REACTION TIME MEASURES OF PERCEPTUAL AND LINGUISTIC FACTORS IN A PHONEME MONITORING TASK

REACTION TIME MEASURES OF PERCEPTUAL AND LINGUISTIC FACTORS IN A PHONEME MONITORING TASK

FRANK DIRK BOKHORST

Thesis submitted to the Department of Psychology, University of Cape Town, in fulfilment of the requirements for the degree of Master of Arts.

SUPERVISOR: PROF. PETER du PREEZ.

CAPE TOWN
South Africa
October, 1982.

ACKNOWLEDGEMENTS

Special thanks to Peter my supervisor, Jane the typist, Gudrun my wife, Andrew the printer, and lastly to UNIVAC 1100.

ABSTRACT

In terms of a model of cognition that distinguishes between automatic and controlled processes, phoneme monitoring is viewed as comprising both consciously controlled target search and automatic speech processing. Facilitation or inhibition can occur between each of these, but previous research has not yet taken this into consideration . In order to isolate these potentially confounding effects, a new method of presenting speech recordings was used. The listener's attention to target location cues in a phoneme monitoring task was directly manipulated, and performance compared relative to a cue-only monitoring task. Temporal parameters of the recorded speech were measured in detail. It was found that on average, phoneme monitoring latency was shorter than the duration of target-bearing words, yet reliably exceeded cue-only monitoring latency. Analysis based on a dichotomy between semantically appropriate and inappropriate target bearing words failed to produce evidence for facilitation or inhibition of performance. Using a quantitative measure of on-line semantic predictability in a multiple regression analysis, stronger evidence of semantic facilitation was found; the length and frequency of target-bearing words had no effect on phoneme monitoring performance. A reliable effect was discovered that could have been the result of backward auditory masking, or target syllable accentuation (stress). The model used in this analysis was evaluated for non-additive components between phoneme and cue-only monitoring tasks. Although not highly reliable, non-additivity did contribute to variance explained by the model. The results were interpreted as evidence that consciously controlled processing was not subject to divided attention deficits. When attention was directed to target location cues, phoneme monitoring performance was facilitated by semantic predictability. This showed that listeners were unable to dissociate phonological and phonetic representations of the speech input. The obligatory nature of automatic speech processing was demonstrated.

Table of contents

1. GENERAL

1.1. INTRODUCTION

2. THE EXPERIMENTS

Table of contents

5. APPENDIX

PREAMBLE

What is the relation between sound and meaning in speech communication? If we restrict the scope of this question to the receiving of communication, then it seems most obvious that the sound serves to specify the meaning. That is, the acoustic input "causes" the perception of meaning in a listener who "receives" the input. On the other hand, the contribution of the listener becomes obvious when we consider the case of two people exposed to the same pattern of sound, for example the word "dog" in Portuguese, but where the one person does and the other does not "know" Portuguese. To one listener the physical sound will be immediately "transparent" (c.f. Polyani, 1964), while for the other it will "fall on deaf ears". Obviously, then, receiving speech communication is an active process. For the present purpose, we shall take the conclusion above as a point of departure for an investigation into another aspect of the opening question. Namely, while the sound may specify the meaning, to what extent is the reverse also true? To what extent does the receiver of spoken communication "hear" just such sounds as are specified by the meaning? With respect to visual perception, experiments abound that illustrate how the visual context of an event influences the perception of that event. For example, Palmer (1975) showed that identification of objects was facilitated by the presence of contextual scenes. The superior identification of letters in words as opposed to letters in non-words was demonstrated by Reicher (1969) and by many others subsequently. In the case of speech perception stronger effects have been reported: the often-quoted study by Warren (1970) showed that certain phonemes in fluent speech were heard, even though these had been physically replaced by a cough.

This kind of finding accords well with the theoretical position that the perceiver "constructs" the perceptual object (e.g. Neisser, 1967, 1972). Transformational-generative grammar describes how in language comprehension it is the phonological component of the grammar that "constructs" the phonetic segments heard. Transposed into the domain of psycholinguistics, this position has been associated with analysis-by-synthesis theories of speech perception (e.g. Stevens, 1960).

Adopting the information-processing metaphor, the terms "top-down" or "conceptually-driven" are used to qualify processes such as analysis-by-synthesis. Conversely, insofar as the physical medium of communication is the datum required for comprehension, "data-driven" or "bottom-up" processes would be responsible for a more directly causal relation between input and percept. Experimental evidence for such a relation is not lacking. In the case of visual perception for example, Johnston and McClelland (1980) argue that their findings are consistent with just such a "bottom-up" processing model of word identification, and explain Reicher's (1969) word-superiority effect without appeal to "top-down" effects. In the field of speech perception, widely divergent findings have been reported. Some of these will be reviewed in later sections, where it will become apparent that there is no simple answer to the opening question above.

Whilst it may be true that the perceiver constructs the preceptual object in an active manner, and is not passively "stimulated" into comprehension of speech, it is doubtful that this applies to all levels, from acoustic energy to semantic interpretation. As Neisser (1967) points out, for analysis-by-synthesis to work, a certain degree of processing must precede construction of the object in order to reduce drastically the number of alternatives to be "hypothesised".

Evidence for the existence of neural detectors for phonetic features (Eimas and Corbit, 1973)suggests that at this early level perception may be more "passive" than "active". Similarly, there is much evidence favouring the existence of passively activated "word-detectors" (Morton's "logogens"; 1969, 1964). It might be true that the only level at which it makes sense to speak of constructive processing is that of semantic interpretation; all the rest would be what Neisser terms the "pre-attentive" processes.

To explain findings such as those by Warren (1970), it should not be said that the missing phonemes were constructed by the listener, but that "top-down" information activated lower-level units coding the missing segments. This suggestion is similar to the explanation of contextual effects on word-identification within the "logogen" model (Morton, 1969). The "logogen" units passively transmit information "upwards", but may be facilitated by actively generated input "downwards" from the contextual system: they are the locus of "top-down" and "bottom-up" interaction. The operation of such facilitation would seem to be of the kind termed "automatic" by Posner and Snyder (1975). By this we mean facilitation of a pathway which occurs with no corresponding inhibition of collateral pathways (see section 1-4 below). It will be seen later that understanding interactions between levels of processing leads to an examination of the terms "automatic" and "controlled" with respect to sequences of mental operations.

Our present concern is with a level prior to word identification, namely the processes that would feed into the "logogens". We ask, can semantic contextual factors influence the identification/perception of phonetic segments or phonemes? If phonetic segments, as opposed to phonemes are not perceptual constructs - the result of phonological transformations, then how could semantic context operate "downwards" on their identification? Perhaps the only way "down" is via word-units, and evidence to be sought would be that which favours a link between the processing of words and the parts they are made up of, i.e. phonemes. Theoretically, such a link is expected, and experimental evidence backs this up (e.g. Morton and Long, 1976). However, what remains unclear is if the same counts for phonetic segments as for phonemes.

Linguistically speaking, the phonetic segment is coded as a "bundle" of distinctive features within a matrix having segments in the columns and features in the rows. Does there exist a simple phonetic percept at an identifiable level in speech processing? Can the listener "hear" a bundle of phonetic features? Or must the phonetic segment be related to its lexical function (i.e. be coded as a phoneme) before it becomes a psychologically real entity, one that can be responded to discriminatively? Dell and Newman (1980) reviewed below, would answer the last question negatively. They provide evidence that phonetic feature bundles form the basis of discriminative responses in their kind of speech monitoring task. However, their "parallel access" hypothesis allows the possibility that listeners might actually "hear" phonemes while also being able to respond discriminatively to their distinctive features. Analogously, people can respond discriminatively to the wave-length of light, but this does not imply that there are psychological entities coding classes of wavelengths, as distinct from colours, that can become part of conscious experience. Wavelength is not experienced except as colour, and perhaps phonetic features are never experienced except as entities that function in a language - that is, phonemes.

Before any of the issues raised thus far can be properly dealt with, some "spade work" must be done in clearing the ground for further discussion. The use of mentalistic terms such as "attention" will have to be justified in the light of a theory of mind-body relations. Concepts such as automatic versus controlled processing must be introduced, and the basis of phonetic perception set out. It will be shown how simple percepts of meaningful entities probably arise from single gnostic units (c.f. Konorski, 1967) which are the end-points of the rapid sequences of neural events we shall call automatic processes. Controlled search for active elements in the sequence is possible, but it is unlikely that the primitive elements themselves can become part of conscious experience. Automatic processing outruns controlled search and elements so located will be experienced as part of the whole, meaningful percept. These are interactions between components of an integrated structure. So long as these interactions are not taken note of, phoneme monitoring research aimed at dissociating the effects of semantic and lexical variables from the operations of a putative phonetic processing stage (as in Dell and Newman) might be inconclusive. We shall deal with some specific shortcomings of the Dell and Newman study. A more general problem is the question of the relation between part and whole, the phonetic segment and its context. How is each processed if the two are not equally accessible to consciousness? It will be shown below that this question has implications for phoneme monitoring research that will be considered in the design of the experiment reported in Section 2.

In a search task that requires identification of part-elements, quite possibly attention will be directed primarily to the highest-level units available (c.f. McNeil and Lindig, 1973) which provide the context be it physical, phonetic, lexical or semantic within which the parts are located. In other words, higher-level contextual entities might structure perception through allocation of attention and, if highly learned, guide search for part stimuli. E.g. Knowledge of the rules of English and of its semantic regularities will structure perception of speech by indicating where certain part-entities can be expected within the whole sentence.

It is suggested here that generally the higher-level "whole" entities develop as a result of rapid automatic processing sequences. Attention will be preferentially directed to certain locations (temporal or spatial) within these sequences where target elements are expected. Thus slower, controlled search for a target phonetic segment would be facilitated by a linguistic structure that develops at very short latency and becomes available to attention. The suggestion is that there is a structural facilitation of target monitoring that operates by directing attention and is distinct from automatic facilitation of pathways. The latter kind of facilitation would explain the contextual effects on word identification in Morton's (1969) model, where logogen thresholds are reduced by input from contextual analysis. This is facilitation of a pathway which passively channels information flow in an automatic word-processing sequence. Structural facilitation operates on consciously controlled processes by directing attention. This is one possible form that the relation between controlled and automatic processing might take in target monitoring of continuous speech.

It is proposed that in speech monitoring research the two kinds of facilitation have been confounded. Processing of part-entities such as phonetic segments might show semantic context effects that are in fact not attributable to facilitation of lexical pathways. Such effects might suggest, if they are mistakenly attributed to lexical pathway-facilitation, that phonetic segments are not identifiable prior to lexical processing (e.g. Morton and Long, 1976). But how are we to know if the one or the other kind of facilitation is involved? Direct access to a phonetic feature-coding of the input might be possible but not easily dissociated from lexical processing while the two kinds of facilitation remain confounded.

What is indicated by this reasoning is that research for evidence of direct access to phonetic feature coding during speech monitoring must dissociate not only phonetic from phonological codes, as Dell and Newman attempted to do, but also isolate the effects due to facilitation of pathways from those due to facilitation of target search.

Following Paap and Newsome (1980), it is suggested that target search performance could be studied by the introduction of a target location cue. This will control the latter kind of facilitation. Contextual effects, if they remain, could then be ascribed with more certainty to facilitation of the pathway to target matching. This pathway would then be seen to lead via lexical access and phonological coding of the input. Absence of context effects in the presence of a target location cue would suggest that the pathway leads directly to a phonetic feature coding stage independently of lexical access.

Finally, it would be possible to gain insight into the relation between the sound and the meaning of speech, especially when experimental research aims not merely at demonstrating the existence of various "top-down" or "bottom-up" effects, but also at disentangling the mechanisms/processes responsible. Part of such a research program must be to identify invisible mental entities or pathways. The Dell and Newman (1980) study cited earlier is a significant contribution along these lines. The experimental research reported in Part Two of this thesis takes the Dell and Newman study as a model and extends it with the aim of dissociating phonetic, phonological and attentional processes in a phoneme monitoring task.

1. GENERAL INTRODUCTION

1.1. Functionalism and the mind-body problem.

The information processing tradition has evolved a methodology and a mode of explanation which Fodor (1981) has called functionalism. Mind is regarded as a symbol-processing structure , and the functionalist abstracts from the physical composition of minds to deal only with the so-called "software". That is, mind is not in principle bound to any particular embodiment. Just as a computer program can exist as holes punched in cards, as magnetic states on a tape or disk, or be stored as an idea in the mind of a person, so too could a mental state have different embodiments. To distinguish between pure mechanism and mind, it is said that mind has both qualitative and intentional content.

(a) The intentionality of mental states is defined by their functional role in the psychological domain. That is, Fodor states, human information processing theories will ultimately have to explain the semanticity of mental representations in terms of their causal inter-relations with each other and with behaviour.

(b) The qualitative content of mental states is hardly explained at all by present theories. We may nevertheless require some statement about the quality that makes certain mental states conscious. Phenomenology may at first sight seem a strange bedfellow for information processing theory, but at least one eminent cognitive psychologist has already attempted a rapprochement between the two (Posner and Rogers, 1978). Given the strong anti-behaviourist trend implicit in theorising about mental states in cognitive psychology, this is perhaps a harbinger of things to come.

Leaving the quality of consciousness as an embarrassing outstanding promissory note, we see that the balance is somewhat redressed by the development of a powerful methodology to study the functional role of mental states in experimental tasks. The use of reaction time (RT) as a dependent variable has become widespread, leading one set of authors to refer to RT as "the measure of a paradigm" (Lachman, Lachman and Butterfield, 1979). By this they mean that agreement as to the nature of RT as a measure and what it is measuring has been reached, lending information processing psychology an air of paradigmatic science. Outstanding examples of the use of RT are found in the work of Chase, Posner and Shepard (all in Estes, 1978); S. Sternberg (e.g. 1976) and R.J. Sternberg (1977) and many others.

The term "mental chronometry" has been coined to characterize the use of this measure; some applications of the methodology are discussed in Section 1.4 below.

The readiness of authors such as Posner and his associates to employ quite freely the terms "attention" and "consciousness" may be contrasted with the usage of the term "attention" by Broadbent (1971). The latter uses the term quite strictly to refer to the modulation of information flow through a processing system. It refers to the changes in "evidence states" that determine the input-output relations of the system. For Posner, on the other hand, "attention" has a variety of meanings. Posner and Boies (1971) and Posner and Snyder (1975) refer to facilitatory and inhibitory functions of attention, while Posner and Klein (1972) and Posner and Rogers (1978) link attention with the operation of a discrete central processing mechanism. This is their way of dealing theoretically with the problem of how elementary processes are co-ordinated and structured into the natural aggregate, or whole mind. An elementary process might be something like "abstraction", defined by Posner and Rogers (1978) as "the recoding of information in a reduced or condensed form" (p.148). It is a transformation of the internal representation of an input, with successive transformations arranged hierarchically in distinct levels.

Part of the problem of co-ordination of elementary processes is the relation between levels of representation of the input - precisely the question that concerns us in this thesis. It is dealt with by Posner in terms of isolable sub-systems (defined in terms of the code or form of representation involved) and the pathways linking them.

Posner and his associates favour a strong reliance within the aggregate on habitual, automatic pathways. Conscious attention is the flexible factor within the system which can alter the routines, but it is limited in its operations:

> "Subjects are heavily influenced by habitual tendencies, even if inappropriate to a new task. Perhaps they can eliminate these tendencies with detailed concentration and practice in the new task. But when additional stresses occur, the habitual processes return."
>
> (p.184)

The interaction between conscious attention-directed processes and habitual pathways constitutes one condition for a description of an aggregate model as required by Simon (1979). This in turn is a condition for understanding the operation of elementary processes. For example, in studying the mental representation of colour, we need to know if the abstract coding of the name of a given colour can be dissociated from the coding of its sensory correlate. Which coding is used by an observer in an experimental task might drastically alter measures on a dependent variable such as RT, with important consequences for the reliability and replicability of the results.

A slightly different approach is found in the work of Chase (e.g. 1978) and of Newell and Simon (1972), who favour a more flexible assemblage of processes on the basis of cognitive strategies which rely on simple operations of a very restricted number. One candidate for a truly elementary operation might be that of comparing two symbols in memory (Chase, 1978). Behaviour in a particular task can be modelled in a computer program in terms of such operations, but only when they are co-ordinated by strategies or "production systems" developed by a control structure.

It should be clear from the above that it is experimentally feasible and perhaps essential to study the mental systems responsible for phenomena relating to consciousness and attention, particularly when attempting to understand the co-ordination of processing levels such as in speech perception. The next section will give details of a theoretical account of the notion of automatic processing that complements nicely the work of Posner introduced above.

In conclusion, functionalism is adopted here as a methodological principle, an aid in operationalising mental states and structures. The mental particulars we shall be dealing with in this research will be defined within a hierarchy of abstraction that has a causal role within a certain task domain. The mental particulars are the levels of representation which code information at progressively higher levels of abstraction from the acoustic input. Each level of abstraction is defined theoretically within linguistics, and can be identified by mental chronometry in the phoneme monitoring task.

1.2 The Shiffrin and Schneider Model

A general model which can account for attention, search and detection in visual tasks and which is flexible enough to allow application to the present research has been proposed by Shiffrin and Schneider (1977). It is rather well tied to experimental findings, also reported by Schneider and Shiffrin (1977) and convergently validated by a wide range of authors (e.g. LaBerge and Samuels, 1974).

The model is summarised in an information-flow diagram which shows the relation between automatic and controlled processes. The structure and its processes, it is claimed, provide a unified account of attention, search, target detection and some aspects of perceptual learning (categorisation).

The following is a summary of the important features of the general framework as presented in Shiffrin and Schneider (1977).

1) Long term store (LTS) is the set of all possible functions of the system (this set presumably includes the functions of the sensory and neuromuscular systems, although the authors are not specific on this point).

2) Any supra-threshold input into the system will activate units in LTS; short term store (STS) is the subset of active LTS units.

3) A sequence of events initiated by such an input is either learned or genetically pre-programmed.

4) LTS is partly organised in levels defined as having a "temporal directionality of processing" (p.155). The levels may be hierarchical.

5) Some sequences of events in LTS occur only in conjunction with a specific input configuration and are so linked as to proceed automatically to completion (i.e. response). These are termed automatic processes.

6) Automatic processes are difficult to establish or to suppress once established. Learning is not always asymptotic.

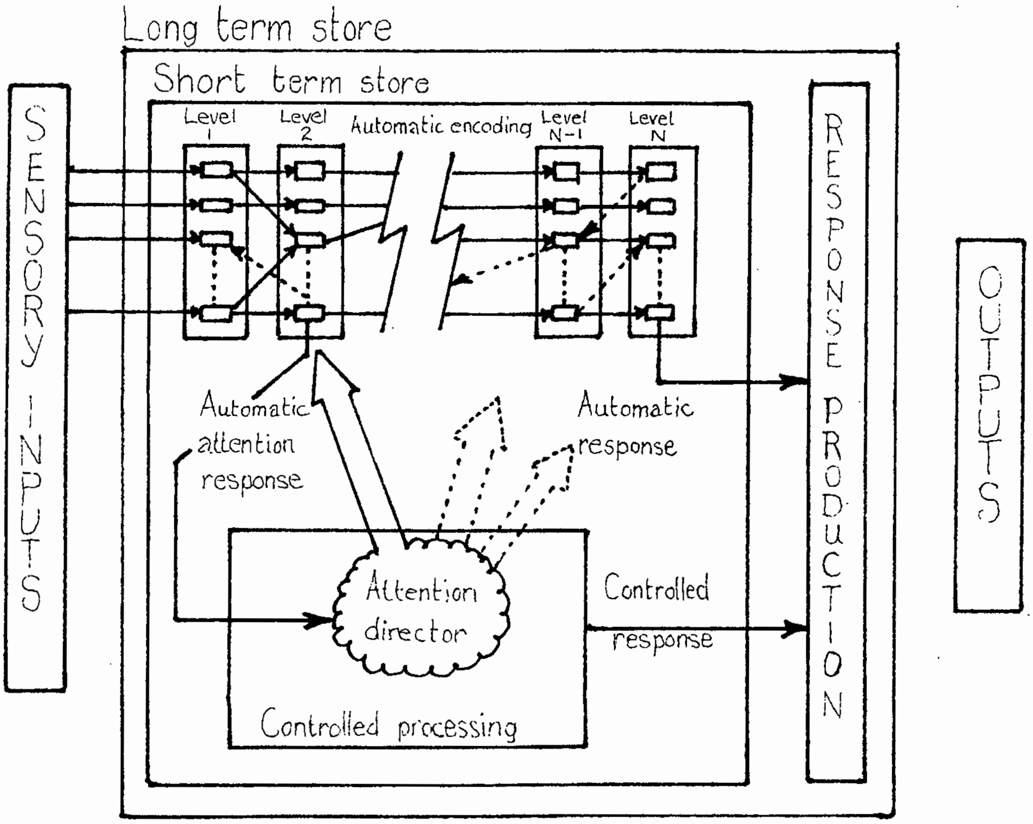


FIGURE 1 : A model for automatic and controlled processing during tasks requiring detection of certain input stimuli. Short-term store is the activated subset of long-term store. N levels of automatic encoding are shown, the activated nodes being depicted within each level. The dashed arrows going from higher to lower levels indicate the possibility that higher level features can sometimes influence the automatic processing of lower level features. The arrow from a node in level two indicates that this node has produced an automatic-attention response, and the large arrow from the attention system to level 2 indicates that the attention system has responded. The arrow from level N to Response Production indicates that this node has called for an automatic overt response which will shortly be executed. The arrow from Controlled Processing to Response Production indicates the normal mode of responding in which the response is based on controlled comparisons and decisions.

(Adapted from Shiffrin and Schneider, 1977)

7) Some LTS units activated into STS remain active longer than others. Some of these may lead to an "attention response". Then the subsets of STS units called controlled processes is activated and directed to the automatic unit concerned.

8) A subset of controlled processes is called the attention director. The function of this system is to activate inactive units in LTS or to maintain or raise the activity of units already active. Sequences of events might be assembled by the attention director, and these are also called controlled processes.

9) Sequences initiated in this manner are temporary and occur slowly relative to automatic processes. Their direction is flexible and maintained by attention. Learning of such sequences is rapidly asymptotic and easily suppressed.

10) The subjective aspect consciousness is associated with controlled processes, but depending on their duration (of activation) they may be more or less accessible to introspection; i.e. rapidly fading sequences of controlled processes are "veiled", others are "accessible".

11) Automatic sequences, while occurring in (through) STS, do not make demands on STS capacity. Consequently they are multiple and simultaneous (parallel). Their duration of activity often preclused access by conscious attention.

!1) Controlled processes are subject to capacity limitations of STS, and consequently serially ordered (take place one at a time).

Experimental testing of the model showed that if the mapping relation from memory to stimulus item was consistent, i.e. if a particular sequence could be consistently and successfully linked to some inputs, then the sequence became automated. If the relation was variable, i.e. if the input was sometimes adequate and sometimes contradictory to the task demands, then attentional control of the sequence became necessary. This difference showed in qualitative (strategy) and quantitative aspects of the resulting performance.

Rapid, almost unlimited capacity processing of input is possible under some conditions, and a small number of conscious, slower operations will occur where flexibility and adaptability are required by the task contingencies. As a model of attention, selectivity of input is denied except under controlled raising or lowering of sensory thresholds. All (or most) input stimuli are processed as far as the presence of automatic sequences allows. Selection takes place after automatic encoding, under controlled attention.

The results concerning attention were summarised as follows: (p.186)

a) Divided-attention deficits arise from limitations on controlled processing. In particular, detection deficits are due to the limited rate of the serial comparison process (resulting from attentional control).

b) Dividing attention is possible when the targets have been consistently mapped during training until automatic detection operates.

c) Focussed-attention deficits arise when the distracting stimuli initiate automatic-attention responses.

d) Focussing attention is possible during controlled processing.

1.3 Evaluation of the Model

The general theory, as it is summarised in the flow-diagram above, does not make explicit how sequences are assembled by the control mechanism according to task demands. The attention director remains a somewhat vague entity, and it is not made very clear exactly how automatic attention responses are effected. More general models have been developed which can be run as computer similations; e.g. Simon (1979) reports that the qualitative differences between automatic and controlled processes have been successfully simulated by at least one program which can also simulate human behaviour in a variety of other tasks: the Shiffrin and Schneider model is not an aggregate model in the full sense used by Simon (1979), but presumably the workings of the attention director are open to experimentation and will be made explicit. Recent physiological work on evoked potentials may successfully complement the kind of methods used by Shiffrin and Schneider (see Pritchard, 1981, for a review). The Posner and Klein experiments (1973) provide another line of evidence for the operation of an isolable control system involved in conscious processes.

It is interesting to note that Posner and Snyder (1975) have developed a theory of attention that echoes the distinction between automatic and consciously controlled processes we have developed above. Their concern is mainly with the facilitation/inhibition dimension of attention, and they have shown that activity of the attention director or central control mechanism can be observed in this way. An asymmetry between the facilitatory and inhibitory effects is the evidence they cite in favour of the automatic versus controlled processing distinction; conscious attention directed to a certain input results in inhibition of co-lateral pathways while automatic facilitation of a pathway is not accompanied by such inhibition. The terminology deliberately borrows from physiology: Posner often seeks physiological evidence to back up his cognitive research (e.g. Posner, 1975). In this respect the work of Shiffrin and Schneider (1977) is more abstract and more comprehensive. These latter authors' model is not deliberately tied to any physiological embodiment. Its strongest theoretical advantage is its generality, in that it encompasses a variety of attentional phenomena in a unified model. At the same time the model is very closely tied to experimental findings in a range of tasks.

In the present context the value of the model is entirely heuristic. The notion of highly overlearned encoding sequences, scanned by a central processing mechanism, provides the conceptual framework for interpreting some of the research to be reviewed below. Since most experimental tasks provide little opportunity for learning to the extent that they can be executed automatically, it can be assumed that "ad hoc" sequences of operations will be involved. Thus performance in the tasks must be interpreted in terms of what is known about controlled memory search, automatic attention responses, divided attention deficits, etc. If, however, the task is such as to incorporate highly automated skills such as those involved in the use of language, then the analysis of performance must take this into account also. It is not surprising, therefore, that the notion of automatic processing is gaining increasing acceptance in the field of speech processing research (e.g. Fischler and Bloom, 1979; Marslen-Wilson and Tyler, 1980; Onifer and Swinney, 1981).

1.4 Subtractive and Additive Methods

The methodology used in the present research is a direct descendent of methods initiated by Donders in 1868 (see Donders, 1969). This subtractive method still forms the basis of present-day research into mental operations. Chase (1978) has pointed out that the logic of the method is sound: assuming that two mental processes are sequential and that each can also operate independently of the other, then the duration of one can be measured by subtraction from the total duration. The implication is that invisible mental processing stages can be isolated.

Donders isolated processes underlying what we now call "sensory discrimination", "simple reaction time" and "response selection", but he had no theory about cognitive functioning in general. If the method requires that a mental process be somehow deleted by changing the task, then it is not certain that the remaining processes are unaffected. There must be a priori reasons for assuming processes are autonomous. Used alone, the method will not be reliable.

Sternberg (1966) revised the foundations of the subtractive method by showing that mental processes can be compared without actual deletion. By varying the number of times a supposedly elementary mental operation occurs in a given sequence, an estimate of its duration is obtained without observing it independently. For example, Sternberg (1966) reported a linear relation between the number of memory items to be scanned (n) and reaction time(RT) in a memory search task. The slope of the regression line of RT on (n) then indicates the increment in RT per unit increase in (n), which is the duration of one memory comparison.

The intercept of the line is the hypothetical duration of the "residual" processes when n = 0.

