Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

# The Purity of Auditory Memory [and Discussion]

R. G. Crowder, N. Harvey and D. A. Routh

*Phil. Trans. R. Soc. Lond. B* 1983 **302**, 251-265
doi: 10.1098/rstb.1983.0053

**Email alerting service**

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

To subscribe to *Phil. Trans. R. Soc. Lond. B* go to: **http://rstb.royalsocietypublishing.org/subscriptions**

This journal is © 1983 The Royal Society

Downloaded from rstb.royalsocietypublishing.org

*Phil. Trans. R. Soc. Lond.* B **302**, 251–265 (1983)
*Printed in Great Britain*



# The purity of auditory memory

By R. G. Crowder†

*Center for Advanced Study in the Behavioral Sciences*, 202 *Junipero Serra Boulevard*, *Stanford*, *California* 94305, *U.S.A.*

Recent evidence from experiments on immediate memory indicates unambiguously that silent speech perception can produce typically 'auditory' effects while there is either active or passive mouthing of the relevant articulatory gestures. This result falsifies previous theories of auditory sensory memory (pre-categorical acoustic store) that insisted on external auditory stimulation as indispensible for access to the system. A resolution is proposed that leaves the properties of pre-categorical acoustic store much as they were assumed to be before but adds the possibility that visual information can affect the selection of auditory features in a pre-categorical stage of speech perception. In common terms, a speaker's facial gestures (or one's own) can influence auditory experience independently of determining what it was that was said. Some results in word perception that encourage this view are discussed.

## Introduction

The late Frank Restle (Restle 1974) authored what may be the most brilliant title ever used for an essay on memory: 'Critique of pure memory'. I admire his attaching this title to the theme that memory storage, putting information into something and getting it out later, is a pseudo-problem, obeying a false metaphor for psychology. I resent Restle's essay, too, of course, because now the title has been used up, as it were, and I have to make do with a less elegant one here. That question of purity is going to be one of the themes of this paper, rejection of memory as a separate faculty of cognition in favour of the view that memory is a by-product, a persistence, of information processing (a lesson that has been made forcefully by Craik & Lockhart (1972)). Evolution from a 'storage' to a 'processing' view of memory will be illustrated by specific reference to investigations of auditory sensory memory by myself and others. Here, the issue of purity becomes the restricted question of whether or not only acoustic information can gain access to auditory memory (as I have claimed in the past). There are four sections. The first two summarize previous models of pre-categorical acoustic storage (PAS), the third shows why these must now be rejected, and the fourth proposes a modified theory capable of handling the new evidence.

## PAS I

It seemed to Morton and me (Crowder & Morton 1969) that there were two facts requiring explanation by a theory of auditory sensory memory. First, it was known that immediate serial recall was affected by presentation modality (Conrad & Hull 1968; Corballis 1966; Murray 1964), with auditory presentation showing a sharp advantage at the terminal end of the list over visual presentation. Secondly, it had been demonstrated that a stimulus suffix (Crowder

† Permanent address: Department of Psychology, Yale University, New Haven, Connecticut 06520, U.S.A.



The Royal Society is collaborating with JSTOR to digitize, preserve, and extend access to *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences.*
www.jstor.org

Downloaded from rstb.royalsocietypublishing.org



1967; Dallett 1965) presented after the list more or less undid the benefit gained by going from visual to auditory presentation. The modality effect was interesting because we had all previously considered memory to be among the higher mental processes and thus protected, as it were, from peripheral bodily functions like vision and hearing. The suffix effect became especially interesting when it was shown that semantic factors had no influence on it but that changes in voice or physical location attenuated the effect (Morton *et al.* 1971).

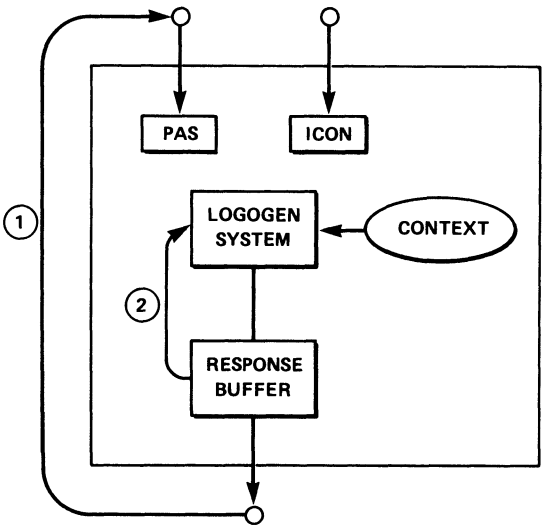


FIGURE 1. PAS I (after Crowder & Morton 1969). 1, Rehearsal aloud; 2, silent rehearsal.

In those days one responded to a new and important cluster of findings by proposing a new memory store. That of Crowder & Morton (1969) is shown in figure 1. It assumed parallel auditory and visual sensory memories occurring before a common lexicon called the logogen system. Access to the PAS system came either from direct auditory stimulation or from the overt (but definitely not covert) vocalization of visual information.

In retrospect, my favourite evidence binding the whole Crowder–Morton PAS package together was research on immediate memory as a function of phonetic class (Crowder 1971). (This research was appreciated by the speech-perception community but was largely overlooked by my colleagues in the field of memory). The evidence was that the occurrence of the suffix and modality effects depended on the type of item being remembered, If the list was assembled from items differing in place of articulation (BA, DA, GA or AB, AD, AG) there was no auditory advantage over visual presentation and the suffix had no selective interference effect at the end of the list. However, if steady-state vowels were the stimuli (BA, BOO, BEE, for example) presented in the lists, both the modality and suffix effects returned. Darwin & Baddeley (1974) showed that the vowel–consonant difference was not all-or-none but that evidence for PAS seemed to depend on the acoustical distinctiveness of the items being remembered.

