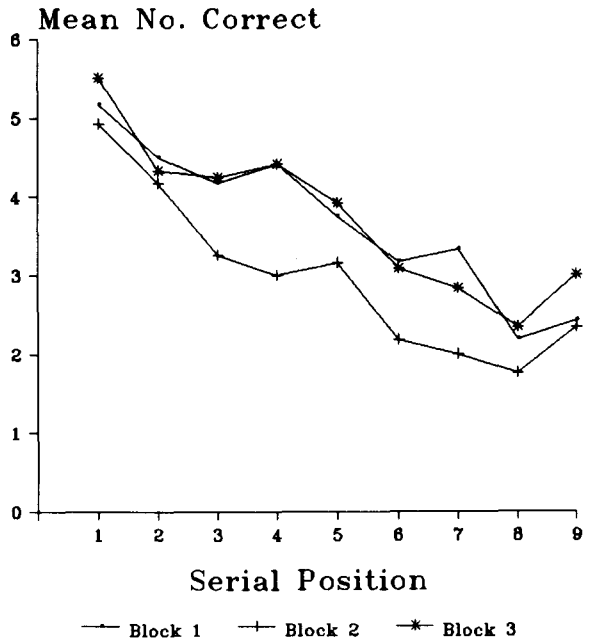


Figure 2. Mean number of consonants correctly recalled in serial order with (a) humming, (b) Italian speech, (c) English speech, and (d) pink noise as habituators.

filter at a given time. This attenuation increases when a continuous flow of speech enters the store and is maintained for some time (at least the duration of a block of trials, 3-4 min). The mechanism of dishabituation is not clear from these findings. It may be the case, for example, that dishabituation occurs when the irrelevant voice changes (e.g., from male to female; Cherry, 1953, found that in "shadowing" experiments a change of voice in the unattended ear was usually detected), so that attenuation may be reduced by other factors in addition to the passage of time.

The second obstacle to entry to the phonological store may be a mechanism that extracts attributes more complex than the phoneme. Such extraction could include, for example, morphemes, some prosodic aspects of speech, and other linguistic attributes. It is clear from the Morris et al. (1989) study, however, that humming is too impoverished to constitute speech with respect to this filter. Thus the phonological store seems to receive the residuum when acoustic input has been processed to exclude nonphonemic and phonemically impoverished material.

Figure 3 presents a possible schematic representation of the phonological store and places it in context with other findings in the working memory literature. On the left, acoustic input passes through the filter if the filter is not attenuated and if the input consists of phonemes. Phonemes that pass through the filter are analyzed by a further filter, which permits entry to the phonological store if the input has "speech-like" qualities. Exactly what constitutes speech-like qualities is in itself an interesting research topic that this study does not address, but it should prove to be an experimentally tractable problem.

Finally, the relationship of the phonological store to the articulatory control process is shown on the right. Note, however, that grapheme-to-phoneme translation is shown dissociated from the process of articulatory rehearsal. In the original Baddeley and Hitch (1974) formulation, grapheme-to-phoneme translation was mediated by articulation of visually presented items, but recent work by Bishop and her colleagues (see, e.g., Bishop & Robson, 1989) suggests that such translation can occur without articulatory rehearsal. Thus, neither obligatory access of speech to the phonological store nor obligatory articulation of visually presented verbal material is a necessary feature of the model.

Rather, a large body of literature now accumulating suggests that the articulatory loop is a more complex mechanism than originally envisaged. Nevertheless, it is probably the case that speech can have obligatory access to the phonological store, and that grapheme-to-phoneme conversion can be mediated by the process of articulation. It therefore seems that the original model requires modification and elaboration, rather than outright rejection. Recent work suggests that a more complex model will encourage experimentation that will lead to a more profound understanding of the nature of verbal processes in working memory.

Finally, these results suggest that some revisionary discussion of the central executive is in order. Baddeley (1981) commented that this is an "area of residual ignorance" (p. 21) about working memory. This component of the model was described by Baddeley (1986) as being the overseer of the memory process and responsible for, amongst other things, the strategic deployment of work-

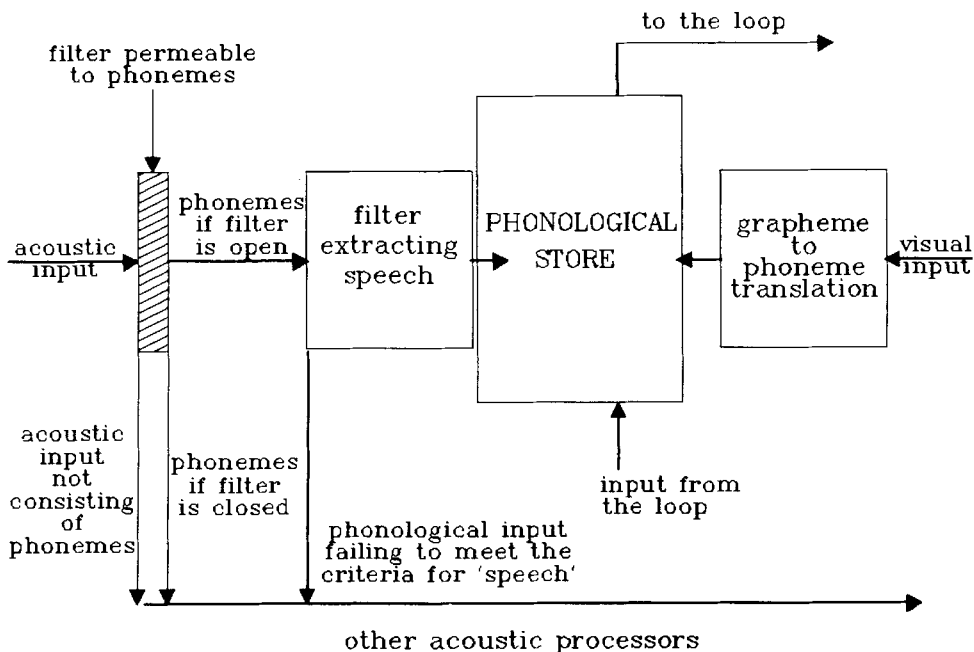


Figure 3. A possible model of the phonological store and its filters.