This method is limited to cases where linear functions can be obtained, and objections have been raised about the assumption of serial exhaustive search that underlies the pattern of results in the 1966 paper. It seems that a more general theory delineating when such a search strategy will be used and when not, would cater for the objections. For example, the Shiffrin and Schneider model presented above could provide such a framework.

Sternberg (1969) proposed a further development of the general idea behind the subtractive method. It is assumed that independent stages will have additive effects on average reaction time measures. Hence a factor which independently varies the duration of one process but not sequentially earlier or subsequent processes, can change the intercept but not the slope of the composite function. This requires that variables be found which selectively change the duration of processing stages and which can be combined or separated within a task to see if their effects are additive or interactive. Additive effects suggest sequential and independent stages of processing. Interactive effects (a change in slope) would suggest that one factor affects two stages at once, and that they are interdependent. This is called the additive factors method for isolating processing stages. An example of the method in use is the study by Dell and Newman (1980) which is presented in later sections of this work.

The subtractive logic has been used extensively by Posner and his associates to isolate mental processing stages. In order to avoid actually deleting or inserting an additional processing stage into a given task, Posner and Klein (1973) varied the time between presentation of two stimuli to be matched, and measured RT to a probe stimulus presented at different points along the continuum. Since different processes were assumed to be occurring at different points along the continuum, the processes were isolated in their effect on the probe monitoring task without being isolated from each other. The effect of on-going processes on probe RT was shown to vary, depending on probe position relative to the other stimuli. It was found that probe RT increased steeply prior to presentation of the second to-be-matched stimulus even when the inter-stimulus interval was two seconds (when encoding of the first stimulus was already finished).

The authors concluded that this effect was due to attention demands of the operations involved in the primary matching task. Moreover, by the subtractive logic, the encoding of a stimulus was shown to have no separate effect on probe RT and hence presumably required no attention; i.e. encoding is automatic, while it is suggested that rehearsal, response priming and execution of the matching process do require attention.

This work was extended by Posner and Snyder (1975). An asymmetry between facilitation and inhibition, associated with a priming stimulus that preceded a visual-matching task, was assumed to indicate the operation of two separate mechanisms in cognition. When the prime stimulus was of low validity to the secondary task, it was presumably not attended to; no cost in performance resulted when the prime mismatched the secondary array; but performance did benefit from a match between prime and secondary array. This happened despite the fact that presumably attention was not involved. When, however, the prime was of high validity (high probability of match with the array)

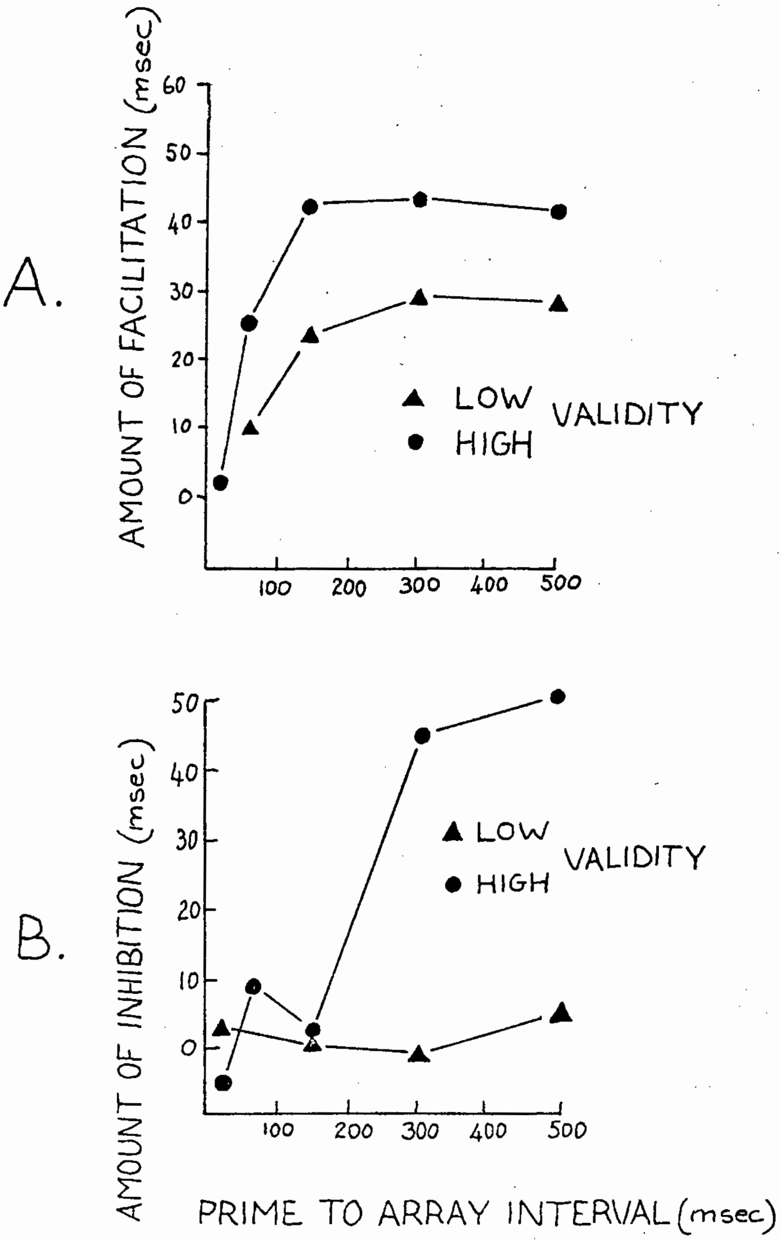


FIGURE 1 - 2 :

Facilitation (A), or inhibition (B) by a prime stimulus that is attended to (high validity) or not attended to (low validity), in a letter matching task. The degree (A) or (B) is calculated from reaction time relative to performance when the prime stimulus is neutral.

(Adapted from : Posner and Rogers, 1978) then inhibition did result from mismatches.

The RT measures in the diagram above were obtained by the subtractive method: a "neutral" prime stimulus provided the baseline measure, and differences between it and those obtained in the "match" and "mismatch" conditions provide the measures of facilitation and inhibition respectively.

1.5 Conclusions

There seems to be ample evidence both theoretical and empirical for the operation of a control process or central processor which is associated with consciousness and attention. It does not matter if the box labelled "control process" in a flow-diagram contains neurons or only symbol-elements. The mental entity has been functionally defined in terms of its causal role in explaining human performance in certain tasks, and in terms of its relations to other mental entities.

Experimental methods exist for isolating the operations of theoretically defined mental processing stages. These make wide usage of reaction time measures. However, there must be a priori reasons for assuming the components of RT reflect the operations of distinct stages. Certain sequences of operations are automatic, rapid and perhaps obligatory, while others are controlled consciously and strategically flexible. The experimental method must take into account what is known about the aggregate functioning of the whole cognitive system. When a task is designed to isolate elements within the system, RT measures may include a component attributable to conscious attention, and which may not be evenly distributed across all mental operations involved.

LANGUAGE COMPREHENSION.

The mental processes/states/structures discussed up to this point have ranged from those involved in problem solving and attention, to the elementary operations such as memory comparisons and sensory encoding. In what follows, a framework will be built up around the central issue of speech perception, so that it will be possible to develop an experimental procedure that dissociates phonetic, phonological and attentional processes in a phoneme monitoring task.

Phonetic and phonological processes are defined linguistically, but may also be distinguished psychologically. Encoding of the auditory input and its transformation into abstract but meaningless elements is the function of phonetic processes. Further transformation and combination of encoded input into meaningful units is mediated by phonological processing.

A semantic interpretation of the input develops as lexical items are "slotted" into a syntactically regular framework. Together with knowledge of pragmatics, these processes, some of which are specific to the auditory mode, comprise the comprehension of language. In what follows, we shall be concerned mainly with the perceptual processes, that is, those not strictly involving communicative content, but that "carry" the content.

1.6 The Mental Particular Defined Theoretically in Terms of Its Linguistic Function

Within linguistic theory functional categories have been developed distinguishing between phonetics, phonology, syntactics and semantics.

These categories are not necessarily relevant to psycholinguistic research. They function within a theory of linguistic competence but should not be uncritically adopted into theories of linguistic performance. However, since it is equally important to have a priori reasons for expecting certain psychological phenomena, we shall adopt here some linguistic definitions.

The phonetic percept should correspond to a representation of the input at a linguistic level of abstraction called a distinctive feature matrix. This specifies phonetic segments in its columns by contrasting distinctive features in its rows. A small number of such distinctive features may be adequate uniquely to describe all the phonetic segments in any given language. The Chomsky-Halle (1968) system numbers only twelve binary features. These have articulatory specifications, which in turn have acoustic correlates which may be realised in the speech signal itself. The phonetic segment is abstract because it cannot be articulated in isolation nor is it represented as a segment in the speech signal. It has no meaning (semantically) and is not specific to any particular language.

The phonological level of abstraction is a transformation of the phonetic feature matrix into language-specific terms. The segments are defined in terms of their linguistic function and are called phonemes. They are abstract in the sense that they are categories of sounds which function in the same way to distinguish between morphemes. The phoneme also has no semantic content, but it is usually defined only in terms of semantic and syntactic structures. Conversely, according to Chomsky and Halle (1968), the listener "hears" the phonological segments and not distinctive feature matrices. The acoustic input itself is actively transformed by way of meaningful structures; the segments which are heard are generated by the listener's expectations.

Thus we say that phonetic segments can be defined purely in terms of articulatory parameters, while phonenemes are the product of a particular grammar. Linguistic theory suggests that the phonetic segment, not yet transformed phonologically, would have no perceptual reality: while it may be structurally real, the acoustic/articulatory segment is only "heard" in terms of its linguistic function, that is, categorically as a phoneme.

1.7 Psychological Reality of the Linguistic Entities

Quite possibly the distinctions which function within linguistic theory also function in linguistic performance and define testable mental states, processes and/or structures. But the categories so defined should not be transposed wholesale into psycholinguistics. Even at an acoustic level, problems with linguistic categories arise. Presently available acoustic analysis has not yet been able to specify unique physical segments corresponding to all phonetic entities. This is not to say that more sophisticated methods than Fourier analysis will not be able to do so in the future. It may even be true that a more exact specification of the input than that provided by the spectrogram could radically change the nature of psychological theories of speech perception (c.f. Neisser, 1976, p.19).

Some researchers have shown that computer simulations can be made to "work" without reference to formal linguistic theory. Wickelgren's (1969) theory of a context-sensitive coding of word segments has been successfully used in a computer simulation of human speech recognition: Marcus (1979) reports a demonstration that his program could learn to recognise novel tokens of synthesised English digit names (one, two, etc.) The "training" input parameters were amplitude, formants and voicing sampled at successive points which were coded as associative chains.

Another example is a report by Lea (1973) that a computer program was devised which could predict 80% of the syntactic boundaries in a range of natural speech samples. The program used an algorithm for detecting changes in fundamental frequency (intonation) which operated without reference to phonetics.

On the other hand, a typical example of research aimed at demonstrating the psychological reality of a linguistic entity is a study by Savin and Bever (1970) which compared RT to target phonemes or syllables. It was found that syllables were identified faster than the initial phonemes of the same syllables. This was taken to indicate that the phoneme was not identified directly, but was derived from the syllable, a finding consistent with the theoretical account of Chomsky and Halle (1968 c.f. above). Savin and Bever concluded that the phoneme is an abstract entity which does have psychological reality although not perceptual or articulatory reality. It functions "by standing neutral between sensory input and articulatory output", interrelating the perceptual and expressive processes (p.301). They argue against Wickelgren (1969, c.f. above) that he mistakenly treats the phoneme as a concrete entity with neurophysiological substrates.

If the phoneme is indeed psychologically real but not perceptually real, does this mean it is not a functionally distinct level of representation in a speech processing hierarchy? In the next section we shall examine some evidence for such a distinct processing stage.

1.8 The Phonetic Percept

Prime evidence for the phonetic percept as distinct from its auditory correlate, is the phenomenon of categorical perception. This has been widely reported (for a review, see Studdert-Kennedy, 1976). The phenomenon is apparently explained by the particular acoustic characteristic of the input and how it is stored. Consonants, which show the categorical effect most strongly, are acoustically low in energy and rapidly transient. Vowels, which are often not perceived categorically, are high in energy and longer in duration. Pisoni (1973, cited in Pisoni, 1978) has shown that discrimination performance depended on delay between token and standard when both were from the same category for vowels but not for consonants. Between-category discriminations were not affected by delays at all. This result indicated that auditory comparisons for vowels were more robust than for consonants, while presumably as a result of their acoustic qualities only, phonetic coding was not subject to rapid decay. We might conclude that certain stimuli, due to their fragile and transitory physical characteristics are rapidly fed forward into an abstract level of representation, thus losing a lot of arbitrary acoustic variability, being encoded as phoneme categories.

Converging on this issue, another area of research has produced evidence for neural substrates to the phonetic processes we have been discussing. Studdert-Kennedy (1976) reviewed a wide range of studies which we can summarise as follows:

(a) The well-known hemispheric specialisation for linguistic processes has been found to apply also to the perceptual processing of speech input. In some cases phonetic processing had a right-ear advantage while auditory processing did not.

(b) Selective adaptation effects on phonetic identification performance have provided evidence for feature-analysing systems which appear to be phonetic and not auditory. By analogy with the feature detectors discovered by Hubel and Wiesel, the feature analysers might be unitary detectors which encode either phonetic segment features, or else phonetic segments as units.

Konorski (1967) has independently developed a physiological theory of unitary perceptions which also takes as point of departure the Hubel and Wiesel discoveries. Konorski elaborates upon the notion of a gnostic unit which codes a single percept, presumably in a single cortical cell. Such units are the highest levels of afferent systems called analysers, each of which has a general function. There are said to be visual, auditory, somesthetic, kinesthetic and emotive analysers in the human brain. The manner in which an afferent system functions resembles strongly the automatic processing of information described earlier (Shiffrin and Schneider, 1977). The gnostic units are partly defined by an "adequate stimulus" which produces the strongest activation, a notion which resembles the concept of "consistent mapping" that defines an automatic sequence in the Shiffrin and Schneider theory.

However, a general theory of how the phonetic percept arises requires more than a physiological description. As Studdert-Kennedy (1976) says - the analyser would still have to be analysed. His own proposal is a sophisticated analysis-by-synthesis theory: phonetic categories are ultimately based on acoustic properties arising from natural constraints on vocal-tract configurations. To perceive a pattern of acoustic events phonetically, is to relate it back to its articulatory source. Thereby the phonetic percept becomes non-arbitrary - a "natural category" (p.282).

The evidence and theory discussed so far certainly favours an abstract linguistic level of processing distinct from the auditory level. Yet itis not certain if a unitary phonetic-segment detector exists. Neural units have been described which could serve the purpose, but they could equally well code words or syllables, or only the distinctive features of phonetic segments. The unitary phonetic percept, insofar as it is psychologically real, might be a construct assembled by conscious attention in an "ad hoc" fashion. It might equally be derived from word or syllable units via phonological rules (see next section). Studdert-Kennedy's theory linking the percept to its articulatory origin allows speculation concerning the link between conscious experience and efference (c.f. Taylor, 1962; Festinger, Ono, Burnham and Bamber, 1967).

Finally, returning to a central point in this thesis, it is once again affirmed that the problem in attempts to identify a unitary phonetic representation is that while it might exist, access to the putative encoding elements might be indirect or a function of other variables. A theory of attention is required to specify under what experimental conditions we can be certain that responses are a direct function of contact with such elements. Furthermore, as will be seen in the next section, interaction between levels in speech processing occurs during normal functioning. The studies reviewed thus far typically use stimuli presented in isolation or in pairs, but not in normal prose context. A stimulus presented in isolation will certainly give rise to an "elementary percept" (c.f. Konorski) if it is processed up to the highest level of encoding that is possible within the automatic sequences available. But if it is part of a greater "whole", then automatic encoding might surpass the elementary level, giving rise to an immediate perception of the higher-level entity. The elements of the whole are then only transiently activated and probably inaccessible to consciousness.

1.9 Interaction of Information Sources and Processing Levels in Perception.

One of the most fascinating aspects of perception is the so-called context effect. The surrounding field in which a stimulus is presented may interfere with or facilitate its perception. In speech, it can be shown that elements of a normally comprehensible signal cannot be recognised when presented in isolation (e.g. Pollack and Pickett, 1964).

We shall not dwell on the various demonstrations of this context effect in psycholinguistic research, since more examples will be presented later. It is important first to review some methodological problems linked to the fact that the context effect also implies that there will be interactions between levels of processing in the mind.

Estes (1975) has shown that an important variable to consider in explaining context effects generally, is the timing of contextual information relative to the primary stimulus. The temporally extended speech input will provide contextual information spread out over time - effects will be due either to prior context or to subsequent information. Given the fact that acoustic/auditory coding of the speech input is transitory and rapidly fed forward into "higher" levels of coding, the timing of contextual information will determine at what level of representation it will interact with the primary input. In visual perception research, this is experimentally manipulated by using tachistoscopic presentations of the stimuli in conjunction with masking methods. The precision of such manipulations has encouraged a vast amount of research.

Two contributions in particular have provided insights that warrant application to speech perception research. Both have some relevance to the automatic vs. controlled processes distinction. The first, by Johnston and McClelland (1980) deals with levels of processing in a strictly hierarchical model of word-recognition. The second, by Paap and Newsome (1980) deals with focussed attention during letter identification.

Letters seen in the context of a word are more easily identified than letters in random strings of arbitrary letters. This is called the word-superiority effect. To demonstrate this, orthographic redundancy must be controlled for and short-term storage effects must be ruled out. Usually sensory processing is also degraded by the use of masks. Reicher (1969) has given the most convincing demonstration of the word-superiority effect. His control condition consisted of a single letter with no context. A single item is unlikely to be lost from STS before it can be reported with greater accuracy than a letter in a word-context. This suggests that the word contains more information than the individual units it is composed of. Such a conclusion would be an embarrassment to a strictly hierarchical theory of information processing: where would the extra information come from, if not from the lower-level units?

Johnson and McClelland (1980) showed how Reicher's (1969) findings could be accounted for within their own model of word identification which consists of a hierarchy of detector units. The explanation rests on a distinction between encoding of input by unitary detectors and usage by the central processor of this encoded information. The probability that an activated detector will result in a response is mediated by the duration of its activity. A response will result only if central attention processes can locate the active unit before it is inhibited by further input or passively decays.

The authors' reply to Reicher is that the level at which processing is disrupted by different kinds of mask, interacts with the level at which the input is coded. In some cases the encoding of a single letter target is disrupted by the mask while the encoding of the word is not. The single letter is not "forgotten" from STS but erased from a lower level store.

The significance of this explanation is that it echoes distinctions made by other theorists between automatic encoding and conscious attending. Although it is not made clear what else besides selective masking could influence access to encoded input, Johnston and McClelland (1980) have set the tone for arguments that will be used later in this report in connection with speech perception.

In an earlier study Johnston and McClelland (1974) had shown that the word-superiority effect did not override cuing. When a target-location cue was provided to letters in words, the word-superiority effect disappeared. By attending to the single letter, people in this experiment apparently no longer processed the whole word and consequently lost the information it provided.

In contrast to this interpretation, Paap and Newsome (1980) describe evidence that the effect is not due to abandonment of a whole-word processing strategy. They distinguish between postcue and predesignated target procedures: the latter kind of experiment involves searching for a small number of well-learned targets and presumably leads to a strong bias favouring these. Partial cues will be sufficient for target detection, but frequent confusions with non-targets will accompany this facilitation. The lexical context for target letters in words provides positional information to eliminate such confusions, but non-words fail to provide this. By manipulating target location cues separately from word and non-word contexts, Paap and Newsome showed that the effect of the location cue is not that the observer abandons a word-processing strategy but that confusions between targets and non-targets in non-words are eliminated.

The implication of this finding would seem to be that even in the presence of a target location cue, lexical access nevertheless takes place. Such an interpretation is suggested by the author's claim that target location cuing did not affect performance on words. This would be plausible in terms of a theory that word processing is obligatory and automatic. Both Konorski (1967) and Shiffrin and Schneider (1977) would predict the same outcome: given an adequate stimulus, the word detector (gnostic unit) will always be activated. This does not mean that a response to the word will result: central processor strategies might select a word, or decompose it to select a letter.

These examples have shown the methodological difficulties involved in demonstrating the inter-relation between letter units and word units, in terms of interactions between mental processes and also of processing strategies. It seems the hierarchical model proposed by Johnston and McClelland (1980) in which the input is first coded as features, than as letter and then as words, might not be incompatible with the context phenomenon; but how could the temporal ordering of these processing stages be tested? I.e. is it really true that within the processing system itself, letters are identified prior to the words they form part of?

By the subtractive logic (Donders) we could test if the time to identify a word is a function of the time it takes to identify its component letters (c.f. Savin and Bever, 1970) but included in the total R.T. will be a component reflecting operation of the central processor. If this is not a constant, then the subtractive method will be unreliable unless the central processor operations are controlled for (this point is made by Johnston and McClelland, 1980, p.519). The Paap and Newsome study has shown how one such aspect of processing can be manipulated or controlled. By providing a location cue, it is certain that controlled processing will be directed to the target, while automatic processes deal with the lexical context. These points will be of great importance in designing the experiment reported later.

1.10 Constructing a Flow-Diagram Model of Speech Processing

To picture a hierarchical structure such as that proposed (Johnston and McClelland above), processing stages are usually modelled in a flow-diagram depicting the stages and their relations in an abstract way.

An early model that gained widespread recognition was Morton's (1964 and 1969) logogen system of word detectors. It was developed primarily to account for experimental findings in word recognition tasks, allowing for precise quantification of the terms rather simplistically depicted in the diagram below. The logogen is the locus of interaction between sensory and semantic/syntactic information sources. It is a passive detector which receives only quantal input and then is triggered if this input exceeds a threshold.

The operation of logogen units has been studied in some detail, and will not be considered here. However, it might be asked to what degree controlled processes partake in the operations of the system. Is it subject to capacity limitations? It has been suggested that the operation of the logogens is "passive" while the context system actively generates predictions (Morton and Long, 1975, p.49). Do these distinctions correspond to the terms "automatic" and "controlled" processing that we have adopted? These questions are important, because if the structure of the model is to be determined by experimentation, then such distinctions will have to be taken into account.

Foss (1969) and Foss and Lynch (1969) claimed that the syntactic, semantic and lexical components of speech comprehension are decision processes utilising short term storage (STS) space. Such a claim is at variance with the depiction of logogens as automatic detectors. It is instructive to see how such opposing claims have been tested.

Foss and his associates used a probe monitoring task to measure processing capacity demands during speech comprehension. The method is an extension of that used by Posner (see above), except that the probe stimulus is not extraneous to the speech comprehension task nor

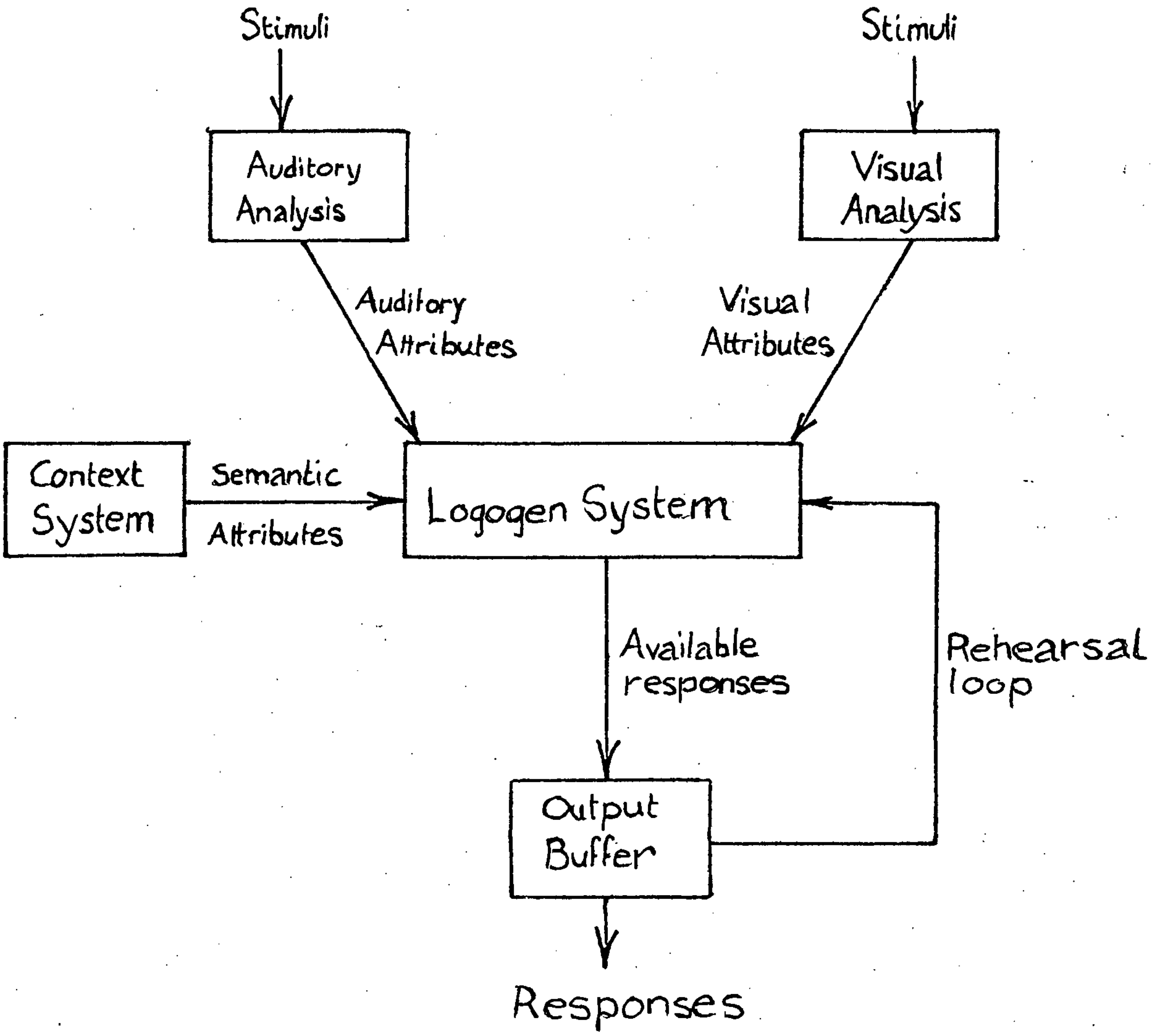


FIGURE 1-3: Flow diagram for the Logogen model.

(Adapted from Morton, 1969)

presented in another modality. The initial phoneme of a word serves as target in the monitoring task and this is secondary to the speech comprehension and recall task. RT measures should reflect varying capacity demands due to syntactic complexity or lexical item difficulty of the speech input.

The results of these studies (Foss, 1969, and Foss and Lynch, 1969) showed that RT to target phonemes increased with complexity of syntactic surface structure surrounding the phoneme. RT was also inversely related to frequency of occurrence of the word preceding the target. These factors were both assumed to affect difficulty of decisions involved in interpretation of the sentence, and it was concluded that phoneme monitoring decisions (target matching), syntactic parsing and lexical access all share a limited-capacity working space.