The weight of these data in supporting the PAS model against alternatives is this: for one thing, the modality and suffix effects come and go together as a function of phonetic class. There are alternative explanations of both phenomena, independently, but this covariation suggests that they depend on the same mechanism, as Crowder & Morton proposed. Secondly, the results with stops and vowels point in the direction of a sensory system rather than some kind of short-term memory system. There has never been a stipulation in short-term memory



PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org



theories that the rules of operation are different depending on what kind of letters are used to make up the stimuli. However, in speech perception research it is a commonplace that there are profound differences between stops and vowels (Pisoni 1973), differences of exactly the sort observed here.

## PAS II

There were two main reasons for the revision of the theory some years later (Crowder 1978, 1981). The first was to assign a specific mechanism to the suffix. To say that the suffix 'masks' auditory traces of the last list item is, after all, no more than a description of the suffix effect itself. Masking could occur by a process of erasure or displacement, where the suffix obliterates the target; by integration, where the suffix combines with the target; by attentional distraction, where a central readout process is diverted; or by two forms of lateral inhibition, to be described below. Experiments presented in Crowder (1978) allowed choice among these in favour of the last.

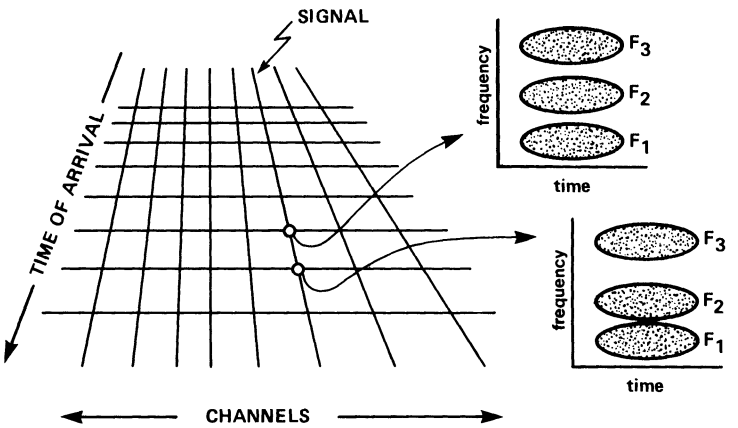


FIGURE 2. The 'grid model' of PAS representation (after Crowder 1978, 1981).

The second reason for proposing a revised PAS was to generalize the theory in the direction of speech perception. Here, there were two results of interest, both documented in (but not originally discovered by) Repp *et al.* (1979). It seemed obvious that A–X (same–different) discrimination of highly similar vowel sounds should depend on a system like PAS. Given two stimuli from the same class, say both tokens of /æ/, separated by 1 s or so, the subject should be able to judge their physical similarity only if he remembers the sound of the first vowel until the second arrives. Accordingly, performance is known to decline with increasing interstimulus interval between two items for comparison. Secondly, in vowel contrast, an ambiguous token, say somewhere between /i/ and /ɪ/, is presented in close proximity to one of the endpoint stimuli (/i/) and consequently sounds more like the other prototype (/ɪ/) than it would have in isolation. It was tempting to wonder whether the contrast might be produced by mechanisms inherent in PAS.

### *The grid model*

Figure 2 shows the working parts of the revised model. The memory representation is assumed to be laid out along a two-dimensional grid organizing auditory input by time of arrival and by source channel. It is straightforward to assume that input from different times could be represented, in some neurally spatial sense, as adjacent. Our understanding of the physical cues underlying source channel is very incomplete. From the old literature on dichotic listening and

Downloaded from rstb.royalsocietypublishing.org



shadowing, we know that two voices are on 'very different channels' if they belong to different genders, less so if they are different speakers of the same gender; we know that the same voice from two spatial locations can simulate two channels; and we know that speech and non-speech are typically different channels. To this point, the representation is informative only to the extent that it marks a channel as having been active at a certain time. This is not very useful in language comprehension or in memory experiments, though. Accordingly the new PAS model proposes that the content of a time–channel intersection, or node, be a rough spectrogram of the speech sound that occurred there. In the figure, there have been two utterances on a single channel, the vowel sounds /a/ and /æ/.

The final assumptions concern interaction of information on the grid. If the two inputs are extremely close in terms jointly of channel and time, it is assumed they will combine with each other, or integrate. Contrariwise, if they are extremely remote from one another in either or both of these dimensions, they will enjoy independent representation. At some intermediate spacing on the grid, however, they are assumed to interact according to principles of *frequency-specific recurrent lateral inhibition*. There are three parts to this last term: 'lateral inhibition', 'recurrent' and 'frequency-specific'. In this context, lateral inhibition simply refers to a process where constituents at some level in a hierarchical organization send out inhibitory connections to other constituents at that same level, as well as excitatory connections to the next level. To say that the lateral inhibition is recurrent means that the inhibition sent from one unit to its neighbour already reflects the inhibition that has come from the neighbour back to the original unit. The detailed workings of these two assumptions and their performance consequences for the suffix effect (disinhibition) are clearly explained in Crowder (1982*a*). The assumption that inhibition is frequency-specific (Crowder 1981) means that units within inhibitory range of each other will undergo lateral inhibition particularly where they share spectral energy. So, for example, if two units have one formant completely in common and another formant in totally different spectral regions, the former two will degrade each other but the latter will remain intact.

### *Application of the grid model to data*

In visual presentation, the grid is assumed to be in repose (but see below). Recency in auditory presentation is produced because the last item (1) has had less time to undergo inhibition that the others, and (2) has only one direction from which inhibition comes, unlike all the other, earlier, items. The second of these circumstances is reversed by the suffix item, provided that the suffix is a qualified inhibitor in terms of channel and time of arrival. Recent evidence indicates that presenting the suffix item about $\frac{1}{2}$ s after the start of the ultimate list item yields maximal lateral inhibition (Crowder 1982*b*). The restriction of PAS effects to certain phonetic classes, particularly their absence with stop consonants distinguished by place of articulation, is consistent with the assumption of crude spectrographic representation. In these, the subtle, fleeting, cues for place would be low in discriminability compared with the prolonged, steady-state formant shifts sufficient to cue vowel distinction (Darwin & Baddeley 1974). The disinhibition result (Crowder 1978, 1982*a*) is, not surprisingly, consistent with the grid model as well, because the result came before the model.