ing memory resources. He further suggested that it could be depicted as a production-system, and he proposed that the model outlined by Norman and Shallice (1986) was an appropriate candidate. In the Norman and Shallice model, most cognitive processing is initiated by existing schemata, which are activated automatically by built-in priorities and environmental cues. In addition to these schemata, the model has a supervisory attentional system that roughly corresponds to the central executive of working memory. One could model the phonological store and its associated filters as a device that has strong links with the executive at the level of the attenuating filter and built-in priorities in the production-system filter (i.e., the second filter). Thus, the attenuating filter would be "transparent" to novel material, but it would become "opaque" when the supervisory attentional system (i.e., the central executive), ceased to constantly monitor vocal input. It is therefore possible to accommodate these findings within the Baddeley (1986) framework by simply removing the obligatory access assumption and replacing it with the plausible assumption that vocal input is monitored by an attentional system that habituates to repetitive input, and it remains to be seen whether or not the recent elaboration of the model by Baddeley (1986) and Norman and Shallice (1986) will be necessary for our understanding of input selection and filtration procedures in short-term memory.

REFERENCES

- BADDELEY, A. D. (1981). The concept of working memory: A view of its current state and probable future development. *Cognition*, **10**, 17-23.
- BADDELEY, A. D. (1986). *Working memory*. Oxford, England: Oxford University Press.
- BADDELEY, A. D., & HITCH, G. J. (1974). Working memory. In G. Bower (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp. 47-90). New York: Academic Press.
- BISHOP, D. V. M., & ROBSON, J. (1989). Unimpaired short-term memory and rhyme judgement in congenitally speechless individuals: Implications for the notion of "articulatory coding." *Quarterly Journal of Experimental Psychology*, **41A**, 123-140.
- BROADBENT, D. E. (1958). *Perception and communication*. London: Pergamon.
- CHERRY, E. C. (1953). Some experiments on the recognition of speech with one and two ears. *Journal of the Acoustical Society of America*, **25**, 975-979.
- COLLE, H. A., & WELSH, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning & Verbal Behavior*, **15**, 17-32.
- GLANZER, M. (1964). The verbal-loop hypothesis: Conventional figures. *American Journal of Psychology*, **77**, 621-626.
- HILLYARD, S. A., SQUIRES, K. C., BAUER, J. W., & LINDSAY, P. H. (1971). Evoked potential correlates of auditory signal detection. *Science*, **172**, 1357-1360.
- JONES, D. M. (in press). Recent advances in the study of human performance in noise. *Environment International*.
- JONES, D. M., MILES, C., & PAGE, J. (in press). Disruption of reading by irrelevant speech: Effects of attention, arousal or memory? *Applied Cognitive Psychology*.
- KIRK, R. E. (1968). *Experimental design and procedures for the behavioral sciences*. Belmont, CA: Brooks/Cole.
- LEVY, B. A. (1971). Role of articulation in auditory and visual short-term memory. *Journal of Verbal Learning & Verbal Behavior*, **10**, 123-132.
- MORRIS, N. (1986). Working memory 1974-1984: A review of a decade of research. *Current Psychological Research & Reviews*, **5**, 281-295.
- MORRIS, N., QUAYLE, A. J., & JONES, D. M. (1988, September). *Irrelevant speech and working memory: Differential effects on a range of tasks*. Paper presented at the British Psychological Society: Cognitive Section Conference, University of Cambridge.
- MORRIS, N., QUAYLE, A. J., & JONES, D. M. (1989). Memory disruption by background speech and singing. In E. D. Megaw (Ed.), *Contemporary ergonomics 1989* (pp. 494-499). London: Taylor & Francis.
- NEWELL, A., & SIMON, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- NORMAN, D. A., & SHALLICE, T. (1986). Attention to action: Willed and automatic control of behaviour. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research and theory*. (Vol. 4, pp. 1-18). New York: Plenum.
- POULTON, E. C. (1973). Unwanted range effects from using within-subject experimental designs. *Psychological Bulletin*, **80**, 113-121.
- REISBERG, D., RAPPAPORT, I., & O'SHAUGHNESSY, M. (1984). Limits of working memory: The digit digit-span. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 203-221.
- SALAMÉ, P., & BADDELEY, A. D. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning & Verbal Behavior*, **21**, 150-164.
- SALAMÉ, P., & BADDELEY, A. D. (1983). Differential effects of noise and speech on short-term memory. In G. Rossi (Ed.), *Proceedings of the 4th International Congress on Noise as a Public Health Problem* (Vol. 2, pp. 751-758). Milan: Centro Ricerche e Studi Amplifon.
- SALAMÉ, P., & BADDELEY, A. D. (1989). Effects of background music on phonological short-term memory. *Quarterly Journal of Experimental Psychology*, **41A**, 107-122.
- SMITH, A. P. (1989). Noise, confidence ratings and recognition memory. In E. D. Megaw (Ed.), *Contemporary ergonomics 1989* (pp. 482-487). London: Taylor & Francis.
- SOKOLOV, A. N. (1972). *Inner speech and thought*. New York: Plenum.
- WICKELGREN, W. A. (1969). Auditory or articulatory coding in verbal short-term memory. *Psychological Review*, **76**, 232-235.

(Manuscript received July 7, 1989;
revision accepted for publication October 2, 1989.)