ECOLOGICAL APPROACHES

TO

SPEECH PERCEPTION

ROBERT BURRELL BYRNES

DOCTOR OF PHILOSOPHY

UNIVERSITY OF EDINBURGH

1982

I hereby declare that this thesis is my own work, having been completed within the normal terms of reference and supervision in the Faculty of Social Sciences, University of Edinburgh.

Robert B Byrnes

ELISABETH

ILONA BRIDGET

CHRISTIAN THOMAS

ABSTRACT

A literature review demonstrates that very general scientific presuppositions which Whitehead regarded as instances of the fallacy of misplaced concreteness and Bohm labelled 'fragmentation' characterise current research in speech perception. It is then argued that the following two hypotheses allow these presuppositions to be tested:-

1 For every exclusively auditory experiment in speech perception, an attempted replication to the audio-visual case can be conducted which will result in a failure to replicate.

2 If an effect that is obtained through dubbing can also be produced with at least contrinsically related optical and acoustic signals, an experiment can be conducted which will result in a failure to replicate from dubbing to the more naturalistic case.

A series of twelve experiments provides strong evidence to support both of the hypotheses. This is taken to establish that future speech research must orientate itself relative to naturalistic speech perception and not the dimensions of Physics. Some implications of this reorientation are discussed.

ACKNOWLEDGEMENTS

"The realms of life are many. For each one, special sciences develop. But life itself is a unity, and the more deeply the sciences try to penetrate into their separate realms, the more they withdraw themselves from the vision of the world as a living whole. There must be a knowledge which seeks in the separate sciences the elements for leading man back once more to the fullness of life. ...

One must be able to confront an idea and experience it; otherwise one will fall into its bondage." (Steiner, 1962).

If this thesis is at all successful in confronting the ideas of current speech research and directing attention to the fullness of the speech event, that is due to the richness of the concepts and thought-pictures which Rudolf Steiner (1861-1925) formulated. All of what is reported constitutes a small step towards experiencing the significance of several of his ideas. The scope of these ideas is indicated in the following words:-

"Monism regards a science that limits itself to a description of percepts without penetrating to their ideal complements as incomplete. But it regards as equally incomplete all abstract concepts that do not find their complements in percepts, and that fit nowhere into the conceptual network that embraces the whole observable world."

I am deeply grateful to J J Gibson (1904-1979) for his lived demonstration of the fruitfulness of a down-to-earth, open-minded commitment to monism.

My warmest thanks go to Mrs D M Klages who introduced me to Steiner's thought, to David Lee who introduced me to Gibson's and to Ed Reed who helped me to appreciate its philosophical importance.

Dr T Myers has devoted much time to helpful discussions which clarified the relationship between monism and current (denied) dualism. So much of what we discussed lies beyond the scope of a thesis.

Mr J Cuthbert has provided much help with filming and explanations of complicated equipment. Mr E Lucey also helped with some filming and kindly allowed me to use equipment belonging to the Film Unit.

CONTENTS

CHAPTER 1: WHOLENESS AND FRAGMENTATION IN SPEECH RESEARCH

#1.1 SPEECH RESEARCH AND THE SCIENTIFIC METHOD

Research into speech perception, as one of many current disciplines, shares the premisses or presuppositions of the prevailing scientific outlook. As Cornford (1950) noted, these premisses 'are not mentioned simply because they are too obvious to be worth mentioning.' A number of thinkers have investigated the "too obvious" of the present age in terms which can be helpful in speech research. Some of the fundamental insights of two of these thinkers, Whitehead and Bohm, will be sketched very briefly in general terms and an attempt made to demonstrate their fruitfulness for speech research. In the final chapter, attention will be paid to a further, more pervasive presupposition, the recognition of which potentially offers even deeper insights than those given by Whitehead and Bohm.

#1.2 SIMPLE LOCATION AND MISPLACED CONCRETENESS

Whitehead (1926) described well a number of the premisses which underlie 'the whole philosophy of nature during the modern period.' These premisses which 'appear so obvious that people do not know that they are assuming them' became established during the seventeenth century, and as Whitehead observes, the resulting

> "Guiding principle of scientific studies ... is still reigning. Every university in the world organises itself in accordance with it ... It is not only reigning, but it is without rival."

One assumption is that matter has the property of 'simple location'. That is,

> "Material can be said to be here in space and here in time, or here in space-time, in a perfectly definite sense which does

> not require for its explanation any reference to other regions of space-time."

This leads to the conclusion that 'the world is a succession of instantaneous configurations of matter', and this is the 'famous mechanistic theory of nature' which has prevailed since the seventeenth century. As Whitehead shows, the concept of simple location is merely the 'accidental error of mistaking the abstract for the concrete;' that is, it follows from regarding the abstract logical constructions of scientific thinking as concrete realities. Thus, it is an instance of what he calls the 'fallacy of misplaced concreteness.'

A second assumption, another instance of the fallacy of misplaced concreteness, is to be found in the 'correlative categories of Substance and quality.' This sounds very old-fashioned, but is very relevant. To see this, it is not necessary to rehearse the notions of primary and secondary qualities or to refer to competing physical and psychological theories, for:

> "Whatever theory you choose, there is no light or colour as a fact in external nature. There is merely motion of material. Again, when light enters your eyes and falls on the retina, there is merely motion of material."

The one outcome of this view, 'however you disguise it', is that

> "Nature is a dull affair, soundless, scentless, colourless; merely the hurrying of material, endlessly, aimlessly."

The second assumption is familiar from contemporary psychology; the first, although less familiar, is more pervasive. For present purposes, the power of Whitehead's rejection of these assumptions is not of primary importance; what matters is that he has shown them to be assumptions. Simply recognising an assumption as such helps in the attempt to examine it afresh.

#1.3 WHOLENESS AND FRAGMENTATION

Bohm's (1980) starting point, which is very similar to Whitehead's views, is formulated in terms which are applicable to speech research. He holds that,

> "Our theories are to be regarded as ways of looking at the world as a whole (ie world views) rather than as 'absolutely true knowledge of how things are' or as a steady approach towards the latter."

Noting that every form of theoretical insight necessarily introduces essential differences and distinctions, he proceeds to another formulation of the fallacy of misplaced concreteness,

> "If we regard our theories as 'direct descriptions of reality as it is', then we will inevitably treat these differences and distinctions as divisions, implying separate existence of the various elementary terms appearing in the theory."

This is true of every theory, but it applies especially to the content of the atomic theory, which was

> "Especially conducive to fragmentation, for it was implicit in this content that the entire world of nature, along with the human being, including his brain, his nervous system, his mind etc., could in principle be understood completely in terms of structure and functions of aggregates of separately existent atoms."

As a physicist, Bohm asserts that both relativity and quantum theory

> "Imply the need to look on the world as an undivided whole, in which all parts of the universe, including the observer and his instruments, merge and unite in one totality. In this totality the atomistic form of insight is a simplification and abstraction, valid only in some limited context."

> "The new form of insight can perhaps best be called Undivided Wholeness in Flowing Movement. This view implies that flow is, in some sense, prior to ... the 'things' that can be seen to form and dissolve in this flow."

Belief in the fragmentary atomistic approach to reality is strongest in

> "The study of life and mind, which are just the fields in which formative cause acting in undivided and flowing movement is most evident to experience and observation;"

and this has a strangely paradoxical effect:

> "Since, in the first instance, fragmentation is an attempt to extend the analysis of the world into separate parts beyond the domain in which to do this is appropriate, it is in effect an attempt to divide what is really indivisible. In the next step such an attempt will lead us also to unite what is not really unitable."

In what follows, Bohm's insights will not be treated as absolutely true knowledge of how things are, but as a means to help identify presuppositions in speech research.

#1.4 DIVIDING AND UNITING IN SPEECH RESEARCH

#1.4.1 SUBSTANCE AND QUALITY

Two straightforward passages, representative of many, reveal that the correlative categories of Substance and quality are common currency. As Whitehead observed, it is indeed one of the 'most natural ideas of the human mind.'

> "The first step in the auditory processing of speech is the conversion of the acoustic speech signal into patterns of activity in the neurons of the auditory nerve" (Young and Sachs, 1981)

> "The speech stimulus consists of changes in the atmospheric pressure at the ear of the listener" (Massaro, 1981).

The extensive application of speech synthesizers is an obvious consequence of this view, and equally obviously an instance of fragmentation. To be acceptable, a synthesizer need only produce the desired changes in atmospheric pressure at the ear of the listener.

#1.4.2 DIVIDING AND UNITING

Even when attention is restricted to the acoustic signal, synthesized speech provides many examples of dividing what is not really divisible. Several examples are:-

- 'simplified' two formant speech (Liberman et al, 1967)
- deleting transients (Liberman et al, 1952)
- isolating chirps (Halwes, 1969)
- deleting bursts (Liberman et al, 1967).

Other examples which involve manipulations of natural speech include:-

- isolating bursts (Schatz, 1954)
- isolating burst and aspiration (Winitz et al, 1972)
- excising 'single words' from connected speech (Pollack and Pickett, 1964)
- separating syllables (Rudnicky and Cole, 1978)
- onset erasures (Fischer-Jorgensen, 1954)
- Phonemic Restoration Effect (Warren, 1970).

Uniting what is not really unitable is obvious in the following techniques:-

- transposing excised segments of phonemes (Dorman et al, 1977)
- altering intersyllabic closure (Rudnicky and Cole, 1978)
- creating tape loops (Warren, 1968)
- compressed speech (Huggins, 1972)
- conflicting cues (Fitch et al, 1981).

It is perhaps appropriate to state explicitly that every instance of dividing what is not really divisible necessarily involves uniting what is not really unitable. For example, a burst which is divided from an acoustic signal is thereby united (in a manner which defies articulation) with adjacent zero energy. Such dividing and uniting is an application of the concept of simple location.

#1.4.3 AUDIO-VISUAL EXPERIMENTATION

#1.4.3.1 WORD RECOGNITION

Following on from O'Neill (1951), an extensive literature has developed which demonstrates that when isolated words are presented in noise, combined audio-visual word recognition is superior to recognition through audition or vision alone (Erber, 1975). Corresponding studies have shown that auditory-visual recognition of filtered speech is similarly superior (Sanders and Goodrich, 1971). It has also often been found that auditory-visual word recognition of hearing impaired listeners is superior to purely auditory recognition (Erber, 1975). Further, Dodd (1979b) reported that detection of mispronunciations in masked speech is improved with audiovisual presentation. These studies typically entail no manipulation of the filmed stimuli beyond masking or filtering of the acoustic signal. Disregarding for the moment the effects of recorded presentation, the division here involves the experimenter rather than the acoustic and optical aspects of the stimuli. Although it is regularly reported that audio-visual presentation improves word recognition, it does not appear to have been reported that the listener's conscious perception of the words has been markedly affected.

Summerfield (1979) modified this technique by employing passages of connected prose to mask test words which were embedded in sentences. A number of further conditions in which only the speaker's lips, only four dots on the speaker's lips or a display of the amplitude envelope of the sentence was visible were also investigated; S/N ratio was adjusted in pilots to a level at which naive listeners could identify about one word in five in the auditory-only condition. It was found that when the

speaker's face was visible, performance improved in a manner comparable to the earlier studies; display of only the speaker's lips resulted in a small decrease in performance compared to the condition in which the full face was visible, Ss attributed this to the absence of teeth and tongue movements, not to the loss of the facial frame; the four dots improved performance only marginally, but it is likely that a more favourable placement of the dots would yield a greater improvement (in the present arrangement, the speaker's mouth never seemed to close); the amplitude-modulated display had no effect on word recognition scores.

It is interesting to compare how Erber (1975) and Summerfield discuss their essentially similar results. Erber considers vision as a source of cues which are useful when the S/N ratio is less than optimal; Summerfield sees the event in more unified terms:

> "Optical concomitants of articulation appear to specify linguistic information rather than merely focussing attention when there is a competing background."

In a further comment, Summerfield expresses the important observation which appears to have eluded all previous researchers in the field. It is not just that word recognition is improved by audio-visual presentation,

> "Phenomenally, the effect is captured by the comments of several subjects who observed that the test sentences sounded clearer when they could see the speaker's lips."

Thus, the form of division which characterised the earlier research in this field has been overcome. The powerful influence of vision on what listeners experience themselves as 'hearing' had, however, already been reported in another context.

#1.4.3.2 DUBBING

McGurk and MacDonald (1976) appear to have been the first researchers to

report that conscious speech perception is influenced by vision. They found that when an utterance of [baba] is dubbed onto the lip movements for [gaga], normal hearing adults report hearing [dada]; with the reversed dubbing process, the majority reported hearing [bagba] or [gaba]. In a later paper (McGurk and MacDonald, 1977), they extended the dubbing combinations and established that the effect is quite general.

Although the statistical analyses yield highly significant effects, McGurk and MacDonald add:

> "Alone, however, the data fail to testify to the powerful nature of the illusions. We, ourselves, have experienced these effects on many hundreds of trials; they do not habituate over time, despite objective knowledge of the illusion involved. By merely closing the eyes, a previously heard [da] becomes [ba] only to revert to [da] when the eyes are open again."

Liberman (1981) confirms this description,

> "My percept was unified in the important sense that I could not have decided by introspective analysis that part was visual in origin and part auditory."

Thus, the 'hearing lips and seeing voices' illusion led McGurk and MacDonald to recognise what had been overlooked in the word recognition studies:

> "Contemporary auditory-based theories of speech perception are inadequate ... a role for vision (that is, perceived lip movements) in the perception of speech by normally hearing people is called for."

In the context of the present discussion, it is clear that dubbing combines the usual dividing of the acoustic aspect of speech production from its optical aspect with the novel feature of uniting it with another, typically conflicting signal. This technique has been applied by other researchers.

Summerfield (1979) combined McGurk and MacDonald's technique with a standard phoneme identification task. He constructed a triangular

arrangement of three 11-member continua of VCV syllables, [aba] to [ada] to [aga] and back to [aba], and dubbed all 11 members of each continuum onto lip movements corresponding to each of its endpoints. Ss were then required to identify the members of each continuum in each of three conditions: two audio-visual conditions as described above and a third purely auditory condition. The response set contained nine members such as [abda], [aTHa], [ava], [aba], being the set of percepts experienced by E and colleagues in informal trials. Although the pattern of data was very complicated, a marked influence of vision was apparent. Two general effects were:-

1 'stimuli (which were) ambiguous in the no-video condition were assimilated into the response category of the phonetic event instantiated in the visual display,'

2 'a bilabial was only perceived when lip closure was specified optically; and, in general, when lip closure was specified optically, a bilabial was perceived (often in a cluster), regardless of what consonant was specified acoustically.'

Summerfield rejects McGurk and MacDonald's (1977) manner-place hypothesis, according to which in face-to-face communication between normally hearing people, manner of articulation of consonantal utterances is picked up by ear, place of articulation by eye and their integration occurs at the level of phonetic features. As Summerfield remarks, all features can be perceived auditorily, so this account demands a special segregated mode of processing for audio-visual speech perception, and thus appears to be extremely implausible. Summerfield argues that the common metric in which the integration of the visual and acoustic information takes place should be articulatory dynamics, not phonetic features. This is indicated by the simple, but very important

observation that,

> "Normally, a talker's articulatory apparatus imposes structure on both light and sound, but the experience of watching and listening is of perceiving one speaker and one message."

Roberts and Summerfield (1981) also applied McGurk and MacDonald's technique to attempt to determine whether the familiar adaptation effect (Eimas and Corbit, 1973) is auditory or phonetic in nature. They constructed a [ba] - [da] continuum of synthetic stimuli and employed several different adaptors - Vb, Vd, Ab, Ad, AbVb, AdVd & AbVg - and reported that visual adaptors produced no adaptation effects, and that in all cases of audio-visual adaptation, the adaptation effect was equal to the effect produced by the isolated auditory component. From this, they concluded that audio-visual adaptation is exclusively auditory and further that adaptation is auditory and not phonetic.

This study will be referred to again in Chapter 3 because it provides an excellent opportunity to test predictions which arise within an wholistic approach. In passing, it may be noted that Samuel (1982) found that a prototypical phoneme produced greater adaptation effects than adaptors from the same phonetic class which lay further from the phonetic boundary. This suggests that adaptation is not exclusively auditory.

#1.4.3.3 ASYNCHRONY

Dodd (1977) and Campbell and Dodd (1980) have employed techniques which produced asynchrony between the acoustic and optical speech signals and found that even in asynchronous conditions, vision aids word recognition and recall. As no reference has been made to conscious perception in these studies, they will not be discussed in detail. Creation of asynchrony is clearly another instance of dividing and uniting. Dodd

(1979a) also found that 10- to 16-week-old infants attended more to in-synchrony than out-of-synchrony nursery rhymes, and took this to indicate that young infants are

> "Aware of the congruence between lip movements and speech sounds."

#1.4.3.4 RECENCY AND SUFFIX EFFECTS

These effects will be mentioned briefly because it has been found that effects which were believed to be exclusively auditory also exist in the visual mode when speech is involved. Klima and Bellugi (1979) report that STM in sign language does yield recency effects; Spoehr and Corin (1978) found that with audio-visual presentation, a silently articulated suffix has the same effect on recall of the last item as a spoken sufix; Campbell and Dodd (1980) found, conversely, that an auditory suffix produced a suffix effect on a lip read list. As Campbell and Dodd note, these effects may not be language specific:

> "They may reflect a general tendency for changing state information to be processed differently than information (usually visual) which can be resolved instantaneously."

#1.4.4 ON FINDING THAT SPEECH IS SPECIAL (LIBERMAN, 1981)

#1.4.4.1 NEITHER LIVING WITH NOR LIVING WITHOUT

Although almost all experimental techniques in speech research can be described in terms of dividing what is not really divisible and uniting what is not really unitable and thus as instantiations of the notion of simple location, it will now be shown that none of the authors who have been cited would accept that this notion is applicable to speech perception. This situation exemplifies Whitehead's (1926) comment on the prevailing presuppositions:

> "The world had got hold of a general idea which it could

neither live with nor live without."

#1.4.4.2 SIMPLICITY AND PSYCHOACOUSTICS

No researchers regard the physicist's 'simple' measures as appropriate for speech perception. Schouten (1981), who finds it 'easy to reject the speech mode', regrets that psychophysics and physiology:

> "Have mainly restricted themselves to pure tones and noise bursts, so that very little is known about the perception of timbre, let alone that of rapidly varying timbre, which is what speech is."

Blumstein and Stevens (1981), proponents of the 'acoustic invariants' approach, stress that the acoustic property must be 'appropriately selected', which is in line with Haggard's (1981) comments about:

> "The desirability of a psychophysiologically veridical measurement system or scaling procedure more adapted to the information-bearing aspects of speech sounds than the applied mathematician's spectrograph."

Haggard commented further that Rosen's (1981) report of range effects demonstrates that the similarity in boundary values in speech and music reported by Cutting and Rosen (1974) - see also Rosen and Howells (1981) and Cutting (1982) -

> "Is not sufficiently fixed to justify strong psychoacoustic determinism - ie the view that all parameter values are set by the inherent structure of the auditory system."