Application of the grid model to A–X discrimination and phonetic contrast in vowel perception is straightforward. By the assumption of frequency-specific inhibition, the more two items resemble each other, the more they should cause mutual inhibition; one can, therefore, model a decision rule were a process indexes how much degradation of the items there has been



PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org



to decide whether or not they were identical. If the two items are presented too far apart in time, the first will no longer be available for interaction with the second, and a decay function may be traced estimating the duration of auditory memory. Results of Crowder (1982*a*) suggest 3 s. One might expect that having the items as close as possible would facilitate this contribution of grid interaction to same–different judgements. The model predicts otherwise – if the two units are too close together, they will integrate rather than inhibit. If they integrate, the subject will lose valuable information (amount of lateral inhibition) sustaining same–different responding, and performance should suffer. This was indeed the result obtained by Pisoni (1973) and Crowder (1982, expt 1) in conditions where the two tokens overlapped. This is a modest victory for the grid model and I did not realize it at the time that I published the A–X delay experiments: in the suffix experiment, lateral inhibition is presumed to *hurt* performance, and overlapping the suffix in time with the target item *helps* performance relative to spacing the suffix by $\frac{1}{2}$ s or so. In the A–X discrimination task, lateral inhibition is assumed to *help* performance (as an index of similarity), and overlapping the two items now *hurts* performance relative to a spacing of $\frac{1}{2}$ s.

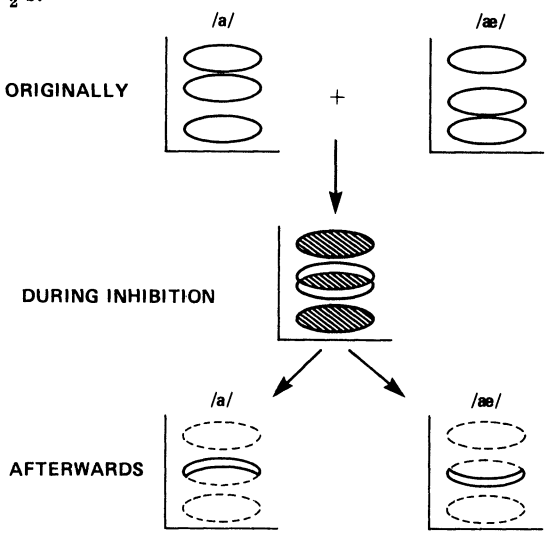


FIGURE 3. Graphic illustration of the prediction of phonetic contrast from the grid model of frequency-specific lateral inhibition in auditory memory.

Figure 3 illustrates predictions of phonetic contrast generated from the grid model. Here, frequency specificity is the important factor. When the vowel distinction in question is cued by a single formant and the two items in question have bandwidths around that formant centre frequency sufficient to overlap somewhat, then it follows that the two will emerge from their inhibitory interaction more distinct than they entered it. This is because the inhibitory process would have degraded their common spectral regions leaving only their distinct formant energy. It follows that if the two items are separated in time, they should exhibit less contrast than if they are placed at a separation maximizing lateral inhibition; this expectation is borne out by recent data (Crowder 1982*b*). It must be added immediately that although this derivation successfully predicts contrast, it predicts too much contrast! For example, the model obviously predicts that contrast should be symmetrical from either direction on the underlying continuum (effects of /i/ on an ambiguous token should resemble in size the effects of /ɪ/ on the same ambiguous token, provided that it is from midway between the two). But the fact is now inescapable that /i/ is a much more potent source of contrast than /ɪ/ (R. G. Crowder &



PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org



B. H. Repp, unpublished; Repp *et al.* 1979; Sawusch *et al.* 1980). Other experiments have shown that other continua do not seem to display contrast at all (Sawusch *et al.* 1980). The grid model provides an initial hypothesis for contrast, but it is clearly going to need adjustment.

Thus in taking seriously the responsibility for specifying the detailed operation of masking by the suffix and to explain how auditory memory articulates with speech discrimination and contrast, we emerge with a proposal (the grid model) that is much less a 'store' than its predecessor. The question arises of what the lateral inhibitory process suggested here is good for. It was always an embarrassment that the original PAS store seemed handy mainly for holding only the last sound before a pause, and this only if the sound were vocalic! The functional significance of the grid model is easier to argue: in vision, a system of recurrent lateral inhibition such as that from which the grid model was copied (Cornsweet 1970) has the obvious adaptive consequence of edge-sharpening. Something quite similar may go on in speech perception. For example, in rapid fluent speech, people rarely achieve the 'target values' of vowels, in terms of formant frequency. A system that could enhance the discriminability of adjacent vowels with high spectral overlap would be handy, especially if it operated at a very early, sensory, level of processing (an therefore required no attention). Admittedly, it is naïve and facile to seize on vowel contrast here, given that this is one of only two mechanisms to have been associated with the grid. The point does illustrate, however, the deeper lesson that it is easier to attach a functional, adaptive, significance to a 'processing view' of memory storage (Craik & Lockhart 1972) than to a 'pure memory' position.

## Speech gestures in memory and perception

Evidence to be presented in this section establishes beyond doubt that the two models for PAS advanced so far are worthwhile psychological theories in so far as they are capable of disproof. Without naming names, I would remind you here that there are some popular models these days whose falsifiability is in question.

In 1978, Spoehr & Corin performed a suffix experiment in which there was one condition that appended a silent, lip-read, suffix to an auditory list. This operation produced an essentially normal suffix affect relative to controls receiving visual–graphemic or auditory suffixes. Two years later, Campbell & Dodd (1980) reported a study of presentation modality in which lip-read sequences exhibited recency comparable with conventional auditory sequences, in comparison with visual-graphemic presentation. Neither data set was notable for its regularity, and so after a period of denial, I decided to replicate some of these operations in my laboratory.