Interference effects of this kind can be interpreted as divided-attention deficits. The listener must both attend to the speech as meaningful input and identify a target within it. This would imply that speech comprehension is not an automatic process. As was explained earlier, automatic processes pass through STS but do not make demands on STS capacity. The alternative interpretation would be that automatic attention responses resulted to the "difficult" items, in which case interference with phoneme target matching could be expected, (c.f. Shiffrin and Schneider, 1977); but a serious drawback in these studies is that they assume the phoneme can be identified directly from the acoustic input. The probe stimulus in these monitoring experiments has a complex acoustic correlate which is spread over at least a whole syllable and probably requires lexical access before it can be recognised. The work of Savin and Bever (1970, above) has shown that perhaps phonemes are not identified directly. The Foss experiments would imply that the listener is able to dissociate the phonetic from the phonological coding of the input. But it cannot simply be assumed that monitoring responses are a direct reflection of an autonomous phoneme detector, nor that such access is not delayed by internal processing interactions having nothing to do with capacity limitations.

In a later study Foss and Swinney (1973) provided some further insight into this issue. They compared RT to phoneme, syllable and word targets, finding that words were consistently identified fastest. To reconcile their findings with those of Savin and Bever (1970) the authors proposed a distinction between "perception" and "identification". The latter was said to involve consciousness and the former not. This means that the temporal structure of basic perceptual processing is not necessarily revealed by measuring identification latencies. "Identification" would be the outcome of speech monitoring, and specific to the experimental task. "Perception" would be the automatic sequence leading up to speech comprehension. What then causes variations in identificaton latencies? Two further studies have a bearing on this question. McNeil and Lindig (1973) came to essentially the same conclusions as Foss and Swinney, showing that the hierarchy of monitoring latencies is not a direct reflection of the temporal hierarchy of perception. They explained their findings in terms of attention to the linguistically highest-order unit present in the search lists. If search lists comprise sentences, then sentences will have the shortest latencies. Such an interpretation is consistent with our view that speech processing is automatic and obligatory. The input will rapidly be processed as far as it can be, depending on the order of unit presented. Attention will be primarily drawn to the outcome of this processing. That is, the monitoring latencies reflect an attentional hierarchy determined by the experimental materials used.

The second study, by Morton and Long (1976), pointed to another methodological difficulty in studies such as Foss (1969). It was shown that given sentences identical up to the point where the phoneme target appears, monitoring RT varied as a function of the subsequent input, i.e., the target-bearing word. Words highly predictable from the preceding context were associated with faster RT than words not predictable from the same context. A compelling implication of this finding is that the semantic predictability effect had not been adequately controlled in earlier studies, making their results suspect. Morton and Long therefore rejected the "time-sharing" hypothesis, saying that their own model (c.f. Morton, 1969, above) could account for the results without invoking limited-capacity effects. Logogens are primed by semantic (contextual) information and will be activated sooner when this information is relevant than when it is not.

This conclusion is compatible with the automatic speech processing view, but it requires the assumption that the link between semantic processing and phoneme monitoring is via lexical retrieval (the logogen). Contrary to Foss (1969), but without any proof, it is claimed that phonemes are responded to on the basis of a phonological coding (i.e. a derivative of the word) and not their acoustic properties. Delay in word recognition will result in delayed phoneme detection responses.

The studies we have reviewed thus far have not succeeded in testing directly the validity of claims they often make about the temporal structure of processes as depicted in a flow diagram. Morton's logogen model is still adequate to account for the experimental findings.

It is interesting to note that a recurring theme is the question of attention in the phoneme monitoring task. Although this has been discussed in various ways by different authors, none has yet adequately tested a hypothesis bearing on the issue directly.

The answer to the question of the structural arrangement of the processes involved in speech comprehension will have to await a full development of the picture, showing how they are accessed by conscious monitoring. Two recent studies have, however, indicated how the structural components of the flow diagram can be better identified. Both use methods related to Sternberg's subtractive and additive factors technique.

1.11 Evidence against Hierarchical Models

A study which directly tested claims about the temporal structuring of speech comprehension processes is reported by Marslen-Wilson and Tyler (1980). The purpose was to contrast predictions derived from a hierarchical model (similar to Johnston and McClelland, 1980) and from a "multiple interactive model". As was seen above, there is some difficulty in accounting for context effects in a hierarchical system. If the term "hierarchy" means that processing stages are serial, each operating on the output of the one below, then higher-level processes, e.g. semantic analysis, cannot intervene in the operation of earlier stages such as lexical access. Semantic information from prior context can only be matched with the semantic marking of a word after the word has been identified and cannot speed the identification itself, (Morton's (1969) model does allow for such interaction, and is not hierarchical in the present sense).

One of the tests Marslen-Wilson and Tyler devised was to examine the dependency between word-identification latency, sensory input and syntactic and semantic information from prior context. The speech materials were built up from words which formed either random strings, semantically anomalous or normal prose sentences. In a word-monitoring task, sensory input was measured as duration of target words in milliseconds, and it was predicted that RT would depend on word length, since more sensory input is needed to specify long words than short words, However, if semantic and syntactic information interact with sensory input during word identification, then the regression of RT on word length would not be the same for the 3 kinds of materials. According to the additive factors method, the 3 kinds of materials should produce additive effects on RT only if they each affect independent stages of processing.

The results obtained, illustrated in the figure below, show that the regression functions for the three kinds of materials do not have equivalent slopes. There were also significant differences between the proportions of variance in RT accounted for by each of the three functions.

These results may be explained using a model such as Morton's (1969) logogen system, but Marslen-Wilson and Tyler describe their own version of the word-identification process as follows: " ... a distributed processing model in which recognition is mediated by a large array of individual recognition elements, each of which can integrate sensory and contextual information in order to determine whether the word it represents is present in the signal." (p.29). It seems this model differs from Morton's only in that the word-detectors are able

Table 1-1

Monitoring reaction time and word length (msec) : Regression coeficients and slopes for three types of prose context. (Adapted from Marslen-Wilson and Tyler (1980).

	NORMAL PROSE	SYNTACTIC PROSE	RANDOM WORD ORDER
r	+0,57	+0,73	+0,93
slope	0,22	0,27	0,49

to output and receive syntactic and semantic information directly, so that there is no autonomous "context system". Also the word detectors do not work in an all-or-none fashion.

The Marslen-Wilson and Tyler (1980) study has demonstrated a powerful methodology for testing the temporal relations between processing stages and sources of information in speech comprehension. It is outstanding in the use made of precise measurements of word length so that sensory input (bottom-up information) could be assessed in its relation to stored information from previous context (top-down information). Together with other findings (e.g. Marslen-Wilson, 1973) the authors claim support for a theory of speech processing that emphasises multiple simultaneous operations which are obligatory and automatic (p.63). They suggest that multiple word-candidates are activated by the earliest outcome of acoustic/sensory analysis (c.f. Onifer and Swinney, 1981). These together form the initial "cohort", the size of which is progressively reduced as more acoustic/sensory information arrives until a recognition point is reached. This can be uniquely defined for single words, but will vary for words in context (c.f. Grosjean, 1980). Semantic and syntactic processing does not operate on the outcome of word recognition, but is fully contiguous with it and follows the same temporal course.

1.12 Parallel Processing of Phonetic and Phonological Information

The previous example dealt with processing subsequent to primary acoustic and phonetic analysis. It still remains to be seen if the phonetic and phonological codes can be experimentally dissociated in the same manner as Posner and Mitchell (1967, cited in Posner and Rogers, 1978) isolated effects due to the physical and phonetic coding of letters in visual tasks.

Newman and Dell (1978) discovered a phonetic variable which independently affected phoneme monitoring latencies. (Similar effects had been reported by Morton and Long, 1976 and by Foss and Swinney, 1973, but these authors did not experimentally analyse them). It was found that the initial phoneme (the critical phoneme) on the word preceding the target phoneme had an effect on monitoring latencies depending on the number of distinctive features common to both (based on the Chomsky-Halle system). Longer RT was associated with highly similar critical and target phonemes (e.g. /p/ and /b/) than when these were quite distinct (e.g. /s/ and /b/). Dell and Newman (1980) used this effect together with the semantic predictability effect (e.g. Morton and Long, 1976) to assess the additivity of their combined effect on phoneme monitoring RT. They reasoned that they would not find additivity but interaction: in the absence of adequate (predictive) semantic contextual information a phoneme could be identified on the basis of its phonetic coding alone, but this would only happen if its similarity with the critical phoneme was low. Dell and Newman (1980) were looking for an access route to the phonetic code which runs parallel to the derivation from lexical processing. Additive effects would have ruled out parallel processing. As the pattern of results illustrated below suggests, a significant interaction was obtained (but only after the recall task was changed to require a paraphrase of each sentence).

In all cases the target was /b/, while critical phonemes were either /s/ (different) or /p/ (similar). The semantic factor was manipulated as in Morton and Long (1976) by altering the target-bearing word following a given sentence frame. Attention to word meanings was

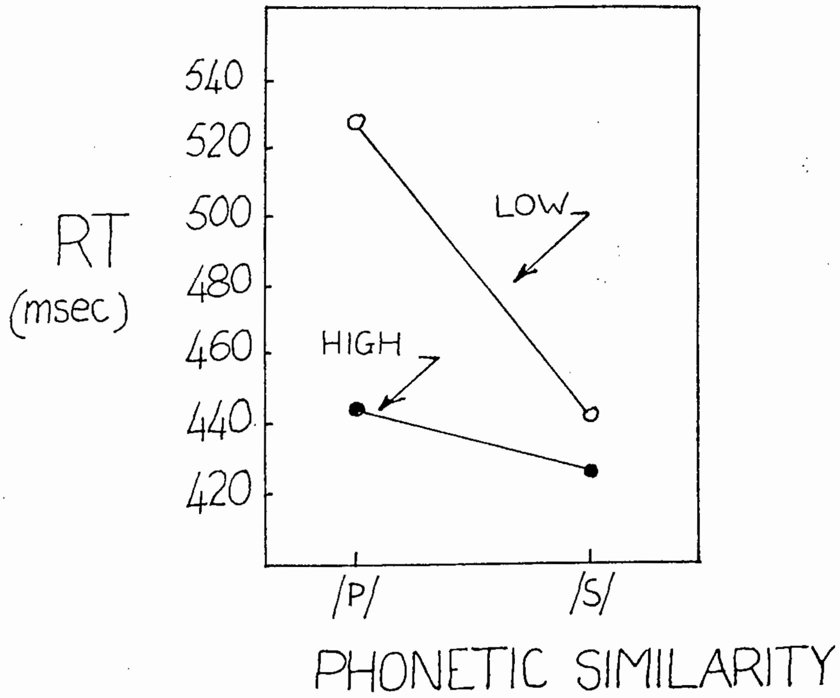


FIGURE 1-4 : Summary of results obtained in a phoneme monitoring experiment. Mean reaction-times are shown for high and low semantic predictability of target-bearing words, and for high and low phonetic similarity between critical and target phonemes, /p/ and /s/ respectively.
(From: Dell and Newman, 1980)

ensured by requiring the listener to paraphrase each sentence after it was heard.

A parallel-access model for phoneme target monitoring was proposed. The figure below shows that a phoneme detection response can be initiated after comparison with either a phonetic or a phonological coding of the input.

It was assumed that the target detection route via feature comparisons is slowed when a highly similar critical phoneme is mistakenly fed into the target-matching system. But the phonological route will not be so affected. Phonetic confusions are impossible after lexical access has taken place, only semantic predictability can speed or slow the phonological route. The detection response will result from whichever route is faster; the phonetic feature route is always faster when there are not "false alarms" to be rejected, while the phonological match is faster when the parallel route is delayed or when predictability is very high.

The general features of this model have been convergently validated by Foss and Blank (1980).

1.13 Critical Evaluation of the Dell and Newman Study

The study we have just described is significant in that it is the first to manipulate phonetic and semantic variables in one task to assess their interaction. As a model of speech comprehension it has not, however, shown whether the processing stages are autonomous and sequential or parallel and inter-related. But it is significant that the model illustrated below was not designed to account for normal speech

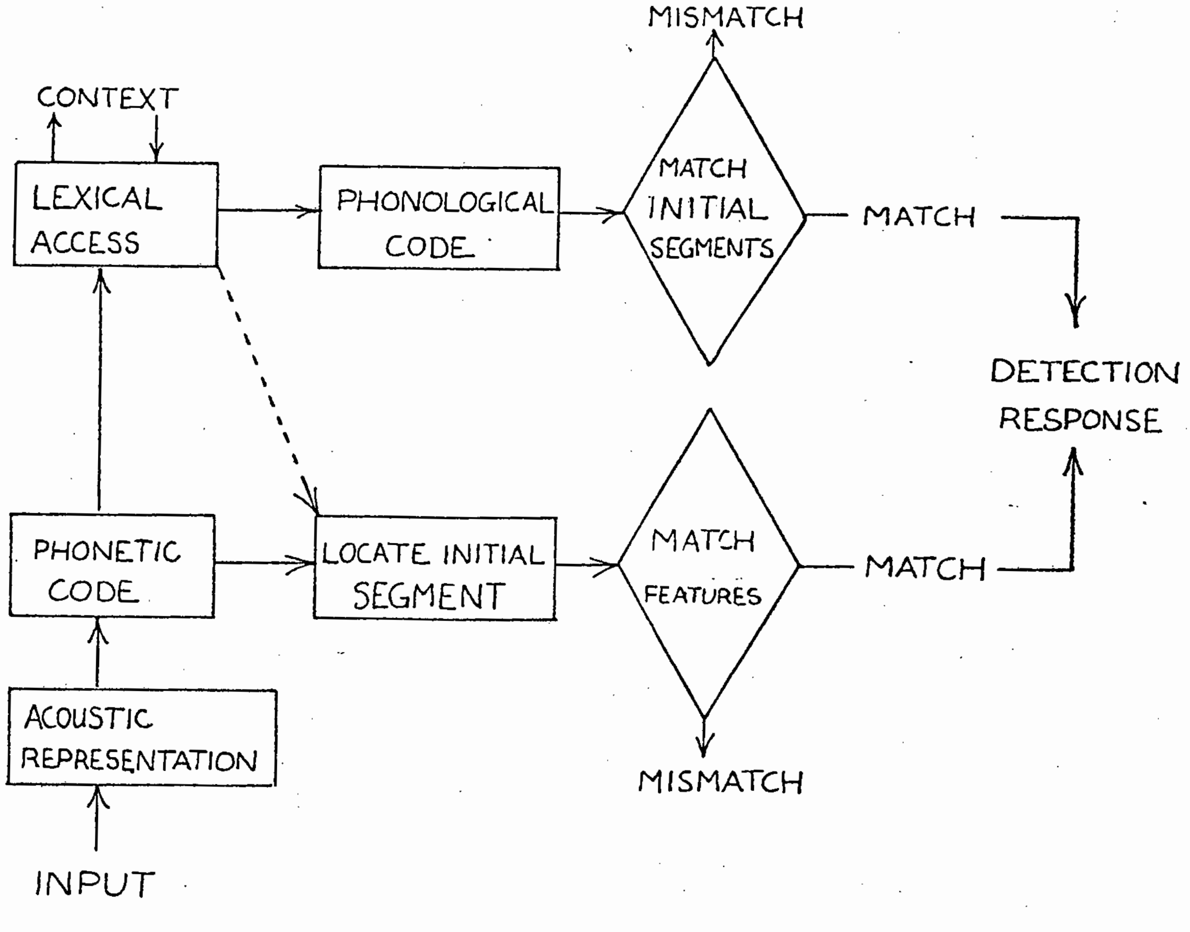


FIGURE 1 - 5 :

The parallel-access model of phoneme monitoring processes. Target detection responses can be based on matching decisions using phonological or phonetic inputs derived from the acoustic representation of the speech input. Phonological segments (phonemes) are marked for their position in words, phonetic segments are unmarked and require prior decisions about their position before matching can be initiated. The broken line has been added between "lexical access" and "locate initial segment" to show this dependency.
(Adapted from : Dell and Newman, 1980)

processing. Instead it represents information flow during target detection, and shows how phoneme monitoring involves access to different stages comprising the normal processing sequence. In other words, the flow-diagram clearly shows the asymmetry between identification and perception, already mentioned by Foss and Swinney (1973). Normal speech perception might involve a sequential ordering between acoustic, phonetic and lexical stages, while target identification could involve parallel sequences. This, then is the kind of explanation we have been seeking, It shows we cannot simply assume that the temporal structure of identification responses mirrors the temporal structure of automatic perceptual sequences.

It must be considered, however, if the phonetic confusion effect could be an artefact. This possibility is suggested by the long duration of phoneme monitoring RT relative to word identification latencies reported elsewhere. The table below illustrates this point.

Table1-2shows that:

(a) In all cases listed, average phoneme monitoring latencies are longer than acoustic duration of the target-bearing word (where this has been measured in milliseconds).

(b) In _all_ cases the average word-monitoring latencies are shorter than the word-targets themselves.

(c) In only one case is an average phoneme-monitoring latency reported (Foss and Blank) which is shorter than the longest word monitoring latency (McNeil and Lindig) in the whole table.

(d) Where direct comparisons have been reported within one experiment, phoneme monitoring latency is 80-100 m.sec.longer than corresponding

TABLE 1-2 Summary of data from selected sources, showing phoneme, word monitoring latencies compared with word-length. Predictability of target words depends on type of context or position in sentence. All measurements in milliseconds.

PHONEMES ONLY

Predictability due to context	Average latency	Average word length	Authors	Materials
Hi	430	?	Dell and Newman (1980)	Sentences
Lo	535	?		
Hi	501	-	Morton and Long (1976)	Sentences
Lo	428	-		
Mean	464	439		
None	546			
	866	?	Foss (1969)	Lists of words
Hi	407	?	Foss and Blank (1980)	Sentences
Lo	453	?	(replication of Morton and Long)	

COMPARATIVE DATA - PHONEMES VS. WORDS

Predictability due to context	Average Latency: Phonemes	Average Latency: Words	Average Length of Words	Authors	Materials
Hi	419	273	369	Marslen-Wilson and Tyler (1980)	Sentence
Lo	492	358	394		
2nd position (Hi)	390	298	?	McNeil and Lindig	Sentence
1st position (Lo)	523	448	?	?	
None	442	340	?	Foss and Swinney (1973)	Lists
None	-	153	413	Grosjean (1981)	Lists
Short	-	245	408		Sentences
Long	-	333	401		Sentences

word monitoring latency.

Obviously, monitoring latency depends on duration of the acoustic input (as was seen above in the Marslen-Wilson and Tyler study). Since Dell and Newman do not report accurate measures of word-length (only given in number of syllables) it is possible that their phoneme monitoring latencies were shorter than the target-bearing words. However, this is unlikely since the reported word-lengths generally fall within the range 350-450 m.sec, while apparently phoneme monitoring latencies fall within the range 390-550 m.sec. The Grosjean (1980) study is especially interesting because it used precise methods for timing the acoustic input duration, and showed that on average words in long sentence contexts could be isolated from other members of the cohort by the time listeners had heard only 37% of the acoustic input. For words alone (without context) still only 83% of the input was required. This was not an "on-line" monitoring task giving RT measures, but it shows exactly how much speech input is needed before an identification response can be initiated. Approximately 100 m.sec more would be needed for simple response execution. In fact, for similar conditions Marslen-Wilson and Tyler (1980) report 273 m.sec. for RT to words while Grosjean (1980) found a mean isolation point for words at 153 m.sec. The difference would be the response component of RT.

These gross comparisons are suggestive: it is plausible to assume that phoneme monitoring responses are initiated only after enough information has been received to allow unique specification of the target-bearing word. If this means that lexical information is used to identify targets, then the phonetic confusion effect reported by Dell and Newman (1980) must be an artefact. Post-lexical coding of the input might allow confusion between allophones, but not between functionally distinctive phonemes, such as /p/ and /b/, as reported by these authors. If, however, the lexical access and phoneme target matching stages are sequential, then semantic and phonetic factors should have additive effects on RT, which was not the case. The alternative is that the phoneme matching stage does operate on phonetic feature information but is slower than word identification because it uses controlled processes or because it suffers divided-attention deficits due to simultaneously on-going speech processes.

Besides empirical evidence for the relatively long duration of phoneme monitoring latencies, there are also theoretical reasons for suspecting that target matching is dependent on lexical access. In most of these studies, the listener must respond only to word-initial phonemes, but as defined in linguistic theory, the phonetic segment is not marked for its position (or function) in the word. How then is the target matching process to exclude non-initial tokens of the phoneme /b/ if it receives only the unmarked phonetic input? (The problem is not solved by constructing sentence materials containing no tokens of the target in non-target positions. In any case, there are two instances of non-initial /b/ in the Dell and Newman materials and seven instances of non-initial tokens in Morton and Long, 1976).

The problem of target location is solved by Dell and Newman through the inclusion of a box labelled "locate initial segment" in the flow diagram (see figure 1-5) A broken line from the lexical access stage to this box has been added to the original diagram to indicate that in fact a word-initial segment in a stream of phonetic segments can only be located on the basis of lexical information. In their own description of this box, Dell and Newman say its function is to locate the initial phonetic segment by checking that the previous segment was word-final. There seems to be some sleight of hand involved here, to hide the phonological origins of this target-matching route!

Given the short latency at which lexical processing is completed, there is good reason to suspect that in fact all target-segment location and identification decisions might follow sequentially on, and make use of phonological information. The result would then be that the two levels of processing have additive effects on reaction time. How could the non-additivity of the Dell and Newman results be explained? This is perhaps simply an empirical question: interactive effects are less compelling as evidence for processing structures than are additive effects. The reason is that the observed interaction could be due to confounding with an uncontrolled third factor which is unevenly distributed across the experimental materials, affecting the two orthogonal factors differently or causing them to be correlated. Sternberg (1969) shows that "pure" additive effects between two factors are much harder to obtain for this reason; but consequently also more interesting. The Dell and Newman effects should thus be regarded with caution, pending further investigation.

Lastly, we argue that Dell and Newman took insufficient precautions to evaluate the degree of semantic predictability in their experimental materials. Both Morton and Long (1976) and Dell and Newman (1980) used sentence completion tasks to assign a transitional probability score to target-bearing words (henceforth called target words). This involves a Cloze procedure where the written sentences are presented to a sample of subjects from the same population sampled for the main experiment. The target words and all following context are deleted and the percentage of correct guesses of the missing target words yields the transitional probability score. The materials in

Dell and Newman had two levels of transitional probability which had to be orthogonally crossed with two levels of phonetic similarity (/p/-initial and /s/-initial critical words). Thus transitional probability for /p/ and for /s/ materials had to be the same, and the authors report 0,462 and 0,459 respectively for "high" transitional probability and 0,008 for both "low" conditions.

The criticism we have of this procedure is that it does not reflect so-called "on-line" predictability of the target words. That is, written or visually presented material follows a different time-course from spoken sentences and offers different cues. Morton and Long (1976) mentioned this difference between what they call "real-time predictability" and their own measurements (p.49). Dell and Newman (1980) also speak of the difference between Cloze predictability and on-line predictability (p.619). If the on-line predictability of their materials is not highly correlated with Cloze predictability, then perhaps their factors are not orthogonal either. The pattern of results they obtained could be explained on this basis alone.

1.14 Rationale for the Present Study

Dell and Newman (1980) posited a parallel-access route to explain their findings. Under some conditions responses are made to a phonological code, and at other times to a phonetic code. They predict performance in terms of the relative speed of these two routes towards completion. Phonetic confusions cause false alarms and delay the phonetic route, while low semantic predictability will delay the other. This explanation was questioned. What further evidence would be required, to show for example that responses can be made to a phonetic code directly? If such evidence were to be found, and we were to assume that in the phoneme monitoring task the listener always responds "phonetically" and never "phonologically", how could we deal with the widely-reported evidence of lexical and semantic effects on monitoring performance? The research undertaken for this thesis is a contribution towards answering these questions.

One suggestion is that divided attention effects could explain variation in monitoring performance: phoneme monitoring is not an automated skill (although it could become so with enough practice) and it may be assumed that target search is a controlled process. While speech processing is highly automated, it is possible that target words having an extremely low predictability demand conscious attention. Perhaps this could interfere with target search, especially when phonetic confusions cause additional demands on attention? Blank, Pisoni and McClasky (1981) examined the effect of performing a monitoring task on speech comprehension. They found there were no differences in comprehension between tasks requiring phoneme monitoring and a control task. Turning this around, does speech comprehension interfere with phoneme monitoring?

Foss (1969) attempted to demonstrate such an effect but failed to provide conclusive evidence because he used a phonetic segment of speech as a probe stimulus. Phoneme monitoring RT was not a pure measure of attention. That is, the probe monitoring task which is supposed to reflect conscious attention devoted to the primary task (speech comprehension), itself required conscious attention.

<u>Monitoring the acoustic source</u>: To avoid these problems, the probe stimulus must be linguistically neutral, and yet exactly co-

incident with the target phoneme. We shall describe a task in which the speech signal is made to switch from one acoustic source to another at a point exactly coinciding with the start of the target phoneme. The phoneme itself remains the probe stimulus, as in Foss (1969), but probe responses are made contingent on where it is heard from and require no target matching decisions. This task is very similar in conception to the "probe task" of Posner and Klein (1973) except that the primary task is to listen to, comprehend and recall spoken sentences, and that reaction time measures are always taken for responses cued by the beginning of a word. Probe RT is measured for acoustic source monitoring instead of phoneme monitoring (Foss, 1969) or an extraneous auditory stimulus (Posner and Klein, 1973). By the logic of the probe task, RT will reflect divided attention effects associated with the primary task.

The acoustic source cue as a target location cue: Our aim is not only to examine possible explanations for variations in monitoring performance, but also to seek evidence that phonological and phonetic codes can indeed be experimentally dissociated. Directly or indirectly, phoneme monitoring RT in Dell and Newman was "contaminated" with phonological information. How can this be avoided?

One way to ensure that target matching will be independent of lexical factors is to introduce a target-location cue in the sense used by Paap and Newsome (1980). If the target phoneme is cued for its location in the speech stream, then lexical information is not needed to distinguish between word-initial and other (non-target) phonemes. The box marked "locate initial segment" in the flow diagram can be dispensed with, and monitoring latencies purged of any confounding phonological components. When manipulation of the predictability of target words affected performance, this was taken as evidence that targets were identified using a phonological code. But it is equally possible that the effects were due to facilitation of target search: the possible location of a target word with initial /b/ may be anticipated using semantic contextual information. This facilitation of search is contrasted with facilitation of a pathway (such as when a logogen threshold is reduced). Provision of an experimentally controlled target location cue should cancel out the former, but not the latter effect. It suggests a way to experimentally separate the two.

Another way to put the above, is to say the location cue will focus the listener's attention on the target, whereby automatic processes are left to deal with the speech comprehension task. Phoneme monitoring will occur under optimally favourable conditions for direct access to the phonetic code. Any effects on RT associated with the semantic factor can then with more confidence be causally linked to it. This should be a powerful test to see if people can dissociate the phonetic from the phonological code of speech input. The acoustic source cue used as "probe" in assessing attention effects, becomes a target location cue when the task is changed to require phoneme monitoring. The initial phoneme of the target word is "marked" by presenting the word over a different channel from the preceding context.

Since the channel switches exactly as the initial (target) phoneme begins, this very switch from one source to another can be the cue on which the listener's target-matching decision can be made contingent. We thus have two different tasks, using the identical materials. In the one task the listener is required to monitor for the switch from one source to another, and RT measures will reflect divided attention effects; in the other task listeners will be required to monitor for phoneme targets coinciding with a particular acoustic source. In that way it will be possible to compare directly probe RT and phoneme monitoring RT. The two tasks will allow independent estimates of the effects due to attentional, perceptual and linguistic factors in the phoneme monitoring task. A better understanding of the processes involved should then follow.

PART TWO

THE EXPERIMENT

2.1 INTRODUCTION

This experiment will introduce a new method whereby listeners' responses in speech monitoring tasks are made contingent on hearing the signal switch between two acoustic sources (loudspeakers). This is called acoustic source-contingent monitoring (ASCM) and may prove useful in the study of speech comprehension within the information-processing paradigm.