This remark applies, independent of the simplicity of the acoustic property which is selected.

Although Stevens (1981) accepts as valid - 'at least under certain speech perception tasks' - the strong hypothesis that phonetic features are marked by a common acoustic property and that this property is employed in the process of speech perception, the following more explicit formulation (Stevens, 1975) must be borne in mind:

> "It is postulated that the child initially utilizes property detectors to classify consonantal speech events in a canonical

> consonant-vowel environment and subsequently associates various secondary context-dependent cues with these phonetic categories.
>
> These properties do not identify the features in all phonetic environments ..."

In view of Schouten's (1980) acknowledgement that he was unable to account for Remez's (1977) adaptation of the speech - non-speech boundary, there does not appear to be any current support for strong psychoacoustic determinism, independent of the simplicity of the acoustic measures which are employed.

#1.4.4.3 THE INTEGRITY OF THE ACOUSTIC SPEECH SIGNAL

Many researchers who have used divided and united stimuli have commented on the integrity (wholeness) of the acoustic signal associated with normal speech. Fischer-Jorgensen (1954) observed:

> "The listener does not compare explosion with explosion and transition with transition, but compares artificial syllables comprising either explosion or transition with natural syllables that always contain both."

Summerfield et al's (1981) statement that:

> "Accounts of speech perception should, presumably, acknowledge the full breadth of a listener's attunement to the acoustics of speech and not only the basic sensitivities that might do the job."

fits in well with Haggard's (1981) discussion of the Haskins laboratories and their 'limited set of caricature features for speech sounds' and the current attempts to preserve the maximum amount of available acoustic information.

Juszyk et al (1981) follow Blumstein and Stevens in working with isolated 30 msec chirps containing varying numbers of formants, an extreme form of dividing, and yet they argue that the information content of the various formants is not indepenent:

> "If one assumes, instead, that it is the relationship that exists between the formants which is critical for perception,

then the first formant information cannot be considered to be redundant."

#1.4.4.4 SPEECH PRODUCTION AND SPEECH PERCEPTION

Although the strongest formulation of the possible relationship between speech production and speech perception, the motor theory, appears to have no active proponents (Studdert-Kennedy, 1981; Liberman, 1981), the view that production is the key to speech perception is supported by at least two different approaches. The event-perception approach associated in various forms with eg Gibson (1966), Neisser (1976), Summerfield (1979) and Studdert-Kennedy (1981) can be characterised by the formulation offered by Studdert-Kennedy (1981):

"The signal carries no message: it carries information about its source."

The speech-mode or phonetic-mode approach associated with eg Liberman (1981) and Repp (1981) asserts that:

"The key to the phonetic code is in its manner of production."

Whereas the event-perception approach holds that speech perception is not esentially different from the perception of any other event and thus that speech perception is only special because its source is special, the speech-mode approach makes a larger claim. Commenting on Summerfield's (1979) statement that the optical and acoustic signals are united in the common metric of articulatory dynamics, Liberman comments:

"I would agree, though I would, of course, prefer to call the common mode "phonetic" ... the important consideration is that, in the ordinary sense of modality, the speech percept is neither visual nor auditory; it is, rather, something else."

#1.4.4.5 RETROACTIVE EFFECTS IN SPEECH PERCEPTION

There have been many demonstrations that the perception of part of an acoustic signal is dependent on later portions of the signal (Rudnicky and Cole, 1978; Repp, 1980; Miller and Liberman, 1979). An extreme

instance is given in the phonemic restoration effect (Warren and Warren, 1970) where the final word of a sentence influences perception of the fifth last word.

All of this is an extension of what Liberman et al (1967) called 'encodedness', and implies the rejection of the notion of simple location.

#1.4.4.6 SIMPLE LOCATION: IMPLICIT AND EXPLICIT

The few examples which have just been given establish conclusively that speech researchers do not accept the concept of simple location as applied explicitly to speech; they also establish equally conclusively that its implicit assumption is endemic. It is indeed an example of an idea which speech research appears to be able neither to live with nor live without.

#1.5 SUMMARY

It has been demonstrated that the presuppositions which Whitehead and Bohm discussed underlie current speech research. The task of the following chapter will be to develop an approach which allows them to be evaluated.

CHAPTER 2: THE WHOLISTIC APPROACH TO SPEECH PERCEPTION

#2.1 NATURALISTIC SPEECH PERCEPTION

The task of the psychology of speech perception is at least to understand how one human being perceives the speech of another. Motor theory (Liberman et al, 1952), analysis by synthesis (Stevens, 1960), information processing (Cutting and Pisoni, 1976), electronic modelling (Chistovich, 1981), speech recognition systems (Klatt, 1977) and all other approaches, no matter how divergent they may be from each other, all agree on this point. The everyday event in which one person speaks to another who is present will be called the naturalistic case of speech perception. An adequate account of speech perception must apply to the naturalistic case; difficulties due to complicating factors such as lack of attention will be referred to in Chapter 4. In the naturalistic case, the speaker is present, audible and visible. As will be established in later chapters, faulty introspection has resulted in the widespread belief that speech is 'heard', and consequently research has concentrated almost entirely on audition and the acoustic signal. Obviously, audition does typically occur during speech perception, and will be assumed during the present discussion. However, the naturalistic case is an inviolably unified event, and it may not be assumed that any one aspect can be isolated as the stimulus for speech perception. Vision and audition are not factors ; they are better viewed as two aspects of a unified event.

#2.2 A SIMPLE METRIC FOR SPEECH PERCEPTION

Disregarding the involvement of the perceiver, it is possible to concentrate on three aspects:- is the speaker present, audible and

visible? To accommodate all research, it will be necessary to refer to the optical and acoustic signals. It is possible to classify any experiment along these dimensions and thus locate it relative to the naturalistic case.

The speaker's physical presence is easily reckoned with, his mental presence less so (see Chapter 4). The nature and extent of the manipulations of the optical and acoustic signals and the relationship between them also need to be established.

In all of the experiments which will be referred to in what follows the speaker is audible. Accordingly, the two normal experimental conditions will be labelled:-

NVis - speaker audible, but not visible

Vis - speaker audible and visible.

This choice of terminology is required by consideration of the naturalistic case. As soon as mention is made of adding an optical signal to an acoustic signal, it is clear that the point of reference is the tradition of purely auditory experimentation.

In the Vis condition, the optical and acoustic signals are intrinsically related when the same source is seen and heard, contrinsically related when two discrete sources produce signals which are (ideally) identical to those in the intrinsically related case, and extrinsically related in all other cases. In an obvious extension of this usage, experiments will be described as, for example, intrinsic if the two energy signals are intrinsically related etc.

Although this metric is very primitive, it can be used to locate any experimental technique. As will be demonstrated in the following section, the location of an experiment within the metric reveals its

relationship to the naturalistic case and can yield an evaluation of it or suggest further experiments to enable it to be evaluated. It is easily possible to apply the metric locally, even without attempting to specify it globally. For example, an experimental technique is nearer to the naturalistic case if it conceals or eliminates less of the original speech event. Thus, audio-visual presentation of a speaker is nearer to the naturalistic case than exclusively auditory presentation.

#2.3 REPLICATION TOWARDS THE NATURALISTIC CASE

The literature review in Chapter 1 was presented in terms of wholeness and fragmentation; it could equally well have been presented in terms of the naturalistic case. Every departure from wholeness is a departure from the naturalistic case and every departure from the naturalistic case is a departure from wholeness. In the naturalistic case, the indivisible is undivided and the ununitable not united. This leads immediately to important insights which can be expressed as a set of postulates for research. Before stating them, it is necessary to introduce a crucial concept. An attempted replication which controls all that was deemed relevant in the original experiment, but conceals less of the original speech event is called a replication towards the naturalistic case, or a beneplication.

1 An account of the psychology of speech perception must apply to the naturalistic case.

2 Non-naturalistic experiments may not be the basis for theories of speech perception.

3 A failure to replicate towards the naturalistic case reveals that the domain of the original experiment is insufficient to provide a satisfactory account of speech perception.

4 Experiments which do not technically permit an attempted replication

towards the naturalistic case cannot be the basis of a theory of speech perception.

It is clear that these postulates do not constitute a new set of presuppositions; they provide a way to examine experiments which is orientated towards the naturalistic case, but remains open on the wholeness-fragmentation issue. The first is quite unexceptional. The second leaves open the possibility that certain non-naturalistic experiments may prove to be special instances of a naturalistic case. In this event, however, the non-naturalistic experiments receive their justification and status from the naturalistic case. The third and fourth are also unexceptional, although it is always possible that researchers will decide that their original view of the domain boundaries was misconceived. These few comments make it clear that the set of four postulates is merely an expansion of the first.

Several simple examples may be helpful. Classical theories of speech perception are exclusively auditory-based. Postulate 2 states that theories of speech perception may not be based on the results of purely auditory experiments unless it has been demonstrated that the auditory domain is a special case of the auditory-visual domain. Postulate 3 states further that a single failure to replicate from the auditory to the auditory-visual domain demonstrates that the domain 'audition' is insufficient to provide a satisfactory account of speech perception, and thus that no auditory theory of speech perception can be adequate. According to Postulate 4, an experiment which employs stimuli which are not naturally consonant with human speech cannot be replicated to even the Vis contrinsically related case because no speaker could utter the required stimuli; this simple observation leads to the rejection of a considerable amount of theorising based on synthetic and manipulated

stimuli. It is always possible, however, that such experiments should suggest others which do allow attempted beneplication. Similarly, the Verbal Transformation Effect (Warren, 1968) may not serve as a basis for theorising, even with vision restored, because no speaker could exactly repeat a word or sentence at exactly regular intervals. The obvious beneplication has not been reported in the literature.

Further consideration of the importance of the speaker's presence will be deferred until Chapter 4. The following sections will be devoted to applying the metric to experimentation which involves both optical and acoustic signals.

#2.4 THE FUNDAMENTAL WHOLISTIC HYPOTHESIS

Within the wholistic approach, it is not appropriate to reify the smallest differences and distinctions of the current world view and regard them as the source of all causes and explanations. For it, the flow is prior to the things. This leads to the expectation that every attempt to explain the whole in terms of the things which can be seen to form and dissolve in the flow will be mistaken and misleading. Similarly, it is to be expected that every attempt to explain any domain in terms of a more restricted domain will be mistaken. In general terms, this yields the fundamental wholistic hypothesis:

FOR ANY EXPERIMENT IN ANY NON-NATURALISTIC DOMAIN, IN EVERY LESS RESTRICTED DOMAIN WHICH CONTAINS THE ORIGINAL DOMAIN, AN ATTEMPTED REPLICATION TOWARDS THE NATURALISTIC CASE CAN BE CONDUCTED WHICH WILL RESULT IN A FAILURE TO REPLICATE.

An apparent objection to this hypothesis is that the principle of restricted domains is so well known as not to require formulation; extending the domain merely means allowing a further - previously

constant - factor to vary with the predictable result that a more complicated pattern of responses will emerge. In terms of the present discussion this would mean that exclusively auditory studies have been intended to investigate the perception of auditorily presented speech, whereas the present study is concerned with audio-visually presented speech.

Two comments are in place:-

1 Earlier studies in audition abound in statements such as,

> "The listener is presented with an acoustic stimulus which can be analysed at a number of levels eg syllables, words, phrases and so on." (Bever and Carroll, 1981)

from which it is clear that all sounds, including speech sounds, were held to be perceived through the impingement of sound waves on the ear. That is, these studies were mistakenly taken to be investigations of the stimulus for audition, not of a specially restricted case.

2 Within psychology, domains are not specified according to the concepts of physics, but according to psychologically relevant criteria such as the naturalistic case.

#2.5 VISION AND SPEECH PERCEPTION - A TEST CASE

#2.5.1 AUDITORY EXPERIMENTATION

Any particular purely auditory experiment in speech perception may differ from the naturalistic case in a number of ways; it will certainly differ from it through the exclusion of the possibility that the listener see the speaker. This one form of divergence from the naturalistic case will now be investigated as a strict test of the principle of replication towards the naturalistic case (beneplication) as applied to the wholistic approach. Corresponding to the general formulation of the fundamental wholistic hypothesis, there results the

specific hypothesis;

FOR EVERY EXCLUSIVELY AUDITORY EXPERIMENT IN SPEECH PERCEPTION, AN ATTEMPTED REPLICATION TO THE AUDIO-VISUAL CASE CAN BE CONDUCTED WHICH WILL RESULT IN A FAILURE TO REPLICATE.

#3.2 and #3.3 together constitute an extensive test of this hypothesis. It is, perhaps, helpful to note that each of these experiments constitutes a further test of the principle of replication towards the naturalistic case. Just one failure to beneplicate is sufficient to establish that speech perception is not exclusively auditory.

#2.5.2 DUBBING AS A RESEARCH TECHNIQUE

#3.4 considers the applicability of dubbing as a research technique. Dubbing employs optical and acoustic signals which are extrinsically related, whereas they are intrinsically related in the naturalistic case; really, they are two aspects of one event. This is merely a more wholistic formulation of Summerfield's (1979) observation (see #1.4.3.2). Accordingly, dubbing constitutes a counter-naturalistic case and can not provide the basis for adequate accounts of speech perception. Although audio-visual, it is a compounding of fragmentation, not a step towards wholeness. Adding another factor to a multiplicity can never be an approach towards wholeness. What is called for is the revealing or becoming aware of another aspect of a unity.

As dubbing does not constitute an impoverishment of the naturalistic case, it is included in no naturalistic domain, however impoverished. Thus, postulate 4 immediately yields the following hypothesis:

IF AN EFFECT THAT IS OBTAINED THROUGH DUBBING CAN ALSO BE PRODUCED WITH AT LEAST CONTRINSICALLY RELATED OPTICAL AND ACOUSTIC SIGNALS, AN EXPERIMENT CAN BE CONDUCTED WHICH WILL RESULT IN A FAILURE TO REPLICATE

FROM DUBBING TO THE MORE NATURALISTIC CASE.

#2.6 SUMMARY

Having demonstrated the fragmentation of current research in Chapter 1, the next step was to sketch some fundamentals of a consistently wholistic approach in order to derive from them some hypotheses which are testable within the fragmentary mode of experimentation. The wholistic approach would of its own accord lead to quite different forms of research. Some further comments about this will be made in Chapter 4; for present purposes, it is sufficient to demonstrate that the wholistic approach can be applied to the investigation of results obtained within the fragmentary approach.

Bohm (1980) has characterised well the difficulties involved in overcoming fragmentation:

> "In the very act in which we try to discover what to do about fragmentation, we will go on with this habit (of confusion around the question of what is different and what is not) and thus we will tend to introduce yet further forms of fragmentation."

CHAPTER 3: AN EXPERIMENTAL APPLICATION OF SOME WHOLISTIC INSIGHTS

#3.1 PRELIMINARY REMARKS

3.1.1 INTRODUCTION

All of the experiments to be described in #3.2 and #3.3 involved presenting the identical acoustic signal to Ss, once Vis-ually and once NVis-ually. In each case, the experimental hypothesis was a specific formulation of the fundamental wholistic hypothesis as applied to audio-visual experimentation. According to normal practice, responses will be scored and subjected to statistical analysis. However, as McGurk and MacDonald (1976) noted, the recorded responses "fail to testify to the powerful nature" of the listener's experience. For any listener, one clear instance of perceiving the identical acoustic signal differently according to whether or not the speaker is visible is far more convincing than any amount of statistical analysis. This point will be considered further in Chapter 5.

#3.1.2 READING LIPS AND WATCHING SPEAKERS

It is essential to note that lip reading was not investigated in the experiments to be described below. Ss were asked to watch the speaker and then write down the word or non-word which they 'heard'; they were not asked to watch the speaker's lips and they were not asked to lip read. During debriefing, Ss often commented that - as was their custom - they had looked primarily at the speaker's eyes and had not really noticed his lip movements. Further Ss expressed no surprise or discomfort when their perception of the utterance conflicted with the perceived lip movements. Indeed, Ss often denied that there had ever been such a conflict. Ss were, in fact, vaguely looking at the speaker

and concentrating on the spoken utterance.

In Polanyi's (1967) terms, lip reading by unpractised Ss requires attending to the speaker's lips, whereas watching the speaker involves attending from the speaker's lips to his utterance. Thus, the important difference between the two activities is captured by the fact that they place the perceiver's relationship to the speaker's lips on opposite sides of the tacit-explicit distinction.

#3.1.3 EXPERIMENTAL MANIPULATIONS

#3.1.3.1 RECOGNISING LEVELS

Because of widespread conceptual confusion and carelessness (Repp, 1981) in the use of acoustic, auditory and phonetic terms, it is difficult to describe manipulations of the acoustic signal simply and unambiguously. Repp made an important contribution towards clarifying this area. Unfortunately, he presupposes a theory of perception which prejudges the relationship between the six levels of description which he lists. He makes a very interesting slip when he refers to silence as an acoustic segment. Just as silence belongs to the auditory, and not the acoustic level, so acoustic energy and not the acoustic signal, can be erased. Periods of zero acoustic energy are crucial parts of the acoustic signal. Zero energy is not identical to zero information. As Bateson (1978) wrote,

> "Zero is different from one, and because zero is different from one, zero can be a cause in the psychological world, the world of communication."

After all, Dorman et al (1979) did publish a paper with the title, "The Sounds of Silence."

Repp distinguishes between the "sounds of speech" (Pilch, 1979) and the

"abstract linguistic segments (the traditional 'speech sounds')". After having discussed the tendency to use linguistic terms such as consonant and syllable as if they were acoustic categories, he adds:

> "Perhaps, this malpractice originated with the time-honoured but quite misleading term, speech sounds. For, patently, we do not normally perceive a sequence of sounds when we listen to speech but a linguistic message in which phonetic segments are the smallest units."

The great weight which Repp and others patently attach to their introspections will be considered in Chapter 5.

#3.1.3.2 ATTENUATION AND DISLOCATION

When the naturalistic case is taken as the point of orientation, it is appropriate to classify manipulations of the acoustic signals as attenuations or dislocations. Used generously, attenuation can be applied to techniques such as filtering and masking. In everyday life, walls and background noises etc attenuate the acoustic signal. Dislocations are manipulations such as excising, inserting or superimposing portions of acoustic signal, and erasing segments of acoustic energy. Such manipulations yield acoustic signals which cannot occur in the natural world.

All of the 'sounds of silence' experiments involve dislocations. An obvious small beneplication is to introduce loud masking noise instead of erasing acoustic energy. Samuel (1981) has done this for the Phonemic Restoration Effect. The further beneplication is to employ naturally occurring masking sounds. Excisions, insertions and superimpositions defy direct beneplication as they constitute violations of 'ecological' time. They are the clearest possible instances of the notion of simple location; they divide what is indivisible and unite what is not unitable in the crassest possible manner.