In figure 4 are shown a set of data from R. L. Green & R. G. Crowder (in press) on both modality and suffix effects with lip-reading. There were two modes of presentation for the nine-digit stimuli and these were crossed orthogonally with three suffix events after the ninth digit, making six conditions altogether. In all cases, the subject watched a t.v. screen and saw a trained speaker pronouncing the stimuli. In half of the conditions, those shown in figure 4*a*, the numbers could be heard as well as seen, whereas in the other conditions (figure 4*b*) the sound was turned off and comprehension required lip-reading. Conditions were tested in blocks of ten trials. For each presentation mode, Visual and Audio-Visual, there was a No Suffix condition, in which the speaker lowers her eyes after the ninth memory item. In the other two suffix conditions, the suffix word was usually the word BEGIN, either presented audio-visually or only visually (lip-read). To ensure that attention was being paid to the suffix in all conditions,



PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org



the subject was warned that the suffix item would occasionally be the word START and that on those trials he should circle the trial number on his answer sheet. (Subjects had no trouble with this extra requirement).

The results show markedly poorer performance in the lip-reading conditions than in those where there was also sound; however, the patterns of results across suffix conditions were similar. The control conditions in each showed a last-position recency effect that rivalled first-position performance. Suffixes had similar damaging effects for both presentation modes, effects that were characteristically largest on the last position. There was a statistically reliable interaction between presentation and suffix mode: the audio-visual suffix was more damaging to audio-visual

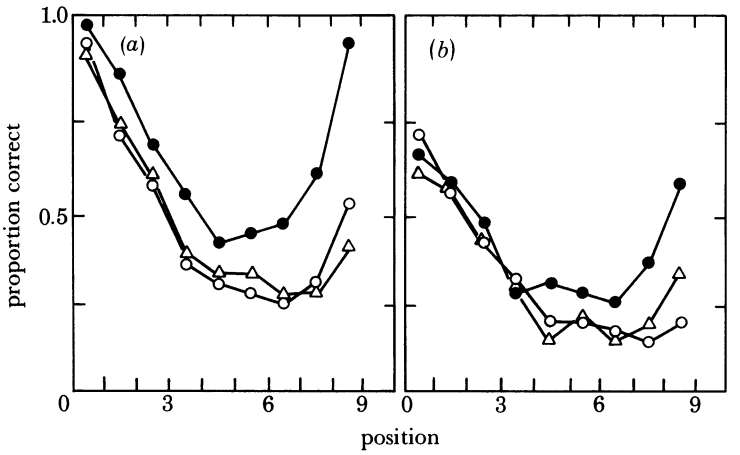


FIGURE 4. Correct items as functions of serial position, presentation modality (audio-visual (*a*) or lip-read (*b*)) and suffix modality (after R. L. Greene & R. G. Crowder, in press). ●, Control; ○ visual suffix; △, audio-visual suffix.

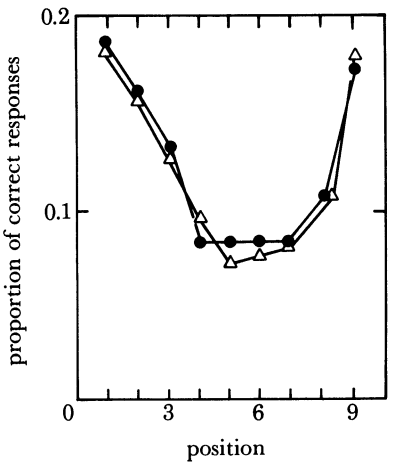


FIGURE 5. The control data of figure 4 normalized. ●, Visual only; △, audio-visual.

presentation and the visual suffix was more damaging to visual presentation than the mismatched suffix-presentation combinations. For the moment, however, it is the similarity of results in the two panels that is of more interest. Figure 5 shows a normalized representation of the two control conditions; here the areas under the two curves are made equal to emphasize comparisons of their shapes. Quite clearly, the lip-reading condition is an excellent approximation



PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org



to the genuine audio-visual condition. As Spoehr & Corin (1978) and Campbell & Dodd (1980) reported, watching someone else pronounce these items is tantamount to hearing it. Note well that in figure 4, a spoken suffix had a full suffix effect on the last item of a silent list (the Visual list, Audio-Visual suffix condition). This directly contradicts the model of Crowder & Morton and the results of their experiment II.

In 1970, I showed that active and passive vocalization worked about the same in modality–suffix experiments. In active vocalization, a written stimulus list is pronounced aloud by the subject and in passive vocalization, he hears the experimenter pronounce it. Both show the recency-sensitive pattern of results that is not evident in 'silent reading' or presumably subvocalized reading. Lip-reading could be described as 'passive silent mouthing' and so the

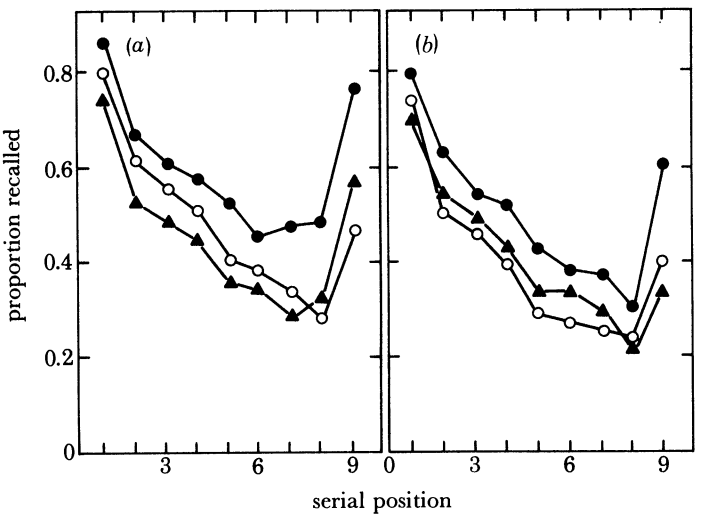


FIGURE 6. Correct items as functions of serial position, presentation modality (aloud (*a*) or mouthed (*b*)) and suffix modality. ●, Control; ○, aloud suffix; △, mouthed suffix.