The aim of the present experiment is twofold. Firstly, the usefulness of the new method will be guaged in an attempt to test some hypotheses relating to certain well-established experimental findings concerning speech monitoring performance. Secondly, as much information as possible is sought about (a) the kinds of variables, perceptual, linguistic or task-specific, that are of importance in this area of research, and (b) what experimental design and analysis of data are most appropriate in this case. These two subsidiary aims are of importance because the work reported here was the first of its kind in the laboratory that was used, and many methodological problems had to be solved or still remain to be solved. Besides the established procedures for data analysis, it was found that multiple regression analysis of the present data proved to be both informative and appropriate in this kind of research. To show this, two experiments are reported. First the monitoring tasks were set up as a factorial analysis of variance design; next a continuous measure of semantic predictability was obtained and the phoneme monitoring data were re-analysed using multiple regression. The latter method incorporated both continuous and categorical independent variables, which were either experimentally manipulated or derived from ex-post-facto measures.

In accordance with the first aim, it was chosen to extend the Dell and Newman (1980) study of phoneme monitoring, the purpose of which was to show that target matching decisions could sometimes follow directly from access to a pre-lexical phonetic feature-coding of the speech input. They did this by manipulating factors affecting both lexical access and phonetic feature matching, and found that these had non-additive effects on phoneme monitoring reaction-times. Accordingly, it was thought that feature-matching and lexical access could not be sequentially ordered independent processing stages (Sternberg, 1969). They hypothesised that a pathway might run directly from phonetic feature coding to target matching, by-passing lexical access. This pathway would run parallel to "phonological" target matching (the parallel access hypothesis - Foss and Blank, 1980).

We have questioned the validity of the parallel access hypothesis, mainly on the grounds that phoneme monitoring latencies reported in the literature quite substantially exceed the reported word monitoring latencies. Does it make sense to speak of direct access to a phonetic (pre-lexical) code, when access to a lexical code is demonstrably faster? Such an anomaly disappears when the task and the mental structures involved are considered in terms of attention and automatic or controlled processes.

To explore this issue further, we partly replicated the Dell and Newman study, using similar materials and experimental variables. Facilitation of lexical access by semantic context was varied in conjunction with phonetic similarity between critical and target phonemes, but the latter phonemes were always marked by a target location cue in the ASCM task. The major innovation in this procedure was that by making all monitoring responses contingent on hearing the signal from a source spatially displaced from the source of the preceding context, ASCM manipulated the listener's controlled search for target segments in a speech continuum that is acoustically unsegmented. The target is presented spatially separate from, but temporally fully contiguous with its preceding context. An automatic attention response to the source of the signal carrying the target is possible without interfering with the continuity of the signal, nor adding any extraneous stimuli. In this way a target location cue introduced in much the same way as was done by Johnston and McClelland (1974) and by Paap and Newsome (1980) in the field of visual perception.

In this kind of task, target monitoring should be possible independently of phonological rules that specify word-initial segments, and direct access to a phonetic feature code facilitated. With the acoustic continuum clearly segmented by the location cue, the phonological dimension of the target is perhaps irrelevant to the monitoring task: a totally foreign language could be successfully monitored for speech sounds. But in normal speech processing with simultaneous comprehension of meaning, phonology may be an inevitable dimension of the sound segments. Could the listener ignore this as an irrelevent dimension and respond only "phonetically"?

An implication of this reasoning is that perhaps effects on phoneme monitoring performance attributed to semantic predictability of target words are not unequivocal evidence that target matching uses a phonological coding of the input. We distinguish between facilitation of pathways due to "top-down" semantic context effects, and facilitation of search due to target location cues arising from semantic or lexical context. In previous speech monitoring research these two kinds of facilitation have always been confounded. Contextual information will activate certain word-candidates in the mental lexicon, and attention responses (see page 17 above) will occur to those items anticipated as having an initial /b/: controlled target search is facilitated because an expectation to hear /b/ at a certain location is generated by the previous context. Such information is likely to be used in the slower, controlled search for phoneme targets.

Thus automatic activation of the lexical items will follow a different time-course to the search for such activated elements by attention-directed processes (Shiffrin and Schneider, 1977). Even though target identification latencies exceed those for words, this need not imply that target matching is "phonological".

<u>Phonetic confusions</u>: As an explanation for the Dell and Newman findings, the above line of reasoning will not suffice. The phonetic similarity effect needs to be explained as well. This effect was reported by Newman and Dell (1978). They showed that phoneme monitoring could be delayed as a direct function of the number of phonetic features in common between critical and target phonemes. Shorter reaction-times were associated with /s/ critical phoneme preceding the

target than with /p/ critical phoneme because /b/ and /s/ differ by many features while /b/ and /p/ are distinguished only by the presence or absence of voicing. The difference is precisely stated in terms of voice onset time (VOT). The acoustic differences between /b/ and /p/ are illustrated in Figure 2-1.

Difficulty in deciding between similar acoustic tokens of different phoneme-categories may cause processing delays. This was shown by Pisoni and Lazarus (1974) who reported that "different" judgements of highly similar synthetic phoneme pairs (/b/, /p/) were made faster as the acoustic differences between the category pairs increased. Such evidence would support an explanation of the similarity effect in terms of phonetic confusions. That is, although phonemes are perceived categorically, physical differences (or similarities) do play a role in cases where discriminations have to be made. As pointed out by Paap and Newsome (1980), since /b/ is always the target, it is expected that the listener will be biased in favour of hearing /b/ even with only partial cues available. Delays might result from rejecting "false alarms" to tokens of /p/ mistakenly perceived as tokens of /b/.

Divided attention effects: An alternative explanation is offered here that could account for the pattern of results in Dell and Newman. If monitoring is a divided attention task, then it might be that deficits would occur whenever automatic processing of speech is disrupted. The controlled processing of phoneme targets requires attention; but if a target word of very low predictability is encountered and also requires attention, then possibly the former process will suffer a deficit. However, this is not always the case. Dell and Newman showed that RT was longer only when the semantic predictability was low and

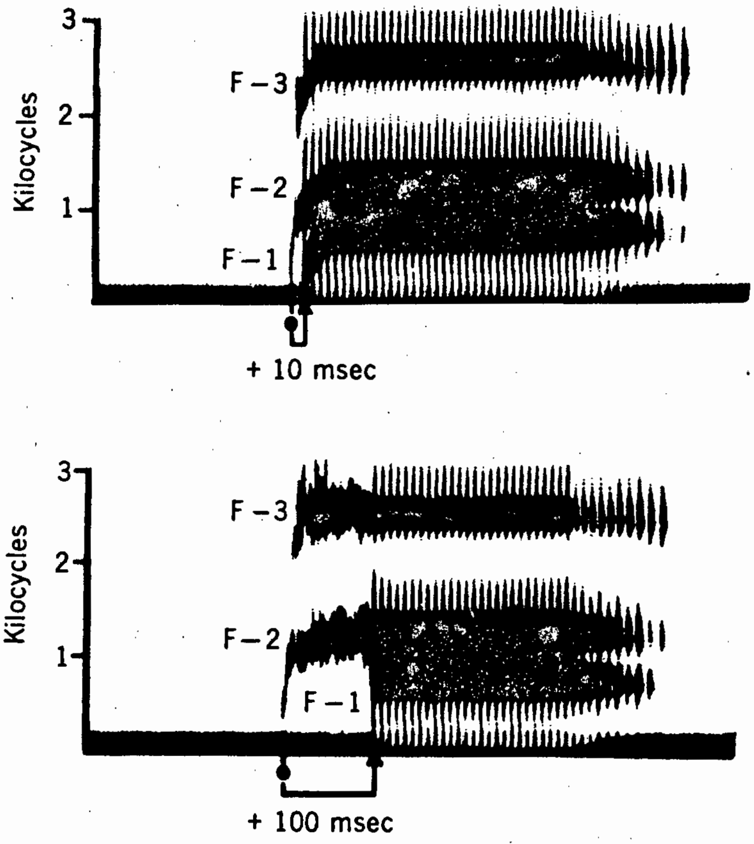


FIGURE 2 - 1 :

Spectrograms of synthetic speech showing two values of voice onset time: The upper figure represents a signal typically heard as [ba] by English listeners and the lower figure represents a signal typically heard as [pa] . The symbols F-1, F-2, and F-3 refer to the three formants characteristic of natural speech.
(From: Eimas and Corbit, 1973)

phonetic similarity was high. We interpret this as the result of attention demands arising from both the target word and the phonetic confusion together. When either one or the other alone requires attention, then no deficit occurs. It might be of assistance to visualize a processing system such as depicted in Figure 2-2.

Without making any claims about reality, the diagram illustrates the point made earlier that target monitoring is a consciously controlled process. Figure 2-1 is really an amalgamation of Figures 1-1 and 1-5. That is, the "target matching" and "locate initial segment" boxes in the Dell and Newman flow-diagram (Figure 1-5) havebeen depicted within the "attention director" component of the Shiffrin and Schneider (1977) model (Figure 1-1). Note that the latter component is subject to divided attention deficits. What this arrangement suggests is that monitoring responses can be initiated at various levels but share a common pathway; latencies will be partly determined by the speed of that pathway, and partly by the speed of the automatic sequences. Divided attention effects should then add a constant component to response latencies, regardless of the level in the automatic sequence where the attention response is directed to. For example, if a monitoring response is executed to an auditory probe (that does not require processing beyond the acoustic level) the model would predict the same divided attention effects as when the monitoring response is contingent on a phoneme-matching decision.

To sum up, phoneme monitoring comprises three components that need to be carefully distinguished:

(a) Consciously controlled target search.
(b) Target matching - comparing input with a stored representation.
(c) Normal speech comprehension.

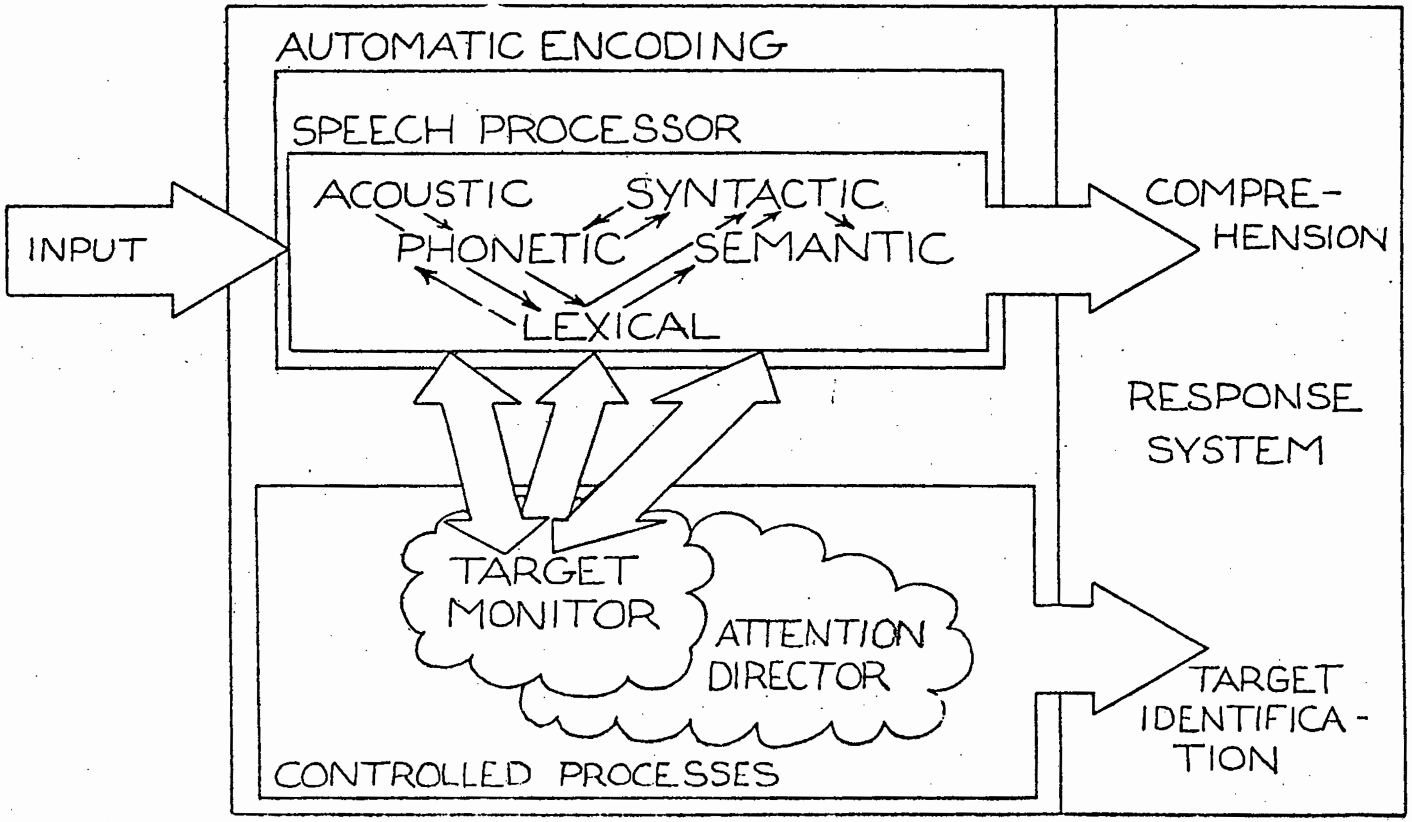


FIGURE 2 - 2 : Flow-diagram of a speech monitoring system, showing the relation between automatic speech processes (with interacting phonetic, lexical, syntactic and semantic elements) and consciously controlled target monitoring. The attention director may activate (prime) elements in the automatic sequences, or these may feed into the monitor, leading to target identification responses. Speech comprehension is simultaneous with the controlled processes.

We propose that monitoring reaction times can vary as a function of all three, while previous research has considered only (b) and (c). The possible causes of variation are:

(1) Divided attention effects between (a) and (c).

(2) "Pathway facilitation" of (c).

(3) Facilitation of (a) by information from (c).

(4) Delays in (b) caused by "false alarms" or phonetic confusions.

The following research strategy has been adopted: firstly, (3) and (4) are controlled by the provision of a target location cue, and a hypothesis is tested that assumes only (1). Secondly, if there is no evidence for (1), hypotheses relating to (2) are tested. If no evidence for (1) or (2) can be found when (3) and (4) are controlled, then it will be assumed that listeners are indeed able to respond purely "phonetically" and to ignore the phonological dimension of the speech targets for the purposes of monitoring.

Testing the hypotheses: In order to produce the location cue that forms the basis of the ASCM method, the recorded speech materials had to be manipulated to cause the signal to switch instantaneously between channels. The figure below illustrates the resulting signal (a). Figure also shows that no "noise" is introduced by the manipulation. A test of this is when the two signals are re-combined and heard over a single channel as in (b). In this case no audible trace of the manipulation remains.

The ACSM task developed for this experiment had two forms. A third variant was not a reaction-time task but was used to obtain a measure of the on-line predictability of the target-bearing words. The three

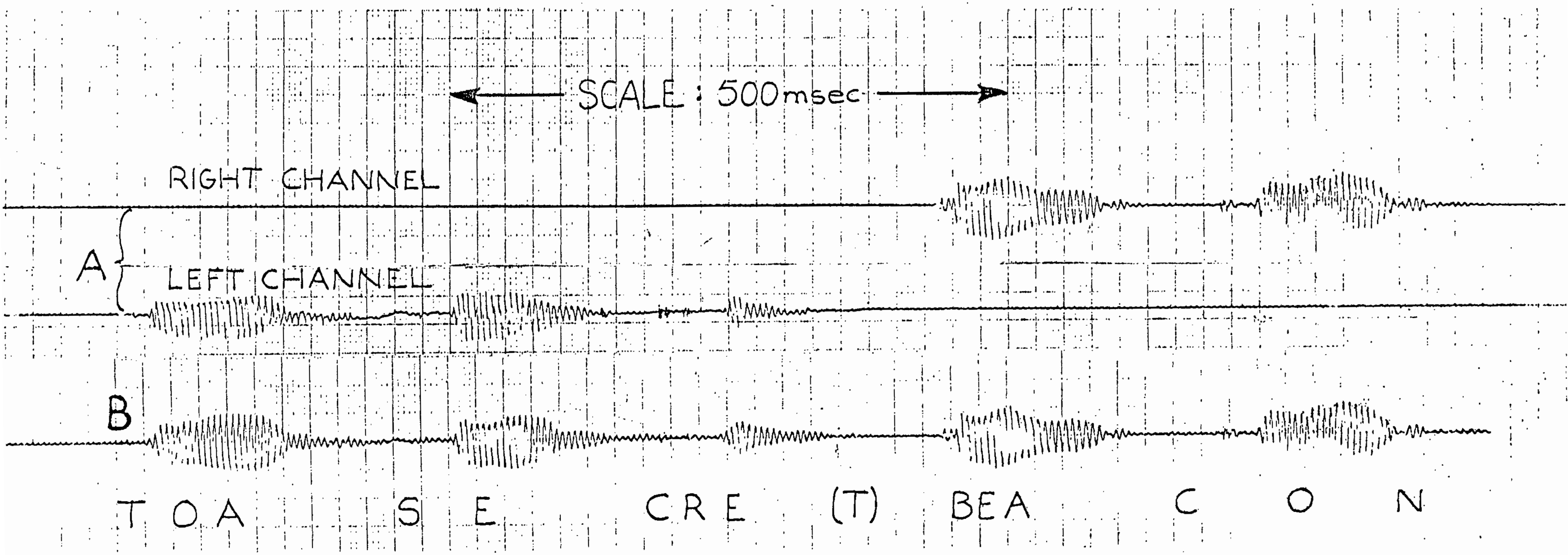


FIGURE 2 - 3 : Waveform-envelope of a signal used in the monitoring tasks. The phrase "to a secret beacon" is shown as the signal appeared with the last word displaced from the left to the right channel as used in the ASCM tasks (A), and with the signal recombined into a single channel (B). The target phoneme was /b/. The signal clearly shows that the final consonant of "secret" was not articulated.

tasks were:

(1) A "control" task in which the listener monitors the signal for a displacement in source while simulatneously attending to its meaning for later recall.
(2) A phoneme monitoring task with targets cued by the same displacement in source and also requiring verbatim recall.
(3) An auditory/verbal sentence completion task in which the target words have been deleted leaving only the preceding input up to precisely the same point where monitoring responses in (1) and (2) were cued to begin.

The same sentences were used in all three tasks. These sentences were adapted from Dell and Newman with some changes. Dell and Newman constructed four sentences from each of 32 sentence frames. For example the frame,

when Henry dropped it the saucer/platter broke/bounced on the floor

yields four sentences corresponding to the four experimental conditions in the design: (Critical phonemes /s/ and /p/ are underlined, as are target phonemes /b/).

(a) When Henry dropped it the saucer broke on the floor.
(b) Wheh Henry dropped it the saucer bounced on the floor.
(c) When Henry dropped it the platter broke on the floor.
(d) When Henry dropped it the platter bounced on the floor.

Semantic predictability is manipulated by inserting different target words (broke, bounced) in the same sentence frame. The critical words (immediately preceding target words) have either /s/ or /p/ initial phonemes to yield the phonetic confusion effect.

The changes made for the present experiment are illustrated in the frame below:

When Henry dropped it the saucer/bowl broke/broached on the floor.

That is, critical phonemes are "identical" or "different" (/b/ or /s/) and target words are "highly predictable" or "semantically inappropriate" (broke, broached), while the initial syllables of both target words are phonetically highly similar to each other instead of sharing only the initial phoneme as in the Dell and Newman materials.

The first hypothesis is formulated so as to allow a test of divided attention effects. It is tested by comparing task (1) and task (2) performance across both the phonetic similarity and the semantic predictability treatments. This will show if it is true that phoneme matching decisons add a constant component to the "control" monitoring latencies. Also, differences in average RT between the two tasks will be of interest; an estimate of the duration of phoneme matching decisions will be possible.

In terms of the second hypothesis, semantic predictability is defined as the probability that the listener will be able to ancitipate correctly the target word, given the sentence up to the end of the critical word. Predictability has two levels as in Dell and Newman, except that "low" predictability here means "semantically inappropriate". This is to maximise the likelihood that a divided attention effect will be observed in the comparison between tasks. "Inappropriate" target words are more likely to demand attention that appropriate but unlikely words. Also, if there is facilitation of monitoring in the presence of predictable target words, then the effect should be quite robust.

At the same time, "high" and "low" target words in this experiment are phonetically more similar to each other than in Dell and Newman. In the example above, the "low" predictability word "broached" is phonetically highly similar to "broke" up to the point where the second consonant is heard. The whole word, or at least more than the first consonant-vowel syllable, must be heard before the semantic anomaly is discovered. It is ensured that the acoustic correlate of the target phonemes in both "high" and "low" conditions is practically the same, and that the subsequent environment of both is maximally similar. In the Dell and Newman materials, as illustrated in the example above, the initial /b/ of "broke" is aspirated while /b/ in "bounced" is not. These differences are phonologically insignificant, but if the listener is responding to a phonetic code, the difference may confound the semantic difference between the words.

The third hypothesis has been formulated as a test that the phonetic similarity effect reported by Dell and Newman will disappear in the presence of target location cues. Half the critical phonemes in the present materials were made the same as the targets (i.e. /b/), and half were unchanged (i.e. /s/) and the same as in Dell and Newman. By replacing the /p/critical phoneme with /b/, the listener would be exposed to two tokens of the target - one of which was marked by the location cue, and the other not. It was expected that the location cue would acquire high relevance as a response cue in this way, since at least half the experimental sentences would have non-target tokens of /b/ (the "identical" critical phonemes). These could only be discriminated from targets by the location cue. Such high relevance would ensure that the listener's attention was directed to the cue, as required by the experimental set-up (c.f. Posner and Snyder, 1975).

Finally, in order that the "on-line" predictability of the target words could be assessed, task (3) was devised. This task was appropriate because (a) the sentence completion norms reported by Dell and Newman are applicable only to the population from which their subjects were drawn (i.e. Canadian college students) and could not be assumed to apply in South Africa, and (b) because Dell and Newman used a written (Cloze) presentation of materials for their assessment of completion norms, while the monitoring task gives an auditory presentation. On-line predictability, operating while the listener comprehends the speech, may not be the same as Cloze predictability. The completion scores obtained from this task were used to verify the effectiveness of the semantic predictability manipulation, and were later used in a multiple regression analysis of the phoneme monitoring data. In the same way, evidence for two further hypotheses will be examined. Ex-post-facto measures of the temporal duration of target words and of their frequency of occurrence in the English language (Kučera and Francis, 1967) will be tested for their relation to phoneme monitoring latencies. It is known that a relation exists between the duration of an acoustic (speech) input and the time it takes to identify it. Both Marslen-Wilson and Tyler (1980) and Grosjean (1980) established such a link in the case of word-identification. Nooteboom and Truin (1980) used a technique similar to that of Grosjean to demonstrate that in a word-identification task, non-native listeners needed to hear longer fragments of spoken words than native listeners. This evidence suggests a link between phonology and acoustics, that was already hinted at in Morton and Long's (1976) finding that monitoring latencies for phonemes varied with the length of the target-bearing words. If for the present materials, phoneme monitoring latencies show such effects, this will suggest that the listener is not able to "switch off" the phonological code to respond purely "phonetically". Similarly, the frequency of occurrence of a lexical item is known to affect its perceptual processing (e.g. Foss and Blank, 1980). Assuming that target-word processing is affected by frequency (of the target word), and if phoneme monitoring performance is found to vary with frequency also, then we may conclude that such performance is not independent of lexical factors.

2.1.4 HYPOTHESES

H-1 : Phoneme-matching decisions add a constant component to acoustic cue-monitoring latencies across all levels of phonetic similarity and semantic predictability treatments.

H-2 : There is an inverse relation between phoneme monitoring reaction time and the semantic predictability of target-bearing words. High predictability will be associated with shorter reaction times than low predictability.

H-3 : Phoneme monitoring reaction time is unaffected by phonetic similarity between critical and target phonemes.

H-4 : There is a positive relation between phoneme monitoring reaction time and length of target-bearing words. Long acoustic duration will be associated with longer reaction time than short duration.

H-5 : There is an inverse relation between phoneme monitoring reaction time and the frequency of occurrence of target-bearing words. High frequency words will be associated with shorter reaction time than low frequency words.

2.2. METHOD

2.2.1. DESIGN

H1,H2 and H3 define 3 experimental treatments, i.e. Task, Semantic Predictability and Phonetic Similarity, each having 2 levels. In the basic design there were three treatments crossed in a factorial design as illustrated diagramatically in Figure 2-4 below.

SENTENCE FRAMES.

The basic design in Figure 2-4 above was obtained by collapsing across a fourth dimension, namely Sentence Frames (D). Within each sentence frame , treatment B consisted of two target words (always with initial /b/), and treatment C consisted of two critical words (with either /b/ or /s/ initial phonemes). These constituted fixed factors, as the four combinations of the set of target and critical words exhausted the possibilities within a particular frame. There were 16 sentence frames all together, which constituted a quasi-random sample of all possible sentence frames. Thus D was a random variable (Clark, 1972), or a non-generalising variable (Coleman, 1979). The complete design illustrated below shows how the two random factors, Sentence Frames and Subjects , were crossed with each other and with treatments B and C.

To avoid carry-over effects, it was impossible to present each sentence frame to a given subject in all of its BC treatment combinations. Thus only 1/4 of the cells are actually filled, with each subject hearing only one BC combination for each level of D. The result is a fractional factorial design. For the purposes of analysis this can be treated as complete design after collapsing one of the dimensions:

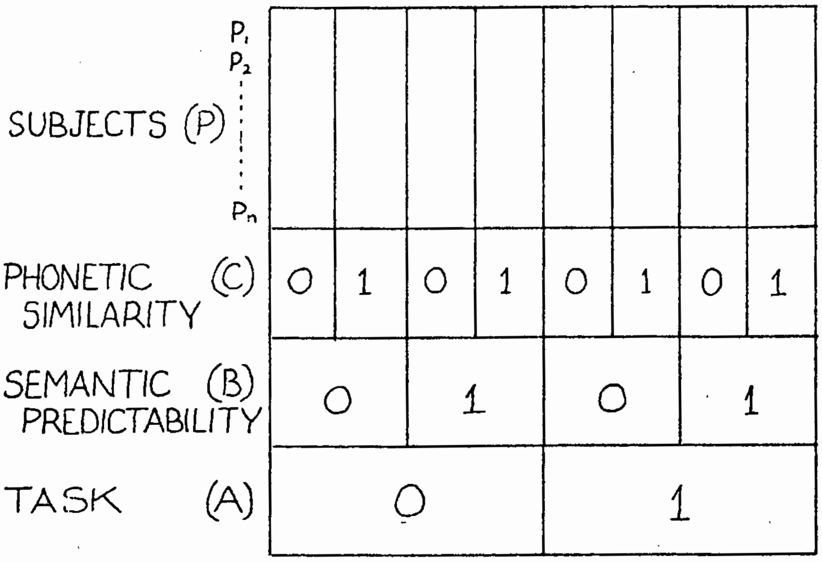


FIGURE 2 - 4 :

The basic design for analysis of variance, showing subjects crossed with treatments.

SUBJECTS	P0	▨						▨			▨						▨
	P1		▨						▨	▨						▨	
	P2			▨		▨							▨		▨		
	P3				▨		▨					▨		▨			
SENTENCE FRAMES		D0	D1	D2	D3	D0	D1	D2	D3	D0	D1	D2	D3	D0	D1	D2	D3
PHONETIC SIMILARITY		C0				C1				C0				C1			
SEMANTIC PREDICTABILITY		B0								B1							

FIGURE 2 - 5 :

The relation of the design in Figure 2 - 4 to a complete factorial design where subjects are crossed with sentence frames and treatments. Shaded cells are those containing observations, and unshaded cells are treatment combinations not observed in the present experiment. The figure shows only one level of A (Task) because both levels A1 and A2 had the same structure. A was crossed with Sentence Frames but nested within Subjects.

First the data are averaged across sentence frames, and then across subjects (c.f.Kirk,1968,pg.381). Two F-ratios are obtained, F1 and F2 respectively, which are combined to obtain the quasi F-ratio,

$$\min F' = F1\ F2 / (F1 + F2) \quad (\text{Clark},1973)$$

When either min F' of both F1 and F2 together are significant, then the results can be generalised to all possible subjects and to all possible sentence frames. Min F' tends to be a conservative test when subject-by-treatment variance is low, but the conjunction F1 F2 my produce high type 1 error rates (Forster and Dickinson, 1976).

2.2.2. MATERIALS AND APPARATUS.

Sixteen sentence frames were selected from 32 frames used by Dell and Newman (1980) (see appendix). The following procedure was used in obtaining this selection:

In order to obtain the two levels of B, defined as "high" and "low" predictability (semantically inappropriate) respectively, certain sentence frames were selected from the Dell and Newman materials . Only frames were chosen for which,

(a) semantically inappropriate target word could be found to replace the "low" predictability words used by Dell and Newman, and

(b) that matched as closely as possible the initial consonant-vowel syllable of the "high" predictability counterpart. Thus for example, the Dell and Newman pair "blushed / blinked" became "blushed / blundered" in the present materials. For three sentence frames no critical word with initial /b/ could be found, and these were left unchanged from Dell and Newman .

The two levels of treatment C, respectively phonetically "identical" and "different", were obtained by selecting from the Dell and Newman materials a list of those sentence frames for which a critical word could be found with initial /b/ instead of initial /p/ . For example, Dell and Newman "pack / sack" became "bag / sack" in the present materials.

Further constraints were imposed on the selection of materials: After a list had been selected that contained the desired combinations of B and C treatments in each sentence frame, the frequency of occurrence, and the length (in syllables) of the target and critical words was checked to ensure that the two treatments would be orthogonal. Frequency of occurrence was defined as the summed frequency of uninflected plus regularly inflected forms of each word (c.f. Cutler, 1981), as appearing in the Kučera and Francis (1967) sample. For example, "beacon" had a frequency of 5, whereas "burn" had a frequency defined as 15 (burn) + 113 (burning,burned, burner,burns) = 128.

For target words, after some adjustments had been made to the list, average frequency (and standard deviations) for the "highly" predictable and the "inappropriate" groups respectively were 45,8 (40,97) and 50,9 (43,56). Word-lengths in syllables provided only a rough guide for preliminary selection. Full details of accurate measurements of the physical duration of words are given in the results section of this report.

As in the Dell and Newman materials, critical and target words always constituted adjective-noun or noun-verb pairs. Additionally, all target phonemes occurred in stressed syllables.

Thirteen of the target words were positioned in the middle, and three at the end of sentences .

PREPARING TAPES FROM SENTENCE LISTS.

32 Sentences were obtained from the four combinations of critical and target words in each of the 16 frames. Four lists were composed , each containing all 16 frames, but each frame appeared in only one of its treatment combinations. The order of the sentences in two lists was the same, but this was reversed in the other two lists. In this way order effects were counterbalanced. Since each sentence contained a particular BC combination, the lists were composed in such a way that (a) a particular treatment level (e.g. /S/ critical phoneme) never occurred more than three times in succession and, (b) so that a particular treatment combination was never repeated more than once. In addition to the 32 experimental sentences, 17 "filler" sentences and 10 "practice" sentences were composed. The latter contained 6 targets and 4 non-targets, in various word-positions. Thirteen of the "filler" sentences constituted "catch" trials in that they contained phonemes other than /b/ in the target position (i.e. co-inciding with the location cue ; see appendix for full details). Four "filler sentences had /b/ targets occurring in the second or third word-position in order to disguise the fact that real targets never occurred near the start of the sentence. RT to these "filler" targets was not recorded.

The "filler" sentences were interspersed between the experimental sentences in each of the four lists. The order of the fillers was the same for all four lists. The arrangement of filler and experimental sentences was such that never more than two experimental sentences and never more than three filler sentences followed each other. (See appendix for an example of the compostion of one such list).

Thus each list contained 33 sentences in all, of which 20 contained targets, leaving 13 "catch" sentences with no targets. Treatment combinations were ordered quasi-randomly within each list.

Tape recordings were made of all the lists composed in this manner, including the 10 practice sentences. Each list constituted one tape. The speaker for these recordings was a female teacher of English who had no knowledge of the purpose of the experiment. She was instructed to articulate the sentences "as if speaking to a class of second-language pupils" with normal intonation. The rate of speaking was approximately 140 words per minute. Recordings were made on an Uher "Report 2000" single-track tape recorder.

PREPARING THE TARGET LOCATION CUE.

The single-track recording was transcribed onto two tracks with a Sony TC 399 stereo recorder, to produce a master tape. Silent switching of the signal from one channel to another was accomplished as follows:

Firstly, the start of each target /b/ was located on the tapes. This was done by moving the tape manually backwards and forwards over the play-back head until the plosive release of energy corresponding to the release of the speaker's lips had been found. The location was marked on the back of the tape. (This method has been reported for locating timing-pulses in a wide range of speech monitoring tasks in the literature). Next, the oxide coating on one track of the recording was scraped off from the point marked on the back of the tape in one direction, and then the same was done to the other track but in the opposite direction. The result is shown in Figure 2-6 below.

Following this physical manipulation, the remaining signal on each track was erased electrically in the appropriate direction (i.e. backwards on one track and forwards on the other). The resulting recording is heard to switch silently and instantaneously from one auditory channel to the other (See Figure 2-3). These recordings were transcribed on onto four experimental tapes. Each tape was made up of ten practice sentences, then 33 experimental sentences.

The tapes were edited so that approximately 5 seconds separated the end of one sentence from the beginning of the next; a warning "bleep" of 1000 Hz was placed at one second before the start of each sentence. Each tape took approximately six minutes to hear. Each subject heard only one tape.

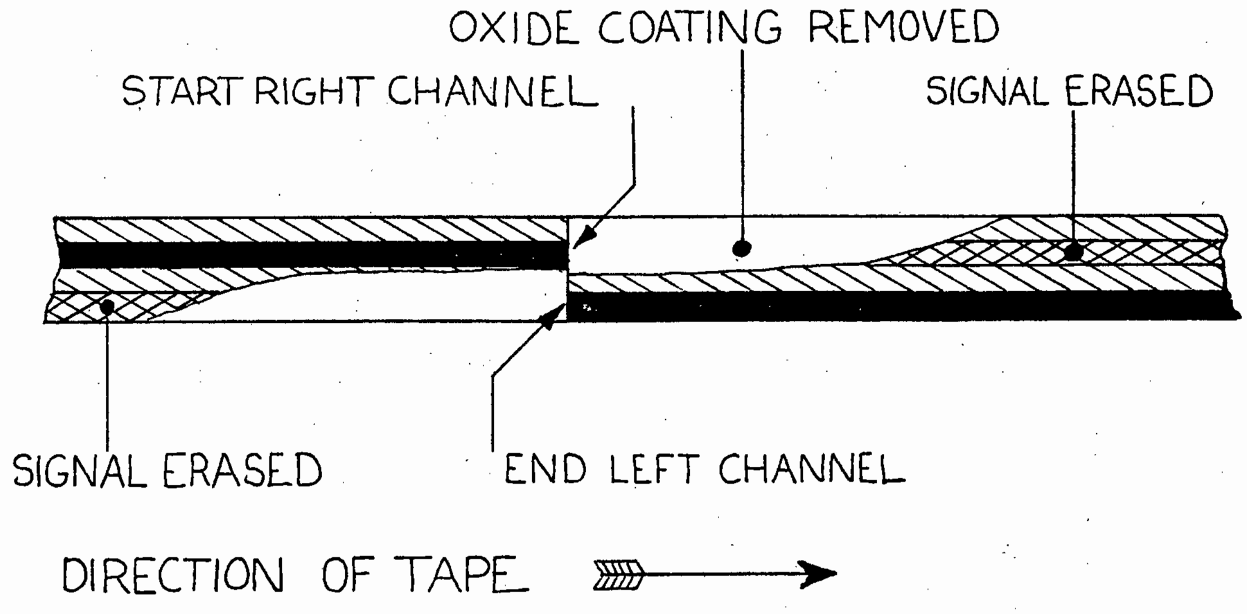


Figure 2-6:

Illustration showing how the two channels of recorded speech (in black) are located on a 4- track tape. The oxide coating is shown removed to obtain an instantaneous and noise-free switch from left to right channels.

PRESENTING THE SPEECH RECORDINGS.

The recordings were presented to the subjects over two loudspeakers of 2 watts each, mounted on a board 60 cm horizontally apart.

TIMING THE RESPONSES.

Target identification responses were made on a standard telegraph key. The key was wired to a relay switch that stopped a Venn Instruments millisecond timing device with digital readout. The timer was started by a voice-activated switch wired to the channel containing the target word. That is, timing was automatically initiated by the start of the accoustic correlate of the target phomene, and was stopped by the subject's response. In addition to the automatic timing, a backup system was used, a second tape recorder running at 19 cm per sec. recorded the righthand channel only,as heard by the subject,(see Figure 2-7) but this recording was automatically interrupted when the listener raised the response-key. The recording on this tape thus consisted of a series of "noise bursts" corresponding exactly in length to the duration of response latency. Played back at 1/4 speed and transcribed onto paper using a Dynograph recorder running at 50 mm per sec., the resulting recording had a scale of 1mm = 5 msec. The length of the signal was measured with a ruler to determine reaction time scores.

2.2.3. SUBJECTS:

24 Male and female volunteers participated in this experiment. Twelve were randomly assigned to each of Task 1 and Task 2.

The voluteers were obtained from the university campus after being approached personally by the experimenter. They were asked if they wished to participate in an experiment on speech perception. All spoke English fluently. Volunteers were not paid for their participation

2.2.4. PROCEDURE:

After obtaining their agreement, volunteers were escorted to the experimental room by the experimenter and instructed about the task (see below). The room was in the basement of a large building and hence well-insulated against noise.

The subjects were seated at a desk facing the two loudspeakers . A microphone was placed on the desk, and an intercom set was placed to one side (for two-way communication with the exprimenter). The experimenter then left the room to operate the equipment in the adjacent room. All except loud noises were inaudible through the separating wall.

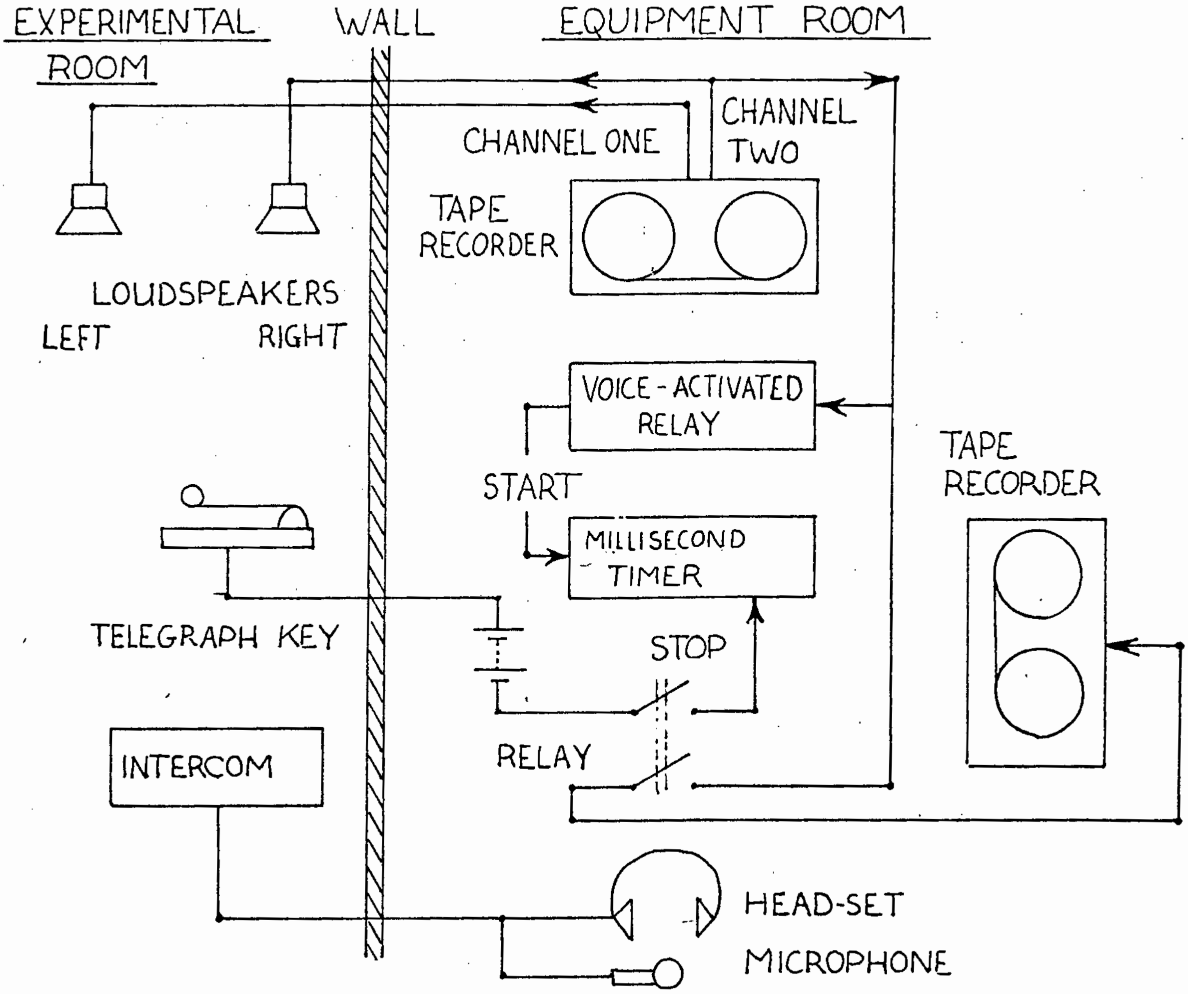


FIGURE 2 - 7 :

A simplified diagram showing the arrangement of experimental equipment.

The ten practice sentences were then presented, during which time it was ensured that the subjects had understood the instructions correctly. If they had not, further clarification was given by intercom before the experimental trials began. At the end of each sentence the tape was stopped momentarily while the subject gave verbatim recall. Accuracy of recall was checked on a score sheet and the reaction time for that sentence recorded from the digital readout. The back-up recorder was switched on and off during these pauses. Most subjects gave their recall response within the four seconds between end of one sentence and the warning bleep for the next.

The experimenter gave feedback on the accuracy of recall by saying "Uh-huh" or "correct" over the intercom before the start of the next sentence.

At the end of the experiment each subject was debriefed. The experimenter asked if he/she had any comments to make, and if they had noticed any pattern or anything unusual in the speech

materials. Finally, the purpose of the experiment was explained.

INSTRUCTIONS TO SUBJECTS:

Every subject was instructed informally so as to set him/her at ease. The instructions had a standard format followed in all cases, but because subjects asked questions, the instructions were not repeated verbatim in every case.

After obtaining their agreement, volunteers were told they would participate in a speech perception task that required devided attention. They had to listen to the recorded speech and repeat what they had heard immediately after each sentence. They were warned to listen carefully as the experimenter would check the accuracy of their recall. At the same time, they had to listen for a key-pressing response cue. Task 1 subjects were told they had to listen for the first sound to appear over the right-hand speaker, and to respond only if this was a target (given as "bee or buh"). Task 2 subjects were told they had to respond every time they heard the speech appearing over the right-hand speaker.

All subjects were told to keep the finger of their preferred hand lightly depressing the telegraph key while they listened to the speech. They were told to respond by releasing the key as quickly as possible when the appropriate cue was heard. They had to keep a balance between speed and accuracy. At the end of each sentence they had to say aloud the sentence exactly as they heard it.

Beyond ensuring that each person understood the instructions, no discussion was entered into.

2.2.5. RESULTS: EXPERIMENT ONE.

The data obtained from tasks one and two (phoneme monitoring and acoustic cue monitoring respectively) were analysed in two ways. First the average score for each subject was calculated over the four sentences heard in each BC combination within each particular task. These data were analysed as a 3-way factorial design with subjects as a random factor and repeated measures on treatments B and C. A summary of this analysis of variance is found in Table 2-1. Secondly, the average score over 3 subjects for each sentence was calculatd within each ABC combination. These data were analysed as a 3-way factorial design with sentence frames as a random factor and repeatd measures on treatments A,B and C. A summary of this second analysis of variance is found in Table 2-2.

Any score exceeding one second was deleted from the record. Deleted scores and omissions (when no response was made to a real target) were replaced by an estimate based on the subject and sentence means obtained from the incomplete data matrix.

Generally, F1 refers to F-ratios calculated for Subjects, and F2 refers to F-ratios calculated for Sentence Frames.

TABLE 2.1

ANOVA SUMMARY TABLE: 3-way analysis comparing treatments across tasks, with subjects as a random factor and repeated measures on B and C.

SOURCE	SS	DF	MS	F
BETWEEN SUBJ.				
A (task)	476862,38	1	476862,38	15,124***
SUBJ.W.G.	693654,0	22	31529,727	
WITHIN SUBJ				
B (phonetic similarity)	4959,656	1	4959,656	1,790
AB	108,186	1	108,186	0,039
B X SWG	60944,25	22	2770,193	
C (semantic)	187,406	1	187,406	0,096
AC	1488,188	1	1488,188	0,761
C X SWG	43039,625	22	1956,347	
BC	7384,781	1	7384,781	2,786
ABC	1855,125	1	1855,125	0,699
BC X SWG	58325,125	22	2651,142	

CHECK ON HOMOGENEITY OF ERROR TERMS:

F MAX (SUBJ. W. GROUPS) = 2,961 df = 2, 11

F MAX (B X SWG) = 1,522 df = 2, 11

F MAX (C X SWG) = 1,322 df = 2, 11

F MAX (BC X SWG) = 3,519 df = 2, 11

*** p 0,001

TABLE 2.2

ANOVA SUMMARY TABLE: 3-way analysis comparing treatments across tasks with sentences as a random factor and repeated measures on A, B and C.

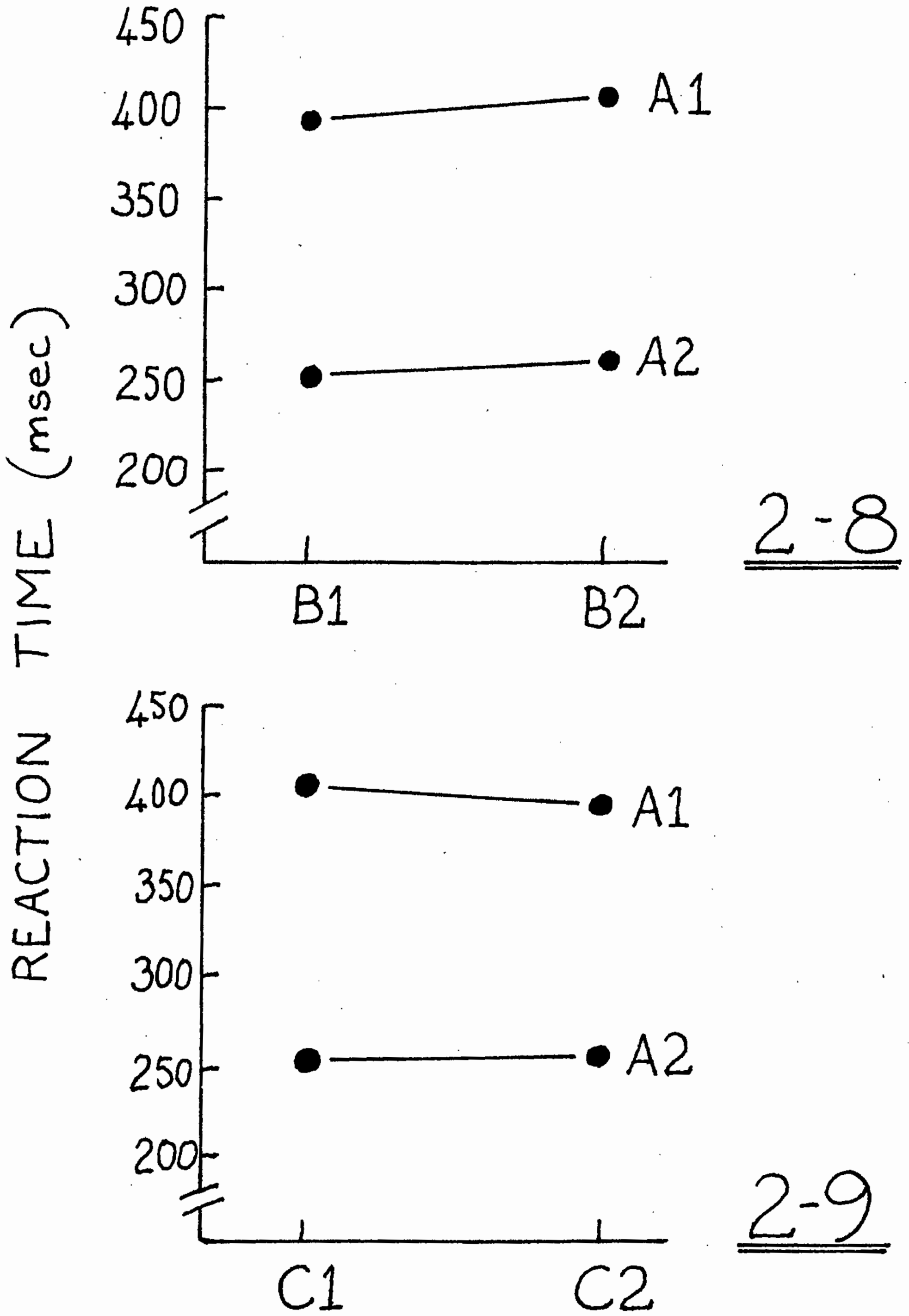
SOURCE	ERROR TERM	SS	DF	MS	F-RATIO
A (TASK)	AS	633797,3	1	633797,3	118,9***
B (PHONETIC SIMILARITY)	BS	6771,6	1	6771,6	0,85
C (SEMANTIC PREDICTABILITY	CS	303,2	1	303,2	0,02
AB	ABS	155,3	1	155,3	0,04
AC	ACS	1747,9	1	1747,9	0,21
BC	BCS	10135,3	1	10135,3	2,89
ABC	ABCS	2104,4	1	2104,4	0,18
S (SENTENCES)		134287,9	15	8952,5	
AS		79951,4	15	5330,1	
BS		120080,1	15	8005,3	
CS		198426,4	15	13228,4	
ABS		54716,3	15	3647,8	
ACS		122344,7	15	8156,3	
BCS		52636,3	15	3509,1	
ABCS		172117,2	15	11474,5	

***p 0,001

THE RELATION BETWEEN TASKS AND TREATMENTS:

The mean reaction times obtained from tasks one and two differed markedly . Phoneme monitoring responses , on the whole , took 404,5 msec. to perform, whereas monitoring responses executed to acoustic cues only took an average of 263,8 msec. The difference of 140,7 msec. between these two means was highly reliable (F1 (1,22) = 15,1 ; F2 (1,15) = 118,9; Min F' (1,27) = 13,4 ; p < 0,001).

Particularly interesting was the entire absence of any interaction between either the "phonetic" (B) or "semantic" (C) treatments and the levels of "task" (A). Figures 2-8 and 2-9 show that, as predicted by H-1, the "task" factor added a constant component of approximately 140 msec. to the baseline RT obtained from task two. Note that the sum of squares for AB, AC and ABC interaction terms together was 4007,6 compared with 10135,3 for the BC interaction term.



FIGURES 2-8 and 2-9 :

Results of Experiment One. Mean reaction-times compared between tasks A1 (phoneme monitoring) and A2 (acoustic cue monitoring and treatments B (phonetic similarity) and C (semantic predictability). B1 = different; B2 = same; C1 = high; C2 = low.

PHONETIC AND SEMANTIC EFFECTS:

Looking next at the other treatment main effects, it is apparent that these were very small indeed.The main effect due to phonetic similarity was only 14,6 msec. (F → 1) and that due to semantic predictabilty was negligeable. Taken across both tasks, there was an indication of some interaction between the B and C treatments. Figure 2-10 shows that when phonetic similarity was low (B1) then responses preceding semantically unpredictable (anomalous) words (C2) tended to be quicker than when phonetic similarity was high (B2). Responses preceding semantically predictable words (C1) seemed to be unaffected by phonetic similarity. However, analysis of variance showed this to be unreliable.

Given the insignificant interaction, the next step was to examine the treatment effects within each task (see Figure 2-11 and Table 2-3). It appeared that monitoring latencies within both tasks were longer when predicatability was low and critical and target phomenes the same than when these were different. On the other hand, when predictability was high then phonetic similarity had little effect. Although this result was not unexpected in terms of the hypotheses formulated, it failed to reach statistical significance and must be regarded as unreliable (F1 (1,22) = 2,78 ; F2 (1,15) = 2,89 ; Min F'(1,35) = 1,42). Turning now to the effect predicted by H-2, we see that within the phoneme monitoring task the effect of semantic predictability was indistinguishable from random error variance. The results indicate no support for H-2.

The phonetic similarity effect, as predicted by H-3, was not observed in this experiment. An examination of the proportions of variance acounted for by each of the experimental treatment factors shows that in the phoneme monitoring task, factor B (phonetic) accounted for more than 20 times as much variance as did factor C (semantic). The relative strength of these two factors is contrary to what was expected in terms of H-2 and H-3.