#3.1.4 TECHNICAL DETAILS

All details of equipment, techniques and stimuli are listed in appendices. This is done to avoid repetition in the following sections. It will, however, be economical to mention several general points before reporting the experiments.

All experiments were conducted in quiet rooms with Ss seated comfortably in front of a video monitor. Headphones were only employed for Expt V and the pilot study mentioned in #3.2.8 because it was not possible to present these stimuli on the equipment which was otherwise employed. No restriction was placed on Ss responses in any experiment; no restricted response sets were employed; it was always stressed that there could be non-word stimuli. As testing was always done individually or in pairs, except for Expt XII, it was possible to discuss Ss responses. These discussions helped to determine criteria for combining data. For example, in Expt III some subjects made some 'thile' responses. As it had been expected that most responses would be 'isle', 'bile', 'vile' or 'file', these Ss were asked to clarify their responses and commented that the 'th' was intended to represent [TH]. Accordingly, when this data was grouped, this response was classified [v] and not [f]. It was seldom necessary to question Ss about their responses, and all such instances are detailed in Appendix 3. All Ss were untrained listeners. Although they did not notice that the stimuli had been manipulated, they could certainly have been trained to do so. As will be apparent from what follows, it is to be expected that such training would provide further support for the argument being developed here.

Statistical analysis was minimal because of the demonstration nature of almost all of the experiments. The Wilcoxon-Mann-Whitney 'd' test was

employed throughout. This statistic allows the easiest possible test of the hypothesis that every member of one group will score better than every member of a comparison group. Obviously, this is a much stronger hypothesis than that each Ss scores will be better in one condition than in another, in which case the Wilcoxon T statistic would be appropriate. It is only when manipulations which are expected to produce different effects are grouped together that the weaker hypothesis is called for (see Expt IV).

#3.1.5 OVERVIEW OF THE EXPERIMENTS

The experiments to be reported below are grouped as follows. All experiments in #3.2 employ dislocated acoustic signals and, thus, extrinsically related optical and acoustic signals. #3.3 reports a set of experiments with attenuated acoustic signals and contrinsically related optical and acoustic signals; they suggest that it will be possible to construct contrinsic and intrinsic experiments to probe the applicability to natural speech of many of the findings of non-naturalistic experimentation. This point will be considered in Chapter 4. The probe's first application will be reported in #3.4. Here, an experiment will be described which confirms the strong wholistic prediction about the results of dubbing techniques failing to replicate towards the naturalistic case.

The fragmentary approach will necessarily find many of the experiments incomplete; they appear to yield phenomena which call for and would obviously reward more detailed investigation. In each case, a moment's pause will reveal that the more detailed investigation would involve adopting the fragmentary outlook, and the aim of the experiments is to establish that this step would be misguided.

#3.2 EXTRINSIC EXPERIMENTS

#3.2.1 INTRODUCTION

Most experiments in speech perception employ dislocated acoustic stimuli and thus most opportunity to make contact with the literature is offered by performing similar experiments. When the acoustic signal is dislocated and the optic signal unmanipulated, the two signals must be extrinsically related. That is the case with all experiments which will be reported here. Expts I-IV involve erasures, V and VI involve excisions. A pilot study investigates a film loop.

#3.2.2 EXPERIMENT I: ABRUPT ONSET ERASURES

Many researchers have investigated the effects of erasing (or slicing off) the onset of an acoustic signal (Fischer-Jorgensen, 1954; Dorman et al, 1977). All of this work has been done with exclusively acoustic stimuli. It is hypothesized that when Ss can see the speaker, their speech percepts will be different. Two instances which can be expected to yield very clear effects will be investigated in this experiment:-

1 When the onset of the acoustic energy of an utterance of, for example, 'blend' is erased, Ss progressively report perceiving 'lend' and then 'end'. The specific form of the present general hypothesis is then that with Vis access to the speaker, Ss will still tend to perceive 'blend' or 'bend'.

2 When successive erasures are made, an utterance such as 'file' is successively perceived as 'vile', 'bile' and 'isle', given that access is NVis. With Vis presentation, the present hypothesis predicts that Ss percept will be nearer to the original utterance.

These hypotheses were tested by subjecting each of six 'blend' (including three 'breach') and each of six 'file' words to two unequal erasures and inserting the resulting twenty-four stimuli in a list of altogether thirty words, subject to the condition that no three consecutive words were of the same type. Interword spacing was approximately 7 secs. Ss, who were tested individually or in pairs, were simply asked to write down what they 'heard'. 'Blend'-words will be considered first. Results are displayed in Tables I.1 and I.2. An example of the manipulation employed is shown in Fig I.1.

		ORDER OF PRESENTATION		
		VIS-NVIS	NVIS-VIS	COMBINED
[b] perceived	VIS	11.6(11-12)	10.0(7-12)	10.8(7-12)
	NVIS	5.8(4-8)	2.8(1-6)	4.3(1-8)
Original word	VIS	4.8(4-6)	5.4(2-8)	5.1(2-8)
	NVIS	0.8(0-2)	0.4(0-1)	0.6(0-2)

TABLE I.1: Means and ranges for 'b' responses and original word responses.
Max score=12. n=5 for each order of presentation.
More 'b' words were perceived with Vis presentation:
d=1, m&n=10, $p<0.001$ one sided.
With NVis presentation, the order effect for perception of 'b' words is significant.
d=3, m&n=5, $p<0.05$ one sided.

NVIS PRESENTATION	VIS PRESENTATION				
	end	lend	bend	blend	total
end	4	1	17	9	31
lend		6	6	35	47
bend	1		34	1	36
blend			1	5	6
total	5	7	58	50	120

TABLE I.2 Response matrix for 'blend'-words.
12 sets of responses for each subject. n=10.

As is clear from these tables, both 'b' responses and original word responses were significantly more common with Vis access. The order effect for 'b' responses with NVis presentation deserves some comment. Ss in the Vis-NVis order of presentation sometimes commented with NVis

presentation that the 'b' was "just there". Having recently perceived the list with Vis presentation, Ss were alerted to the presence of onset 'b's and, with this set they could sometimes 'just manage' to perceive them. Inspection of the response matrix reveals that, as was predictable, stimuli which were perceived as 'end' with NVis presentation tended to be perceived as 'bend' when presented Vis-ually, whereas words which had been perceived as 'lend' tended to be perceived as 'blend'. This difference is significant ($d=15.5$, $m\&n=10$, $p<0.01$ one sided).

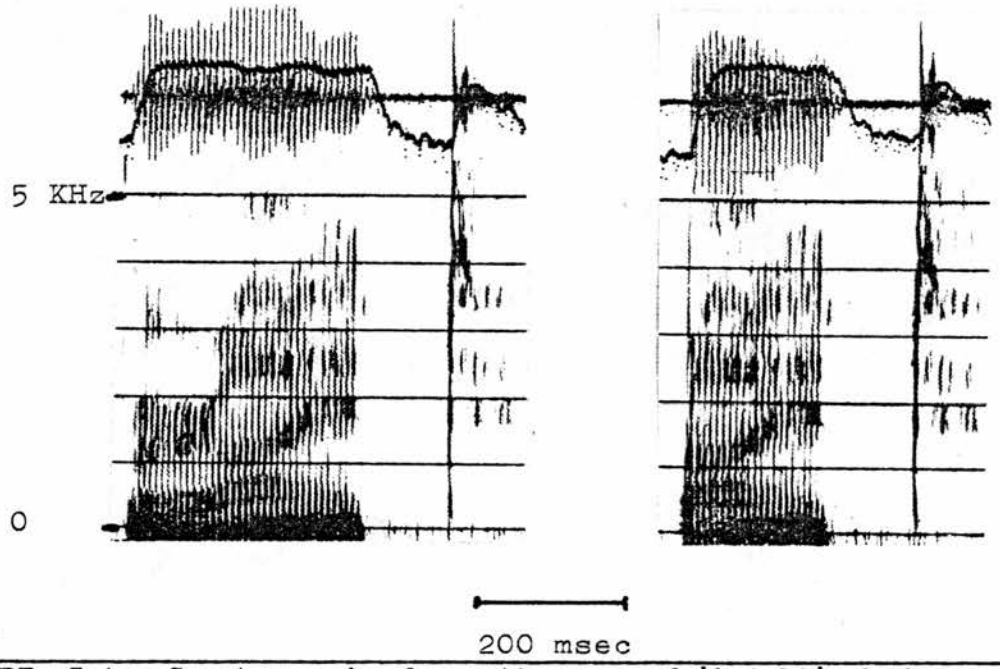

FIGURE I.1: Spectrograph of an utterance of 'bright', before and after abrupt onset erasure.

The many NVis 'bend' responses indicate that the abrupt onset after erasure still "carries information" about the onset of the original word. It is quite possible that the onset 'b's are dependent on where the erasure ends relative to voice pulsing. That would be a good fragmentary non-phenomenon (Ades, 1981). A technique which largely eliminates the 'bend' responses was developed and will be described in Expt II.

The 35 instances of a stimulus being perceived NVis-ually as 'lend' and Vis-ually as 'blend' are not to be described as visual restorations; they are rather visual preservations. With NVis-ual presentation, access to the originally uttered 'b' has been eliminated for many Ss; with Vis-ual presentation, access is still attainable.

		ORDER OF PRESENTATION		
		VIS-NVIS	NVIS-VIS	COMBINED
Original word	Vis	7.6(3-12)	5.6(2-8)	6.6(2-12)
	NVis	2.8(2-4)	1.8(1-3)	2.3(1-4)

TABLE I.3: 'File' responses in both conditions and both orders of presentation.
Max score=12. n=5 in each condition.
More original-word responses were made with Vis presentation:
d=8, m&n=10, p<0.001 one sided.

NVIS RESPONSES	VIS RESPONSES				
	isle	bile	vile	file	total
isle	9		20	19	48
bile		1	11	19	31
vile			10	6	16
file			1	22	23
total	9	1	42	66	118

TABLE I.4: Response matrix for 'file' words. Results for both orders of presentation have been pooled. Two responses missing because of difficulties with tape.
n=5 in each condition, total=10.

Responses for the 'file' words were equally interesting. They are summarised in Tables I.1 and I.2. Erasures are as shown in Figure III.1.

Again the main effect is clear and the order effects, although not attaining statistical significance, are in the expected direction. The order effects make it likely that strong range effects (Rosen, 1981) could be obtained with such stimuli. As is shown in Table I.4, there were 76 instances in which Ss responses to the identical acoustic signal differed across the two conditions; in only one instance was the

difference not in the predicted direction. The remarkable pattern of 'bile' responses, 31 with NVis presentation and only 1 with Vis presentation, accords well with Summerfield's (1979) findings. Further discussion of these results will be reserved until Expt III.

For both 'blend' and 'file' words it is not appropriate to attach significance to any of the actual numbers which are shown in the various tables because a different set of erasures would have yielded quite a different set of results. Only the overall statement that the perception of the presented stimuli was not exclusively auditory is justified.

#3.2.3 EXPERIMENT II: TRANSIENT ONSETS AFTER ERASURE

The many 'bend' responses to 'blend' words were essentially eliminated by a new technique which produced erased stimuli with transient onsets. An example stimulus is shown in Figure II.1.

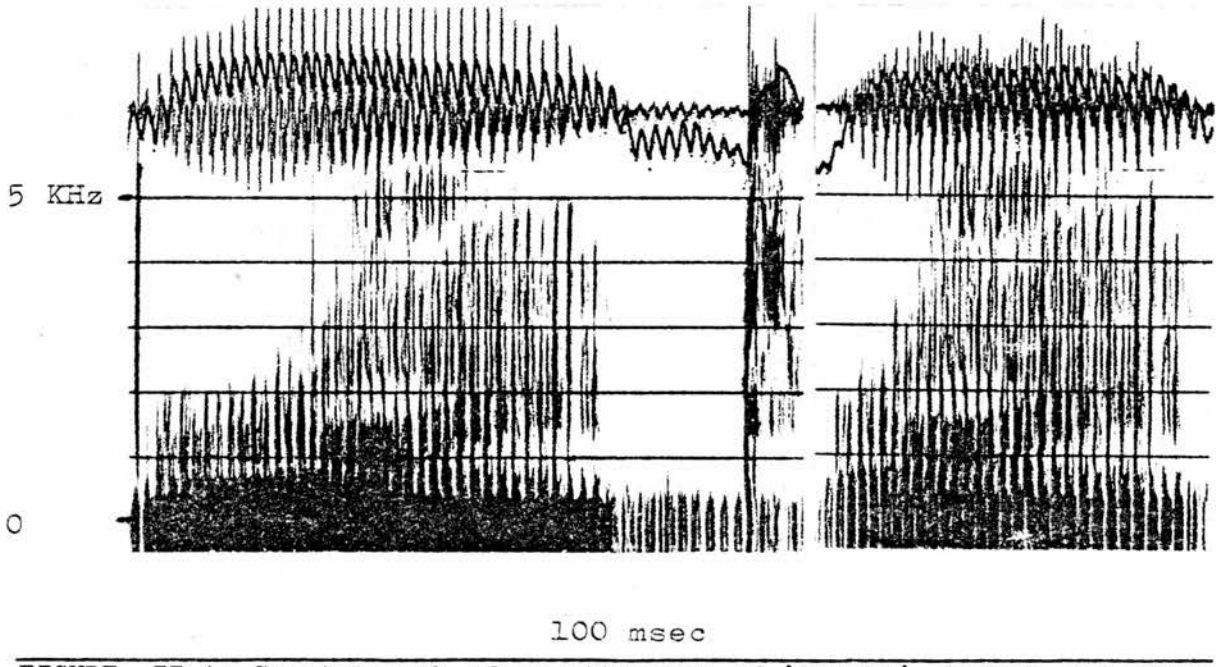

FIGURE II.1: Spectrograph of an utterance of 'bright', before and after transient onset erasure.

As informal testing demonstrated that the revised technique was very successful, it was decided to employ it to investigate the robustness of

the involvement of vision in speech perception. Three new Vis conditions were studied:-

1 Vis-Inv(erted), in which the speaker's face is shown rotated through 180 degrees,

2 Vis-Res(tricted), in which only a vertical 3/8"-wide strip down the centre of the speaker's face is visible,

3 Vis-Ter(minated), in which the the film of the speaker's face is terminated before the onset of the (partially erased) acoustic signal.

Mean 'b' and 'bend' responses for each condition are shown in Table II.1.

	CONDITION			
	NVIS	VIS-INV	VIS-RES	VIS-TER
'b' responses	0.8	5.6	5.4	5.3
'bend' responses	0.1	0.4	0.9	0.1

TABLE II.1: Means of 'b' and 'bend' responses.
Max score=6. n in each condition = 7, same subjects tested in NVis and Vis-Inv conditions.
All Vis scores differ from NVis score.

Table II.1 shows that the 'bend' response category was almost eliminated with NVis presentation, as also in all Vis conditions. Each of the Vis conditions proved to be very effective. This is not to assert that there are no differences between them. The wholistic hypothesis would predict that there are differences which appropriately selected stimuli could reveal. The Vis-Inv results are striking because this is a most unusual setting of which Ss would have had little or no experience. The Vis-Res results relate to Summerfield's (1979) 'four dot' condition, but do not allow direct comparison because with his four-dot setting the speaker's mouth never seemed to close. The Vis-Ter scores reveal a truly remarkable effect; vision influences speech perception even when there is absolutely no temporal overlap between the periods of non-zero optic

and acoustic energy. As is shown in Table II.2, the average gap between offset of the optic and onset of the acoustic signal was 70 msec. Table II.2 also displays responses for each of the six stimulus words in each experimental condition.

STIM	RESPONSES													
	NVIS				VIS-INV			VIS-RES			VIS-TER			GAP
	E	L	B	BL	L	B	BL	L	B	BL	L	B	BL	
1		6		1	1		6		1	6			7	60
2	1	6					7	1	3	3		1	6	160
3		7					7			7	1		6	-20
4		6	1		2	3	2	2	2	3	5		2	40
5		4		3			7			7			7	60
6		7					7	1		6	2		5	40

TABLE II.2: Responses and zero energy times for each stimulus.
n in each condition = 7.
E=END, L=LEND, B=BEND, BL=BLEND.

It would certainly be possible to covary the extent of erasure, form of onset of the acoustic signal and the temporal relationships bewteen onset and offset of the acoustic signals and thus derive a function to predict Ss response patterns; at least it would be possible to do this with 'simplified' synthetic stimuli. The overall argument, however, is that such model-making would be misguided.

#3.2.4 EXPERIMENT III: EMBEDDED PHONEMES

This experiment is a more systematic study of the 'file-vile-bile-isle' effect which was reported in Expt I. A stimulus set of 48 words was prepared by recording an utterance of 'file', making 6 different onset erasures and quasi-randomly ordering 8 instances of each of the 6 stimuli at intervals of 7 secs. A spectrograph of the original utterance showing the positions of the 6 erasures is given in Figure III.1.

As there appears to be an order effect (see Expt I), it was decided to minimise the statistical effect of vision by testing all Ss first with

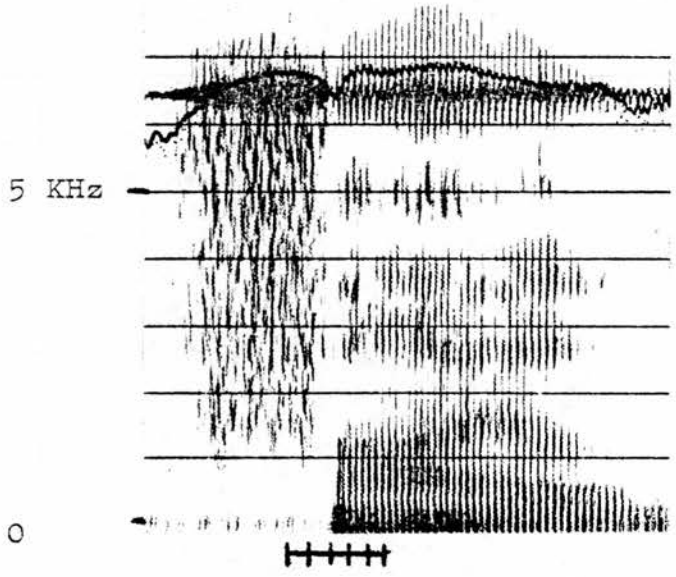

FIGURE III.1: Spectrograph of an utterance of 'file' showing the the termination points for the six erasures.

NVis presentation. Results are displayed in Figure III.2 and Tables III.1 and 2.

NVIS RESPONSES	VIS RESPONSES				
	Isle	Bile	Vile	File	Total
Isle	12	-	35	83	130
Bile	2	-	49	65	116
Vile	1	-	13	16	30
File	-	-	7	101	108
Total	15	-	104	265	384

TABLE III.1: Response matrix for 'file' words, combining results for all Ss and all stimuli. n=8. There were 8 presentations of each of 6 stimuli to each S.