question arises as to what would be the results of 'active silent mouthing' conditions, in which a written simulus would be translated into overt, but silent, speech movements by the subject. Recent results by Nairne & Walters (1983) have indicated that this modality, too, can effectively simulate audition. R. L. Green and I (Green & Crowder, in press) have replicated the Nairne & Walters finding too, with the results shown in figure 6. The experiment was laid out exactly as was our previous (figure 4) one, the three suffix conditions being crossed with two presentation modalities. This time, the subject either read the printed items aloud from a t.v. screen (left panel) or engaged in exaggerated silent mouthing of them (right panel). In this latter condition, they were to remain silent but to make lip movements sufficient for the experimenter to observe their accuracy of encoding the items, if he chose to. The same two conditions were applied to the suffix item, orthogonally, and compared with a no-suffix condition.

The results for active mouthing, here, were very similar to what they had been for passive mouthing (lip-reading) in the previous experiment. Mouthing lowered overall performance somewhat, presumably because it is an unnatural form of reading (for most of us); however, on the last item it produced a sharp recency effect, comparable in size with that of the ordinary spoken presentation condition. Again, the suffix interacted with the list-presentation mode such



PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org



that matching modalities led to somewhat larger decrements than mismatching. The normalized control data are shown in figure 7, emphasizing once again that once levels differences are corrected for, the condition with silent speech gestures is tantamount to real auditory presentation.

These are challenging results. They clearly refute the Crowder–Morton contention that the auditory system must be engaged for PAS to operate. I strongly believe that our first priority should be in settling the facts of the matter empirically, before getting committed to, and blinded by, strongly inferential hypotheses about theory. However, that said, it would be

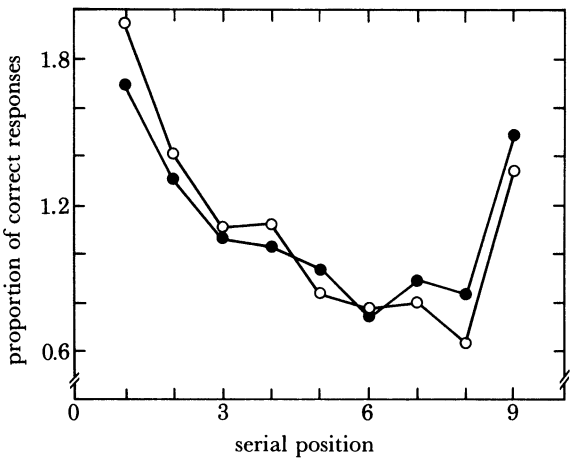


FIGURE 7. The control data of figure 6 normalized. ●, Aloud; ○, mouthed.

derelict to ignore other hypotheses pertinent to these experiments. Campbell & Dodd (1980) explained their suffix-modality effects with lip-reading by appealing to stimulus modalities that entail information that is slowly changing state. In this respect, audition and lip-reading are comparable, and visual–graphic information is excluded. But the mouthing data are inconsistent with any simple form of this hypothesis: in mouthing, the stimuli (letters on a screen) are presented in a fashion that allows instantaneous resolution. Indeed to mouth these items, the subject must already have fully categorized them. Thus, an extended categorization process must not be the critical factor. It would need to be assumed that the changing-state information comes from internal feedback produced by the subjects mouthing-responses and not from a primary pattern-recognition operation that is stretched out in time. Although Campbell & Dodd never addressed this issue, they would have to enlist the same feedback assumption to account for modality–suffix effects with overt mouthing of visually presented letters. Shand & Klima (1981) have argued from somewhat similar data on gestures from American Sign Language that the modality–suffix effects depend on items' being presented in a 'primary linguistic code'. According to this appealing idea, the translation from a visual–graphic to a (primary) speech-related code is what deprives ordinary visual presentation of the recency effect. However, again the mouthing data show a full recency effect and suffix effects too, even with a compulsory recoding stage present. Thus these authors, too, need to add assumptions about feedback from self-generated cues. In the next section, I explore one possible modification of the PAS model.



PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org



### PAS III

Figure 8 presents a modified system, which retains attractive features of earlier versions and yet allows a rationalization of the lip-reading and mouthing data. It should be obvious by now that I intend no implication of serial stages or stores by this layout; the diagram suggests possibilities for logical information flow, not spatial or chronological flow.

There are two major changes from earlier formulations. One that does not especially concern us here separates the logogen system into three components: two modality-specific input logogens and one output logogen. Morton (1979) was obliged to make this change in response to data showing poorer cross-modality priming than within-modality priming. The modification

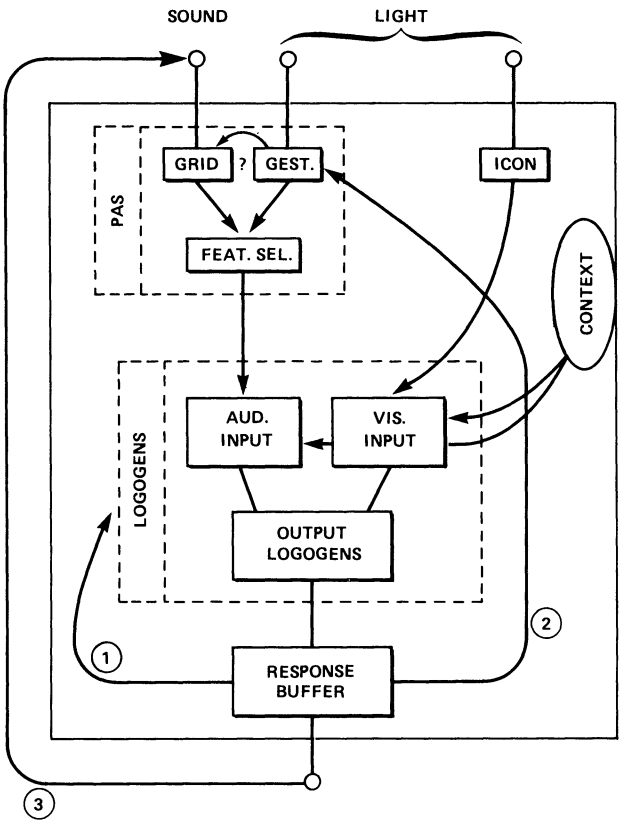


FIGURE 8. PAS III.

of interest here is the subdivision of PAS into three components: the grid, gestures, and auditory feature selection. Addition of feature selection is no particular departure, at least from what I now understand Morton to have envisaged all along for PAS. Feature selection was unmarked but assumed to be a part of the auditory analysis system from the start. (In this sense, PAS was always a process as well as a store.)

