# EXPERIMENTAL STUDY OF AURAL DISCRIMINATION BETWEEN SPEECH AND NON-SPEECH.

BY

LUIS FELIPE MUJICA TORO

A Thesis submitted for the Degree of Doctor of Philosophy, in the Faculty of Engineering, University of London.

Department of Electrical Engineering,
Imperial College of Science and Technology,
University of London.

JULY 1984

TO MY FATHER'S MEMORY

TO MY MOTHER

TO MY SONS.

ABSTRACT

This thesis describes the study of some characteristics of the processes involved in aural discrimination between speech and nonspeech signals. Particular emphasis is placed upon measurements of reaction time and accuracy of discrimination.

It is a working hypothesis of this study that the incoming stimuli are discriminated prior to any form of linguistic "decoding". This decision prepares the human receiver to deal with another person, not with an object. The prima facie grounds for this hypothesis are the capacity of any person to tune a radio receiver regardless of the receiving conditions or the language of the broadcasting, and in the study of young infants' reaction to speech sounds.

Subjects who had normal hearing and were right-handed were employed in a series of three experiments. Reaction Times and Accuracy of their discrimination between BBC English, cello music and pink noise were measured.

The relations of Reaction Time and Accuracy to stimulus duration were obtained by varying the duration of the stimuli while controlling for the attention condition of the subjects by varying the interval between stimuli in a random manner. These results are then compared with Reaction Times and Accuracy obtained using a constant interstimuli interval.

Possible brain hemispheric specialisation in discriminating the incoming stimuli was explored by conveying the set of stimuli to different ears.

The Study of the three factor ANOVA and the graphs of the data obtained led to the following conclusions.

1. The average stimulus duration necessary for a correct decision in 90% of presentation varied for subjects, averaging 70 msecs for noise, 110 msecs for music and 90 msecs for speech.

2. No significant tendency for errors when discriminating between speech, music and noise.

The duration of the stimuli is significant at $p < 0.5\%$ .Variation between subjects is also significant at $p < 0.5\%$.

3. Random variation of the interstimuli interval prolonged the subjects' Reaction Time and diminished the accuracy of the Speech/Non-speech discrimination.

4. No ear advantage was detected in the Speech/Non-Speech discrimination.

In the light of the review of the literature and the theoretical discussion the following further conclusions are claimed.

1. The heuristic suspicion that rapid discrimination occurs is sustained by the results obtained.

2. Attention is a relevant variable as regards discriminatory capacity.

3. Lack of right ear advantage and the relation between the brevity of the stimuli and the average duration of a phoneme indicate that neither meaning nor phonological structure play a part in the discriminatory process.

It is conjectured that the significance of these conclusions is that Speech/Non-speech discrimination is a prelinguistic capacity, both in the sense that very young infants possess it, and in that the segment of speech necessary for it to occur is insufficient in duration to obtain any significant "linguistic" component.

## ACKNOWLEDGEMENTS

The author is deeply grateful to his supervisor, the late Professor Colin Cherry for his assistance and inspiring guidance throughout the course of this research. Thanks are also due to my second supervisor, Doctor Robert Spence for his quiet encouragement in completing this work. The author wishes to acknowledge the friendship and solidarity of many, without which this work could not have been possible. The financial support for this work came from World University Service (UK).

TABLE OF CONTENTS.

RESUME:

EXPERIMENTAL STUDY OF AURAL DISCRIMINATION
BETWEEN SPEECH AND NON-SPEECH.

The material presented in this thesis resulted from an attempt to study an obvious and specific operation of speech perception: its detection as a signal different from noise.

This thesis falls basically into two parts. The first three chapters will lead the reader from the heuristic realization of the existence of a basic perceptual problem, through the psychophysical and abstract framework of the Speech and Non-Speech discrimination. Evidence which supports the presumption of such an operation is extracted from the literature which, though not dealing directly with our problem, nevertheless gives some clues about it.

The remaining chapters present the search for an experimental tool, give an account of the experimental work and present the conclusions. The alternatives to gain insight into this perceptual problem are to study the amount of information which needs to be absorbed in the performance of the Speech/Non-Speech (S/NS) discrimination, measure the time needed to perform such a task, or to measure the amount of disruption upon this task that a simultaneous task causes. The selection of reaction time techniques leads to the exploration of the problems this tool presents when working with natural stimuli and subsequently to the description of the experimental stage.

CHAPTER I is a description of this thesis in non-specialist language and poses the problem under scrutiny using everyday experiences. The main issues in the thesis are briefly anticipated and the framework, tools and the experimental exploration should become clear for the reader in this chapter.

CHAPTER II is an overview of the psychology of speech. Three major elements are described as belonging to one system. Speech production and the hearing system are linked through an appreciation of their neurological interdependence.

CHAPTER III attempts to deal with the concatenated dynamics of the elements presented in Chapter II. Apprehension and planning of a word or phrase are explored and experimental evidence is gathered in support of the need to study the aural discrimination of Speech and Non-Speech.

Ontogeny and phylogeny of the S/NS discrimination is also reviewed to obtain more insights into the processes which contribute to the mastering of language.

CHAPTER IV reduces the problem of Speech/Non-Speech discrimination into measurable proportions and after reviewing the available techniques, explores the measurement of reaction time as the most suitable for our series of experiments.

The problems associated with reaction time measurements when using natural speech as a stimulus are discussed in this chapter.

CHAPTER V is a brief account of a series of preliminary experiments carried out to gain familiarity with reaction time techniques and to estimate the magnitude of the variables under study. The effect of the signal bandwidth upon speech detectability and a replication of a particular case of Hick's experiment are discussed in this chapter. The results of these experiments are presented and discussed.

CHAPTER VI presents the approach, justification, hypothesis, experimental procedures and analysis of the data of the main set of experiments. The concluding remarks are complemented by proposals for further research.

The reaction time to a three choice experiment using random samples of natural speech, cello music and pink noise - is studied while varying the stimulus duration as a means to vary the information content of the stimuli. Interstimuli interval is varied in a random manner in order to assess the effect of the readiness of the subjects upon the reaction time and accuracy of their responses.

The same set of stimuli was then relayed monoaurally to the left and right ear to explore the presence of right ear advantage in the Speech/Non-Speech discrimination.

The study of the statistics and graphs of the reaction time and accuracy of the subjects' responses when varying the duration of the stimuli lead us to the following conclusions:

1. The shortest time when discriminating between Speech and Non-Speech stimuli with an accuracy of 90% is 80 msecs $\pm$ 20 msecs.

2. No preference was detected in the accuracy of identification of speech and of music. Noise was detected more readily when shorter stimuli were relayed. The longer reaction times and lower accuracy can be the result of the short exposure of the subjects to the stimuli; nevertheless, the fact that discrimination was possible with such short stimuli is still valid.

3. The changes in attention state arising from variation of the interstimuli interval affect the reaction time and the accuracy of the subject's responses.

4. No right ear advantage was detected, eliminating therefore the possibility of hemispheric specialization in Speech/Non-Speech discrimination.

5. Aural discrimination between speech and Non-Speech signals is a non-linguistic operation of our perception apparatus.

## STATEMENT OF ORIGINALITY

As far as the author is aware, the opinions and techniques presented in this work are his own unless otherwise acknowledged by making specific reference. The main contributions are, in the author's opinion, as follows.

1. The study of the obvious in perception is a difficult exercise. This thesis is focused on the exploration of a basic perceptual problem identified as important but not previously pursued by experimental procedures, namely the classification or recognition of an aural stimulus as Speech (Chapter 1, 3).

2. The hypothesis that the incoming aural stimuli are identified as speech prior to any form of linguistic decoding is an original starting point for this thesis. The extreme brevity of the stimuli so identified, as well as the lack of REA (Right Ear Advantage) in effecting the discrimination, would both be an indication that this hypothesis might be correct.

3. A critical review of current literature, in which there was no direct reference to the central problem. A wide range of sources was consequently reviewed in order to discover indirect indications of the nature of the problem and the relevant consideration for constructing an experimental technique for its investigation. (Chapter 2,3,4).

4. The application of a Three Forced Choice task using direct labelling techniques to avoid the use of short term memory in carrying out this discrimination (Chapter 4).

5. Measurement of Reaction Times and Accuracy using natural speech under different attention conditions of the subjects. This was achieved by changing the interstimuli interval (Chapters 4,5,6).

6. Measurement of possible REA in carrying out the Speech/Non-Speech discrimination (Chapters 5,6).

7. Inference of conclusions which may affect current proposals for models of speech perception and an identification of the outstanding problems in the field of Speech/Non-Speech discrimination (Chapter 6).

# CHAPTER I

## 1.1 Motivation

The prima-facie attraction of perception work for a telecommunications engineer lies in the fact that this work is a compulsory avenue in the intergration of the final objective of any communication system, the human being and his/her characteristics into the specification of the telecommunication machine. Mr. Morse designed his code with the help of a print worker who showed him which characters were most used. Telephones are designed with a certain capacity to convey the frequencies of human speech which does not impair intelligibility. Television, a one way communication system, uses the inability of the human perception system to detect the non-continuous portrayal of movement. This is achieved by forming a picture out of two interlaced "frames" transmitted every 20 milliseconds. This trick saves channel space and makes the equipment less expensive.

For an engineer educated not to trespass the boundaries of his/her small province of knowledge, the integration of the human element into the design of any telecommunication system seems, at first sight, a fairly straightforward step. It appears that it is only necessary to add another box at the end of a series of squares and circles linked by arrowed lines.

The naivety of that approach resulted in the introduction of the author to the complexities of the human "communication system" and led him into the exploration of an obvious "super problem" in speech perception studies: How does the human brain discriminate between speech and other stimuli reaching the auditory system....?

The first steps in the specification of the research problem came into the author's mind when he was faced with a seemingly trivial question:

Can we build a machine able to discriminate between speech and other signals, for example music?....

The natural response of an engineer is to avoid any dealings with the unknown and inquire into the signal's properties which could lead to mathematical description of it. Is it a periodical signal...? For if it is periodical it could be modelled by expressions in the frequency domain or time domain. Panic!... Speech is not a periodical signal. The obstinate scientist can still pretend that the signal is periodic in sections and apply the "magic" ointment, Fourier transformation, which enables him to characterise the section of the speech waveform in the frequency domain. The other avenue to explore is the long term analysis of the waveform. The statistical data can be elicited by correlating the speech waveform with a known signal or autocorrelating the speech.

This approach demands an arsenal of instrumentation. The literature offers ample evidence of this approach. One of its clearest result is the work with visible speech. This is the real time transformation of the speech waveform into its frequency components. See Fig. 1.1

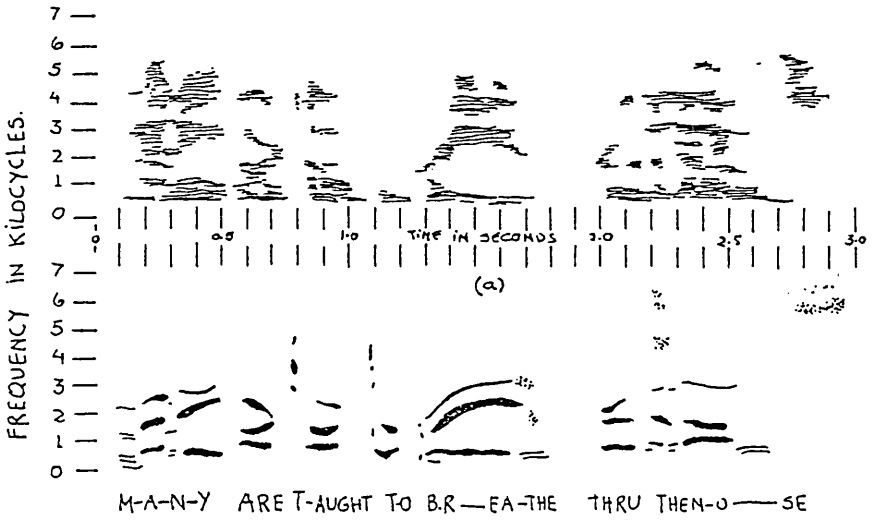



Fig. 1-1. A "visible speech " stream and its formants.

The darker areas of the top graph are hand drawn in the bottom one. These areas of accumulation of energy are called formants. The lowest is the first formant, and the next lowest is the second formant. These two formants correspond to the two major cavities in the speech production system - the pharynx and the oral cavity. See Chapter 2. This work now lies behind the infinite variety of talking computers, toy cars, etc., yet it has failed to produce the easy characterization of the speech waveform and its differences with a music waveform....

Turning our attention to nature's response to the problem it can be noticed that it is extremely fast, reliable and uses a portable "machine", our brain...

The problem of discriminating between speech and other noises is already solved by our auditory apparatus. The next section of this Chapter will present some examples of this operation.

## 1.2 The Speech/Non-Speech Discrimination

Cherry (1978) describes the Speech/Non-Speech discrimination using the following example: "Imagine yourself sitting in an armchair, absorbed in a book; a sound falls upon your ears. Immediately a fundamental decision is made in your brain, which could be described as answering the question: Was somebody speaking to me, or was it the wind...? Usually the decision is easy, natural, rapid and unconscious but has to be decided, for all else that follows depends upon that decision....."

"This primary classification of a stimulus as (a) a communicative sign, or (b) a casual sign, made within your brain so rapidly, naturally, unconsciously, and (usually) reliable, is an assumption, or hypothesis".....

There are more examples which throw some light on the characteristics of this decision. Any person tuning a short wave receiver is able to tune it into a very weak station. The murmur coming from the loudspeaker can be buried in all sort of noises, obscured by another station yet the listener will be able to tune into it precisely.

The broadcasting could be in Chinese and/or in a single side band transmission, nevertheless, the tuning will be achieved....

In this simple example the auditory apparatus is not extracting meaning yet is classifying the incoming sound as speech from a signal with a negative signal to noise ratio. There is no machine capable of that technological feat!

When one detects speech originating at a distance, the frequency-intensity characteristics of the incoming speech have been changed by the transmission media. Vowels will convey most of the energy. The detection of the signal as something different from traffic noise will be made using the information conveyed by the vowels. The opposite is also true. We do detect whispered speech, this time at shorter distances when somebody is whispering to us. This time the energy conveyed by vowels and consonants is approximately the same.

## 1.3 The Scheme of This Thesis

It should be emphasized that this work gives an experimental insight into a problem which is so obviously present in our every day experiences that it seems to have been left aside by linguists and psychologists.

The analysis of the speech processes have, until recently, followed the linguistic heterarchy when trying to elicit the mechanisms which are assumed to be present in our perceptual system. The focus of the research has been in the transformation of auditive level material into phonological cues. The work of Haskins Research Laboratories is an example of this trend. Psychologists nowadays tend to reject this approach (Broadbent, 1981). See Chapter III.

In contrast with the parallelism with the linguist's description this thesis postulates the existence of an "intermediate step of cognition" (Cherry, 1978) which is a vital decision for the subsequent analysis of the signal as speech, or a meaning related sign; and as a sound which has a cause....

This thesis falls basically into two parts. The first three chapters will lead the reader from the heuristic realisation of the existence of a basic perceptual problem, through the biology, psychophysical and abstract framework of the Speech and Non-Speech discrimination. Evidence which supports the presumption of such operation is extracted from the literature which does not deal with this problem in a direct manner, yet gives some clues about it.

The biology of the system which contributes to the production and perception of spoken utterance is presented in Chapter II. These systems are linked through an appreciation of their neurological interdependence. The functional links and concatenated dynamics of these elements are explored in Chapter III. Apprehension of a word or phrase is explored and experimental evidence is gathered in support of the need to study the discrimination of Speech and Non-Speech at a very basic level. Ontogeny and phylogeny of the Speech/Non-Speech discrimination is also reviewed to obtain more insight into the processes which contribute to the mastering of language.

Chapter IV reduces the dimensions of the problem under study to measurable proportions and after a review of the available techniques in the psychophysicists's arsenal, explores reaction time measurements as the most suitable tool for the experimental stage. The problems associated with the use of reaction time techniques when using natural speech as a stimulus are also discussed.

The timid approach to the experimental stage led the author to design a series of preliminary experiments which are presented and discussed in Chapter V. These were carried out in order to gain familiarity with reaction time techniques and to estimate the magnitude of the variables and parameters under study.

The main set of experiments is presented in Chapter VI together with the approach, justification, the techniques used and the results of the experiments.

The reaction time and accuracy of the responses to three forced choice experiments using samples of natural speech (as far as we can define BBC accent as natural...), Cello music and pink noise are studied while varying the stimuli duration as a means to vary the information content of the stimuli.

This method of determining an absolute threshold for the labelling of an incoming stimuli is described as one of the "classical" methods used by psychophysicists as the method of the limits. (Plutchik, 1974).

The state of readiness of any person when asked to perform a given perceptual task affects the person's performance. In other words, paying attention is a more receptive condition than day dreaming... This effect is studied in one of the main experiments by means of varying the timing of the presentation of the stimuli in a random manner. Human beings have the amazing capacity to predict empty intervals of time within a certain range. The drummer soldier times the pace of the presentation of the flag to the Queen with an amazing precision.

The fact that humans have two ears yet perceive one word (Cherry, 1953), has been under the scrutiny of a multitude of scholars for a long time. The effects of the specialisation of one of the brain hemispheres to deal with speech signals, are another important source of hard facts in the formulation of models for the perception of speech.

The study of the presence of hemispheric specialisation effects upon the Speech/Non-Speech discrimination is attempted in the last experiment.

The results of the preceeding experiments are presented and discussed in the closing sections of the last Chapter.

## 1.4 An Operational Paradigm

The preceeding sections of this chapter pointed out the fact that there is a heuristic and also a logical need to formulate a very simple operational heterarchy for our abstract formulation of the Speech/Non-Speech discrimination. The decision made about the nature of the incoming stimuli is made without reference to any meaning, it is very rapid and therefore must logically be placed before any linguistic analysis carried by our central processing capabilities.

The operational description of the Speech/Non-Speech discrimination is therefore quite simple in terms of boxes, squares and arrows. See Fig. 1.2

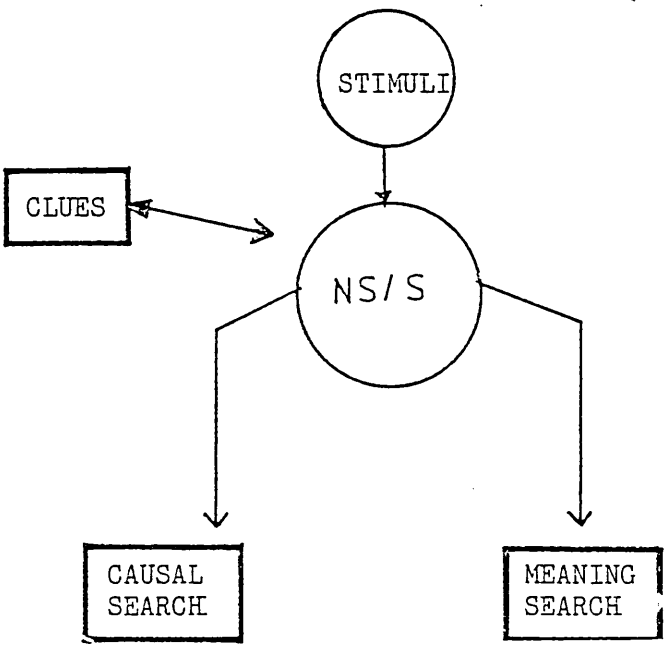



Fig. 1-2. Operational description of Speech/Non-speech Discrimination.

An incoming signal is analysed with reference to some of its clues and channeled into two different types of analysis. In one of these analysis-branches the meaning of the incoming signal will be sought, while the other type of analysis will carry out the search for a cause.

This model is an abstract representation of a logical decomposition of a natural phenomena. It is not a physical representation of two different modes of neurological processes. It is a simple addition of another box in the serial representation of the language mechanisms. See Chapter III.

The short and pedestrian description of the elements of our anatomy which play a part in the perception and production of a word is considered necessary as a contrast to the abstract account of the inner mechanisms of their product, the language. This is characterized in Chapter III as one of the dynamics which have crystalised after, and in direct, interrelation with the developmental and historical processes by which the human organism has acquired the unique characteristic of speech. " Man is human because he can say so ". Lieberman (1970)

## 1.5 More Evidence

Speech perception is a special mode of perception, different from the perception of other aurally conveyed signs (Webster et al, 1968). Notwithstanding this, speech perception is one of the mechanisms humans use to apprehend the surrounding world. This fact imposes some general characteristics upon its modus operandi. Some comparisons of speech perception with respect to other modes of perception provide some idea of the workings of speech perception itself. See Chapter III.

The developmental study of the acquisition and mastering of our speech faculties is another avenue to explore. Are we born with a dictionary in the brain?.. If we are not given such a facility, are there any prewired rules which enable any human to master a language? Is one of these "rules" the discrimination between speech and other noises?....

This process of acquisition of language is another "ongoing" process which links the elements described in Chapter II. There is evidence that the two cerebral hemispheres are at birth equally capable of acquiring language. (C. A. Fowler, 1975). There is also evidence that new-born babies of 4 weeks are capable of discriminating between speech and noises, and furthermore, are able to employ some of the mechanisms used by adults in the recognition of consonants. See Chapter III.

The evolutionary process resulted in the shaping of highly specialised organs in the production and reception of speech. The constant reinforcement of production and perception of speech in young infants is also a result of the evolutionary process, which should be considered as the deciding criteria in the discussion of the predominance of production or perception. The extreme viewpoint of some of the scholars who support the Motor Theory of speech perception gives predominance to the production over perception. The basic tenet of the motor theory is that speech perception and articulatory control involve the same (or closely linked) neurological processes. (B. Repp, 1982). As these processes are interdependent, the assignment of some evolutionary supremacy to one side of the process would be comparable to giving supremacy to the chicken or to the egg.

If we want to discover whether man is specialized to process speech and to receive phonetic segments, appropriate comparisons with animals must be made. See Chapter III.

The extremely accurate and complex song detection seen in some species of animals is a limited example of the sort of operation on which this work has focused its attention. The limitation of the comparison is given by the consequences this decision has upon our behaviour. The complex song of a mating bull-frog does not bring into play linguistic mechanisms nor moral codes. The character of this decision as a new-born baby will carry it out, is obviously something to explore. Is it a "reflex" or is it learnt through the intense contact with the mother?

There is a peculiar embarrassment about trying to telephone a friend and suddenly realising that one has been talking to an answering machine... The author feels deeply cheated when this happens. Another example of moral codes being brought to bear by the Speech/Non-Speech decision is our attitude when dressing and being interrupted by a strange voice. Towels and dressing-gown come quickly to help. A muslim woman will cover her hair and face.

The multifaceted processes in which the Speech/Non Speech decision is immersed have been explored by psychologists in their search for mechanisms specific to speech. The work on the perception of consonants and vowels using artificial speech yielded some collateral results which are of some relevance to this work. Subjects used in these experiments were trained to accept those stimuli as speech. Some patterns of the behaviour which were elicited using this technique are thought to be specific to the decoding of speech. If the subject believes that the stimulus is speech the perception will be categorical, i.e. continuous manipulation of some characteristics of the consonants will yield discrete classification of the consonant as "g" "b" or "d". This pattern of perception is not achieved if subjects are not informed that the stimuli are speech.

The problem of designing a machine that could discriminate between speech and music has been transformed into a search for the clues humans use to determine that a certain signal they hear is speech...

## 1.6 What and How to Measure?

Oscilloscopes, measuring tape, electronic time meters, vocoders etc. are of little use in the measurement of mental activities. The object of our attempted measurements lies somewhere between the ear and the subject's response. There is a decision as to the nature of the incoming stimuli somewhere between the auditory level and phonological level of analysis of the stimuli. The activity which is triggered by this decision is mental and this collection of operations takes time to be carried out.

The proposed tools to measure this perception work have been described as follows; (Westhoff, 1963).

a) The measurement of the amount of information which must be absorbed and processed in the performance of a given task.

b) The measurement of the time needed to absorb and process the information, called reaction time, or the time needed to perform a given task.

c) The measurement of the extent to which the performance in one task is reduced when another task is carried out at the same time.

The selection of the measurement tool should be consistent with the approach to the perception of signs which are subjectively significant to the subject. The information theory approach which assigns equal values to the different lamps, flashes or letters is discarded for our purposes in Chapter III. Humans directly choose the stimuli to be decoded, in contrast with a selection by exhaustive comparison to pick one stimulus out of an ensemble. Particular perceptual capacities exercised on natural stimuli will not be revealed by experiments of that sort. The Speech/Non-Speech decision is one such specific capacity.

Shadowing experiments, which are among the tools at our disposal described at the beginning of this section, are reported in Chapter III. These experiments have also demonstrated that the Speech/Non-Speech decision was always carried out, regardless of the amount of central load imposed to the subjects. What we are engaged on is the study of a peripheral mechanism not susceptible to disruption however much the subject's attention is otherwise engaged.

The choice of tool is therefore limited to reaction time measurements. When using reaction time techniques, one is using the subject's knowledge of the clues conveyed by natural speech for the determination of the signal's nature. The manipulation of the stimuli

can adopt two paths, the use of artificial or of natural stimuli. The use of electronically generated speech presupposes the knowledge of the crucial dimensions and parameters which are used by the perceiver in his/her decision on the nature of the incoming stimuli. At the time this work was carried out, commercially available electronic speech did not sound natural at all, and furthermore, the commanding software was of American origin imposing a striking American accent to the very American utterances it produced... This fact could disturb British born subjects when carrying out detection tasks.

The use of natural BBC English as the stimuli introduced some difficulties in the measurement of the reaction time and accuracy when deciding the signal they are hearing is speech. The onset of the stimuli would not be very well defined when using random samples of speech. See Chapter IV. The information content of the stimuli can easily be changed by shortening the duration of the stimuli, until the clues that the subjects use are no longer present or are too short to provide an accurate decision.

The comparison of the reaction times and accuracy of response to the speech stimuli with that to other types of stimuli such as music and electronic noise can provide some extra clues as to the manner this decision is carried out.

In order to make things more difficult for the subjects when directly labelling the stimuli being conveyed, Cello solo music was recorded and random samples of these solo passages were prepared. The electronic noise, which engineers call white noise, was filtered to equalize its long term frequency content to the speech and cello. This noise is called "pink noise".

There are difficulties in defining the beginning of the stimuli, and therefore the timing of the subject's responses, are avoided by using the subject's ability to predict empty periods of time. The stimuli can be conveyed at fixed periods of time to maximize readiness to respond. See Chapter IV.

The modus operandi of the experimental stage is beginning to emerge from the considerations above. A three forced choice type of experiment, using direct labelling while reducing the information content of the stimuli to the minimum, and the use of the regularity of the period of the interval between stimuli as the signalling clue brings the subjects close to a real situation when they are deciding that somebody is speaking to them. Whether this realisation comes about in the middle of a word, vowel or consonant, we do not know. It appears to be irrelevant.

## 1.7 The First Steps

Will the reaction times of the experiments proposed in the previous section be of one second, half a second, a few milliseconds?. Will the bandwidth of the stimuli affect its detectability?. The author was not aware of the magnitude of the variables or parameters under study. In order to acclimatise myself to psychophysical experimentation, a set of preliminary experiments was carried out to answer the questions which open this section. The method of limits was used. A detectability test was carried out comparing the responses to speech filtered down to 180 Hertz and comparing these responses to the accuracy in detecting pink noise. A replication of an experiment carried out first by Merkel and then by Hick (1952) was carried out in order to estimate the purely motor element of reaction time in a three choice experiment. The results and experience gained in these experiments was then used in the preparation of the main set of experiments.

## 1.8 The Experimental Stage

The reaction time and accuracy of the responses to a three way forced choice of natural speech, (BBC English); music, (cello solo); and noise, (pink noise) is recorded while varying the duration of the stimuli as a means to change the information content of the stimuli.

The experimental setup is briefly described in Chapter V and the apparatus, procedure, data gathered and its statistical analysis is presented in Chapter VI.

The interstimuli interval is then varied in order to study the effect of the readiness of the subjects on their performance. A simple comparison between the averages of reaction times (t-test) and accuracy of the responses in both conditions, using a fixed and variable interstimuli interval should reveal this effect, if any.

It has been recognised since the times of Broca, i.e. 1861, that one hemisphere of man's brain (usually the left) is specialised in speech function. In the studies of possible mechanisms used by humans in the decoding of the speech stream an advantage of the right ear over the left when processing some of the components of the speech waveform has come to light. This phenomenon, REA, is associated with the linguistic "decoding" of the speech stream.

REA is attributed to the stronger neural connections which exist between the ear and its corresponding opposite hemisphere. Speech received in the left ear will have to be routed through the hemispheric "bridge", the corpus callosum; to the speech related areas of the left hemisphere. This should result in a time difference for the processes of the decoding of speech relayed to the left and right ear. (Fry, 1974).

The study of presence of REA in Speech/Non-Speech discrimination can provide some more evidence for the "location" of this decision in an operational paradigm and at the same time, some estimate of the possible level of cortical involvement when we carry out this operation. See Chapter VI.

The suspicion of the speed and accuracy with which this operation is accomplished is confirmed by the results presented and discussed in Chapter VI.

The Speech/Non-Speech decision is carried out extremely fast under favourable conditions of attention and is the prior operation setting in motion all the subsequent speech decoding levels.

The brevity of the stimuli which can be described as speech and the lack of difference between the reaction times to speech, music and noise stimuli point to the fact that this comparison of stimulus category is carried out by a single operand, which is not "central" in its character.