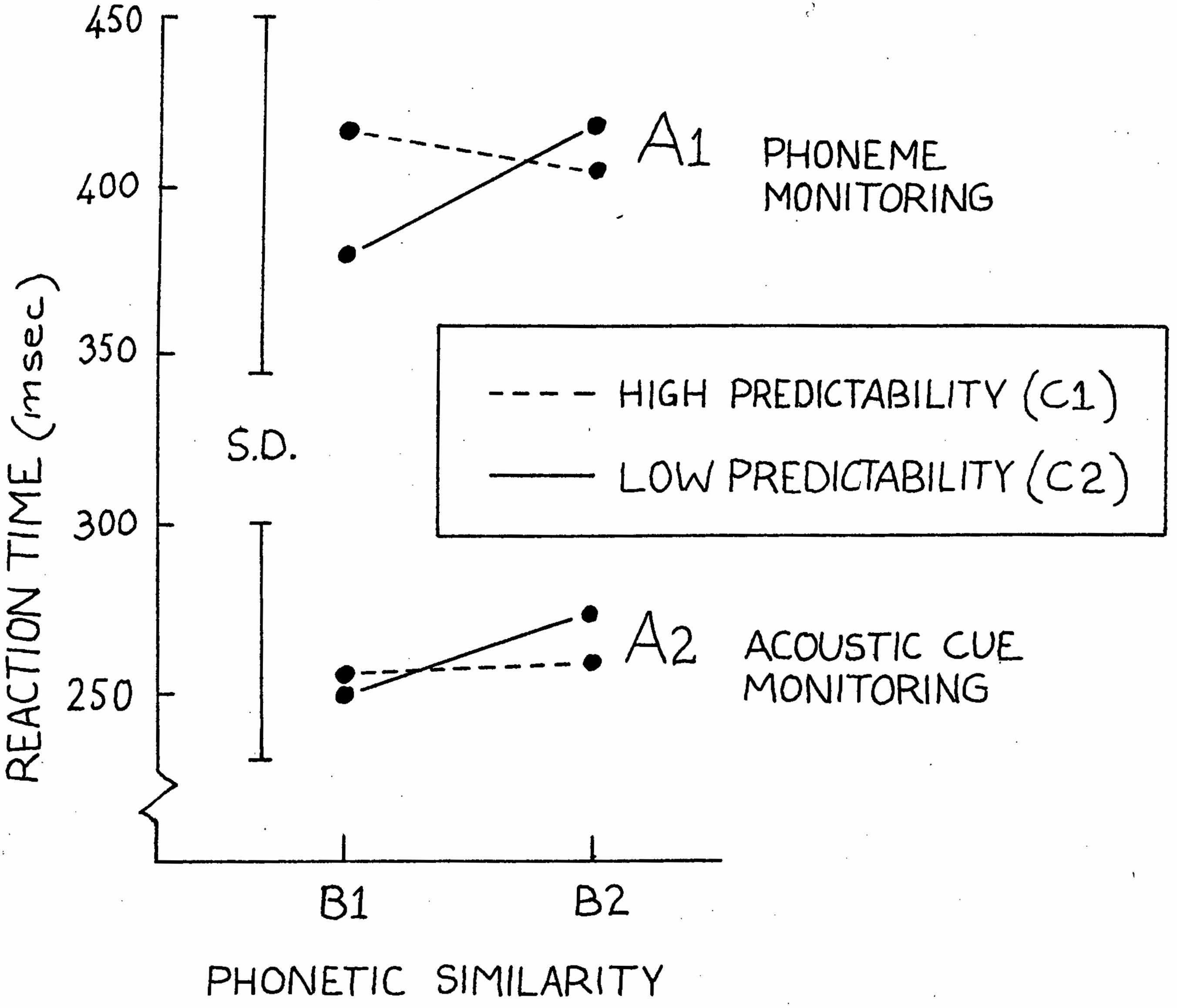


FIGURE 2 - 11 :

Results Experiment One. Cell mean profiles for tasks and treatments. B1 = different critical and target phonemes; B2 = same critical and target phonemes. The standard deviation of reaction-times for each task is shown.

TABLE 2 - 3

Cell means and treatment main effects with standard deviations for phoneme monitoring task.

		PHONETIC SIMILARITY		
		B1	B2	
SEMANTIC PREDICTABILITY	C1	$\bar{X}$ = 416,8 s = 146,8	$\bar{X}$ = 402,8 s = 94,3	409,8
	C2	$\bar{X}$ = 379,8 s = 115,7	$\bar{X}$ = 418,4 s = 114,2	399,1
		398,3	410,6	$\bar{\bar{X}}$ = 404,5 s = 101,8

TABLE 2 - 3 (cont.)

Cell means and treatment main effects with standard deviations of reaction times for acoustic cue monitoring task.

		PHONETIC SIMILARITY		
		B1	B2	
SEMANTIC PREDICTABILITY	C1	$\bar{X}$ = 257,1 s = 62,3	$\bar{X}$ = 264,8 s = 67,6	260,9
	C2	$\bar{X}$ = 253,4 s = 76,3	$\bar{X}$ = 278,7 s = 81,9	266,1
		255,3	271,8	$\bar{\bar{X}}$ = 263,5 s = 69,3

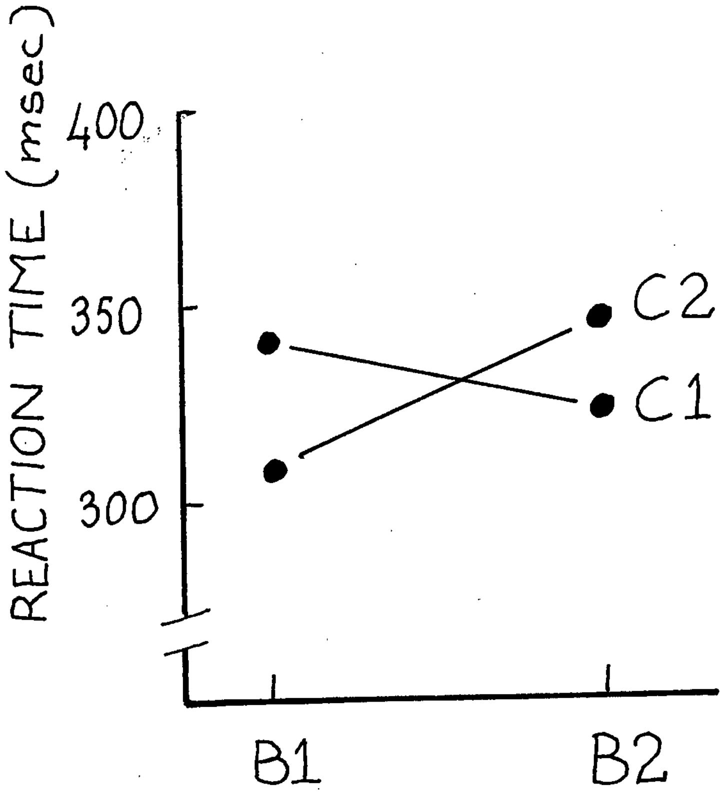


FIGURE 2 - 10 :

Results Experiment One: Treatment main effects across tasks. B1 and B2 = Phonetic Similarity (different, same) C1 and C2 = Semantic Predictability (high, low)

FALSE ALARMS, ERRORS IN RECALL AND OMISSIONS.

In the phoneme monitoring task any positive ("target") response to a phoneme other than /b/ in the catch trials was recorded as a false alarm. Such responses totalled 10 % of all responses recorded. Their small number does not allow of any generalisations, but it is interesting to note that /ɔ/ in "awfully" (catch trial number) gave rise to the same number of false alarms as did /p/ in all other catch trials. The false alarms were roughly equally spread over all subjects. Errors in recall were also recorded. These totalled only 1,5 % of all the recall responses given by subjects in the phoneme monitoring task. No such errors were observed in the acoustic monitoring task.

Errors of ommission occurred in both tasks when subjects did not respond to a real target. In Task one these totalled 3,6 % and in Task two 1,5 % of all responses. In Task one there were 3 scores exceeding one second in duration. These were deleted from the record. No scores were deleted from Task Two records.

WORD LENGTHS:

The critical and target words selected for each treatment level within a sentence frame were roughly matched in length on the basis of number of syllables. During preparation of the tapes, the beginning and end of each critical and target word was marked on the back of the tape. The length of the acoustic correlate of all the words could then be measured directly in centimetres and converted to milliseconds on the basis of tape speed (19 mm/sec. 1mm = 5,26 msec.). Tables 2-4 and 2-5 show the mean duration of critical and target words for each experimental treatment level.

Table 2-4. Critical word length.

	B1	B2	
C1	507,3	394,3	450,8
C2	546,4	408,1	477,3
	526,9	401,2	464,1

Table 2-5. Target word length.

	B1	B2	
C1	438,8	444,1	441,5
C2	491,7	511,6	501,7
	465,3	477,9	471,6

Mean physical duration of critical and target words respectively was 464,1 and 471,6 milliseconds. Note that these word lengths were not equal across all treatment levels despite efforts to match the number of syllables.

In addition to the duration of critical and target words, the duration of the first consonant-vowel syllable (target syllable) for all target words was measured, as well as the length of the pause between critical and target words (i.e. the pause occurring before target phonemes). The average duration of the target syllable was 236,7 msec. and mean pause-length was 129,5 msec. Tables 2-6 and 2-7 give details of these measurements.

Table 2-6. Target syllable length.

	B1	B2	
C1	255,9	255,9	255,9
C2	231,3	203,6	217,5
	243,6	229,8	236,7

Table 2-7. Pause length.

	B1	B2	
C1	113,5	112,1	112,8
C2	128,8	163,3	146,1
	121,2	137,7	129,5

consistently across sentence frames, but that the experimental treatments had widely different effects across sentence frames.

The treatment by sentence frame interactions have considerable implications for the interpretation of the results. Particularly, the observation that the CS interaction mean square (Table 2-2) is almost double the BS interaction mean square leads to the conclusion that sentence frames within treatment C differed widely in the degree of semantic predictability of the target words selected for each frame. In a complete factorial design the mean square for Subjects X Sentence Frames X Treatments could be used to evaluate the significance of the Treatments X Sentence Frame interaction term, but given the fact that this was only a fractional replication of the complete design, some of the interactions involving subjects and sentence frames were confounded . It was thought inappropriate to calculate an error term on this basis . Instead, it was decided to investigate further the actual properties of the linguistic materials used in the experiment.

A FURTHER EXPERIMENT.

Task three was devised in order to measure directly the semantic predictability (treatment C) of target words. Task 3 would give the listener all the same acoustic and semantic information up to precisely the point where monitoring responses on Tasks one and two were cued to begin. The target words, however, were deleted. To measure semantic predictabilty, listeners would be required to guess the appropriate ending to each sentence under normal listening conditions . The percentage of correct responses would be an indication of the on-line predictability of target words under exactly the same conditions and for the identical materials as in the two monitoring tasks.

A MODEL FOR MULTIPLE REGRESSION ANALYSIS.

Initially the purpose of these additional measurements was to verify that the on-line predictability equalled the Cloze predictability reported by Dell and Newman (1980) for similar materials, and that the two levels of treatment C did really differ in degree of semantic predictability. However, the availability of a continuous measure of predictability would also allow a more powerful test of both H-1 and H-2 than was possible with the orthogonal factorial design used above.

As indicatd by the high Treatments X Sentence Frame interaction variance , the C1 treatment level grouped together a set of target words that were not all equally high on predictability. Even if the on-line predictability should turn out to be equivalent to the Cloze measures, a lot of variance would remain "unused" within this group of sentences, with a consequent loss of information.

Would multiple regression allow a more powerful analysis of the monitoring data ? The aim of the next experiment was to obtain a set of scores to be used as a continuous independent variable in a model for multiple regression analysis.

This discussion raises another issue: recall that within a given sentence frame, high and low predictability target words had been matched with respect to their initial consonant-vowel syllables. Perhaps the lack of a semantic effect in the present results could be explained on the basis of this matching. Subjects heard almost exactly what they might have expected to hear up to the point where the inappropriate target word deviated from its appropriate counterpart. If the monitoring response was initiated before this point, then acoustically speaking the difference between appropriate and inappropriate words would effectively have been absent. Perhaps the phonetic matching procedure actually nullified the semantic treatment effect.

Looking at the duration of the target syllable, we see that it varied between 126 and 400 milliseconds, with a mean of 236 msec. Comparing this with the phoneme monitoring RT mean of 404 msec., it is apparent that target matching decisions could have been made before the end of the first syllable was heard, although not in all cases. The influence of target syllable length can be statistically controlled by including it as a co-variate in the analysis of results. Also, the regression of RT on target syllable length will give an indication of the degree to which target matching decisions were influenced by this factor. A model for RT scores is proposed :

$$Y=M+a(A)+b(\mathrm{TARGSYLL})+c(\mathrm{COMPLETN})+d(A*\mathrm{TARGSYLL})+e(A*\mathrm{COMPLETN})$$

where: Y = Monitoring RT

M = Y-intercept

A = Task

TARGSYLL = Target syllable co-variate

COMPLETN = Semantic predictability

a,b,..e = the regression coefficients for each variable in the equation.

The full model states that an RT score is a non-additive function of the nature of the task (phoneme vs. acoustic cue monitoring) , length of target syllable and predictability of the target word. While this model is not intended to give a complete account of the structure of monitoring scores, it is sufficient to allow an evaluation of the first two hypotheses formulated earlier. H-1 will be supported by the finding that a reduced, purely additive model is sufficient to account for the data, i.e. with no interactive component carrying the non-additive effects of the variables under consideration. To evaluate the non-additivity of the variables in the equation,product vectors are generated by multiplying A X TARGSYLL and A X COMPLETN. The separate contribution of these product vectors to the equation is analysed directly by multiple regression (Cohen and Cohen, 1975). TARGSYLL in this model is a covariate used to exercise statistical control over the effect of COMPLETN on the D.V. H-2 will be tested on the basis of the separate contribution of variable C to the equation after the effect of target syllable length has been held constant. If it makes no significant cotribution, then H-2 will be supported.

Not included in this model are the measures of target word length and frequency. These will be examined for their relation to monitoring latency on the basis of a separate multiple regression analysis for which no model has been specified. In terms of H-4 and H-5,RT is expected to be positively correlated with target length and negatively correlated with target word frequency.

2.3. EXPERIMENT TWO : METHOD.

2.3.1. MATERIALS AND DESIGN.

The experimental tapes used in the sentence completion task were the identical tapes used to compile the materials for experiment one. Two changes were made:

(a) all target words were deleted by disconnecting the channel carrying the target words and rest of the sentence thereafter. Thus subjects in this experiment heard everything that appeared over the left-hand channel in the previous experiment. Since the left channel was only erased (see 2.2.2 above) from immediately prior to the initial plosive release of the speaker's lips for /b/, the signal in the present materials retained the acoustic correlate of lip-closure prior to release. Lip closure prior to plosive release often occurred as much as 20 msec. earlier.

(b) The experimental tapes were recorded from the master tape which contained no fillers or catch trials between experimental sentences. The order of presentation of these sentences was the same, however. That is, the order of presentation in two of the tapes was the reverse of the other two. Each tape began with the same set of practice sentences. In this experiment each subject heard all the sentence frames, exactly as in the monitoring tasks, but never heard the words constituting levels of semantic predictability . Therefore, each sentence frame produced two semantically distinct forms, depending on the critical word (that was always heard last). This gave altogether 32 distinct sentences, each having 2 physically different but semantically equivalent forms (because the target word was missing).

PRESENTATION OF THE MATERIALS.

The single (left-hand) channel was presented to both ears of the subjects.

RECORDING THE RESPONSES.

Subjects were required to say aloud their responses and these were recorded over a microphone placed in front of the subjects. A tape recording was made of each session. This included the initial part of each sentence as heard by the subject, followed by the subject's response.

2.3.2. SUBJECTS :

24 male and female volunteers from the same population as in the first experiment were randomly assigned to each of the four experimental tapes. All subjects spoke fluent English. Two volunteers who turned out not to speak English as a home language were not included in the sample.

2.3.3. PROCEDURE :

Each volunteer was approached in the same way as in the previous experiment. Subjects who volunteered were told they we~~k~~re participating in a sentence completion task . They were instructed to listen carefully to each sentence and immediately after the last word to say aloud whatever word occurred to them as an appropriate continuation of the incomplete sentence. They were encouraged to be spontaneous but not consciously original. Only one word was needed, and this should be their first reaction. They were encouraged to respond as quickly as possible and to refrain from giving any thought to their responses.

The subject was seated in the experimental room and the experimenter left to operate the equipment. The 10 practice sentences were presented first, during which time the volume level in the earphones was adjusted to suit the listener. The instructions were repeated if the subject did not seem to respond quickly enough. The experimental sentences were then presented and no further dialogue was engaged in until the end of the experiment. The duration of each session was approximately 10 minutes. At the end of the session the subject was debriefed in a manner similar to the first experiment.

2.4. RESULTS : EXPERIMENT TWO.

The number of correct guesses was found for each sentence. Within each sentence frame responses were totalled for "same" versus "different" phonetic treatments. Since there were 24 subjects, a maximum of 12 correct responses could be obtained for each sentence in this way. In addition to the number of correct responses, the latency of each response was recorded. Latency was measured directly from the tape by marking the distance between offset of the final word and onset of the subject's response. Distance was converted to milliseconds on the basis of tape speed (19 cm per sec.: 1mm = 5,26 msec.) .

COMPLETION SCORES :

The average number of correct responses to sentences with "different" versus "same" critical words were 5,06 and 5,13 respectively. Converted to percentages these scores give 42,2 % and 42,7 % semantic predictability. The results show that the level of predictability was substantially the same for both levels of phonetic similarity in the experimental materials used in the previous monitoring tasks, i.e. that treatments B and C in Exp.One were indeed orthogonal.

The scores for each sentence at each level of phonetic similarity were examined to see if they departed significantly from the mean for each group. A Chi-squared test was performed of the null-hypothesis that these scores did not differ from 5,06 and 5,13 respectively. It was found that the observed proportions of correct responses differed significantly from a chance distribution (χ^2 = 68,14; df = 31; p < 0,001). On the basis of this test it was concluded that the completion scores obtained using Task 3 provided a real measure of the semantic predictability of target words.

RESPONSE LATENCY:

Average response latency was 1167,2 milliseconds, with a standard d deviation of 424,6 milliseconds. These data were retained for inclusion in a regression analysis.

DEBRIEFING:

None of the subjects reported noticing that many of the appropriate responses had an initial /b/ phoneme. However, when prompted, some subjects did agree that they noticed the cue to initial /b/(lip closure). If this was indeed perceived it could have influeced the pattern of repsonses given, but there was no evidence that this happened.

2.5. RESULTS : MULTIPLE REGRESSION ANALYSIS.

In the analyses reported below the dependent variable was always monitoring RT. This had 127 degrees of freedom based on the number of distinct sentences used in the materials multiplied by the two levels of Task . Within each task, sentences were really repeated measures of each sentence frame. However, in order to simplify the regression analyses the sentences were regarded as if they were 128 independent cases. The known high variance due to sentence frames would procuce conservative significance tests in this way, since it was not partialled out of the error term (Cohen and Cohen, 1975).

As a preliminary to multiple regression analysis, 11 variables were entered into a correlation matrix. The monitoring dependent variable was split into two distinct variables (called Y-PHONEM and Y-ACOUST) for this analysis (these were the same as the A-1 and A-2 task dependent variables in the earlier analysis). Thus there were 64 records for each D.V. , these being the average scores over three subjects for each sentence.

In all the regression analyses that follow, categrical variables were coded orthogonally and interaction variables were obtained from the product of the variables concerned (Overall and Spiegel, 1969; Kerlinger and Pedhazur,1973).

The independent variables IVPHON, IVSEMANT and "11 X 12" were the Phonetic Similarity and Semantic Predictability variables and their interaction as used in previous analyses. The other variables were the continuous measures of the linguistic and acoustic variables already described.

In order to normalise the sentence completion proportions, these were first submitted to an arcsine transformation (Edwards,1951). The disribution of word frequency scores was found to be heavily skewed, so that a square-root transformation was applied to these.

An inspection of the correlation matrix (Table 2-8) showed that the zero-order correlations of the independent variables with the phoneme monitoring scores were much higher than with the acoustic cue montoring scores. For example, target syllable length (TARGSYLL) , completion scores (COMPLETN) and their interaction vector ('7*8') correlated -0,3387, -0,2873 and -0,3692 with Y-PHONEM but only -0,0337, -0,0997 and -0,0618 respectively with Y-ACOUST. Thus, high completion scores (semantic predictability) and long target syllables were associated with short phoneme monitoring reaction times, but apparently acoustic cue monitoring did not vary in the same way. The correlation between the two dependent variables was only 0,1007. This finding was worthy of further investigation.

		PHONEME MONITORING REACTION TIME 1 Y-PHONEM	ACOUSTIC CUE MONITORING REACTION TIME 2 Y-ACOUST	CRITICAL WORD LENGTH 3 CRITLGTH	TARGET WORD LENGTH 4 TARGLGTH	SQUARE ROOT OF CRITICAL WORD FREQUENCY 5 CRITFREQ	SQUARE ROOT OF TARGET WORD FREQUENCY 6 TARGFREQ	LENGTH OF TARGET WORD SYLLABLE 7 TARGSYLL	ARCSIN TRANSFORMED SENTENCE COMPLETION SCORES 8 COMPLETN	LATENCY OF SENTENCE COMPLETION RESPONSES 9 LATENCY	LENGTH OF PAUSE BETWEEN CRITICAL & TARGET WORDS 10 PAUSE
Y-PHONEM	1	1.0000									
Y-ACOUST	2	.1007	1.0000								
CRITLGTH	3	-.1001	.0889	1.0000							
TARCLGTH	4	-.0034	.0776	-.0151	1.0000						
CRITFREQ	5	-.0612	.0023	-.1430	.0986	1.0000					
TARGFREQ	6	-.1280	.4164	.1797	.0322	.0968	1.0000				
TARGSYLL	7	-.3387	-.0337	.1231	.0834	-.2045	.1532	1.0000			
COMPLETN	8	-.2873	-.0997	-.0486	-.1849	.2004	.0885	.1160	1.0000		
LATENCY	9	.1400	.1337	-.1332	.0410	.0250	.0685	.0903	-.2270	1.0000	
PAUSE	10	-.0218	.2645	.0883	.3592	-.0059	.4709	-.0453	-.2326	.2384	1.0000
7*8	11	-.3692	-.0618	.0169	-.1061	.0407	.1787	.5748	.8511	-.1828	-.1857

TABLE 2 - 8 :

Matrix of zero-order correlations between all variables for which measurements were obtained. The dependent variables are 1 and 2. Variable 11 carries the interaction between variables 7 and 8; it is the product of these two variables.

Note that the product vector '7*8' was very highly correlated with TARGSYLL and COMPLETN. It was not expected to make a significant unique contribution to an analysis of variance components. None of the other independent variables correlated very highly with the dependent variables. Completion response latency (LATENCY) correlated only 0,1400 with Y-PHONEM ; pause length (PAUSE), correlated -0,0218 with Y-PHONEM, and these two variables did not contribute enough to monitoring variance to warrant further investigation.

The eleven variables together accounted for only 24 % of the phoneme monitoring variance (multiple R = 0,4930; F (10,53) = 1,70).

On the basis of the observed correlations it was decided to proceed further with analysis of the model as previously specified. Note that the next analysis grouped Y-PHONEM and Y-ACOUST into one dependent variable.

TESTING THE MODEL:

The model was tested by multiple regression analysis using a method which orthogonalised all the variance components so that the unique contribution of each variable to the equation could be estimated. This was accomplished by calculating the regression of the D.V. on all the other variables to obtain the residual mean square. Next, each variable in turn was removed from the full model to obtain its squared semi-partial correlation with the D.V., but replaced before removing the next variable . The squared semi-partial coefficient is the unique proportion of variance in the D.V. accounted for by the variable in question (in the context of the other independent variables). For the variable i,

$$r^2_{y(i.123..(i)..k)} = R^2_{y.123.....k} - R^2_{y.123....i}$$

Using this method, the order of entry of the variables into the equation is not important. The significance of the squared semi-partial coefficient is calculated by an F-ratio

$$F_{\text{to remove}} = \frac{(R^2_{y.12...i.k} - R^2_{y.12..(i)..k})}{(1 - R^2_{y.12...i..k})} \times \frac{N - k - 1}{k - j}$$

where $R^2_{y.12..i.k}$ is the squared multiple correlation coefficient for the full model; $R^2_{y.12...j}$ is the squared multiple correlation coefficient for the full model minus the variable of interest; k=the number of I.V's in the full model; j=the number of I.V's in the reduced model. In the present case, k=5 and k-j=1 always.

The equation obtained for the full model was

$$Y = 427,51 - 143,37\ (A) - 0,241(\text{TARGSYLL}) - 25,75(\text{COMPLETN}) + 0,22\ (3*A) + 15,03(4*A).$$

Substituting the maximum and minimum observed values for TARGSYLL and COMPLETN, and +1 and -1 respectively for Task 1 and Task 2, the equation can be expressed graphically. Figure 2-12 shows monitoring RT against the co-variate TARGSYLL. The mean value of RT expected for each level of Task is shown at two levels of Semantic Predictabilty (after adjustment for target syllable length) , by the Y-intercepts of the four regression lines (Cohen and Cohen, 1975).

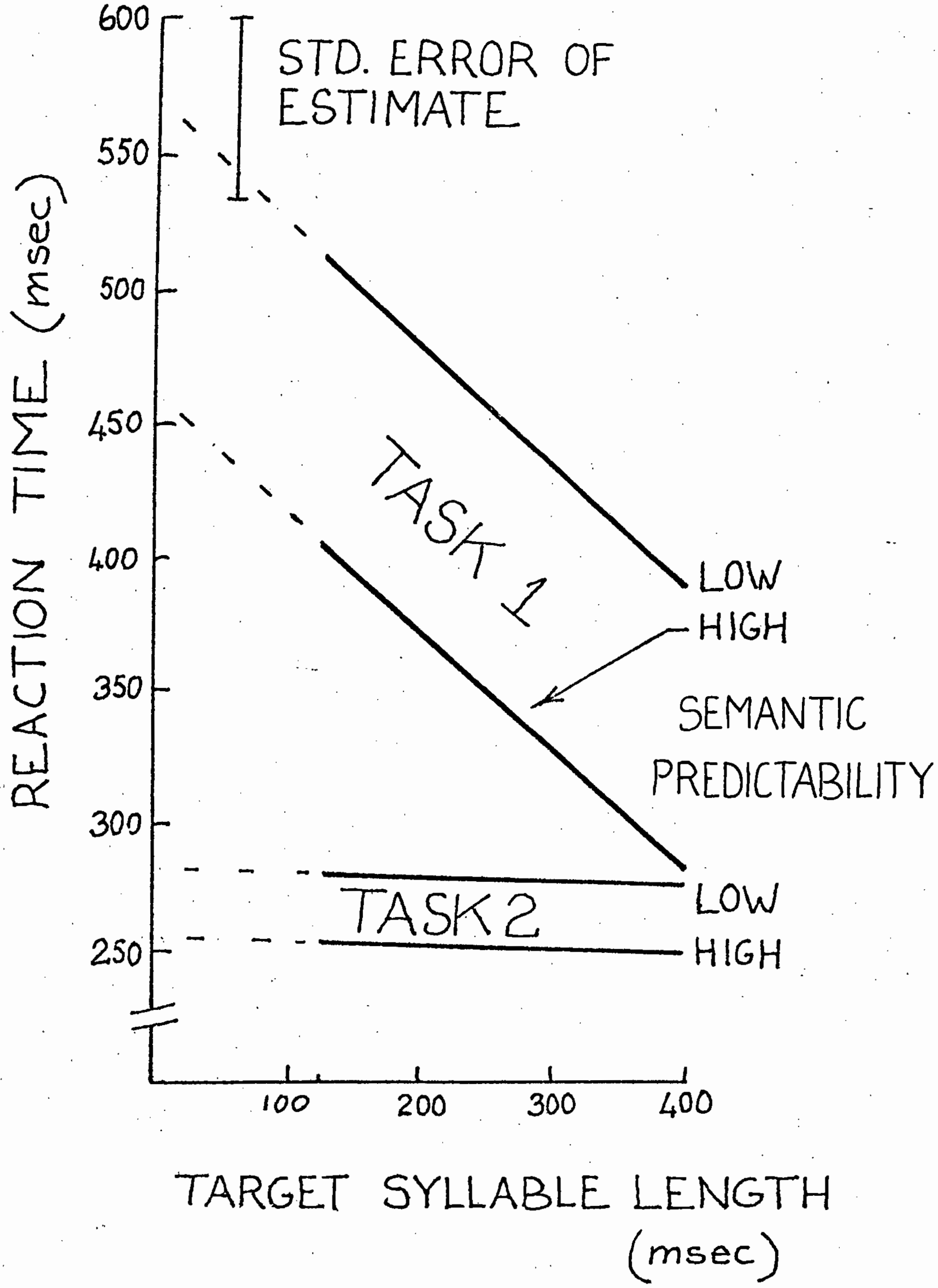


FIGURE 2 - 12 :

Illustration showing the effect of, and the relation between variables in the model for reaction time scores. The range of semantic predictability is 0 - 92 % from high to low. Target syllable length ranges between 126 and 400 msec. The standard error of the estimate for the model is shown.

The difference between the Y-intercepts of the two regression lines shown for each level of Task indicates the range of the semantic predictability effect after adjustment for target syllable length.

For the phoneme monitoring task the difference between intercepts was 104,3 msec. and for the acoustic cue montoring task this was 27,4 msec. Note that these are not experimental treatment effects. For this reason the difference between the intercepts was not tested for statistical significance. Instead, the specific contribution of each variable to the model was tested. Table 2-9 shows the multiple regression analysis results for the full model.

The proportion of variance accounted for in monitoring scores by the full model was 47 % ($R^2 = 0,4735$; F (5,122) = 21,94; $p < 0,001$). The standard error of the estimate was 82,825, which was a substantial reduction from the standard deviation of 111,87 in the dependent variable .

TABLE 2-9

Analysis of variance summary table: The unique contribution of each variable in the full model.

SOURCE	SS	DF	MS	F
0 Y	1589575,1	127	12516,34	
1 A	157985,1	1	157985,1	23,03
2 TARGSYLL	34025,45	1	34025,45	4,96
3 COMPLETN	32516,2	1	32516,2	4,74
4 '2*A'	27910,0	1	27910,0	4,07
5 '3*A'	11070,6	1	11070,6	1,61
7 Residual	836916,34	122	6859,97	
Check on homogeneity of the covariate with respect to levels of COMPLETN:				
8 '2*3'	14783,05	1	14783,05	2,18
9 Residual	822082,06	121	6794,06	

TABLE 2-10

Multiple regression analysis: Coefficients of variables in the full model.

SOURCE	COEFFICIENT	STD. ERROR OF COEFFICIENT
0 Y	427,51	
1 A	-143,375	29,879
2 TARGSYLL	-0,241	0,108
3 COMPLETN	-25,754	11,834
4 'A*2'	0,2183	0,108
5 'A*3'	15,034	11,834

$R^2 = 0{,}4735$

$R = 0{,}6881$

Shrunken $R^2 = 0{,}4519$

Standard error of estimate = 82,825

	SS	DF	MS	F
Regression	752658,75	5	150531,7	21,94
Residual	836916,3	122	6859,97	

Looking at the results for each variable in the model, it is apparent that despite the high statistical significance of the overall regression coefficient, not all the variables contributed equally to it. The Task variable (A) had an F-to remove ratio of 23,03, while the variable 'A*4' which carried the interaction between Semantic Predictability and Task did not reach statistical significance (F = 1,61; df = 1,126). The other three variables all had acceptable F-ratios. Note that for each variable except 'A*4' the standard error of the coefficient is less than half the coefficient itself (see Table 2-10). A test of the homogeneity of the covariate with respect to Semantic Predictability was performed by calculating the increment (in the proportion of variance accounted for) due to the interaction vector '2*3' after all the other variables had entered the equation (Kerlinger and Pedhazur, 1973) . It was found that this increment was insignificant (F = 2,18; df = 1,121) and the variable was dropped from further analyses, and not included in the model.

ADDITIVITY OF THE COMPONENT VARIABLES:

The result concerning variable A confirms the earlier results; there was a substantial difference between phoneme and acoustic cue monitoring RT. However, looking at Figure 2-12 it is apparent that semantic predictability did have a much stronger effect on acoustic cue monitoring performance than on phoneme monitoring performance. The present results seem to be inconsistent with H-1. That is, Task was apparently not an additive component in the model for RT scores. This non-additivity is carried by the variables 'A*3' and 'A*4' (Cohen and Cohen,1975 ; pg. 295) . The latter was not statistically significant, but the former was, indicating that Task was additive with respect to semantic predictability but not with respect to target syllable length.

In order to estimate better the main effect of the three variables in the additive model, a calculation of the variance components was done using the method commonly applied to experimental designs where the main effects are first tested against the model disregarding interactions, and the interactions tested against the additive components. Table 2-11 shows the composition of the sums of squares obtained in this way. The table also shows the results of these calculations. The dramatic rise in the F-ratio for variable A (using the same error term as before) from 23,03 to 92,39 is an indication of the fact that the latter figure includes a substantial portion of variance due to interaction. Calculated in this way without the interactions partialled out, there was little change in the results for variables 3 and 4. The F-ratios (4,96 and 4,72 respectively) remain essentially the same. Without the effect of the other interaction term partialled out, variable 'A*4' now accounts for more variance, but remains statistically insignificant (F = 2,29; df = 1,126) . Taken together, however, the two variable carrying the interactions in the full model accounted for just under 3 % of the variance (r^2 = 0,0275) which was statistically reliable enough to consider the model to be non-additive (F = 3,19; df = 2,125; p < 0,05).

TABLE 2-11

Anaysis of variance summary table: "Experimental Design" method.

SOURCE	SS	DF	MS	F
0 Y	1589575,1	127	12516,34	
1 A	633797,4	1	633797,4	92,39
2 TARGSYLL	34009,07	1	34009,07	4,957
3 COMPLETN	32427,3	1	32427,3	4,72
4 A*2	32573,4	1	32573,4	4,75
5 A*3	15736,8	1	15736,8	2,29
6 Residual	836916,8	122	6859,97	

Composition of Sums of Squares for the analysis of variance

SOURCE	
1 A	$SS_{total} (R^2_{y.123} - R^2_{y.23})$
2 TARGSYL	$SS_{total} (R^2_{y.123} - R^2_{y.13})$
3 COMPLETN	$SS_{total} (R^2_{y.123} - R^2_{y.12})$
4 A*2	$SS_{total} (R^2_{y.12345} - R^2_{y.1235})$
5 A*3	$SS_{total} (R^2_{y.12345} - R^2_{y.1234})$
6 RESIDUAL	$SS_{total} (1 - R^2_{y.12345})$

The finding that the non-additive component in the full model contributed significantly to total variance accounted for (in the RT scores) was incompatible with H-1, although not highly reliable.

PHONEME MONITORING PERFORMANCE:

Given the significant interactions in the full model it was appropriate to examine separately the phoneme monitoring task.

Table 2-12 shows that phoneme monitoring performance varied significantly as a function of semantic predictability of target words (F = 4,61; df = 1,60; p < 0,05). This was the result obtained after variance attributable to target syllable length had been partialled out of the predictability effect (that is, treating TARGSYLL as a co-variate). The covariate was also significantly related to phoneme monitoring performance (F = 7,01; df = 1,60; p < 0,05). The interaction between the covariate and semantic predictability did not contribute to the variance accounted for , indicating that the effect of the co-variate was homogeneous with respect to all observed levels of the independent variable. This result lends support to H-2 : There was a significant inverse relation between phoneme monitoring RT and the level of semantic predictability of target words.

TABLE 2-12

Multiple regression and analysis of variance of phoneme monitoring data.

SOURCE	SS	DF	MS	F
1 Y	653299,85	63	10369,8	
2 TARGSYLL	61951,6	1	61951,6	7,01
3 COMPLETN	40741,3	1	40741,3	4,61
4 2*3	7335,2	1	7335,2	0,83
5 RESID.	530255,5	60	8837,6	

R = 0,4208

R^2 = 0,1771

Shrunken R^2 = 0,1501

Standard Error of Estimate = 93,87

	SS	DF	MS	F
Regression	115702,8	2	57851,4	6,56
Residual	537597,1	61	8813,1	

Regression equation:

Y = 570,88 - 0,459(TARGSYLL) - 40,78(COMPLETN)

Standard error of coefficients	0,174	18,97

THE EFFECT OF TARGET WORD FREQUENCY AND LENGTH:

In order to determine the unique contribution due to target word frequency and length, a stepwise regression analysis was performed with eight independent variables . These were

(1) Semantic Predictability (COMPLETN)
(2) Phonetic Similarity (IVPHON)
(3) Semantic Predictability (IVSEMANT)
(4) Target word frequency (TARGFREQ)
(5) target word length (TARGLGTH)
(6) Target Syllable length (TARGSYLL)
(4) Two interaction terms.

The results of the first 5 steps are shown in Table 2-13. The F-ratios for both TARGFREQ and TARGLGTH remain consistently insignificant throughout. These two variables did not have any discernable effect on phoneme monitoring performance. Note that the F-ratios for COMPLETN and for TARGSYLL remain consistently high throughout the stepping procedure, and that none of the other variables reached significance. These results were inconsistent with H-4 and H-5, and again confirm H-2 and H-3. Note that the variable carrying the interaction between 1 and 3 remains insignificant at all the steps. This means that the regression coefficients for RT on COMPLETN are homogeneous with respect to both levels of the categorical variable IVSEMANT. The application of the sentence completion scores to all target words, both appropriate and inappropriate, was justified.

TABLE 2-13

Step-wise regression analysis: Variables and the significance of their contributions to multiple regression analysis at each step (F-to-remove), or their potential significance if they were to enter at the next step (F-to-enter).

F-to-enter or F-to-remove Ratios

STEP	1 COMP-LETN	2 IV-PHON	3 IV-SEMANT	4 1*2	5 1*3	6 TARG-FREQ	7 TARG-LGTH	8 TARG-SYLL
0	5,58	0,23	0,17	0,52	0,00	1,46	0,00	8,04
1	5,58	0,25	0,18	0,60	0,00	1,05	0,22	7,01
2	5,51	0,25	0,18	0,62	0,00	1,03	0,25	6,73
3	5,44	0,24	0,18	0,61	1,00	1,01	0,14	7,94
4	5,44	0,24	1,18	0,61	1,00	0,61	0,13	6,93
5	5,52	0,26	1,17	0,61	0,99	0,78	0,17	8,34

F-RATIOS BELOW THE STEPPED LINE REPRESENT VARIABLES INCLUDED AT THAT STEP

2.6. DISCUSSION OF RESULTS.

The auditory-verbal completion task devised for this experiment was useful to determine the real on-line predictability of target words. Within the a priori semantic predictability treatment levels, words had widely different levels of on-line predictability. On average, it was shown that the difference in degree of predictability between levels C1 and C2 in Experiment One was similar to that reported by Dell and Newman (1980). For the present materials the levels were 42 % and 0,0 % respectively, while Dell and Newman give 46 % and 0,8 % for their materials.

The scores obtained from only a small sample (N = 24) of listeners varied consistently: one "highly predictable" target word was never guessed correctly ("At the lending library the banker borrowed ..."), while three target words were guessed correctly 11 times out of 12, (... solid brass; ...the bowl broke; ... the pretty bird..."). Clearly the completion scores provided information about variations in predictability within the semantic treatment levels. That this was successful is demonstrated by the observation that the orthogonal component of phoneme monitoring variance attributed to the quantitative I.V. (COMPLETN) had a sum of squares = 40741,3 (Table 2-11) while for the categorical I.V. the sum of squares was 1722,25 (Appendix).

This demonstates the validity of the point made by Kerlinger and Pedhazur (1973) and by Cohen and Cohen (1975) that where a quantitative I.V. is available, a conventional analysis of variance might be inappropriate.

How must the model for RT scores be interpreted? The model is an abstraction from the data which tells us that there is a certain relationship between the variables included in the equation. Figure 2-12 is an aid to visualising this relationship, but should not be regarded as a summary of the data in the same way as Figure 2-1. The results favour the conclusion that the relationship between the variables is non-additive, and therefore the model is inconsistent with H-1. However, this is only the case if target syllable length is regarded as an integral part of the model. On its own the variable representing semantic predictability did not interact significantly with Task. But with target syllable length as a covariate, the degree of interaction between COMPLETN and Task became more significant, and taken together, the pooled interactions just reached an acceptable level of significance.

Given the basis for postulating H-1, when it is false the conclusion must be that there were no divided attention effects in the phoneme monitoring task. The acoustic cue monitoring task gave no reliable indication of any systematic effects due to the semantic or phonetic treatments. It could therefore be regarded as a "neutral" control task, against which to compare phoneme monitoring performance: semantic effects in task 1 were specific to phoneme monitoring, and most probably mediated by lexical access. The conclusion is that performance in task 1 depended on target word processing as well as on target phoneme matching. Attentional effects were ruled out or controlled.

This conclusion is in accord with the findings of Morton and Long (1978) and of Foss and Blank (1980).

However, this interpretation of the results should be moderated by the consideration that the covariate TARGSYLL is not very well understood. The criterion for this variable was suggested by the notion "recognition point" introduced by Marslen-Wilson and Tyler (1980; 1981), and refined by Grosjean (1980). Starting form the beginning of a word, this is the point "at which the word in question becomes uniquely distinguished from all of the other words in the language, beginning with the same sound sequence, that are also compatible with the available context". (Marslen-Wilson and Tyler,1981; page 17). Closely allied to this is the "non-word point" used by these authors as a criterion in experiments with sound sequences that started as familiar words but turned into non-words at a certain point. By analogy, TARGSYLL should tell us how long it took before listeners would discover that some target words starting with an appropriate sound sequence were in fact inapropriate. Yet the evidence from the experiments does not bear this out. In the phoneme monitoring data, TARGSYLL had its strongest effect when target words were semantically appropriate (i.e. at C1). The precaution was taken of checking for an interaction between TARGSYLL and the categorical I.V. (C), but this was found to be negligeable. Although defined in terms of the dimension appropriate /inappropriate (semantically), in the data no such relationship was found. TARGSYLL interacted much more reliably with Task: the slope of the regression lines for Task 2 was almost zero, showing that TARGSYLL had no effect during acoustic cue monitoring. The observed main effect was therefore almost entirely due to variations in phoneme monitoring. The reason for this could be that when two consonants follow each other closely, backward auditory masking will occur. The range of such effects might vary between 150 and 250 msec (Studdert-Kennedy,1976). The shorter target syllables (120 msec was the shortest syllable observed) in the present materials could plausibly have involved some masking of the target phoneme by the following consonant (end of target syllable). A negative correlation (as observed) between duration of target syllable and reaction time would be expected, but there was no other evidence to back up this conjecture.

Another possibility is that TARGSYLL measured not masking but stress. This would also accord with the observed negative correlation.

Stress in English speech is primarily realised through sentence intonation, secondly by syllable duration and lastly by intensity (Lehiste,1976). On its own, duration does not determine perceived stress. The syllable must have the phonological potential to receive stress. In the present materials, target syllables were all in the "stressed" position, but the degree of stress they received could have varied, leading to systematic variations in acoustic duration. The measure of target syllable length reported here could have reflected these differences.

Respiratory effort has so far been the most reliable objective correlate of perceived stress; Lehiste (1976) uses the term "accent" to refer to prominence achieved by means other than respiratory effort, but given the intimate relation between stress / respiratory effort, intensity and duration, it is plausible that longer target syllables received prominence also through greater intensity, making them generally more perceptible ; hence they were responded to more quickly.

Returing to the model for reaction time scores, TARGSYLL could now be interpreted as a statistical control for differences in degree of accentuation in target syllables. A substantial amount of variance in the dependent variable was successfully controlled by including this variable in the model. Note also that TARGSYLL and COMPLETN operated largely independently of each other: there was no significant interaction between them. Target syllables were accentuated regardless of the degree of semantic predictability of the target words.

3. GENERAL DISCUSSION AND CONCLUSIONS.

Turning again to the Dell and Newman (1980) study, we see that the phonetic similarity effect they reported was absent under conditions where target location cues were provided; it must have resulted from confusions in target search and could not have operated at the level of target matching. Once the former was controlled in the ASCM tasks, it no longer could delay monitoring responses. However, a study comparing performance both with and without location cues would be required to settle this issue definitively.

The fact that semantic effects on phoneme monitoring were nevertheless observed in the presence of a location cue suggests that monitoring decisions did not occur on the basis of phonetic feature matches as claimed by Dell and Newman. Direct access to a phonetic feature code would by-pass lexical effects and would not vary as a function of semantic predictability. On the basis of the present evidence we are obliged to deny the validity of the parallel access hypothesis proposed by Dell and Newman (1980) and by Foss and Blank (1980).

Addressing the issue of the long duration of phoneme monitoring latencies relative to word monitoring latencies, we attempted to use the subtractive method to determine the duration of the target matching component in phoneme monitoring performance. However, the interaction discovered within the model depicted in Figure 2-12 invalidates the 140 msec estimate given in Section 2.2.5 for the duration of this component. Very short target syllables might inflate the estimate, and when the target syllable approaches 400 msec, no real difference is expected between monitoring latencies on the two tasks. The latter finding is especially interesting, because it indicates that acoustic source cue monitoring might not be a "pure" baseline measure of response latency after all. The high average latency in this task itself needs to be accounted for: 263 msec is not typically the duration of "pure" reaction time (100 msec would be a more likely estimate), and this measure probably contains within it a "cost" component due to dividing attention between monitoring , speech comprehension and rehearsal for later recall. The interaction observed by Dell and Newman between phonetic similarity and semantic predictability, as remarked earlier, could have been an artifact caused by a third variable uncontrolled in their experiments . Quite possibly variations in accentuation (prominence) of target syllables, or backward auditory masking caused by the consonant immediately following the target could have been responsible for the pattern of results these authors observed. If either of these interpretations of TARGSYLL in the present study is valid, then it must take credit for having discovered yet another potentially confounding variable to add to the long list already known. In the same way as these authors (Newman and Dell,1978) reported the potentially confounding effect of phonetic similarity between target and critical phonemes in speech monitoring studies, we report here a reliable and substantial (nine percent of total variance) effect not previously taken directly into account. We have shown that despite efforts to ensure that all target syllables were stressed, some evidence was found that stress (more precisely accentuation) was not evenly distributed. Having made available precise measurements of the stimulus materials presented to listeners in this study, it showed that variation existed where none was expected. Not only was target syllable accentuation found to be unevenly distributed, it was also found that target word length varied widely despite control over the number of syllables in these words. Multiple regression analysis showed this to have been an unimportant factor, but the lack of control has been demonstrated.

Comparing the results of Experiment One with the results of the multiple regression analysis, it appears that not appropriateness but on-line predictability of target words affected phoneme monitoring performance. The lack of an effect due to inappropriate target words was earlier assumed to be the result of phonetic matching between target syllables of inappropriate and appropriate words. The rationale for the matching was that it would rule out phonetic differences that could confound the semantic effect. If, as claimed by some researchers (e.g. Marslen-Wilson and Tyler, 1980) the listener requires to process the target word before identification of its phonetic segments is possible, then phonetically matched target words should nevertheless produce a semantic effect on phoneme monitoring. This effect was not observed in Experiment One. Further, under ASCM conditions, listeners responded to phonemes with a mean latency shorter than the acoustic duration of target words (404 msec versus 472 msec; compare these results with those summarised in Table 1-2 above).

We conclude therefore, that processing the entire acoustic input of target words was not a prerequisite for target phoneme matching. Moreover, we have evidence showing that listeners could have initiated phoneme monitoring responses prior to hearing the end of the first syllable of target words. Consider that,

(a) mean target syllable length was 237 msec and,

(b) acoustic cue monitoring responses were executed within 264 msec on average and that,

(c) there was no significant difference between the zero-order correlations of COMPLETN with each of the two categorical levels of the semantic predictability treatment.

Given (a) and (b) it is plausible to assume that phoneme monitoring decisions were made within the duration of target syllables (as defined) . Given the acoustic input up to that point the inappropriate words hardly differed from their appropriate counterparts. As shown by (c), COMPLETN correlations with RT for the two categories of target words can be assumed not to differ in reality. Altogether, this suggests that on-line predictability of target words had its effect on monitoring RT before processing of all the acoustic input of target words was completed. Therefore it is possible to maintain that semantic or lexical factors had an effect on monitoring RT, even though the duration of the acoustic signal for target words had no such effect. This conclusion is compatible with the point illustrated in Table 1-1 above, namely that given increasing semantic constraints , word-processing dependency on acoustic input declines (Marslen-Wilson and Tyler, 1980; Grosjean, 1980).

An alternative interpretation of the results might be as follows: One difference between the categorical and the quantitative semantic independent variables is that the former would have had both facilitatory and inhibitory effects, and the latter only facilitatory effects. Absence of an effect due to hearing semantically inappropriate words might be evidence that these had no inhibitory effect on speech processing. The listener's attention was controlled in the present experiment to ensure that conscious attention would not be directed to target search. In

ASCM the target location cue and not the target word had high relevance to locating targets. Thus target word processing was not consciously attended to. We further assume that there was no conscious attention involved because we found no inhibition but facilitation by on-line semantic predictability. This is characteristic of automatic processing (Posner and Snyder,1975). This asymmetry was also observed by Fischler and Bloom, (1979) in a visual processing task. They found that inhibition did result from the presentation of anomalous words, but this depended on the deployment of attention. Moreover, inhibition was not found with unlikely but semantically appropriate words, leading to the conclusion that sentence processing is automated to the extent that all possible word candidates are activated by a given context, not only the most likely ones (c.f. the present finding of no relation between performance and target word frequency). Their conclusion recalls Marslen-Wilson and Tyler's (1981) emphasis on the obligatory nature of linguistic processing. The present findings tend to confirm this.

We conclude that the encoding and/or matching of target phonemes was facilitated by top-down effects of a phonological or lexical nature. While subjects were attending to the target location cue, they were not able to "switch off" the automatic speech processes. Though irrelevant for the purposes of the task at hand, the phonological coding of the target stimulus intruded on performance. Subjects were unable to respond purely "phonetically" even under optimal conditions, because the phonological coding of speech is inevitably present as a result of automatic processing sequences.

3.1. REFLECTIONS.

Anne Cutler (1981) somewhat mischievously described the gloomy prospects facing a psycholinguist in search of appropriate language materials for use in experiments. The sheer number of potentially confounding variables makes experimental control a virtual impossibility. One aspect of this problem has been noted here also: it is apparent that the researcher who takes a sample of words or sentences from any particular corpus really faces the same difficulties associated with the sampling of subjects. Any particular word, like any particular person, netted and brought into the laboratory, carries with it a host of confounding variables; unlike a person, it is the deceptively unlikely container of a host of known and unknown factors. Rather than the psycholinguist, the poet would be the most qualified to assess each word for its unique interaction with other words and above all, its effect on a particular listener. For it is when in the laboratory persons and words are brought into intimate contact, that a particularly fascinating chemistry begins to occur. The psycholinguistics laboratory turns out to be a kind of test-tube where unknown compounds are rather carelessly blended : not surprising that the results are frequently unpredictable. Although not confronted with sudden explosions in the chemistry of psycholinguistics, we are familiar with the occurrence of a few scores exceeding one second in a reaction time experiment where the average is only 400 msec. A standard deviation of 100 msec in a set of data with a mean of 400 msec as in the present case, is a sign of fulminations better avoided.

Unfortunately the psycholinguist is not in the position of the analytical chemist who is able to vary judiciously the proportions of the particular elements brought into the the test-tube in the first place. The difficulty is that the elements are not elementary or else not independently observable. Wishing to hold constant word-frequency, the psycholinguist is faced with variations in transitional probability (semantic predictability) wholly inseparable from the particular context used. An interesting example of this in the present materials is the word "beak", which had a remarkably high semantic predictability, but more remarkably, did not occur even once in the Kučera and Francis (1967) sample of 1015232 words. In the same sample "Beauclerk" had a frequency of eight. Not surprisingly therefore, control over psycholinguistics experimantal materials is difficult!

3.2. STRATEGIES FOR RESEARCH:

In the light of the foregoing, we note the comments of Kerlinger (1973) in an introductory textbook on research design. This author points out, and we agree, that multiple causality lies in the very nature of the phenomena of interest to the behavioural scientist. The methodology employed will inevitably have to reflect this. Multivariate analysis, says Kerlinger, is by no means the "bête noir" it has been made out to be by a long tradition of hard-headed experimantal researchers. The major qualifier to these comments is that a complex analysis is blind with respect to its own interpretation (like any statistical analysis, but far more difficult to interpret correctly) . Only when guided by sophisticated theory can complex phenomena be understood through complex analysis. Perhaps some of the uneasiness we feel at reading Cutler's (1981) paper is linked to the particluar status of theory in the area of research she is reviewing, and not so much to the properties of the verbal materials involved. The proliferation of confounding variables she refers to is a clear warning for the theorist to keep abreast.

Ades' (1981) dissatisfaction with the state of cognitive experimental research arises partly from the lack of theoretical coherence in the field. In the case of phoneme monitoring research, such a lack is painfully obvious, as evidenced by the widely conflicting results and interpretations, some of which we have reviewed already. What is needed above all, is an integrated theory of cognitive functioning. A move in the right direction has been the readiness to add a subset of conscious processes to the more conventional flow-diagrams (e.g. Broadbent , 1971) that assumed only the minimal components neccessary. The Shiffrin and Schneider (1977) model is a more general structure within which "mini-models" such as the flow-diagram in Dell and Newman (1980) should be located. Figure 2-2 above gives some indication of how this could be done. It locates phoneme monitoring processes (e.g. target search and matching) within a more general model of cognition. Automatic perceptual and linguistic processes are shown in relation to the consciously controlled processes required for target monitoring.

Some aspects of the model were tested by means of multiple regression analysis. The model for reaction time scores (section 2.2.5.3) included variables representing perceptual effects (TARGSYLL) and linguistic effects (Task and COMPLETN) . The effects associated with attention were assessed on the basis of the interactions between these variables. The precise significance of the variables was not always clear, and their application to the reaction-time data was sometimes based on questionable assumptions, but it is claimed that even if the results obtained in this way were unreliable, at least the general approach to the problem of understanding cognitive processes in the phoneme monitoring task is worthy of consideration.

Further research will have to address directly the question of what TARGSYLL really measured. A separate study should be made of the semanically appropriate versus inappropriate target words using ASCM. A major problem with the present study is that too many changes were made from the original Dell ane Newman (1980) design, so that it is impossible to draw unequivocal conclusions bearing directly on their results. Nevertheless, this study was overtly exploratory, and it cannot be denied that a rich amount of information has been gathered for future use.

REFERENCES

Ades, T. Time for a purge. *Cognition*, 1981, *10*, 7-15.

Blank, M.A., Pisoni, D.B. and McClasky, C.L. Effects of target monitoring on understanding fluent speech. *Perception and Psychophysics*, 1981, *29*, 383-388.

Broadbent, D.E. *Decision and stress*. N.Y.: John Wiley, 1971.

Chase, W.G. Elementary information processes. In W.K. Estes (ed.) *Handbook of Learning and Information processing*, Vol.5. Hillsdale, N.J.: L.E.A., 1978.

Chomsky, N. and Halle, M. *The sound pattern of English*. N.Y.: Harper and Row, 1968.

Clark, M.M. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behaviour*, 1972, *12*, 335-339.

Cohen, J. and Cohen, P. *Applied multiple regression/correlation analysis for the behavioural sciences*. Hillsdale, N.J.: Lawrence Erlbaum, 1975.

Coleman, E.B. Generalisation effects vs. random effects: Is σ^2_{TL} a source of Type 1 or Type 2 error? *Journal of Verbal Learning and Verbal Behaviour*, 1979, *18*, 243-256.

Cutler, A. Making up materials is a confounded nuisance, or: Will we be able to run any psycholinguistic experiments at all in 1990? *Cognition*, 1981, *10*, 65-70.

Dell, G.S. and Newman, J.E. Detecting phonemes in fluent speech. *Journal of Verbal Learning and Verbal Behaviour*, 1980, *19*, 608-623

Donders, F.C. On the speed of mental processes (1865). Acta Psychologica, 1969, 30, 412-431.

Edwards, A. Experimental design in psychological research. N.Y.: Rinehart and Co., 1951.

Eimas, P.D. and Corbit, J.D. Selective adaptation of linguistic feature detectors. Cognitive Psychology, 1973, 4, 99-109.

Estes, W.K. The locus of inferential and perceptual processes in letter identification. Journal Exp. Psychology : General, 1975, 104 (2), 122-145.

Estes, W.K. Handbook of learning and cognitive processes, Vol.1-6. Hillsdale, N.J.: Lawrence Erlbaum, 1978.

Fischler, I. and Bloom, P.A. Automatic an attentional processes in the effects of sentence contexts on word recognition. Journal of Verbal Learning and Verbal Behaviour, 1979, 18, 1-20.

Festinger, L., Ono, H., Burnham, C.A. and Bamber, D. Efference and the conscious experience of perception. Journal Exp. Psychol. Monographs, 1967, No.4

Fodor, J.A. The mind-body problem. Scientific American, 1981, 244 (1), 124-132.

Forster, K.I. and Dickinson, R.G. More on the language-as-fixed-effect fallacy: Monte Carlo estimates of error rates for F_1, F_2, F' and min F'. Journal of Verbal Learning and Verbal Behaviour, 1976, 15, 135, 142.

Foss, D.J. Decision processes during sentence comprehension: effects of lexical item difficulty and position upon decision times. Journal of Verbal Learning and Verbal Behaviour, 1969, 8, 457-462.

Foss, D.J. and Blank, M.S. Identifying the speech codes. Cognitive Psychology, 1981, 12, 1-31.

Foss, D.J. and Lynch, R.H. Decision processes during sentence comprehension: Effects of surface structure on decision times. Perception and Psychophysics, 1969, 5, 145-148.

Foss, D.J. and Swinney, D.A. On the psychological reality of the phoneme: Perception, identification and consciousness. Journal of Verbal Learning and Verbal Behaviour, 1973, 12, 246-257

Grosjean, F. Spoken word recognition and the gating paradigm. Perception and Psychophysics, 1980, 28, 267-283

Johnston, J.C. and McClelland, J.L. Perception of letters in words: seek not and ye shall find. Science, 1974, 184, 1192-94.

Johnston, J.C. and McClelland, J.L. A hierarchical model of word identification. Journal of Verbal Learning and Verbal Behaviour, 1980, 19, 503-524.

Kerlinger, F.V. and Pedhazur, E.J. Multiple regression in behavioural research. New York: Holt, Rinehart and Winston, 1973.

Kirk, R.E. Experimental design: Procedures for the behavioural sciences. Belmont, California: Brooks/Cole, 1968.

Konorski, J. Integrative activity of the brain. Chicago: Univ. Chicago Press, 1967.

Kučera, H. and Francis, W.N. Computational analysis of present-day American English. Providence, Rhode Island: Brown Univ Press, 1967.

La Berge, D. and Samuels, S.J. Towards a theory of automatic information processing in reading. Cognitive Psychology, 1974, 6, 293-323.

Lachman, R., Lachman, J.L. and Butterfield, E.C. Cognitive psychology and information processing: An introduction. Hillsdale, N.J.: Lawrence Ertbaum, 1979.

Lea, W.A. An approach to syntactic recognition without phonemics. IEEE Transactions Audio and Electroacoustics., 1973, Au-21 (3), 249-258.

Lehiste, I. Supresegmentals. In N. Lass (Ed.) Contemporary issues in experimental phonetics. N.Y.: Academic Press, 1976, pp.243-293.

Marcus, S.M. Contex-sensitive coding in speech perception. Instituut voor Perceptie Onderzoek. Annual Progress report, 1979, 14.

Marslen-Wilson, W.D. Linguistic structure and speech shadowing at very short latencies. Nature, 1973, 244, 522-523

Marslen-Wilson, W.D. and Tyler, L.K. The temporal structure of spoken language understanding. Cognition, 1980, 8, 1-71.

Marslen-Wilson, W.D. and Tyler, L.K. Central processes in speech understanding. To appear in Philosophical transactions of the Royal Society, 1981, B.

McNeill, D. and Lindig, K. The perceptual reality of phonemes, syllables, words and sentences. Journal of Verbal Learning and Verbal Behaviour. 1973, 12, 419-430.

Morton, J. The effects of context on visual duration threshold for words. Brit. Journal of Psychology, 1964, 55, 165-180.

Morton, J. Interaction of information in word recognition. Psychological Review, 1969, 76, 165-178.

Morton, J. and Long, J. Effect of word-transitional probability on phoneme identification. Journal of Verbal Learning and Verbal Behaviour, 1976, 15, 43-51.

Neisser, U. Cognitive psychology. N.Y.: Appleton-Century-Crofts, 1967.

Neisser, U. Cognition and reality. San Francisco: W.H. Freeman, 1976.

Newell, A. and Simon, H.A. Human problem solving. Englewood Cliffs, N.J.: Prentice Hall, 1972.

Newman, J.E. and Dell, G.S. The phonological nature of phoneme monitoring: A critique of some ambiguity studies. Journal of Verbal Learning and Verbal Behaviour, 1978, 17, 359-374.

Nooteboom, S.G. and Truin, P.G.M. Word recognition from fragments of spoken words by native and non-native listeners. IPO Annual Progress Report, 1980, 15.

Onifer, W. and Swinney, D.A. Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. Memory and Cognition, 1981, 9, 225-236.

Overall, J.E. and Spiegel, D.K. Concerning least squares analysis of experimental data. Psychological Bulletin, 1969, 72, 311-322.

Paap, K.R. and Newsome, S.L. A perceptual-confusion account of the WSE in the target search paradigm. Perception and Psychophysics, 1980, 27, 444-456.

Palmer, S.E. The effects of contextual scenes on the identification of objects. Memory and Cognition, 1975, 3, 519-526.

Pisoni, D.B. Auditory and phonetic memory codes in the discrimination of consonants and vowels. Perception and Psychophysics, 1978, 13, 253-260.

Pisoni, D.B. and Lazarus, J.H. Categorical and non-categorical modes of speech perception along the voicing continuum. Scientific American, 1974, 55, 328-333.

Pollack, I. and Pickett, J.M. Intelligibility of exerpts from fluent speech: Auditory vs. structural context. Journal of Verbal Learning and Verbal Behaviour, 1964, 3, 79-84.

Polyani, M. Personal knowledge: towards a post-critical philosophy. N.Y.: Harper and Row, 1964,

Posner, M.I. and Klein, R.M. On the functions of consciousness. In S. Kornblum (Ed.), Attention and performance IV. N.Y.: Academic Press, 1973.

Posner, M.I. and Mitchell, R.F. (cited in Posner and Rogers). Chronometric analysis of classification. Psychological Review, 1967, 74, 392-409.

Posner, M.I. and Rogers, M.G.K. Chronometric analysis of abstraction and recognition. In W.K. Estes (Ed.), Handbook of Learning and Cognitive Processes. Vol.5. Hillsdale, N.J.: Lawrence Erlbaum, 1978.

Posner, M.I. and Snyder, C.R.R. Facilitation and inhibition in the processing of signals. pp.669-682. In P.M.A. Rabbitt and S. Dornic (Eds.), Attention and Performance II. London: Academic Press, 1975.

Pritchard, W.S. Psychophysiology of P300. Psychological Bulletin, 1981, 89, 506-

Reicher, G.M. Perceptual recognition as a function of meaningfulness of stimulus material. Journal of Experimental Psychology, 1969, 81, 274-280.

Savin, H.B. and Bever, T.G. The nonperceptual reality of the phoneme. Journal of Verbal Learning and Verbal Behaviour, 1970, 9, 295-302.

Schneider, W. and Shiffrin, R.M. Controlled and automatic human information processing: I. Detection, search and attention. Psychological Review, 1977, 84, 1-66.

Shiffrin, R.M. and Schneider, W. Controlled and automatic human information processing: II. Perceptual learning, automatic attenting, and a general theory. Psychological Review, 1977, 84, 127-190.

Simon, H.A. Information processing models of cognition. Annual Review of Psychology, 1979, 30, 363-396.

Sternberg, R.J. Component processes in analogical reasoning. Psychological Review, 1977, 84, 353-378.

Sternberg, S. High-speed scanning in human memory. Science, 1966, 153, 652-654.

Sternberg, S. The discovery of processing stages: Extension of Donders' method. Acta Psychologica, 1968, 30, 276-315.

Stevens, K.N. Towards a model for speech recognition. Journal of the Acoustical Soc. of America, 1960, 32, 47-55.

Studdert-Kennedy, M. Speech perception. In: N. Lass (Ed.), Contemporary issues in experimental phonetics. N.Y.: Academic Press, 1976, pp. 243-293

Taylor, J.G. The behavioural basis of perception. New Haven, Conn.: Yale University Press, 1962.

Warren, R.M. Perceptual restoration of missing speech sounds. Science, 1970, 167, 392-393.

Wickelgren, W.A. Context-sensitive coding, associative memory and serial order in (speech) behaviour. Psychological Review, 1969, 76, 1-15.

APPENDIX

List of the sentence frames used in the experiments.

1. The surfers drove to a secret/beautiful beach/beacon.
2. While her toast was still warm Susan/Betty buttered/bubbled it with dairy spread.
3. When Henry dropped it the saucer/bowl broke/broached on the floor.
4. The mosquito left a stinging/painful bite/bind on her arm.
5. Musical instruments like the trumpet are made of solid/pure brass/branch
6. When Mary kissed him Sam/Bob blushed/blundered and turned away.
7. The dynamite in the sack/bag blew up/blueprinted the entrance.
8. Strong forces make steel/brass bend/bench very easily.
9. After filling the tub Sally/Barbara bathed/baked in the water.
10. Moving high in the sky the swift/pretty bird/birth was easily seen.
11. After pouring water over the kneeling woman the saint/bishop blessed/blended her and directed her to pray.
12. At the lending library the student/banker borrowed/bottled the damaged text.
13. The sparrow landed on the cedar/pine branch/brass and rested awhile.
14. On her wrist Sarah's/Barbara's bracelet/breaker looked rather strange.
15. The juicy worm caught in the swallow's/bird's beak/beach provided a nourishing meal.
16. The very hot oven made the sauce/bread burn/birth.

APPENDIX

LIST OF SENTENCES COMPRISING ONE EXPERIMENTAL TAPE.

CATCH TRIALS ARE MARKED WITH AN ASTERISK.

TARGET PHONEMES AND NON - TARGETS IN TARGET POSITIONS ARE MARKED.

1. The surfers drove to a beautiful beacon.
2. While her toast was still warm Betty buttered it with dairy spread.
3. * Sally loved the exciting thriller.
4. When Henry dropped it the saucer broached on the floor.
5. * Brave ships are always very young.
6. The mosquito left a painful bite.
7. Musical instruments like the trumpet are made of pure branch.
8. * The rich buyers were impressed with the estate.
9. * Many people prefer to ride a motor car.
10. When Mary kissed him Bob blushed and turned away.
11. * Happily he took another draught of beer.
12. The dynamite in the sack blueprinted the entrance.
13. * There are many things in the world we do not understand.
14. Strong forces make steel bend very easily.
15. * The happy bells chimed in the frosty morning.
16. After filling the tub Barry baked in the water.
17. * The polite gentleman greeted the old lady.
18. * The large terrier was digging a hole in the street.
19. * Beverly disguised her books on the table.
20. Moving high in the sky the swift bird was easily seen.
21. After pouring water over the kneeling woman the bishop blended her and directed her to pray.
22. * The star outfielder used his own bayonet.
23. * Some brown books are very exciting to read.
24. * When the car appeared it was awfully big.
25. At the lending library the banker borrowed the damaged text.
26. The sparrow landed on the cedar brass and rested awhile.
27. * David grew up to be a serious parent.
28. * He was surprised when the plant slowly died.
29. On her wrist Saraha's breaker looked rather strange.
30. The juicy worm caught in the swallow's beak provided a nourishing meal.
31. * The sailor went to sea in a beautiful ship.
32. * The hungry peasant chewed the brown bread.
33. The very hot oven made the bread burn.

APPENDIX :

The Dell and Newman (1980) sentence materials.

The critical and target words associated with the four experimental conditions are given in italics, the /p/-initial critical word first, followed by, in order, the /s/-initial critical word, the highly predictable target word, and the unpredictable target word.

1. The surfers drove to a (*private/secret beach/bay*) to try out the waves.
2. When Mary kissed him (*Paul/Sam blushed/blinked*) and turned away.
3. The fisherman's lead sinker dropped to the (*pond's/stream's bottom/bank*) and was lost.
4. After filling the tub (*Peter/Sally bathed/basked*) in the water.
5. The sparrow landed on the (*pine/cedar branch/bed*) and rested awhile.
6. For winter weather the shoe department featured (*plastic/suede boots/belts*) as well as leather ones.
7. Musical instruments such as the trumpet are made of (*pure/solid brass/bronze*).
8. While her toast was still warm (*Polly/Susan buttered/brushed*) it with dairy spread.
9. After pouring water over the kneeling woman the (*priest/saint blessed/bound*) her and directed her to pray.
10. The muscular Mr. Canada contestants had (*perfect/strong biceps/blood*) which indicated good health.
11. Moving high in the sky the (*pretty/swift bird/ball*) was easily seen.
12. The mosquito left a (*painful/stinging bite/blister*) on her arm.
13. On her wrist (*Paula's/Sarah's bracelet/brooch*) looked rather strange.
14. On his shirt the man sewed a (*plain/single button/badge*).
15. When Henry dropped it the (*platter/saucer broke/bounced*) on the floor.
16. The red cape placed over his horns enfuriated the (*proud/savage bull/buck*).
17. Strong forces make (*platinum/steel bend/buckle*) very easily.
18. The juicy worm caught in the (*pigeon's/swallow's beak/belly*) provided a nourishing meal.
19. The cook mixed the flour and eggs in the (*purple/silver bowl/basket*).
20. Johnny is a very (*polite/silly boy/baby*) which is unusual for someone his age.
21. The very long and tedious (*play/sermon bored/baffled*) its audience.
22. That well-constructed house is the one that (*Patrick/Steven built/bombed*).
23. The insect with the handsome wings was a (*passing/sleeping butterfly/bee*).
24. Always looking for a scapegoat the (*press/schools blamed/blasted*) the federal government.
25. The driver asked the passengers to get off the (*parked/stalled bus/bicycle*) for their own safety.
26. The animal gnawing on the tree was a (*plump/sluggish beaver/bear*).
27. In order to hide their treasure the (*pirates/sailors buried/boarded*) it in the cave.
28. At the tavern the longshoreman enjoyed a (*pleasant/soothing beer/brandy*) and good company.
29. The dynamite in the (*pack/sack blew up/blocked*) the entrance to the castle.
30. After reading the (*picture/sad book/bill*) the whole family discussed it.
31. At the lending library the (*pupil/student borrowed/bought*) the damaged text.
32. The very hot oven made the (*pie/sauce burn/bubble*).

			RAW DATA							
CASE NUMBER	PHONEME MONITORING REACTION TIME Y-PHONEM	ACOUSTIC CUE MONITORING REACTION TIME Y-ACOUST	CRITICAL WORD LENGTH CRITLGTH	TARGET WORD LENGTH TARGLGTH	SQUARE ROOT OF CRITICAL WORD FREQUENCY CRITFREQ	SQUARE ROOT OF TARGET WORD FREQUENCY TARGFREQ	TARGET WORD SYLLABLE LENGTH TARGSYLL	ARCSIN TRANSFORMED SENTENCE COMPLETION SCORES COMPLETN	LATENCY OF SENTENCE COMPLETION RESPONSES LATENCY	LENGTH OF PAUSE BETWEEN CRIT. & TARGET WORDS. PAUSE
1	338	272	505	437	10.630	8.718	252	.586	1211	121
2	429	188	379	284	6.557	5.196	126	2.094	909	25
3	375	200	579	421	1.414	11.619	342	2.301	880	116
4	249	157	658	400	3.162	4.899	316	1.231	985	42
5	274	292	452	568	10.198	4.583	284	2.556	686	147
6	455	205	532	500	9.110	3.606	200	1.571	702	25
7	522	332	495	468	3	6.708	174	.841	1160	295
8	280	350	616	484	7.071	5.745	400	1.047	1130	84
9	415	248	421	500	3.873	5.099	305	1.231	1026	100
10	247	345	432	337	7	9.950	295	1.403	1745	105
11	460	290	463	342	5	4.583	195	1.403	1748	226
12	487	225	474	479	18.735	5.745	132	1.403	950	105
13	768	193	353	474	3.162	1	205	.586	1260	79
14	632	382	558	563	5.385	4.583	200	.586	1840	95
15	347	238	579	358	3.464	5.477	289	2.301	853	25
16	386	212	621	405	5	11.314	379	1.231	1432	226
17	266	318	447	442	12.207	8.718	274	1.911	1039	58
18	292	228	358	326	2	5.196	168	1.571	1286	53
19	433	523	405	405	5.099	11.619	305	2.556	748	95
20	482	182	516	400	6.245	4.899	311	1.403	862	25
21	403	217	258	532	9.434	4.583	300	1.403	2460	100
22	466	250	137	463	7.937	3.606	221	.841	1395	121
23	380	335	489	484	7.483	6.708	195	1.403	1472	500
24	347	227	615	411	4.583	5.745	384	1.047	1640	174
25	403	218	305	584	1.732	5.099	342	1.047	957	121
26	361	190	347	395	10.488	9.950	326	2.556	1420	63
27	631	260	387	342	5.099	4.583	184	.841	1395	5
28	357	268	516	421	4.472	5.745	142	0	0	63
29	362	198	379	516	4	1	179	1.047	667	5
30	580	268	468	589	3.162	4.583	200	1.403	1171	32
31	337	255	342	400	9.950	5.477	326	2.094	707	63
32	347	305	337	395	6.403	11.314	237	1.231	987	316
33	421	160	456	452	10.630	2.236	258	.586	1211	153
34	378	260	595	337	6.557	5.657	189	2.094	909	68
35	212	241	516	495	1.414	1.732	279	2.301	880	53
36	438	333	684	510	3.162	5	326	1.231	985	47
37	320	177	389	563	10.198	2.449	237	2.556	686	226
38	428	253	547	605	9.110	6.307	184	1.571	702	5
39	268	293	432	710	3	13	253	.841	1160	353
40	473	268	631	463	7.071	6.708	253	1.047	1130	216
41	366	177	447	468	3.873	4.472	163	1.231	1026	94
42	538	267	474	426	7	8.367	226	1.403	1745	126
43	308	198	621	411	5	4	147	1.403	1748	200
44	386	245	558	579	18.735	8.888	147	1.403	950	105
45	382	237	421	537	3.162	1.414	332	.586	1260	25
46	341	373	474	542	5.385	8.124	316	.586	1840	105
47	453	218	853	395	3.464	8.718	211	2.301	853	47
48	370	343	642	416	5	8.367	179	1.231	1432	237
49	308	217	432	526	12.207	2.236	158	1.911	1039	63
50	445	192	289	353	2	5.657	184	1.571	1286	105
51	482	200	353	579	5.099	1.732	252	2.556	748	25
52	385	367	542	563	6.245	5	211	1.403	862	111
53	412	233	305	553	9.434	2.449	237	1.403	2460	36
54	497	293	274	511	7.937	8.307	163	.841	1395	147
55	320	242	458	816	7.483	13	289	1.403	1472	647
56	406	322	584	526	4.583	6.708	232	1.047	1640	253
57	301	182	374	437	1.732	4.472	279	1.047	957	84
58	442	362	284	411	10.488	8.367	184	2.556	1420	100
59	382	358	447	458	5.099	4	142	.841	1395	132
60	588	375	600	579	4.472	8.888	168	0	1214	484
61	371	203	379	495	4	1.414	237	1.047	667	25
62	436	300	500	579	3.162	8.124	253	1.403	1171	95
63	373	295	374	421	9.950	8.718	200	2.094	707	74
64	546	325	326	379	6.403	8.367	168	1.231	987	231

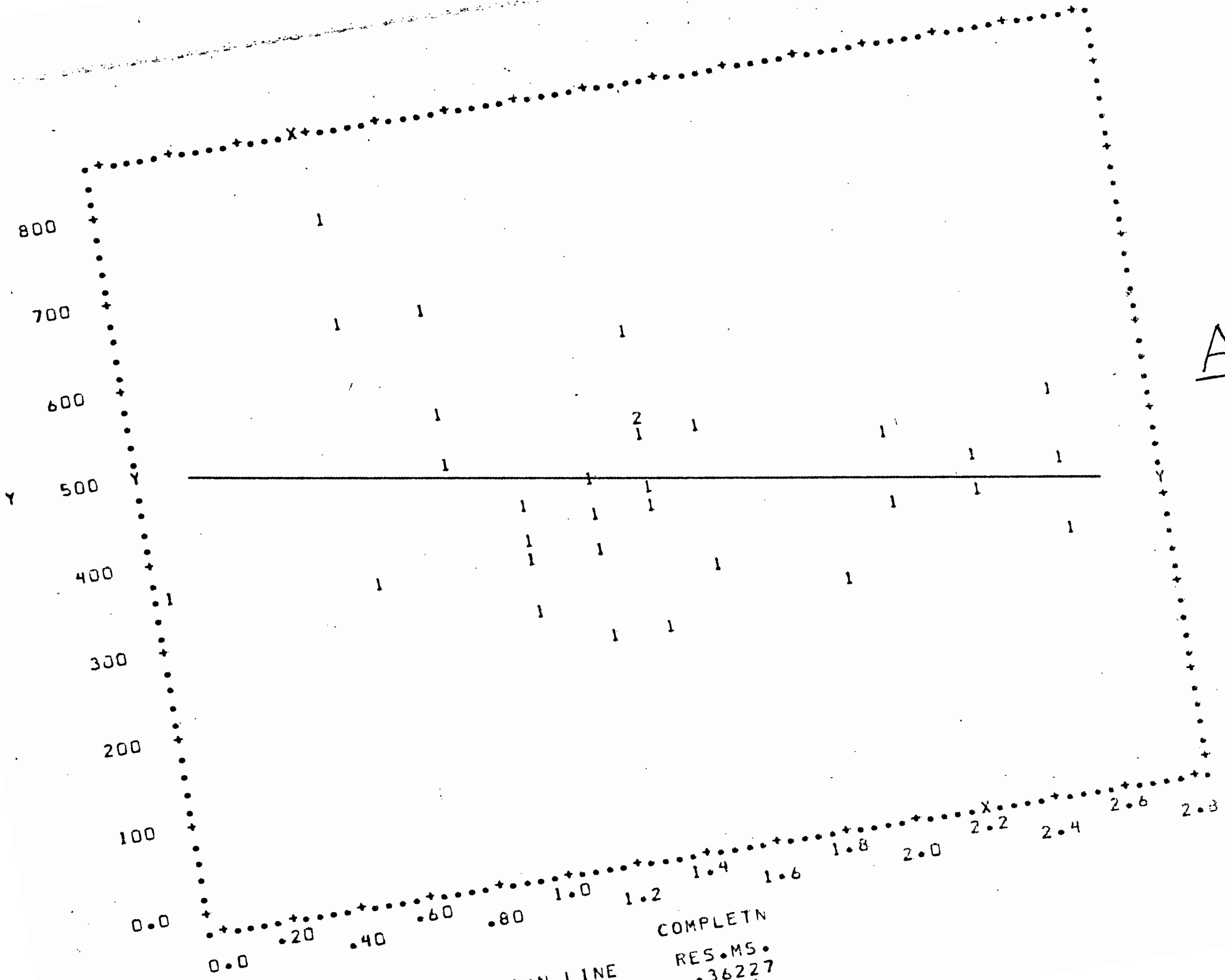
APPENDIX :

The regression of RT (Y) on sentence completion scores (c);

COMPLETN = $2 \arcsin \sqrt{c/12}$.

A : RT only to sentences with semantically appropriate target-bearing words (N = 32).

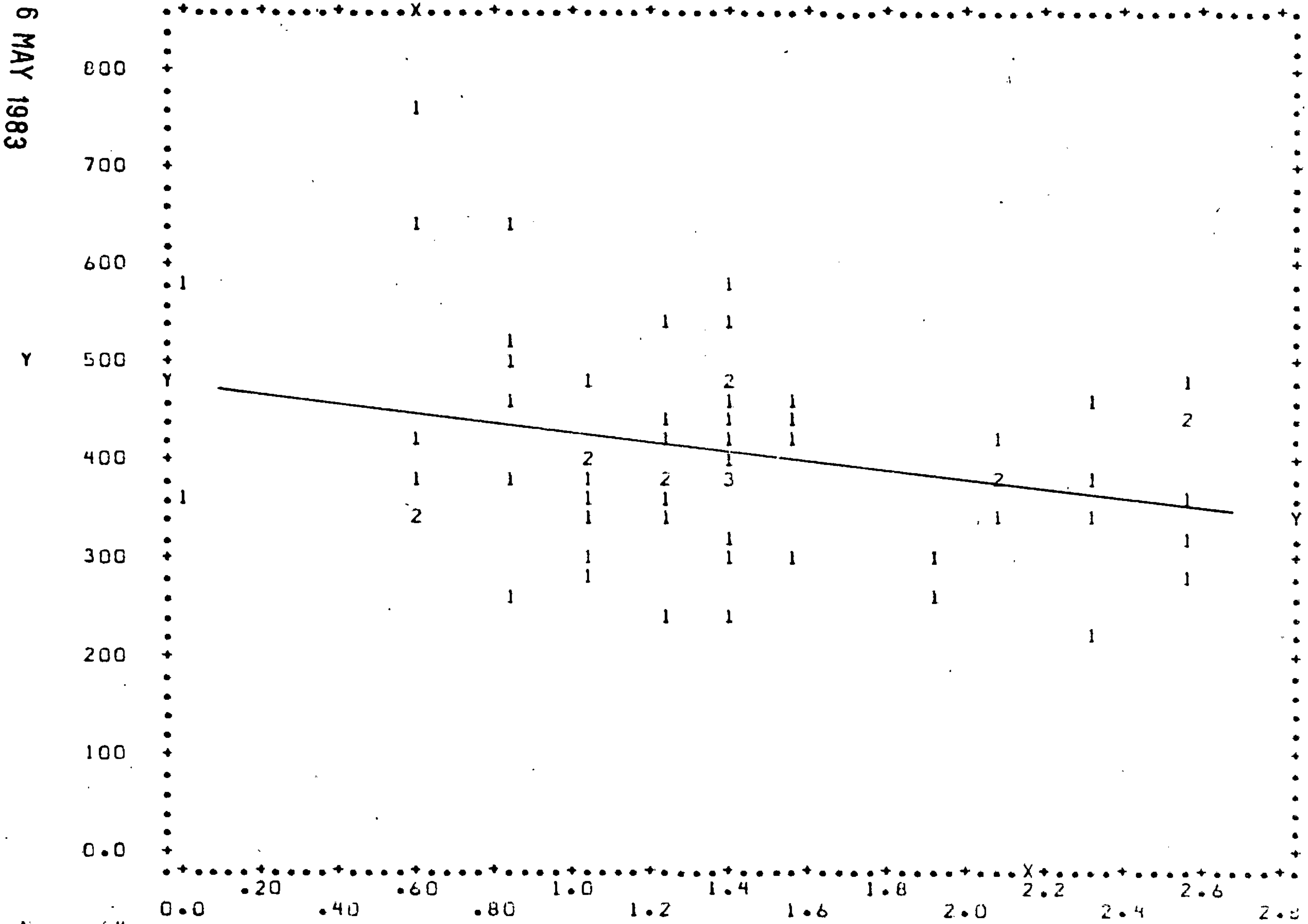
B : RT to sentences with both semantically appropriate and inappropriate target-bearing words (N = 64).



A

```
                                                   RES.MS.
                        REGRESSION LINE            .36227
N=     32                     *Y+  2.1629           12921.
   .3527                           502.82
```

26 MAY 1983



B

```
N=    64
COR=-.2873                                            COMPLETN

    MEAN      ST.DEV.     REGRESSION LINE      RES.MS.
X   1.3977     .62773    X=-.00177*Y+ 2.1141    .36734
Y   404.48    101.83    Y=-46.611*X+ 469.63    9667.2

VARIABLE    4  COMPLETN VERSUS VARIABLE  1  Y
```