The new assumption follows a suggestion by Morton *et al.* (1981) allowing visual (gesture) information to enter into the selection of auditory features. Graphemic information is also visual, and it would of course control categorical selection (perception of letters, words, and so on), but it would never contribute to auditory experience in the way that perceived speech gestures would. In the absence of gestural information, the system works just as before, as a pure auditory



PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org



system. Now, however, the proposal is that the auditory feature selection can still operate in the absence of sound, namely in lip-reading.

Fundamental to this perspective is a distinction between (1) a decision as to *what the listener heard* and (2) a decision as to *what the speaker said*. The site of (1) is the PAS system, working on evidence from the grid, from perceived visual gestures, and possibly from other, as yet undiscovered, sources. The site of (2) is in the auditory input logogen system, accepting contextual information and information from PAS pertaining to auditory features. Ordinarily, the selection of auditory features would be controlled by auditory information on the grid. In exceptional cases such as lip-reading, supplementary information about gestures would control auditory feature selection. Lip-reading is not conceived here as an immediate visual-to-categorical process; rather, it is mediated by a sort of 'phantom audition'. (Of course, this applies to normally hearing individuals, such as those used in these studies, not to the deaf). That is, if a person is seen to make articulatory gestures appropriate to the word PIP, we first calculate the sound that these gestures would have produced and only then attach these sounds, via the logogen system for audition, to the target word.

Three forms of feedback are distinguished: in loop 1, the traditional form of subvocal (and unmouthed) rehearsal is given. It is assumed, because this 'normal rehearsal' occurs close to the logogen system, that it is at a high level of abstraction, perhaps at the phonological level (systematic phonemes). Like Salamé & Baddeley (1982) I associate the Conrad (1964) 'acoustic confusion' effects with this level. Loop 3 is the familiar vocalized rehearsal, which produces effects on PAS indistinguishable from external speech by another speaker (Crowder 1970). The remaining loop 2 is required by the data presented here on mouthing: it is assumed that when one silently reads and then mouths overtly, he perceives his own overt speech gestures as if he were watching those of someone else (in lip-reading). Thus two forms of silent articulation, if you will, lead to different behavioural consequences. The internal, abstract speech of loop 1 does nothing to set up activity in the sound analysis PAS system. The externalized but still silent speech of exaggerated mouthing from loop 2 does result in functional sounds, however, and yields us the classical PAS data.

Loop 2 and the gestural component have indeed destroyed the 'purity' of the PAS system. These aspects of the new model make PAS resemble in many ways what Massaro (1975) has called Synthesized Auditory Memory, a suggestion of his that I have not always so warmly endorsed. In particular, Massaro has anticipated a feedback loop allowing knowledge of rules to provide input to Synthesized Auditory Memory via rehearsal (Massaro 1975, pp. 599–602). It should be stressed that Massaro's store and the PAS system now being advanced remain authentically pre-categorical in Morton's original sense of operating before distinctions based on learned linguistic categories. The gesture-to-sound and sound-to-gesture rules that are now allowed into the auditory system rely presumably on associations formed through infantile babbling and not the categorization process. (But note that Kuhl & Meltzoff (1982) have discovered a very similar gesture–sound compatibility, for vowels, with infants 18–20 weeks old!)

A natural question for the model of figure 8 is where, now, we are to place the loci of 'all-auditory' and cross-modal suffix effects (auditory list with either auditory or lip-read suffix, respectively). If the detailed operations of the grid model are to remain intact in the new system, there seem to be only two options. First, we could allow gesture information to invade the grid itself; this is shown by the arrow with a question mark in figure 8. The functional basis for



PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org



this connection would be the associations between speech gestures and acoustic patterns cited above. By this hypothesis, a silent speech gesture (including one's own) could actually result in an entry on the channel–time coordinates of the grid. This obviously would surrender the last remaining 'purity' of the sensory system to audition and the grid itself would immediately become a far more abstract piece of machinery than originally conceived.

A second option for explaining the cross-modal suffix effect in terms of the revised PAS model would be to replicate some properties of the grid within the feature-selection system. This alternative has the awkwardness of overpredicting ordinary suffix interference: it would have to be claimed that whereas the auditory–auditory suffix effect occurs within the grid system, the auditory–lip-read suffix effect – although it is all but identical empirically – occurs one stage further down in the system, at feature selection. It would then remain to be seen which of the effects rationalized by the grid would also occur with the cross-modal suffix effect. For example, would the timing of a lip-read suffix following an auditory list follow the same inverted-U shaped function established for the intramodal effect (Crowder 1978)? Would the cross-modal effect show disinhibition, channel differences, and so on? Pending resolution of some of these issues, it must be said that the model of figure 8 is more a framework for revision of the PAS model than it is a way out of its difficulties.