CHAPTER II

# THE SPEECH CHANNEL SYSTEM

This chapter is the first of three which attempt to describe the framework of the problem under study. The ability to discriminate speech and nonspeech sounds is an operation of our processes of perception. The functional link of the perception of speech with its production are numerous and obvious. A profoundly deaf person will have great difficulties uttering a speech sound since he has never heard it before. Injury to the left hemisphere of the brain will affect the comprehension and/or the production of speech sounds. This intricate chain of functions and apparatus is the physical universe in which the Speech/Non-Speech decision is immersed. The description of this framework as a single system formed by three elements - the production apparatus, the brain or the relevant areas of it, and the hearing system - seems to the author to be the most profitable approach since no prejudgement of the 'location' of the Speech/Non-Speech decision is desirable for this work.

A brief description of the functioning of the production apparatus, some insights into the known relevant areas of the brain and of the hearing apparatus will lead us naturally to the next chapter which deals with the links and dynamics of these elements of the Speech Channel System.

## 2.1 Speech Channel, A System

To speak of the Speech Channel as a system implies that speech perception and its production form a distinctive part of the overall system of human perception. As an element of the human perception system, speech perception fulfills a particular role in the relationship of the human being with the perceived world. At the same time, it should present some of the operational characteristics of other varieties of perception.

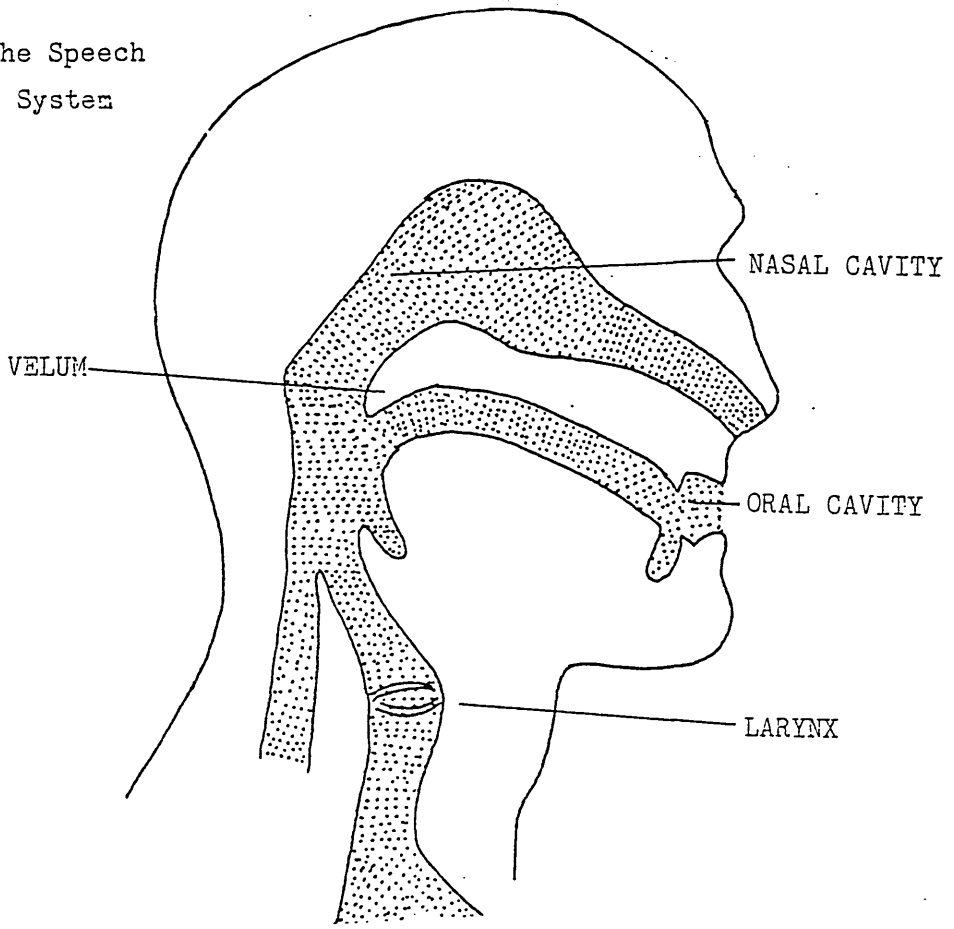
Most authors when discussing the speech perception and production processes treat these "extremes" or ends of the speech channel chain as isolated subsystems of the perception process e.g. Hayes (1973).

This chapter introduces the physical framework of the Speech/Non-Speech decision as an interrelated system formed by three elements: the production apparatus, the brain or the relevant areas of it and the hearing system.

## 2.2 The Production of Speech

Speech production system is the set of all those elements of our anatomy (non-neural) which contribute to producing an utterance.

Fig. 2.1 The Speech production System




It is possible to describe its functioning by breaking it down into two major components. (a) A series of resonant cavities and (b) Different sources of acoustic energy.

The sounds of speech are produced by air flowing from the lungs passing through the larynx up to the pharynx. This is coordinated by numerous muscular movements. The abdomen, thorax, larynx, tongue, lips, velum etcetera act in a perfectly co-ordinated fashion in order to produce a "simple" sound. The velum controls the amount of air flux passing through the nasal and oral cavity. When selecting the oral cavity more controls might be activated. It is then possible to change the resonant characteristics of the oral cavity by moving the lips, jaw and tongue.

One source of excitation of this system is the vibrations of the vocal folds in the larynx. This vibration produces a sequence of puffs of air which then pass into the vocal tract. This mode is called voicing or phonation and it is used to produce the sounds like /a/, /m/, etc.

The sound source of excitation in the production of our speech sounds is the turbulence caused by air passing through our articulatory system while keeping the vocal folds taut. This mode of operation is called "voiceless" and is used to produce the sounds like /h/ and when whispering.

The third sound source for producing speech sounds is a pressure built up behind some point of closure in the vocal tract. A sudden release of this pressure is used to generate the "stop consonants" e.g. /p/, /t/, /k/.

Nowadays, this simple analysis of speech production is behind the current trend of electronic "talker toys". In the past, Wolfgang Ritter von Kempelen (1734 - 1804), (Cherry 1978), produced an instrument capable of imitating human speech. Its operation, although very complicated, contained the source of energy (bagpipe) and the "articulatory controls" were the fingers varying the characteristics of the resonant cavities.

The "talker toys" of the present day are constructed on the same principles. A straight forward replacement may be deduced from the following diagram. (Texas Inst., 1978)

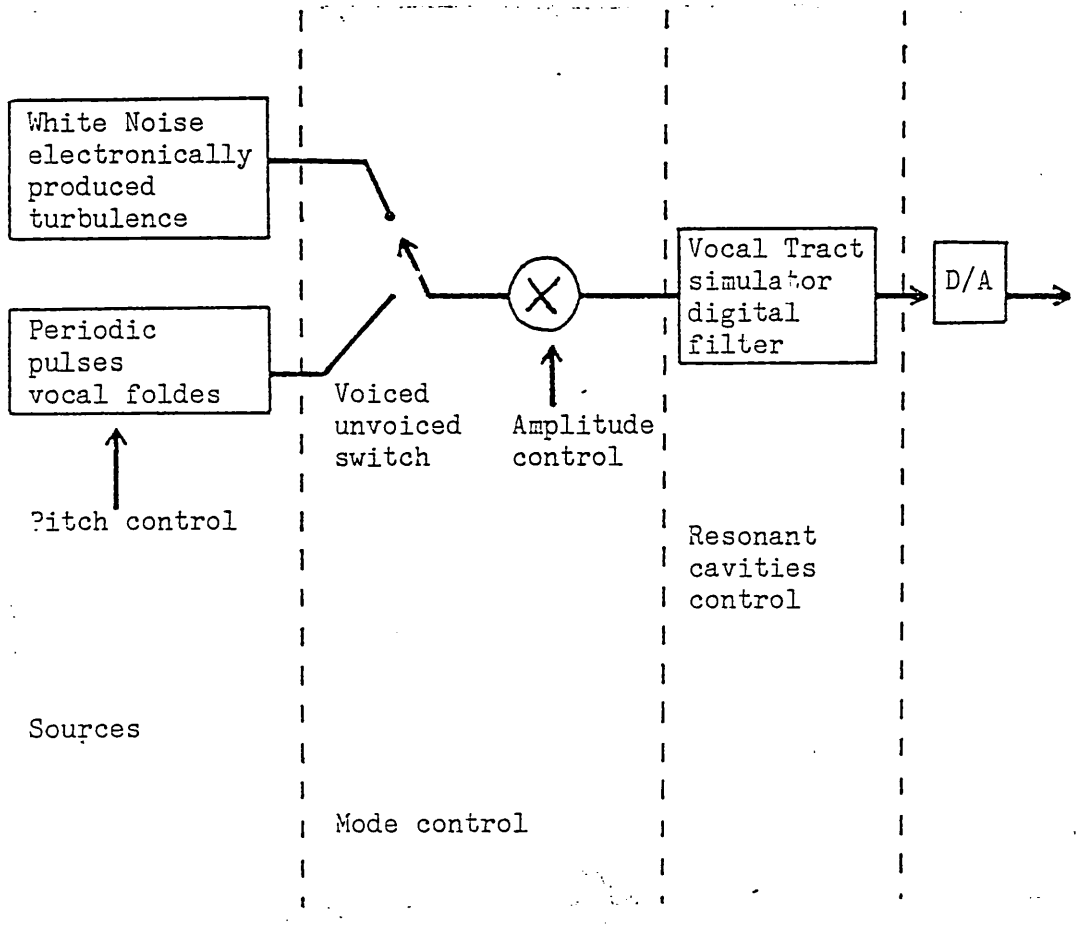



Fig.2-2. Block diagram of the single chip L.P.C Synthesiser.

The bagpipe and fingers are nowadays replaced by a silicon chip, yet the principle is the same as 200 years ago.

The understanding of this principle has allowed not only profit-making but also the implementation of therapy for people whose larynx has had to be removed. They are trained to use their oesophagus as a vibration source and the stomach as the air reservoir.

### 2.2.1 The Sounds of Speech

The speech production mechanism is capable of producing a great variety of acoustically different sounds. In a given language some of these sounds are interpreted by the listener to be linguistally equivalent, i.e. the sounds n and ɳ in Italian are equivalent because there are no two different words which have different meanings and differ in just these two letters. In English the words sign and sin are of course different. Any set of sounds which are different but linguistically equivalent is classified in a single unit called phoneme. A Phoneme is the minimum element which distinguishes one word from another in the language.

The international Phonetic Alphabet shown in Fig. 2.3 provides a standardised set of symbols for classifying speech sounds. Their pronunciations are given by the key words of figure 2.4.

Linguists depart from these elements in their construction of a logical set of rules which define the various levels of grammar.

The description of the production apparatus and its product could be extended, but as the hypothesis of this work is that the Speech/Non-Speech decision has no linguistic character, we abandon the linguistic universe. The speech metalanguage is useful insofar as it allows us to grasp some of the conceptual universe which is present in the characterization of speech as a product of human beings.

| | | Bi-labial | Labio-dental | Dental and Alveolar | Retroflex | Palato-alveolar | Alveolo-palatal | Palatal | Velar | Uvular | Pharyngal | Glottal |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CONSONANTS | Plosive | p b | | t d | ʈ ɖ | | | c ɟ | k g | q ɢ | | ʔ |
| | Nasal | m | ɱ | n | ɳ | | | ɲ | ŋ | N | | |
| | Lateral Fricative | | | ɬ ɮ | | | | | | | | |
| | Lateral Non-fricative | | | l | ɭ | | | ʎ | | | | |
| | Rolled | | | r | | | | | | R | | |
| | Flapped | | | ɾ | ɽ | | | | | ʀ | | |
| | Fricative | ɸ β | f v | θ ð \| s z \| ɹ | ʂ ʐ | ʃ ʒ | ɕ ʑ | ç j | x ɣ | χ ʁ | ħ ʕ | h ɦ |
| | Frictionless Continuants and Semi-vowels | w \| ɥ | ʋ | ɻ | | | | j (ɰ) | (ʍ) | ʁ | | |

| | | Front | Central | Back |
|---|---|---|---|---|
| VOWELS | Close | (y u ʉ) | i y ɨ ʉ | ɯ u |
| | Half-close | (ø o) | e ø ɘ | ɤ o |
| | Half-open | (œ ɔ) | ɛ œ ɜ ɞ | ʌ ɔ |
| | Open | (ɶ) | a | ɑ ɒ |

(Secondary articulations are shown by symbols in brackets.)

OTHER SOUNDS.—Palatalized consonants: ʲt, ʲd, etc.; palatalized ʃ, ʒ : ʃʲ, ʒʲ. Velarized or pharyngalized consonants: ɬˠ, ðˠ, zˠ, etc. Ejective consonants (with simultaneous glottal stop): pʼ, tʼ, etc. Implosive voiced consonants: ɓ, ɗ, etc. ɾ fricative trill. σ, ʚ (labialized θ, ð, or s, z). ʃʷ, ʒʷ (labialized ʃ, ʒ). ǂ, ǁ, ɓ (clicks, Zulu c, q, x). ɺ (a sound between r and l). ɳ Japanese syllabic nasal. ʂ (combination of x and ʃ). ʍ (voiceless w). ɪ, ʏ, ʊ (lowered varieties of i, y, u). ɜ (a variety of ə). ɵ (a vowel between ø and o).

Affricates are normally represented by groups of two consonants (tʃ, dʒ, etc.), but, when necessary, ligatures are used (ʦ, ʧ, ʤ, etc.), or the marks ‿ or ⁀ (t͡s or t‿s, etc.). ‿ ⁀ also denote synchronic articulation (m͡ŋ = simultaneous m and ŋ). ɕ, ʑ may occasionally be used in place of tʃ, dʒ, and ɜ, ƨ for ts, dz. Aspirated plosives: ph, th, etc. r-coloured vowels: eɹ, aɹ, ɔɹ, etc., or eʳ, aʳ, ɔʳ, etc., or e˞, a˞, ɔ˞, etc.; r-coloured ə: əɹ or əʳ or ɹ or ə˞ or ɚ.

Fig. 2-3 The International Phonetic Alphabet (taken from a publication by the INTERNATIONAL PHONETIC ASSOCIATION, 1964)

| Phoneme category | Symbol | Key word | Relative frequency |
|---|---|---|---|
| Vowel | i | eve | 1.82 |
| | e,eI | hate | 1.94 |
| | ɛ | met | 2.92 |
| | a | cat | 2.00 |
| | ɑ | father | 2.06 |
| | ɔ | all | 2.13 |
| | o,oʊ | obey | 1.93 |
| | u | boot | 1.20 |
| | I | it | 9.22 |
| | ʊ | foot | 1.53 |
| | ʌ | up | 1.32 |
| | ə | about | 7.26 |
| | ɚ | over | 1.77 |
| | ɝ | bird | 0.28 |
| Dipthong | aI | die | 2.60 |
| | aʊ | out | 0.68 |
| | Iu | new | 0.22 |
| | ɔI | boy | 0.06 |
| Fricative | f | for | 1.48 |
| | θ | thin | 0.81 |
| | s | see | 3.74 |
| | ʃ | she | 0.49 |
| | h | he | 1.49 |
| | v | vote | 1.76 |
| | ð | then | 3.62 |
| | z | zoo | 2.13 |
| | ʒ | azure | 0.00 |
| Stop | p | pay | 1.37 |
| | t | to | 9.11 |
| | k | key | 2.93 |
| | b | be | 1.44 |
| | d | day | 3.78 |
| | g | go | 1.39 |
| Nasal | m | me | 3.16 |
| | n | no | 6.43 |
| | ŋ | sing | 1.27 |
| Glide | w | we | 2.74 |
| | j | you | 2.40 |
| Semi-vowel | r | read | 3.31 |
| | l | let | 3.62 |
| Affricate | t | chew | 0.31 |
| | d | jar | 0.28 |

Fig.2-4 Phoneme pronunciation guide (key words from FLANAGAN, 1965)

## 2.3 Perception and Production of Speech

When producing speech sounds, a complex system of feedback appears to be essential for the production of speech. This feedback occurs through our speech perception system.

It is a very well known fact that distorted or non-existent feedback disturbs our production of speech. (Békésy 1971, Cherry 1978).

The profoundly deaf child is not aware of the loudness pitch of his/her own speech. Various methods have been tried in order to give to the child the necessary feedback. W. Edmondson (1973) studied the improvement of the child's speech with tactile feedback.

A simple experiment to assess the effect of aural feedback upon speech might be tried. The reader should obtain or borrow a tape recorder which can simultaneously record and monitor his own speech. By varying the delay between the recording and monitoring head the reader will find his own speech disturbed. There are sometimes benefits from the distraction of the aural feedback. A device which picks up the vibrations of the larynx and relays a loud pitch to the stutterer's ears has been recently launched into the market. An impressive fluidity of the stutterer's speech is claimed. This effect was demonstrated by Cherry 30 years ago. (Cherry, 1953).

The vibrations of our vocal chords not only produce sounds which are fed back to our hearing system through the air but also induce the jaw to vibrate. The vibrations are in turn transmitted to the ear canal (Békésy 1971).

The self monitoring of one's speech is done through two different paths. Other listeners to our speech hear only our air conducted sounds. These sounds lack some of the low frequency components produced by the vocal folds. This is the reason why recordings of our own voices sound strange to us.

## 2.4 THE "RECEIVER", THE HEARING SYSTEM

As a telecommunication engineer one tends to compare our hearing system with a receiver (a black box) which does marvellous things. In this section it will be shown that the comparison is not at all appropriate.

In section 2.3 a few links between the production and perception of speech have been mentioned. In the description, the working of the hearing system appears naturally as one of the components of the speech perception production system. It should be noticed that until now this chapter has merely been pointing out the links between both systems. It is a working assumption of this thesis that there is no supremacy of speech perception over its production. It is also assumed that both the hearing system and production systems are related together like "the chicken and the egg".

From observations of the child's speech we infer that the link between speech perception and its production is the base upon which we learn to produce and perceive a speech sound (Cherry 1978).

The human ear and its associated systems constitute one of the wonders of our anatomy. Its performance is far from being challenged by any modern technological achievement. From Békésy (1971) we learn that the hearing system is so sensitive that it can detect the random run of air molecules bouncing against the eardrum, yet it can cope with an enormous amount of acoustic energy e.g. a pop concert.

At some audible frequencies the deflection of the eardrum when vibrating is as small as one billionth of a centimeter ($9 \times 10^{-10}$ inches) - about one tenth the diameter of the hydrogen atom. Deeper inside the ear at the point of the Mechanical (Analog)/Neural (digital) converter these vibrations are around 100 times smaller in amplitude.

One of the most popular ways to investigate the performance of the human ear is the pure tone audiogram.

In this measurement a pure tone is used to assess the subject's ability to detect it at various frequencies. Despite the fact that this measurement provides a method to measure potentiality to perceive it is very difficult to draw a line between normal and abnormal audiograms (Licklider, 1951).

One must not be discouraged if one's audiogram does not follow a "normal" one. Statistically, there is a fall of audible frequency with age. In childhood some of us could have heard frequencies up to 40,000 cycles. For those who are in their forties the upper limit drops 80 cycles per second every six months in a period of 5 years (Békésy 1971).

A diagram of the internal structure of our ear is given in Figure 2.5.

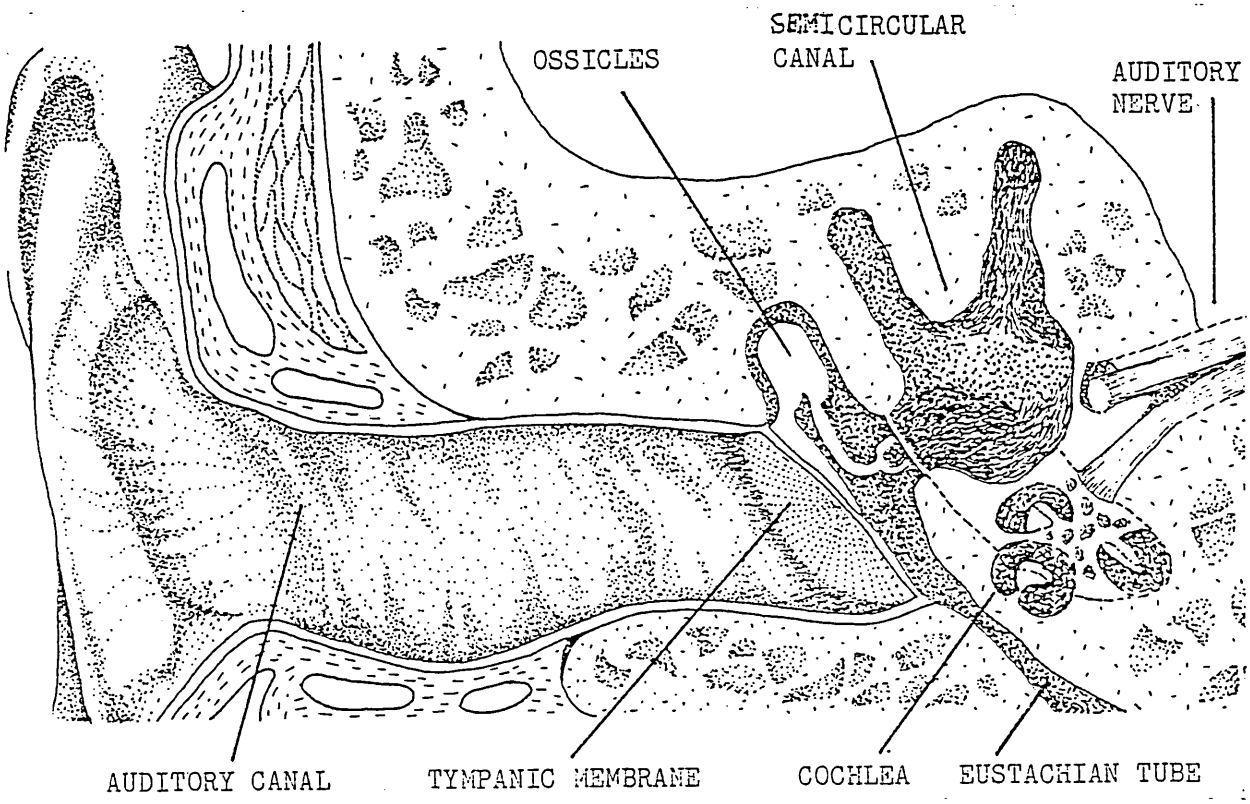



Fig. 2-5. Cross section of the inner ear.

Its working is presented in a rather systematic fashion in Figure 2.6.

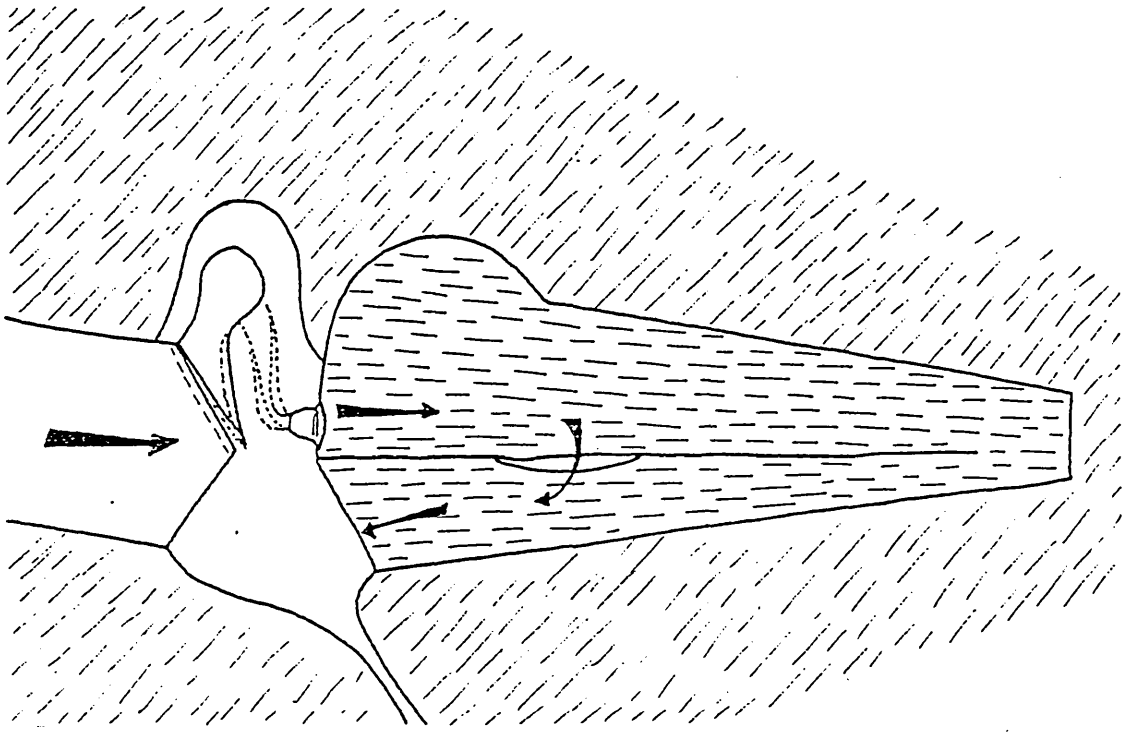


Fig. 2-6 Aural to Neural transformation of the stimuli in the Ear.

The sound waves set the eardrum (tympanic membrane) vibrating, these vibrations are then transmitted by a series of small bones to the fluid of the inner ear.

The pressure waves in the inner ear form a faithful representation of the sound waves, which in turn acts upon the walls and the membrance of the cochlea. At this point the acoustic/neural translation transformation is done. More detailed information can be found elsewhere in the literature cited.

The neural impulses are then "routed" to the various areas of the brain involved in our aural perception. It must be pointed out that the foregoing account is just the tip of the iceberg. A multitude of theories are still being confronted with various empirical aspects of hearing being newly discovered every day.

## 2.5 A Third Element

In this chapter we have presented the most prominent components of a system which we call the speech perception/production system.

Our "receiver end" has been linked with our "productive end" by their common product, the language. The physical counterpart that links both ends is obviously our brain. In the following sections a brief glimpse of the known aspects of this link will be presented.

A hundred years ago a noted French surgeon, Paul Broca, performed post mortem examinations on the brains of patients who had developed serious speech defects. These studies supplied the first indications that damage to specific portions of the brain can be responsible for speech disturbances (Wooldridge 1963).

Nowadays, some areas of the brain have been detected as being mostly responsible for some functions of the speech production-perception chain.

Alexander Luria distinguished three main blocks of the brain (Calder 1970). Fig. 2-7

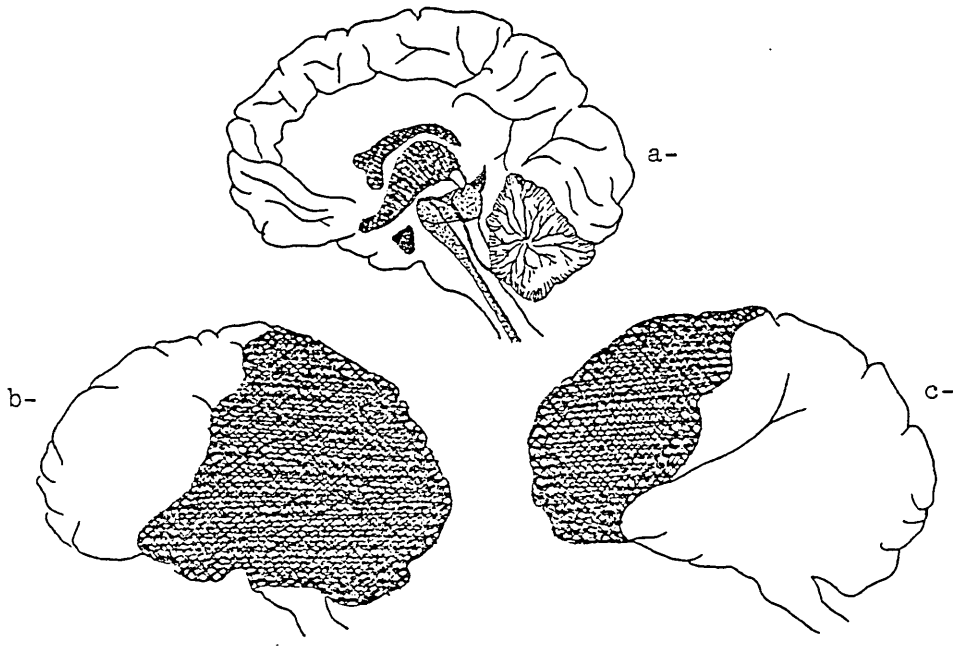



Fig.2-7. Main blocks of the brain.

(a) The brain stem, commanded by the thalamus is concerned with the "triggering" of action of the whole brain. It is a sort of maingate to the brain which connects us with the outside world.

(b) The rear part of the cerebral cortex concerned with analysis and storing information.

(c) The front of the brain, of elusive function but probably concerned with intentions and plans.
A more specific mapping of some functions of the left hemisphere has been taken again from Calder (1970). Fig. 2-8.

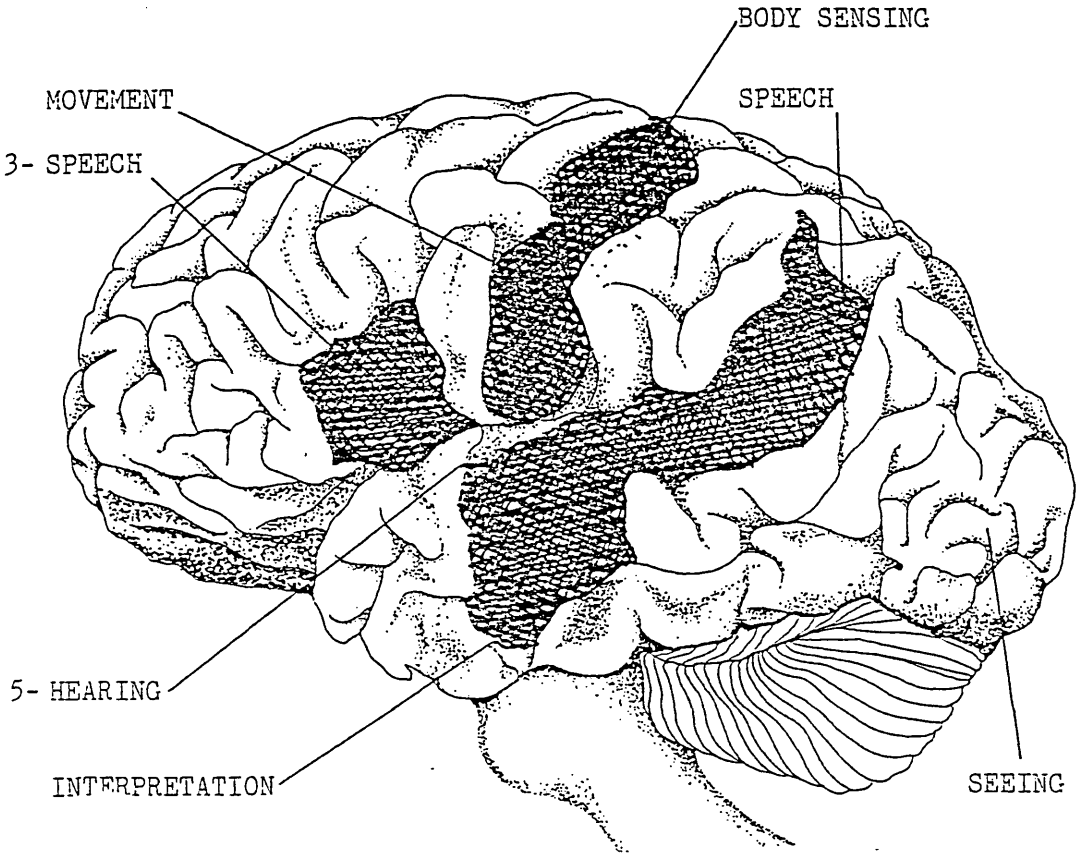



Fig. 2-8. Some of the functions mapped in the left hemisphere of the brain.

Some other scholars name the areas 3 and 5 as Broca's Area and Wernicke's Area, which are mainly concerned with speech.

The control of movement in different parts of the body as allocated along the motor strip is depicted in Fig. 2.9.

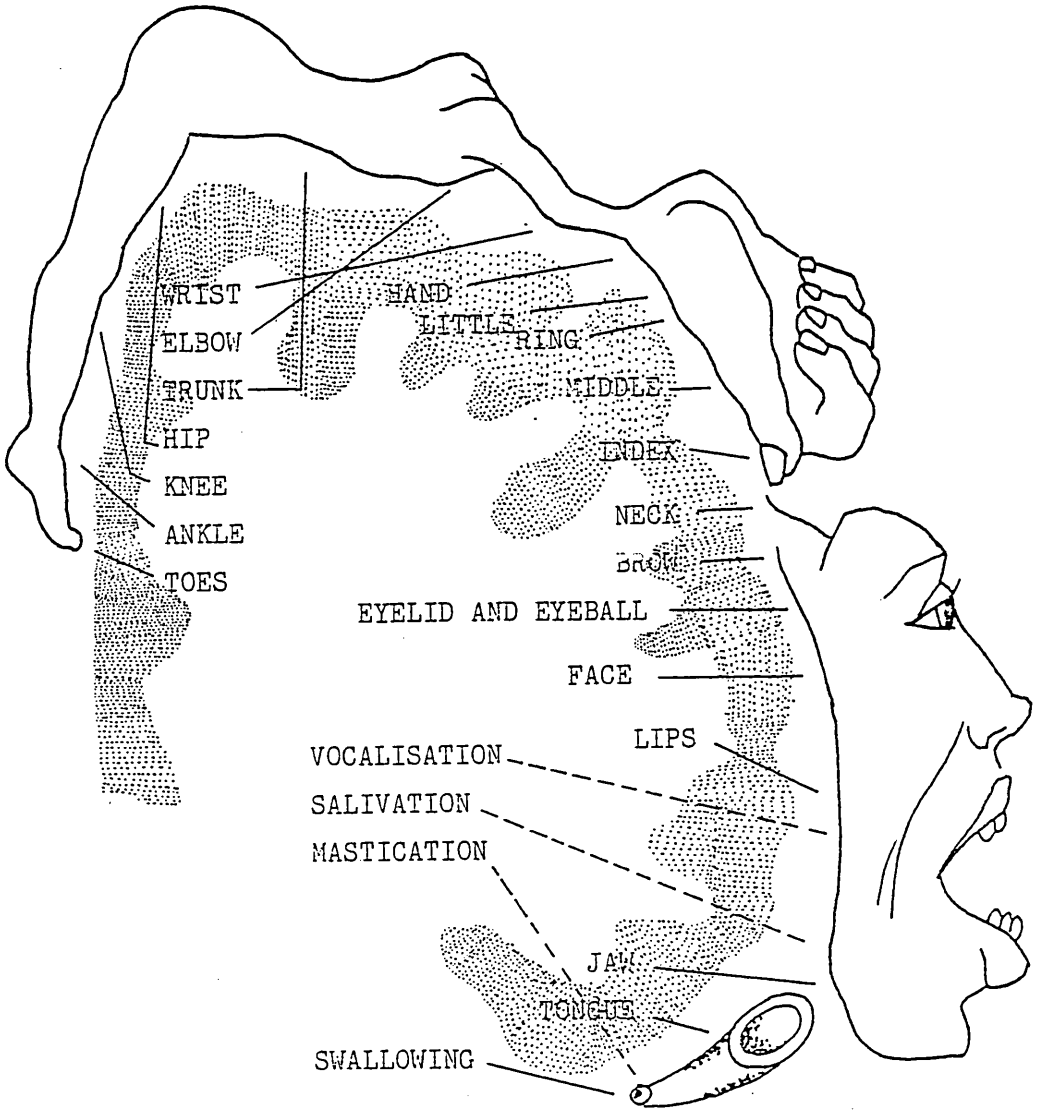



Fig. 2-9. How control of movements in different parts of the body is allocated along the motor strip.

The reader will notice the large areas dedicated to the hand and mouth. These areas are mapped in a rather crude technique. A tiny electrode is applied at the area under investigation while the human guineapig, with the skull opened, is conscious to describe the effect of such applications. Penfield reports that an electrical stimulation of some parts of the speech area leaves the patient able to speak but unable to name the object which he identified as presented to him.

Further observations of the correlation of brain injuries and word perception are reported by (Luria 1973). Certain patients were

able to attach meaning only but not words to simple and common sounds, like the knock on a door, fire bells etc. No attempt tc investigate the Speech/Non-Speech decision is reported.

This technique in searching for areas of the brain must not lead us to believe that there is a one to one correspondence between areas and functions of the cortex. Although the speech area is located in the left part of the brain in most of the population, this can be changed. The speech zone can be transferred to the right hemisphere if the left is damaged. This transfer is more successfully accomplished when the patient is younger than 14/15 years old.

This plasticity appears to be a property of the brain tissue involved in all higher intellectual processes; in marked contrast to the rigid assignment of specific areas of the brain for the receipt of sensory input and reflect processes (Nathan, 1969).

## 2.6 The Brain Stem

The crucial role in the integration of the functioning of the areas mentioned in foregoing sections appear to be performed by the thalamus (Calder 1970). Fig. 2-10.

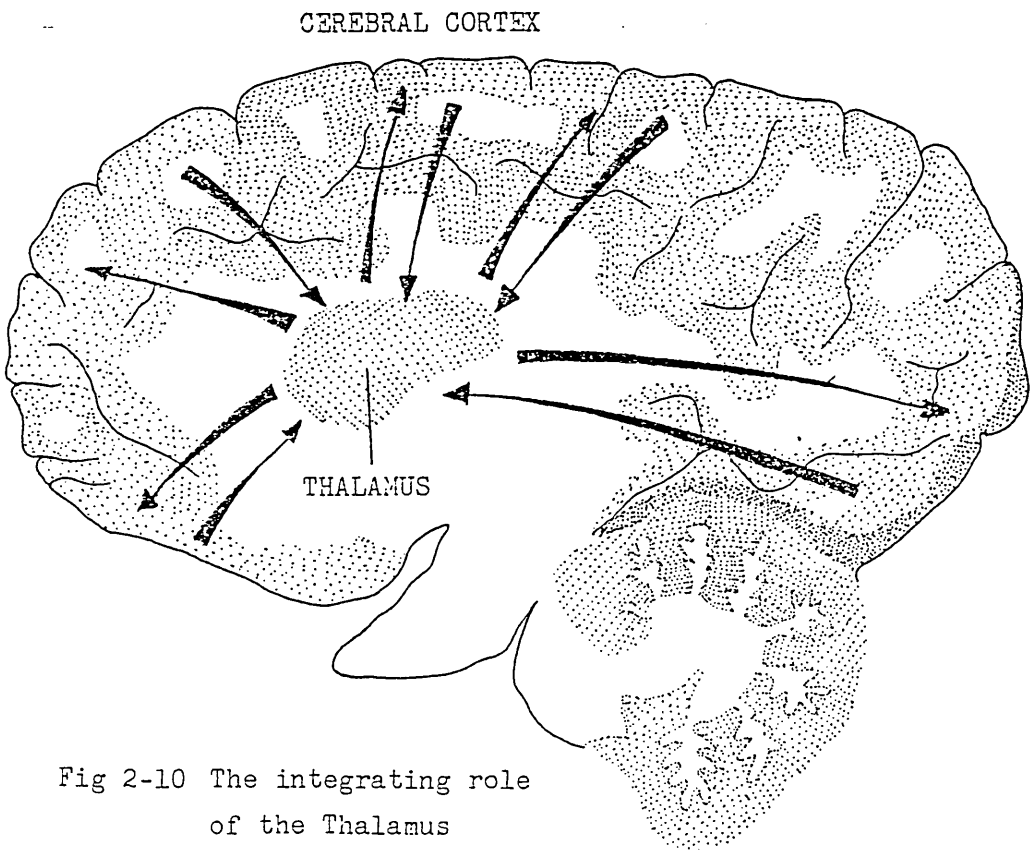



Fig 2-10 The integrating role of the Thalamus

Zangwill (1960) also mentions the components of the motor system, together with those areas of the cortex that subserve the functions of speech.

The neural pathway between the auditory cortex and the hearing apparatus has been under scrutiny for a long time but it is far from having been described as thoroughly as the neural pathway of vision. No feature detector has been successfully located in the aural neural pathway.

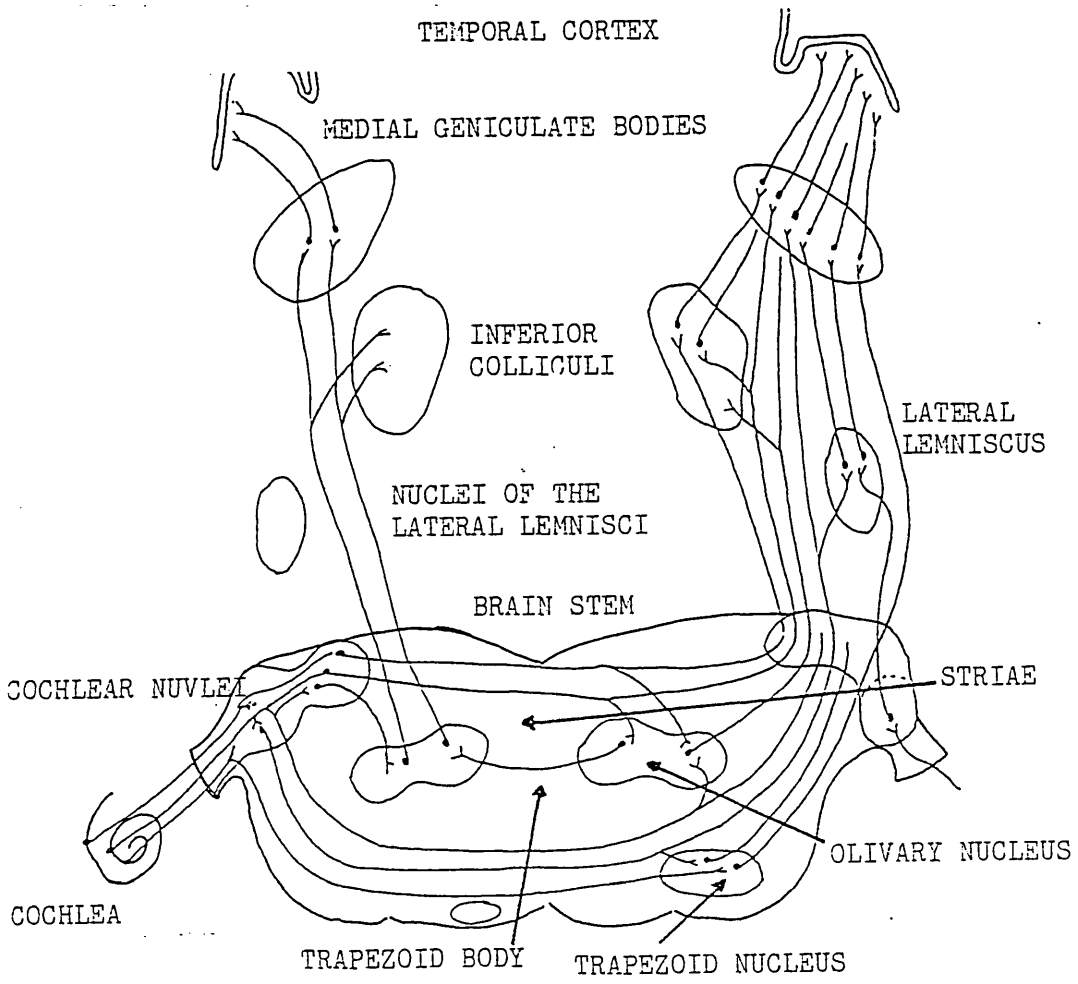



Fig. 2-11. The central auditory connections.

The diagram represents the stronger neural connection between opposite ear and hemisphere. The implications of such a physiological feature are still under detailed investigation. The role of the

various nuclei along the auditory neural pathway is not clear at all. Audiometry researchers have developed a method to measure the electrical activity of the auditory neural pathway by detecting the instantaneous electrical potentials generated along the path and averaging their values. Hood (1975)

This technique allows the researchers to obtain an intensity - time representation of the E.P. (Evoked Potential) which in turn might represent the "activity level" of the various nuclei. Audiometry of non co-operative patients can be obtained using these techniques.

This chapter has introduced the basic aspects of speech production, its product, the sounds of speech, the hearing apparatus and the linking organ, the brain. Special reference is made to its speech related areas. The incorporation in just one chapter of these three elements is not only an overall simplification as its brevity suggests, but it underlies the aim of the chapter which is to integrate the universe in which Speech/Non-Speech decision might lie.

This overview will lead us to the next Chapter in which an attempt is made to describe the various processes and dynamics the speech production and perception system undergoes in order to produce and understand a spoken word.

CHAPTER III

DYNAMICS OF THE SPEECH PERCEPTION SYSTEM

3.1 In this chapter the various processes involved in the production and perception of a "human made" speech utterance are explored.

In the previous chapter an overview of the main elements of the system under study was presented. In contrast, this chapter aims to examine the relationships established between the elements making up the speech perception system which enables a person to perceive or utter a word.

Speech perception is a special perception mode, different from the perception of other orally conveyed signs (Webster et al 1968). Notwithstanding this specifity, speech perception is a component of our macro system for apprehending the surrounding and inner world. This fact imposes some general characteristics upon its operation.

The short-time dynamic (the processes involved when an oral sign is meaningfully decoded) of word sign apprehension has been under close investigation for a long time. A very schematic presentation of these decodification - codification chains is attempted in the first section of this chapter.

The developmental study of the acquisition and mastering of our speech, called ontogeny of speech, is another avenue to explore. Are we born with a dictionary somewhere in the brain? Does a new born child always perceive a human voice as such? By exploring these problems some characteristics of the Speech/Non-Speech decision might come to light.

If we want to discover whether man is specialized to process speech so as to receive phonetic segments, we must make the appropriate comparisons with animals. (Liberman 1976). Some insight into species-song detection in other than the human animal, could provide some clues to characterize our problem.

Indulgence is sought from the scientists and scholars affected by this crude presentation of their complex province of knowledge.

## 3.2 On Perception

Perception of speech has nothing to do with photography nor with templates. When we perceive a word, we make an inference based upon a mediator, our language. To paraphrase Gregory (1968) and Cherry (1978), the perception of an object goes beyond the sensory information. This elevation, transformation of the incoming data into a different category of information, is the result of the interaction of all the dynamics mentioned in the foregoing section of this chapter. In Desheriev's (1979) words, "language and its perception is the result of a long process which has absorbed and condensed all the historical data collected by the individual and his society". This crystallisation of concepts which becomes a mechanised operation in our development is decribed by Craik (Blakemore, 1977) as a model of the world. As a scientist or engineer one is tempted then to model our perception processes in an engineer like fashion. Information theory provided the basis for this trend. Using a left-to-right approach to the generation and apprehension of an aural sign, the likelihood of a word or syllable being followed by another word or sound came to be seen as the mechanism in the creation or understanding of a word or phrase. This argument is a literal translation of the serial or parallel modes of thinking. Neisser (1963), from Greene (1975), points out that both modes of operation are present in human thinking, but one process can be "focused" at a time by our conscious attention.

The fact that we have one mouth which forces any parallel or serial construction of a phrase to become a serial sequence of sounds, has nothing to do with the way we apprehend or plan a given concept conveyed in an oral manner.

We have adopted Cherry's approach for the development of this concept. Perception cannot be described by a probabilistic analysis of all the incoming data being gathered by our sensory channels. Human beings do choose among a great variety of data.

Our attention is directed to a voice picked up by us in a "cocktail party" situation. No assessment of the significance of all the voices is made. In fact, we choose rather than select our stimulus (Cherry 1978). This statement implies a conceptual difference in the use of the words 'choice' and 'selection' which in our context, are correlated with a still mysterious mechanism of attention focusing and Bayesian selection.

There are features of this process which might lead us in to deep waters which are more easily navigable by philosophers. This thesis does not attempt to enter the philosophers' province. It will merely be based upon the basic assumptions mentioned above.

## 3.3 Language and Speech Channel

The study of linguistics has achieved a major goal in its efforts to characterise the object under study, the natural language.

Luria (1974), points out two significant steps which have taken linguistics into the category of the so called exact sciences.

Trubetzkoy and later Jakobson proposed a set of binary characteristics related to our speech production organs' modes, with which it is possible to reduce the entire wealth of sounds of all languages.

A similar approach was adopted by Chomsky (1972) who demonstrated that the same principle of reducibility can be applied to syntactic structures, to the manner of conveying meaning through a structure formed by nouns, verbs, articles, etc.

The construction of the model of the generation of our language is started then at the acoustic level when the object to be perceived has already been treated as speech, as a stream of sounds which need linguistic decoding.

Our account of the linguistic formalisation of our language does not need to go further than the serial description of the various stages of language decoding Fig. 3.1.

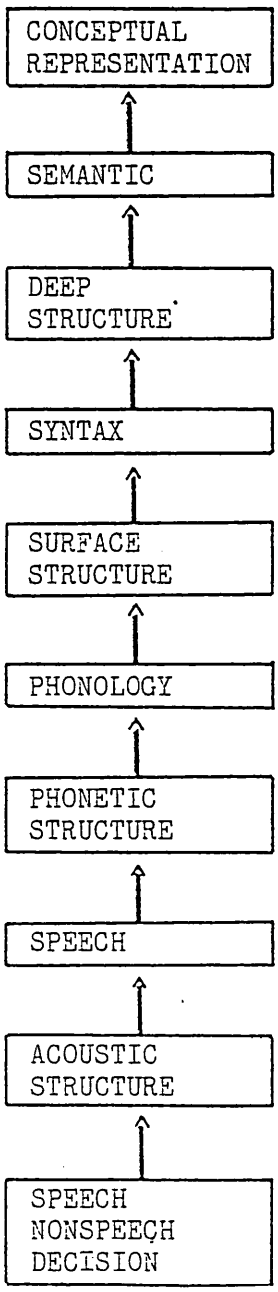



Fig.3-1. The speech chain.

The translation of the linguistic heterarchy into the psychologist's world also underlies Hirch's description of the speech chain: "The understanding of speech requires that the sounds be audible (i.e. detectable), that the different sounds of speech be discriminable one from another, that the listener add to this auditory discrimination some word memory so that he can recognise speech sounds, and finally, that the entire series of processes be stored and organised in time to permit the comprehension of long sequences of words". (from Bench 1970).

Cutting (1976) also approaches the study of speech perception as a process based on the following premises:

(i) Speech production and its perception are processes that require time to be accomplished.

(ii) A logical analysis of the language is carried out, breaking it into units (or the reverse, a synthesis).

(iii) Some transformation rules (grammar) are followed when the incoming material is transferred from one category to the next.

All these accounts, in "black box" style, seem to be also based on the assumption of symmetry between the speech perception process and the process of generation of speech.

Haggard (1977) pointed out the excessive partisanship of some researchers when interpreting perception experiments based on some particular linguistic theory. Luria (1974) warns us of the dangers in transforming an abstract structure into the general functioning of the mind. Desheriev (1979) reacts strongly against those "who assign to grammar the role of a model of the reality, over-estimating the role of the linguistic categories in the reflection of the reality". Benjamin Whorf represents the opposite trends in the polemic in assigning to the language the role of conditioning the perception of the world. (Cherry 1978).

In some of these abstract descriptions of speech it is possible to find a place for some functional "black box" which may set our "linguistic processors" into action. The ommission of such a crucial operation from the psycholinguist's scheme seems due to the non linguistic nature of the speech nonspeech decision.

The "black box" accounts of the speech production-perception chain does not provide a location for the perception function which is the concern of this thesis. It will be therefore, an hypothesis of this work that the speech nonspeech decision is at the limit of the proposed psycholinguistic chain. It affects the subsequent steps in the linguistic decoding of an aural stimulus. For example the disruption on the categorisation of CV sounds when subjects are told that the stimuli is not speech. A major review of these speech nonspeech related effects can be found in B. Repp (1982).

## 3.4 Psycholinguistic Experimentation

Psycholinguistic experimentation has approached its goals through a linguistic frame-work. In this section of this chapter a few experiments which have some relevance to our work will be presented. Their results are still contentious but they reflect some facts which bear on our problem.

Our research launched us into a jungle of experiments, results and mini-theories which are attempts to link the dynamics of some generative grammar with observations of speech behaviour. Notwithstanding this fact, there are some collateral observations which are present in some of these experiments that can be interpreted in support of our thesis.

Psycholinguists have been trying to discover the mechanisms which might help us to define a given linguistic "black box" in psychological terms.

An example of this attempt is the work with "visible speech", the frequency time-intensity characteristics of the different phonemes. Nowadays, it is a daily occurrence to find talking calculators, computers, etc., whose operation is based on their results.

Thus M. T. Turvey and S. Sears (1976), characterise three different "modes" of perceiving.

(i) A semantic mode in which we perceive the meaning of what we hear.

(ii) A phonological mode in which what we experience sounds distinctively.

(iii) An acoustic mode in which we experience certain non-linguistic aspects of speech.

At the phonological level a phenomenon called "categorical perception" demonstrates the specifity to discrete categories (of "analog variations") of some acoustic variables affecting speech perception. This is a mechanism present in all our perceiving channels.

Cutting (1975) analysing a list of six phonological mechanisms presented by Wood (1975) in support of the specificity of the speech perception, reduced these to four. Three of Wood's mechanisms are related to the same fact: The speech perception system is mainly controlled by the left hemisphere of the brain. Cutting gave some examples of Non-Speech perception modes where these mechanisms were also present (i.e. music), and concludes that there is no known mechanism at the phonological level which is specific to speech perception. Haggard (1977) considers that these interpretations of psycholinguistic experiments, "are based upon partisan emphasis derived from traditional concepts in linguistic theory". They certainly do not help in the search for the location of the mechanism for the Speech/Non-Speech decision.

Cutting's conclusion is indeed specific to the "phonological black-box-decoder" but in his analysis he pointed out that speech is a special mode of perception.

Our search for relevant works in this field of literature gave a negative result: It is still necessary to continue the search for a mechanism in the auditory system which is used specifically to perceive speech as distinct from nonspeech. From Chapter I we recall the heuristic suspicion that the speech nonspeech discrimination is a non-linguistic event. A search in the boundaries of this field therefore might yield some useful information.

Cherry (1953) in a study of the fusion of information from both ears comments upon some "oddities" of the shadowing experiments. Subjects were presented with two messages through headphones, one to each ear. They were to shadow, or repeat aloud one and to ignore the other. The experiment was designed to measure the amount of disruption caused by the ignored message upon the shadowed one. His subjects were always able to detect the presence of speech in the channel. They also were capable of discriminating male from female voices yet they were not able to make sense of the non-target message. As the attention of the subjects was forced on to the apprehension of the message-to-be-reproduced and therefore occupied, we conclude that speech nonspeech detection is a peripheral function.

Lindsay (1977) confirms Cherry's observation and adds some characteristics to the detection of the unattended message. Lindsay's subjects reported that:

(i) They remembered whether a voice was present or not.

(ii) They noticed other classes of stimuli such as a whistle.

(iii) They could not remember the content of the non-target message nor the language in which it was spoken, nor its semantic coherence.

Again these results reinforce our suspicion regarding the non-linguistic characteristics of the speech nonspeech decision, since nothing else specific to the perception of language was recalled, other than the presence of voices.

In the foregoing we have presented some characteristics of the speech nonspeech decision which have been inferred from results of experiments not specifically designed to study the Speech/Non-Speech function of our perception system. In order to build a theory about a perception phenomenon there is an obvious need to characterise its singularities. Haggard (1977) presents three propositions:

a) "The sounds of speech are themselves so radically unlike any other class of sound or visual object that the question about differences in their mode of processing barely needs to be asked.

b) Qualitative phenomena shown in speech experiments are quite different from those shown in other experiments.

c) General purpose computer power in the brain is used in a specific way or special centres are activated when a combination of stimulus and task factors precipitate a switch into speech mode".

Most researchers adopted the path described in proposition (b) to conclude after years of hard work and a lot of papers, that mechanisms such as categorical perception, right ear advantage, etc., are also used in the perception of other classes of aural signs i.e. "plucked" and "bowed" sounds.

From the multitude of results produced at Haskins Laboratories also, following Haggard's proposition (c) and taking into account the guidelines contained in the first sections of this chapter, we will continue the review of results which show some characteristics of the Speech/Non-Speech decision.

This "Switching" action as presented by Haggard affects the results of psycholinguistic experimentation. Subjects respond differently to speech-like stimuli after a degree of experience that enables them to hear the stimuli as speech (Haggard 1977). This "Switching" cannot be modelled on a telephone exchange, as demonstrated in the shadowing experiments, and in the following experiments.

Whistles, knocks and some other kinds of sound "go through" the attention channel. Memory recall can be an explanation for these results which will be explored in the following sections of this chapter.

Continuing with the shadowing experiments in order to study further aspects of the perceptual decision under study, we find a very intriguing one. Lindsay (1977) reports the following experiment. A subject is asked to attend (listen) and repeat immediately a message which is conveyed to his left ear. A different message is relayed to his right ear, the messages possess a cross complementary "meaningfulness". Fig. 3-2.

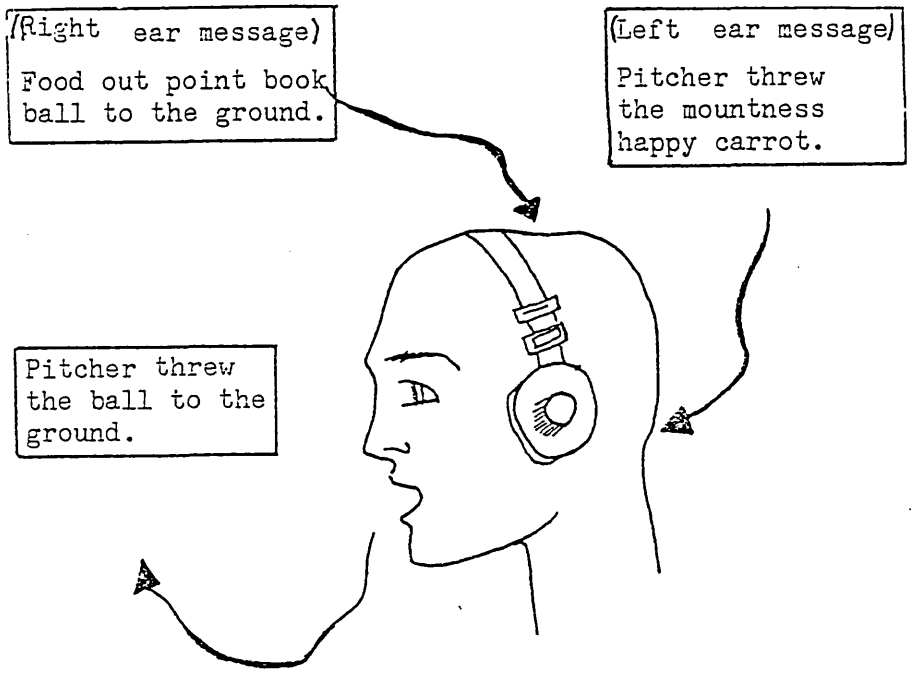



Fig. 3-2 . A shadowing experiment.

In this experiment subjects are not only able to distinguish whether or not the shadowing message is speech or noise but they are also capable of detecting the meaning - complementation and override - the conscious will to repeat the object-message. This fact adds some more difficulties to the serial psycholinguistic heterarchy model of speech perception. "Central" decisions are made upon the shadowing message yet no phonological analysis of it was detected in other

experiments. For the purpose of this thesis the fact that a decision was made about the nature of the non-target stimuli still remains valid.

## 3.5 Memory and Speech Perception

Every human being is well aware of his capabilities to store aurally conveyed stimuli - "It sounds like ....." is a common phrase in our every day relation with the aural world. This capability enables us to speak, write, listen, etc. Our memory system is capable of the most complex operations. One can remember sounds, tastes, feelings, images and so on through our highly detailed record of sensory experiences.

Memory, although a unitary process can be described as having different aspects in its operation.(Clark,1977)

i) A memory system which has common characterists across our perception system and its various "channels". This system is called "Sensorial Storage" and it holds the information for about 100 to 500 milliseconds. (Lindsay 1977).

This memory comes into action when we ask our interlocutor "What did you say" and simultaneously with this phrase, our interlocutor's word comes to mind.....

A.D. Baddeley (1971) distinguishes between the experimental results and models of memory systems: Sensory memory, Primary memory and Secondary memory which differs in the linguistic category of the information stored.

Sensory memory is that memory which holds the incoming stimuli in its original form. In the auditory system this memory has been called "echoic" or "precategorical memory". The terms used are "echoic" memory because the aural stimuli is recalled like an echo and "precategorical" because no linguistic categorization operates upon this material.

Primary and Secondary memory are those stores of our memory system which hold a linguistically converted version of the incoming stimuli i.e., phonemic and semantic coding (Baddeley et al, 1971). These expressions of the memory system are experimentally measured and named Short Term Memory. Its "holding" time is limited if the material is not rehearsed, at the same time its capacity as regards the amount of information held in it, is also quantifiable.

Long Term Memory appears to have no limits in its capacity. All learned experiences, including the rules of languages, must be an important part of such memory. (Lindsay 1977).

The role of the memory system in the dynamics of language is currently thought to be a sort of inter-stage buffer in the linguistic chain. The serial heterarchy of the linguistic black boxes chain requires some sort of latching action in order to carry out the analysis - synthesis of the incoming stimuli due to the fact that this stage action is not instantaneous.

Goldman-Eisler (1972) incorporated into the store - boxes analysis some strategic planning action carried out during the interval between clauses of spontaneous speech. This is also a logical reason to assume the existence of memory capability across the linguistic-boxes chain. The limited processing capacity of each of these "black-boxes" requires a "buffer" memory, in order to carry out its particular analysis-synthesis.

The finding that a disruption of the material held in the "pre-categorical store" is achieved by recalling it with material acoustically similar to the stored item ( Craik, 1971) leads us to the study of the behaviour of this specific memory when processing Speech and Non-Speech material.

Crowder, (1971) reported neither suffix effects nor advantage effects when the material presented differed in consonants whereas both effects were present when the stimuli varied in a vowel. This effect is stronger when using speech-like material, Baddeley(1971). Day (1972) reported a strong effect of the speech nonspeech nature of

the stimuli upon the ear report in dichotic memory tests. This finding increases the suspicion that a distinction takes place somewhere in the aural perception system even before the material is stored in the precategorical memory.

## 3.6 The Speech/Non-Speech Decision Effects

Bailey et al (1977) reported disruption which may be interpreted as caused by a Speech/Non-Speech categorization. He attempted to determine whether the existence of some attribute of the auditory system predisposes the categorization of acoustic patterns into groups bearing a direct relationship with their phonetic label. To achieve this computer controlled synthesis of speech sounds. was used. The slope of the variations of the second formant was varied. to obtain phoneme categorization from the subject. These results were compared with the results of categorization induced for pure tone stimuli which followed the same frequency-time pattern as the speech stimuli. Subjects were induced to perform similar categorizations in both cases by suggesting to them that both sets of stimuli were speech sounds. These results differed remarkably when the subjects heard the pure tone stimuli as Non Speech material. They classified them as whistles.

The authors stated: "The pattern of the results appeared to relate not so much to the spectral structure of the stimuli, as to the way in which the stimuli were heard". Bailey et al, also concluded that speech is perceived when the acoustic elements in a sound stream can be interpreted by reference to an internalized representation of a possible structure of our speech production apparatus.

Bailey's first observation appears to correspond to a particular case of a general behaviour of our perception macrosystem.

Turvey (1976) points out a rather similar characteristic present in our perception of space. Our attention capabilities can be focused at a spatial location with greater efficiency when the modality of the stimuli at that spatial point is known. Preference is given to the knowledge of the messenger rather than knowing from

where the message is coming. A less equivocal example is given. When a letter "o" is embedded in a list of digits, it can be found more rapidly if the observer is told that he is looking for a letter. Conversely, if the sign "o" is a member of a list of letters, latency of search is considerably shortened if one is looking for a digit zero rather than a letter "o".

## 3.7 To Detect or not to Detect.

The experimental evidence gathered in the first sections of this chapter is not the result of experiments directly dealing with the problem of this thesis. Notwithstanding this, a few observations collateral to the experiment described, are valuable for our purpose.

There are effects which act in a negative manner upon the process under study i.e. disruption of the material held in the precategorical memory. Our viewpoint is that a disruption should imply the existence of a pre-categorical operation, the result of which bears some influence upon the memory functioning. That operation is in our model, the Speech/Non-Speech decision.

Other mechanisms could be playing a role in such disruption. Lack of capacity of the pre-categorical memory could be another feasible alternative to explain the disruption. One of the possible roads ahead of us is to try to explain every particular observation and reinforce our empirical basic assumption that there is an operation which sets our perception mechanisms in a "speech decoder mode". The other alternative is to observe the underlying trends of all the evidence presented in the foregoing, and from this formulate a working hypothesis: there is an operation of our aural perception mechanism which lies at the root of a linguistic decoding. It was therefore incorporated as another box in the speech chain mechanics, This conclusion is reached from:

(i) General consideration showing that a probabilistic analysis of perceptual processes cannot explain their selectivity.

(ii) Consideration regarding the components of the "speech chain" which all depend on the outcome of the Speech/Non-Speech decision.

(iii) The empirical results of shadowing experiments.

(iv) Empirical results of memory experiments.

(v) The results of perception experiments.

The introduction to this chapter incorporated two more (very important) dynamics, which might have some longer term effects on the building up of our perception processes. The ontogenetic and phylogenetic development of our speech abilities could add some extra evidence and clues of the existence and functioning of the Speech/Non-Speech operation.

## 3.8 Ontogeny of Speech Perception

The study of the processes used by man to acquire the ability of speech is another dimension to the study of such skills.

There is plenty of evidence that the skills of speech are acquired during a critical period in a man's life. If a child has had no contact with speaking people before the age of seven he will have the greatest difficulty in mastering a language later on. More evidence of this fact comes from studies of brain injuries. A child, the speech area of whose brain has been damaged, will recover his language capability fairly rapidly. The speed of this recovery is also a function of the amount of damaged area. The speech function can even be transferred to the non-dominant hemisphere of the brain. Woolbridge (1963). In contrast, the same damage to the speech area of an adult's brain will result, as a rule, in an irreparable loss of his capabilities. (Blakemore 1977).

The amazing processes which enable us to talk with our children have been observed by generations of parents and scientists. Lev Semenovich Vygotsky (1962) will lead us into this world. Vygotsky studied the development of the relationship of thought to language: From a pre-linguistic phase in the use of thought and a pre-intellectual phase in the use of speech our children begin discovering the symbolic function of speech.

For our purposes, the characterization of a child's thoughts are pre-linguistic and of the child's babbling as pre-intellectual speech, might well apply to the speech non-speech function as one stage of the different dynamics of the speech perception. Although Vygotsky does not specifically refer to the Speech/Nonspeech discrimination, he mentions this operation as present in one of the stages of the development of a child's speech. He distinguishes two characteristics of the development of Speech/Non-Speech discrimination. The mere reaction to a human voice as something different from noises is reported as appearing during the third week of life.

A social reaction to a human voice has been detected during the second month. Neonatal expressions such as laughter, inarticulate sounds, movements, etc., are described as indicating social contacts (ibid p.43).

Although the origin of the major linguistic features which develcp or emerge after the child has started his social life through sound emission and reception is something contentious, most authors hold the view that the acquisition of linguistic structures is rooted in the child's praxis. This praxis acts as a background and as an inseparable component of the first forms of a child's speech. (Luria 1974, Morehead 1974, Wadsworth 1971).

The observations of Luria, Piaget and Vygotsky reinforce the basic assumption of an early appearance of the Speech/Non-Speech discrimination function in the ontogenic development of speech.

Some other researchers also report discrimination between (speech) voices and other auditory sounds during the second week of life. Cutting and Eimas (1974) and Eimas et al (1971) report that month old children were capable of discriminating formant transitions and steady state vowel information in a manner rather similar to the adult's perception of this feature of the speech wave form.

From the foregoing review, we infer the following points:

(i) Speech Non-Speech discrimination is a mechanism which is formed since birth or even before it.

(ii) Ability to perform such discrimination should not lead us to the conclusion that those mechanisms are used in a linguistic-oriented analysis during the first stages of a child's speech development.

## 3.9 Phylogeny of Speech Perception

The study of the evolution of man's abilities to communicate in an oral manner presents the difficulty of the lack of material to work with. Communication (to share a message) in an oral fashion has however left some archaelogical clues. Traditions of stone tool manufacture as reflected in the rise of successive stone-tool industries, persisted for hundreds of thousands of years. Then came the great acceleration of language development of about 40,000 years ago (Washburn 1978).

There is consensus in ascribing to human language the same impact upon human development as upright walking. Tool use and tool manufacture were possible because hands were free. The mutual reinforcement of cognitive abilities and bipedal locomotion led to the foundation of the basis for the developing of our language (Lieberman 1970).

The theory of descent with modification through variation and natural selection put forward by Darwin opens another avenue to the study of the evolution of man's abilities to speak.

There is a limit to the application of this line of thought. The enormous span in time of the evolution of our abilities to speak and the short period of time spent in teaching a chimpanzee to speak are not comparable. Notwithstanding this, great efforts have been made in the laboratory. The case of Washoe, a chimpanzee who was taught sign language, is an interesting case. Cherry (1978) assessing Washoe's achievements in respect to the characteristics of human speech, comes to the conclusion that Washoe does not demonstrate the use of communication skill as used by the human race. Furthermore, if a parallel is sought between Ontogenetic and Phylogenetic development, this must be set against Vygotsky's point of view which is that thought and speech follow different types of development.

The study of the speech nonspeech discrimination function borders these two developments. The pragmatic "tuning" of the auditory system of the crickets when mating resembles the same effect reported by Chernigovskaya (1974) on human hearing tuned to the characteristics of human speech. Monkey's reactions to species vocalization have been measured and even specific groups of neurons identified as more sensitive to species vocalizations than other noises. Bull-frogs show the same specific responses to mating songs of their species companions. (Lieberman 1974). Revzin, (1979) assumes that human language derives from a reasonably well developed communication system, visualising the birth of language as a hominifying process. Species song identification is deeply rooted in any animal evolution. It's a matter of survival, association etc. Speech has been characterized as a "species-specific song" (Cutting, 1975) but the role of this function upon the decoding of it has been taken for granted and not studied as a necessary prerequisite.

The complexities of our "song" are correlated to the structure and development of the apparatus for its reproduction. In turn, there is also a correlation of the complexity of the hearing organs of an animal with its ability to produce sounds.

Archeological comparative studies also confirm this development of speech production organs in the ancestors of Homo Sapiens.

As a conclusion of this section we repeat Lieberman's words, " Man is human because he can say so....". The conversion of this simple and every day question into a super-problem again borders the philosopher's provinces.

Before mankind walked upright we assume they were living in gregarious social units. The mutual enhancement of perceptive capacity and sound production capabilities led to a more rapid transformation when our primitive ancestors liberated their hands. Tool manufacture can be compared with the planning of a phrase (grammar). One has to keep in mind only two things - The last tool that one made and the final form of the tool one is trying to make (Lieberman, 1974).

As well as developing the base mechanisms of grammar, a human was aurally aware of the presence of his/her tribe mates. Speech/Non-Speech discrimination is therefore of as much survival value as signalling for mankind as it is for frogs. This mechanism is also developed in an ontogenic perspective. The amazing performance of the aural perception of a contemporary new born child also locates the appearance of the Speech/Non-Speech discrimination function at a very early stage in a child's speech development.

Turvey (1976) states the loci of the function under study in a rather general and elegant manner.

"The set of constraints of <u>how</u> information is processed is by necessity linked with the intent of the perceiver as well as with <u>what</u> information exists in the surrounding medium",

Approaching the same problem from a different point of view, Cherry (1978) develops the speech chain in a rather distinctive manner: "Suppose you are sitting quietly, when a sound falls upon your ears. Such stimulus when perceived, is perceived <u>as</u> something. Once accepted as significant, it becomes a sign. Next, a second and vital decision is made in your brain: Was that sound uttered by another human being or not, for if it is accepted as somebody speaking you will seek to interpret its meaning".

The basic claim of this thesis is that the Speech/Non-Speech decision is crucial to the manner the incoming sound is "decoded". From this chapter we have learnt that this function of our speech perception system is deeply rooted in our development as mankind, as a child and also in the manner we start the apprehension of the aural reality.

The task is therefore, the measurement of such decisions with the obvious recognition that this is a function present in human beings.

CHAPTER IV

# WHAT AND HOW TO MEASURE

## 4.1 Introduction

The last two chapters were dedicated to the presentation of the speech production-perception system, its various models and their development. The need to study a more basic and rather obvious function present in this system is, however, the primary task of this thesis.

From the philosophers we borrowed Cherry's, Chomsky's and other scholars' points of view that there is a logical necessity in any speech perception model for an assessment of the significance of the incoming stimuli. This judgement must be made prior to the linguistic decoding of the speech stream.

In a different field, psychophysicists have produced a multitude of publications trying to identify the mechanisms of the auditory system specific to speech perception. There still is a very partisan polemic concerning the validity of these results. Notwithstanding this, a few observations relevant to our investigation were gathered. The categorisation of various speech-like and nonspeech stimuli is affected by the subjective assignment of the class of stimuli being heard. Different patterns of results were obtained if the subject believed that the stimuli to be heard was speech rather than pure tones or noise.

Neurophysiologists have identified areas of the brain dedicated to the decoding of speech stimuli. There are suggestions that the reticular formation of the brain stem is responsible for the "waking up" action of the cortex and some routing of the incoming stimuli. (French 1971). This is somewhat more complex than the simplistic telephone-exchange analogy of the functioning of the brain.

The study of the processes which lead to the formation of speech abilities in neonates shows that the Speech Non-Speech (mother-other) discrimination appears as early as the third month of life, before any other form of linguistic communication.

The search for clues to the origin of our aural communication skills has shown a close resemblance between the ontogenetic and phylogenetic development of our speech abilities. The rapid discrimination of the "human species song" was vital for our ancestors and would have been present in their behaviour well before the development of any form of language.

Having identified three distinct temporal dynamics in the problem under scrutiny, the following sections of this chapter will explore the availability of tools in the psychophysicist's arsenal in order to measure or elicit some characteristics of the Speech Non-Speech discrimination.

The determination of the set of acoustic cues which the perceiver uses in his/her decision as to the nature of the incoming stimuli is one avenue to explore. The measurement of the amount of information which must be acquired and processed in the performance of the Speech/Non-Speech discrimination and the comparison of these numbers with other relevant scales provide us with another working avenue.

The measurement of the amount of time that person takes to perform the Speech/Non-Speech discrimination is yet another measurement tool which uses the subject's inner mechanisms in order to characterize his/her responses to the stimuli.

Masking techniques have been used in the past to measure the amount of central involvement and the capacity of the attention "channel" in a given task. The amount of information that gets through has not been quantified.

The use of "virgin" material or neonates for work in the exploration of the acquisition of the Speech/Non-Speech skill is not feasible for an engineer with no special talents to work in a hospital environment.

## 4.2 The Approach

Reports on the work of psychophysicists appear to obscure perception by a series of measurements upon non-significant tasks. Pure tones, their harmonics, white noise, pink noise, bursts, flashing lights and so on structure the reality when a human guinea-pig enters a perception laboratory. Particular perceptual capacities exercised on natural stimuli will not be revealed by such experiments - the Speech/Non-Speech decision being one such specific perceptual capacity.

The latest addition to the psychophysicist's paraphenalia is the computer. Macro, mini or micro-computers, no longer the brain itself, are helping the experimenter to produce the stimuli, control and measure the subject's responses. Through the use of synthetic speech generated by electronic means, psychophysicists are disclosing the role played by formants, silences, etc., in the perception of phonemes. Yet there are still problems in the characterization of the relationship between physical characteristics of the speech stream and their perception. A given formant-time pattern is heard as a sound 'b' but a different array can also be heard as a 'b' in a different phonemic "environment".

The foregoing approach cannot be applicable to the study of our problem for the obvious reason that the use of artifical speech precludes the testing of a decision made upon real speech. The use of artificial speech presupposes the knowledge of the crucial dimensions or parameters used by the perceiver in its decision about the nature of a given incoming stimuli.

The investigations carried out using artificial speech assume the subject's acceptance of such stimuli as speech. If a subject does not accept a given distorted stimuli as speech, he is trained to do so. (Haggard 1977).

This example of a current approach to psychophysics exhibits the experimenter's dilemma. A given physical scale is arbitrarily produced (ie. slope of F2), the subject's percepts measured (different phonemes) and then this categorization of the scale $\frac{(df2)}{dt}$ is applied to different subjects to measure their percepts. Circular research is broken by the statistical application of results plus a constant feeding back of results to improve the original scale.

In our case, we would generate the psychophysical functions of the detectability of speech by manipulating some of its characteristics (bandwidth , signal to noise ratio etc.).

Westhoff (1963) summarizes various measurement techniques to obtain some means of comparison between different perceptual tasks:

(a) By measuring the amount of information which must be absorbed and processed in the performance of a task.

(b) The measurement of the time needed to absorb and process the information (reaction time) or the time needed to perform the task (performance time).

(c) The measurement of the extent to which the performance of the target task is reduced when another task is carried out simultaneously.

## 4.3 Measuring the Amount of Information

The amount of information is a concept which emerged at the juncture of two scientific trends. Mathematicians were applying quantitative descriptions to problems of signal transmission.

Kolmogoroff and Shannon formulated the statistical theory of communication which treated the message and noise as a statistical series (Gabor 1957).

Information from a source of N outcomes is expressed as:

$$I = \log_{10}(N) \text{ (Hartleys)}$$

$$I = \log_{2}(N) \text{ (bits)}$$

For a brief survey of the theoretical background of the communication theory see i.e. Cherry, (1978), Gabor, (1957).

In psychophysics, communication theory has its counterpart in Detection Theory. This is based upon a combination of decision theory and the theory of ideal observers. Decision theory recognises

that a priori probabilities, values and costs of incorrect decisions, as well as physical parameters of the signal, play a decisive role in establishing whether or not a subject reports hearing a signal. It is assumed, then, that a detection role is adopted comparing different scales which represent ratios of probabilities assigned to the members of the ensemble.

Cherry (1978) contradicts this approach by arguing that a person does not know the probabilities as a relative frequency expressed numerically but rather knows them subjectively as various forms of judgement. In a more general approach this point has bee.. discussed in Chapter 3, section 3.2.

## 4.4 Reaction Time Measurement

Westhoff (1963) sees reaction time measurement as another alternative way of measuring perceptual load.

Reaction time measurement uses the knowledge and experience of the subjects when performing a given task. It is an experimental measurement of the minimum time required for a recipient to respond, by some voluntary movement, to one of a number of alternative signals.

The concept of a finite time elapsing between a given perception and a human reaction began to emerge when astronomers of the 19th Century were faced with an intriguing problem. Measurements on the timing of the planets showed inexplicable discrepancies. The conception of the link between perception and action was placed under revision, the idea at that time being that such links were infinitesimal in time. ( Doesschate,1963)

In 1850 Von Helmholtz succeeded in measuring the speed of various perception processes. Donders suggested and demonstrated that it was possible, using Reaction Time Measurements, to determine approximately the duration of isolated mental processes, such as discrimination and selection.

Hick (1952) modelled some of the basic assumptions of information theory with the reaction time of subjects. The relationship found is expressed as:

$$RT = K \ln (n+1)$$

Where 'n' is the number of alternatives of a prearranged ensemble of events. In these experiments Hick used a number of lamps arranged in a small circle. His results were obtained using no significant stimuli, therefore limiting the application of such results to that psychological "domain".

Unlike Hick, Donders worked out a method to distinguish the various logical stages of a perceptual process. The measurements of a simple reaction time (one stimulus-one response) was compared with a disjunctive reaction time*.

Donders was differentiating the various stages of a perceptual process in which a decision was in response to an ensemble of stimuli. Firstly, it is necessary to discriminate between the sources or stimuli, then a choice is to be made between the signalling alternatives (push a key, etc.). Donders also assumed that these two different mental processes were additive in time.

Using these assumptions he concluded that the time that a human being takes to distinguish between two stimuli is approximately 46 msecs and the time to select a response is approximately 42 msecs. (Doesschate,1963)

Donder's assumption and methodology are still contentious. Wood (1975) separated his measurement of reaction time into two classes:

1. Pure motor reaction time which is related to muscular movements.

---

* a Disjunctive Reaction Time is the timing of a response to two different stimuli.

2. Pre-response time related to brain processing time.

To study the time dedicated to the classification of acoustic and phonemic-like sounds, Wood related the motor reaction time to the time point after which 99% of the button-press responses occurred. This point was the result of statistical analysis of evoked EEG with respect to time.

As an overall expression of a series of single mental processes an aural reaction time is formed by the following components:

1. Transmission time: the time taken to convert an acoustic stimuli into a neural representation.

2. Discrimination time: processing of the class, nature of the stimuli-source.

3. Choosing of an appropriate response following the outcome of 2, (above).

4. Motor reaction time.

According to a number of theorists as reported by M.W. Van der Molen et al (1979), at least three distinct stages are required in the analysis of the choice reaction time process, which may be labelled (a) stimuli encoding; (b) response choice and; (c) response initiation. This analysis of the reaction time dynamics correspond to the stages proposed above. The separation of latency time (transmission delay in the hearing mechanism) and discrimination time is only an idealisation of the model. Yet some "decoding" might be carried out while the neural version of an aural stimuli "travels" along the neural pathway towards the cortex.

There is some evidence for presuming this is so i.e. it is known to be so for visual pathways.

The treatment of the various components of the R.T. is as numerous as the number of scholars using R.T. techniques. The aims of the experiments reported in the preceding sections of this chapter are different. Hick did not need to separate the various mechanisms and found a mathematical expression which has applications in a given perceptual mode, a non-significant environment. Wood used a pragmatic criterion and measured the evoked potentials of the first stages of the reaction process to an aural stimuli. Phoneme discrimination tasks produced reaction times which were employed in differentiating between phonemes and acoustic processes in the speech decoding mechanism, as well as in measuring brain hemisphere specialization. (Rubin 1975; Fry 1974, 1979).

## 4.5 Reaction Time and The Speech/Non-Speech Discrimination.

From the foregoing overview of possible techniques to measure the characteristics of the Speech/Non-Speech discrimination, Reaction Time techniques offer the best alternative. Its implementation is easily achievable in an electronic engineer's laboratory and uses the inner mechanisms of the subjects without assuming a Bayesian model for the speech perception. The choice of Reaction Time measurement techniques should provide some estimate of the performance of the human brain when dealing with such discrimination, and provide us with some overall characteristics of the acoustic clues of the stimuli which trigger the decision on the nature of the incoming stimuli.

### 4.5.1 The Problems

Reaction time measurements are known to be affected by some characteristics of the incoming stimuli. (Fry 1975). Intensity and duration of the stimuli do affect the Reaction Time in a simple reaction time experiment. This seems to be a general pattern of our perception abilities. Sperling, 1963, reported that the number of letters correctly reported after variable duration of exposure falls dramatically when the letters were exposed less than 100 msecs. (Fig. 4.1) (from Fitts, 1973).

Another characteristic of the detection of speech as a class of stimuli is its seeming independence of linguistic (phonetic) features. From Chapter I the radio tuning example can be recalled. The listener suddenly realises that, in the midst of a barrage of noise, there is somebody speaking. This realisation came about in the middle of a word, despite the language of the broadcaster. The design of a choice reaction time experiment which tries to simulate these conditions cannot define the onset of the stimuli with the precision that is easily achieved in an experiment which uses artificially generated speech.

A random selection of speech samples will balance the onset of the stimuli across all the sounds of speech. Vowels, fricatives, plosives, etc. will constitute the stimuli onset.

Figure 4.2 illustrates the difficulties in defining the intensity and frequency characteristics of the stimuli onset.

In experimental work on pitch discrimination, most authors indicate the intensity of the stimuli as an average level expressed in dBs. Fry (1975) normalised for his measurements the peak amplitude(instantaneous) as a more realistic parameter for use in the characterisation of the stimuli in experiments using short samples of speech.

Pitch discrimination is affected by the intensity of the aural stimuli as reported by Van der Molen et al (1979). Their conclusion is that intensity seems to affect the timing of the choice of response. The discrimination function does not appear to be affected.

### 4.5.2 Temporal Uncertainty

Another factor which has an effect upon the reaction time to an aural stimulus . . is the uncertainty of its onset. Simple reaction time may approach zero if it is possible for an individual to predict the onset of an event. A batsman in cricket who knows the speed of the bowler can time his swing almost perfectly. A forced reaction time can also be disturbed if the subject is given the

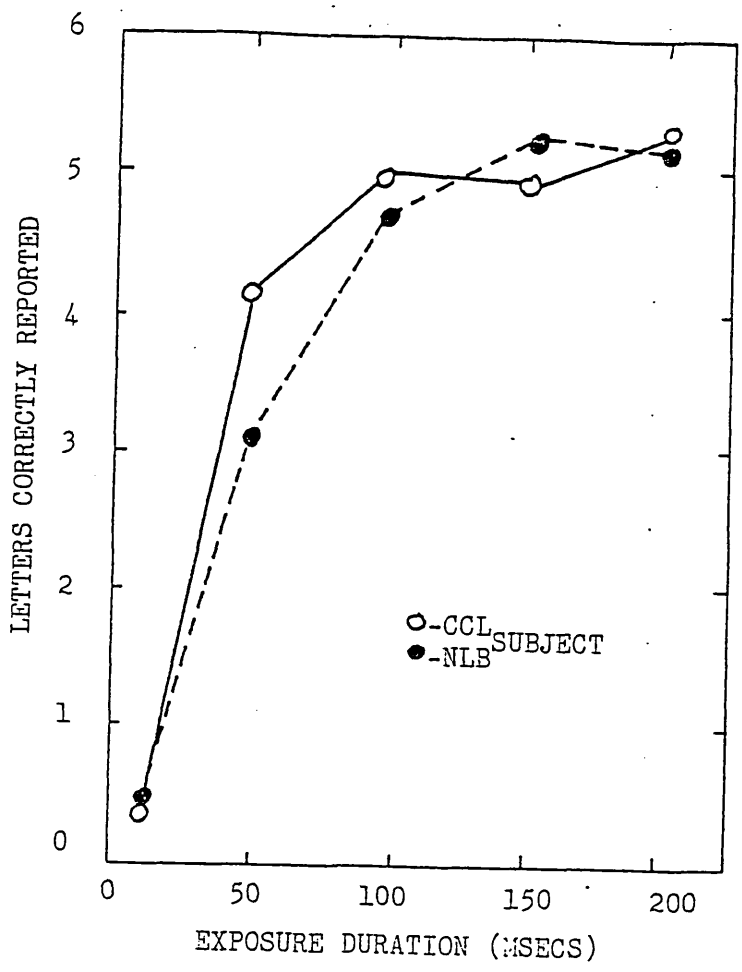



Fig 4-1. Letters correctly reported as a function of the exposure duration

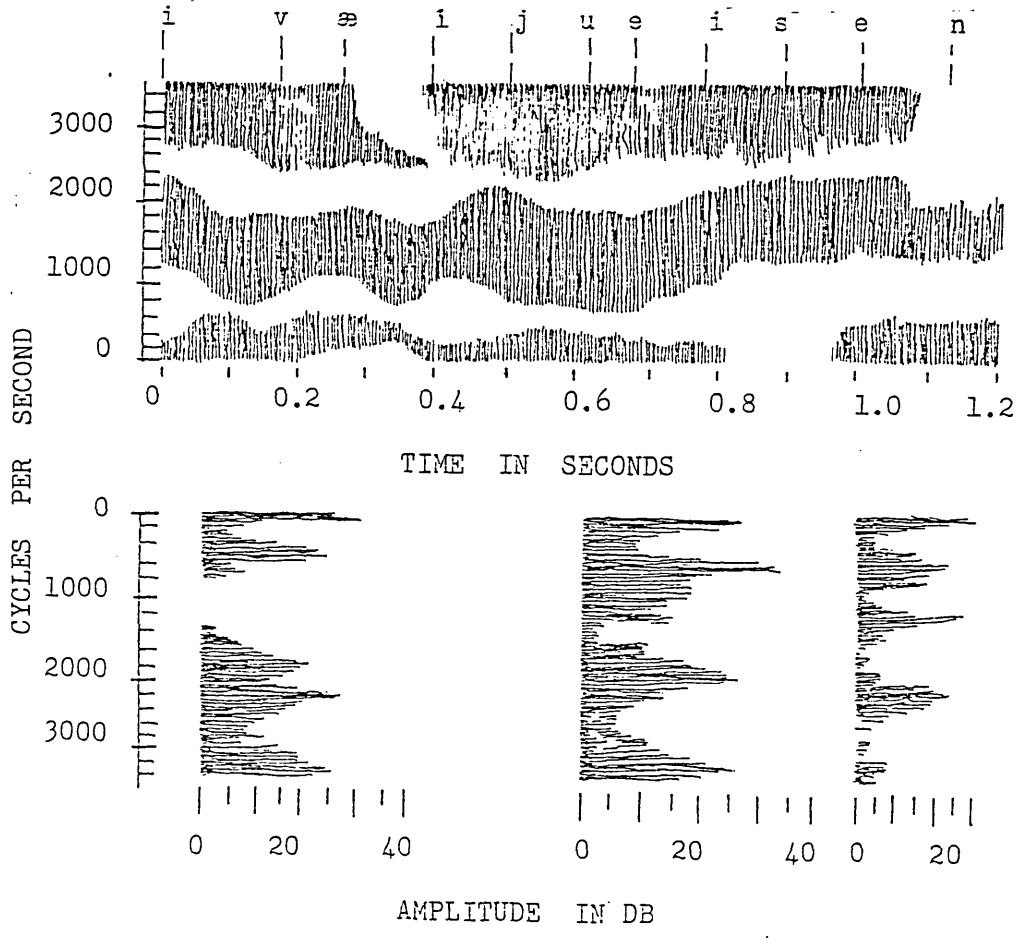



Fig. 4-2. Formants and amplitude-time characteristics of some speech sounds.

stimuli at random intervals. The efficiency with which a person can prepare to receieve a signal varies with his certainty about the time of its arrival.

The use of temporal uncertainty as a parameter related to reaction time experiments was studied by Klemmer . He varied the foreperiod of the stimuli and also the predictability of it. An increase in reaction time with the amount of variation of the foreperiod was found. (Fitts 1973).

4.5.3. Number of Alternatives

The variation of the number of alternatives and its effect upon the reaction time was studied by Merkel in 1885, who extended Donder's data on choice reaction time experiments. He found a logarithmic increase in the reaction time as the number of stimuli and responses increased. Hick (1952) replicated some of Merkel's experiments and extracted a relationship between the number of alternatives and the reaction time. Fig. 4.3.

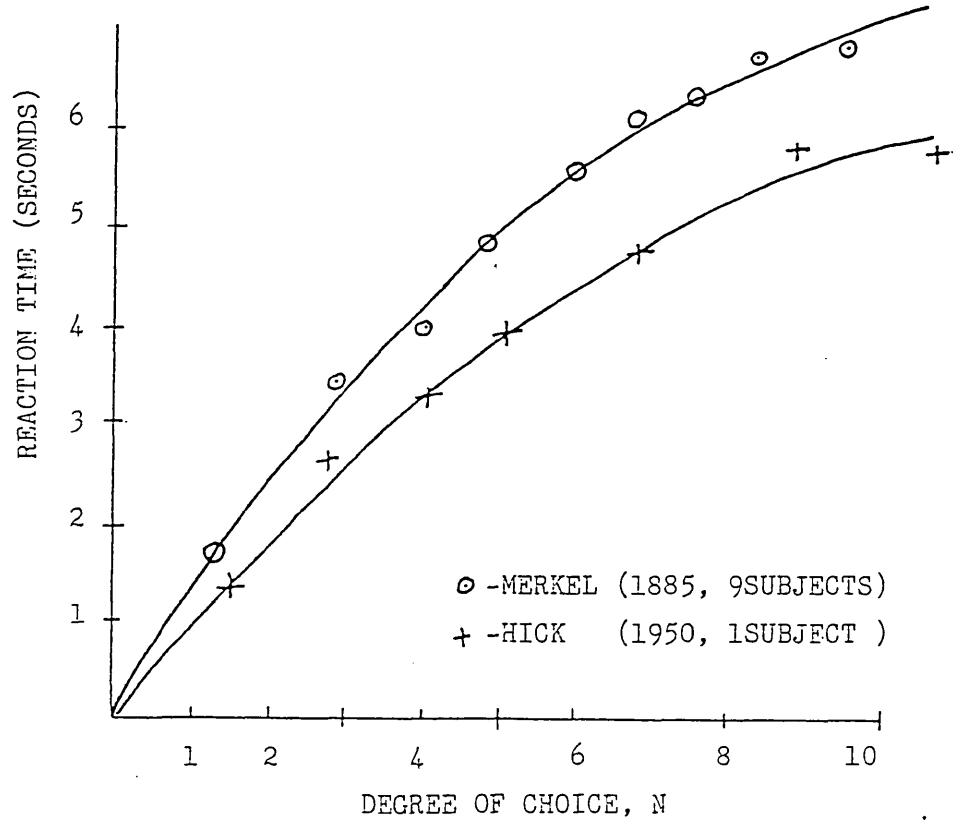



Fig.4-3. Reaction time as a function of the degree of choice.

Hick's and Merkel's results have been extended to cover more response modes and stimuli. These results are applicable to the reaction to stimuli which have no processing other than their detection. They are of interest because they show the mechanisms which affect reaction time.

Fig. 4.4 from Fitts (1973).

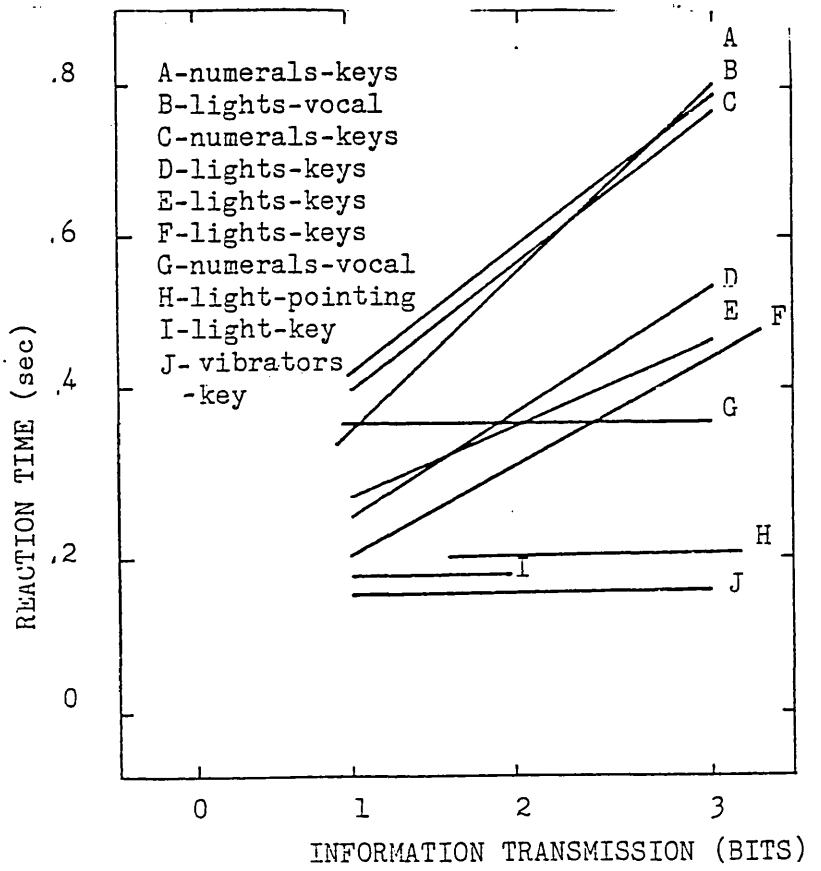



Fig. 4-4. Reaction Time as a function of information transmission of experiments of varying compability.

4.5.4. Hearing Mode

It has been recognised since the time of Broca, that one of the hemispheres of the human brain is specialised for language and speech functions. Reaction time experiments have shown insignificant right ear advantage (REA) for key pressing response to natural utterances of word pairs when the stimuli was relayed monoaurally. (Fry, 1975).

The explanations of the origin of REA are still contentious, yet most authors link REA with speech-like stimuli being decoded in the left hemisphere.

## 4.6 Our Measurement Tool

The choice of reaction time techniques for investigating the Speech/Non-Speech decision carries some difficulties which have been underlined in the preceding sections of this chapter. The use of natural speech in reaction time experiments adds some extra complications which can be overcome by use of statistical techniques in the analysis of the resulting data. Also an experimental paradigm should be chosen to try to elicit some characteristics of the Speech/Non-Speech discrimination.

The comparison of reaction times and accuracy of key pressing response to natural speech, music and noise, with the reaction times achieved, to a three alternative experiment, such as the ones carried out by Hick and Merkel, provides the first approximation to the magnitude of the reaction time, especially its motor response. The results which a three forced choice reaction time experiment could yield are expected to be different, given the "significance" of the stimuli to the subjects, and the amount of processing suspected to be carried out.

The use of the subjects' capacity to predict the onset of the incoming stimuli can also run into difficulties when using random samples of speech. The comparison of the results obtained using a fixed and a random interstimuli interval should elicit this effect, if there is one.

The existence of REA for speech signals conveyed monoaurally when subjects are asked to perform speech related tasks, has led to the hypothesis of the existence of specialised processors dealing with certain aspects of the speech waveform in the left hemisphere of the human brain. The heuristic suspicion that the Speech/Non-Speech decision is carried out before this stage is reached, could present

problems for the current explanation of REA. This is based on the difference in neurological paths that stimuli conveyed to different ears have to follow before reaching the speech associated areas of the cortex.

The extreme efficiency and accuracy required to carry out the Speech/Non-Speech decision, its speed and automatic characteristics, are prima facie grounds to suspect the non-existence of REA when carrying out this decision.

The preceding sections of this chapter have studied the various alternative approaches to study the Speech/Non-Speech decision carried out by humans. The measurement of reaction time appears to be the most advantageous measurement, it is easily implemented in an electronic laboratory, uses the subject's capacities without assuming a too narrow serial signal processing approach, and does not use the subject's memory, therefore avoiding any signal linked disruptive effect.

The tentative approach to the experimental stage of this work is presented in the next chapter, together with the description of three preliminary experiments carried out in order to gain experience in dealing with human experimental material, and to assess the magnitude of the variables and parameters of the main set of experiments.

CHAPTER V

# EXPERIMENTAL APPROACH TO SPEECH/ NON-SPEECH DISCRIMINATION

This chapter serves two purposes. The first is to briefly describe experiments carried out to obtain experience in reaction time experiments and to assess the magnitude of the variables under study. This will lead to the description of the main set of experiments which will study the reaction time and accuracy of subjects' responses to the discrimination between speech, music and noise.

## 5.1 Preliminary Experiments

A detection test was conducted to study the effect of the speech bandwidth upon its detectability. At the same time the interstimuli interval was varied.

The sample material was natural speech as a more secure alternative to electronically generated speech. The samples of speech and noise were edited in a random sequence and were separated using different lengths of non magnetic tape.

The speech samples were recorded and then edited using the following criteria:

- Speech has to be audible during the initial 100 msecs. Pauses of conversational speech were avoided.

- No attempt to keep the meaning of the samples was made.

- Brief exclamation comments such as ahh...ehh.. etc. were excluded.

- Interstimuli separation was varied between 1 and 4 seconds in order to estimate the best compromise point between subjects' ability to predict and the recording of his/her responses.

Six British born male subjects, all right-handed and with no hearing difficulties; were asked to discriminate between pink noise and speech. The signals were recorded following a random sequence and the speech signal was filtered using an Pekel active filter type TF823. This filter has 8 preset frequencies with independent control of attenuation. The slope of the attenuation could be varied up to 60 db per octave.

Each subject made 32 responses to a set of 16 samples of speech and 16 samples of noise per bandwidth. The stimuli were recorded from a standard BBC transmission, male voice, using a Revox 77 recorder. This apparatus was then used to relay the signals to the subjects.

Subjects were seated in a soundproof room, under comfortable lighting and temperature conditions and monitored from the experimenter's room through a window. Fig. 5-1.

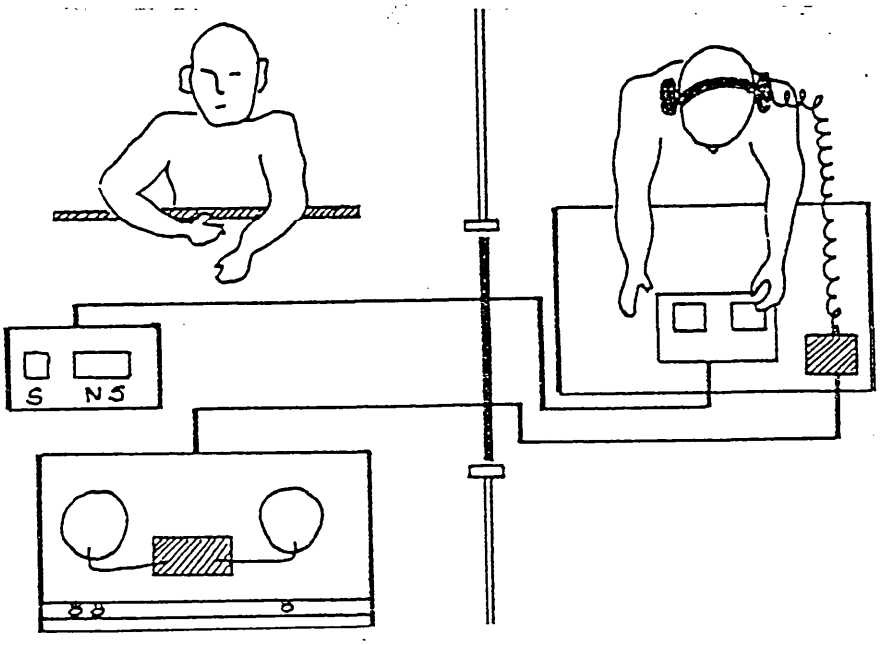
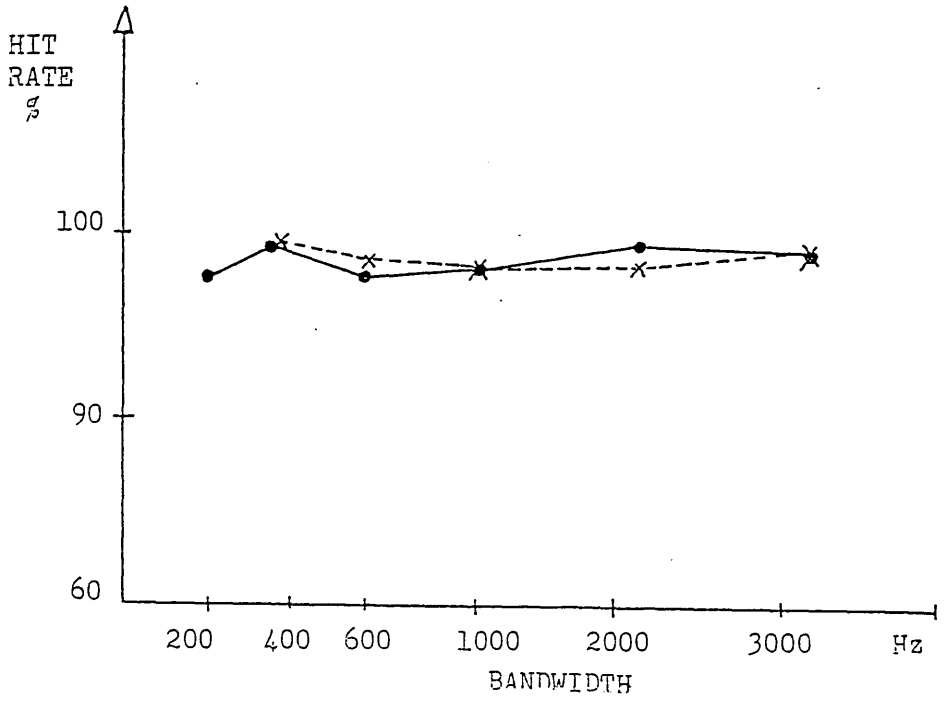



Fig. 5-1. Experimental set-up for the study of speech detectability as a function of its bandwidth.

The hit rate was virtually unchanged across the different bandwidths. The minimum frequency was 180 Hz and the maximum 4000 Hz. The interstimuli interval found to best suit the subjects' and experimenter's purposes was 2 secs. See table 5.1 and graph 5.2.

| BANDWIDTH (Hz) | 180 | | 250 | | 500 | | 1000 | | 2000 | | 3000 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| STIMULI | S | N | S | N | S | N | S | N | S | N | S | N |
| ERROR | | | | | | | | | | | | |
| SUBJECT | | | | | | | | | | | | |
| M.P. | 2 | 2 | 1 | 1 | 2 | 1 | 3 | 2 | 2 | 1 | 1 | 1 |
| M.Y. | 2 | 1 | 0 | 1 | 1 | 1 | 1 | 2 | 1 | 3 | 0 | 1 |
| J.D. | 3 | 2 | 2 | 1 | 1 | 2 | 2 | 3 | 2 | 2 | 1 | 0 |
| A.K. | 3 | 3 | 1 | 2 | 3 | 2 | 0 | 1 | 2 | 1 | 2 | 2 |
| C.C. | 1 | 0 | 2 | 3 | 1 | 1 | 3 | 2 | 2 | 1 | 1 | 0 |
| $\Sigma \epsilon$ | 13 | 11 | 6 | 9 | 10 | 9 | 10 | 12 | 9 | 10 | 7 | 7 |
| $\hat{\epsilon}$ | 2.7 | 1.8 | 1.0 | 1.5 | 1.7 | 1.5 | 1.7 | 2.0 | 1.5 | 1.7 | 1.1 | 1.1 |

TABLE 5-1.




GRAPH 5-1

## 5.2 Estimation of the Motor Response

From Chapter 4 it can be recollected that the longer component of a choice reaction time is its motor response. Hick concluded that the reaction time is a logarithmic function of the number of alternatives presented to the subjects. This applies if the subjective "weight" is equal from all the alternatives. In order to estimate the motor component of a three forced choice reaction time experiment and in order to assess the effect of the relative position of the keys, a replication of Hick's experiment was carried out.

A keyboard which has three keys, separated by 1 inch and with green LED indicators was positioned in front of the subjects. The room used was the same as in the previous experiment, under dimmer lighting conditions. Fig. 5-2.

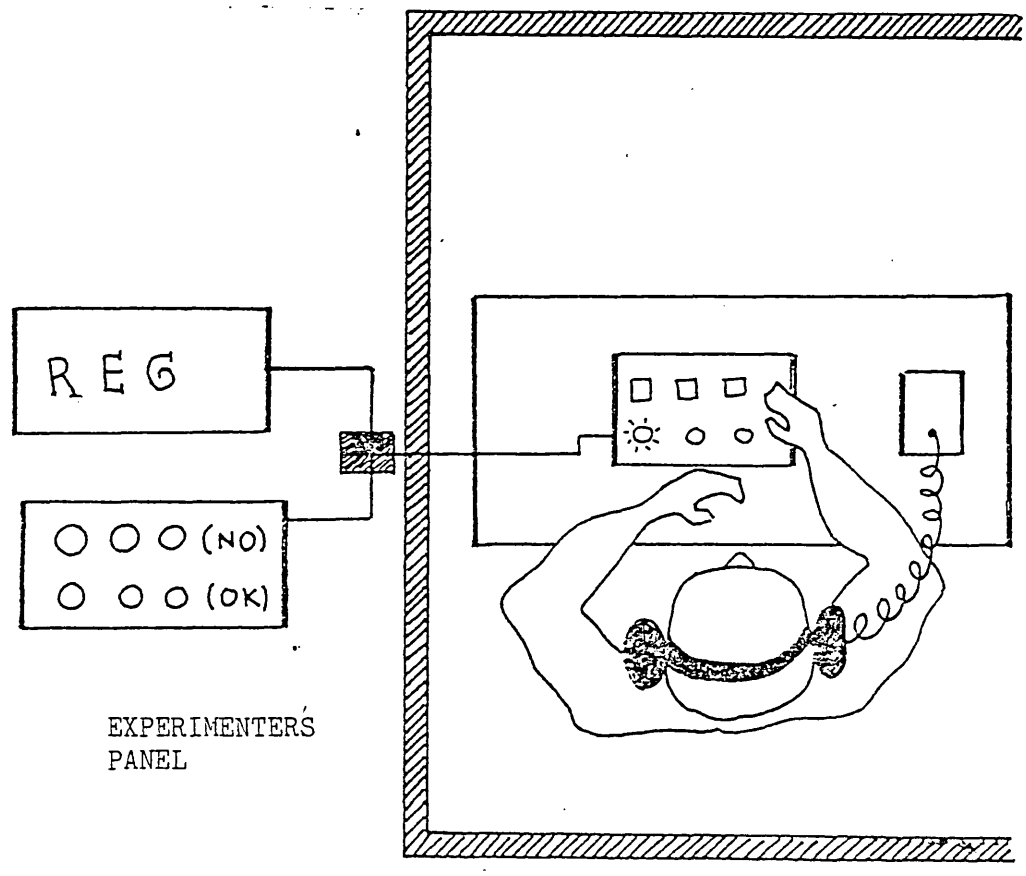



Fig.5-2. Three forced choice reaction time experimental configuration.

The indicators were driven from the random event generator which also produced the waveforms to drive the gated oscillator and the timer. The random event generator is an "electronic dice" which is basically a white noise source, a small amplifier, and counters. The state of the counters at the end of an arbitrary period has the same probability distribution as the white noise in frequency.

| Subject | RT (msecs.) | $\sigma$ | e % |
|---|---|---|---|
| M.P. | 302 | 49 | 8 |
| M.Y. | 375 | 86 | 5 |
| J.O. | 368 | 49 | 4 |
| A.K. | 326 | 67 | 7 |
| D.T. | 306 | 38 | 8 |

The statistics of the above table are in agreement with the figures given by Hick (1952). The average reaction time among the 5 subjects is 349 msecs. Figure 5.3 shows the distribution for the responses of one subject.

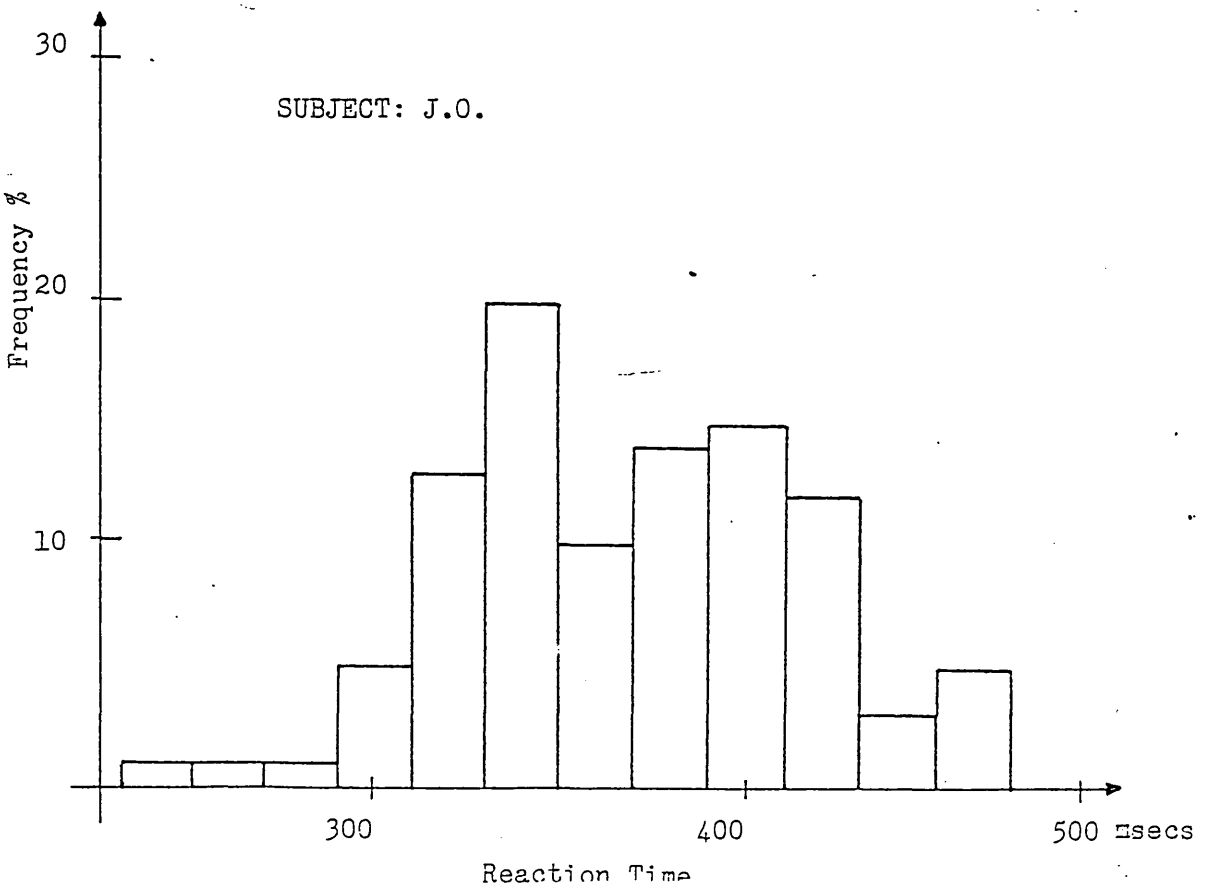



Fig. 5-3. Frequency distribution of the reaction times of one subject.

Each subject made 120 responses to the stimuli. The stimuli were relayed with a 2 secs interval and a duration of 0.5 secs. The first 20 responses were eliminated from the statistical analysis, avoiding the introduction of the learning process of the subjects into the statistics. The keyboard position (vertical or horizontal) did not affect the results. The distribution of the single subject responses shown in Fig. 5.4 presents a fairly good approximation to the normal distribution.

## 5.3 Preliminary Conclusions

The results of the detection test which studied the detectability of speech with respect to its bandwidth, yielded an estimate of the best interstimuli interval and provided the experimenter with some clues with respect to the length of the training processes. The replication of a particular case of Hick's experiment yielded results which are comparable to his data based on one subject and Merkel's results which are an average over eight subjects.

The experiments described in the preceding sections justify the following conclusions:

- A "telephone" quality of speech will not effect its detectability.

- Estimating that the difference between reaction times to aural and visual stimuli to be of 20% (Fitts 1973), the motor component of a three forced choice reaction time varies around 350 msecs.

- An interstimuli interval of 2 seconds allows the observer an easy recording of the subjects' responses. At the same time this period of time is short enough to enable the subjects to use their capacity to predict the onset of the incoming stimuli without producing anxiety. Woodrow (1951) pointed out that the greatest accuracy in the discrimination and reproduction of empty intervals could be used to maximise the subjects' attention and therefore speeding their responses.

## 5.4 Experimental Insights

In order to assess and characterize some of the mechanisms used by the human organism when discriminating speech from non-speech, three sets of experiments were conducted. The results of these experiments will enable the author to estimate the role of Speech Non-Speech discrimination in relation to current models of speech perception. The aim of the following sections is to describe these experiments in a brief manner.

## 5.5. Experiments 1 and 2

1 & 2 The study of the psychological functions of Reaction Time and Accuracy vs Duration of the stimuli.

Speech/Non-Speech discrimination is not an instantaneous process. A basic assumption is that the reaction time is a measure of the psychological "load" or amount of processing being carried out by our speech or auditory system. This processing time, which is one component of the overall reaction time, will be assumed to be a function of the amount of information that the aural perception system receives.

The transmission time and the motor reaction time are postulated to remain constant during experimental manipulation of the amount of information conveyed to the subjects. The easiest way to vary the amount of information being relayed to the subjects is to shorten the duration of the stimuli until the clues conveyed by the speech waveform as intensity, frequency variations and timing patterns are insufficient to reach an accurate and speedy decision.

The stimuli set comprises samples of BBC English, Solo cello music and pink noise. The music stimulus is selected because it is an interesting alternative to natural speech and the control stimulus, pink noise. The spectral variation of the fundamental notes of the cello lies between 65 and 660 Hz. Music is not as significant a sign as speech for a human listener, yet it can convey meaning.

Feelings and images evoked by music might be thought to be decoded in the same manner as the meaning of a phrase. The stimulus set was divided in six blocks of different durations. Within each block equal numbers of samples of every class of stimuli were edited in a random sequence.

The speech samples were edited following the same criteria as in the experiment described in section 5.1 and band-passed between 80 and 3000 Hz to equalize the frequency characteristic of the three stimuli.

The first experimental attempt yielded no results which could be considered significant. The capacity of the subjects to detect very short samples of speech, music and noise was underestimated. Speech samples of 80 msecs. were nearly always detected as speech...! The stimuli were then shortened to the range of 0.030 to 0.5 seconds to investigate a sharp variation of the reaction time and accuracy when the stimuli duration was less than 100 msecs.
The experimental layout is depicted in Fig. 5-4.

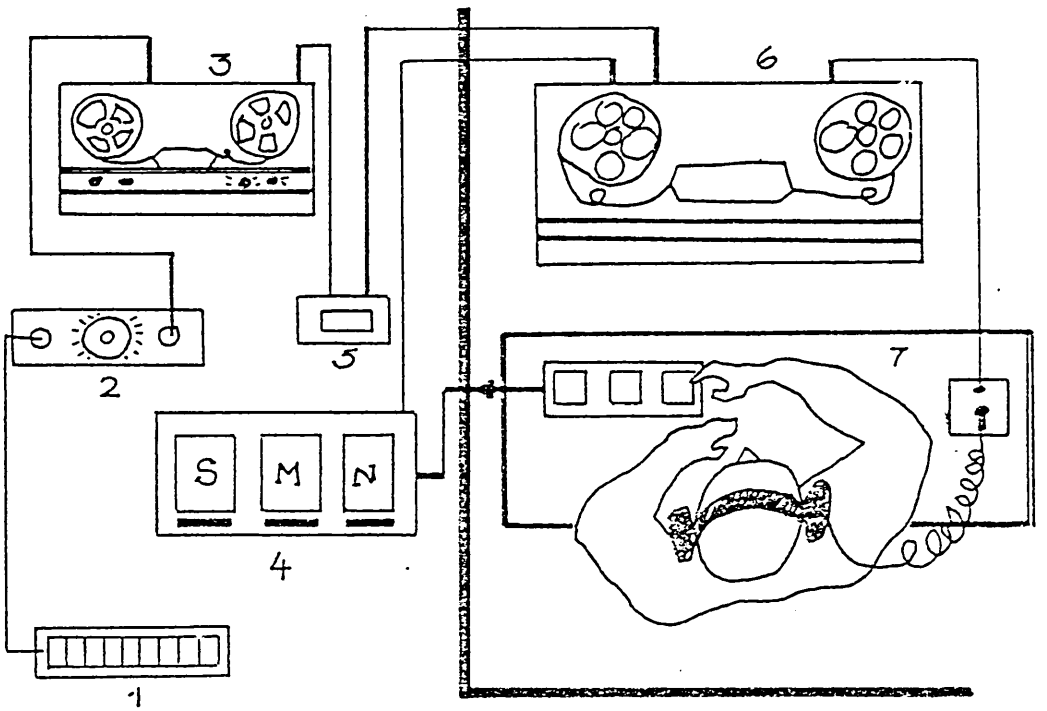



Fig 5-4. Experimental set-up for the study of the reaction time and accuracy, as a function of the stimuli class and its duration.

The equipment numbers in Fig. 5-4 are as follows:

1 - Electronic Timer/Counter
2 - Gated Oscillator
3 - Revox Professional 77
4 - Observer's display
5 - Remote On/Off switch
6 - Revox Professional Hs 77
7 - Subject's Keyboard

The stimuli are relayed to subjects via headphones (Koss, Pro.44.A). As soon as a key is depressed, the corresponding light will indicate the response in the observer's display. This in turn generates the signal interrupting the counter and high frequency tone which were set by the onset of the stimuli. In this manner the responses of the subjects are recorded for accuracy and delay. The copy of the tape which contains the tone is used to re-check the times written down by the experimenter.

A more comprehensive explanation of the experimental procedures is given in Chapter 6. The replication of the first experiment this time using shorter stimuli (0.03 to 0.5 seconds) will be referred to as experiment 2.

## 3.6 Experiment 3

3 - Study of the Reaction Time and Accuracy of Speech/Non-Speech discrimination under different attention conditions.

This experiment is complementary to experiments 1 and 2. The subjects' attention or readinesss was perturbed by varying the interstimulus interval from a constant 2 seconds to a range of random intervals between 1 and 4 seconds. The stimuli, experimental set up and subjects are the same as in experiment 2, although a period of two months was allowed between experiments. A variation in the reaction time and accuracy is expected under these conditions in respect to the results obtained in experiments 1 and 2. This result could provide extra clues about the position, in a hierarchical model, of Speech/Non-Speech discrimination.

## 5.7 Experiment 4

4 - Hemispheric Brain Specialisation and aural Speech/Non-Speech Discrimination.

From the description of the speech perception system given in Chapter 2 and the psychophysical experiments detailed in Chapter 3, it is clear that the "hardware" of the speech perception system is mainly located in the left hemisphere of the human brain. Since Broca correlated brain injuries with speech difficulties, more than 100 years have elapsed, and the quantitative effects of this characteristic of our brain are just beginning to be elicited.

This specialisation has come to the phoneticians attention resulting in a series of experiments which demonstrate an advantage of the right ear (REA) when processing certain speech sounds. Stop consonants (p,t,k) are perceived with a clear advantage by the right ear over the left, (Cutting 1978).

In a typical experimental paradigm, a pair of phonemes are relayed to the subjects in "competition" to both ears. Most listeners report the item presented to the right ear more easily than the one conveyed to the left. D.B. Fry (1974), also detected REA for speech presented monoaurally, i.e. one ear stimulated at a time.

REA is an effect not only related to speech. Cutting (1978) points out a few other classes of stimuli where REA is also present. Music stimuli, relayed as plucked and bowed sounds also elicit REA (Blechner 1976). Morse code signals presented to naive subjects also elicited REA (Papcum 1974).

The works just mentioned and their results are important considerations in the formulation of any speech perception model. The discussion is centred on the isolation of mechanisms of our auditory apparatus that are specific to speech. Phonological decoding follows, in our working framework, the realisation that the stimulus being heard is speech.

The search for REA in the Speech/Non-Speech decision might provide additional clues for the specification of the boundaries of a model for speech discrimination.

The first version of this experiment was arranged in such a way that every cell, subject-ear-stimuli-duration, was replicated the same number of times as in experiments 2 and 3.

The stimuli were relayed to each ear in a random manner. Subjects were asked to depress the key which corresponded to the class of stimuli being heard as fast as they could. Some difficulties were encountered when using this modality. Subjects tended to depress the key which corresponded to the ear being stimulated. This very effect was reported by Hammond and Barber (1978), in a study of three forced choice reaction time tasks.

This effect was overcome by rearrangement of the order of the stimuli to the left and right ear and by reversing the spatial arrangement of the keyboard every session with each subject.

The experimental layout is similar to the experiment 2 and 3. This time the tape was edited using both channels of the tape player.

# CHAPTER VI

## Procedures, Data Processing and Concluding Remarks

### 6.1 Introduction

Chapter I gives rise to the heuristic suspicion that the Speech Non-Speech discrimination is an operation of our auditory apparatus which presents characteristics explored and formalized in Chapter III. The framework and history in the literature which refers to this problem are described in Chapters II and III. The Speech/Non-Speech (S/NS) decision is not a "linguistic" operation yet its outcome is crucial to the operation of linguistic related operations i.e. phonological decoding.

The Speech/Non-Speech decision seems to be a peripheral function. Its outcome is not disturbed by central load and can be stored and recalled, as proved by shadowing experiments. See Chapter III.

The first measurement of a phenomenon, such as the S/NS discrimination, which can easily be done is the measurement of the speed and accuracy with which it is carried out under optimal conditions. The hypothesis here is that the S/NS decision is carried out faster than the duration of an average syllable. (250 msecs). Direct comparison with music and artificial noise stimuli should not yield significant differences. Discrimination between different classes of stimuli should be equally fast for all the stimuli, as a logical requirement of this operation.

The role of the attention condition of the subjects when carrying out this discrimination could provide some more clues in the characterization of the Speech/Non-Speech operation.

The attention condition can be seen as a "catalyst" for perceptual processes. It is clearly more than a go/no-go valve as demonstrated by shadowing experiments. In these experiments

the Speech/Non-Speech discrimination is carried out regardless of the amount of central load imposed on the subjects. At the same time its outcome is stored in memory and can be recalled with no problems which could be associated with the fragility of the pre-categorical memory.

A random variation of the interstimuli interval should prolong the subjects' responses and reduce their accuracy if the attention "pointer" has some effect upon the Speech/Non-Speech discrimination. A replication of the first experiment, carried out this time using random foreperiod, and using the same subjects as in the first experiment should bring this effect to light when comparing the subjects' responses.

Right ear advantage in the decoding of some consonants during the recognition of word pairs (i.e. /splei: Sprei/, Fry, 1974) and the fact that REA has been linked with the peculiarities of the speech waveform (Haggard, 1977) have prompted the author to try to elicit REA when subjects are asked to carry out the Speech/Non-Speech decision.

The hypothesis for this experiment is that REA should not be present in the Speech/Non-Speech discrimination. In a sense this type of discrimination has the same category as the attention function of our brain activities. It is also playing a sort of pointer role. It should not therefore be associated with hemispheric specialization.

The search for REA in the Speech/Non-Speech discrimination might provide additional clues for the specification of the psychological boundaries of this operation.

## 6.2 Experimental Method

### 6.2.1. Task Characteristics

This section provides the details of the methods actually used in each of the experiments described in the previous Chapter

and in the introduction of this Chapter. The various components of the experimental situation are described in detail to allow replication. The task is an approximation under controllable conditions of the examples described in Chapter I. The experiments adopt a three forced choice form in which the subjects are asked to respond as quickly as possible when the incoming stimulus has been classified as speech, music or noise. They respond via a keyboard which signals the subject's response to the observer. The stimulus information is varied by shortening its duration. The stimuli are relayed to the subjects in blocks of 6 different durations. Each block contains 12 samples of speech, music and noise stimuli making a total of 36 samples per block. The samples are distributed at random and separated by a constant length of non-magnetic tape in the experiments 1, 2 and 4. Experiment 3 uses random lengths of tape which produce an inter-stimuli interval varying between 1 and 4 seconds.

Subjects were asked to make their responses as soon as they were reasonably certain of the class of stimuli being heard without waiting for the end of the incoming stimuli. Speed and accuracy was encouraged by rewarding the best subject scoring $[(1/R.T.) \times 100 \times Acc.\ (\%)]$ , where RT. is the reaction time and Acc., the accuracy of the subject's responses. A period of two months was allowed between experiments 2 and 3 to avoid possible training effects. Each subject was allowed a rehearsal-training period of approximately 15 minutes. At this stage a practice block was relayed to accustom the subjects to the keyboard and allow the adjustment of the volume of sound in their headphones. The practice block contained 36 samples; 6 different durations and 2 classes of stimuli per duration.

The experimental set up is common to all the experiments and is depicted in Figure 5.4.

Three forced choice experiment combined with direct labelling and the use of the limit of stimuli duration are the main features of this experimental stage. Each subject made 216 responses making 1728 responses in total for experiments 1, 2 and 3. Experiment 4 doubles the figure.

### 6.2.2. Stimuli Factors

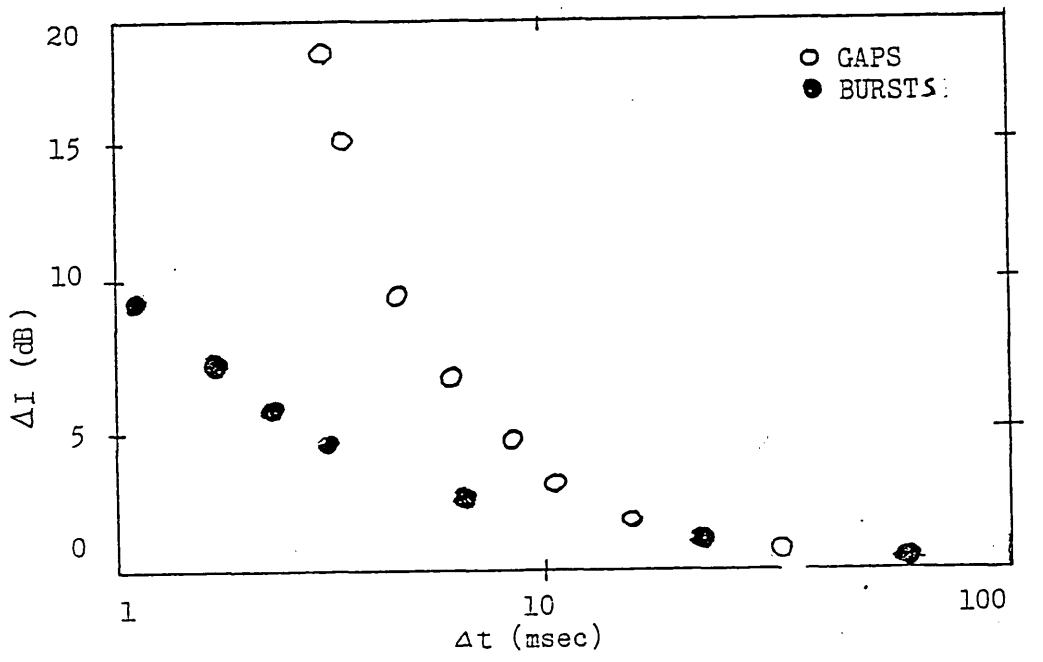
The difficulties of using natural speech in reaction time experiments have been outlined in Chapter IV. The editing techniques aim to achieve a random sampling and at the same time ensure that the speech is audible in the first 100 msecs. This was checked by ear and storage scope measurements with the tape running at half of its normal speed (17 cm/sec). To avoid the "click" sound of the transition of non-magnetic tape into the magnetic part, the tapes were spliced with a slope which produced a smooth onset ramp of 5 msecs.

The speech sample was taken from a standard BBC transmission, Radio 3, F.M., male voice. The music stimuli was taken from a Casals interpretation of a solo cello piece. The noise stimulus is white noise from a standard laboratory generator. All the stimuli were band filtered between 80 and 3000 Hz to equalize their long term frequency characteristics.

The stimulus duration was the variable in all the experiments. The minimum detectable duration of a stimulus has been investigated by Irwin and Purdy (1982). Their results are depicted in Fig. 6.1. The detection of bursts and gaps by the auditory system seems to be finely tuned to the characteristics of phonological features. For our purposes, it is enough to notice that bursts and gaps of 20 msecs and signal noise ratio of 3 dB are detected with a 75% accuracy. This figure might improve when gaps and bursts are part of a spoken syllable. For the first experiment the minimum duration was 80 msecs. The guessing level in accuracy was not reached in this experiment forcing the stimuli duration lower limit to change to 30 msecs for the subsequent experiments.

Shadowing experiments have shown that some people can reproduce speech only lagging by 250 msecs. The simple classification of the stimuli being heard should be achieved in a shorter period.

Fig. 6-1:
'The signal to noise ratio as a function of the minimum detectable duration for bursts and gaps.




The detection of simple and complex sounds using sinusoids and mixture of tones of different durations have been studied by D.M. Green (1958). The results presented are in conflict with the figures given above and the trend of the figures given by Repp (1982).

The averages in table 6.1 have been taken from Green's (1958) graphs.

TABLE 6.1

| Stim. Durat. (msecs) | St1 | St2 | St3 | St4 | St5 | St6 | St7 | St8 | St9 | St10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 50 | 69.3 | 68.3 | 66.3 | 69 | 78.7 | 83 | 83 | 80.3 | 87 | 85 |
| 200 | 58.3 | 56.7 | 62 | 62.3 | 74 | 70.3 | 73 | 66 | 73 | 75 |
| 1000 | 60 | 60 | 56 | 55 | 72 | 63 | 71 | 72 | 72 | 62 |

From table 6.1 it can be seen that the detection of complex tones (stimuli 5 to 10) is significantly higher than the pure tones. The fact that detectability increases with shorter stimulus duration is intriguing. The conditions of the subjects' responses are not specified in Green's paper.

An interesting point from Green's data is that detectability is improved with the complexity of the stimuli in the frequency plane.

### 6.2.3 Subjects Factor

The language of the subjects although a factor which was thought to be of no relevance to the Speech/Non-Speech discrimination, was eliminated from the experimental design by selecting British-born subjects in all the experiments. Subjects in experiment 1, 2, 3 were all male and only one of them was left-handed. There were no history of hearing difficulties among the subjects. Their average age varied between 21 and 27 years.

## 6.3 Experiment 1

The methodology, hypothesis and experimental design have been discussed in the previous sections of this Chapter. A three forced choice experiment, using Speech, Music and Noise as the stimuli, with a constant inter-stimuli interval while the stimulus duration was shortened was carried out on 8 subjects. The minimum duration is 80 msecs and the maximum is 1 second. Table 6.2 shows the average per subject-stimuli-Duration cell. Each average represents 12 replications less the responses which were longer than 1000 msecs

and were eliminated from the analysis. This represented about 1% of the total responses. T(i) represents the different durations: 1000, 500, 250, 125, 100, 80 msecs. P(i) represents the different subjects.

Table 6.3 is an auxiliary table used in the statistical analysis of the results to calculate the sum of the squares of the Stimuli and Duration factors. Table 6.4 is used to calculate the Subjects' factor sum of squares. Table 6.5 is used in graph 6.2, the psychological functions of reaction time and accuracy of the subjects' responses vs stimuli duration. The dotted lines are the reaction times for speech (red), music (green) and noise (blue). The continuous lines represent the accuracy.

Table 6.6 is the summary of the statistical analysis carried out on the reaction times. The analysis was performed in two steps due to the limited capacity of the statistical package ISIS. In order to calculate the error, an ANOVA analysis was carried out in each of the time blocks. The harmonic average of these errors was then used when the ISIS package was entered, the averages of the 12 replications as a single element in each Subject-Duration-Stimuli cell. The two factor variance was checked following the procedures described by Edwards (1972), using F values from Lindley (1953).

The significance of the duration factor is a confirmation of the hypothesis described in the introductory sections of this Chapter. The fact that the Subjects factor is also significant merely tells us that there are unequal performances when human beings carry out these tasks. The mean reaction times and accuracy suggests a pattern which is confirmed by the ANOVA. The duration of the stimulus is a significant factor, $F(5,\infty) = 3.02$, $p<0.01$. The subjects are also a significant source of variation. $F(7,\infty) = 2.64$, $p<0.01$. This variation is mainly due to the sharp increase in reaction time for stimuli shorter than 100 msecs. At the same point in duration, the accuracy of the responses begin to drop. As this reduction in accuracy does not fall to the guessing level, and is different from the accuracies for Music and Noise, the experiment can be said to be inconclusive.

The lack of significant variation due to the stimuli factor again confirms the hypothesis that there is no difference in the discriminatory performance for different types of stimuli regardless of the duration of the stimuli.

TABLE 6-2

| Duration / Subject / Stimuli | | T1 | T2 | T3 | T4 | T5 | T6 | Total |
|---|---|---|---|---|---|---|---|---|
| P1 | S | 706 | 679 | 664 | 664 | 669 | 759 | 4141 |
| | M | 717 | 665 | 658 | 647 | 609 | 743 | 4043 |
| | N | 692 | 636 | 637 | 681 | 730 | 735 | 4110 |
| P2 | S | 514 | 535 | 559 | 613 | 537 | 760 | 3518 |
| | M | 596 | 497 | 529 | 545 | 529 | 715 | 3411 |
| | N | 504 | 548 | 481 | 699 | 533 | 707 | 3472 |
| P3 | S | 597 | 631 | 543 | 590 | 685 | 706 | 3664 |
| | M | 621 | 555 | 620 | 678 | 564 | 738 | 3746 |
| | N | 585 | 606 | 514 | 707 | 698 | 708 | 3818 |
| P4 | S | 688 | 662 | 667 | 739 | 775 | 661 | 4192 |
| | M | 721 | 634 | 610 | 660 | 769 | 703 | 4097 |
| | N | 544 | 755 | 624 | 737 | 774 | 701 | 4138 |
| P5 | S | 558 | 540 | 612 | 590 | 501 | 789 | 3570 |
| | M | 666 | 574 | 677 | 576 | 517 | 657 | 3667 |
| | N | 536 | 547 | 643 | 547 | 536 | 638 | 3447 |
| P6 | S | 697 | 718 | 633 | 754 | 823 | 819 | 4444 |
| | M | 719 | 685 | 625 | 699 | 669 | 806 | 4203 |
| | N | 660 | 633 | 571 | 699 | 714 | 745 | 4022 |
| P7 | S | 646 | 617 | 600 | 633 | 697 | 783 | 3976 |
| | M | 685 | 573 | 688 | 741 | 814 | 821 | 4322 |
| | N | 751 | 522 | 547 | 587 | 700 | 700 | 3807 |
| P8 | S | 651 | 684 | 652 | 836 | 780 | 901 | 4484 |
| | M | 553 | 626 | 641 | 758 | 882 | 887 | 4307 |
| | N | 676 | 572 | 680 | 846 | 807 | 864 | 4405 |
| Total | | 15283 | 14998 | 14594 | 16166 | 16212 | 18051 | 95304 |

TABLE 6-3

| Duration Stimuli | | T1 | T2 | T3 | T4 | T5 | T6 | Total |
|---|---|---|---|---|---|---|---|---|
| | S | 5057 | 5066 | 4890 | 5399 | 5367 | 6180 | 31959 |
| | M | 5278 | 5113 | 5028 | 5281 | 5353 | 5968 | 32021 |
| | N | 4948 | 4819 | 4670 | 5483 | 5492 | 5903 | 31315 |
| Total | | 15283 | 14998 | 14594 | 16166 | 16212 | 18051 | 95304 |

TABLE 6-4

| Subj. Stim. | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | Total |
|---|---|---|---|---|---|---|---|---|---|
| S | 4141 | 3518 | 3664 | 4192 | 3570 | 4444 | 3976 | 4484 | 31959 |
| M | 4043 | 3411 | 3746 | 4097 | 3667 | 4203 | 4322 | 4307 | 32031 |
| N | 4110 | 3472 | 3818 | 4138 | 3447 | 4022 | 3807 | 4405 | 31315 |
| Tt. | 12294 | 10410 | 11228 | 12427 | 10684 | 12669 | 12105 | 13196 | 95295 |

TABLE 6-5

| Duration Stimuli | | T1 | T2 | T3 | T4 | T5 | T6 |
|---|---|---|---|---|---|---|---|
| | S | 632 | 633 | 611 | 675 | 671 | 772 |
| | M | 660 | 639 | 629 | 660 | 669 | 746 |
| | N | 619 | 602 | 585 | 685 | 687 | 738 |
| | S | 94 | 98 | 96 | 99 | 96 | 33 |
| | M | 96 | 98 | 99 | 98 | 96 | 88 |
| | N | 99 | 99 | 99 | 96 | 99 | 92 |

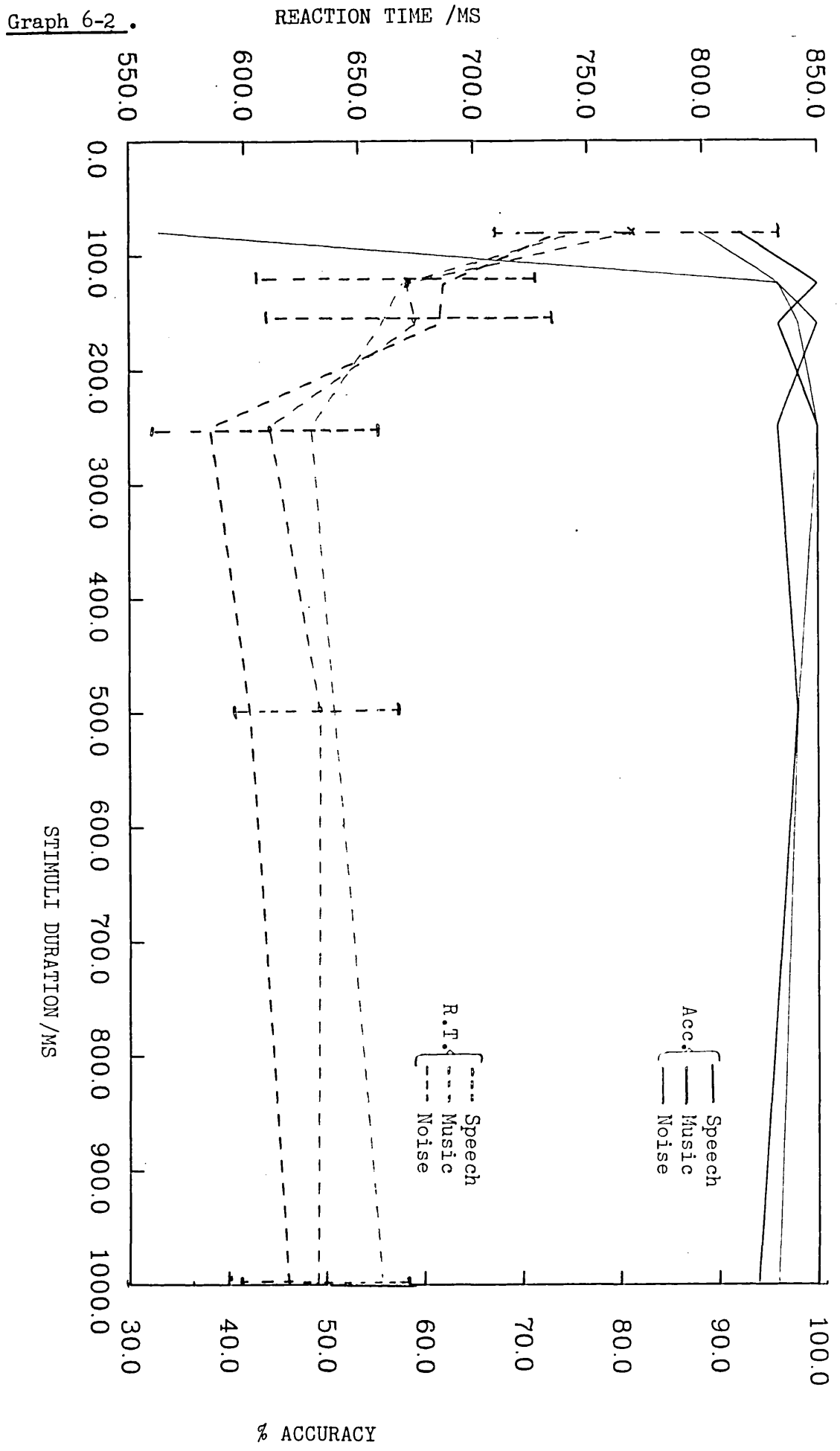
TABLE 6-6

ANOVA , Experiment 1.

| Source of Variation | Sum of Squares | D.F. | MSS | F |
|---|---|---|---|---|
| Stimuli | 6368 | 2 | 3184 | |
| Subjects | 384356 | 7 | 54908 | 2.75 |
| Duration | 330469 | 5 | 66093 | 3.31 |
| Duration- Subjects | 213608 | 35 | 6103 | |
| Stimuli-Subjects | 15031 | 10 | 1503 | |
| Duration-Stimuli | 42750 | 14 | 3504 | |
| Subj.-Dur.-Stimuli | 164831 | 70 | 2355 | |
| Within error | | | 19983 | |

Graph 6-2 depicts the reaction time, in dotted lines; and accuracy in continuous lines, for speech, noise and music . The red vertical bars represent the standard deviation for reaction times to speech stimuli. The deviations for music and noise are omitted for clarity. They are not significantly different from that for speech.

Graph 6-2 •


REACTION TIME /MS
550.0
600.0
650.0
700.0
750.0
800.0
850.0
0.0
100.0
200.0
300.0
400.0
500.0
600.0
700.0
800.0
900.0
1000.0
STIMULI DURATION /MS
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0
% ACCURACY
EXPERIMENT No 1
Acc.
Speech
Music
Noise
R.T.
Speech
Music
Noise

The next experiment is a complement to the first and the discussion of its results is postponed so as to incorporate them into the final discussion.

## 6.4 Experiment 2

The introduction to the hypothesis of this experiment is described in the preceeding Chapter and the opening section of this Chapter.

Table 6.7 shows the averages per subject-stimuli-duration cell. The same procedures as experiment 1 was used. T(i) represent the different durations of the stimuli; 30, 60, 100, 125, 250, 500 msecs. P(i) are the subjects which are not the same as in the preceding experiment. Tables 6-8, 6.9 are again auxiliary tables to calculate the main factors sum of squares. Table 6.10 is represented in graph 6.3. Table 6.11 is the summary of the ANOVA analysis carried out following the same procedures as in experiment 1.

## 6.5 Experiment 3

This experiment employed the same subjects and procedures as experiment 2. Eight subjects were asked to respond, by key depressing, with their classification of the stimuli being heard as Speech, Music or Noise. The inter-stimuli interval was varied between 1 and 4 seconds in a random manner. The reaction times are expected to be longer than in Experiment 1 and the Accuracy is expected to vary accordingly. Subjects, although encouraged to be fast and accurate, might trade these terms, producing a pattern which might not be consistent with the nearly monotonic relationship between accuracy and reaction time seen in experiments 1 and 2.

### 6.5.1 Results

The averages of the 12 replications per subject-stimuli-duration cell are in table 6.12. The summary of the ANOVA is in table 6.14 and the total averages of reaction time and accuracy are plotted against the stimuli duration in graph 6.4.

TABLE 6-7

| Duration Stimuli Subject | | T1 | T2 | T3 | T4 | T5 | T6 | Total |
|---|---|---|---|---|---|---|---|---|
| P1 | S | 695 | 662 | 667 | 670 | 762 | 780 | 4236 |
| | M | 705 | 693 | 603 | 634 | 740 | 750 | 4125 |
| | N | 701 | 710 | 659 | 670 | 836 | 866 | 4442 |
| P2 | S | 614 | 639 | 592 | 564 | 640 | 640 | 3689 |
| | M | 589 | 616 | 682 | 607 | 670 | 660 | 3824 |
| | N | 618 | 534 | 606 | 571 | 670 | 658 | 3657 |
| P3 | S | 640 | 618 | 613 | 557 | 619 | 692 | 3739 |
| | M | 570 | 644 | 591 | 563 | 622 | 675 | 3665 |
| | N | 648 | 600 | 646 | 624 | 745 | 812 | 4075 |
| P4 | S | 545 | 600 | 626 | 613 | 668 | 782 | 3834 |
| | M | 554 | 588 | 600 | 606 | 576 | 742 | 3646 |
| | N | 623 | 516 | 630 | 568 | 601 | 782 | 3720 |
| P5 | S | 654 | 712 | 675 | 656 | 688 | 780 | 4165 |
| | M | 776 | 763 | 780 | 782 | 679 | 826 | 4506 |
| | N | 711 | 734 | 748 | 765 | 658 | 769 | 4385 |
| P6 | S | 747 | 750 | 724 | 681 | 829 | 858 | 4589 |
| | M | 718 | 694 | 693 | 710 | 857 | 792 | 4461 |
| | N | 628 | 654 | 631 | 724 | 754 | 909 | 4300 |
| P7 | S | 636 | 617 | 653 | 650 | 859 | 916 | 4331 |
| | M | 617 | 605 | 729 | 704 | 864 | 939 | 4468 |
| | N | 577 | 524 | 626 | 640 | 804 | 787 | 3940 |
| P8 | S | 643 | 631 | 637 | 695 | 856 | 868 | 4330 |
| | M | 672 | 674 | 681 | 755 | 865 | 811 | 4456 |
| | N | 624 | 682 | 666 | 664 | 834 | 879 | 4349 |

TABLE 6-8

| Duration Stimuli | T1 | T2 | T3 | T4 | T5 | T6 | Total |
|---|---|---|---|---|---|---|---|
| S | 5174 | 5229 | 5187 | 5086 | 5921 | 6316 | 32913 |
| M | 5181 | 5285 | 5359 | 5261 | 5770 | 6195 | 33151 |
| N | 5130 | 4954 | 5167 | 5226 | 5902 | 6462 | 32886 |
| Total | 15485 | 15468 | 15713 | 15573 | 17593 | 18973 | 98950 |

TABLE 6-9

| Subj. Stim. | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| S | 4236 | 3689 | 3739 | 3834 | 4165 | 4589 | 4331 | 4330 | 32913 |
| M | 4125 | 3824 | 3665 | 3646 | 4506 | 4461 | 4468 | 4456 | 33151 |
| N | 4442 | 3657 | 4075 | 3270 | 4385 | 4300 | 3940 | 4349 | 32868 |
| Tot. | 12083 | 11170 | 11479 | 11190 | 13056 | 13350 | 12739 | 13135 | 98950 |

TABLE 6-10

| Duration Stimuli | | | | | | |
|---|---|---|---|---|---|---|
| S | 647 | 654 | 648 | 636 | 740 | 790 |
| M | 648 | 661 | 670 | 658 | 721 | 774 |
| N | 641 | 619 | 652 | 653 | 738 | 808 |
| S | 98 | 94 | 90 | 94 | 58 | 46 |
| M | 98 | 98 | 90 | 88 | 64 | 50 |
| N | 99 | 99 | 94 | 96 | 88 | 62 |

TABLE 6-11

| ANOVA, Experiment 2 . | | | | |
|---|---|---|---|---|
| Source of Variation | Sum of Squares | D.F. | MSS | F |
| Stimuli | 326 | 2 | 163 | |
| Duration | 453249 | 5 | 90645 | 4.44 |
| Subjects | 331443 | 7 | 47349 | 2.32 |
| Stimuli-Duration | 18801 | 10 | 1880 | |
| Subjects-Duration | 206303 | 35 | 5895 | |
| Subjects- Stimuli | 68860 | 14 | 4919 | |
| Subj.-Stimuli-Dur. | 82553 | 70 | 1179 | |
| Within error | | | 20423 | |

Graph 6-3.

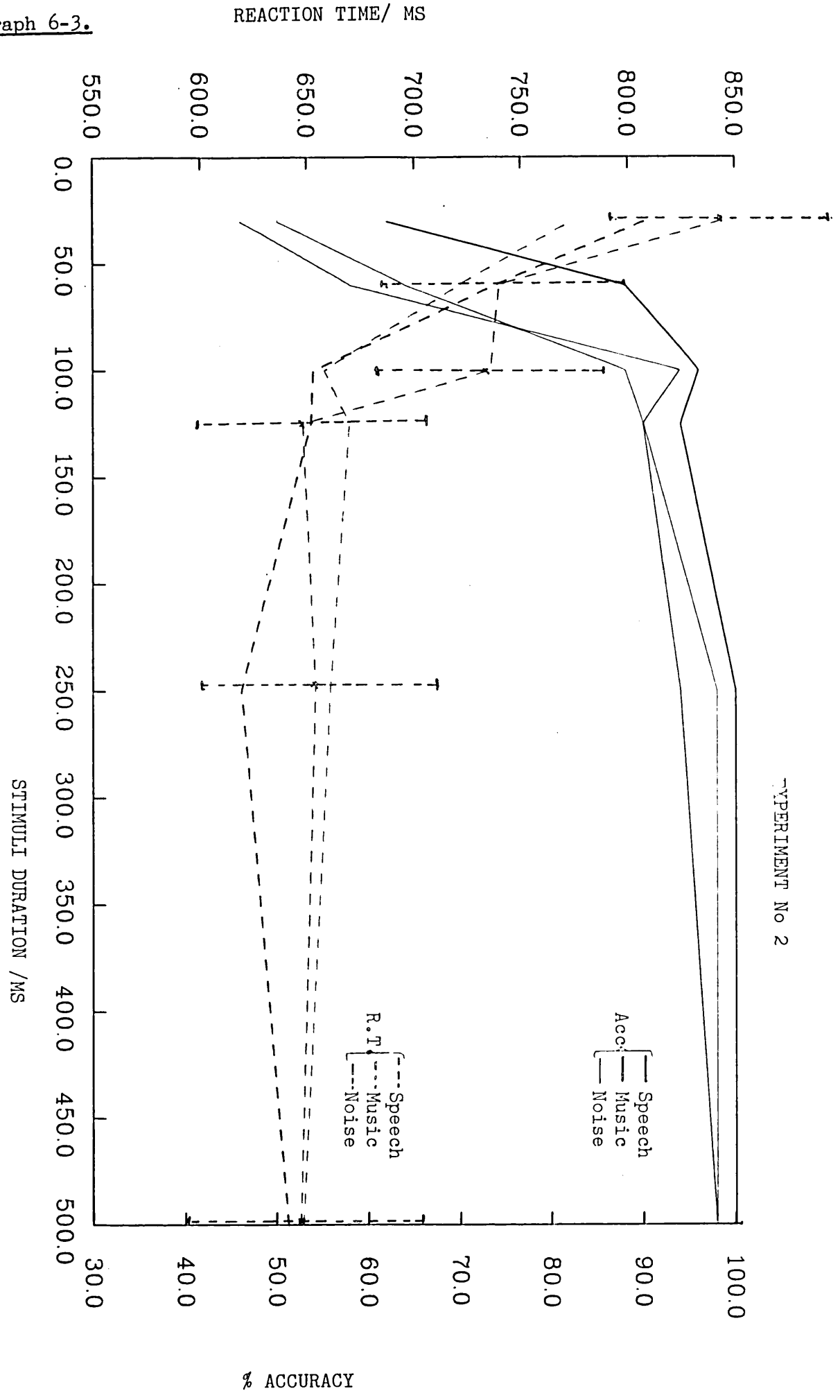

REACTION TIME/ MS
550.0
600.0
650.0
700.0
750.0
800.0
850.0
0.0
50.0
100.0
150.0
200.0
250.0
300.0
350.0
400.0
450.0
500.0
STIMULI DURATION /MS
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0
% ACCURACY
Acc:
Speech
Music
Noise
R.T
Speech
Music
Noise
EXPERIMENT No 2

TABLE 6- 12

| | | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| T1 | S | 659 | 684 | 634 | 644 | 666 | 721 | 688 | 735 | 5431 |
| | M | 649 | 634 | 645 | 649 | 571 | 695 | 693 | 743 | 5279 |
| | N | 634 | 628 | 689 | 564 | 685 | 662 | 662 | 708 | 5242 |
| T2 | S | 571 | 608 | 649 | 568 | 613 | 742 | 688 | 730 | 5169 |
| | M | 606 | 617 | 570 | 623 | 695 | 769 | 659 | 782 | 5321 |
| | N | 541 | 619 | 575 | 580 | 592 | 732 | 620 | 710 | 4974 |
| T3 | S | 663 | 670 | 661 | 527 | 664 | 787 | 695 | 656 | 5323 |
| | M | 708 | 731 | 637 | 570 | 798 | 735 | 813 | 740 | 5732 |
| | N | 565 | 641 | 669 | 560 | 755 | 700 | 656 | 703 | 5249 |
| T4 | S | 790 | 660 | 647 | 693 | 685 | 795 | 777 | 748 | 5795 |
| | M | 734 | 669 | 644 | 605 | 676 | 765 | 795 | 794 | 5702 |
| | N | 602 | 637 | 601 | 615 | 728 | 719 | 712 | 725 | 5339 |
| T5 | S | 738 | 726 | 700 | 610 | 779 | 879 | 899 | 872 | 6203 |
| | M | 821 | 738 | 610 | 683 | 814 | 897 | 883 | 907 | 6353 |
| | N | 682 | 665 | 625 | 577 | 786 | 798 | 771 | 869 | 5783 |
| T6 | S | 904 | 752 | 815 | 671 | 866 | 918 | 864 | 963 | 6753 |
| | M | 983 | 775 | 764 | 749 | 940 | 899 | 977 | 862 | 6949 |
| | N | 783 | 707 | 826 | 773 | 883 | 937 | 731 | 900 | 6540 |

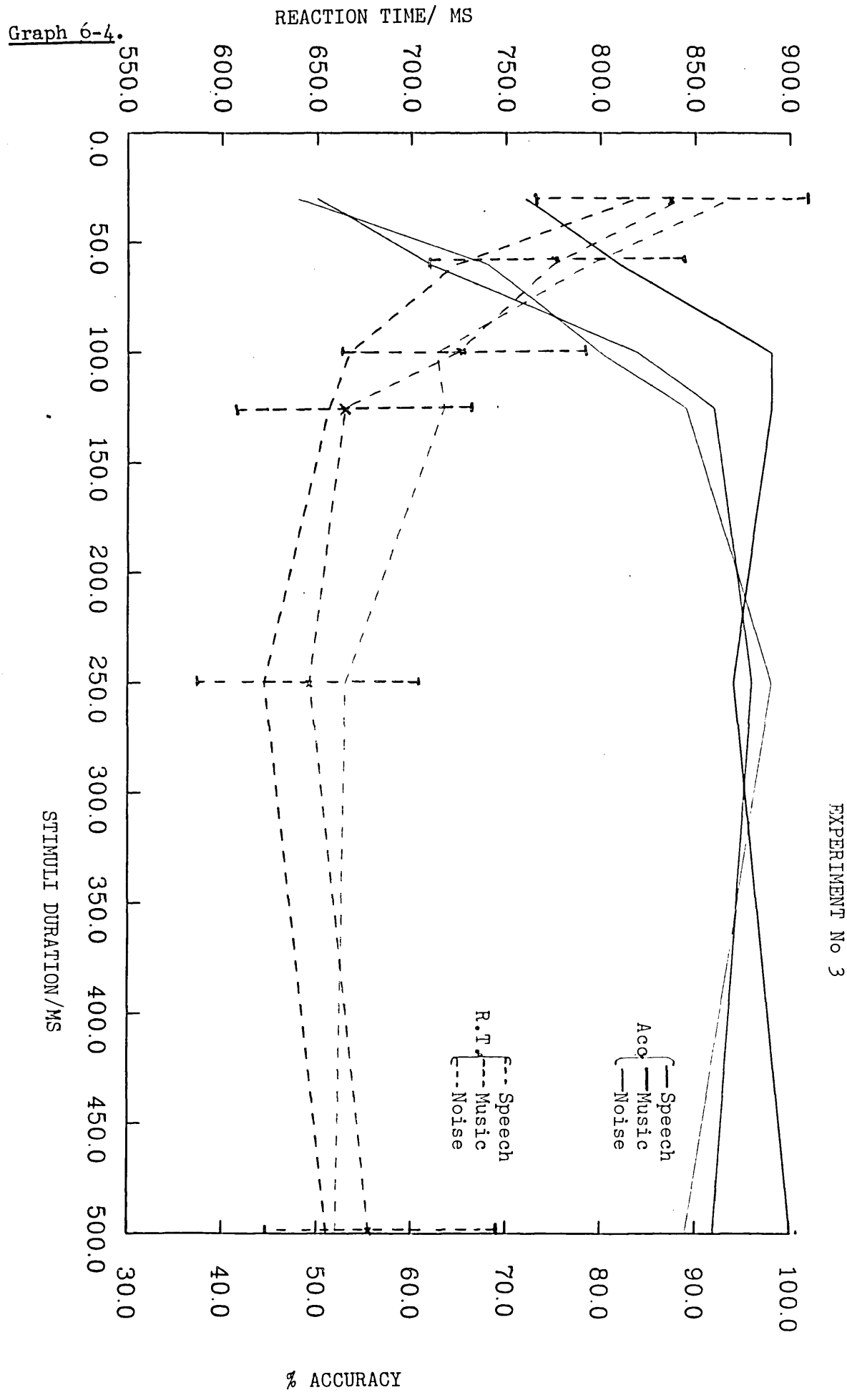
TABLE 6- 13

| Duration Stimuli | | T1 | T2 | T3 | T4 | T5 | T6 |
|---|---|---|---|---|---|---|---|
| R.T. (mscs) | S | 679 | 646 | 665 | 724 | 775 | 844 |
| | M | 660 | 665 | 717 | 713 | 794 | 869 |
| | N | 655 | 622 | 656 | 667 | 723 | 819 |
| Acc. (%) | S | 92 | 96 | 92 | 84 | 62 | 50 |
| | M | 89 | 98 | 89 | 80 | 68 | 48 |
| | N | 99 | 94 | 98 | 98 | 82 | 72 |

TABLE 6-14

| Analysis of Variance , Experiment 3 | | | | |
|---|---|---|---|---|
| Source of Variation | Sum of Squares | D.F. | M.S.S | F |
| Subjects | 435692 | 7 | 62242 | 3.2 |
| Stimuli | 44516 | 2 | 22258 | 1.1 |
| Duration | 669039 | 5 | 133808 | 6.9 |
| Duration-Subjects | 150499 | 35 | 4300 | |
| Stimuli-Subjects | 59901 | 14 | 4279 | |
| Duration-Stimuli | 20242 | 10 | 2024 | |
| Stimuli-Durat.-Subj. | 97173 | 70 | 1388 | |
| Error | | | 19467 | |

Graph 6-4.


REACTION TIME/ MS
550.0
600.0
650.0
700.0
750.0
800.0
850.0
900.0
0.0
50.0
100.0
150.0
200.0
250.0
300.0
350.0
400.0
450.0
500.0
STIMULI DURATION/MS
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0
% ACCURACY
Acc
Speech
Music
Noise
R.T.
Speech
Music
Noise
EXPERIMENT No 3

The dotted lines are the reaction times for speech, red; music, green and noise, blue. The continuous lines represent the accuracy. Note that the drawings cannot represent the 99% accuracy faithfully enough.

The psychological functions of accuracy and reaction time follow the same pattern as that obtained in experiments 1 and 2. The duration of the stimuli is highly significant. $F(5, \infty) = 4.10$, $p < 0.001$. The subjects factor is still significant $F(7, \infty) = 2.64$, $p < 0.01$.

A 't' test comparing the grand means for speech and noise for a duration of 500 msecs gives some results which are not significant.

For a stimuli duration of 500 msecs.

Mean Reaction Time:

Speech = 679 msecs.

Noise = 654 msecs.

$t = 1.22$ $t(14,0.1) = 2.624$

The above results which present a non-significant difference for the mean reaction times for speech and noise, differs from the difference of reaction times for a vowel /a:/ and the sound of a bell obtained by Fry (1975).

In Fry's experiment, subjects were asked to detect two sounds. The stimuli were not described to the subjects prior to the experiment. The subjects did not need to classify the signal in order to detect it. They were merely to react to a sound...

The subject's pre-experimental subjective priming (state of attention and knowledge of the nature of the stimuli) plays a significant role in the outcome of experiments which, by their lack of "speech environment", are designed to elicit mechanisms related to speech.

A very clear example of this phenomenon is given by Remez and Rubin (1982). In their experiment a seven word phrase waveform was stripped of its "speech likeness" by reducing the bandwidth of its three first formants until they became a pure tone, centred in the original formant's bandwidth. Subjects were then asked to infer the original phrase.

Remez and Rubin conclude that: "The use of a sinusoidal replicas of speech signals reveals that listeners can perceive speech solely from temporally coherent spectral variations of non-speech acoustic elements"... "Listeners told nothing in advance about the three tone signals, heard them simply as three simultaneous tones, modulated asynchronously as if three part counterpoint. However, the simple instruction to listen for a sentence enabled almost 70% of naive listeners to detect a sizeable chunk of the information that was exclusively time-varying in nature, in the absence of short-time spectra characteristic of vocalization".....

From these considerations, the results of Fry's experiment might be seen as a consequence of the higher accuracy in detecting more complex sound elicited by Green (1958), and described in section 6.4 of this Chapter. Fry's subjects reacted to the more complexed /a:/ with all its bandwidth formants' complexities more rapidly than to the bell but in doing so, they need not have detected the former as speech at all. The reaction times recorded were not for the performance of <u>any</u> discrimination.

A "t" test carried out comparing the reaction times for experiments 2 and 3 yield significant differences for the averages of the responses to speech and music, yet the difference for the average response times for noises is not significant. $t(47,0.002)= 3.2$.

| | Speech (E3-E2) | Music (E3-E2) | Noise (E3-E2) |
|---|---|---|---|
| D | 34.4 | 47.7 | 8.2 |
| sd | 71.5 | 70.3 | 72.1 |
| t | 3.35 | 4.5 | 0.71 |
| D.F. | 47 | 47 | 47 |

The results confirm the hypothesis that attention is a relevant parameter in the discrimination of speech and non-speech. The lack of significance for the difference in the averages of reaction time for the Noise stimuli is puzzling given the lack of significance of the stimulus factor in all the ANOVAs carried out so far.

The analysis of the data obtained in experiments 1, 2 and 3, and observation of the graphs 6.2, 6.3, lead to the following conclusions:

1. The average stimulus duration necessary for a correct decision in 90% of presentation, varied for subjects, averaging 70 msecs for noise, 110 msecs for music and 90 msecs for speech.

2. No significant tendency for errors was found when discriminating between speech, music and noise. The duration of the stimuli is significant at $p < 0.5\%$. The variation between subjects is also significant at $p < 0.5\%$.

3. Random variation of the inter-stimuli interval prolonged the subject's reaction time and diminished the accuracy of the Speech/Non-Speech discrimination. This is clear when comparing the graphs of the accuracy of the responses in experiment 2 and 3.

4. The difficulties in assessing the speed of the discriminatory stage of the overall reaction time arise from the reluctance of the author to consider the various processes which make up the reaction time as additive in time. The simple subtraction of the motor reaction time elicited in experiment described in Chapter V will not yield a clear result. The average reaction time for stimuli durations above 100 msecs in experiments 1 and 2 is 644 msecs. The average reaction time for the replication of Merkel's experiment is 336 msecs. Reaction time to visual stimuli are generally 20% longer than auditory stimuli. This is thought due to the photochemical processes which convert light into electrical energy. (Fitts, 1973).

## 6.6 Experiment 4

Right Ear Advantage in the decoding of consonants and other speech related acoustic patterns have prompted the author to investigate the presence of REA in the Speech/Non-Speech discrimination. The hypothesis and introductory remarks to this experiment are contained in the introduction to this Chapter and in the preceding Chapter.

Since this operation is carried out with such speed that it could not involve the cortex, it is logical to expect no brain hemispheric advantage.

Each subject made 12 replications per stimuli-ear-duration cell making a total of 432 responses. The first version of this experiment varied the ear to be stimulated in a random manner, while the subjects made their responses using the same keyboard used in previous experiments. A link between the ear being stimulated and the position of the key was detected, forcing a change in the experimental procedure. The stimuli being relayed to the different ears were grouped so each duration block had equal numbers of stimuli relayed to the right and left ear. At the beginning of each ear block, six responses were allowed to inform the subject of the change of ear. These results were eliminated from the analysis. The subjects made their responses in two sessions separated by a week. The keyboard orientation was changed between sessions. The experimental conditions are the same as in previous experiments.

The averages of the Stimuli-Subject-Ear-Duration cell are shown in table 6.15. Tables 6.17, 6.18 and 6.19 are auxiliary tables which are used to check the main and multi-mode factors of the 4 way ANOVA.

Tables 6.20 and 6.19 are represented in graphs 6.5, 6.6 and 6.7, the solid lines represent the accuracy and the broken lines the reaction times. The graphs depict these functions separately for speech, music and noise to avoid the presentation of too much information in a single graph.

The duration of the stimuli is again highly significant. $F(5.00) = 4.83$, $p < 0.001$. The subjects factor is also significant at $p < 0.01$. The lack of significance of the Ear factor is clear. This is also noted from the graphs. The stimulus factor is again not significant.

The differences between the averages of the reaction times for the right and left ear are shown in table 6.22.

RT(R-L) msecs.

| Duration | T6 | T5 | T4 | T3 | T2 | T1 |
|---|---|---|---|---|---|---|
| Stimuli | | | | | | |
| S | 40 | 14 | -20 | 17 | -8 | 20 |
| M | 18 | -11 | 71 | 17 | -26 | 10 |
| N | 25 | -60 | -26 | -50 | -12 | -26 |

TABLE 6-15

| | | | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| T1 | R | S | 647 | 689 | 552 | 657 | 645 | 642 | 715 | 585 | 5132 |
| | | M | 672 | 674 | 580 | 635 | 727 | 659 | 708 | 671 | 5326 |
| | | N | 585 | 64o | 523 | 590 | 638 | 654 | 680 | 570 | 4880 |
| | L | S | 678 | 768 | 512 | 562 | 675 | 706 | 736 | 665 | 5302 |
| | | M | 634 | 665 | 541 | 675 | 748 | 673 | 658 | 654 | 5248 |
| | | N | 660 | 645 | 509 | 640 | 670 | 697 | 651 | 619 | 5091 |
| T2 | R | S | 554 | 601 | 566 | 652 | 694 | 582 | 642 | 701 | 4992 |
| | | M | 610 | 596 | 550 | 686 | 638 | 675 | 667 | 626 | 5048 |
| | | N | 509 | 660 | 496 | 696 | 675 | 656 | 654 | 629 | 4975 |
| | L | S | 572 | 713 | 568 | 639 | 715 | 557 | 647 | 645 | 5056 |
| | | M | 684 | 728 | 624 | 627 | 689 | 611 | 608 | 679 | 5250 |
| | | N | 655 | 587 | 558 | 592 | 695 | 642 | 697 | 649 | 5075 |
| T3 | R | S | 615 | 672 | 581 | 741 | 666 | 605 | 606 | 740 | 5226 |
| | | M | 606 | 730 | 609 | 696 | 674 | 725 | 722 | 779 | 5541 |
| | | N | 627 | 603 | 498 | 599 | 617 | 577 | 575 | 604 | 4700 |
| | L | S | 629 | 595 | 560 | 676 | 684 | 611 | 684 | 652 | 5091 |
| | | M | 618 | 755 | 619 | 740 | 619 | 684 | 631 | 745 | 5411 |
| | | N | 657 | 751 | 612 | 668 | 651 | 577 | 584 | 607 | 5107 |
| T4 | R | S | 547 | 608 | 652 | 730 | 737 | 575 | 701 | 619 | 5169 |
| | | M | 675 | 765 | 669 | 686 | 853 | 617 | 683 | 807 | 5755 |
| | | N | 587 | 701 | 644 | 674 | 637 | 667 | 667 | 681 | 5258 |
| | L | S | 611 | 734 | 675 | 677 | 674 | 546 | 750 | 659 | 5326 |
| | | M | 624 | 690 | 649 | 540 | 650 | 561 | 695 | 777 | 5186 |
| | | N | 579 | 679 | 731 | 570 | 703 | 707 | 774 | 626 | 5369 |
| T5 | R | S | 625 | 763 | 597 | 638 | 780 | 657 | 857 | 748 | 5665 |
| | | M | 727 | 756 | 678 | 733 | 793 | 782 | 835 | 762 | 6066 |
| | | N | 623 | 664 | 591 | 512 | 702 | 656 | 620 | 656 | 5024 |
| | L | S | 711 | 699 | 625 | 676 | 760 | 712 | 789 | 580 | 5552 |
| | | M | 795 | 831 | 690 | 658 | 838 | 670 | 855 | 813 | 6150 |
| | | N | 771 | 736 | 633 | 590 | 725 | 690 | 641 | 723 | 5509 |
| T6 | R | S | 816 | 884 | 701 | 696 | 855 | 760 | 726 | 729 | 6167 |
| | | M | 642 | 863 | 672 | 815 | 853 | 663 | 819 | 610 | 5935 |
| | | N | 700 | 725 | 633 | 812 | 930 | 797 | 810 | 786 | 6193 |
| | L | S | 898 | 759 | 708 | 737 | 707 | 655 | 689 | 680 | 5833 |
| | | M | 635 | 845 | 709 | 678 | 819 | 698 | 783 | 627 | 5785 |
| | | N | 669 | 762 | 702 | 722 | 808 | 680 | 802 | 849 | 5994 |

TABLE 6-16

| Stimuli / Ear | S | M | N | Total |
|---|---|---|---|---|
| Left | 32160 | 33030 | 32145 | 97335 |
| Right | 32351 | 33671 | 31030 | 97052 |
| Total | 64511 | 66701 | 63175 | 194387 |

TABLE 6-17

| Duration / Ear | T1 | T2 | T3 | T4 | T5 | T6 | Total |
|---|---|---|---|---|---|---|---|
| Right | 15338 | 15015 | 15467 | 16182 | 16755 | 18295 | 97335 |
| Left | 15641 | 15381 | 15609 | 15881 | 17211 | 17612 | 97052 |
| Total | 30979 | 30405 | 31076 | 32063 | 33966 | 36216 | 194387 |

TABLE 6-18

| Subject / Ear | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | Total |
|---|---|---|---|---|---|---|---|---|---|
| Right | 11367 | 12594 | 10792 | 12248 | 13112 | 11952 | 12687 | 12303 | 97335 |
| Left | 12080 | 12942 | 11225 | 11667 | 12830 | 11677 | 12674 | 12249 | 97052 |
| Total | 23447 | 25536 | 22017 | 23915 | 25942 | 23629 | 25361 | 24552 | 194387 |

TABLE 6-19

| Duration / Ear-Stimuli | | T1 | T2 | T3 | T4 | T5 | T6 |
|---|---|---|---|---|---|---|---|
| Right | S | 642 | 624 | 653 | 646 | 708 | 771 |
| | M | 666 | 631 | 693 | 719 | 758 | 741 |
| | N | 610 | 622 | 588 | 657 | 628 | 774 |
| Left | S | 663 | 632 | 636 | 666 | 694 | 733 |
| | M | 656 | 657 | 676 | 648 | 769 | 723 |
| | N | 636 | 634 | 638 | 671 | 688 | 749 |

TABLE 6-20

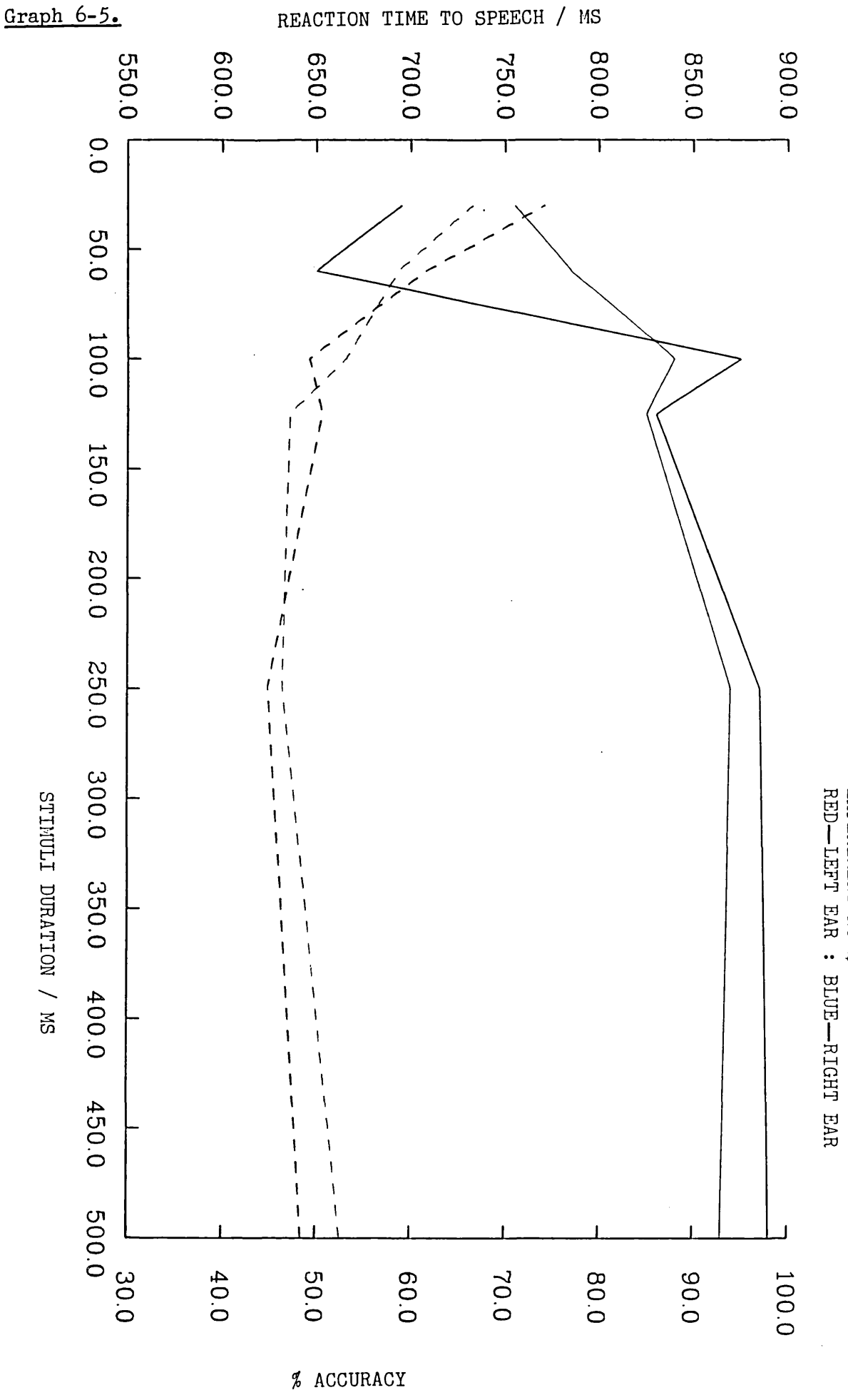
| Duration Ear- Stimuli | | | | | | | |
|---|---|---|---|---|---|---|---|
| Right | S | 99 | 97 | 86 | 95 | 50 | 59 |
| | M | 98 | 97 | 92 | 90 | 70 | 72 |
| | N | 99 | 97 | 97 | 95 | 99 | 67 |
| Left | S | 93 | 94 | 85 | 88 | 77 | 71 |
| | M | 98 | 95 | 94 | 88 | 78 | 81 |
| | N | 99 | 99 | 98 | 96 | 87 | 74 |

TABLE 6-21

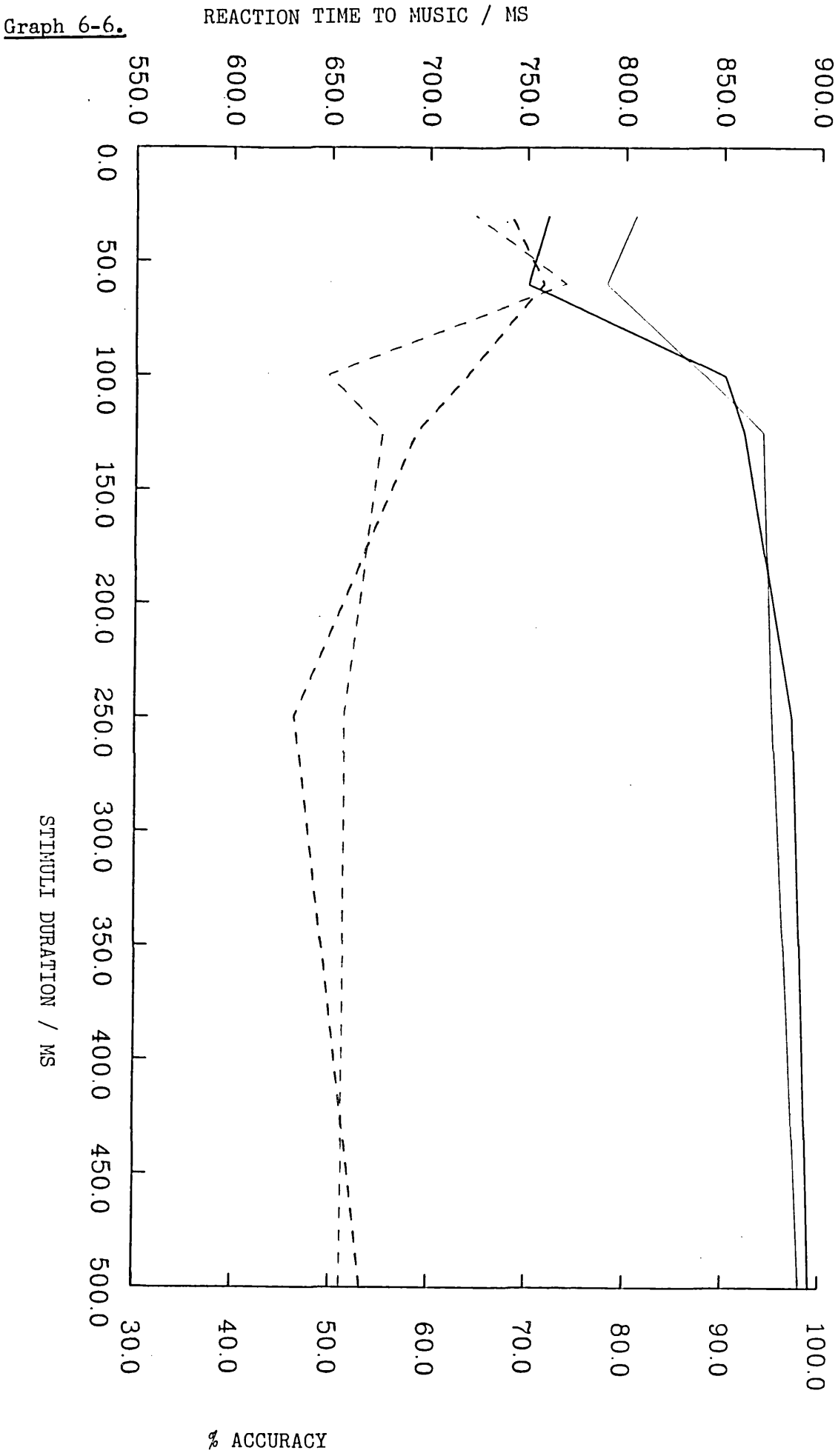
Analysis of Variance , Experiment 4.

| Source of Variance | Summ of Squares | D.F. | M.SS | F |
|---|---|---|---|---|
| Ear | 278 | 1 | 278 | |
| Stimuli | 66019 | 2 | 33009 | 1,69 |
| Subjects | 329728 | 7 | 47104 | 2,41 |
| Time | 471941 | 5 | 94388 | 4,83 |
| Ear-Stimuli | 17331 | 5 | 3466 | |
| Ear-Subjects | 38440 | 7 | 5540 | |
| Ear-Time | 20783 | 5 | 4156 | |
| Stimuli-Subjects | 31437 | 14 | 2245 | |
| Stimuli-Time | 93162 | 10 | 9316 | |
| Subjects-Time | 256189 | 35 | 7319 | |
| Ear-Stimuli-Subjects | 8527 | 14 | 609 | |
| Ear-Stimuli-Time | 31893 | 10 | 3189 | |
| Stimuli-Subjects-Time | 236753 | 70 | 3382 | |
| Ear-Subjects-Time | 52048 | 35 | 1487 | |
| Time-Ear-Subject-Stimuli | 147560 | 70 | 2108 | |
| Error | | | 19527 | |

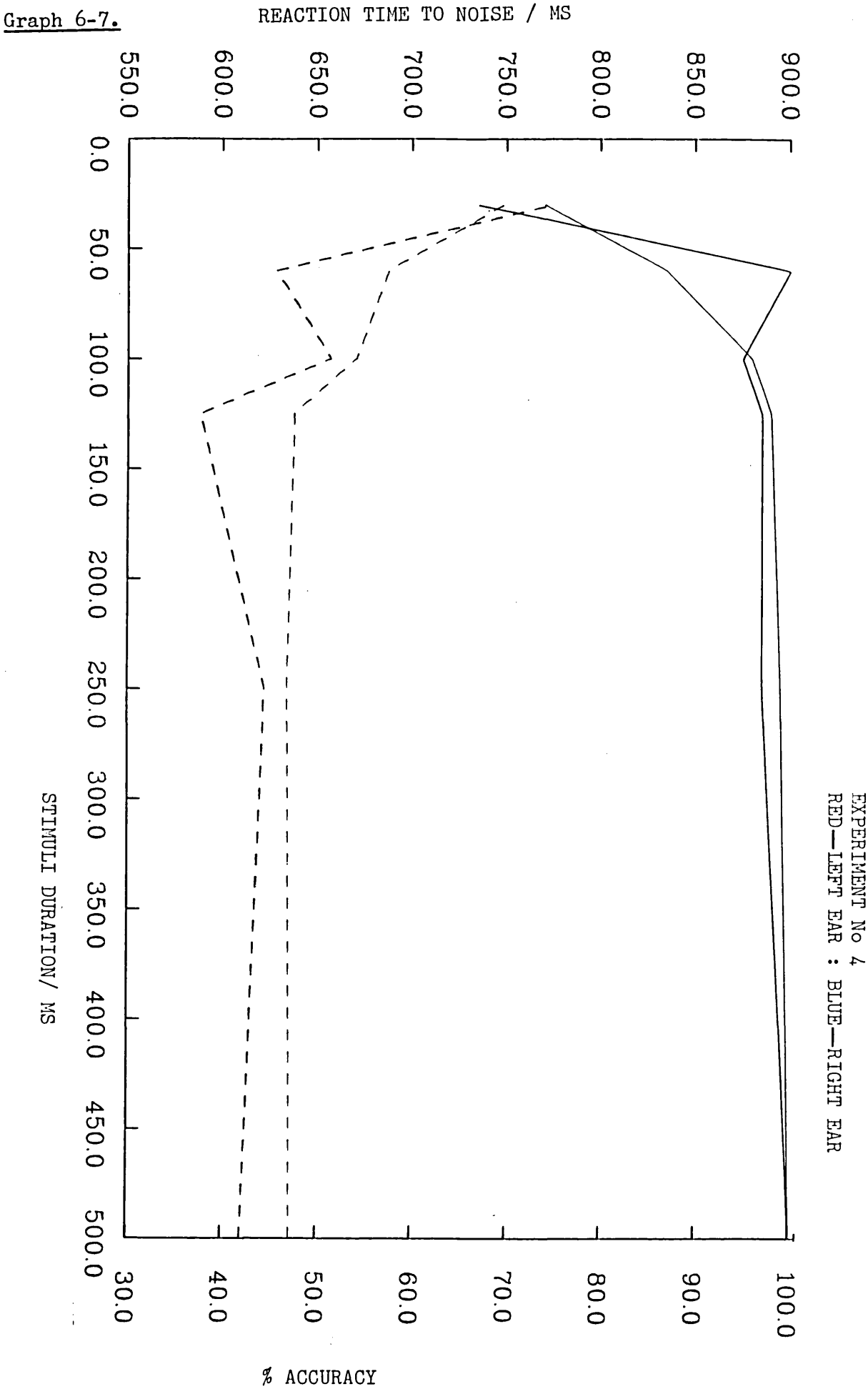
Graph 6-5.


REACTION TIME TO SPEECH / MS
550.0
600.0
650.0
700.0
750.0
800.0
850.0
900.0
0.0
50.0
100.0
150.0
200.0
250.0
300.0
350.0
400.0
450.0
500.0
STIMULI DURATION / MS
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0
% ACCURACY
EXPERIMENT No 4
RED—LEFT EAR : BLUE—RIGHT EAR

Graph 6-6.


REACTION TIME TO MUSIC / MS
550.0
600.0
650.0
700.0
750.0
800.0
850.0
900.0
0.0
50.0
100.0
150.0
200.0
250.0
300.0
350.0
400.0
450.0
500.0
STIMULI DURATION / MS
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0
% ACCURACY
EXPERIMENT No 4
RED—LEFT EAR : BLUE—RIGHT EAR

Graph 6-7.


REACTION TIME TO NOISE / MS
550.0
600.0
650.0
700.0
750.0
800.0
850.0
900.0
0.0
50.0
100.0
150.0
200.0
250.0
300.0
350.0
400.0
450.0
500.0
STIMULI DURATION/ MS
30.0
40.0
50.0
60.0
70.0
80.0
90.0
100.0
% ACCURACY
EXPERIMENT No 4
RED—LEFT EAR : BLUE—RIGHT EAR

## 6.7 Concluding Remarks

The theoretical "hole" in the current models of speech perception discovered through an engineer's approach to a perceptual problem was picked out through the heuristic suspicion that the perception of an aural stimulus is carried out with reference to the class of stimuli to which it belongs as a communicative sign. This intermediate stage of knowledge is present in the examples given in Chapters II and III. The trend of psychologists to "migrate" to the linguistic province and the exhaustion of the work on "visible speech" created this vacuum.

The description of the neurological links between speech perception and production contrasted with the linguist's heterarchical description of the production of speech in the form of language.

The Speech/Non-Speech discrimination is carried out extremely fast under favourable conditions of attention and is the prior operation setting in motion all the subsequent speech decoding levels. Phonetic categorization is affected by the subjective "priming" of subjects to the class of stimuli to be heard. Semantic decoding is also affected by the prior knowledge of the stimuli category.

The storage in memory of the Speech/Non-Speech discrimination carried out in shadowing experiments also demonstrates lack of disruption of this discrimination by the subsequent stimulus category.

Descriptions of ontogenic development points to the fact that this discrimination is present in the early weeks of human life.

The prior requirement that the Speech/Non-Speech discrimination occur before the decodification of the speech stream suggests that an extra 'box' needs to be added to the three-tier processing model which is currently used to describe speech perception.

The experimental study of the Speech/Non-Speech discrimination led us to the following conclusions:

The brevity cf the Speech/Non-Speech discrimination is confirmed by the results obtained in the experiments. The shortest duration of a speech stimuli to be correctly discriminated with an accuracy of 90% is approximately 80 msecs. This time is comparable with a third of the duration of an average syllable.

The results presented by Wood (1975) ,indicate statistically significant differences between EGC readings related to phonetic and auditory tasks. Such differences appear initially at 60-80 msec. after aural stimulation. This is another indication that the differentation between auditory and phonetic type of "decoding" in speech perception is achieved during the early stages of the decodification of the stimuli.

Direct comparison between the data produced by Wood (1975), Fry (1974,1975), Hammond (1978) and Fitts (1973) with the results of this work is not posible due to the differences in experimental procedures. Wood used a 2x2 paradigm in which subjects were informed of the nature of the stimuli prior to the experiment. The stimuli duration used by Wood was constant,300 msec. Fry used simple reaction times to measure statistically significant differences between reaction times to a vowel and to the sound of a bell. As pointed out earlier in this chapter,Fry's results can be atributed to the preference for detection of complex stimuli measured by Green.

The reaction times are no different for the three classes of stimuli used. This reflects the fact that the same operation is carried out upon the different types of stimuli. This should not lead to the conclusion that subsequent decoding cannot be carried out within the period of the reaction times measured in experiments 2, 3 and 4. The author is reluctant to subtract motor reaction times from the discrimination related period of the overall reaction times, because of the possibility of parallel processing.

The lack of REA found in this discrimination and the brevity of the stimuli which the subjects were able to discriminate suggests a lack of cortical involvement and consequently its non-linguistic character. Note that the lack of REA allows this conclusion, yet the contrary cannot be inferred. Presence of REA in linguistic related tasks does not imply that the existence of REA is exclusive to speech related mechanisms.

The role of the attention upon the Speech/Non-Speech discrimination emerges from the comparison of the results of experiments 3 and 2. The significance of the differences between averages for the discrimination of speech and music allow the conclusion that the changes in attention, arising from variations of the interstimuli interval, affect the reaction time and the accuracy of the subjects' responses. The fact that the comparison of the averages of responses to noise did not yield a significant result is puzzling given the lack of significant differences related to the stimulus factor found in experiments 1, 2, 3 and 4. This could be due to the coarseness of the ANOVA analysis carried out across the different levels of the duration of the stimuli, yet a visual inspection of the graphs depicting the accuracy and reaction times as a function of the stimuli duration lead to the same conclusion . The reaction time functions for the different stimuli are not parallel.

The analysis of the reaction times and accuracy from the graphs 6-2, 6-3 and 6-4 , indicate that the trade off between speed and accuracy changed when the stimuli was shorter than 100 msec. The accuracy was maintained but the reaction time began to increase.

The extent of individual differences in the reaction time data may be evaluated by examining the main effect of subjects and the interaction of the subjects factor with other factors. In all four experiments the main effect of subjects was highly significant indicating that the individual subjects differed considerably in their reaction time. The analysis of the multifactors indicate that the subjects were consistently slow or fast in their reactions to the diferent type and duration of the stimuli.

These conclusions suggest further research. An investigation of the presence of REA in relation to attention could be achieved by a monaural presentation of the stimuli used in experiment 3 and comparing the data obtained with the data of experiment 4.

A change of task of Fry's experiment described in Chapter 6, could provide some clues concerning the effect of the number of alternatives upon the Speech/Non-Speech discrimination. This could be achieved by instructing subjects to react to a specific class of stimuli, incorporating their Speech/Non-Speech discrimination operation into the overall reaction time.

The prospect of obtaining some insights into the Speech/Non-Speech discriminatory processes in both first and foreign language can be obtained by repeating these series of experiments in different languages. The role of vowels and consonants in this discrimination can also be studied by a careful editing of the tapes, incorporating this factor into the ANOVA analysis.

The manipulation of the stimuli characteristics is the avenue chosen by Haskins Research Laboratories for research into the differences between the two modes of processing the auditory system. Studies of "duplex" perception indicates the existence of these two modes of processing.

The results obtained in the series of experiments and the theoretical elements extracted from the literature converge with the result of recent studies at Haskins Research Laboratories. Bentin and Mann (1983) provide some results which give some "insight into how, and at what level of information processing, speech is recognised as such, and starts to be processed differentially".

They conclude "that speech perception involves activation of a central mechanism, while non-speech perception is more dependent on peripheral auditory processes".

The conclusions also require the existence of a pre-perceptual mechanism that enables the central processing of the speech stream by signalling its presence as speech.

The value of distinguishing between auditory and phonetic modes is speech perception is not only methodological. The evidence now being gathered , to which this work is a contribution, is conducive to the construction of a model for speech perception in which the auditory and phonetic analysis of the speech stream is carried out in a parallel manner, after being "pre-decoded" by a common auditory stage. The common analysis of both aspects of the speech stream "decodification", which is a peripheral process, is the locus for the operand upon which this work has been focused.

# REFERENCES

BAILEY P.J. (1977)
"On the Identification of Sine-Wave Analogies of certain Speech Sounds"
Haskins Laboratories Status Report on Speech Research, SR-51/52.

BADDELEY A.D. (1971)
"Relation between long term and short term memory"
British Medical Bulletin V 27, No.3

BENCH J. (1970)
"On the Assessment of Hearing and Intellect: Some Conceptual Problems"
"International Symposium on Speech Communication Ability and Profound Deafness"
Ed. G. Fant; Publ., A.G. Bell Asoc. for the Deaf, 1972

BENTIN S. AND MANN V.A. (1983)
"Selective effects of masking on Speech and Non-Speech in the Duplex perception paradigm"
Haskins Laboratories Status Report on Speech Research, SR-76.

BEKESY G. von (1971
"The Ear"
From: The Biological Bases of Behaviour, Open University Publications.

BLAKEMORE C. (1977)
"Mechanics of the Mind"
MIT Press, New York.

BLECHNER M.J. (1976)
"REA for Musical Stimuli Differing in Rise Time"
Haskins Laboratories Status Report on Speech Research, SR-47.

BROADBENT D.E. (1981)
"Perceptual Experiments and Language Theories"
Phil. Trans. Roy. Soc. (1077): 375-385

CALDER N. (1970)
"The Mind of Man"
BBC Publications.

CHERRY C. (1953)
"Some Experiments on The Recognition of Speech, with one ear and with two ears"
J. Acoust. Soc. of America 25, 975-979

CHERRY C.,SAYERS,B.McA, MARLAND P.M. (1955)
"Experiments on The Complete Suppression of Stammering".
Nature, 176: 874, 1955

CHERRY C. (1978)
"On Human Communication "
MIT Press, New York.

CHERNIGOVSKAYA, T.V. AND MOROZOV V.P. (1974)
"Link between thresholds of human hearing and amplitude modulated sounds and the amplitude modulation characteristics of Speech"
Biofizika 19, No.6, 1104-1105

CHOMSKY N. (1972)
"Language and Mind"
Harcourt, Brace & World, Inc. New York.

CLARK H.H. AND CLARK E.V. (1977)
"Psychology and Language"
Harcourt, Brace Jovanovich Inc., New York

CUTTING J.A. AND EIMAS P.D. (1974)
"Phonetic feature analysers and the processing of Speech in Infants"
Haskins Laboratories Status Report on Speech Research, SR-37/38.

CRAIK F.I.M. (1971)
"Primary Memory"
British Medical Bulletin, V 27, No.3

CROWDER R.G. (1971)
"The Sounds of vowels and consonants in immediate memory"
Journal of verbal Learning and Verbal Behaviour, 10, 587-596.

CUTTING J.E. (1975)
"On the Relationship of Speech and Language"
Haskins Laboratories Status Report on Speech Research, SR-41.

CUTTING J.E. (1975)
"Auditory and Linguistic processes in Speech perception. Inferences from six fusion dichotic listening"
Haskins Laboratories Status Report on Speech Research, SR-44.

CUTTING J.E. (1976)
"An information-processing approach to Speech perception"
Haskins Laboratories Status Report on Speech Research, SR-48.

CUTTING J.E. (1978)
"There may be nothing peculiar to perceiving in a Speech mode"
J. Requin (ed.) Attention and Performance" VII Hillsdale, N.J. Erlbaum, 1978, pp. 229-244

DAY R.S. (1972)
"Memory for dichotic pairs: Disruption of ear report performance by the Speech Non-Speech Distinction"
Haskins Laboratories Status Report on Speech Research, SR-31/32.

DESCHERIEV Y. (1979)
"El desarrollo de la Sociolinguistica en las condiciones de la R.C.T."
Ciencias Sociales 1, Academia de Ciencias de la URSS.

DOESSCHATE G. (1963)
"Notes on the history of Reaction Time Measurements"
Philips Technical Review, V 25

EIMAS P.D. et al (1971)
"Speech perception in infants"
Science, 1971 303-305

EDMONDSON W. (1973)
"An aid for the talking deaf" Phd. Thesis, University of London

EDWARDS A.L. (1972)
"Experimental design in Psychological Research"
Holt, Rinehart and Winston, Inc.

FITTS P.M. AND POSNER M.I. (1973)
"Human Performance"
Prentice/Hall Int., Inc., London.

FLANAGAN J.L. (1965)
"Speech analysis synthesis and perception"
Springer Verlag, Berlin.

FOWLER C.A. (1975)
"A system approach to the cerebral hemispheres"
Haskins Laboratories Status Report on Speech Research, SR-44.

FRENCH D.J. (1971)
"The Reticular Formation"
From "The Biological Basis of Behaviour"
Ed. N. Chambers et al, Open University Publ.

FRY D.B. (1974)
"Right ear advantage for speech presented monoaurally"
Language and Speech, V 17, part 2, p. 142-151.

FRY D.B. (1975)
"Simple Reaction Times to Speech and Non-Speech Stimuli"
Cortex, V XI, pp. 355-360.

FRY D.B. (1979)
"The Physics of Speech"
Cambridge Press

GABOR D. (1957)
"A summary of communication theory"
London Symp. of Communications.

GOLDMAN-EISLER F. (1972)
"Pauses, Clauses and Sentences"
Language and Speech, 15, pp. 103-113.

GREEN D.M. (1958)
"Detection of Multiple Component"
J. Acoust. Soc. of America V30, 904-911

GREENE J. (1975)
"Thinking and Language"
Methuen and Co. Ltd.

GREGORY R.L. (1968)
"The Brain"
BBC Publications

HAGGARD M. (1977)
"Do we want a theory of Speech perception?"
Research Conf. of Speech processing aids for the deaf, Gallaudet College, Washington D.C.

HAMMOND N; BARBER P. (1978)
"Evidence for abstract response codes: Ear-hand correspondence effects in three choice R.T. task".
Quarterly Journal of Exp. Psychology 30 Feb: 71-82

HICK H.E. (1952)
"On the rate of gain of information"
Quart. J. Exp. Psychol., IV

HOOD D.C. (1975)
"Evoked Cortical Response Audiometry"
From L.J. Bradford (ed) "Physiological Measures of the Audio-Vestibular System".
Academic Press Inc.

INTERNATIONAL PHON. ASSOC. (1964)
"The principles of the international phonetic association"
University College, London.

IRWIN R.J. AND PURDY S.C. (1982)
"The minimum detectable duration of auditory signals for normal and hearing impaired listeners"
J. Acoust. Soc. of America V 71. No. 4.

LIBERMAN A.M. AND PISONI D.B. (1976)
"Evidence for a special speech perceiving subsystem in the Human"
Haskins Laboratories Status Report on Speech Research, SR-48.

LIEBERMAN P. AND CRELIN E.S. (1970)
"On the speech of the Neanderthal Man"
Haskins Laboratories Status Report on Speech Research, SR-21.

LIEBERMAN P. (1974)
"The evolution of speech and Language"
Haskins Laboratories Status Report on Speech Research, SR 37/38.

LICKLIDER J.C.R. (1951)
"Basic Correlates of the Auditory Stimulus"
From "Handbook of Experimental Psychology"
J. Wiley and Sons, New York.

LINDLEY D.V. AND MILLER J.C.P. (1953)
"Cambridge Elementary Statistical Tables"
University Press, Cambridge.

LINDSAY P.H. AND NORMAN D.A. (1977)
"Human information processing"
Academic Press, Inc. London.

LURIA A. (1973)
"The working brain"
Penguin Books, London.

LURIA A. (1974)
"Scientific perspectives and philosophy dead ends in modern linguistics"
Cognition, 3(4) pp. 377-385.

MOREHEAD D.M. AND MOREHEAD A.
"From signal to sign: A Piagetion view of thought and language during the first 2 years"
Schiefelbusch R.L. and Lloyd L.L. (eds)
"Language, perspectives and acquisition".
Macmillan, New York

NATHAN P. (1969)
"The nervous system"
Harmondworth, Middlesex, Penguin.

PAPCUN G. et al. (1974)
"Is the left hemisphere specialized for speech, language and/or something else?"
J. Acoust. Soc. of America V 55. No.2.

PLUTCHIK R. (1974)
"Foundations of experimental Research"
Harper & Row, New York.

REMEZ R.E. AND RUBIN P.E. (1982)
"The stream of Speech"
Haskins Laboratories Status Report on Speech Research, SR-69.

REPP B.H. (1982)
"Categorical perception: Issues, methods findings."
Haskins Laboratories Status Report on Speech Research, SR-70.

REVZIN I. (1974)
"From animal communications to human speech."
From "Pragmatic aspects of human communication"
Ed. C. Cherry.
D. Reidel Publishing Co.

RUBIN et al (1975)
"Initial phonemes are detected faster in spoken words than in spoken non-words"
Haskins Laboratories Status Report on Speech Research, SR-44.

TEXAS INSTRUMENTS. (1978)
"Single Chip L.P.C. Speech Synthesizer and companion 131 K bit ROM."
Texas Instrument, Computer Division.

TURVEY M. AND SEARS S. (1976)
"Modes of perceiving: Abstract, comments and notes."
Haskins Laboratories Status Report on Speech Research, SR-47.

VAN DER MOLEN AND KEUSS P.G. (1979)
"The relationship between Reaction Time and Intensity in discrete auditory tasks".
Quart J. of Exp. Psychol. 31. pp. 95-102.

VYGOTSKY I.S. (1962)
"Thought and Language"
MIT Press, New York.

WADSWORTH B.J. (1971)
"Piaget's Theory of cognitive development"
D. McKay Co. Inc., New York.

WASHBURN S.L. (1978)
"The Evolution of Man"
Scientific American, Sept.

WEBER E.G. (1961)
"Theory of Hearing"
J. Wiley & Sons, New York.

WEBSTER J.C. AND CARPENTER A. (1968)
"Perceiving steady state vowel, musical and meaningless sounds"
J. Speech and Hearing Research, Sept. V11, No.3.

WESTOFF J.M. (1963)
"In search of a measure of perceptual work"
Philips Technical Review, V 25.

WOODROW H. (1951)
"Time Perception"
From "Handbook of experimental psychology".
J. Wiley & Sons, New York.

WOOLDRIDGE D.E. (1963)
"The Machinery of the Brain"
McGraw -Hill, London.

ZANGWILL O.L. (1960)
"Speech"
From "Handbook of Physiology"
American Physiological Soc. V2.