RESPONSE	STIMULUS											
	1		2		3		4		5		6	
	V	NV	V	NV	V	NV	V	NV	V	NV	V	NV
Isle	-	5	-	2	-	7	3	17	4	54	8	45
Bile	-	6	-	6	-	47	-	34	-	10	-	13
Vile	3	3	10	15	22	3	24	5	20	-	25	4
File	61	50	54	41	42	7	37	8	40	-	31	2

TABLE III.2: Grouped responses to 'file' words of 8 Ss for 8 presentations of each of 6 erasures.

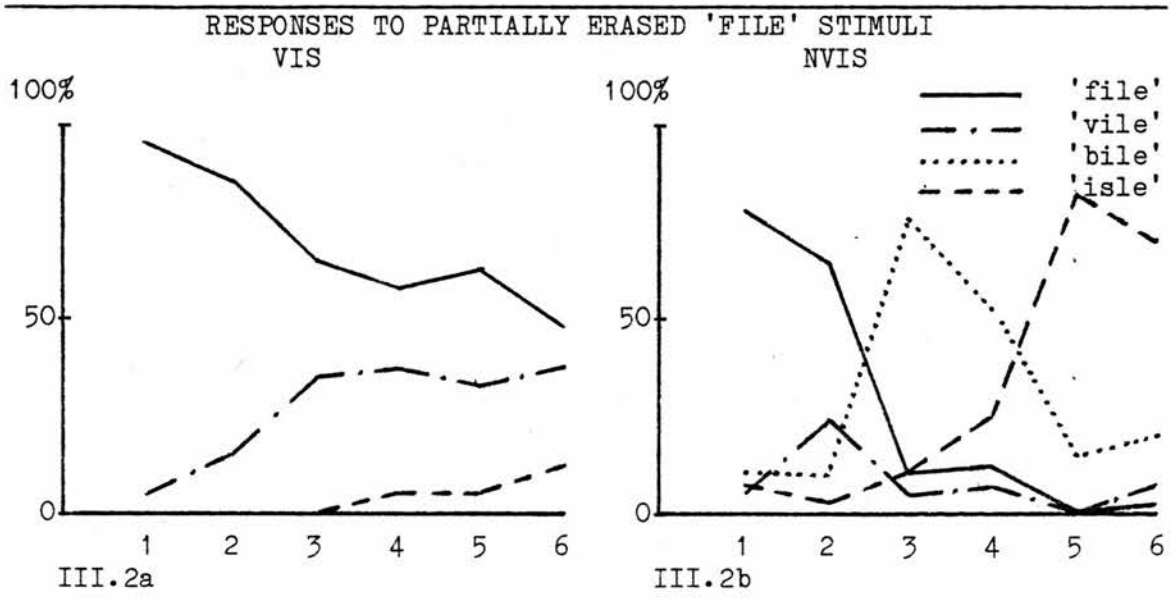

FIGURE III.2: % Responses of 8 Ss to partially erased 'file' stimuli. 8 responses from each S to each of 6 erasures.

The NVis curves are very much as would be expected from Cole and Scott's

(1974) description of the series, 'sha-cha-ja-da', which results from successive erasures of the onset of an utterance of 'sha'. It is possible that a further stimulus between the second and the third erasures would have shown 'vile' to be the consensus NVis response. The Vis curves, however, reveal the major effects which are obtained when S is not permitted to see the speaker. A whole response category, 'bile', appears only when the speaker is not visible. All of the other response categories are also affected. Repeated Wilcoxon tests (not all of them are independent) show the differences between the two sets of responses to differ for each response category for each of the final four erasures.

Ss responses revealed that they perceived differences between the stimuli which are finer than the results already shown indicate. For example, one differentiated between 'file' and 'phile', the second response including a "touch of p". With Vis presentation, each presentation of Stimulus 1 was identified as 'phile', one presentation of stimulus 2 was so labelled and there were twenty 'file' responses. All of this indicates that although not one S suspected that the stimuli had been manipulated, improved perceptual judgement could well be expected to be able to detect manipulated stimuli. The question of improved perceptual judgement will be important in Chapter 5.

Also here, it is obvious that it would be possible to construct a bi-modal model to accommodate these results. Its overall structure is indeed obvious. With NVis presentation, shortening the fricative aperiodicity decreases 'file' responses, abrupt onsets with varying formants lead to 'bile' responses, and 'isle' responses are made when the varying formants have been erased. Vis presentation permits bilabial responses only when lip closure is shown, and the visible articulatory

movement in its time-course corresponds to the gradual onset of "file" not the more abrupt onsets of 'vile' or 'isle'. It is remarkable that the comparatively fine distinction between the articulatory gestures for utterances of 'file' and 'vile' should prove to be so perceptually significant for untrained Ss.

#3.2.5 EXPERIMENT IV: THE SIGHT OF SILENCE

The first three experiments have all investigated onset erasures; this experiment will consider an interior erasure. The much-investigated 'slit-stlit-split' sequence (Dorman et al, 1981; Summerfield et al, 1981) suggests that erasing acoustic energy at about the end of the fricative aperiodicity of the acoustic signal of an utterance of 'slit' will be comparable to 'varying the period of silence' (Summerfield et al, 1981) between the fricative energy and the remainder of the acoustic signal. As this proved to be the case, several predictions follow immediately from the wholistic hypothesis:

1 There will be less 'split' responses with Vis presentation,

2 There will be more 'slit' responses with Vis presentation

It is difficult to separate predictions 1 and 2, although they really are distinct. 1 asserts that the 'p' will not be perceived when the optical signal contraindicates its presence. 2 asserts that it is not just the gross optical information corresponding to lip closure which is significant, but that the overall gesture corresponding to an utterance of 'slit' will tend to preserve perception of the original word. The 'say-stay' distinction would offer the possibility to test this prediction directly. Fortunately, however, 2* allows the same effect to be tested.

2* There will be more 'split' responses with NVis presentation than

'stlit' responses with Vis presentation.

This asserts that with Vis presentation, not only will the 'p' not be perceived, but there will also be less tendency for the experimental manipulation to be perceived at all as a stop consonant. Thus, when 1 and 2* are analysed statistically, it is possible to conduct independent tests of both of the original predictions. Typically for such a fragmentary experiment, there is the likelihood that 2* is confounded with a tendency to perceive words.

Fifteen stimuli were prepared by making erasures of 50, 100, 150, 200 and 250 msec ending at each of the three positions marked on Fig IV.1 which is a spectrograph of an utterance of 'slit'. The fifteen stimuli were randomly ordered at 7 sec intervals. All Ss were tested in the order NVis-Vis with a short pause between the two conditions.

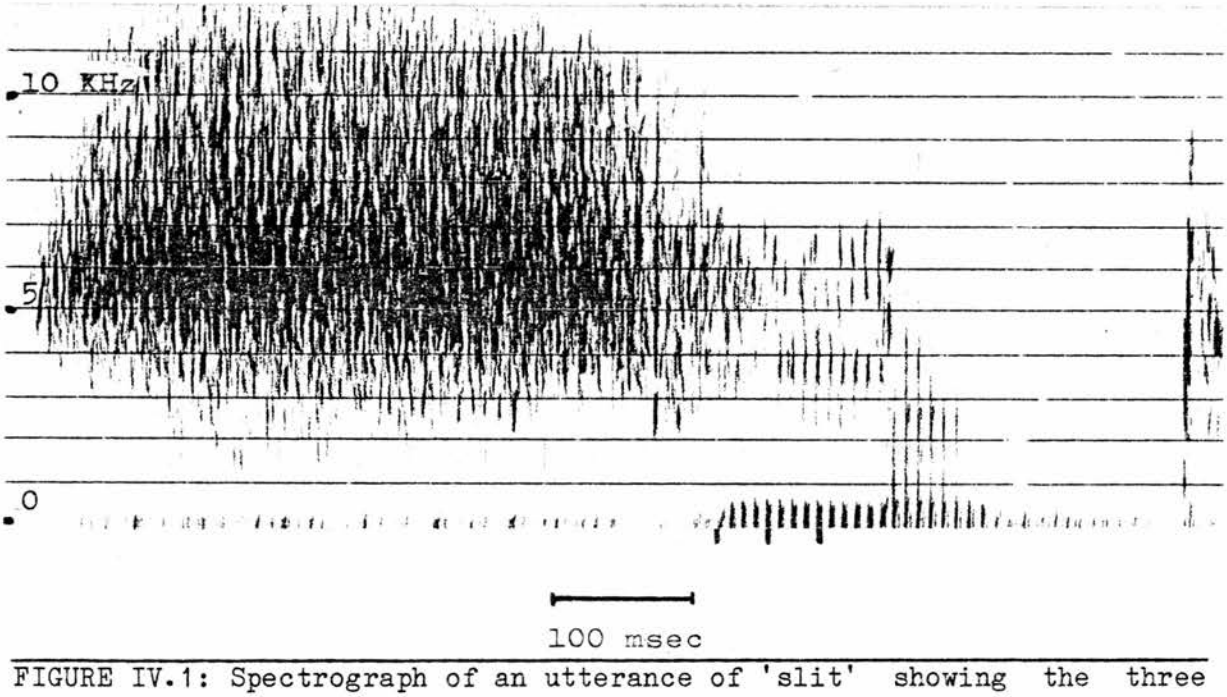

FIGURE IV.1: Spectrograph of an utterance of 'slit' showing the three termination points for the internal erasures.

ERASURE		POSITION OF ERASURE											
		A			B			C			Total		
		p	t	l	p	t	l	p	t	l	p	t	l
50	NVis	3	-	7	2	-	8	1	-	9	6	-	24
	Vis		-	10		-	10		1	9		1	29
100	NVis	6	1	3	4	-	6	5	-	5	15	1	14
	Vis		3	7		1	9		1	9		5	25
150	NVis	4	2	4	4	-	6	7	-	3	15	2	13
	Vis		6	4		4	6		2	8		12	18
200	NVis	8	1	1	5	2	3	8	-	2	21	3	6
	Vis		8	2		2	8		4	6		14	16
250	NVis	8	1	1	8	-	2	8	1	1	24	2	4
	Vis		8	2		8	2		7	3		23	7

TABLE IV.1: Responses for 10 Ss for each of 15 partially erased 'slit' stimuli.
p=split, t=stlit, l=slit.

	VIS RESPONSES			
NVIS RESPONSES	p	t	l	Total
p	-	39	42	81
t	-	6	2	8
l	-	10	51	61
Total	-	55	95	150

TABLE IV.2: Responses matrix for 'slit' stimuli with interior erasures.

Results are shown in Tables IV.1 & 2 and Fig IV.2. Fig IV.2b displays a typical response curve for a 'sound of silence' experiment, confirming that the erasure technique was satisfactory. Comparing 'split' responses for 50 and 100 msec erasures with those for 200 and 250 msec erasures confirms that the proportion of 'split' responses increases with erasure length ($d=0$, $m\&n=6$, $p<0.01$). With Vis presentation, the response pattern is radically altered. Whereas there were 81 'split' responses with NVis presentation, there were none at all with Vis presentation; this difference is significant ($d=0$, $m\&n=15$, $p<0.001$). Inspection of Table IV.1 reveals further that 2* was also confirmed ($T=1.5$, $n=9$ (6 equal),

p<0.001). This powerful finding is evidence that even fine articulatory distinctions such as that between 'slit' and 'stlit' may be perceptually potent.

RESPONSES TO INTERNALLY ERASED 'SLIT'

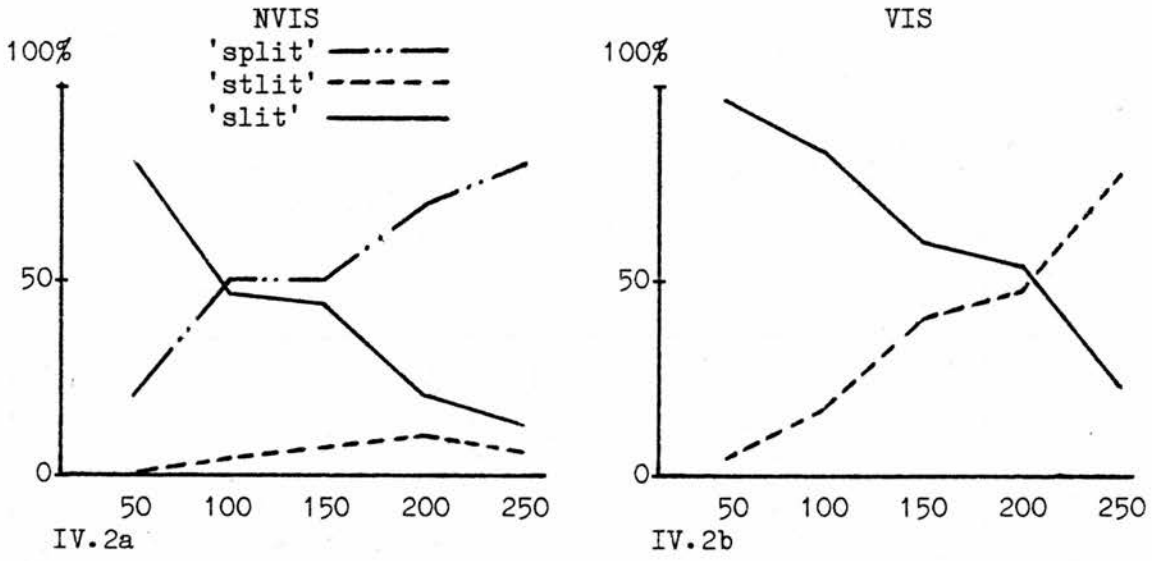

FIGURE IV.2: Responses of 10 Ss to each of 15 stimuli with varying internal erasures.
Results for all 3 erasure positions have been pooled.

#3.2.6 SUMMARY

These four experiments together give a clear picture of the major effects which result when the speaker is not visible. In Expts III and IV, whole response categories only appeared with NVis presentation. It can now be taken as established that thorough justification will be required before any conclusions at all about speech perception are drawn from NVis experimentation with dislocated stimuli. McGurk and MacDonald (1977) obtained significant effects only when at least one of the energy signals corresponded to an utterance of a bilabial. The effects of visible utterances of 'file' and 'slit' extends the range of established effects of vision on speech perception. Because the slit-stlit distinction is very fine indeed, it appears likely that every distincion

can be perceptual effective. This leads to the possibility that improved visual capabilities will be influential in improving speech perception.

Lip separation was measured for a number of stimuli, but was found to be too crude a measure to merit inclusion here. It is, however, of some interest that in the Vis-Ter(minated) condition of Expt II, lip opening had not occurred for stimuli 1,2 & 5 before the film was terminated.

As has already been stated, it would have been relatively straightforward to construct a bi-modal model of speech perception by extending several of these experiments. There is, however, no reason at all to think of speech perception as a bi-modal activity. At present, the question of modalities is quite unanswerable and it is more advantageous for investigations to approach the naturalistic case before allowing the old presuppositions to express themselves in new assumptions about a number of modalities.

#3.2.7 EXPERIMENTS WITH DISLOCATED ACOUSTIC SIGNALS

#3.2.7.1 INTRODUCTION

As excised acoustic signals appear to constitute extremely non-naturalistic cases and it is impossible to retain the original optical signal without creating severe asynchrony between even the envelopes of the two signals, only two small investigations of them will be reported here.

#3.2.7.2 EXPERIMENT V: INTERSYLLABIC CLOSURE

Many studies have shown that the length of the appropriate period of zero acoustic energy is related to how and whether the consonants C1 and C2 are perceived in [CVC1 C2V] combinations (Rudnicky and Cole, 1978).

The wholistic hypothesis predicts that perception of C1 and C2 will be related to the speaker's being visible. In the simplest case, it can be predicted that when C1 is a bilabial, it will be perceived more often with Vis presentation.

To test this hypothesis, seven stimuli consisting of utterances such as [gabga] and [papka] were prepared and dislocated by removing 5 frames of zero acoustic energy from the sound track (approx 210 msec). The stimulus list with interstimulus interval of 5 sec was then presented to Ss in both conditions, with a short pause between conditions. As there was no apparent order effect, pooled results only are shown in Table V.1.

CONDITION	BILABIAL RESPONSES
Vis	2.7(0-4)
NVis	0.0(0-0)

TABLE V.1: Mean bilabial responses to bisyllables with excised inter-syllabic acoustic signal.
Max score=7. n=6.
More bilabials were perceived with Vis presentation.
d=3, m&n=6, $p<0.05$ one sided for Ss.
d=0, m&n=7, $p<0.001$ one sided for stimuli.

	POSITION OF BILABIAL	
STIMULUS	1st Syllable	2nd Syllable
/gabga/	1	1
/babga/	-	1
/babda/	2	-
/tapta/	2	1
/kapka/	4	1
/papta/	1	1
/papka/	1	-
Total	11	5

TABLE V.2: Bilabial responses to each of 7 stimuli, grouped according to position.

The effect of vision is small, but consistent. Obviously, more detailed experimentation would have succeeded in locating excision values for which the effect was much stronger (eg stimulus 4). The result, however,

has another significance precisely because of the small proportion of bilabial responses. No S expressed surprise or discomfort because lip closure was not associated with the perception of a bilabial. As ever, Ss were not aware of attending to the speaker's lips. The high proportion of non-bilabial responses establishes that Ss do not have a simple bias to respond with a bilabial whenever lip closure occurs; the relationship between optical and acoustic signals is clearly very important.

Table V.2 shows that there are at least two different effects involved here. Bilabial responses in the first syllable are a form of visual preservation, whereas second syllable responses are possibly instances of the McGurk and MacDonald illusion.

Thus, this little experiment has established that the influence of bisyllabic closure is related to the speaker's being visible and that bilabial responses are not a simple artefact due to response bias. Ss reports that they write down what they have 'heard' receive independent confirmation.

#3.2.7.3 EXPERIMENT VI: 'SLIT-SPLIT' REVISITED

It has been reported (Dorman et al, 1975) that when the fricative aperiodicity of the acoustic signal of an utterance of 'split' is moved nearer to the remainder of the acoustic signal, there is a tendency for Ss to perceive 'slit'. In the terms of the present report, this procedure involves excising a portion of the appropriate zero energy segment of the acoustic signal, or superimposing the fricative aperiodicity on a part of the remainder of the signal. The obvious prediction is that with Vis presentation, Ss will continue to perceive 'split' even when NVis presentation results in the perception of 'slit'.

Ten stimuli were prepared with various excisions; an example is shown in Figure VI.1. As is clear from Figure VI.1, part of the acoustic signal was also erased during preparation of the stimuli. Thus, this experiment, although relevant to the hypothesis under consideration, does not constitute a beneplication of the Dorman et al study. The stimuli were randomly ordered with inter-syllabic interval of 10 secs. Each S was tested in both conditions, order always NVis-Vis, with a short pause between conditions. Results are shown in Tables VI.1 & 2.

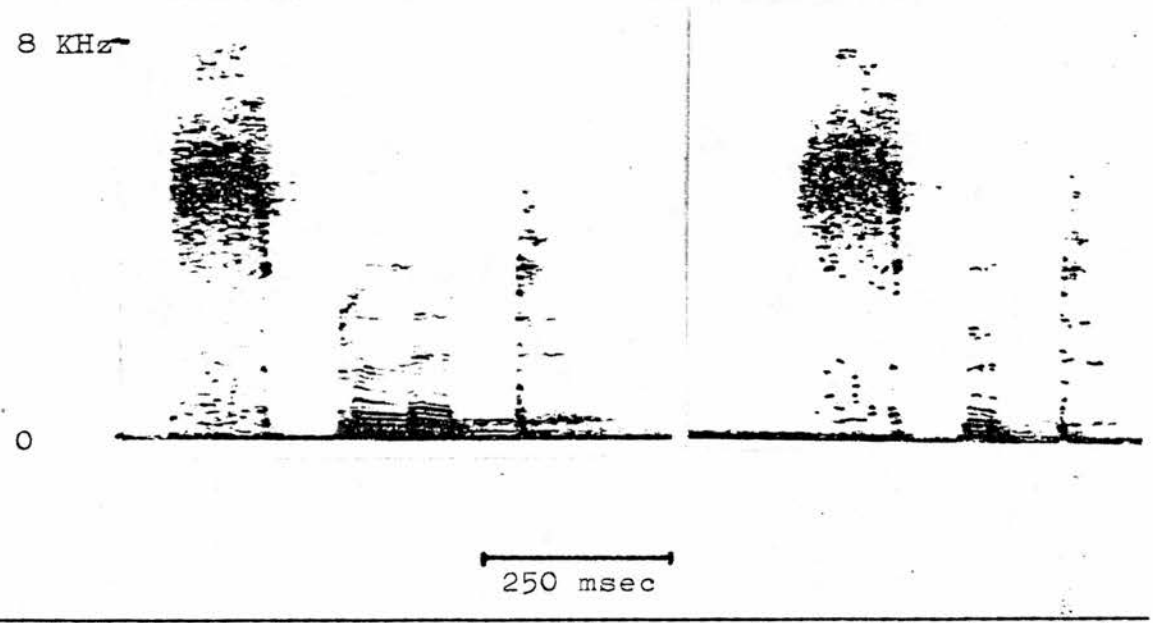

FIGURE VI.1: Spectrograph of an utterance of 'split', also showing a manipulated stimulus.

CONDITION	BILABIAL RESPONSES	'SPLIT' RESPONSES
Vis	9.9(9-10)	7.7(6-10)
NVis	7.7(6-10)	3.1(2-6)

TABLE VI.1: Bilabial and 'split' responses to 'split' stimuli.
Max score=10, n=8.
More bilabials were perceived with Vis presentation.
d=12.5, m&n=8, $p<0.01$ one sided for Ss.
d=13.5, m&n=10, $p<0.01$ one sided for stimuli.

NVIS RESPONSES	VIS RESPONSES				
	Sit	Slit	Spit	Split	Total
sit	-	-	3	-	3
slit	-	-	1	14	15
spit	-	-	29	8	37
split	-	1	-	24	25
Total	-	1	33	46	80

TABLE VI.2: Response matrix for 'split' stimuli. Ten stimuli presented to each of eight Ss. All responses pooled.

As shown in Table VI.1, the experimental hypothesis was confirmed. Table VI.2 shows further that whereas there were 26 instances in which the identical stimulus was perceived as containing a bilabial with Vis presentation, but not with NVis presentation, there was only one instance of the opposite response pattern. The many 'spit' responses constitute a curious fragmentary phenomenon; the eight NVis 'spit' & Vis 'split' responses appear to be instances of another form of visual preservation.

#3.2.7.4 CONCLUSION

These two short experiments demonstrate that speech perception is not independent of vision when the acoustic signal has been severely dislocated. Because excision is such a severe form of dislocation, it is not appropriate to pursue the reported findings any further. Again, it can be stated that any experimentation with dislocated acoustic signals requires special justification before it can be considered as a basis for an account of speech perception.

#3.2.8 PILOT STUDY: AUDITORY SEGMENTATION

Cole and Scott (1974b) report that when Ss are presented with repetitions of utterances of 'fah' at the rate of two per second, they

soon report that the stimulus segments into a cat's hiss and repetitions of 'bah'. When the transitions have been removed from the stimulus, segmentation occurs after just two or three repetitions of 'fah'; when they have not been removed, more than sixty repetitions are required. The wholistic prediction is that segmentation will be less likely to occur when the speaker is visible.

Damage to the stimulus film prevented completion of this experiment. However, as the results of a pilot study are very interesting they will be reported here. The stimulus consisted of a film and tape loop of fifteen utterances of 'fah' with a duration of 4.25 secs. 6 Ss (including E) were tested, controlling for order of presentation. The stimulus was presented for approximately one minute (225 repetitions of 'fah') in each condition. Not one S experienced segmentation with Vis presentation, although several experienced minor verbal transformations such as 'ahf'. With Vis presentation, 3 Ss (including E) experienced a form of segmentation. The stimulus was felt to become dehumanised, degenerating into a distressing throbbing. Looking at the speaker converted the throbbing into speech again. It is justified to report this small study here because the experience of it is extremely powerful. Seeing the speaker does not just influence how individual phonemes are perceived; it influences whether or not such an artificial stimulus is even perceived as speech.

Segmentation is obviously a form of verbal transformation (Warren, 1968). Vis-ually presented loops are a very interesting case because the optic and acoustic signals are everywhere locally contrinsically related and yet their relationship is globally impossible. The closed loop which produces this situation is an excellent symbol which suggests that the verbal transformation effect is to be viewed as a form of sensory

(information) deprivation.

#3.2.9 SUMMARY

It has now been confirmed for a considerable number of experimental techniques that NVis-ual findings obtained with dislocated acoustic signals do not replicate to the Vis-ual case. It would be possible to continue the demonstration by beneplicating further experiments. It is, however, more desirable to grasp why such demonstrations can only have the negative purpose of directing attention to the naturalistic case.

#3.3 EXPERIMENTS WITH ATTENUATED ACOUSTIC SIGNALS

#3.3.1 INTRODUCTION

Following all of the above experimentation, there remains the possibility that the non-exclusively-auditory nature of speech perception is restricted to artificial settings in which the acoustic signal or its relationship to the optical signal has been tampered with. If this were to be true, it would lead not just to the rejection of the experiments reported above, but equally to the rejection of almost all exclusively auditory experimentation. In either case, the tradition of exclusively auditory experimentation has to be recognised as inadequate for the psychology of speech perception.

The following five experiments involve no manipulation of the acoustic signal beyond masking and filtering; in one of them, there is not even any attenuation. All five reveal clear differences in speech perception across the Vis and NVis conditions.

#3.3.2 EXPERIMENT VII: MASKING WITH WHITE NOISE

It has long been known that audio-visual word recognition is superior to auditory word recognition when the speech signal is masked. Ss in the present experiment reported that their percept was quite different in a number of cases when the speaker was visible.

Minor versions of elliptic speech (Miller and Nicely, 1953) were constructed by mispronouncing one phoneme in a proverb. For example, 'A fool and his mummy are soon parted' involves replacing the [n] of money with [m]. As Miller and Nicely report, such pronunciations are virtually undetectable when the substitutions are chosen appropriately. For the construction of the stimuli for this experiment in which the masking was at a S/N ratio of approximately -10db, the substitutions suggested by Miller and Nicely were selected, with the condition that each involved only one phonetic feature, either coronal or anterior. The stimuli comprised sixteen proverbs at 10 sec intervals, each of which contained one mispronunciation. Each S was tested with 8 proverbs in each condition. As there were no order effects only pooled data is presented.

CONDITION	MISPRONUNCIATIONS DETECTED
Vis	4.6(3-7)
NVis	0.4(0-1)

TABLE VII: Mean number of mispronunciations detected (and ranges) when elliptic speech is presented at S/N=-10db.
Max score=8, n=8.
More mispronunciations were detected with Vis presentation.
d=0, m&n=8, $p<0.001$ one sided

Miller and Nicely commented that elliptic speech and normal speech 'sounded just the same' when heard under the appropriate conditions. The

NVis condition is thus a replication of their result and the Vis condition a demonstration that elliptic speech and normal speech do not sound the same when the speaker is visible.

#3.3.3 EXPERIMENT VIII: PHONEMIC RESTORATION EFFECT

Warren (1970) reported that when portion of an acoustic signal is erased and a 'noise' inserted, Ss tend to report having heard the original utterance and often report the intruding noise as having occurred at another point in the utterance. Warren and Sherman (1973) found further that when the erased acoustic energy corresponded to a mispronunciation, Ss reported perceiving the correctly pronounced form. The wholistic hypothesis predicts that this will only be the case for NVis presentation

Six sentences such as 'This letter arrived with the first p(k)ost this morning', each of which contained a single mispronunciation involving one of the features, coronal or anterior were prepared, and a 140 msec buzz excised from a cough used to mask the mispronunciation at S/N of approximately -6 db.

Results are shown in Table VIII.

CONDITION	MISPRONUNCIATIONS DETECTED
Vis	5.5(4-6)
NVIS	0.8(0-2)

TABLE VIII: Mean number of mispronunciations detected (and ranges) in the Phonemic Restoration Effect.
Max score=6, n=6 in each condition.
More mispronunciations were detected with Vis presentation.
d=0, m&n=6, $p<0.01$ one sided.

Again, the statistical difference was significant. In this case, however, the phenomenological report was not as straightforward as

usual. Some Ss who were tested in the Vis condition were not certain whether or not they had actually heard all of the mispronunciations. They could, however, locate the buzzes exactly. This suggests the interesting possibility that Ss will be able to locate buzzes and clicks more accurately when the speaker is visible. On the other hand, some Ss who were tested in the NVis condition and then informally shown the tape with Vis presentation would not believe that the acoustic signals were identical.

The main result of this experiment is, however, that the Phonemic Restoration Effect is not dependent only on audition and cognition; vision is clearly also involved.

#3.3.4 EXPERIMENT IX: BINAURAL SHADOWING

Cherry (1954) reported that binaural shadowing of one of two messages spoken by the same speaker is extremely difficult. The obvious wholistic hypothesis is that shadowing will be much easier when the speaker is visible.

To test this hypothesis, E read passages from two books onto the two auditory channels of a videotape, the optical track corresponding to one of the auditory channels. Instead of shadowing, Ss were required to write down what they had 'heard' whenever there was a pause in the auditory tracks. These pauses occurred at convenient points in the shadowed message. Each excerpt contained approximately ten words. The number of correctly recorded words was then scored. Average correctly recorded words are shown in Table IX.

This experiment is an independent replication of Summerfield (1979) and yielded comparable, although less striking results. Summerfield

presented his experiment as more naturalistic than masking with white noise; this experiment was conceived as a Vis beneplication of binaural shadowing. The lower scores for the Vis condition may be due to the length of the message to be recalled and the fact that the two messages often started together, whereas Summerfield allowed the unshadowed message to run for a time before beginning the shadowed message. This possibility is supported by the comment made by some Ss in the Vis condition that they could hear distinctly what had been said, but could not remember it. As memory is clearly involved in this experiment, it is not as easy to evaluate as the others. Certainly, however, Ss confirmed that the message 'sounded clearer' when the speaker was visible.

CONDITION	CORRECTLY RECORDED WORDS
Vis	59 (38-80)
NVis	33 (22-47)

TABLE IX: Percentage correctly recorded words (and ranges) for Vis and NVis shadowed messages.
More words are recorded with Vis presentation.
d=4, m&n=11, $p<0.001$.

#3.3.5 EXPERIMENT X: FILTERED SPEECH

The three previous experiments have demonstrated that speech perception and not just word recognition is aided by the speaker's being visible when there is a masking noise. The present experiment seeks to establish that this is also the case when the original utterance is attenuated by being filtered.

This experiment is identical to Expt VII, with the sole exception that the proverbs were filtered to remove frequencies above approximately 750Hz. Mean error detection scores are shown in Table X.

As is clear from the table, the detection of mispronunciations is

significantly easier when the speaker is visible. Again, however, the data do not capture the effect. With NVis presentation, the mispronunciations were almost undetectable, even with repeated presentation; with Vis presentation Ss often commented that the mispronunciations seemed to 'jump out at you.'

CONDITION	MISPRONUNCIATIONS DETECTED
Vis	6.0 (5-7)
NVis	0.2 (0-1)

TABLE X: Mean mispronunciations detected (and ranges) for Vis and NVis presentation of proverbs.
Max score=8. n=11.
More mispronunciations were detected with Vis presentation.
d=0, m&n=11, $p<0.001$ one sided.

Experiment XII contains an even stronger demonstration of the involvement of vision in the perception of filtered speech. An adaptor which consisted of twelve repetitions of 'two ones are two, two twos are four' in which each 't' had been mispronounced as 'p' was perceived Vis-ually as 'poo ones are poo etc' and NVis-ually as 'two ones are two etc' by nine Ss, without any weakening during five presentations of the adaptor. In this case, Ss were not listening for mispronunciations.

#3.3.6 EXPERIMENT XI: NATURAL SPEECH

The mispronunciations in the proverbs were so obvious that they were immediately apparent when the acoustic signal was unattenuated. Cole (1973) has reported that mispronunciations in fluent speech are difficult to detect. It was decided, therefore, to record short passages which contained mispronunciations which were extremely difficult to detect with NVis presentation, in order to test the hypothesis that even with entirely unmanipulated speech the speaker's being visible speaker can influence speech perception.

Twelve sentences were recorded, eight of which included mispronunciations of the final phoneme of an unstressed syllable of a three-syllable word. Examples are 'ab(d)normal' and 'naked(b)ness'. Again, all mispronunciations involved the features anterior or coronal. Mean detection scores are shown in Table XI.

CONDITION	MISPRONUNCIATIONS DETECTED
Vis	6.9 (3-8)
Nvis	0.4 (0-2)

TABLE XI: Mean numbers of mispronunciations detected in recorded natural speech.
Max score=8. n=8 in each condition.
More mispronunciations were detected with Vis presentation.
d=0, m&n=8, $p<0.001$ one sided.

Again, the difference in detection across the two conditions is significant. In this case, the phenomenological report was not so much that errors tended to 'jump out' at you, but rather that Ss felt a little surprised to notice the mispronunciation. All were adamant that they 'heard' them and 'heard' them clearly.

#3.3.7 CONCLUSION

These five experiments have demonstrated that the influence of vision on speech perception is not restricted to very abnormal cases in which the acoustic signal has been dislocated. For all five, statistically significant results were obtained, as were reports that the speaker's utterance sounded quite different when the speaker was visible. All of these experiments, however, still involve a distortion of the naturalistic case - deliberate mispronunciations. The naturalistic case appears to defy experimentation.

#3.4 DUBBING AND THE WHOLISTIC APPROACH

#3.4.1 INTRODUCTION

The experiments reported above had been planned and partially completed when the Roberts and Summerfield (1981) study became available. This study will reward detailed consideration because it is a precise, thorough investigation - described by Repp (1981) as elegant and ingenious - which exemplifies all of the ideals of current research and which led to conclusions which are diametrically opposed to the expectations of the wholistic insight. This will be especially important because dubbing appears to have become the principal technique in the audio-visual study of speech perception, and the foregoing discussion suggests that it will inevitably lead to inappropriate experimental findings being used to support future theorising.

#3.4.2 AUDIOVISUAL ADAPTATION

Several quotations from the study by Roberts and Summerfield will show that all of the presuppositions which were outlined in Chapter 1 inhere in their conceptualisation.

> "A wholly rigorous test of either claim would require that the acoustical structure of speech stimuli be dissociated from the phonetic percepts they engender. That is, it would be necessary to create stimuli that either possessed completely non-contiguous spectrotemporal specifications but produced the same phonetic percept or possessed identical acoustics but were perceived as belonging to different phonetic categories."

> "It may not be possible to isolate absolutely the acoustic and phonetic components of the adaptation process using purely acoustic stimuli, especially with regard to the dimension of place of articulation."

> "Nevertheless, with this technique, an utterance, although specified quite unambiguously in acoustical terms can be modified perceptually for the majority of observers without changing its acoustical structure."

All references to 'dissociation', 'components', 'engendering', 'noncontiguous spectrotemporal specifications', 'isolating' and 'specified quite unambiguously' in this context exemplify familiar presuppositions:- dividing what is not divisible, simple location and the substance-quality form of the fallacy of misplaced concreteness.

The use of the word 'dissociation' can be taken as a simple example. That a portion of acoustic energy can be associated with two or more phonetic percepts in no way dissociates it from any of them. Zero acoustic energy can be associated with many different phonetic percepts (Dorman et al, 1981); it is dissociated from none of them. Not any random acoustic signal would have 'engendered' the desired phonetic percept for Roberts and Summerfield; Samuel (1981) has studied in some detail what he calls the 'role of bottom-up confirmation' in the phonemic restoration effect. Thus, in no sense at all was a dissociation between acoustic signal and phonetic percept achieved. It would appear that the term would only be applicable in the case of auditory hallucinations being experienced quite independently of simultaneous auditory stimulation. The other terms which were mentioned are equally obviously expressions of fragmentation.

Ades (1981) attributes the prevalence of non-phenomena to the legacy of Behaviourism, for which there is no essential difference between a phenomenon and a non-phenomenon. Correspondingly, for this study, it is not considered significant or surprising that conclusions about speech perception are drawn on the basis of synthesised (unutterable) speechlike sounds which were dubbed more or less accurately onto simulated conflicting articulatory gestures (mouthed syllables). For the wholistic approach, the gap between such experimental conditions and natural speech precludes the drawing of any conclusion at all about

speech perception. In this context, the reference to synchronising synthesised syllables and conflicting articulatory gestures is instructive. Obviously, it is impossible to synchronise conflicting acoustic and optical signals. What McGurk and MacDonald (1976) showed is that it is possible to temporally align two such signals to yield a third phonetic percept, but this is not synchronisation.

Even within the fragmentary mode, it is clear that the authors' conclusion is premature. Several specimen reasons for this are-

1 No audiovisual test syllables which may well have shown an adaptation effect were used, even though Summerfield (1979) had shown that this could be done.

2 Roberts and Summerfield comment that their purely optical adaptors Vb and Vd produced no measurable adaptation and state that,

> "Our conditions Vb and Vd have shown that lipreading does not tap whatever aspects of acoustical speech perception there are that might be influenced by graphical presentation. The result suggests that further searches for such effects would not be rewarding."

Further analysis of their published data reveals that the mean adaptation effects for Vb and Vd presentation, although very small and not significant, were at least in the predicatble direction and, more importantly, that the variances, especially for Vb, were larger than for Ab and Ad. This seems to accord well with Cooper's (1979) findings. Further, although the variance for Vb is not significantly greater than for Vd, it is sufficiently large to suggest that the difference may be a real effect. This leads immediately to the prediction that the variance with AbVb presentation will be greater than with AbVg adaptation. This prediction is confirmed ($F(11,11)=4.13$, $p<0.02$). This shows that audio-visual adaptation is not exclusively auditory. It is, of course, not necessary for vision to reverse the direction of the adaptation

effect for it to be involved in it.

3 The rejection of the existence of two sites of phonetic processing is also not convincing. The inadequacy of McGurk and MacDonald's (1977) specific two site hypothesis (their manner-place hypothesis) is no argument for the rejection of all such hypotheses. The Erber and De Filipo (1978) experiment with its buzzes and simulated faces and its absence of reference to Ss conscious perception is not directly relevant.

In summary it can be said that this study, in spite of being audio-viual, is prototypically fragmentary in its conceptualisation; that the experimental hypothesis was not fully tested because of the absence of audio-visual test stimuli; that the published data show an audio-visual adaptation effect; and that the argument against two sites of phonetic adaptation is faulty.

However, even if the conclusion that adaptation is auditory rather than phonetic was unjustified, the data would still appear to support the claim that the mean adaptation effect of audio-visual adaptation is attributable entirely to auditory adaptation. It could perhaps be argued that the visual component served as a distractor and that therefore the increased variance is not related to adaptation as such. Should this be true, as it may well be, it would be another strong argument against dubbing, as a comparison with Expt VIII will show. During binaural shadowing, S finds the two spoken messages too much to cope with and performance is poor. From the fragmentary viewpoint it would be expected that a third channel would further overtax S; from a wholistic viewpoint it is to be expected that if less of the speaker's activity is concealed, shadowing will be easier. This is the case. Whereas seeing dubbed lip movements may be a distraction, seeing the speaker is not.

Nevertheless, the mean adaptation effect was equal for Vis and NVis presentation of the identical acoustic signal; and that, although not inconceivable, is vanishingly unlikely to wholistic thinking. If the speech event is a unified whole, it is difficult to conceive of one aspect independently producing adaptation effects. An obvious first thought is that dubbed acoustic and optic signals really are independent of each other, even though they may intersect in the phonetic percept, and so it is not at all surprising that they should "fall apart" in their effects on the perceiver. (Similarly, it is possible that the unutterable synthetic stimuli are not unitable with human lip movements.) From this it would follow that a beneplication which employed neither dubbing nor synthetic syllables would be expected to show audio-visual adaptation to be audio-visual. The experiment now to be reported does just that.

#3.4.3 EXPERIMENT XII: ADAPTATION

Rudnicky and Cole (1977) demonstrated that connected prose can serve as an adaptor. Because of this, Expts VII and IX, ie filtering and masking of elliptic speech, offer two techniques to construct contrinsic adaptors which are perceived phonetically differently across Vis and NVis presentation. It was found that repetitions of 'Two ones are two, two twos are four' in which each /t/ had been replaced by /p/ was perceived quite differently across conditions. NVis-ually presented, it was perceived as the familiar arithmetical table; Vis-ually presented, its mispronunciations were laughably apparent. Twelve repetitions of the message lasting altogether 35 secs served as the adaptor.

There appear to be no references to adaptation effects in the perception of natural or attenuated speech. As the confusion matrices published by

Miller and Nicely (1953) suggest that filtered speech may indeed be susceptible to adaptation, it was decided to employ the syllables /pu/, /tu/ and /ku/ as test stimuli with the adaptor which has already been described.

A set of 25 syllables, 10 each of /pu/ and /tu/ and 5 of /ku/, was recorded with interstimulus interval of 6 secs. These syllables served as the pre-adaptation test syllables and in five sets of five during post-adaptation testing. Test syllables and adaptor were filtered to remove frequencies above approximately 790 Hz. A sample syllables is shown below.

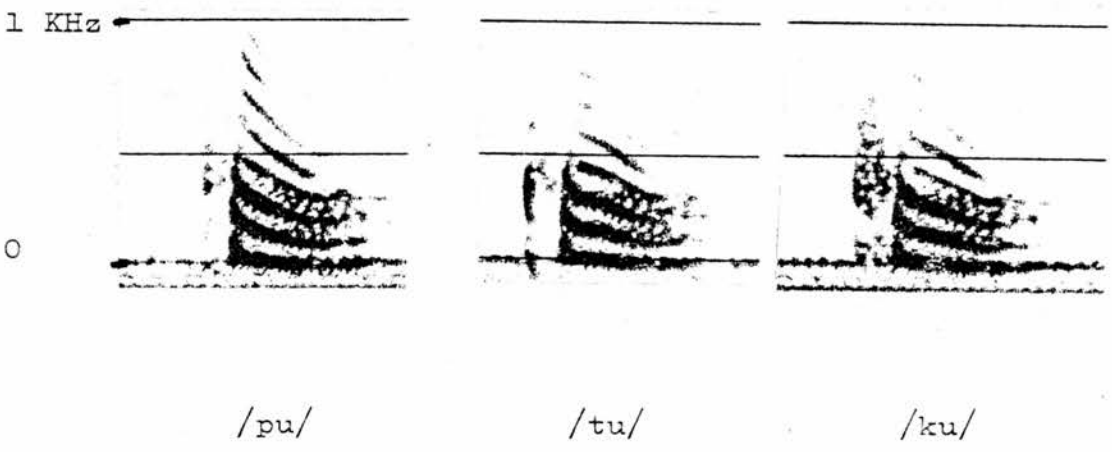

FIGURE XII: Spectrograph of /pu/, /tu/ and /ku/ test syllables which were filtered with a Barr and Stroud Variable Filter set at 790 Hz.

Before trying to predict the possible adaptation effects for the various test syllables, E ran a trial with an enlarged set of 50 test syllables and himself as S. Because the test syllables do not lie along a continuum in the normal manner, the experimental hypothesis will be that identification of the test syllables will differ across the two adaptation conditions. Results are shown in Table XII.1.

TEST SYLLABLE	RESPONSES								
	Pre-Ad			NVis			Vis		
	p	t	k	p	t	k	p	t	k
/pu/	19	1	-	4	16	-	19	-	1
/tu/	3	13	4	-	19	1	11	6	3
/ku/	2	-	8	1	1	8	-	-	10

TABLE XII.1. Responses of one subject in each of the three conditions: pre-adaptation, Vis- and NVis-adaptation.
Correct identifications of both /pu/ and /tu/ syllables differ across adaptation conditions. In each case, $p<0.001$, two sided Fisher's exact probability test.

Correct responses to /pu/ and /tu/ test syllables both differ across the adaptation conditions. Although, there are other significant differences in the data, these two appear to be suitable for testing as the adaptor was prepared by mispronouncing /tu/ as /pu/. As will be reported below, both of these differences were significant with naive Ss. The specific experimental hypotheses will be that there will be more correct 'p' responses after Vis adaptation and more correct 't' responses after NVis adaptation. The main experiment will be described in two parts; a short demonstration and a somewhat more detailed study. The demonstration follows from Es experience that the first test syllable of each set was clearly perceived, but later syllables sometimes produced doubt about the accuracy with which earlier syllables had been identified. This led to the radical hypothesis that a single presentation of a single test syllable could show that audiovisual adaptation is not purely auditory.

First the demonstration. After a period of familiarisation with the test syllables (Vis-ual presentation of twelve syllables for both conditions and a pre-adaptation set of 25 syllables for the NVis group), Ss were presented with the adaptor and the test syllable /pu/. Nine Ss were tested in each condition; testing being done in groups of two or three. Results are displayed in Table XII.2.

	RESPONSES ADAPTOR					
SYLLABLE	VIS			NVIS		
	p	t	k	p	t	k
/pu/	9	-	-	1	6	2

TABLE XII.2: Responses to a single /pu/ test syllable following Vis and NVis adaptation.
n=9 in each condition.
There were more 'p' responses after Vis adaptation.
$p<0.001$, Fisher's exact probability test, one sided.

It has now been demonstrated that audiovisual adaptation is not purely auditory.

Now the more detailed study. The NVis group was presented with the full set of NVis-ually adapted test syllables, and then after a pause of five minutes with the same syllables Vis-ually adapted. Testing of the Vis group was terminated after the presentation of the first set of five test syllables. All responses for the five test syllables which were presented to both groups are shown in Table XII.3.

	RESPONSES ORDER OF PRESENTATION								
SYLLABLE	NVIS-VIS						VIS		
	NVIS			VIS					
	p	t	k	p	t	k	p	t	k
/pu/	1	6	2	8	1	0	9	0	0
/ku/	1	2	6	0	5	4	2	5	2
/pu/	3	4	2	6	2	1	6	2	1
/tu/	1	7	1	2	5	2	3	3	3
/ku/	-	4	5	1	2	6	2	2	5

TABLE XII.3: Responses of both experimental groups to five test syllables.

The similarity between the responses of both groups after Vis adaptation is striking. There is, however, an indication that the adaptation effect of Vis adaptation is smaller in the experimental group which had already been subjected to NVis adaptation.

Full results for the NVis-Vis group are shown in Table XII.4.

STIMULUS	RESPONSE								
	PRE-AD			NVIS			VIS		
	p	t	k	p	t	k	p	t	k
p	37	24	29	22	49	19	60	16	14
t	15	18	57	9	41	40	27	18	45
k	8	7	30	3	16	26	7	12	26

TABLE XII.4: Responses for 9 Ss in each of three test conditions. Each S received 10 presentations of /pu/ and /tu/ and 5 of /ku/ in each condition.

As little or nothing is known about the adaptation of filtered speech and as the confusion matrices in Miller and Nicely (1953) display a great variety - even for /p/, /t/ and /k/ - depending on the exact level of filtering and masking, there are no standards with which to compare these results. However, both hypotheses were confirmed as shown in Table XII.5.

STIMULUS	CORRECT RESPONSES CONDITION	
	VIS	NVIS
p	6.7	2.4
t	2.0	4.6

TABLE XII.5: Correct responses to /pu/ and /tu/ test syllable after Vis and NVis adaptation.
Max score=10, n=9.
Both adaptation effects are significant.
For 'p' stimuli:
d=10.5, m&n=9, $p<0.005$, one sided for Ss;
d=6, m&n=10, $p<0.001$ one sided for stimuli.
For 't' stimuli:
d=13.5, m&n=9, $p<0.005$ one sided for Ss.
d=14, m&n=10, $p<0.005$ one sided for stimuli.

The size of the adaptation effect decreases within each set. If just the initial test syllable of each of the five sets are considered and then compared with the other four a clear picture emerges. This is shown in Table XII.6. As nothing is known about the time-course of this adaptation effect, it is not possible to assess how these differences are related to time and interactions between different stimuli. For

present purposes, this is not important.

	AVERAGE /P/ RESPONSES			
	NO	VIS	NVIS	ADAPT EFFECT
Init p	2	8.0	1.5	6.5
Non-Init p	8	5.5	2.4	3.1
Init t	3	5.0	1.3	3.7
Non-Init t	7	1.7	0.7	1.0

TABLE XII.6: /p/ responses for group initial /pu/ and /tu/ test syllables.
For both test syllables, adaptation towards /pu/ is greater with group initial syllables.
For /pu/ test syllables:
d=0, n=2, m=8, $p<0.05$ one sided.
Fpr /tu/ test syllables:
d=2, n=3, m=7, $p<0.05$ one sided.

Interestingly, it is almost certainly incorrect to describe this experiment as a study in audio-visual adaptation; it is almost certainly a study in audio-visuo-cognitive adaptation as it is probable that a non-english-speaker would not have perceived the audio-visual adaptor as an excerpt from the two-times table and may well have adapted differently. Such a possibility is, of course, grist to the wholistic mill. Any suggestion that audio-visual adaptation is somehow more basic than audio-visuo-cognitive would merely be an expression of fragmentation.

The most important conclusion to draw from this experiment is that audio-visuo-cognitive adaptation is not exclusively auditory, and thus that a second-degree beneplication of the Roberts and Summerfield experiment has shown that it fails to replicate towards the naturalistic case.

#3.5 OVERVIEW

This chapter has been intended as a wide ranging demonstration of the

efficacy of the wholistic approach. The fundamental wholistic hypothesis as applied to speech perception has been shown to generate testable (and confirmed) hypotheses, and it is now reasonable to assert that when brought to bear on any fragmentary experiment it will yield a testable prediction.

The principle of replication towards the naturalistic case (beneplication) was proposed as a research strategy because, although open on the fragmentary-wholistic issue, it led to experimentation which continually tested the wholistic approach. In this respect, most current research is irrelevant. In view of the uniform success of the wholistic approach, it is necessary to recognise that the presuppositions which Whitehead as a philosopher and Bohm as a physicist detected and rejected must now be rejected within psychology.

In #3.2, the fundamental wholistic hypothesis was applied to experiments with dislocated acoustic signals and showed them to be incapable of beneplication and thus unable to bear the weight of theorising which has been placed on them. The wholistic approach can be seen to clarify even the most fragmentary experiment. The little study which was described in #3.2.8 is a first demonstration that the difference between Vis and NVis presentation goes beyond the perception of single phonemes. Whether or not an event is even perceived as speech can be influenced by vision.

The five experiments of #3.3 have shown that the striking involvement of vision in speech perception is equally clearly demonstratable with masked, attenuated and unmanipulated speech. As a consequence, it was clear that psychological research in speech perception must be at least audio-visual and that dubbing is a psychologically inadmissible technique.

#3.4 brought the full confirmation of this apparently radical statement. The clear results obtained in a dubbing experiment were shown clearly not to apply in a more naturalistic case. The added probability that adaptation was not only not auditory, but also not audio-visual because it was at least audio-visuo-cognitive was an extra bonus. The apparently inevitable lengthening of hyphenated words indicates that a reorientation is called for.

Speech research needs to orientate itself with respect to the naturalistic case. Ades (1981) expressed this as clearly and as simply as possible:

> "The point of natural science is to take a naturally occurring phenomenon under some characterisation and then try to understand it."

Ades attempts to explain why such an obvious point is disregarded:

> "One possibility is that natural phenomena are more complex than the non-phenomena beloved of experimental psychology. The lure of the non-phenomenon lies in its tractability. One has a sense that natural phenomena are, relatively speaking, either understood or not: progress is by sudden leaps of insight. Whether or not this is true, the belief that real phenomena are too difficult to study is widespread."

In Psychology, only the real phenomena are worth understanding. As the experiments of this chapter show, the non-phenomena contribute nothing to the understanding of the real phenomena; all else is artefact. Orientating thought relative to the naturalistic case (the real phenomena) reveals the irrelevance of the fragmentary experiments (the non-phenomena) and constitutes a first step towards the natural science of Psychology. Every other orientation can only introduce yet further forms of fragmentation (Bohm, 1980).

CHAPTER 4: CANONICAL SPEECH PERCEPTION

#4.1 REPLICATION TOWARDS THE NATURALISTIC CASE AS A RESEARCH STRATEGY

Although the primary purpose of the preceding chapters was to identify some of the prevailing presuppositions in speech research as expressions of fragmentation and demonstrate that an examination of them indicates that a more wholistic approach is called for, a secondary aim was to establish the value of the principle of replication towards the naturalistic case as a research strategy. Although devoid of all empirical content, this principle is an expression of commitment to the naturalistic case, and as such readily suggests experiments which relate to it. The earlier discussion of dubbing is a powerful demonstration of the principle's efficacy; the principle suggested a clear, easily tested prediction which was contrary to a fragmentary result and was subsequently confirmed. The remaining experiments show that the principle can be applied with great generality.

In addition to the earlier conclusions which related specifically to speech perception, a further more general conclusion appears to be indicated:

NO THEORISING CAN BE SUPPORTED BY THE FINDINGS OF EXPERIMENTS WHICH DO NOT REPLICATE TO THE NATURALISTIC CASE. REPLICATION TOWARDS THE NATURALISTIC CASE THUS CONSTITUTES A MINIMAL RESEARCH STRATEGY FOR PSYCHOLOGY.

#4.2 THE CANONICAL CASE

Mention was made in Chapter 2 of complications within the naturalistic case, for example lack of concentration on the part of speaker or listener. This suggests that another focus is required and that the

naturalistic case is only a transitory focus for research. The simplest and purest form of speech perception will be called the canonical case.

In the canonical case, a capable fully-involved listener perceives the speech of a capable fully-involved speaker under ideal conditions. The nature of the canonical case is of course largely unknown:-

- canonical speech sounds may be essential. This is perhaps related to the question of phonetic prototypes (Samuel, 1982).
- it may be impossible for a listener to attend fully to speech sounds while also attending fully to content.
- it is not even known which senses are implicated; some researchers have also suggested touch (Erber and De Filipo, 1978); there has been considerable discussion of a 'speech mode' (Morton and Chambers, 1976; Liberman and Pisoni, 1976; but see Schouten, 1980); the relevance of the live speaker has also been mentioned (von Raffler-Engel et al, 1980). All of these modes will need to be investigated. It is obvious that, in any common sense of the term, touch is not involved in speech perception, but the fact that it can be involved demands consideration as it promises to provide insight into the relationships between the various sensory modalities.

In spite of all of this ignorance, it is essential to recognise that the canonical case is the simplest case and logically prior to the naturalistic case and all experimental findings. A commitment to the canonical case, even more than to the naturalistic case, helps to ensure that no presuppositions are allowed to restrict the enquiry.

#4.3 THE LOGICAL PRIORITY OF THE CANONICAL CASE

Although the canonical case will be extremely difficult to attain,

understanding it is a precondition to understanding the naturalistic case and all experimental settings. In any non-canonical case, unknown, uncontrolled and uncontrollable influences will be operative; their presence renders the non-canonical case ununderstandable by itself. Obviously, it will always be possible to obtain certain behavioural measures such as word recognition scores which may suffice for many practical purposes, but it will not be possible to attain understanding. The history of so many 'parsimonious' theories which profligate under the pressure of new findings reveals the difficulties which must confront all non-canonical studies.

The fact that the canonical case is difficult to attain, and may be beyond many speakers and listeners, is no argument against its logical primacy. This is unassailable.

#4.4 FUTURE RESEARCH

The most general formulation of a research programme must be to approach the canonical case. Much is still possible, and perhaps necessary, within the prevailing framework:-

1 It may be desirable to extend the range of phenomena in which it has been established that NVis experimentation does not replicate to the Vis case. It may also be desirable to demonstrate the applicability of the principle of replication towards the naturalistic case to areas other than speech research.

2 The converse of this is to attempt to discover whether the effects described here are specific to speech. Campbell and Dodd (1980) consider the possibility that their findings are not specific to speech, but are rather aspects of movement generally. This very interesting thought, which can be interpreted as calling for a reconsideration of how

psychologists partition the world, will be considered further in the following chapter.

3 It has been shown in Chapter 3 that differences in lip movement as fine as that between articulations of 'file' and 'vile' are perceptually potent. It would be helpful to test the fundamental articulatory hypothesis that every utterance has a characteristic gesture and all articulatory gestures are perceptually potent. A sufficient variety of filtering and masking procedures would allow this hypothesis to be subjected to extensive investigation.

4 It will be necessary to extend the scope of investigation beyond audio-visual presentation. One extension will be to study audio-tactile presentation, not in the style of Erber and De Filipo (1978), but by allowing the perceiver to touch the speaker's lips. Pilot studies indicate that touching the speaker's lips does indeed influence conscious speech perception. Combining filtering or masking with allowing S to touch the speaker's lips allows great scope for experimentation. This form of investigation will be important, not because listeners often touch speaker's lips in daily life, but because it will extend the range of ecologically coherent (intrinsic) "inter-modal" effects. It will be important to study "inter-modal" effects with ecologically coherent stimuli because it can now be predicted that for all combinations of sensory modes, results obtained with extrinsically related signals will not beneplicate to the intrinsically related case.

5 A step towards the naturalistic case could be to replicate standard experiments, including those described in Chapter 3, with a live speaker. Various forms of background noise and speaking through various types of transparent material should allow all experiments with contrinsically related stimuli to be beneplicated to the case of the

live speaker. As has already been argued, those which do not beneplicate to the contrinsic case do not call for beneplication to the case of the live speaker.

6 The main task is to develop means to conduct psychological experiments with the unattenuated speech of a live speaker. The experiments suggested in the previous paragraph can only be seen as a step towards this goal. All current experimentation involves manipulating the acoustic signal or its relationship to the optic signal; current presuppositions do not appear to encourage or even offer the possibility of such research. This point will be considered again in the following chapter.

#4.5 THE FUNDAMENTAL WHOLISTIC INSIGHT

The foregoing discussion in this chapter and the previous experimental chapter has been intended to demonstrate that wholistic insights can be fruitful within the fragmentary approach - all of the experiments of Chapter 3 are recognisably fragmentary experiments. Further, it was intended to show that the fruitfulness of the wholistic approach inheres in its commitment to the naturalistic case. This commitment results in an improved orientation which reveals the flaws and presuppositions of the fragmentary approach. The fragmentary approach appears to lead inevitably to manipulation of ever smaller, ever less realistic stimuli (see Bohm's (1980) discussion of atomism and Ades' (1981) discussion of "non-phenomena"). The wholistic approach with its principle of beneplication demonstrates the inapplicability of these stimuli to the understanding of the naturalistic case. This insight is apparent throughout J J Gibson's work. In the introduction to his final book (Gibson 1979), he makes the following comments, which establish the same point with respect to vision. After describing standard experiments in

terms of snapshot and aperture vision, he characterises ambient and ambulatory vision as the forms of natural vision which need to be understood. His commitment to the naturalistic case is succinctly stated:

> "It is not true that the laboratory can never be like life. The laboratory must be like life."

For understanding of the naturalistic case:

> "The vast quantity of experimental research in the textbooks and handbooks is concerned with snapshot vision, fixed-eye vision, or aperture vision, and it is not relevant."

An important corollary of commitment to the naturalistic case, for Gibson the ecological case, is found in the overall plan of the book:

> "Picture vision comes last because it can only be understood when ambient and ambulatory vision have first been understood."

Turvey et al (1981) capture a similar point:

> "It is only when the ecologically relevant measurement principles are developed that we will ever be able to comprehend what went on in these sorts of studies (ie those which study illusions or use ecologically uninteresting or unrepresentative displays) in the first place."

The really fundamental wholistic insight, however, is that consistent commitment to the naturalistic (ecological) case leads to commitment to the canonical case.

As has been stated above, the naturalistic case defies experimentation. Researchers are bound to either create and study artefact or approach the canonical case. Gibson et al's (1969) delicate little study of objects going out of sight or out of existence is a model in another area of a gentle approach towards the canonical case. This is a rewarding instance of improving perceptual judgement.

#4.6 SUMMARY

It has been shown in Chapter 1 that the presuppositions which Bohm and Whitehead regard as typical of current scientific thinking also prevail within speech research. These presuppositions can be characterised as expressions of fragmentation. In fact, even the almost universal acceptance of the obvious "truth" that speech is heard was shown to be a specific expression of fragmentation. In Chapter 2 it was shown that acceptance of these presuppositions can be queried within Psychology. Indeed, simply orientating thought relative to the naturalistic case indicates strongly that a more wholistic approach is called for. A very general wholistic hypothesis was formulated and a research principle established which, although open on the fragmentation-wholeness issue enabled the wholistic hypothesis to be tested experimentally within the fragmentary mode of experimentation. The experiments described in Chapter 3 all confirmed the wholistic hypothesis. This now requires that for psychological research in speech perception the naturalistic case be regarded as the point of orientation. When this is accepted, much experimentation with manipulated and synthetic acoustic signals is seen to be "not relevant"; all purely auditory experimentation is shown to constitute an impoverishment of the naturalistic case, not a simplification; all audio-visual research based on extrinsically related acoustic and optic signals is recognised as compounded fragmentation and not a step towards wholeness. The obvious and necessary extension of audio-visual experimentation in which the speaker is visible will be to conduct experiments in which the speaker is present. In the earlier sections of this chapter it was shown that commitment to the naturalistic case entails commitment to the canonical case. This is the fundamental wholistic insight. The wholistic insight will continue to be

able to design experiments within the fragmentary mode which will help to order the data, but the primary task of these experiments will be to help prepare for research in the wholistic mode. A few preliminary thoughts will be sketched in the following 'Outlook'.

CHAPTER 5: OUTLOOK

#5.1 INTRODUCTION

The foregoing discussion, conducted within the fragmentary approach, has provided a strong argument that prevailing presuppositions lead to misformulations of basic questions. It is being recognised increasingly that experimental results alone do not determine researcher's theoretical positions; something deeper is involved. Liberman (1981) notes this with exasparation:

> "But the auditory theory is not so easily disposed of, because it can always fall back on the assumption that ... It matters little that there is nothing in what we know about the perception of complex sounds to suggest that ... Nor does it necessarily matter how implausible it is to suppose that ... Such considerations make an explanation based on auditory interaction endlessly ad hoc, but they do not, in principle, rule it out."

So does Norman (1978):

> "Forget the behaviourists ... You cannot prove the existence of mental events to the behaviourist, and you need not prove it to the mentalist."

Schouten (1981) displays greater equanimity:

> "Whether or not one believes in the speech mode is just that: a question of belief."

But the message is the same. James (1907) offers an explanation. He is referring explicitly to philosophers, but what he says applies much more generally.

> "Of whatever temperament a professional philosopher is, he tries, when philosophising, to sink the fact of his temperament. Temperament is no conventionally recognised reason; so he urges impersonal reasons only for his conclusions. Yet his temperament really gives him a stronger bias than any of his more strictly objective premisses ... in the forum he can make no claim, on the bare ground of his temperament, to superior discernment or authority. There arises thus a certain insincerity in our philosophic discussions: the potentest of our premisses is never mentioned."

Several aspects of our "potentest premisses" will be sketched briefly in this Outlook because they help in the recognition of the most fundamental presupposition of current scientific thought.

#5.2 MINIMAL INTROSPECTION

Many, indirectly all, academic disciplines are based on what can be called minimal introspections. Thus, optics and acoustics have their origin in our ability to experience colour and sound. No matter how much these disciplines may reject introspection, disregard subjective experience and adopt mathematical and physiological language, their origin lies in the naive experience of colour and sound; that is, in an almost universally available minimal introspection. Clearly, faulty minimal introspections will lead to faulty formulations of problems.

Speech research rests on the minimal introspection that speech is 'heard', and this has been a minimal introspection of probably millions of people throughout thousands of years. And yet it is wrong. The experiments of Chapter 3 constitute an extended demonstration of the faultiness of one of our simplest introspections. One such instance casts doubt on all else.

In the present case, it proved to be quite straightforward to create a situation in which an unusual introspection established the faultiness of the common minimal introspection, and was still available for all. Difficulty arises for the scientific community when an important introspection is not minimal.

An example of the far reaching revisions required by the recognition of a new introspection is provided by Weiskrantz (1978). Finding Ss who could locate spatial positions and identify forms without any visual

awareness led him to conclude that:

> "The whole domain of visual field defects associated with brain damage, which has been thought to be more or less a closed book since Gordon Holmes's classical studies during the First World War, must now be studied almost from scratch."

This example is not felt to be so problematic because the Ss were brain damaged and their introspections were less extensive than those of their investigators. The difficulty arises when the investigator cannot share Ss introspection.

#5.3 TACIT INSIGHTS

Polanyi (1967) has offered a challenging attempt to understand the nature of our potentest premisses. His concept of "tacit knowledge" is a form of conceptual equivalent to the perceptual minimal introspection. J J Gibson is an excellent example of a researcher recognising the power of a new tacit insight. His commitment to monism led him to the insight that Psychology requires an ecological optics (Gibson, 1961) rather than physical optics and to the insight that the truth of ecological optics would compel a revision of sensory physiology (and much else). As he noted (Gibson, 1973):

> "Anyone who considers the senses to be channels of sensation has to be a mentalist when it comes to perception. That is, he has to assume a mind that can copy, store, compare, match, decide, and issue commands - a man in the brain.
>
> The whole idea of sensory signals and motor commands is wrong, together with the psychology that goes with it. The brain does not receive messages nor does it issue orders. Even the use of the terms sensory and motor is mentalistic. The concept of a perceptual system is in sharp contrast to the old notion of a sensory channel inasmuch as the brain takes its proper place as part of the system and is no longer the seat of the mind."

These few examples demonstrate that minimal introspections and tacit insights are amongst our potentest premisses.

#5.4 SPEECH AS AN EXPRESSION OF THE MIND

It is apt that speech research should yield such a clear demonstration of faulty introspection. What is perceived in speech perception is an expression of another mind; it is here that perceiver and perceived are of the same kind. Of course, the foot which is seen is an aspect of a person, but the speech which is perceived is a direct expression of his mind. In this sense, speech is special.

This is not a new or difficult introspection - related formulations by Summerfield (1979) and Repp (1981) have already been quoted - but it is one which is difficult to adhere to. The prevailing fragmentation appears to be incompatible with it, and so it is overlooked in research even by those who have expressed it.

Phonetic utterance is not the only means through which the mind is expressed. Gestures, facial expressions, prosody etc are related phenomena. In terms of physical energy, utterance and gesture are clearly distinct factors to be detected and combined; in wholistic terms, they are two aspects of an expression of the mind. Orientating thought relative to physical energy or relative to the naturalistic case leads inevitably to different ways of partitioning the world. Psychology's point of reference must be at least the naturalistic case.

#5.5 EVENT PERCEPTION

A central tenet of the event-perception approach is that the world is perceived, not merely its influence on the perceiver. This immediately raises all of the questions related to direct perception. As Gibson (eg 1976-77) has repeatedly shown, the fragmentary approach is incompatible with direct perception; within it, the simple "fact" of the spatial

separation between the brain of the perceiver and the perceived event necessitates that perception be mediated. His work with ecological optics, the nature of the stimulus, affordances and animal-environment complementarity constitutes an extended and incomplete attempt to confront this problem (Gibson: 1960,1961,1979).

Consciousness poses a problem for this approach which is reflected in Gibson's reference to consciousness as an "incidental" dimension of sensitivity (Gibson, 1963) and his statement:

> "So I have to admit that the study of sensations is important for an understanding of one's awareness of the self even when I deny that it is basic to an understanding of one's awareness of the world."

The quotations in #5.6 suggest that this distinction is even subtler than Gibson thinks it to be.

The approach being outlined here could be viewed as a form of event-perception, but a form in which it is stressed that the event to be perceived in speech perception is not sufficiently described when reference is made to movements of the articulators (Studdert-Kennedy, 1981; Summerfield, 1979) which structure optic and acoustic energy. The real event being perceived is one human being expressing something of his inner life.

Although it is true that current presuppositions cannot accommodate either direct perception or conscious perception, the potency of new (minimal) introspections and tacit insights is a constant reminder that current limits are not absolute. A short discussion of possibly the most deep-seated scientific presupposition will underline this point.

#5.6 THOUGHT AND PERCEPTION

Model making is the current mode of scientific 'explanation'. The

sensory data are grouped and manipulated according to concepts which are quite external to them. Wigner's (1969) formulation can almost be taken as a definition of the received scientific method:

> "Science gives us only a different view of these sensory data; it creates pictures from which they can be correlated in novel fashions. The primitive sensory data are the material with which science deals, which it orders and illuminates."

Wigner also expresses clearly that the scientific method, for all its instrumentation and formalism, is dependent on everyday knowledge (ie the prevailing presuppositions). Whatever equipment is used the scientist will at least be required to read a dial or examine a photographic plate etc, and this activity although unavoidable:

> "Is nevertheless an element which is foreign to the otherwise precise and clearly articulated framework of the theory."

All of this, and the further examples which Wigner mentions, merely constitute examples of the ubiquity of the driving power of minimal introspections. This point is underlined by Popper's (1972) statement that all observations are "theory impregnated". The theories which he refers to are often simply the prevailing presuppositions.

By now it is clear that there is no reason beyond habit and fleeting introspection to presuppose that concepts are essentially external to perception and have no power beyond creating pictures or ordering and illuminating the sensory data. A conceivable alternative is that we can become aware of the <u>one</u> world through perception and through thought. This formulation accords well with Gibson's (1966) (and Dewey's (1896) and Whitehead's (1926) and Kantor's (1978) and Bohm's (1980)) insight that the current presuppositions, in spite of their often being clothed in materialistic terms, are inescapably dualistic; they are a pernicious form of residual dualism.

Just as new minimal introspections were required to establish that

speech perception is not a purely auditory event, so new introspections will be required before it is possible to decide between the current dualistic presuppositions and the monistic view which holds that the world is accessible to us through both percept and concept, which taken together can first be viewed as a reality.

The crucial function of minimal introspections and tacit insights as potentest premisses in scientific research has been described. It will require introspections and insights which are equal to the canonical case to decide between the monistic and dualistic approaches. The psychology of speech perception will be quite different according to which is correct. The whole of this thesis is an argument in favour of monism.

Three final quotations (Steiner: 1960,1974,1962) give an indication of potential (minimal) introspections. The whole of this thesis is a preliminary to understanding them. The first provides a starting point for attaining insight into the new minimal introspection in speech perception which has been described here; the second and third offer insight into the extreme subjectivism of current scientific thinking (see #5.1 and any issue of The Behavioural and Brain Sciences) and can help in overcoming it.

> "In the hearing of human words and the understanding of them as thoughts a threefold activity is involved, and each component of this threefold activity requires separate consideration, if we are to conceptualise in a scientifically valid way. One of these activities is "hearing". But "hearing" per se is no more a "becoming aware of words" than "touching" is a "seeing". And just as it is proper to distinguish the sense of "touch" from that of "sight", so it is to distinguish the sense of "hearing" from that of "being aware of words", and again from that of "comprehending thoughts". A starveling psychology and a starveling epistemology both follow as consequences from the failure to distinguish the "comprehending of thoughts" from the activity of thinking, and to recognise the "sense" character of the first process." (Steiner, 1960).

"If cognition strives only to model what has been observed before cognitive activity, it attains not an experience of a full reality, but a picture of a half reality." (Steiner, 1974).

"The single individual is not actually cut off from the universe. He is a part of it, and between this part and the totality of the cosmos there exists a real connection which is broken only for our perception. At first we take this part of the universe as something existing on its own ... Whoever remains at this standpoint sees a part of the whole as if it were actually an independently existing thing, a monad which receives information about the rest of the world in some way from without. Monism, as here described, shows that we can believe in this independence only so long as the things we perceive are not woven by our thinking into the network of the conceptual world. As soon as this happens, all separate existence turns out to be mere illusion due to perceiving. Man can find his full and complete existence in the totality of the universe only through the experience of intuitive thinking. Thinking destroys the illusion due to perceiving and integrates our individual existence into the life of the cosmos." (Steiner, 1962).

APPENDIX 1: DETAILS OF EQUIPMENT

All equipment used in the experiments described in Chapter 3 is listed here. The numbers in brackets specify the experiments for which each piece of equipment was used.

I: PREPARATION OF STIMULI

CAMERA: i Hitachi FP 20 SK. Colour (I,II,III,IV,VI,VII,VIII, IX,X,XI&XII).

ii ARRIFLEX BL.

VIDEO RECORDERS AND TAPE: i Panasonic NV-9240 "U"-matic. (II,VI,VII, VIII,IX,X,XII).

ii Sony AV-3670 ACE 1/2" high density monochrome reel to reel EIA5 system (I,II,III,IV,VII,VIII,X,XI,XII).

TAPE RECORDER: ARRI-TANDBERG for lip-sync recording.

MICROPHONE: Uher M534.

EDITING TABLE: KEM Editing table plus headphones.

FILM: Ektachrome commercial, 24 frames per sec.

FILTER: Barr and Stroud Variable Filter Type EF2 (X,XII)

MIXER: Uher Stereo-Mix 500 Type A124 (VII,IX).

WHITE NOISE GENERATOR: Dawe Type 419C (VII).

II: ANALYSIS OF STIMULI

SONAGRAPH: i Kay Electronics Corp Digital Sonagraph 7800.

ii Kay Electronics Corp Spectrum Analyzer 7029A.

APPENDIX 2: PREPARATION OF STIMULI

In all cases, the speaker was filmed in full face so that he was visible from below his mouth to the top of his head.

EXPT I: The onset of the acoustic signal was located approximately by playing the tape at slow speed and noting the time trace at the onset of sound. Erasures were made by allowing the tape to pass slowly across the erase head with the recorder on record, but with the record head disconnected. The time trace and bar lines were used to control the extent of the erasure. Exact values were then determined by making sonagrams of the original and erased stimuli.

EXPT II: The stimulus tape was manipulated while being transcribed from one Sony AV-3670 to another. The audio level of the receiving recorder was held at zero and then quickly raised manually to comfortable listening level just as a 'blend' word was uttered. Again, exact values were determined when necessary by making sonagrams of the original and manipulated stimuli. The Vis-inverted condition was created simply by inverting the monitor; the Vis-restricted condition by attaching two sheets of paper to the screen of the monitor. To prepare the stimuli for the Vis-terminated condition, it was necessary to transcribe the erased stimuli onto a "U"-matic tape and then erase the appropriate portion of the video signal.

EXPT III: As in Expt I.

EXPT IV: As in Expt I.

EXPT V: The intersyllabic period of zero acoustic energy was located by playing the soundtrack at slow speed. In all cases five frames of the sound track were excised. To maintain overall synchrony, five frames of

the film were removed midway between stimuli.

EXPT VI: Using the "U"-matic editor, the fricative aperiodicity of an utterance of 'split' was transcribed to another tape; the original aperiodicity was then erased and the removed aperiodicity retranscribed to a position nearer to the remainder of the acoustic signal. As noted earlier, this technique also erased portion of the acoustic signal.

PILOT STUDY: Twenty repetitions of 'fah' at the rate of approximately two utterances per second were filmed, and a loop of 15 utterances (length 4.25 secs) constructed by joining two frames in which the speaker's lips were nearly closed. The sound track was joined correspondingly.

EXPT VII: The proverbs were recorded onto track 2 of a "U"-matic tape and broadband white noise was simultaneously recorded onto track 1. The two tracks were mixed while they and the video track were transcribed onto a reel to reel video tape. The two audio levels were set so that the S/N ratio was approximately -10db.

EXPT VIII: The stimulus sentences were filmed and recorded onto track 2 as before. The to-be-masked phoneme was located by playing the stimuli at slow speed. The masking stimulus which was an approximately 140 msec burst which had been excised from a cough was then edited onto the appropriate position on track 1. The two tracks were mixed during presentation of the stimuli.

EXPT IX: The to-be-shadowed message was filmed and the audio signal recorded onto track 2 of a "U"-matic tape. The masking message was then recorded onto track 1. The mixing was done as in Expt VII.

EXPT X: The same recording as described in Expt VII was passed through a filter to remove frequencies above 750 Hz while being transcribed onto a reel-to-reel tape.

EXPT XI: No manipulation was required. The sentences containing mispronunciations were recorded directly onto a reel-to-reel tape.

EXPT XII: The adaptor and tests syllables were recorded onto a "U"-matic tape. They were filtered with the filter set at 790 Hz (determined as the value at which E always perceived the filtered adaptor as 'two ones are two etc') while being edited onto a reel-to-reel tape.

APPENDIX 3: SUBJECTS AND DATA

All Ss were students or staff of the University of Edinburgh. All were volunteers in the sense that they did not refuse.

In chapter 3, all responses were presented in concise, sometimes grouped form. All instances in which an actual response did not accord exactly with its classification will be listed here. As an example, in Expt I, two NVis 'mend' responses were tabulated in the 'bend' category. This will be recorded as follows:-

NVis bend - mend(2).

EXPT I: Ss: 5 male, 5 female (20.0-24.0).
Vis
blend (breach)
br.. - bw.. (3); b... - p... (7)
file (fear)
ear (isle) - year (1).
NVis
blend (breach)
b... - p... (3), - m... (2); r... - fr... (1), - h... (1).
l... - th... (3), - d... (1).
file (fat)
b... - p... (3); v... - th... (3); at - hurt (1); v... - w... (1).

EXPT II: Ss: 7 male, 14 female (19.6-44.9).
Vis blown - balloon(1)
NVis lend - thend(2), thone(1), thand(1).

EXPT III: Ss: 3 male, 5 female (19.10-43.10).
Vis file - v/file (4), trial (4)
NVis bile - mile (2); vile - thile (10).

EXPT IV: Ss: 2 male, 8 female (20.3-45.0)
Vis stlit - sklit(4), sdlit (2)
NVis split - sblit(2).

EXPT V: Ss: 2 male, 4 female (18.10-20.0).

EXPT VI: Ss: 1 male, 6 female (19.7-37.9).
Vis ...t - ...d(3); split - sblid (1).
NVis ...t - ...d(3); slit - snit (1).

PILOT STUDY: Ss: As EXPT V.

EXPT VII: Ss: 4 male, 4 female (20-24)

EXPT VIII: Ss: 3 male, 9 female (22.4-50.6)

EXPT IX: Ss: 4 male, 7 female (20.3-22.4).

EXPT X: Ss: 3 male, 8 female (20.4-43.10).

EXPT XI: Ss: 7 male, 9 female (19.4-42.8).

EXPT XII: Ss: 2 male, 16 female (19.4-50.6).
NVis k - th(5)

Expts VII, IX and X were conducted partly during practical classes.

REFERENCES

ADES, A.E. 1981 Time for a purge COGNITION 10:7-15.

BATESON, G. 1978 Steps to an Ecology of Mind GRANADA PUBLISHING, LONDON.

BEVER T.G., & CARROLL J.M. 1981 On some continuous properties in language IN T. MYERS et al (1981).

BOHM, D. 1980 Wholeness and the Implicate Order ROUTLEDGE & KEGAN PAUL, LONDON.

BLUMSTEIN, S.E., & STEVENS, K.N. 1981 Phonetic features and acoustic invariance in speech COGNITION 10:25-32.

CAMPBELL, R., & DODD, B. 1980 Hearing by eye QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY 32:85-99.

CHERRY, E.C. 1953 Some experiments on the recognition of speech, with one and with two ears JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 25:975-979.

CHISTOVICH, L.A. 1980 Auditory processing of speech LANGUAGE & SPEECH 23:67-73.

COLE, R.A. 1973 Listening for mispronunciations: a measure of what we hear during speech PERCEPTION & PSYCHOPHYSICS 1:153-156.

COLE, R.A., and SCOTT, B. 1974a The phantom in the phoneme:invariant cues for stop consonants PERCEPTION & PSYCHOPHYSICS 15:101-107.

COLE, R.A., & SCOTT, B. 1974b Toward a theory of speech perception PSYCHOLOGICAL REVIEW 81:348-374.

COOPER, W.E. 1974 Adaptation of phonetic feature analyzers for place of articulation JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 56:617-627.

CORNFORD, F.M. 1950 The Unwritten Philosophy UNIVERSITY PRESS, CAMBRIDGE.

CUTTING, J.E. 1982 Plucking and bowing are phonetically perceived, sometimes PERCEPTION & PSYCHOPHYSICS 31:462-476.

CUTTING, J.E., & PISONI, D.B. 1976 An information processing approach to speech perception STATUS REPORT ON SPEECH PERCEPTION, HASKINS LABORATORIES, SR-48:287-325.

CUTTING, J.E., & ROSNER, B.S. 1974 Categories and boundaries in speech and music PERCEPTION AND PSYCHOPHYSICS 16:564-570.

DEWEY, J. 1896 The reflex arc concept in psychology PSYCHOLOGICAL REVIEW 4:357-370.

DODD, B. 1977 The role of vision in speech perception PERCEPTION 6:31-40.

DODD, B. 1979a Lip reading in infants: attention to speech presented in- and out-of-synchrony COGNITIVE PSYCHOLOGY 11:478-484.

DODD, B. 1979b The integration of auditory and visual information for speech perception PAPER TO THE EXPERIMENTAL PSYCHOLOGY SOCIETY, LONDON.

DORMAN, M.F., RAPHAEL, A.M., LIBERMAN, A.M., & REPP, B. 1975 Maskinglike phenomena in speech perception JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 57 SUPPLEMENT 1, S46(A).

DORMAN, M.F., STUDDERT-KENNEDY, M., & RAPHAEL L.J. 1977 Stop-consonant recognition: release bursts and formant transitions as functionally equivalent, context-dependent cues PERCEPTION & PSYCHOPHYSICS 22:109-122.

EIMAS, P.D., & CORBIT, J.D. 1973 Selective adaptation of linguistic feature detectors COGNITIVE PSYCHOLOGY 4:99-109.

ERBER, N.P. 1975 Auditory-visual perception of speech JOURNAL OF SPEECH AND HEARING DISORDERS XL:481-492.

ERBER, N.P., & DE FILIPO, C.L. 1978 Voice/mouth synthesis and tactual/visual perception of /pa,ma,ba/ JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 64:1015-1019.

FISCHER-JORGENSEN, E. 1954 Acoustic analysis of stop consonants MISCELLANEA PHONETICA 2:42-49.

FITCH, H.L., ERICKSON, D.M., & LIBERMAN, A.M. 1981 Perceptual equivalence of two acoustic cues for stop-consonant manner PERCEPTION & PSYCHPHYSICS 27:343-350.

FOWLER, C.A., RUBIN, P., REMEZ, R.E., & TURVEY, M.T. 1978 Implications for speech production of a general theory of action IN B. BUTTERWORTH (ED), LANGUAGE PRODUCTION, ACADEMIC PRESS, NEW YORK.

GIBSON, J.J. 1960 The concept of the stimulus in psychology AMERICAN PSYCHOLOGIST 15:694-703.

GIBSON, J.J. 1961 Ecological optics VISION RESEARCH 1:253-262.

GIBSON, J.J. 1963 The useful dimensions of sensitivity AMERICAN PSYCHOLOGIST 18:1-15.

GIBSON, J.J. 1966 The Senses Considered as Perceptual Systems HOUGHTON-MIFFLIN, BOSTON.

GIBSON, J.J. 1968-69 Are there sensory qualities of objects? SYNTHESE 19:408-409.

GIBSON, J.J. 1973 Direct visual perception: a reply to Gyr PSYCHOLOGICAL BULLETIN 79:396-397.

GIBSON, J.J. 1974 A note on ecological optics IN E.C. CARTERETTE & M.P. FRIEDMAN (EDS), HANDBOOK OF PERCEPTION, ACADEMIC PRESS, NEW YORK.

GIBSON, J.J. 1976-77 The myth of passive perception: a reply to Richards PHILOSOPHY AND PHENOMENOLOGICAL RESEARCH XXXVII:235-238.

GIBSON, J.J. 1979 The ecological approach to visual perception HOUGHTON MIFFLIN, BOSTON.

GIBSON, J.J., KAPLAN, G.A., REYNOLDS, H.N. & WHEELER, K. 1969 The change from visible to invisible: a study of optical transitions PERCEPTION & PSYCHOPHYSICS 5:113-116.

HAGGARD, M. 1981 Chairman's Comments IN T. MYERS et al (1981).

HALWES, T.G. 1969 Effects of dichotic fusion on the perception of speech UNPUBLISHED DOCTORAL DISSERTATION, UNIVERSITY OF MINNESOTA.

HUGGINS, A.W.F. 1972 On the perception of temporal phenomena in speech JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 51:1279-1290.

JAMES, W. 1908 Pragmatism LONGMANS, NEW YORK.

JUSCZYK, P.W., SMITH, L.B., & MURPHY, C. 1981 The perceptual classification of speech PERCEPTION & PSYCHOPHYSICS 30:10-23.

KANTOR, J.R. 1978 Cognition as events and as psychic constructions PSYCHOLOGICAL RECORD 28:329-342.

KLATT, D. 1977 Review of the ARPA speech understanding project JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 62:1345-1366.

KLIMA, E., & BELLUGI, U. 1979 The Signs of Language HARVARD UNIVERSITY PRESS, CAMBRIDGE, MASSACHUSETTS.

LIBERMAN, A.J., & PISONI, D.B. 1977 Evidence for a special speech-perceiving subsystem in the human IN T.H. BULLOCK (ED), RECOGNITION OF COMPLEX ACOUSTIC SIGNALS. BERLIN: DAHLEM KONFERENZEN.

LIBERMAN, A.M., COOPER, F.S., SHANKWEILER, D.P., & STUDDERT-KENNEDY, M. 1967 Perception of the speech code PSYCHOLOGICAL REVIEW 74:431-461.

LIBERMAN, A.J. 1981 On finding that speech is special STATUS REPORT ON SPEECH RESEARCH, HASKINS LABORATORIES, SR-67/68:107-144.

McGURK, H., & MacDONALD, J. 1976 Hearing lips and seeing voices NATURE 264:746-748.

McGURK, H., & MacDONALD, J. 1977 Hearing lips and seeing voices: a new illusion PAPER PRESENTED AT THE ANNUAL CONFERENCE OF THE BRITISH PSYCHOLOGICAL SOCIETY, UNIVERSITY OF EXETER, APRIL, 1977.

MASSARO, D.W. 1981 Sound to Representation: An Information-Processing Analysis IN T. MYERS et al (1981).

MILLER, G.A., & NICELY, P.E. 1955 An analysis of perceptual confusions among some english consonants THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 27:338-352.

MILLER, J.L., & LIBERMAN, A.M. 1979 Some effects of later occurring information on the perception of stop consonants and semivowels PERCEPTION & PSYCHOPHYSICS 25:457-465.

MORTON, J., & CHAMBERS, S.M. 1976 Some evidence for 'speech' as an acoustic feature BRITISH JOURNAL OF PSYCHOLOGY 67:31-45.

MYERS, T., LAVER, J., & ANDERSON, J. 1981 The Cognitive Representation of Speech NORTH-HOLLAND PUBLISHING COMPANY, AMSTERDAM.

NORMAN, D.A. 1978 Stop already, my mind is made up THE BEHAVIOURAL & BRAIN SCIENCES 1:589-590.

O'NEILL, J. J. 1951 Unpublished doctoral dissertation QUOTED IN O'NEILL, J.J. (1954).

O'NEILL, J.J. 1954 Contributions of the visual components of oral symbols to speech comprehension JOURNAL OF SPEECH AND HEARING DISORDERS 19:429-439.

PICKETT, J.M., & POLLACK, I. 1963 The intelligibility of excerpts from fluent speech: auditory versus structural context LANGUAGE AND SPEECH 3:151-165.

PILCH, H. 1979 Auditory phonetics WORD 29:148-160.

POLANYI, M. 1967 The Tacit Dimension ROUTLEDGE & KEGAN PAUL, LONDON.

POPPER, K.R. 1972 Objective Knowledge CLARENDON PRESS, LONDON.

REMEZ, R.E. 1977 Adaptation of the category boundary between speech and non-speech: a case against feature detectors STATUS REPORT ON SPEECH RESEARCH, HASKINS LABORATORIES, SR-50:151-167.

REPP, B. 1980 Bidirectional contrast effects in the perception of VC-CV sequences STATUS REPORT ON SPEECH PERCEPTION, HASKINS LABORATORIES, SR-63/64:157-175.

REPP, B.H. 1981a On levels of description in speech research STATUS REPORT ON SPEECH RESEARCH, HASKINS LABORATORIES, SR-65:223-232.

REPP, B.H. 1981b Perceptual equivalence of two kinds of ambiguous speech stimuli STATUS REPORT ON SPEECH RESEARCH, HASKINS LABORATORIES, SR-66:79-84.

ROBERTS M., & SUMMERFIELD, Q. 1981 Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory PERCEPTION & PSYCHOPHYSICS 30:309-314.

ROSEN, S.M. 1981 Discussant's contribution IN T. MYERS et al (1981).

ROSEN, S.M., & HOWELLS, P. 1982 Plucks and bows are not phonetically perceived PERCEPTION & PSYCHOPHYSICS 30:156-168.

RUDNICKY, A.I., & COLE, R.A. 1977 Adaptation produced by connected speech JOURNAL OF EXPERIMENTAL PSYCHOLOGY: HUMAN PERCEPTION & PERFORMANCE 3:51-61.

RUDNICKY, A.I., & COLE, R.A. 1978 Effect of subsequent context on syllable perception JOURNAL OF EXPERIMENTAL PSYCHOLOGY: HUMAN PERCEPTION & PERFORMANCE 4:638-647.

SAMUEL, A.G. 1981 The role of bottom-up confirmation in the phonetic restoration illusion JOURNAL OF EXPERIMENTAL PSYCHOLOGY: HUMAN PERCEPTION AND PERFORMANCE 7:1124-1131.

SAMUEL, A.G. 1982 Phonetic prototypes PERCEPTION & PSYCHOPHYSICS 31:307-314.

SANDERS D.A., & GOODRICH, S.J. 1971 The relative contributions of visual and auditory components of speech to speech intelligibility as a function of three conditions of frequency distortions JOURNAL OF SPEECH & HEARING RESEARCH 14:172-178.

SCHATZ, C. 1954 The role of context in perception of stops LANGUAGE 30: 47-56.

SPOEHR, K.T., & CORIN, W.J. 1978 The stimulus suffix effect as a memory coding phenomenon MEMORY & COGNITION 6:583-589.

SCHOUTEN, M.E.H. 1980 The case against a speech mode of perception ACTA PSYCHOLOGICA 44:71-98.

SCHOUTEN, M.E.H. 1981 Speech perception is timbre perception IN T. MYERS et al (1981).

STEINER, R. 1960 Von Seelenraetseln VERLAG DER RUDOLF STEINER-NACHLASSVERWALTUNG, DORNACH. TRANSLATION IN O. BARFIELD (1970), THE CASE FOR ANTHROPOSOPHY, RUDOLF STEINER PRESS, LONDON.

STEINER, R. 1962 Die Philosophie der Freiheit VERLAG DER RUDOLF STEINER-NACHLASSVERWALTUNG, DORNACH. TRANSLATION IN M. WILSON (1972), THE PHILOSOPHY OF FREEDOM, RUDOLF STEINER PRESS, LONDON.

STEINER, R. 1974 Die Raetsel der Philosophie (Zweiter Band) VERLAG DER RUDOLF STEINER-NACHLASSVERWALTUNG, DORNACH. OWN TRANSLATION.

STEVENS, K.N. 1975 The potential role of property detectors in the perception of consonants IN G. FANT & M.A. TATHAM (EDS), AUDITORY ANALYSIS AND THE PERCEPTION OF SPEECH ACADEMIC PRESS, NEW YORK.

STEVENS, K.N. 1981 Constraints imposed by the auditory system on the properties used to classify speech sounds: data from phonology, acoustics and psychoacoustics IN T. MYERS et al (1981).

STUDDERT-KENNEDY, M. 1981 The emergence of phonetic structure COGNITION 10:301-306.

STUDDERT-KENNEDY, M. 1980 Speech perception LANGUAGE & SPEECH 23:45-65.

SUMMERFIELD, Q. 1979 Use of visual information for phonetic perception PHONETICA 36:314-331.

SUMMERFIELD, Q., BAILEY, P.J., SETON, J., & DORMAN, M.F. 1981 Fricative envelope parameters and silent intervals in distinguishing 'slit' and 'split' PHONETICA 38:181-192.

TURVEY, M.T., SHAW, R.E., REED, E.S., & MACE, W.M. 1981 Ecological laws of perceiving and acting: in reply to Fodor and Pylyshyn (1981) COGNITION 9:237-304.

WARREN, R.M. 1968 Verbal transformation effect and auditory perceptual mechanisms PSYCHOLOGICAL BULLETIN 70:261-270.

WARREN, R.M. 1970 Perceptual restoration of missing speech sounds SCIENCE 167:392-393.

WARREN, R.M., & SHERMAN G.L. 1974 Phonemic restorations based on subsequent context PERCEPTION & PSYCHOPHYSICS 16:150-156.

WARREN, R.M., & WARREN R.P. 1970 Auditory illusions and confusions SCIENTIFIC AMERICAN 223:30-36.

WEISKRANTZ, L. 1977 Trying to bridge some neurophysiological gaps between monkey and man BRITISH JOURNAL OF PSYCHOLOGY 68:431-445.

WHITEHEAD, A.N. 1926 Science and the Modern World UNIVERSITY PRESS, CAMBRIDGE.

WIGNER, E. 1969 Epistemology of quantum mechanics: its appraisal and demands IN M. GRENE (ED) THE ANATOMY OF KNOWLEDGE, ROUTLEDGE & KEGAN PAUL, LONDON.

WINITZ, H., SCHEIB, M.E., & REEDS, J.H. 1972 Identification of stops and vowels for the burst portion of /p,t,k/ isolated from conversational speech JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA 51:1309-1317.

VON RAFFLER-ENGEL, W., NEWMAN, K., FOSTER, R., & GANTZ, F. 1980 The relationship of nonverbal behaviour to verbal behaviour in the evaluation of job applicants IN W. VON RAFFLER-ENGEL (ED), ASPERCTS OF NONVERBAL COMMUNICATION, SWETS, NORTH AM.

YOUNG, E.D., & SACHS, M.B. 1981 Processing of speech in the peripheral auditory system IN T. MYERS et al (1981).