Now finally, what can be said to rescue the proposed PAS III model from the stinging criticism that it is nothing but a gratuitous epicycle? A particularly impressive illustration of the distinction between auditory feature selection and categorization has been arranged by McGurk & MacDonald (1976; MacDonald & McGurk 1978). A series of acoustically uniform CV syllables, perhaps /ga/, is roughly synchronized with a video segment showing a speaker pronouncing a varied list of CVs such as /ba, fa, ma, da, .../. The striking phenomenal impression is that one *hears* a varying series matching the video input up to the point where the video syllables are normally produced in a concealed manner (/ga, ka, .../). By looking away from the screen, one easily verifies that the acoustics are invariant but the gestural information makes them sound different. It is not that one works out that the speaker must have said /fa/ given his mouthed gestures, it is rather, to everyone I know who has seen the demonstration, a genuine auditory experience. Our immediate memory studies in a way confirm that the experience is auditory, for the visual gestures are shown empirically to be tantamount to auditory stimulation.

An experiment by Ayres *et al.* (1979) has always been troublesome for the PAS model. They showed that whether or not the same sound (WAH) was labelled as a muted trumpet or as speech made all the difference as to whether it interacted with auditory memory. Because the sound was physically identical in the two cases, it should have operated identically on the grid. If, however, the PAS system is now considered specialized for the selection of speech features, as the argument here increasingly points to, then it would simply lie dormant when feature selection is not at stake, as with non-speech sounds (see also Morton *et al.* 1981).

Thus, in concluding summary, we have seen the purity of the PAS system fundamentally compromised; however, in return, the short-term memory phenomena that have driven that PAS hypothesis from the beginning are now even more a part of a normally functioning perceptual processing system.

This paper was prepared while I was at the Center for Advanced Study in the Behavioral Sciences and supported by N.S.F. grant no. BNS8206304. The research itself was supported



PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org



by N.S.F. grant no. BNS800538 to R. Crowder and by grants nos NICHD HD01994 from the National Institutes of Health and BRS RR05596 from the National Science Foundation to Haskins Laboratories. I appreciate the comments of Robert L. Green and Dominic Massaro on an earlier version of this paper.

## References

Ayres, T., Jonides, J., Reitman, J. S., Egan, J. C. & Howard, D. A. 1979 Differing suffix effects for the same physical suffix. *J. exp. Psychol.: hum. Learn. Memory* **5**, 315–321.

Campbell, R. & Dodd, B. 1980 Hearing by eye. *Q. J. exp. Psychol.* **32**, 85–99.

Conrad, R. 1964 Acoustic confusions in immediate memory. *Br. J. Psychol.* **55**, 75–88.

Conrad, R. & Hull, A. J. 1968 Input modality and the serial position curve in short-term memory. *Psychonom. Sci.* **10**, 135–136.

Corballis, M. C. 1966 Rehearsal and decay in immediate recall of visually and aurally presented items. *Can. J. Psychol.* **20**, 43–51.

Cornsweet, T. N. 1970 *Visual Perception*. New York: Academic Press.

Craik, F. I. M. & Lockhart, R. S. 1972 Levels of processing: a framework for memory research. *J. verb. Learn. verb. Behav.* **11**, 671–684.

Crowder, R. G. 1967 Prefix effects in immediate memory. *Can. J. Psychol.* **21**, 450–461.

Crowder, R. G. 1970 The role of one's own voice in immediate memory. *Cogn. Psychol.* **1**, 157–178.

Crowder, R. G. 1971 The sounds of vowels and consonants in immediate memory. *J. verb. Learn. verb. Behav.* **10**, 587–596.

Crowder, R. G. 1978 Mechanisms of auditory backward masking in the stimulus suffix effect. *Psychol. Rev.* **85**, 502–524.

Crowder, R. G. 1981 The role of auditory memory in speech perception and discrimination. In *The cognitive representation of speech* (ed. T. Myers, J. Laver & J. Anderson), pp. 167–179. Amsterdam: North-Holland.

Crowder, R. G. 1982*a* Disinhibition of masking in auditory sensory memory. *Memory Cogn.* **10**, 424–483.

Crowder, R. G. 1982*b* Decay of auditory memory in vowel discrimination. *J. exp. Psychol.: Learn. Memory Cogn.* **8**, 153–162.

Crowder, R. G. & Morton, J. 1969 Precategorical acoustic storage (PAS). *Percept. Psychophys.* **5**, 365–373.

Dallett, K. 1965 'Primary memory': the effects of redundancy upon digit repetition. *Psychonom. Sci.* **3**, 237–238.

Darwin, C. J. & Baddeley, A. D. 1974 Acoustic memory and the perception of speech. *Cogn. Psychol.* **6**, 41–61.

MacDonald, J. & McGurk, H. 1978 Visual influences on speech perception processes. *Percept. Psychophys.* **24**, 253–257.

McGurk, H. & MacDonald, J. 1976 Hearing lips and seeing voices. *Nature, Lond.* **264**, 746–748.

Massaro, D. M. 1975 *Experimental psychology and information processing*. Chicago: Rand McNally.

Morton, J. 1979 Facilitation in word recognition: experiments causing changes in the logogen model. In *Processing of visible language*, (ed. P. A. Kolers, M. E. Wrolstad & H. Bouma), vol. 1, pp. 259–268. New York: Plenum.

Morton, J., Crowder, R. G. & Prussin, H. S. 1971 Experiments on the stimulus suffix effect. *J. exp. Psychol.* **91**, 169–190.

Morton, J., Marcus, S. & Ottley, P. 1981 The acoustic correlates of 'speechlike': a use of the suffix effect. *J. exp. Psychol. gen.* **110**, 568–593.

Murray, D. J. 1966 Vocalization-at-presentation, auditory presentation and immediate recall, with varying recall methods. *Q. J. exp. Psychol.* **18**, 9–18.

Nairne, J. S. & Walters, V. L. 1983 Silent mouthing produced modality-and suffix-like effects. *J. verb. Learn. verb Behav.* (In the press.)

Shand, M. A. & Klima, E. S. 1981 Nonauditory suffix effects in congenitally deaf singers of American Sign Language. *J. exp. Psychol.: Hum. Learn. Memory* **7**, 464–474.

Pisoni, D. B. 1973 Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Percept. Psychophys.* **13**, 253–260.

Repp, B. H., Healy, A. F. & Crowder, R. G. 1979 Categories and context in the perception of isolated, steady-state vowels. *J. exp. Psychol.: Percept. Perform.* **5**, 129–143.

Restle, F. 1974 Critique of pure memory. In *Theories of cognitive psychology: the Loyola symposium* (ed. R. L. Solso), pp. 203–217. Potomac, Maryland: Lawrence Erlbaum Associates.

Salamé, P. & Baddeley, A. D. 1982 Disruptions of short-term memory by unattended speech: implications for the structure of working memory. *J. verb. Learn. verb. Behav.* **21**, 150–164.

Sawusch, J. R., Nussbaum, H. C. & Schwab, E. C. 1980 Contextual effects in vowel perception: evidence for two processing mechanisms. *Percept. Psychophys.* **27**, 421–434.

Spoehr, K. T. & Corin, W. J. 1978 The stimulus suffix effect as a memory coding phenomenon. *Memory Cogn.* **6**, 583–589.



PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org



*Discussion*

N. Harvey (*Department of Psychology, University College London, U.K.*). The McGurk effect appears to be a phenomenon restricted to stop consonants. Professor Crowder explicitly excluded analysis of stop consonants from the processing carried out by the crude spectrographic grid in pre-categorical acoustic storage. Thus I cannot accept Professor Crowder's claim that the McGurk effect can be explained by arguing that it results from a coalition between the grid and a gesture processor.

R. G. Crowder. Dr Harvey's point is well taken: the McGurk effect is indeed a demonstration about the perception of consonants whereas the PAS 'story' seems to be about vowels, for the most part. However, I do not know that the McGurk effect is absent for vowels when one tries for it. In a recent report by Kuhl & Meltzoff (1982) infants of from 18 to 20 weeks were shown to be sensitive to acoustic and articulatory–visual correspondences for the vowels /a/ and /i/. Moreover, we should expect that such bimodal support for speech perception with consonants would show up in immediate perception but not persist long, or at all, in PAS, because of the well known differences in auditory decay for vowels and consonants (Pisoni 1973). In other words, consonants are, after all, *heard*, and the auditory experience would not have to last long for the grid–gesture coalition to occur and affect categorization.

*Reference*

Kuhl, P. S. & Meltzoff, A. N. 1982 The bimodal perception of speech in infancy. *Science, Wash.* **218**, 1138–1141.

D. A. Routh (*Department of Psychology, University of Bristol, U.K.*). Sometimes, it strikes me that theories are not unlike battleships. If they survive for long enough, then they tend to be patched up to be come floating museums, or training ships. However, some of them really ought to be sunk.

First of all, let me deal with Professor Crowder's recurrent lateral inhibition model. If we consider the pattern of results that have emerged from a range of experiments using the delayed suffix paradigm, together with the existence of what look like 'full-blown' suffix effects at rather slow rates of presentation, that it seems fairly clear that the maximal suffix is obtained when a suffix is delivered on the next rhythmic beat after a list of items. In addition, my colleague Dr Clive Frankish and his student Dr Judith Turner have recently obtained the usual delayed suffix function with high-speed speech, presented at a rate of around 10 digits per second. This finding completes the pattern nicely, and there can be little room for doubting that the delayed suffix function depends upon *relative* rather than absolute time. Professor Crowder's theory appears to be irrevocably committed to processes operating in absolute time.

Next, turning to Professor Crowder's second theory, he appears to have modified the original conception of PAS in order to accommodate recent findings obtained by using video-speech (lip-reading). However, this change is only necessitated if one retains a commitment to the idea that strong recency is a trustworthy symptom for PAS. Professor Morton and I have a collection of experiments, as does Frankish, that demonstrate that strong recency is fairly mobile. One may observe it at non-terminal serial positions under conditions where a PAS theory could not possibly apply. This suggests that strong recency might better be regarded as resulting from a suprasegmental process, and that one should look elsewhere for the symptoms of an accessory acoustic representation. My own preference of course is to look in the direction of 'across-



PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org



the-board' modality effects, and it is interesting that comparisons of serial recall for audio- and video-speech sequences do throw up just this sort of effect. Of course, there may well be a greater psychological cost in deriving a segmental representation in the case of video-speech, and this possibility will have to be investigated. However, this seems much more preferable than camouflaging the effect, as Professor Crowder has done with his data, by resorting to normalized serial position curves.

R. G. Crowder. First, I agree with Dr Routh that the results of delayed suffix experiments are important for the revised theory of PAS. That is of course why, in the article introducing that theory (Crowder 1978), I reported a large experiment with orthogonal variation in presentation rate and suffix delay. That experiment seemed to show some influence of rhythm at slow rates of presentation, which was acknowledged in the paper. The more striking (and larger) finding was that for all presentation rates, there was an inverted-U shaped masking function with the maximum effect at around $\frac{1}{2}$ s. I certainly would not have offered the theory publicly without that empirical encouragement. Naturally, I am anxious to see the unpublished work by Frankish & Turner on this problem, especially if their experiments included even a larger range of orthogonally and parametrically varied rates and delays.

Again, without access to the unpublished data by Routh & Morton and by Frankish, I cannot comment in detail on the other point (that recency is not 'a trustworthy symptom for PAS'). Surely Dr Routh would agree that there are relatively trivial reasons that one could obtain an 'across-the-board' modality effect. For example if information coming over one modality is much harder to understand than another, we would expect such a difference. Well, lip-reading *is* much harder than listening to clearly spoken speech with the speaker's face in view. Dr Routh implies that he would require experimental evidence on this point. I rest my case on (1) the overwhelming intuitive plausibility that turning the t.v. sound off makes the message harder to understand, (2) the unanimous testimony of subjects who have been asked, and (3) the empirical fact that memory was badly impaired even for the first list position, which to me has always indicated a perceptual rather than a memory difficulty.

The practice of normalization is justified by the hypothesis that PAS adds a new source of information to 'regular short-term memory'. We do indeed want to correct for, or 'camouflage', the factors, like difficulty, that greatly affect the level of short-term memory but not the PAS contribution.



PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES