A NEW FREQUENCY TRANSPOSITION DEVICE FOR THE DEAF; A SIMULATION AND A VALIDATION STUDY.

by

Max Leopold Velmans.

A dissertation submitted in fulfilment of the requirements for the degree of Doctor of Philosophy.

Psychology Department
Bedford College
University of London

October, 1972.

Abstract

To increase the high frequency speech information available to the sensory-neural deaf, with low-frequency residual hearing only, a frequency "recoding" device was constructed which "shifts" a selected band of high frequency speech information and superimposes it on the low frequency range, in a manner designed to maintain the 'speechlike' nature of the "recoded" input signal (patent applied for).

The design and evaluation of the "recoding" device are considered in the context of factors likely to be involved in the acquisition of "recoded" speech, e.g. the separation of sounds that are 'speechlike' from those that are not, by the ear-brain system, the interaction in speech processing of auditory, visual and kinaesthetic cues, and the influence of already established strategies for processing "non-recoded" speech on the acquisition of altered strategies for "recoded" speech.

The 'speechlike' nature and the utility of the "recoding" were assessed (a) in a simulation study involving normal hearing subjects under simulated deafness conditions, and (b) in a validation study with sensory-neural deaf children.

In the simulation study significant improvements in the ability to imitate CVC nonsense syllables were brought about both by "recoding" and by visual cues (from articulatory movements) without formal discrimination or

imitation training, the "recoded" high frequency information contributing in particular to imitation of "manner" and "place" of articulation of phonemes with major energy components in the "recoded" High Frequency region (HF phonemes).

Further, in the validation study, "recoding" produced a significant improvement in the articulation learning of HF phonemes, indicating (together with the simulation study findings) that the "recoded" signals were sufficiently 'speechlike' to be of use to the ear-brain system in speech processing.

It was concluded therefore that the generality of utility (to the hearing impaired) of the "recoding" mode proposed, merits serious further investigation.

CONTENTS

pp

Section 1. Introduction 10

Speech information available to the normal hearing child. 11

Speech information available to the deaf child. 12

Alternative sources of speech information for the deaf child. 13

Frequency "transposition" or "recoding" devices. 14

Psychological assumptions governing the design of "recoding" devices. 15

Design of the frequency "recoding" device (FRED) built for the present study: 19

(a) description of the FRED 19

(b) the "recoding" process 29

Similarities and differences between the FRED and the "Johansson transposer". 35

A review of the relevant experimental literature. 37

Some psychological variables involved in validation procedures for "recoding" devices. 40

Section 2. Initial evaluation study using normal hearing subjects under simulated deafness conditions. 45

Aims of the experiment 46

Method:

(a) the simulation of deafness 47

(b) test material 50

(c) subjects and design 51

(d) procedure 52

pp

Results:

(a) imitation scores 56

(b) "distinctive feature" analysis 57

(c) decision times and required repetitions 59

Discussion:

(a) the effects of visual cues 61

(b) the effects of "recoding" 62

Section 3. A validation study with sensory-neural deaf children. 65

Aims of the experiment 66

Method:

(a) subjects 67

(b) test materials 67

(c) procedure and experimental design 69

Results:

(a) articulation learning 73

(b) transfer-of-training 80

(i) articulation transfer 80

(ii) 'discrimination' transfer 81

Discussion:

(a) articulation learning and transfer-of-training 82

(b) discrimination learning and transfer-of-training 83

(c) supplementary 'discrimination' test. 86

Summary of experimental findings. 90

Conclusions. 94

	pp
References.	96
Appendix 1. Simulation study data.	106
"Logatom" test list	107
Phonetically balanced practice list	108
Confusion matrices of "judged match" imitation responses.	109
Appendix 2. Training and test stimuli used in the validation study.	117

Enclosed Publications.

1. Velmans, M. (1971). Information theory measures of the ear/brain system - some doubts. Sound, 5, 58-61.

2. Velmans, M. (1973). Aids for deaf persons. Patent Specification No. 1340 105. British Patent Office.

3. Velmans, M. (1973). Speech imitation in simulated deafness, using visual cues and 'recoded' auditory information. Language and Speech, 16 (part 3), 224 - 236.

4. Velmans, M. (1974). The design of speech recoding devices for the deaf. British Journal of Audiology, 8, 1 - 5.

FIGURES

pp

Figure 1. Flow diagram of the transposer used in the present study. 21

Figure 2. Spectograms of the words "ship" and "sip" (a) unfiltered (b) low-pass-filtered and (c) "recoded" and low-pass-filtered. 33

Figure 3. (a) Schematic diagram of the simulated deafness situation (b) Spectrographic representation of "recoding", filtering, and white noise masking operations, performed on the input speech signal. 49

Figure 4. Articulation learning curves of six sensory-neural deaf children under (a) "recoding", selective amplification and amplitude compression, and (b) selective amplification and amplitude compression (only) conditions. 75

Figure 5. Proportional articulation learning curves for each treatment group (of deaf children) under (a) "recoding", selective amplification and amplitude compression conditions and (b) selective amplification and amplitude compression (only) conditions. 78

TABLES

pp

Table 1. Percentage correct phoneme (or phoneme cluster) imitation scores, for all subjects (in the simulation study) under all conditions. 56

Table 2. Percentage of HF-component phoneme imitations with correct "manner" and/or "place" of articulation (in the simulation study) under all conditions. 58

Table 3. Mean decision time and required repetitions per CVC stimulus (in the simulation study) under all conditions. 59

Table 4. Description of subjects (validation study). 68

Table 5. Pretest and transfer-of-training test scores (validation study). 79

Table 6. Supplementary 'discrimination' test scores (validation study). 88

PLATES

Plate 1. The frequency "recoding" device: front view, showing the instrument panel. 23

Plate 2. The frequency "recoding" device: angled view, showing the frequency generator (mounted on top) supplying an external "modulating" frequency. 25

Plate 3. The frequency "recoding" device: top view with cover removed, showing electronic circuitry, and internal power supply. 27

Plates 4 to 9. Training and test stimuli: used in the validation study (Appendix 2). 118

ACKNOWLEDGEMENTS

I wish to express my thanks to Professor Colin Cherry and to Mr. Archy White of Imperial College, London for technical advice on the construction of "frequency shifting" apparatus, to Mr. Martin of the R.N.I.D. for the use of calibration and control equipment, and to Mrs. Rosemary Westley of Bedford College, London for organising general technical assistance in particular that of Mr. David Finn who carried out the photographic work and Mr. Geoffrey Gonella who built the machine chassis and instrument panel. Further, my thanks are due to Miss Quinn of the I.L.E.A. for her assistance in organising the validation study, and in particular, to Mr. Davis of the I.L.E.A. for training the children. In addition I wish to thank Dr. John Wilding and Dr. Mary Pickersgill of Bedford College for reading various sections of this manuscript, and Mr. John Valentine of Bedford College and Mr. Michael Stone of Goldsmiths College for their valuable comments on the statistical design of both evaluation studies. Finally, I wish to express my extreme gratitude to Dr. Monica Lawlor of Bedford College for her constant support, advice and encouragement in her supervision of this study.

INTRODUCTION

Speech information available to the normal hearing child.

In the normal hearing child, the acquisition of speech involves at least three interrelated processes:
(a) auditory information, and to a lesser extent visual cues (arising from the articulatory movements of the speaker), permit the identification and discrimination of the speech sounds of the child's language community,
(b) auditory and tactual-kinaesthetic feedback allow the child to monitor his articulatory attempts to imitate those sounds, and
(c) the link thus established by auditory feedback between self produced sounds and the articulatory movements required to produce those sounds itself facilitates learning to discriminate functional segments in the speech of others, e.g. phonemes, syllables, etc. (as suggested from differing points of view by Liberman, Harris, Eimas, Lisker and Bastian, 1961; Mecham and Arbess, 1970; Whetnall and Fry, 1964).

Speech information available to the deaf child.

In the deaf child however, auditory information in the speech signal may be reduced (a), by impairment of the mechanical system involving the tympanic membrane and ossicles of the middle ear, which translate the sound energy impinging on the outer ear into a pattern of vibrations at the "oval window" of the inner ear, thereby leading to conductive deafness, and/or (b), by impairment of the structures of the inner ear or auditory nerve which convert the mechanical energy impinging on the "oval window" into a coded pattern of electrochemical impulses terminating in the auditory cortex, thereby leading to perceptive or sensory-neural deafness (see Davis and Silverman, 1970; Fraser, 1970 and Friedmann, 1970, for detailed descriptions of the aetiology and pathology of deafness).

Such a decrease in the profoundly deaf childs auditory input commonly leads to:

(a) an increased dependence on visual information in monitoring the speech of others,

(b) an increased dependence on tactual-kinaesthetic information in monitoring his own articulations, and

(c) the virtual elimination of the facilitatory interaction between speech articulation learning (by imitation) and learning to discriminate the speech of others, for the reason that the sense modalities used to monitor these are now largely different.

Alternative sources of speech information for the deaf child.

Rehabilitation techniques, therefore, either train the child to make better use of the now dominant modality, e.g. teaching him to discriminate speech solely by lip-reading or finger-spelling, or, in the "multisensory" approach, to make simultaneous use of information available to different sense modalities, e.g. by teaching a joint reliance on lipreading and any low frequency residual hearing.

Such retraining can be supplemented by electronic devices, designed to recode the speech signal into a form capable of being more efficiently utilised by the remaining input sensors. Such a recoding potentially re-establishes an articulation-discrimination interaction, as information regarding the speech of others is transformed in the same way as that relating to self-produced speech. This will apply whether the recoding be to a visual display (e.g. Børrild, 1967; Hudgins, 1935; Martony, 1967; Risberg, 1967) to a pattern of vibrotactile-signals (e.g. Geldard, 1957, 1960; Kringlebotn, 1967; Pickett, 1963) or into some acoustic form intended to make greater use of any residual hearing, (the approach followed in the present study).

The suitability of any given auditory recoding technique is however dependent on the nature of the deafness. The linear or selective amplification of specified frequency ranges, used in conventional hearing aids, compensates (in principle) for the loss of sensitivity to given frequency ranges, occurring in "conductive" deafness. As pointed out by Davis and Silverman (1970) and Pimonow (1968) such selective amplification may be quite inappropriate in a case of "sensory-neural" deafness where the neural circuits mediating frequencies within a given range are largely inoperative (as opposed to merely lowered in sensitivity). Such impairments, occurring in approximately 30% of deafness cases (Friedmann, 1970) generally involve a loss of function in sensors mediating the higher frequencies of speech, e.g. above 1000cps (see Dale, 1967 and Friedmann, 1970 for reviews of the distribution of different types of hearing impairment in the deaf population).

Frequency "transposition" or "recoding" devices.

Following a proposal generally accredited to Perwitzschy (1925), to make better use of the remaining low frequency residual hearing in such cases, various "frequency transposition" (or "frequency recoding") devices have been built, whose function is to "map" either a part, or the whole of the inaccessible speech spectrum, on to a set of correlated signals lying within the residual hearing range. This "recoded" high frequency speech energy may then be "mixed" with whatever low frequency speech energy already lies within the residual hearing range.*

* Following the classification scheme used by Risberg (1969) the various means of recoding may be broadly divided into
(a) modulation systems (e.g. Biondi and Biondi, 1968; Johansson, 1959),
(b) distortion systems (e.g. Risberg and Spens, 1967)
(c) vocoder systems (e.g. Lafon and Isaac, 1963; Ling, 1968; Ling and Druz, 1967; Pimonow, 1965) and
(d) frequency division systems (e.g. Kringlebotn, 1962; Guttman and Nelson, 1968).

While the technical details of "recoding" devices need not be discussed here, having been recently reviewed by Ling (1968) and Risberg (1969), in order to provide a context for the rationale governing the design of the transposer used in the present study, it is necessary to consider briefly the psychological assumptions which have to-date been implicit in the design of such devices.

Psychological assumptions governing the design of "recoding" devices.

"Recoding" of the types outlined above necessarily involves alterations in the input speech spectra. Accordingly the output from the different transposers have varying degrees of resemblance to 'normal' speech (see Risberg, 1969) the general assumption appearing to be, that acquisition of any given set of speech sounds (or cues) will be assisted, if for each such sound a unique discriminable correlate is provided within the residual hearing range, irrespective of whether that correlate has a 'perceptual resemblance' to the original sound.

Further, as has been argued by Roworth (1970) (following Mazeas, 1968 and Pimonow, 1968) in order to provide a recoding of speech that does not exceed the "channel capacity" of the deaf ear, it is desirable that an "information reduction" is brought about in the recoding process, by filtering out redundant elements in the original speech, and transmitting correlates of 'essential' cues only.

In assessing the utility of such an information theory approach to the design of a "transposer" it has been argued by the present author (in a paper prepared in the course of this study) that assertions about the "channel capacity" of the deaf ear, the "transmission rate" of speech, etc., affect a precision which is highly misleading in the present state of knowledge of processes underlying speech perception (see enclosed publication, Velmans, 1971a) when in fact, "... the complex interactions of the ear-brain system (and our poor understanding of them) make it difficult to isolate an objectively demonstrable property of the mechanism, to which an information measure can be made to correspond," (p.60). Such reservations do not however, detract from the argument that transposition devices should provide correlates (within the residual hearing range) of whatever speech cues are found to be essential for intelligibility.

It will be further argued however, that the providing of such correlates does not in itself ensure the utility of the "recoded" speech to the deaf. In fact, evidence against the sufficiency of such a design principle converges from a number of sources.

Most importantly, in spite of two decades of experimentation with transposition devices, the few controlled studies comparing speech acquisition either with or without "recoding" have been largely inconclusive (see Ling, 1968; Risberg, 1969). Revealingly, Liberman et al (1967) report a parallel lack of success in the devising of those 'reading' machines for the blind which operate by converting print to a "sound code"; in spite of 50 years of experimentation with various "sound codes", the best auditory 'reading' speed achieved by blind subjects, is little more than one tenth of the rate at which they can decode speech. Conversely, although speech decoding presents no difficulties, e.g. to the ear-brain system of a normal five year old child, it is extremely difficult for even well trained adults to read <u>visual</u> transforms of the speech (see Liberman et al, 1968) or to build machines that will do so (see e.g. Lindgren, 1965a, 1965b; Wathen-Dunn, 1967, for reviews).

These findings, together with considerable additional evidence reviewed by Liberman et al (1967) indicate the existence of a specialised "speech decoder" in man, possibly involving distinctive feature analysers similar to those found, e.g. by Altman (1968), Evans and Whitfield (1964) and Nelson et al (1966) in the auditory cortex of the cat (see Abbs and Sussman, 1971, for a review of the relevant evidence). Further, the superior identification of words and digits presented to both ears simultaneously by the dominant hemisphere (e.g. Kimura, 1961a, 1961b; Bryden, 1965) and the finding that aphasia following on lesions to the dominant hemisphere are more severe (see e.g. review by Geschwind, 1965) implicates the primacy of the dominant hemisphere in speech functioning. On the other hand the non-dominant hemisphere appears to be specialised in the identification of 'non-speech' sounds (e.g. Curry, 1967; Kimura, 1964; Webster and Chaney, 1967).

It may be inferred from the above evidence that the ear-brain system is predisposed to separate sounds that are 'speechlike' from those that are not, in the decoding process (although as argued by Bever (1971), and Studdert-Kennedy and Shankweiler (1970), this may occur at the level of "linguistic" as opposed to "auditory" analysis). Further there is considerable evidence (reviewed by Lenneberg, 1964, 1967) that the acquisition of speech by the child cannot be entirely accounted for in terms of schedules of rewarding stimuli reinforcing the appropriate speech responses in the child (e.g. as suggested by Skinner, 1957) but requires the postulation of a "species specific" maturation process, biologically programmed to 'extract' the rules of speech (and language) from the corpus of speech surrounding the child.

It is argued here therefore, that in order that "transposed" speech may activate those mechanisms predisposed to the acquisition of speech, the "recoded" signals produced by a "transposer" must be not only discriminable, but also accepted by the ear-brain system of the deaf person as 'speechlike' (a conclusion also arrived at by Roworth, 1970).

To achieve this end however, more research is required, (a) on the nature of cues deaf persons can use in making auditory discriminations (as pointed out by Ling, 1968) and (b) on the nature of cues deaf persons can use to distinguish between speech and 'non-speech'. In the absence of such research* the utility of any given auditory recoding must ultimately be examined post hoc, by assessing the ability of the deaf to utilise the "recoding" in question, in properly designed validation studies (validation procedures are discussed in a later section).

Design of the frequency recoding device built for the present study.

In the present study the Frequency REcoding Device built (henceforth referred to as FRED) was adjusted to maintain the 'speechlike' nature of the input signal by minimising the distortions performed on it in the recoding process. A brief description only of (a) the construction of the FRED and (b) the mode of recoding used, is given below, further details and options being outlined in the enclosed patent specification (Velmans, 1971b).

(a) Construction of the FRED

The prototype FRED is illustrated in Plates 1, 2 and 3. A flow diagram of the essential operations performed by the FRED is given in Figure 1.

*The investigation of the discriminability of various types of synthetic speech stimuli to sensory-neural deaf listeners, e.g. filtered white noise signals (Pickett and Martin, 1968) synthetic formant transitions (Pickett and Martin, 1970) synthetic vowel formants (Pickett and Martony, 1970) and their masking effects (Martin and Pickett 1970a, 1970b), may however provide useful beginnings in this direction.

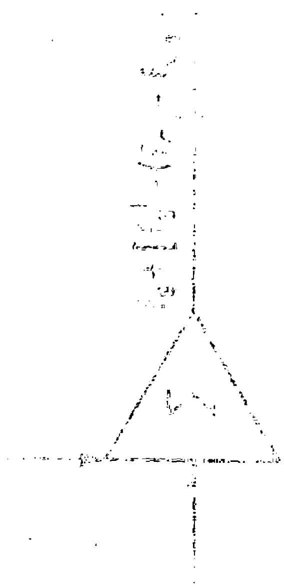


FIGURE 1. Flow diagram of the transposer used in the present study, where fs is the input signal, fs' is a selected high frequency band of the input signal, fc_1 is the modulating frequency, fc_2 is the demodulating frequency, $(fc_1 - fc_2)$ is the frequency "shift" performed on each frequency component of fs', and $fs + (fs' - (fc_1 - fc_2))$ is the output signal.

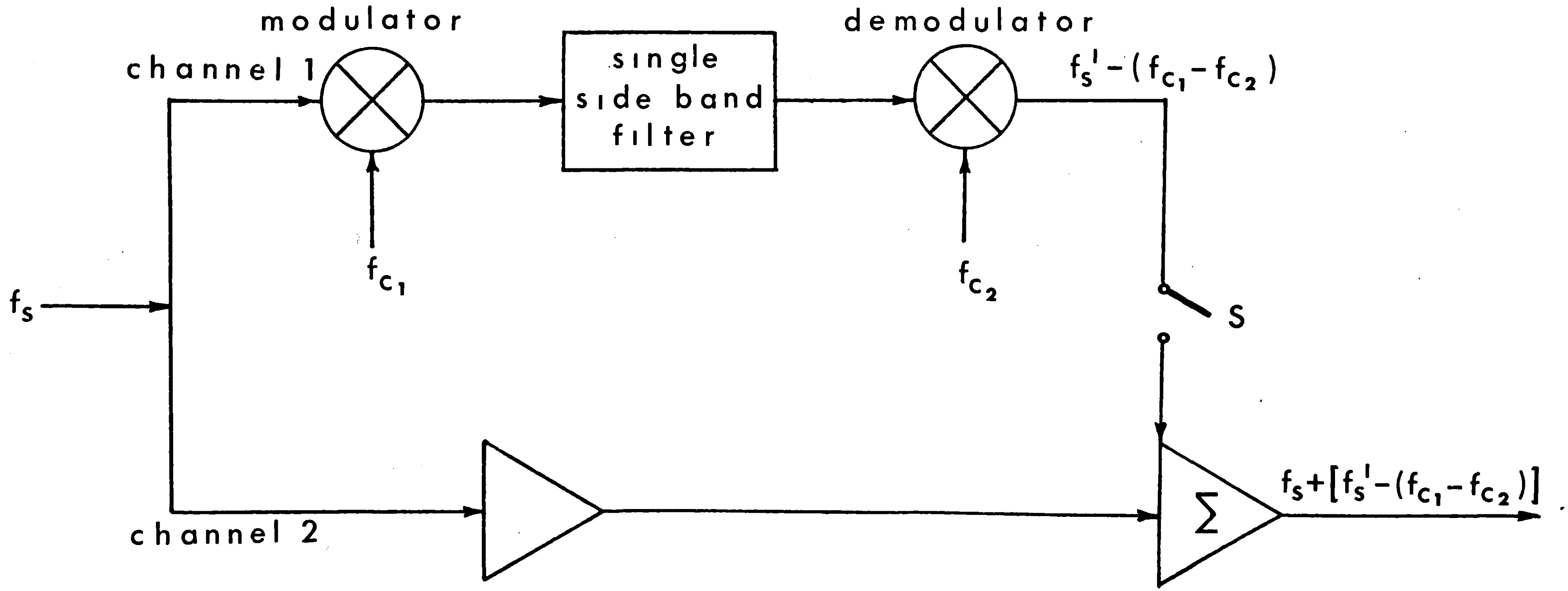

modulator
channel 1
single side band filter
demodulator
$f_s' - (f_{c_1} - f_{c_2})$
f_{c_1}
f_{c_2}
f_s
S
channel 2
Σ
$f_s + [f_s' - (f_{c_1} - f_{c_2})]$

Plate 1. The frequency "recoding" device: front view, showing the instrumental panel.

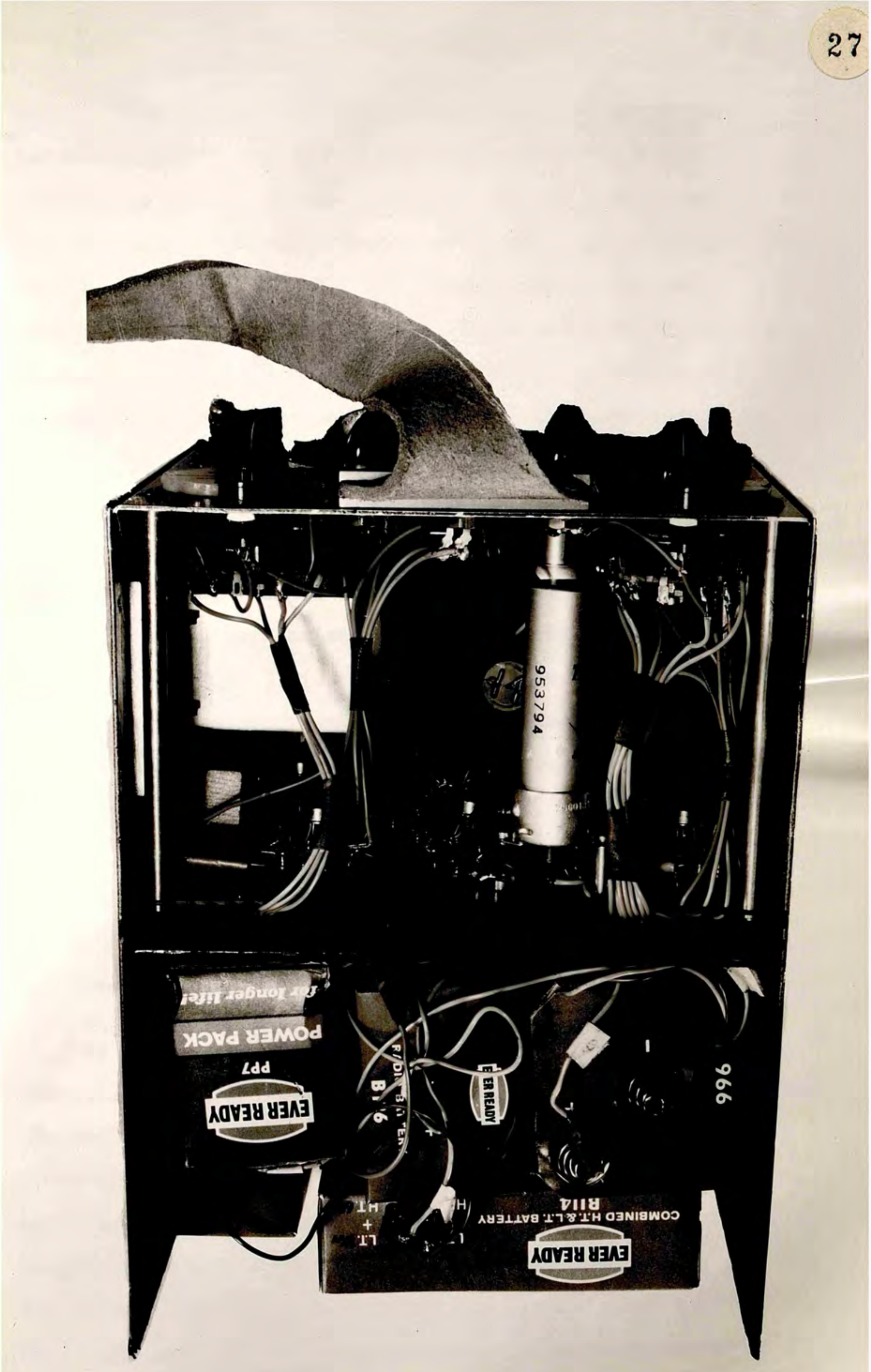
23


INSTRUMENT
OFF
ON
GAIN dB
43
18
38
33
28
23
MIKE
OFF
ON
ATTEN(D) dB
-17·5
-10
-5
-27
-5
-25
-15
-10
OFF
ON
ON
MIKE.1
MIKE.2
INPUT
TAPE
INPUT
OUTPUT
OSC.1
OSC.2

Plate 2. The frequency "recoding" device: angled view, showing the frequency generator (mounted on top) supplying an external "modulating" frequency into channel 1, via "osc 1".


MIKE
OFF
ON
GAIN dB
ATTENUATE dB
INSTRUMENT

Plate 3. The frequency "recoding" device:
top view with cover removed, showing
electronic circuitry, and internal power supply.


953794
for longer life
POWER PACK
PP7
EVER READY
EVER READY
COMBINED H.T.&L.T. BATTERY
B114
EVER READY
996

As can be seen in Plate 1, the top row of the instrument panel consists of an on-off switch for the instrument, an overall gain control, and an independent microphone on-off switch connected to an internal power supply (shown in Plate 3). The middle row of the panel consists of an on-off switch and attenuator for a "recoding" channel (channel 1 in Figure 1) as well as an independent on-off switch and attenuator for a linearly amplified channel (channel 2 in Figure 1). The bottom row of the panel has two microphone input sockets (allowing a given subject and experimenter to have separate microphones) a tape or signal generator input socket, and the FRED output socket (e.g. to tape or earphones). In addition, the socket marked "Osc 1" (Oscillator 1) allows an external "modulating" frequency and the socket marked "Osc 2" allows an external "demodulating" frequency to be supplied to the "recoding" channel(as shown in channel 1, Figure 1); an input to either "Osc" socket automatically bypasses a corresponding (but fixed) modulating or demodulating frequency which is otherwise automatically supplied to the recoding channel by an internal crystal oscillator.

As described (in "The recoding process") below, by suitably adjusting the modulating and demodulating frequencies, the machine can select any band of frequencies with a bandwidth of 4kc (the pass band of the single side band filter shown in Figure 1) within the frequency range 0 cps to 40,000 cps (the range of input frequencies provided by the B & K condenser microphones used) and "shift" that band to any other part of the range (although in auditory research this must be within the frequency response of the output transducer, e.g. the earphones). If desired the "shifted" signal may be "folded" or "mirror imaged" (a technique used in the Johansson transposer described below, but not in the present study).

The frequency response (FR) of the linearly amplified channel was limited only by the output transducer (e.g. earphones) whereas the FR of the "recoding" channel was limited by the "single-side-band" filter (see Figure 1). Both channels were however, substantially "flat" (± 3dB) in the residual hearing range of concern in the present study (i.e. 60 cps to 1000 cps).

(b) The "recoding" process

As can be seen from the flow diagram of the FRED (Figure 1) the input signal (f_s) is fed to two parallel circuit paths (channel 1 and 2). In channel 2 the signal is linearly amplified; in channel 1, using a modification of conventional radio engineering practice, the "modulating" (or "carrier") frequency (fc_1) is selected so that the following "single-side-band filter" passes only a selected band (fs^1) of the original (but now modulated) input signal (fs). The selected signal is then "demodulated" (from radio to audio frequencies) with a demodulating frequency (fc_2).

By setting fc_2 at a different frequency to fc_1 a "frequency shift" is performed on fs^1, producing an output from channel 1 of $fs^1 - \Delta$, where Δ is a constant frequency substracted from each frequency component (fs_a^1, fs_b^1, ...etc.)

within the selected band of frequencies (fs^1), and where $\Delta = fc_1 - fc_2$. The "shifted" signal is then "mixed" with the original linearly amplified signal (fs) in channel 2 (provided that both channels are switched "on") to produce an output signal of $fs + (fs^1 - \Delta)$.

In this way a given high frequency band of the input speech spectrum (in channel 1) can be (a) separated from the lower frequency components (b) lowered in frequency (by an amount Δ) so as to lie within the residual hearing range of the deaf, and (c) "mixed" with the linearly amplified but otherwise unaltered input speech signal (on channel 2).

It should be emphasised that unless the selected high frequency band fs^1 is shifted to the 'negative' frequency region (causing "folding" or "mirror imaging") the only spectral properties to be altered by the "frequency shift", are the "frequency ratios" (e.g. $\frac{fs_a^1}{fs_b^1}$) of the spectral components. All other spectral interrelations, e.g. the "frequency differences" ($fs_a^1 - fs_b^1$) remain undisturbed (ignoring minor distortions produced by any such device).

It follows that to the extent that (a) the "frequency ratios" are not essential to the 'speechlike' nature of the selected high frequency band, and (b) interference between the "shifted" signal (on channel 1) and the amplified signal (on channel 2) is avoided, the output from the device will retain the 'speechlike' nature of the input.

On the basis of a series of informal investigations of the effects of different modes of "recoding" on phoneme intelligibility, it was hypothesised that condition (a) above, would be satisfied for those "noisy" High Frequency fricative, sibilant and stop consonant components in the region 4000 cps and above (henceforth termed HF components) and further, that if this region (having minimal vowel energy) were "shifted" down and "mixed" with the region 0 cps and above (setting Δ equal to 4000 cps) that HF component information in the latter region would be increased without interfering with the predominantly vowel information in the low frequency band, thereby satisfying condition (b); i.e. given that conditions (a) and (b) were met it was envisaged that such a "recoding" could in principle compensate for the typical loss (in sensory-neural deafness) of high frequency fricative, sibilant and stop consonant cues (essential for speech intelligibility) without interfering with or 'distorting' speech information already within the residual hearing range.

This particular mode of "recoding" is shown (in the context of a simulated deafness situation) in Figure 3(b), p 49. In addition, the spectograms in Figure 2 illustrate the effect on the words "ship" and "sip" of (a) simple linear amplification, (b) amplification followed by "low-pass-filtering" with the cut-off frequency of the filter set at 1000 cps (to simulate the loss of HF component information in a deaf ear with negligible hearing above 1000 cps) and (c) "recoding" (as proposed above) followed by "low-pass-

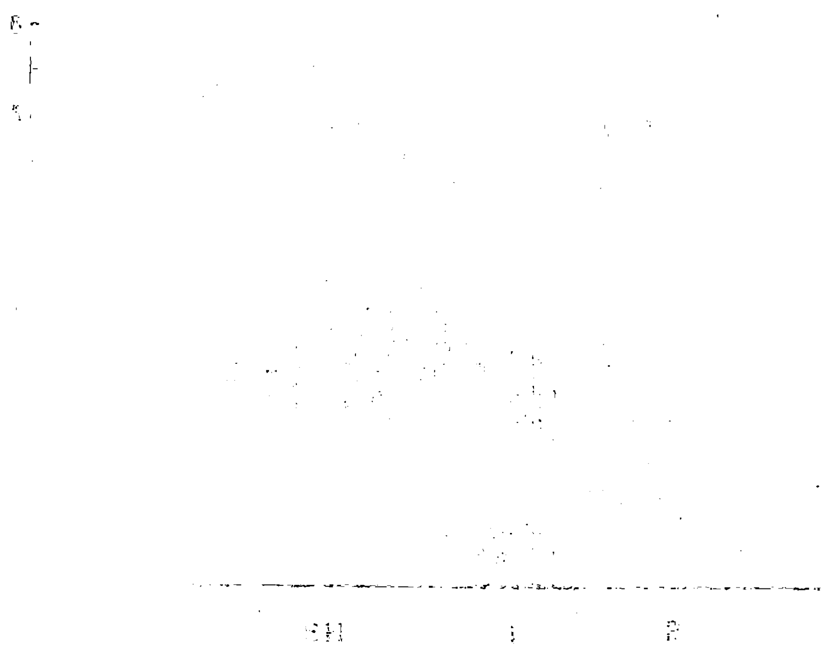
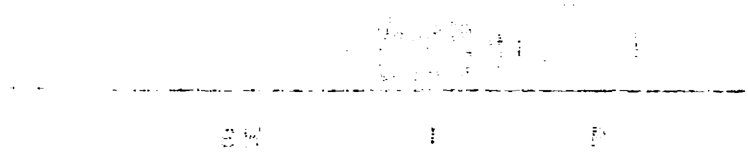
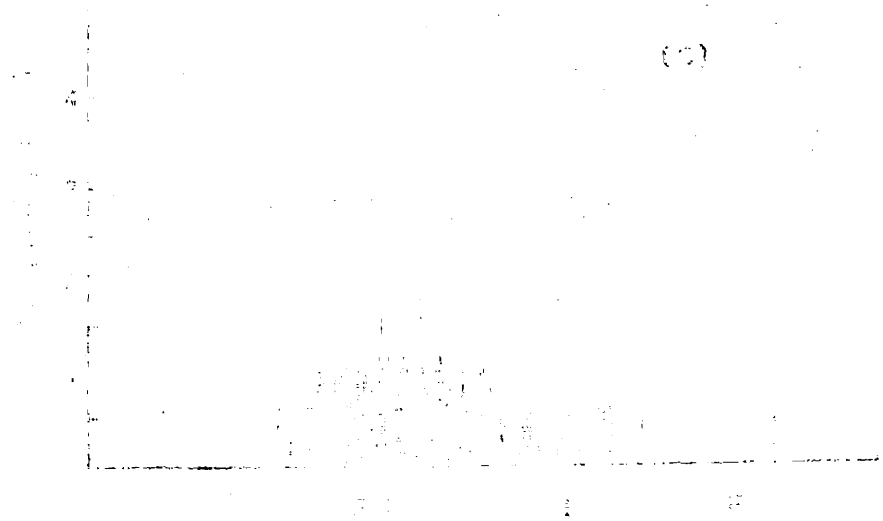
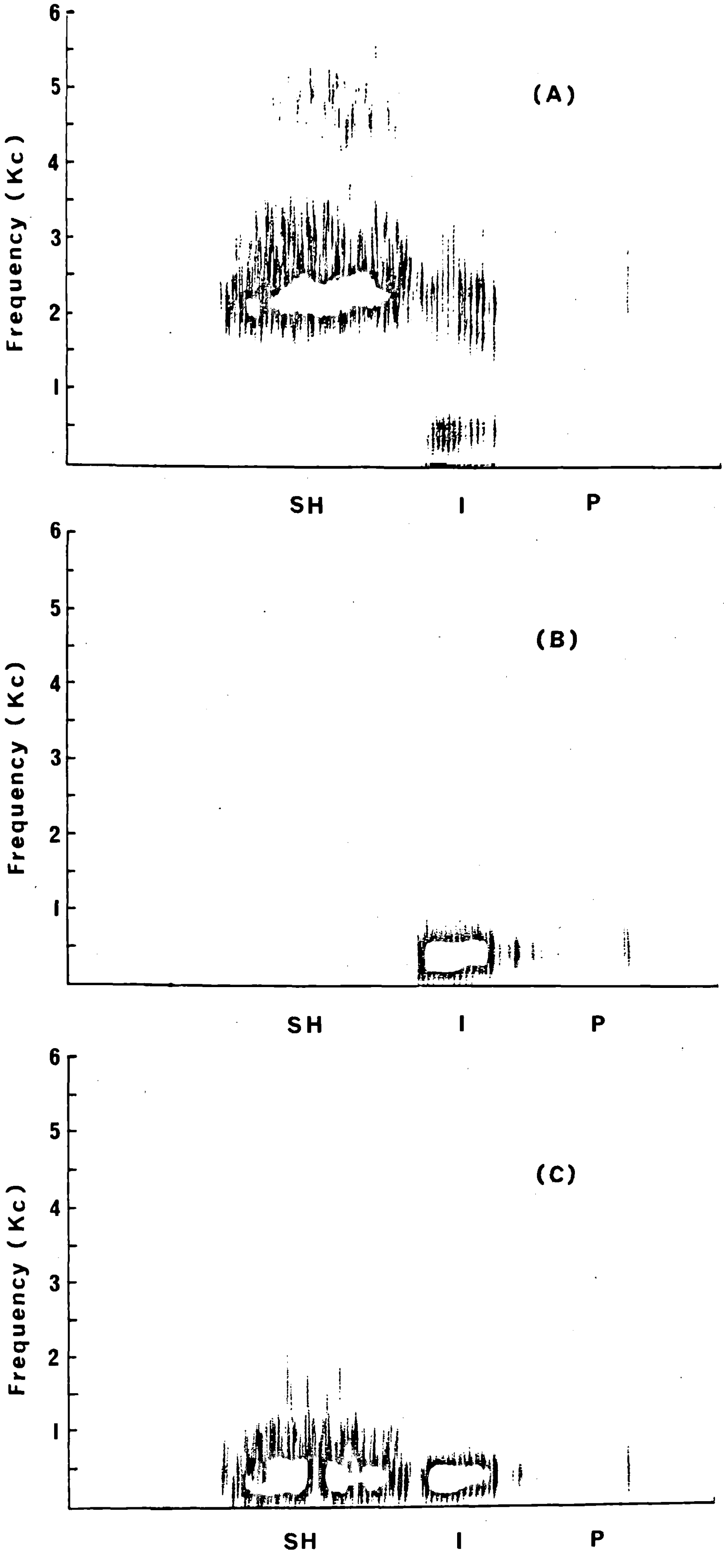


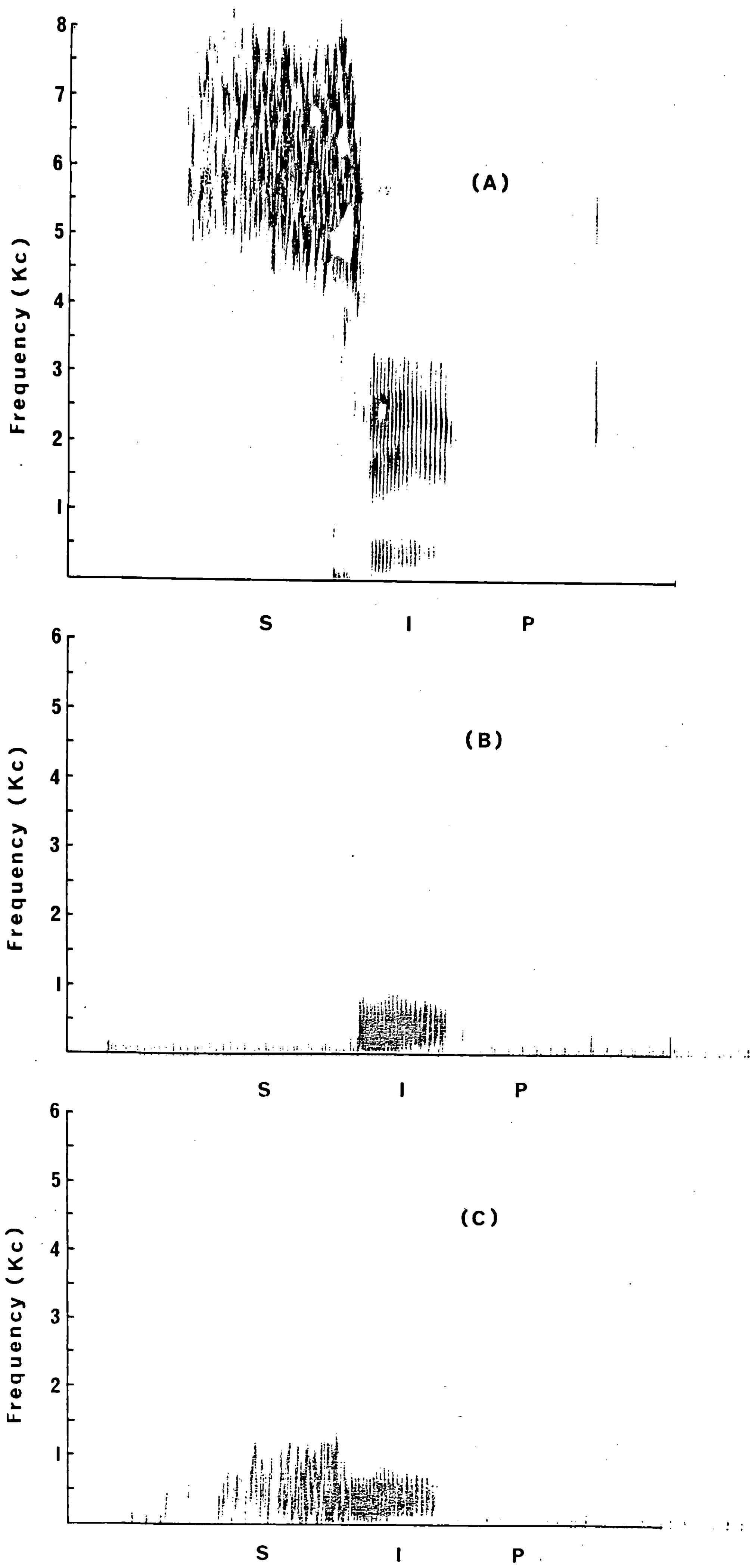
FIGURE 2. Spectograms of the words "ship" and "sip" (a) linearly amplified, (b) linearly amplified and then "low-pass-filtered", with filter cut-off frequency set at 1000 cps and (c) linearly amplified and "recoded", before low-pass-filtering.







6
5
4
3
2
1
Frequency (Kc)
(A)
SH
I
P
6
5
4
3
2
1
Frequency (Kc)
(B)
SH
I
P
6
5
4
3
2
1
Frequency (Kc)
(C)
SH
I
P


(A)
Frequency (Kc)
8
7
6
5
4
3
2
1
S
I
P
(B)
Frequency (Kc)
6
5
4
3
2
1
S
I
P
(C)
Frequency (Kc)
6
5
4
3
2
1
S
I
P

filtering", to demonstrate the restoration of HF component information (in particular that relating to the phonemes /s/ and /ʃ/ in the range below 1000 cps (and in principle, now available to the deaf ear)).

Similarities and differences between the FRED and the "Johansson transposer"

The "recoding" proposed above is similar in many respects to that employed by the Johansson transposer (Johansson, 1955,1959). In particular the later version of the device (see Risberg, 1969) which "recodes" the 4000 cps to 6500 cps band, superimposing it on the linearly amplified low frequency region below 1500 cps. The major difference in principle between the devices is in the means of recoding the selected high frequency band. The Johansson transposer 'folds' the high frequency signal into the low frequency range (see Wedenberg, 1959, for a description) mirror imaging the 4000 cps to 5000 cps band (so that a 5000 cps tone now correlates with a 0 cps tone, and a 4000 cps tone with a 1000 cps tone) and superimposing it on the "recoded" 5000 cps to 6500 cps band (where a 6000 cps tone now also correlates with a 1000 cps tone) before "mixing" with the linearly amplified speech band below 1500 cps. Such 'folding' allows a wider band of high frequency information to be "compressed" within the range below 1500 cps, but destroys spectral interrelationships such as the "frequency differences" and "frequency ratios", as well as causing the central portion of the "recoded" high frequency band to be removed by the low frequency "cut-off" of the earpiece into which the "recoded" signal is fed, e.g. an earpiece with a low frequency cut-off of 300 cps will remove the "recoded" 4700 cps to 5300 cps band.

The mode of operation proposed for the FRED on the other hand, merely "shifts" the high frequency signal, without "mirror imaging" or "compression", thereby preserving all spectral interrelations, apart from the "frequency ratios" (see Hopner and Andrews, 1957 for a description of both techniques, as possible ways of improving high-frequency consonant reproduction, in narrow-band telecommunication systems).

As pointed out by Risberg (1965), for fricative and stop consonants, the information regarding "manner" of articulation is likely to be transmitted by the duration and timing of the noise components (the frequency spectrum being unimportant) whereas the "place" of articulation is related to the frequency spectrum of both the noise components, and the formant transitions. It seems plausible to argue therefore that whereas cues relating to "manner" of articulation may be provided by both techniques, the preservation of spectral interrelations in the mode of operation proposed for the FRED, may provide added information regarding "place" of articulation (this point is examined again in the discussion following the "simulation" study).

A review of the relevant experimental literature

Unfortunately the variety of validation procedures used (rather than the variety of transposition techniques) makes the relative merits of the various means of "recoding" proposed, difficult to assess (on the basis of the reported literature). In particular, evaluation studies previous to 1968 in the main failed to provide adequate control groups, making it impossible, (a) to determine whether any improved performance after training with a given device was due to the additional "recoded" information or to the training itself, and (b) to establish any advantage of "recoding" over conventional amplification techniques.

It is worth noting however that of the controlled studies carried out with sensory neural deaf children few workers report a significant advantage for any transposition technique. If we ignore the single positive finding of those experiments using tape slowed material (reported by Bennett and Byers, 1967) on the grounds that the technique is not in any case suitable for a hearing aid (involving an unacceptable delay between the occurrence of, and the perception of speech) only Guttman, Levitt and Bellefleur (1970) report a small but significant advantage for transposition; using an instrument that produces surrogate low frequency fricative sounds by dividing the rising "zero crossings" of high frequency fricative sounds, they found a slight advantage in the articulation of /s/ and a "substantial" advantage in the articulation of /ʃ/

(compared to controls) after subjects had been given approximately 70 sessions of individual training (20 minutes per session).

A number of uncontrolled studies have also reported positive results using the "Johansson transposer" (described earlier). As the functioning of this device is similar (but not identical) to the mode of "recoding" used in the present study (see p 35) these findings will be considered in some detail.

Improvements in the discrimination of fricative sounds by sensory-neurals after training with the Johansson transposer have been reported in uncontrolled studies by Wedenberg (1959), Johansson and Sjögren (1965) and Johansson (1966). A controlled study by Ling (1968) however, produced no significant difference in the discrimination of "recoded" consonants and linearly amplified consonants after training with the Johansson transposer.

At first glance, the study of Ling (1968) indicates that whereas improved discrimination may occur when training is given with the Johansson transposer (as reported in other studies) this improvement is not significantly different from that brought about by conventional amplification (in fact, it may be the "linearly amplified" channel of the transposer (see p 35) rather than the "recoded" high frequency signal which is providing the useful information).

However, Wedenberg (1959), Johansson and Sjögren (1965) and Johansson (1966) claimed improved discrimination only of those consonants with major energy components in the "transposed" high frequency region (e.g. the fricatives) whereas of the three tests used by Ling (1968) in studies relating to the Johansson transposer, two measured vowel discrimination, while the third, a consonant discrimination test, included many consonants, which, having no major energy components in the "transposed" region, would be expected to be unaffected (beneficially) by "recoding". As there was no separation of scores on consonants predicted to improve, from consonants not predicted to improve, in the analysis of results, the reported findings remain inconclusive.

The position is not clarified unfortunately by the controlled studies with the same device using normal hearing subjects under "simulated deafness" conditions, i.e. whereas Risberg (1965) found that even without training, "recoding" with the Johansson transposer improved the identification of "manner", but not "place" of articulation, of fricative sounds (seeming to support the positive findings above), Amcoff (1971) found no significant difference, after discrimination training, between "recoded" and "low-pass-filtered" consonants (seeming to support the findings of Ling, 1968).

The utility of the Johansson transposer therefore remains and open question. In fact the main conclusion to be drawn from such conflicting results (with the same "recoding" technique) is that the validation procedures themselves must become a topic for further discussion and research. Accordingly, before proceeding to the evaluation studies relating to the present device (the FRED) some of the factors governing the rationale of the validation procedures used will be outlined (in the following section).

Some psychological variables involved in validation procedures for "recoding" devices.

The controlled validation studies reported to-date (e.g. those referred to in the previous section) all share to some extent the following design features:
(i) The "recoded" and/or linearly amplified speech is presented to deaf subjects either on tape or by means of a teaching machine. This has the advantage of standardising the test material over both experimental groups and training sessions. However, it is pointed out by Krug (1960) that auditory information and visual information provided by the articulatory movements of the speaker, interact in ways that may be non-additive, the mode of interaction varying with the nature of the hearing impairment. Further, Siegenthaler and Gruber (1969) review evidence that in sensory-neural deafness, the effectiveness of each sense modality is in fact heightened by the use of the other.

It is argued here therefore, that any test program designed to evaluate the utility of a given auditory "recoding" should include a condition where the speech stimuli to be acquired are presented in a face-to-face situation, to allow any facilitory visual-auditory interaction to occur.

(ii) The intelligibility of either "recoded" or linearly amplified material is assessed in training or test sessions either by use of a forced-choice response (e.g. pointing, or pressing one of a series of buttons on a teaching machine) or by requiring the subject to write down what he hears, after which appropriate error-feedback is provided.

However, as argued earlier (p11) in the acquisition of speech, not only are auditory and visual cues used to monitor the speech of others, and auditory and kinaesthetic cues used to monitor self produced speech, but there is also an interaction, made possible by the auditory feedback involved in each function, between learning to articulate speech and learning to discriminate the speech of others, which may facilitate both functions. This interaction is ignored in such evaluation procedures, e.g. as pointed out by Ling (1969) "In no study reported in the literature pertaining to transposition has articulation training been specifically used to enhance auditory discrimination". (p.301)

One indication of the potency of the articulatory-auditory interaction is the finding of Ling and Maretic (1971) that deaf children given combined articulation and discrimination training on "recoded" stimuli (produced by a "vocoder" technique) have the ability to generalise their discrimination learning to the same stimuli presented under "non-recoded" conditions, in a post-training test (and vice versa). Such transfer of training does not take place when discrimination training only is given (e.g. as found by Ling and Druz, 1967; Ling, 1968).

It is further argued therefore, that any test program designed to evaluate the utility of a given recoding technique must at some time provide opportunity for joint articulation learning and discrimination learning to take place, thereby allowing any facilitatory articulation-discrimination interaction to occur.

The controlled "simulated deafness" studies follow a similar pattern to the above with the exception that the "recoded" or linearly amplified speech material is
(a) passed through a "low-pass-filter" with a frequency "characteristic" resembling that of a typical pure tone audiogram of the deaf population under consideration, and
(b), in general "mixed" with white noise to further decrease the "signal-to-noise ratio" and mask any remaining high frequency components (but see the suggested modification in the use of white noise on p 47).

Such simulation studies have value in providing some prognosis as to the suitability of a given auditory recoding, e.g. as suggested later in the present study (p 46) it may be possible to establish whether a given recoding is 'speechlike' or not. In addition, a comparison of the results of such studies with the results of corresponding validation studies with the deaf may be valuable in further defining the differences and similarities in the speech-hearing functioning of deaf and normal hearing subjects (see e.g. the discussion on pages 62, 84, 94).

Any assumption of functional equivalence of simulated deafness to actual deafness must however be treated with caution. Evidence for such an equivalence is reviewed by Pickett, et al (1970) indicating that, "the phonetic features used by deaf listeners to perceive and discriminate speech sounds are the same as those used by normal listeners".

However normal hearing subjects and e.g. congenitally deaf subjects, differ (by definition) in the linguistic (particularly the phonemic) knowledge they bring to any training situation involving "recoding". It is likely for instance that in studies involving discrimination training of "recoded" phonemes over a series of training sessions, that the possession by normal hearing subjects of well established discrimination-identification strategies (for the "non-recoded" test phonemes) is likely to interfere both within and (in particular) between training sessions, with the acquisition of altered discrimination-identification strategies for the "recoded" phonemes, tending to favour any stimulus presentation which resembles the already established "non-recoded" pattern, e.g. as provided by linear amplification (even if under simulated deafness conditions such linear amplification provides less "information" than "recoding"). Such interference from well established phonemic discrimination-identification strategies will be largely absent in the learning of the deaf child, who will therefore be free to develop new strategies (but see p 84 for a discussion of an analogous interference of well established "word labels" for 'known' concepts, with the formation of new "recoded" word labels for those concepts in deaf children).

It can be argued that to minimise the potential interference referred to above (in simulation studies) the experimental task should investigate whether the normal hearing subject can in <u>principle</u> attach his 'old' identification responses to the 'new' "recoded" stimuli, by altering his discrimination-identification strategies within an experimental session, rather than whether the subject can 'build-up' such altered discrimination-identification strategies over a series of training sessions. If "recoding" can be shown to be <u>in principle</u> of use to normal hearing subjects, it may <u>in practice</u> be of use to deaf children also.

An initial evaluation study with the FRED (using normal hearing subjects under simulated deafness conditions) designed to accommodate the considerations outlined above, is reported in the following section.

Initial evaluation study using normal hearing subjects under simulated deafness conditions.

Aims of the experiment.

The experiment reported below was intended to provide an initial evaluation of (a) the utility of the speech "recoding" (proposed on p 29) and (b) the role played by visual cues (e.g. in interaction with such a "recoding") in assisting untrained normal hearing subjects under simulated deafness conditions, to imitate the articulations of an experimenter. As the experiment involved no formal discrimination training, the utility of the "recoding" was intended to provide a test of the hypotheses (outlined on p 30) that the "recoded" HF components (a) would be 'speechlike' and (b) would not interfere with the predominantely vowel information in the low frequency band (with which the "recoded" components are "mixed").

The experiment was designed to simulate both "sensory-neural" deafness, and a 'spontaneous' speech imitation situation as closely as possible, and in accordance with the arguments outlined in the previous section (p 40) opportunity for the occurrence of
(a) any articulation - discrimination interaction and
(b) any visual - auditory interaction, was provided.

Specifically, it was predicted,
(a) that the simulated deafness conditions (by removing the HF - components) would make phonemes or phoneme clusters containing a major HF - component, more difficult to imitate than those containing no major HF - component.

(b) that imitation of phonemes or clusters containing a major HF - component would be improved, by "shifting" the HF - component information, down to the low frequency region (where it would avoid elimination by the simulated deafness conditions), and
(c) that imitation of phonemes and clusters both with and without HF - components would be improved, if the experimenter's face were visible to the subject (by virtue of the visual cues provided by the experimenter's articulatory movements).

METHOD

The simulation of deafness

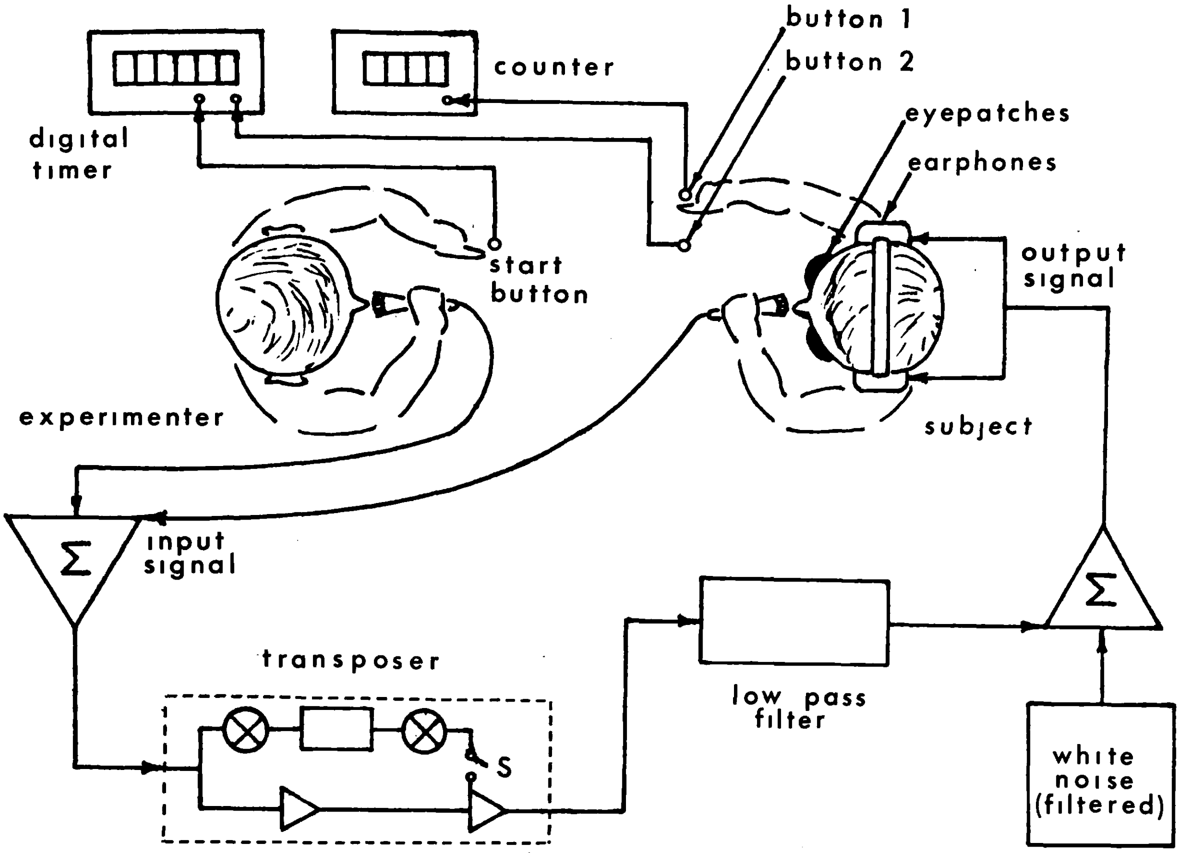
The simulation of deafness was accomplished as illustrated in Figure 3 (a).

The subject was seated facing the experimenter across a table 0.6 metres wide. The speech of both experimenter and subject were passed through B & K 4132 condenser microphones, mixed and amplified, before being processed by the transposer.

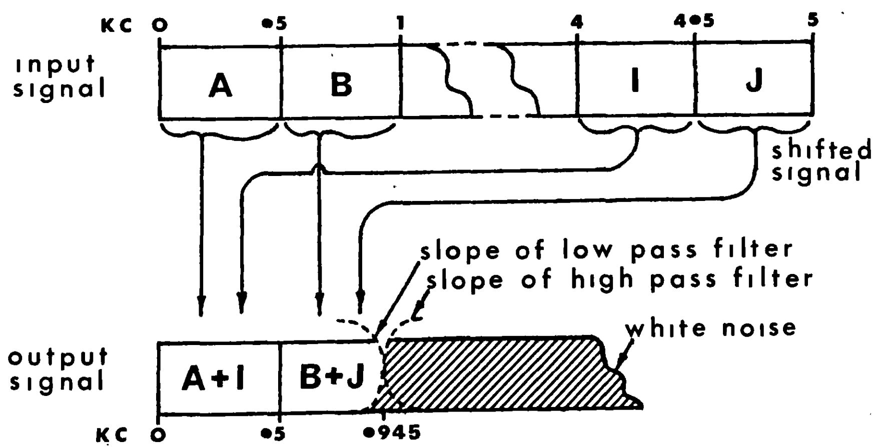
The processed speech signal was then passed through a Mullard GSS 001/01, variable "low pass" filter, with cut-off frequency set at 900 cps (to simulate the pure-tone hearing loss of a sensory-neural, partially-hearing child). The filtered signal was then mixed with preprocessed white noise (at $\frac{S}{N} \triangleq 15$ dB), before being fed to a pair of Sharpe HA - 10A earphones, worn by the subject.

FIGURE 3. (a) Schematic diagram of the simulated deafness situation.

(b) Spectographic representation of "recoding", filtering, and white noise masking operations, performed on the input speech signal.


button 1
button 2
counter
digital timer
eyepatches
earphones
start button
output signal
experimenter
subject
input signal
transposer
S
low pass filter
white noise (filtered)

(A)


KC 0
•5
1
4
4•5
5
input signal
A
B
I
J
shifted signal
slope of low pass filter
slope of high pass filter
white noise
output signal
A+I
B+J
KC 0
•5
•945

(B)

The white noise had been previously passed through a Mullard GSS 001/02, variable "high pass" filter, with cut-off frequency set at 990 cps, and had the function of masking any high frequency components of either the experimenter's speech or of the subjects own speech, which was:

(a) not eliminated by the "low pass" filter,
(b) transmitted by air conduction (through the sealed earphone cups), or
(c) transmitted by bone conduction.

The white noise components below 990 cps had been removed, to avoid interference with both the residual low frequency information, and the "recoded" HF - component information, passed by the "low-pass" filter. The various operations performed on the speech signal are shown in Figure 3 (b).

Test Material

The test material had to satisfy the contrary requirements of being both representative of English speech (the utility of the device being assessed for English language speakers), and being meaningless, to orient subjects away from guessing behaviour, and towards a dependence on the auditory and visual input in guiding their articulatory imitations (as would be the case with a deaf child, acquiring speech).

Accordingly a list of the most frequently occurring CVC nonsense syllables (drawn from a sample of English telephone conversation), was generated from the "logatom" tables of Richards and Berry (1952), by

(a) selecting the 21 initial consonants or consonant clusters occurring with relative frequency > 1% (which accounted for 90.8% of all initial consonants),

(b) selecting the vowel most frequenctly combined with each respective initial consonant, and

(c) selecting the final consonant or consonant cluster, most frequently combined with each respective vowel (or, if the resultant CVC combination were meaningful, the next most frequently combined final consonant, etc.)

The test stimuli are listed in Appendix 1 (p 107).

Subjects and Design

16 normal - hearing psychology undergraduates and staff, were randomly divided into two groups of eight subjects each.

Both groups were placed in the simulated deafness situation, and required to perform an imitation task (described in the Procedure), under both visual and non-visual conditions, in a balanced design. The experimental group however, received "recoded" HF - component information under all conditions, whereas the control group received no "recoded" HF - component information. The 21 CVC stimuli were presented in a different random order, for each subject, under each condition.

Procedure

Each subject was informed that wearing the earphones would produce simulated deafness, and that he should familiarise himself with how words 'sounded' under this condition by reading out the words from the phonetically balanced list supplied into his microphone (see Appendix 1, p 108). Each subject was also encouraged to alter the gain control on the recoding channel of the transposer, to "note any effects", in order to direct his attention to the "recoded" HF - components (although with the control group, the gain control had no effects). After familiarisation both channels on the transposer were set to the same 'comfortable hearing' level.

Each subject was then given the following written instructions:

"In the following task you must try to imitate the <u>nonsense</u> syllables I am going to say to you.

<u>You may say each syllable to yourself as often as you like</u> (into your microphone).

If you wish <u>me</u> to repeat the syllable press BUTTON 1*

When you consider that the sound you are making is the same as the sound I produce, press BUTTON 2*.

* See Figure 3a.

I will then say the syllable again, after which you must repeat it for the last time. The first trial will be a practice trial. Indicate when you have understood the instructions and are ready to begin."

Then, on each trial, the experimenter pressed a Venner timer "start" button*, before presenting the CVC syllable. The pressing of BUTTON 2 by the subject, stopped the timer, thereby giving the total 'decision time' to make a 'judged match' response, to each CVC stimulus.

The pressing of BUTTON 1 triggered a digital counter, which therefore recorded the number of repetitions required by the subject, of each CVC stimulus.

When the subject had decided that his own articulations were the same as those of the experimenter (by pressing BUTTON 2), the latter then, (a) tape recorded both his own final presentation of the stimulus and the subject's final 'judged match' response, (b) made an immediate phonetic transcription of the subject's response, thereby taking account of the visual cues provided by the subject's articulatory movements (to be cross-checked against the taped response), and

* See Figure 3a

(c) recorded the 'decision time', and the 'required repetitions' of each stimulus, (to obtain a measure of the 'subjective difficulty' of obtaining a 'judged match').

The visual condition differed from the non-visual condition, only in that in the latter, the subject wore eyepatches*.

The "recoded" HF - component condition differed from the "non-recoded" condition, only in that in the former, switch S (on the transposer)*, was closed.

* See Figure 3a

It was important:

(a) to the simulation of a 'spontaneous' speech imitation situation, that the imitation procedure was largely under the subject's own control, allowing him to adapt the auditory and/or visual information to his own learning strategies, and

(b) to the evaluation of the 'speechlike' nature of the recoded signal, that no information regarding the correctness or incorrectness of an imitation attempt, was supplied by the experimenter (as discrimination training might be expected to produce some improvement even with signals which were not 'speechlike').

RESULTS

Imitation Scores

Each final imitation response (constituting the subject's 'judged match' to the CVC stimulus) was scored for the correctness of initial consonant (or consonant cluster), middle vowel, and final consonant (or consonant cluster) respectively.

To assess the predicted effects of (a) the simulated deafness conditions, and (b) the recoded high frequency information, on the imitation of phonemes (or clusters) containing major HF - components, the imitation scores for these (identified from tables provided by Fletcher, 1958), were separated from the imitation scores for phonemes (or clusters) containing no major HF - components. The percentage of correct imitations (for all subjects) of HF - component phonemes (or clusters), and of no - HF - component phonemes (or clusters) are given in Table 1.

TABLE 1.

Percentage correct phoneme (or phoneme cluster) imitation scores, for all subjects under all conditions.

Condition	Non-Visual		Visual	
	HF - Component	no - HF - component	HF - component	no - HF - component
No Recoding	32.9	51.9	66.5	86.3
Recoding	52.3	59.5	79.6	89.0

A split-plot three-way analysis of variance was carried out on the results. As predicted:-

(a) the elimination of high frequency information by the simulated deafness conditions, made the imitation of phonemes or clusters with major HF - components more difficult than those with no major HF - components ($\underline{F}$ = 20.466, $\underline{d.f.}$ = 1,14, $\underline{p}$ < .0005),

(b) the reinsertion of HF - component information into the low frequency band (by the FRED), improved overall imitation scores ($\underline{F}$ = 5.791, $\underline{d.f.}$ = 1,14, $\underline{p}$ < .025), with a significantly greater improvement taking place for HF - component phonemes, than for no - HF - component phonemes (for the interaction, $\underline{F}$ = 3.2, $\underline{d.f.}$ = 1,14, $\underline{p}$ < .05), and

(c) allowing subjects to see the experimenters face, brought about a highly significant improvement in imitation performance ($\underline{F}$ = 137.788, $\underline{d.f.}$ = 1,14, $\underline{p}$ << .0005).

Apart from the interaction between the effects of recoding, and whether a phoneme (or cluster) had a major HF - component (mentioned in (b), above), interaction effects were not significant.

"Distinctive feature" analysis

To obtain more precise data on the nature of the speech information conveyed by (a) the "recoded" auditory signal, and (b) the visual cues, those phoneme stimuli with major HF - components, and their corresponding 'judged match' responses, were grouped into confusion matrices (see Appendix 1).

Making use of the "phoneme parameter" table, provided by Peterson (1970, p 168), the matrices were then examined for how well information regarding "manner" of articulation (sibilant, fricative, stop etc.), and "place" of articulation (unilabial, alveolar, palatal etc.) had been transmitted (see Table 2).

TABLE 2

Percentage of HF - component phoneme imitations, with correct "manner" and/or "place" of articulation under all conditions.

Condition	Non-Visual		Visual	
	Manner	Place	Manner	Place
No Recoding	50.5	41.3	71.2	71.7
Recoding	69.0	62.5	89.7	89.1

A split plot two-way analysis of variance was carried out for individual "manner" of articulation scores and "place" of articulation scores respectively. The "recoding" of HF - components was found to produce a significant improvement in both the imitation of "manner" of articulation ($\underline{F}$ = 8.609, d.f. = 1,14, $\underline{p} < .025$), and the "place" of articulation ($\underline{F}$ = 8.566, d.f. = 1,14, $\underline{p} < .025$), for HF - component phonemes.

Visual cues provided by the experimenter's articulatory movements, further improved both imitation of "manner" of articulation ($\underline{F}$ = 14.237, d.f. = 1,14, $\underline{p} < .01$) and "place" of articulation ($\underline{F}$ = 41.42, d.f. = 1,14, $\underline{p} < .001$), for HF - component phonemes.

There were no significant interactions between improvements brought about by recoding, and improvements brought about by visual cues.

Decision times and required repetitions.

The decision times, and required repetitions could only be assessed (under the procedure used), for each CVC stimulus as a whole, although it is plausible to attribute any differences occurring between the "recoding" and "no-recoding" conditions, mainly to perceived differences in the HF - components of the respective CVC syllables.

The mean decision time per stimulus, and the mean number of required repetitions per stimulus are given in Table 3.

TABLE 3

Mean decision time, and required repetitions, per CVC stimulus under all conditions.

Condition	Non-Visual		Visual	
	Decision Time (seconds)	Required Repetitions	Decision Time (seconds)	Required Repetitions
No Recoding	12.0	1.8	9.1	1.4
Recoding	19.2	3.3	11.4	1.8

A split-plot two-way analysis of variance was carried out on the decision times and "required repetitions" respectively. When stimuli were "recoded", the times required for subjects to make a "judged match" imitation response was significantly greater ($F = 5.609$, d.f. $= 1,14$ $p < .05$), but the number of requests for the stimuli to be repeated, while greater, was not significantly so ($F = 3.54$, d.f. $= 1,14$, $p > .05$).

When the experimenter's face was visible, a significant decrease occurred in both the decision times ($F = 15.301$, d.f. $= 1,14$, $p < .01$), and in the required repetitions ($F = 12.237$, d.f. $= 1,14$, $p < .01$).

Discussion

The effects of visual cues

The facilitatory visual-auditory interaction effects described by Siegenthaler and Gruber (1969), were not significant in the present experiment, where the improvements brought about by recoded auditory information, and visual cues were largely additive (see Tables 1 and 2).

This may have been due to the fact that in the experiments reported by Siegenthaler and Gruber, the sum of speech intelligibility scores (a) with "voice" information only, and (b) with "lipreading" information only, was compared with the intelligibility score when subjects received simultaneous "voice" and "lipreading" information, whereas in the present experiment, only the interaction of an increment in "voice" information ("recoded" HF - components), with the addition of visual cues (provided by the experimenter's articulatory movements), was investigated.

The provision of visual cues brought the percentage correct imitation score (for all phonemes and clusters), up from 42.4% to 76.4% under "no-recoding" conditions and from 55.9% up to 84.3% under "recoding" conditions (see Table 1), with a significant improvement occurring in imitation of both "manner" of articulation ($p < .01$) and "place" of articulation ($p < .001$) of the HF - component phonemes.

As indicated by the highly significant value of $\underline{F}$ ($\underline{F}$ = 137.788, $\underline{d.f.}$ = 1,14, $\underline{p} \ll .0005$), the overall improvement (on both HF - component and no - HF - component phonemes), occurred with great consistency for individual subjects, under the various conditions of the experiment.

Further, when visual cues were permitted, a significant decrease took place in both the time required to make a 'judged match' response ($\underline{p} < .01$), and the number of times the subject required the CVC stimulus to be repeated ($\underline{p} < .01$), indicating that although subjects had received no lipreading training, the integration of auditory information and visual cues required no 'familiarisation', but on the contrary, simplified the task of making a 'judged match' response.

These findings (based on 'judged match' imitation responses), considered together with the findings of Erber (1969), Neeley (1956), O'Neill (1959), Sanders (1967) and Sumby and Pollock (1954) (using "recognition" responses), provide strong grounds for concluding that for CVC syllables, normal hearing subjects have a well developed (although untrained) ability, to combine auditory and visual speech information, under difficult hearing conditions.

The effects of "recoding"

The "recoding" of HF - components brought about a significant improvement in imitation performance, particularly for those phonemes (or clusters) containing major HF - components (/s/, /ʃ/, /z/, /f/, /t/, /st/, etc.) bringing their percentage correct imitation score up from 32.9% to 52.3% under non-visual conditions and from 66.5% to 79.6% under visual conditions (see Table 1).

These findings may best be evaluated in the light of the similar experiment carried out by Risberg (1965) in which the effects are evaluated, of the "recoding" provided by the Johansson transposer on the recognition of CV syllables by untrained, normal hearing subjects, under simulated deafness conditions.

Risberg (1965) found that the Johansson transposer improved subjects ability to recognise "manner" of articulation of fricative and stop consonants, but had no effect on the recognition of "place" of articulation. In the present experiment however, the confusion matrix analysis indicated a significant improvement in both "manner" of articulation ($\underline{p} < .025$) and "place" of articulation ($\underline{p} < .025$) in the "recoded" condition (see Table 2).

The different findings could be due to differences in the validation procedures used, e.g. white noise is not high-pass-filtered in the Risberg (1965) experiment, which may have interfered with HF - component information relating to "place" of articulation "recoded" by the Johansson transposer into the low frequency region. Further, in the Risberg (1965) experiment no articulation - discrimination interaction was possible (see p 41). However, as pointed out on p 35 , the mode of recoding used by the FRED preserves spectral interrelations which are destroyed by the Johansson transposer, and as Risberg (1965) points out, cues regarding "place" of articulation are generally held to be provided by the spectral pattern.

It is argued here therefore, that the improved perception of "place" of articulation in the present experiment, may plausibly be attributed to the preservation by the FRED of spectral interrelations in the "recoded" high frequency signal, whereas the improved perception of "manner" of articulation may be attributed to the preservation of the temporal attributes of the "shifted" signal (as in the case of the Johansson transposer).

The mean time required to make a "judged match" response (in the present experiment) was significantly greater however ($\underline{p} < .05$) for "recoded" stimuli (see Table 3), indicating a need for subjects to familiarise themselves with the HF - component information, in its "recoded" form.

Nonetheless, the improved imitation performance, obtained without any experimentally guided articulation or discrimination training, leads to the conclusion that although the recoding process alters the HF - component information, to the extent of requiring some familiarisation, the "recoded" signal is still sufficiently 'speechlike' to be effectively utilised by the ear-brain system, in the imitation task, (and <u>may</u> therefore be of use in the rehabilitation of the "perceptively" deaf).

A validation study with
sensory-neural deaf children.

Aims of the experiment

In the following validation study with sensory-neural deaf children it was reasoned that to the extent that the simulated deafness situation (outlined on p 47) has functional equivalence to that of the deaf child, a similar improvement in learning to articulate (by imitation) those consonants with major HF components (henceforth referred to as HF consonants) may be expected to be brought about by the "recoding" used by the FRED.

The validation study constituted therefore a test of the assumption that the simulated deafness and actual deafness situations were in certain senses functionally equivalent as well as providing an initial evaluation of the utility of the "recoding" technique to the deaf.

As in the simulation study opportunity was provided throughout the validation study for:

(a) any articulation - discrimination interaction, and

(b) any visual - auditory interaction, to occur.

Specifically, it was predicted that ability to articulate the HF consonants would improve with imitation training, and in particular, that this improvement would be greater when "recoded" high frequency information was superimposed on the selectively amplified, and amplitude compressed, low frequency region (henceforth referred to as the R + SA condition) than on those training sessions where selective amplification and amplitude compression only were used (henceforth referred to as the SA only condition).

Further, it was predicted that performance on (a), articulation and (b), 'discrimination' transfer-of-training tests (involving the HF consonants embedded in different phonemic contexts and uttered in a different voice to that used in training sessions) would be better when both training sessions and following transfer tests were carried out under the R + SA condition, than when both training sessions and following transfer tests were carried out under the SA only condition.

Method

Subjects

Six sensory-neural, partially-hearing children were selected on the basis of having audiograms indicating little hearing above 1000 cps, but sufficient low frequency residual hearing to benefit in principle from HF information transposed by the FRED. A summary of the audiogram data, age, Snijders-Oomen IQ (Snijders-Oomen, 1959) and additional debilities (to perceptive deafness) of each subject is given in Table 4.

Test Materials

The HF phonemes were incorporated as initial consonants into three sets of six CVC words, each set constructed on the following criteria:

(a) that the middle vowel and final consonant be identical within the /s/, /ʃ/, /t /, /tr/ and /t/, /k/ groups

TABLE 4

Description of Subjects

Subject	Ear	Hearing Loss in dB 125 Hz	250 Hz	500 Hz	1kHz	2 kHz	4 kHz	8 kHz	Age (yrs-mths.)	Snijders-Oomen IQ	Debilities Additional to Perceptive Loss
S_1	L	50	70	85	100	>110	>110	>110	10-4	93	
	R	45	65	80	95	100	95	90			
S_2	L	>110	70	90	80	75	95	85	9-3	102	Some conductive loss
	R	50	60	55	70	80	95	>110			
S_3	L	45	75	95	100	105	95	85	7-10	95	
	R	60	75	100	110	>110	105	>110			
S_4	L	40	60	85	100	105	>110	>110	7-8	94	
	R	40	60	90	90	90	105	>110			
S_5	L	30	40	60	100	105	95	80	11-1	119	Slight cerebral palsy
	R	40	50	90	100	105	>110	>110			
S_6	L	5	10	45	75	65	60	60	10-1	83	Slight cerebral palsy
	R	5	10	50	65	80	65	50			

respectively (to eliminate the possibility of within group discrimination on any basis other than the initial consonant) and (b) that the meanings of the words so formed be both known to the subjects and possible to illustrate.

The words and corresponding illustrations were drawn on cards to form a pretraining or "transfer-of-training" test (set 1 cards) and two sets of "training" stimuli (card sets 2 and 3). The training and test stimuli are illustrated in Plates 4 to 9 (Appendix 2).

The stimuli used were as follows:

Set 1 - ship, sip, chip, trip; tap, cap;

Set 2 - sheet, seat, cheat, treat; tab, cab;

Set 3 - skis (pronounced /ʃ/ /i:/ /s/), seas, cheese, trees; tar, car.

Procedure and Experimental Design

To obtain a measure of the initial ability of subjects to articulate the HF consonants, the experimenter (E) held up each card of the pretraining test (set 1) twice, in random order, on each occasion stating the name of the card, which then had to be repeated by the subject (S). To obtain a separate measure of the initial ability of subjects to 'discriminate' the HF consonants, the set 1 cards were placed in front of S, after which E named each card twice, in random order, with S now required to point to the card (when named). The tests were administered to each subject through a Peters (Westrex) group hearing aid (providing both individually adjusted selective amplification for each ear, and amplitude compression).

On all subsequent tests and training sessions the FRED (consisting of both a "recoding" channel and a linearly amplified channel) was linked in series with the group hearing aid (using the "auxiliary input" of the latter).

On the basis of scores obtained on the above tests, subjects were divided into two roughly matched groups (of three subjects each). Subjects in group 1 then received seven individual articulation training sessions (using set 2 words) on successive school days, with the "recoding" channel on the FRED open circuited, i.e. with speech signals preamplified, selectively amplified, and subjected to amplitude compression (the SA only condition). Subjects in group 2 were trained similarly, but with both channels of the FRED operating, the "recoded" high frequency speech information being "mixed" with the linearly amplified channel before selective amplification and amplitude compression (the R + SA condition); the recoding channel was set 5dB above the linearly amplified channel.

Both groups were then tested for "transfer-of-training" to performance on the original articulation and 'discrimination' tests, now administered to each subject under the same conditions of amplification as those used in the articulation training sessions.

Group 1 subjects then received a further seven articulation-training sessions, but under the R + SA condition; whereas group 2 subjects now received training under the SA only condition. This was followed by a final "transfer-of-training" measurement on the original articulation and 'discrimination' tests, now administered to each subject under the same conditions of amplification as those used in the second series of training sessions.

To determine whether articulation learning (of the HF consonants) would "transfer" not only to words other than those used in the training sessions, but also to a different voice, the "pretraining" and "transfer-of-training" tests were administered by the experimenter, whereas articulation training was carried out by a teacher (of-the-deaf).

The "split-plot" design (balanced for order effects) used above, had the advantage that with respect to the main effects of interest (i.e. articulation learning under R + SA conditions as against SA only conditions) each subject served as his own control.

The experimenter converted the SA only condition to the R + SA condition by simply pressing a switch on the FRED. This made it possible to leave the teacher unaware of the amplification conditions and experimental design. It was permissible therefore, to leave certain details of the training technique most suited to the needs of a given child to the judgement of the teacher (the purpose of the experiment being to assess any advantage of "recoding" to articulation learning, in a 'realistic' teacher-pupil relationship).

In the first training series, the teacher held up each card of set 2 in turn (the order being randomised over training sessions) and trained the child to articulate the name of the card (by imitation) with particular emphasis placed on the correctness of the initial consonant. Techniques such as encouraging the child to produce a "hissing" sound (to train the articulation of /s/) or placing the back of the child's hand near his mouth (to 'feel' the differences in aspiration between /t/ and /k/) were used at the teachers own discretion. However, the child was permitted to make only three attempts at imitating each of the six words (in each training session) and only the 'acceptability' of the articulated initial consonant was scored.

The procedure followed in the second training series was the same as that of the first series, with the exception that set 3 cards were used as training stimuli.

The 'acceptability' of subjects' articulation responses (under the various experimental conditions) was separately assessed by both teacher and experimenter. When judgements differed, the subject was required to repeat the response, after which the judgement of the teacher was taken to be decisive (the latter being unaware of the experimental design).

Results

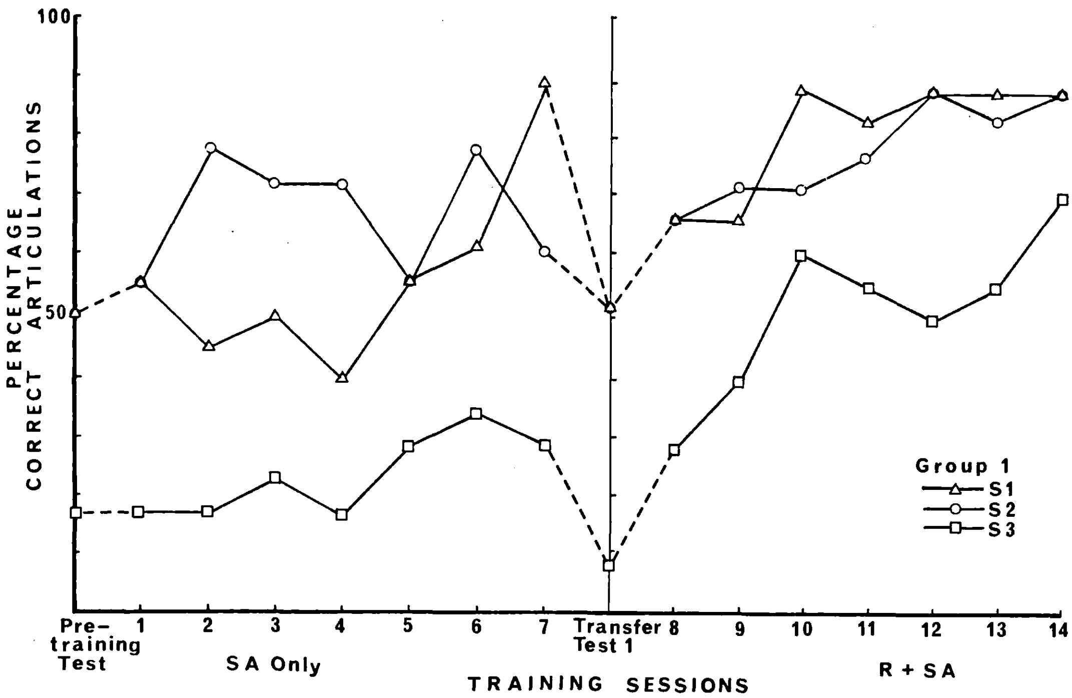
Articulation Learning

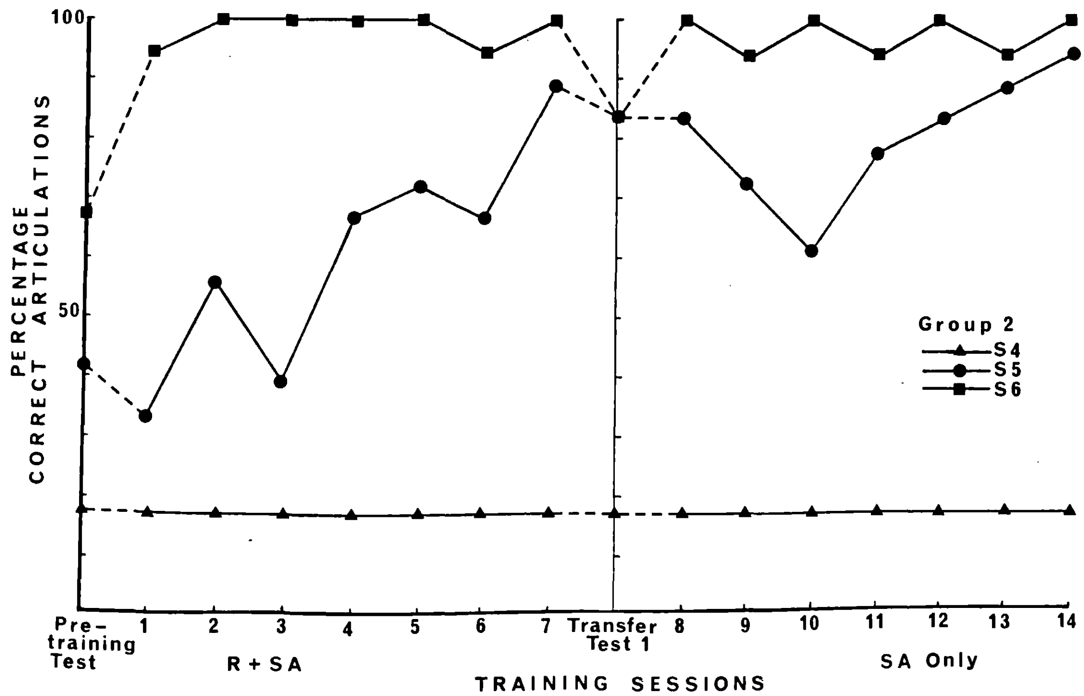
The articulation learning curves for subjects in group 1 and group 2 respectively under each condition of amplification are given in Figure 4. A split-plot, two-way analysis of variance was carried out on the raw scores.

As predicted, articulation scores improved significantly over training sessions ($\underline{F}$ = 3.34; d.f. = 6,24; $\underline{p} < .0125$) and, as indicated by a significant training-session x condition-of-amplification interaction ($\underline{F}$ = 2.62; d.f. = 6,24; $\underline{p} < .025$) the improvement under R + SA conditions was greater than under SA only conditions. The only other significant effect was the order-of-presentation x condition-of-amplification interaction ($\underline{F}$ = 10.57; d.f. = 1,4; $\underline{p} < .025$) indicating assymetrical carry-over effects, i.e. there was a greater transfer-of-training from training series 1 under R + SA conditions, to training series 2 under SA only conditions (as in group 2) than when the order of conditions (within training series) was reversed (as in group 1).

The analysis above provides the most direct interpretation of the performance curves in Figure 4. However, to compensate for "end-of-scale" effects, and to provide an analysis of articulation learning (or loss of learning) which takes into account the articulation performance of each subject prior to a given training series, a further

FIGURE 4. Articulation learning curves for each subject under (a), "recoding", selective amplification and amplitude compression (R + SA) conditions, and (b) selective amplification and amplitude compression (SA only) conditions.


100
50
PERCENTAGE CORRECT ARTICULATIONS
Group 1
S1
S2
S3
Pre-training Test
1
2
3
4
5
6
7
Transfer Test 1
8
9
10
11
12
13
14
SA Only
R + SA
TRAINING SESSIONS


100
50
PERCENTAGE CORRECT ARTICULATIONS
Group 2
S4
S5
S6
Pre-training Test
1
2
3
4
5
6
7
Transfer Test 1
8
9
10
11
12
13
14
R + SA
SA Only
TRAINING SESSIONS

computation was carried out, with the performance scores transformed to proportional learning (and proportional loss of learning) scores.

The proportional learning score (L_s) for each subject, in each training session was obtained from equation 1:

$$L_s = (A_s - A_I)/(A_{max} - A_I) \quad (1)$$

- where A_s is the articulation score on a given training session (as plotted in Figure 4), A_I is the articulation performance prior to the given training series (see below) and A_{max} is the maximum possible articulation score (18, in the procedure used).

When performance (A_s) fell below the reference level (A_I) a proportional loss of learning score was obtained from equation 2:

$$-L_s = (A_s - A_I)/(A_{min} - A_I) \quad (2)$$

- where A_{min} is the lowest possible articulation score (zero, in the procedure used).

The proportion of correct articulations on the pre-training test, provided the reference level (A_I) for each subject on training series 1. The proportion of correct articulations on transfer-of-training test 1 (identical in form to the pretraining test) provided the reference level (A_I) for each subject on training series 2.

The proportional learning curves for each group under each condition of amplification are plotted in Figure 5.

FIGURE 5. Proportional articulation learning curves for each group under (a) "recoding", selective amplification and amplitude compression (R + SA) conditions, and (b) selective amplification and amplitude compression (SA only) conditions.

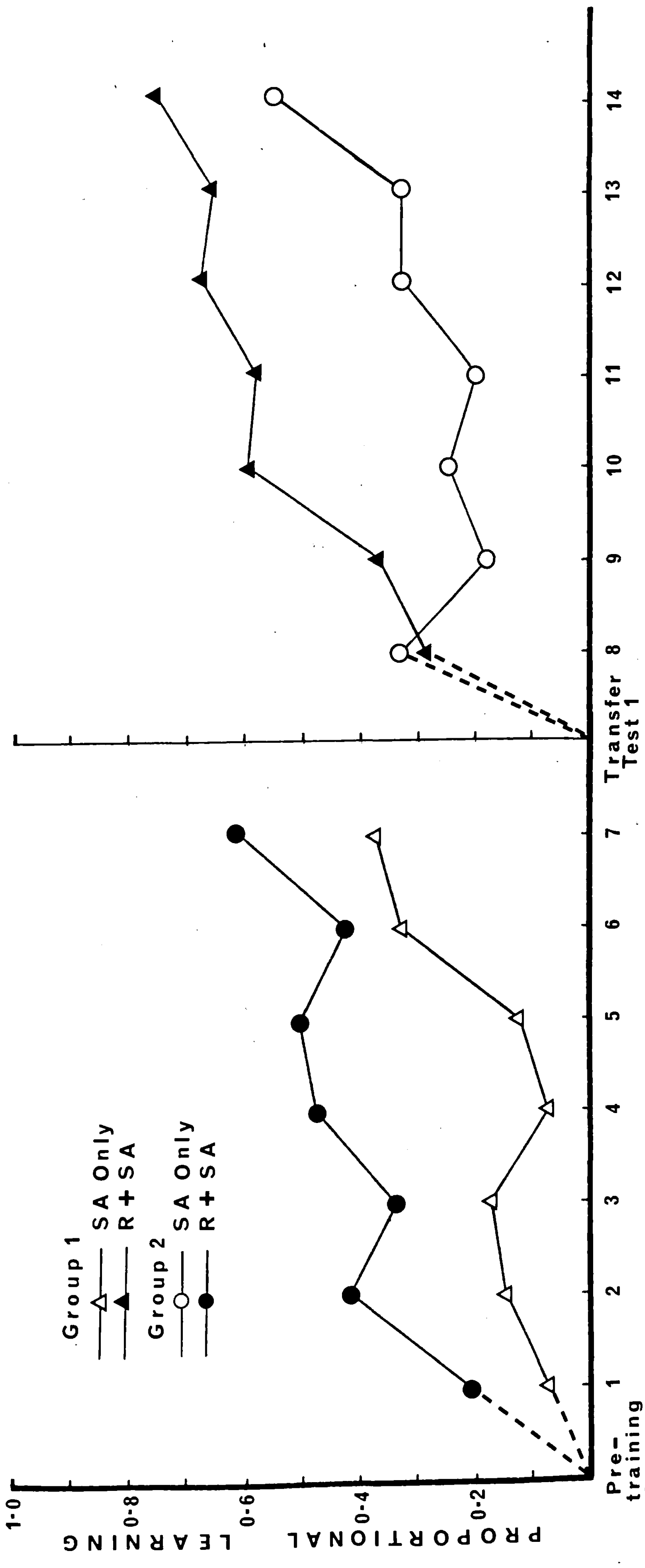

PROPORTIONAL LEARNING
1·0
0·8
0·6
0·4
0·2
Group 1
SA Only
R + SA
Group 2
SA Only
R + SA
Pre-training Test
1
2
3
4
5
6
7
Transfer Test 1
8
9
10
11
12
13
14
TRAINING SESSIONS

TABLE 5

Pretest and Transfer-of-Training Test Scores

Treatment	Conditions of Amplification	Pretest		Transfer Test 1		Transfer Test 2	
		Manner[a]	Place[b]	Manner	Place	Manner	Place
				Articulation Scores			
1	SA only	49.0	44.0	61.1	38.9	-	-
	R + SA	-	-	-	-	68.1	54.2
2	SA only	54.2	48.6	-	-	67.0	63.9
	R + SA	-	-	70.8	65.3	-	-
				'Discrimination' Scores			
1	SA only	74.0	61.1	81.9	81.9	-	-
	R + SA	-	-	-	-	75.0	77.8
2	SA only	75.0	67.0	-	-	88.9	80.6
	R + SA	-	-	69.4	61.1	-	-

Note - Scores are expressed as mean percentage correct

[a] "manner" of articulation

[b] "place" of articulation

A split-plot, two-way analysis of variance of the transformed scores revealed a significant increase in proportional learning scores over training sessions ($F = 3.22$; d.f. $= 6,24$; $p < .01$) the proportional learning under the R + SA condition being significantly greater than under the SA only condition ($F = 20.16$; d.f. $= 1,4$; $p < .006$). No other effects were significant (for the transformed scores).

Transfer-of-training

Responses in the pretraining test and transfer-of-training tests 1 and 2, were analysed in confusion matrices for correct identification of "manner" and "place" of articulation, either by an imitation response (the articulation test) or by a pointing response (the 'discrimination' test). A summary of the mean percentage correct identification of "manner" and "place" of articulation for each group under each test condition is given in Table 5.

Scores under the SA only condition versus the R + SA condition for (a) the articulation and (b) the 'discrimination' tests, were separately analysed in split-plot, two-way analyses of variance.

Articulation transfer. As predicted, articulation performance on the articulation transfer-of-training tests was significantly better under R + SA test conditions (following training under R + SA conditions) than under SA only test conditions (following training under SA only conditions) the mean percentage improvement of 5.4% for "manner" and 8.4% for "place" of articulation (see Table 5) being small, but consistent over subjects (F = 20.09; d.f. = 1,4; $p < .0125$). Further, this superiority of articulation transfer under R + SA conditions had in fact been partially obscured (i.e. diminished) by a significant order-of-presentation x condition-of-amplification interaction (F = 7.23; d.f. = 1,4; $p < .05$) i.e. following the pattern of the articulation learning scores (reported under "Articulation learning") the carry over effects from transfer test 1 to transfer test 2 were greater when the R + SA condition preceded the SA only condition, than vice versa.

The only other significant effect was the superior articulation of "manner" of articulation to that of "place" of articulation, over the various test conditions (F = 7.54; d.f. = 1,4; $p < .05$)

'Discrimination' transfer. Contrary to expectations, the 'discrimination' transfer-of-training test scores tended in the opposite direction to those of the articulation transfer-of-training tests, the mean percentage correct "manner" of articulation being 13.2% better under SA only conditions, and the mean percentage correct "place" of articulation being 11.8% better under SA only conditions, than under R + SA conditions (see Table 5). The subject variation within groups was much greater however, than in the articulation transfer tests, bringing the significance of the superiority of scores under the SA only condition into question ($F = 7.49$; d.f. = 1,4; $p < .06$, on a two-tailed test).

Nonetheless, given that the implications of the 'discrimination' transfer test findings could be regarded as contrary to both the predictions of the experiment, and the implications of the articulation learning and transfer-of-training results, this 'non-significant' effect should not be ignored and will be considered further in the "Discussion" (under "Discrimination").

All effects other than those due to the conditions of amplification fell well below significance.

Discussion

Articulation learning and transfer-of-training

As predicted, articulation learning of the HF consonants was significantly better when a recoded high frequency signal (provided by the FRED) was superimposed on the selectively amplified residual hearing range, than when selective amplification only was used. Further, the ability of sensory-neural children to imitate the same HF consonants occurring in different phonemic contexts and spoken in a different voice (to the "training" words) was significantly improved when "recoding" supplemented selective amplification (in both testing and training).

The improvement brought about by the FRED occurred relatively quickly (compared for example with the improvements reported by Guttman, Levitt and Belle Fleur, 1970) subjects having been given only seven training sessions of five to ten minutes, under each condition of amplification. Further, the improvement was sufficiently large to be manifest over a small group of children.

These results are encouraging, and support the implications of the "simulation" study that the "recoding" provided by the transposer is sufficiently 'speechlike' to be used with effect by the ear-brain system. It must be borne in mind however, that individual differences in response to recoding (or selective amplification) are likely to be large; e.g. subject four (Figure 4) exhibited no articulation learning under either condition of amplification (being unable to produce any consonants other than the "back" consonants /g/ and /k/). Further, as both the subject sample and training material were carefully selected on a priori grounds (see "Method") the generality of utility of the FRED as a supplementary articulation training (and/or hearing) aid for the HF consonants requires further investigation.

Discrimination learning and transfer-of-training

Production of a correct articulatory imitation response requires (a) the correct discrimination and identification of the physical speech stimulus and (b) the integration of the appropriate articulatory movements. The improved articulation (by imitation) brought about by "recoding" in both training and transfer-of-training implies therefore, a correspondingly improved discrimination and identification of previously inaccessible, high-frequency, consonant information.

A separate 'discrimination' transfer-of-training test was included (in addition to the articulation indicator) on the grounds that whereas correct articulation implies correct discrimination, it is not the case that incorrect articulation implies incorrect discrimination, i.e. subjects may be able to discriminate amongst speech stimuli without being able to articulate them.

Although the 'discrimination' test (requiring the subject to point to one of six pictures, when that picture is named) gives an indication of discrimination ability independent of articulation ability, it is no more a 'pure' measure of discrimination than the articulation response. This was demonstrated for instance on a number of occasions, where subjects undertaking the 'discrimination' test repeated the required stimulus <u>correctly</u> to themselves (although this was not a requirement) and then pointed to an <u>incorrect</u> picture. In fact a correct 'discrimination' response required (a) discrimination and identification of the physical speech stimulus (b) identification of the corresponding picture (or concept) and (c) the appropriate pointing response.

It is useful to make these underlying processes explicit in order to account for the finding that whereas the articulation learning and articulation transfer scores imply superior discrimination and identification of high frequency information under the R + SA condition, scores on the 'discrimination' transfer-of-training test tended to be better under the SA only condition.

These seemingly inconsistent findings may be simply accounted for if it is remembered that all stimuli used were "known" words, i.e. for each picture (or concept) subjects had a pre-experimental phonemic label (or word form), the latter being the conventionally spoken word form as perceived by the deaf ear fitted with a conventional hearing aid (involving selective amplification only). Further, articulation training oriented subjects entirely to the discrimination and identification of physical differences in the training stimuli (sets 2 and 3) i.e. although the corresponding training pictures were visible, the procedure did not require subjects to form new picture-word form correspond~~e~~ences, either for the training or the transfer test (set 1) stimuli.

Training under the SA only condition therefore, which improved utilisation of already available low frequency information, reinforced the already established picture-word form correspondences. The recoded high frequency information (in the R + SA condition) on the other hand, altered the perceived word form, and although more information regarding the physical speech stimulus was available (as indicated by superior articulation performance) the added information confused the picture identification task (operating on already established picture-word form correspondences).

It would appear therefore, that in order to obtain both improved discrimination (and identification) of the physical speech stimuli, and an altered set of picture-word form correspondences, training with "recoded" stimuli should combine an articulation (imitation) procedure, with some other task designed to encourage a replacement of established phonemic labels (of both training and transfer test stimuli) by "recoded" phonemic labels (e.g. a forced-choice pointing response, as used by Ling, 1968, experiment 2).

The above interpretation gives a consistent account of the various experimental findings, however to be certain that discrimination of the physical speech stimulus was not in fact worse under R + SA conditions (as might have been implied by the 'discrimination' transfer test taken in isolation) a supplementary 'discrimination' test was given to each subject on the day following the final transfer tests.

Supplementary 'Discrimination' Test

Each subject was required to make a series of two-choice discriminations, between stimuli randomly chosen from within each of the confused groups of the three sets of previously used stimuli (see "Test Materials" above).

Each stimulus set was presented to each subject under one of three conditions of amplification, these being (a) SA only, (b) R_1 + SA, where R_1 involved the recoding channel (on the FRED) set 5dB below the linearly amplified channel, and (c) R_2 + SA, where R_2 involved the recoding channel set 5dB above the linearly amplified channel. Conditions of amplification were balanced over subjects for order effects (using a Latin square design) each stimulus set occurring with equal frequency under each condition of amplification.

To place greater emphasis (than in the 'discrimination' transfer test above) on ability to discriminate (and identify) the physical attributes of speech stimuli, and to minimise dependence on already established picture-word form correspondences (in the performance of the experimental task) the teacher held up and <u>named</u> each of the two cards in turn, then immediately placed the cards in front of S and named the card to which S was required to point.

Pointing responses for each S were analysed in confusion matrices for correct identification of "manner" and "place" of articulation respectively. The mean percentage correct identification of "manner" and "place" of articulation of all subjects, under each condition of amplification, is given in Table 6.

TABLE 6

Supplementary 'Discrimination' Test Scores

Feature	Condition of Amplification		
	SA only	R_1 + SA	R_2 + SA
Manner of Articulation	91.7	95.8	95.8
Place of Articulation	91.7	93.5	90.4

Note - Scores are expressed as mean percentage correct.

Inspection of Table 6 suggests slightly superior performance under "recoding" conditions, however, a two-way analysis of variance (same subjects design) of the confusion matrix scores, indicated no significant treatment or interaction effects.

These results considered in conjunction with those of the articulation learning and transfer-of-training tests, provide strong evidence that the trend towards superior performance under the SA only condition in the 'discrimination' transfer-of-training tests, was not due to superior discrimination (and identification) of the non-recoded speech stimuli, but to some other variable, e.g. the reinforcement of already established picture-word form correspondences.

Finally, in addition to the conventional distinction between discrimination and articulation ability, the discrepancy between scores on the two types of 'discrimination' test (used above) indicates the need in studies of this type, to use test procedures which clearly separate measures of how well subjects have learnt to discriminate (and identify) physical attributes of the "recoded" speech stimuli, from measures of the extent to which established phonemic labels (for known pictures or concepts) have been replaced by the "recoded" phonemic labels.

Summary of experimental findings.

In order to increase the otherwise inaccessible high frequency speech information in the low frequency residual hearing range of the sensory-neural deaf, a "recoding" mode was proposed which "shifts" the 4000 cps-and-above speech band down to the region 0 cps-and-above, superimposing the high frequency signal on the residual low frequency range, in a way intended to retain the 'speechlike' nature of the "recoded" signal. The utility of the "recoding" and other factors related to the acquisition of "recoded" speech were examined in two studies, the first involving normal hearing subjects under "simulated deafness" conditions and the second, sensory-neural deaf children.

The initial "simulation" study assessed the contribution of (a) "recoding", and (b) visual cues provided by the articulatory movements of the experimenter, to the ability of subjects to imitate CVC nonsense syllables in a face-to-face, individual, articulation imitation task. The experimental conditions were designed to approximate "sensory-neural" deafness and a 'spontaneous' speech acquisition situation as closely as possible, providing opportunity for any articulation-discrimination interaction and/or any visual-auditory information interaction to occur.

Visual cues brought about an improved imitation of both "manner" and "place" of articulation of certain fricative, sibilant and stop consonants with major high frequency energy components in the region 4000 cps and above (termed HF consonants) as well as a highly significant improvement in overall phoneme imitation scores. This indicated that the normal hearing subjects, although untrained in lipreading, had a well developed ability to combine visual and auditory speech information under the difficult hearing conditions created by the "simulated deafness" situation.

"Recoding" further improved the imitation of both "manner" and "place" of articulation of the HF consonants without deteriorating the imitation of those phonemes with major energy components in the low frequency region, (on which the "recoded" high frequency signal is superimposed) e.g. the vowels.

No interaction occurred however, between improvements brought about by visual cues, and those brought about by "recoding", the effects being largely additive.

The improved imitation performance under "recoding" conditions, occurring without formal discrimination or imitation training with the "recoded" stimuli, was taken to be evidence of the 'speechlike' nature of the "recoded" stimuli, and the utility of the "recoding" to normal hearing subjects, as sufficient grounds for carrying out further studies with the deaf.

Accordingly, a validation study, assessing the effects of "recoding" on articulation and discrimination learning of HF consonants, was carried out with sensory-neural deaf children, again providing opportunity for any articulation-discrimination interaction and/or any visual-auditory information interaction to occur, in a face-to-face, individual, articulation training situation.

"Recoding", combined with selective amplification and amplitude compression, led to significantly better articulation learning of HF consonants, than selective amplification and amplitude compression only, the improvement being manifest within only seven training sessions of five to ten minutes each (under "recoding" and "no-recoding" conditions respectively). Further, "recoding" produced a small but significant improvement in articulation transfer-of-training, measured by ability to imitate the same HF consonants embedded in different CVC words to the training words and uttered in a different voice (i.e. that of the experimenter as opposed to that of the teacher-of-the-deaf who trained the children).

On the other hand, scores on a 'discrimination' transfer-of-training test, requiring subjects to point to one of six cards illustrating the words in which the HF consonants were embedded (an "absolute identification" task) 'tended' to be better under "no-recoding" conditions. However, in a further two-choice 'discrimination' task where the experimenter firstly named both cards and then named the card to which the subject was to point (thereby lessening the dependence on well-learnt "non-recoded" labels in the performance of the experimental task) this 'tendency' disappeared.

The scores on the initial 'discrimination' (transfer-of-training) test indicated that superior discrimination and articulation of "recoded" HF consonants (evidenced by articulation training and transfer-of-training scores) did not suffice to establish the "recoded" words (in which the consonants were embedded) as substitute labels for the corresponding pictures (or concepts) in preference to already known "non-recoded" word labels. This, considered together with the discrepant scores on the two types of 'discrimination' task (described above) indicated the need, in experiments involving "recoding", to separately investigate the effects of a given training procedure or "recoding" technique on the learning of new "recoded" phonemic labels, for already known pictures (or concepts) in addition to the conventional separation of effects of "recoding" on discrimination and articulation learning.

Conclusions

The improved imitation of "manner" and "place" of articulation of HF phonemes in the simulation study, considered together with the improved articulation learning of HF phonemes by sensory-neural deaf children in the validation study, supported the hypothesis that the "recoding" mode proposed would produce signals sufficiently 'speechlike' in nature to be of use to the ear-brain system in speech processing.

Further, the congruity between findings in the simulation and validation studies, gives some indication that the "simulated deafness" situation described, is functionally similar in certain respects to the situation of the deaf person, and may be of some practical use therefore, in obtaining an initial prognosis of the utility of a given recoding technique.

It should be emphasised however, that the controlled variables of any given evaluation study involving "recoding", produce variations in performance
(a) between subjects under "simulated deafness" conditions, and a given deaf population (e.g. where pre-experimental experience with language influences performance of the experimental task)
(b) between deaf persons with different types (and degrees) of hearing loss (see e.g. Krug (1960) for an account of how the manner in which auditory and visual speech cues are combined, varies with the type of hearing loss) and

(c) between deaf persons with similar audiograms, but other relevant individual differences, e.g. I.Q. (see e.g. the articulation learning curves shown in Figure 4).

Although the improvements brought about by "recoding" (in particular to articulation learning in the "validation" study) are encouraging, therefore, considerable further work is required

(a) to elucidate the mechanisms involved in the acquisition of "recoded" (and indeed 'normal') speech, and

(b) to establish the generality of utility of the mode of "recoding" used by the FRED in the present study.

References

Abbs, J.H. and Sussman, H.M. (1971). Neurophysiological feature detectors and speech perception: a discussion of theoretical implications. J. Speech Hearing Res., 14, 23-36.

Altman, J.A. (1968). Are there neurons detecting direction of sound source motion? Exp. Neurol., 22, 13-25.

Amcoff, S. (1971). Uppfattbarhet av frekvenstransponerade stavelser vid simulerad hörselskada. Rapport NR 15, Pedagogiska Instituten, LHU, Uppsala.

Bennett, D.N. and Byers, V.W. (1967). Increased intelligibility in the hypacusic by slow-play frequency transposition. J. Aud. Res., 7, 107-119.

Bever, T.G. (1971). The nature of cerebral dominance in speech behaviour of the child and Adult. In Huxley, R. and Ingram, E. (Eds.) Language acquisition: models and methods. Academic Press: London, New York.

Biondi, E. and Biondi, L. (1968). The sampling of sounds as a new means of making speech intelligible to the profoundly deaf. Alta Frequenza, 37, 728-739.

Børrild, K. (1967). Experience with the design and use of technical aids for the training of deaf and hard of hearing children. In Pickett, J.M. (Ed.) Proceedings of the conference on speech-analysing aids for the deaf.* Gallaudet College, Washington, D.C., pp. 168-177.

Bryden, M.P. (1965). Tachistoscopic recognition, handedness, and cerebral dominance. Neuropsych., 3, 1-8.

* Reprinted in full in Am. Ann. of the Deaf, 1968, Vol.113, pp. 116-327.

Curry, F.K.W. (1967). A comparison of left-handed and right-handed subjects on verbal and non-verbal dichotic listening tasks. Cortex, 3, 343-352.

Davis, H. and Silverman, S. (Eds.) (1970). Hearing and Deafness (3rd. Edn.) Holt, Rinehart and Winston, New York.

Dale, D.M.C. (1967). Applied Audiology for Children. (2nd. Edn.) Chales C. Thomas, Springfield, Illinois.

Erber, N.P. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. J. Speech Hearing Res., 12, 423-425.

Evans, E.F. and Whitfield, I.C. (1964). Classification of unit responses in the auditory cortex of the unanesthetised and unrestrained cat. J. Physiol., 171, 476-493.

Fletcher, H. (1958). Speech and Hearing in Communication. Van Nostrand Co., New York.

Fraser, G.R. (1970). Profound childhood deafness. In Wolstenholme, G.E.W. and Knight, J. (Eds.) Sensorineural Hearing Loss. A CIBA Foundation Symposium.

Friedmann, I. (1970). The pathology of deafness. In Wolstenholme, G.E.W. and Knight, J. (Eds.) Sensorineural Hearing Loss. A CIBA Foundation Symposium.

Geldard, F. (1957). Adventures in tactile literacy. American Psychologist, 12, 115-124.

Geldard, F. (1960). Some neglected possibilities of communication. Science, 131, 1583-1588.

Geschwind, N. (1965). Disconnection syndromes in animals and man, Part I. Brain, 88, 237-294.

Guttman, N., Levitt, H. and Bellefleur, P.A. (1970). Articulatory training of the deaf using low-frequency surrogate fricatives. J. Speech Hearing Res., 13, 19-29.

Guttman, N. and Nelson, J.R. (1968). An instrument that creates some artificial speech spectra for the severely hard of hearing. Am. Ann. of the Deaf, 113, 295-302.

Hopner, E. and Andrews, M.C. (1957). Consonant response in narrow band transmission. U.S. Patent No. 3,069,506.

Hudgins, C.V. (1935). Visual aids in correction of speech. Volta Review, 37, 637-643 and 703-704.

Johansson, B. (1955). Förfarande och anordning för uppökning av överförd information inom begränsade frekvensband. Swedish Patent Application No. 1353/1955.

Johansson, B. (1959). A new coding amplifier system for the severely hard of hearing. Proceedings of the Third International Congress on Acoustics. Stuttgart, Vol.2, pp. 655-657.

Johansson, B. (1966). The use of the transposer for the management of the deaf child. J. Internat. Audiol., 5, 362-371.

Johansson, B. and Sjögren, H. (1965). The effect on normal hearing subjects and profoundly deaf children by coding voiceless consonants using frequency transposition. Proceedings of the Fifth International Congress on Acoustics. Liege, paper B69.

Kimura, D. (1961a). Some effects of temporal lobe damage on auditory perception. Canad. J. Psychol., 15, 156-165.

Kimura, D. (1961b). Cerebral dominance and the perception of verbal stimuli. Canad. J. Psychol., 15, 166-171.

Kimura, D. (1964). Left-right differences in the perception of melodies. Quart. J. Exp. Psychol., 16, 355-358.

Kringlebotn, M. (1962). Nivabegrensning og transponering. Nord. Tidskr. för Dövundervisning, 66, 161-167.

Kringlebotn, M. (1967). Experiments with some visual and vibrotactile aids for the deaf. In Pickett, J.M. (Ed.) Proceedings of the conference on speech-analysing aids for the deaf. Gallaudet College, Washington, D.C., pp. 311-317.

Krug, R.F. (1960). Effects and interactions of visual and auditory cues in oral communication. Ph.D. dissertation, University of Oklahoma, U.S.A., L.C. Card No. Mic 60-5189.

Lafon, J.C. and Isaac, R. (1963). Prothèse acoustique adaptée aux restes auditifs du sourd profond. Compt. Rend. Acad. Sci., 257, 3707-3709.

Lenneberg, E.H. (1964). A biological perspective of language. In Lenneberg, E.H. (Ed.). New Directions in the Study of Language. Cambridge: M.I.T. Press, pp. 65-88.

Lenneberg, E.H. (1967). Biological Foundations of Language. New York: Wiley.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P. and Studdert-Kennedy, M. (1967). Psychological Review, 74 (6), 431-461.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P. and Studdert-Kennedy, M. (1968). Why are speech spectograms hard to read? Am. Ann. of the Deaf, 113, 127-133.

Liberman, A.M., Harris, K.S., Eimas, P., Lisker, L. and Bastian, J. (1961). An effect of learning on speech perception: the discrimination of durations of silence with and without phonetic significance. Language and Speech, 4, 175-195.

Lindgren, N. (1965a). Automatic speech recognition. IEEE Spectrum, March, 114-136.

Lindgren, N. (1965b). Theoretical models of speech perception and language. IEEE Spectrum, April, 45-59.

Ling, D. (1968). Auditory discrimination of coded speech by deaf children. Ph.D. dissertation, Institute of Otolaryngology, McGill University, Montreal, Canada.

Ling, D. (1969). Speech discrimination by profoundly deaf children using linear and coding amplifiers. IEEE Transactions on Audio and Electroacoustics, AU-17 (4), 298-303.

Ling, D. and Druz, W.S. (1967). Transposition of high frequency sounds by partial vocoding of the speech spectrum: Its use by deaf children. J. Auditory Res., 7, 133-144.

Ling, D. and Maretic, H. (1971). Frequency transposition in the teaching of speech to deaf children. J. Speech Hearing Res., 14 (1), 37-46.

Martin, E.S. and Pickett, J.M. (1970). F1 Masking of F2 transmission discrimination in hearing-impaired listeners. J. Acoust. Soc. Amer., 48 (1), (Part 1), 93.

Martin, E.S. and Pickett, J.M. (1970). Sensorineural hearing loss and upward spread of masking. J. Speech Hearing Dis., 13, 426-437.

Màrtony, J. (1967). On the correction of the voice pitch level for severely hard of hearing subjects. In Pickett, J.M. (Ed.). Proceedings of the conference on speech-analysing aids for the deaf. Gallaudet College, Washington, D.C., pp. 195-202.

Mazéas, R. (1968). Hearing capacity, its measurement and calculation. Am Ann. of the Deaf, 113 (2), 268-274.

Mecham, S.R. and Arbess, J. (1970). The effect of syllable repetition on an auditory discrimination task in deaf children. Paper delivered to the International Congress on the education of the deaf. Stockholm, August 17-21, 1970.

Neeley, K. (1956). Effect of visual factors on the intelligibility of speech. J. Acoust. Soc. Amer., 28, 1275-1277.

Nelson, P.G., Erulkar, S.D. and Bryan, S.S. (1966). Responses of units of the inferior colliculus to time-varying acoustic stimuli. J. Neurophysiol., 29, 834-860.

O'Neill, J.J. (1954). Contributions of the visual components of oral symbols to speech comprehension. J. Speech Hearing Dis., 19, 429-439.

Perwitzschky, E. (1925). Ein neues prinzip der hörverbesserung. Zschr. Hals. Nas. Ohrenhk., 12, 593-602.

Peterson, G.E. (1970). The speech communication process. In Malmberg, B. (Ed.). Manual of phonetics. North-Holland Publishing Co., Amsterdam, pp. 153-172.

Pickett, J.M. (1963). Transmitting speech sounds by a tactual vocoder and by lip-reading. R.I.T., Stockholm: Speech Transmission Laboratory Report No. 27.

Pickett, J.M. and Martin, E.S. (1968). Some comparative measurements of impaired discrimination for sound spectral differences. Am. Ann. of the Deaf, 113 (2), 259-267.

Pickett, J.M. and Martin, E.S. (1970). Discrimination of formant transitions by listeners with sensorineural hearing loss. J. Acoust. Soc. Amer., 47 (1), Pt. 1, 128(A).

Pickett, J.M., Martin, E.S., Johnson, D., Brand Smith, S., Daniel, Z., Willis, D. and Otis, W. (1970). On patterns of speech feature reception by deaf listeners. In Fant, G. (Ed.). Proceedings of the international symposium on speech communication ability and profound deafness, Stockholm. A.G. Bell Association, Washington D.C., 1972 (in press).

Pickett, J.M. and Martóny, J. (1970). Low-frequency vowel formant discrimination in hearing-impaired listeners. J. Speech Hearing Res., 13, 347-359.

Pimonow, L. (1965). La parole synthétique et son application dans la correction auditive. Ann. Télécommun. 20, (7-8), 151-171 and 175-187.

Pimonow, L. (1968). Technical and physiological problems in the application of synthetic speech to aural rehabilitation. Am. Ann. of the Deaf, 113 (2), 275-282.

Richards, D.L. and Berry, J. (1952). Tables for the generation of logatoms based on conversational speech. Research report no. 13596, Post Office Research Station, London.

Risberg, A. (1965). The 'transposer' and a model of speech perception. In Quarterly Progress and Status Report. R.I.T. Speech Transmission Laboratory, Stockholm, Jan. 15, pp. 26-30.

Risberg, A. (1967). Visual aids for speech correction. In Pickett, J.M. (Ed.). Proceedings of the conference on speech-analysing aids for the deaf. Gallaudet College, Washington, D.C., pp. 177-194.

Risberg, A. (1969). A critical review of work on speech analysing hearing aids. IEEE Transactions on Audio and Electro acoustics, Vol. AU - 17, No. 4, 290-297.

Risberg, A. and Spens, K.E. (1967). Teaching machine for training experiments in speech perception. In Quarterly Progress and Status Report. R.I.T. Speech Transmission Laboratory, Stockholm, No. 2-3, pp. 72-75.

Roworth, D.A.A. (1970). The recoding of speech. Sound, 4 (4), 95-105.

Sanders, D.A. (1967). The relative contribution of the visual and auditory components of speech to speech intelligibility. In International conference on oral education of the deaf. Northampton, Massachusetts, June 17-21, 1967, Vol. 1, pp. 865-873.

Siegenthaler, B.M. and Gruber, V. (1969). Combining vision and audition for speech reception. J. Speech Hearing Dis., 34, 58-60.

Skinner, B.F. (1957). Verbal Behavior. Appleton-Century-Crofts, Inc: New York.

Snijders-Oomen, J. Th. (1959). Non-verbal Intelligence Tests for Deaf and Hearing Subjects. Wolters, Groningen.

Studdert-Kennedy, M. and Skankweiler, D. (1970). Hemispheric Specialisation for speech perception. J. Acoust. Soc. Amer., 48 (2), Part II, 579-594.

Sumby, W. and Pollock, I. (1954). Visual contributions to speech intelligibility. J. Acoust. Soc. Amer., 26, 212-215.

Velmans, M. (1971a). Information theory measures of the ear/brain system - some doubts. Sound, 5, 58-61.

Velmans, M. (1971b). Aids for deaf persons. British Patent Application No. 19476/71.

Wathen-Dunn, W. (Ed.) (1967). Models for the Perception of Speech and Visual Form. Cambridge: M.I.T. Press.

Webster, J.C. and Chaney, R.B. (1967). Information and complex signal perception. In Wathen-Dunn, W. (Ed.). Models for the Perception of Speech and Visual Form. Cambridge: M.I.T. Press.

Wedenberg, E. (1959). Auditory training of the severely hard of hearing using a coding amplifier. _Proceedings of the Third International Congress on Acoustics_. Stuttgart, Vol. 2, pp. 658-660.

Whetnall, E. and Fry, D.B. (1964). _The Deaf Child_. Springfield, Illinois: Charles C. Thomas.

Supplementary Bibliography of references relating to frequency transposition.

Roworth, D.A.A. (1971). Study of requirements for a transposing deaf aid: Bibliography. STL Ltd., Harlow, England. (body of the report still in preparation).

Appendix 1. Simulation experiment data.

		pp
1.	"Logatom" test list.	107
2.	Phonetically balanced practice list.	108
3.	Confusion matrices of "judged match" imitation responses.	109

1. "Logatom" test list (generated from tables supplied by Richards and Berry, 1952, in accordance with the procedures outlined on p 50). Phonemes or phoneme clusters with major high frequency energy components in the region 4000 cps and above are underlined (imitation scores for these having been separated from imitation scores for phonemes containing no major HF components, in the analysis of results).

p	a:	f
w	o	f
f	a:	z
r	ai	v
m	ʌ	n
ð	a	v
b	a	v
n	ou	nt
s	i:	v
t	u:	n
dʒ	ʌ	n
θ	i	z
k	ʌ	n
l	e	n
tr	ei	k
d	au	ndz
st	a:	nt
j	e	v
h	a	n
g	o	k
ʃ	i	z

2. <u>Phonetically balanced practice list</u> (generated from the "logatom" tables of Richards and Berry, 1952) used for pre-experimental "familiarisation" with "recoding" (see p 52).

path
bat
tool
doubt
cut
got
mud
note
less
fall
thin
that
seek
right
hat
what
yes
trait
jump
staff
shin

3. <u>Confusion matrices</u> of imitation responses (to CVC syllables) under "simulated deafness" conditions, for the group receiving "no-recoding" and the group receiving "recoding" respectively, either with or without visual cues available. Imitation responses for initial consonants (or clusters), middle vowels, and final consonants (or clusters) are shown in separate confusion matrices.

Note:- each recorded response is a "judged match" to the stimulus spoken by the experimenter, arrived at after both attempted imitations of the stimulus and repetitions of the latter by the experimenter, as required by the subject (see p 52).

(a) Initial consonants, under non-visual, "no-recoding" conditions.

	p	w	f	r	m	ð	b	n	s	t	dʒ	θ	k	l	tr	d	st	j	h	g	ʃ	Other responses
p	7									1												
w		4		2																		wr, dw
f			5				1			1		1										
r				7																		z
m					7	1																
ð				4		2																v, z
b							7							1								
n					4			2						2								
s	1								6			1										
t										2	1		3			1						kl
dʒ		1		2					1		0							1				dr, vr, z
θ	1		4									1	1			1						
k										2			5						1			
l										1				6								z
tr													5		1							kr, kw
d							2									5						br
st			1									1				1	4					v
j				5														3				
h																			8			
g							3									5				0		
ʃ		1	2						1				1						2		1	

(b) Initial consonants, under non-visual, "recoding" conditions.

	p	w	f	r	m	ð	b	n	s	t	dʒ	θ	k	l	tr	d	st	j	h	g	ʃ	other responses	unidentifiable
p	5									1			2										
w		7		1																			
f			8																				
r				7										1									
m					8																		
ð						3																v(x4), z	
b							8																
n					8			0															
s			2						6														
t										5			2						1				
dʒ											4									1		dr(x2), ds	
θ			1									2	1						4				
k	1									2			5										
l														8									
tr										4			1		1				1			kr	
d							1									4				2		v	
st																	7					sp	
j				3														5					
h													1						6			g(guttural)	
g						1	2									3				1		v	
ʃ									1										4		2		1

(c) Initial consonants, under visual, "no-recoding" conditions.

	p	w	f	r	m	ð	b	n	s	t	dʒ	θ	k	l	tr	d	st	j	h	g	ʃ	other responses
p	8																					
w		8																				
f			8																			
r				8																		
m					8																	
ð						8																
b							8															
n								6						2								
s									5	1		1				1						
t										6			2									
dʒ				1							3							1				dr (x3)
θ										1		7										
k													8									
l														8								
tr															5							kr (x2), sr
d									1							7						
st									1				1			3	3					
j						1								1				6				
h																			8			
g														1		2				5		
ʃ															2						0	tʃ (x6)

(d) Initial consonants, under visual, "recoding" conditions.

	p	w	f	r	m	ð	b	n	s	t	dʒ	θ	k	l	tr	d	st	j	h	g	ʃ	other responses
p	8																					
w		8																				
f			8																			
r				8																		
m					8																	
ð						7						1										
b							8															
n								7						1								
s									8													
t										7												kw
dʒ											5											dr (x3)
θ												7							1			
k													7						1			
l														7					1			
tr										2					5							str
d																8						
st																	8					
j																		8				
h																			8			
g																3				5		
ʃ																					6	tʃ (x2)

(e) Vowels, under non-visual, no-recoding conditions

	a:	o	ai	ʌ	ɑ	ou	i:	u:	i	e	ei	au	other responses	unidentifiable
a:	9		2		2				1	1		1	eə(x4),oə,ə:	2
o		7		3	2				1	3				
ai			4		2							1	ə:	
ʌ	5	3		12		1				3				
ɑ	5		1	1	14							1	ə:, oue	
ou		2		3	1	0						1	oə	
i:							3	5						
u:					1			7						
i		1				4		4	5				ə: (x2)	
e	1	1		5	3	1				4	1			
ei						4					1		ə: (x2),oə	
au	1				2					1		4		

(f) Vowels, under non-visual, "recoding" conditions

	a:	o	ai	ʌ	ɑ	ou	i:	u:	i	e	ei	au	other responses	unidentifiable
a:	13		2		8						1			
o		7		3	3					1	2			
ai	2		3		1	1						1		
ʌ				17	2					4	1			
ɑ	3				20			1						
ou	1			1		0				3			o: (x2)	1
i:							3	2					ju: (x2)	1
u:							1	6					ju:	1
i		1					1		7				ju:(x4),e:(x2)	1
e				6						10				
ei						4					4			
au					4						1	3		

(g) Vowels, under visual, "no-recoding" conditions.

	a:	o	ai	ʌ	ɑ	ou	i:	u:	i	e	ei	au	other responses
a:	18				5							1	
o		14		1	1								
ai			6		1					1			
ʌ				19	2					3			
ɑ	1			1	21					1			
ou		2				4							o:, oə
i:							6		2				
u:								7					ju:
i									14	1			eə
e				2	2					12			
ei										1	7		
au												8	

(h) Vowels, under visual, "recoding" conditions.

	a:	o	ai	ʌ	ɑ	ou	i:	u:	i	e	ei	au	
a:	18		1		4						1		
o		14		1						1			
ai	1		5		2								
ʌ	1			23									
ɑ					24								
ou						6						1	o:
i:							7		1				
u:								7					ju:
i							1		15				
e										16			
ei											8		
au												8	

(i) Final consonants, under non-visual, "no-recoding" conditions.

	f	z	v	n	nt	k	ndz	other responses	omissions	unidentifiable
f	4		2			1		tʃ,t,ð, g	4	1
z		7	7	2				r, l (x3)	4	
v			21	1				m(x2), l,s g(x2), d(x2)	10	
n				33				na,l,kwm,mv	3	
nt				3	5	2		t(x2),nk,θ	1	1
k	1					12		t(x3)		
ndz				5			1	nz, nθ		

(j) Final consonants, under non-visual, "recoding" conditions.

	f	z	v	n	nt	k	ndz	other responses	omissions	unidentifiable
f	11		2					lf	2	
z		9	8					l,lz(x2)	2	1
v			20	5				l(x2),d(x3), md	8	1
n			3	27				m(x8),ŋ(x7)	3	
nt				1	10	1		nk,mp(x3)		
k	1					10		p(x4), t		
ndz			1	2			2	nv,nz,nd		

(k) Final consonants, under visual, "no-recoding" conditions.

	f	z	v	n	nt	k	ndz	other responses	omissions
f	13					1		θd, θ	
z		14		1				l(x2), ŋ, d, ld(x2), lz(x2)	1
v	1		36					vd, l	1
n				44				ŋ(x2)	2
nt					8			ts, ndz nz, t(x5)	
k						15			
ndz							4	nd(x4)	

(l) Final consonants, under visual, "recoding" conditions.

	f	z	v	n	nt	k	ndz	other responses	omissions
f	9		1					lf(x4), θ	1
z		18	1	3				l	1
v	1		37					ð (x2)	
n			1	35				ŋ(x10)	2
nt				1	10			nk, nl, t(x2)	1
k						15		tʃ	
ndz				3			5		

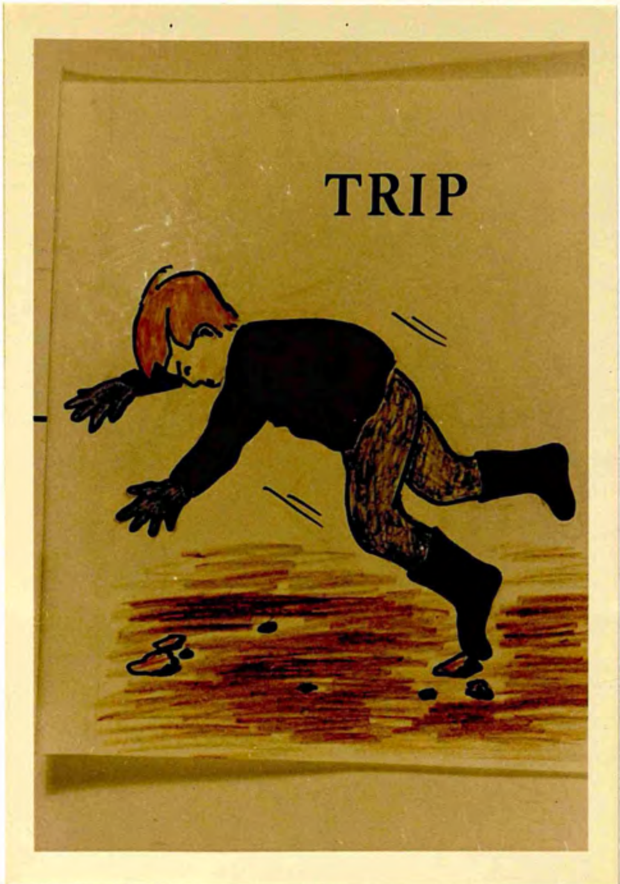
Appendix 2

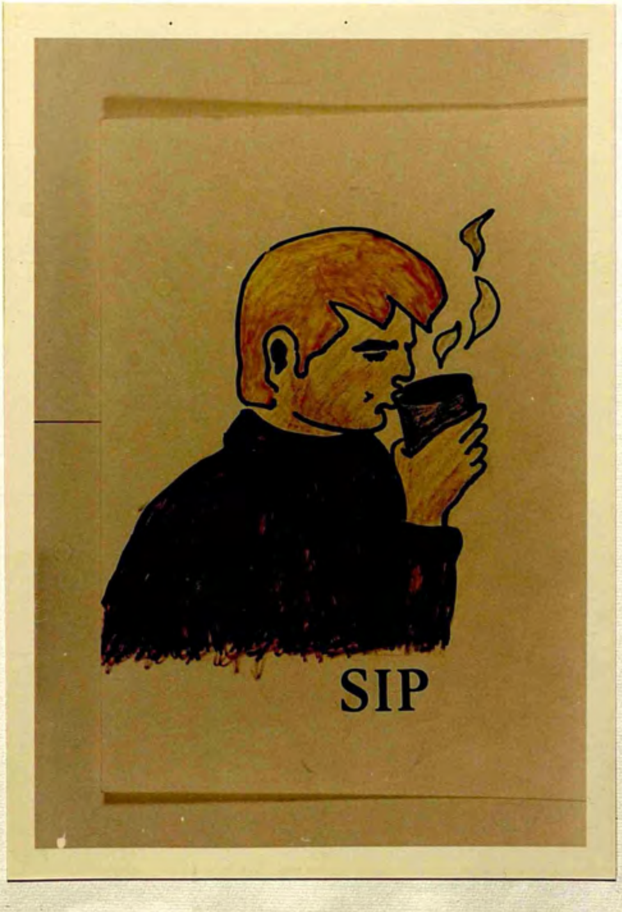
Training and test stimuli used
in the validation study.

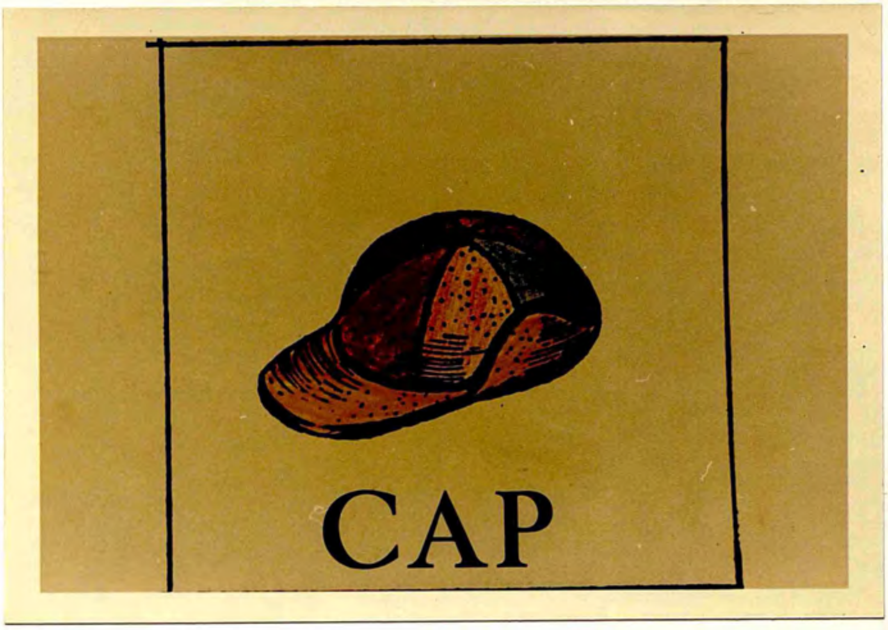
Plates 4 and 5. Test stimuli used to obtain a measure of ability to articulate and 'discriminate' the HF consonants (a) prior to articulation training (b) after articulation training series 1, and (c) after training series 2 (see "Procedure", p 69).

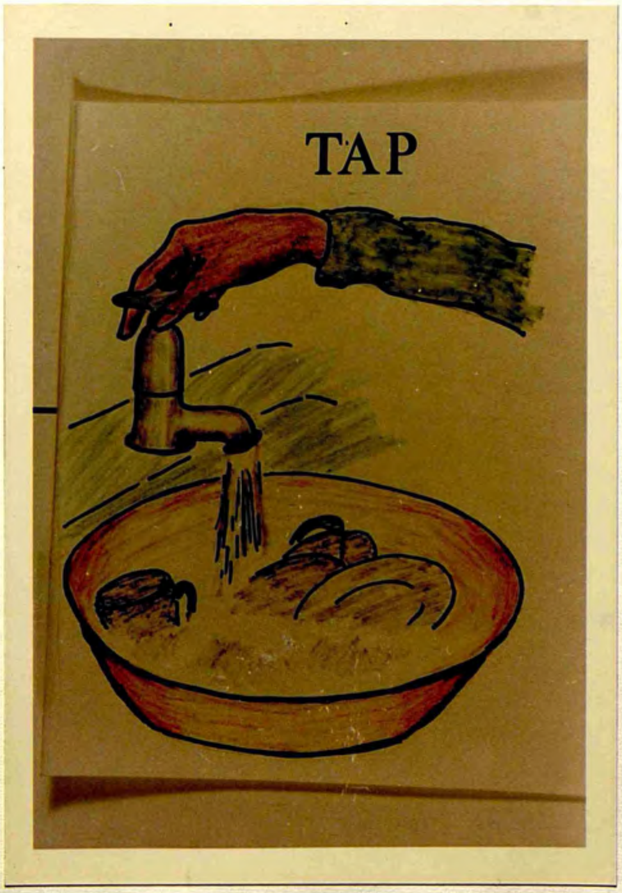

S.O.N.
SHIP


some chips


TRIP


SIP


CAP


TAP

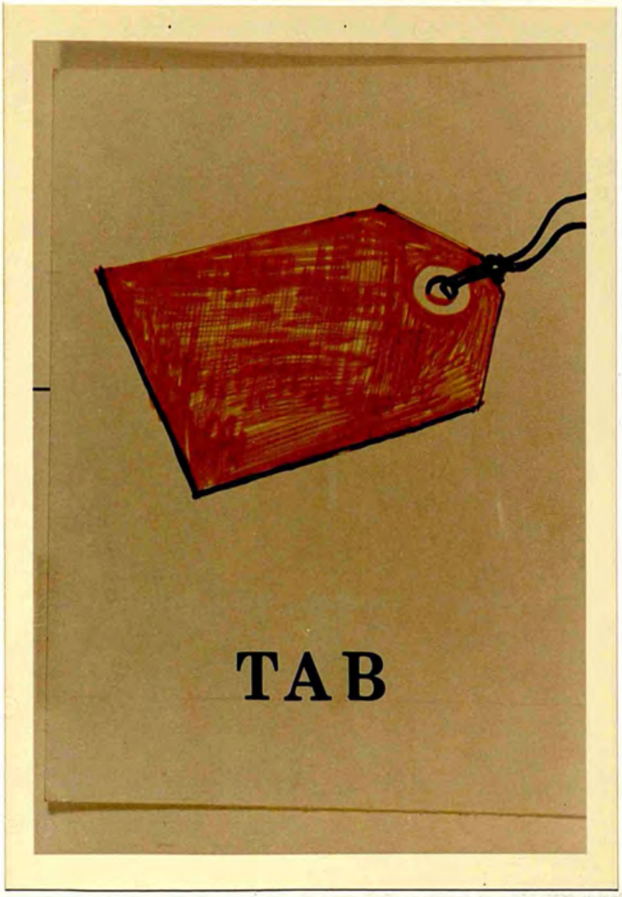
Plates 6 and 7. Training stimuli used in articulation training series 1, under "no-recoding" conditions for group 1 subjects, and "recoding" conditions for group 2 subjects.

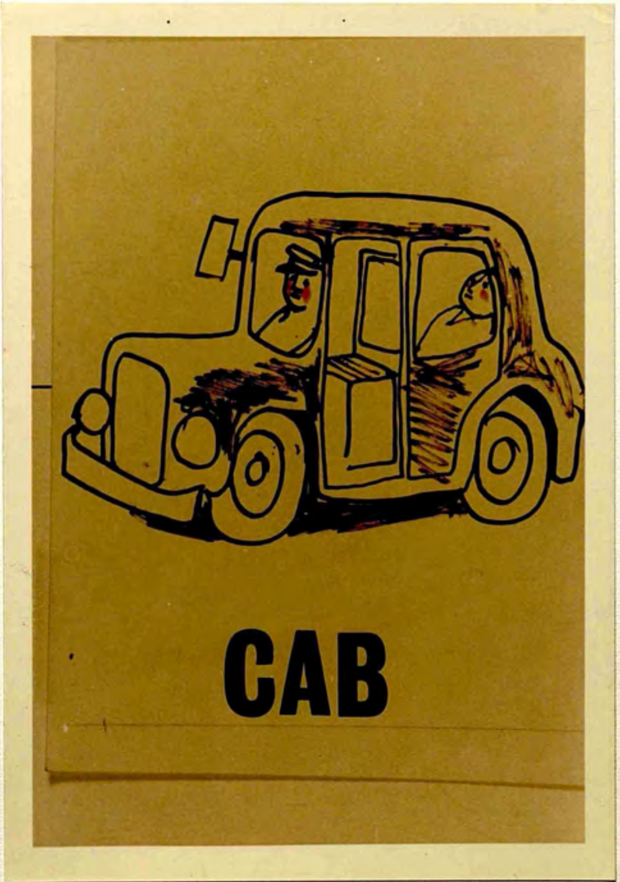

SEAT


SHEET


CHEAT

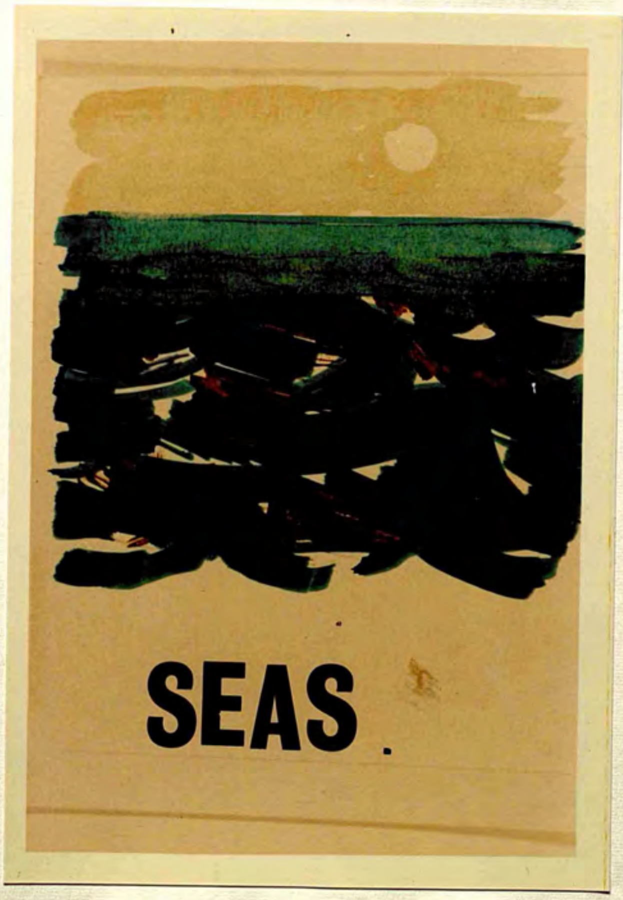

TREAT

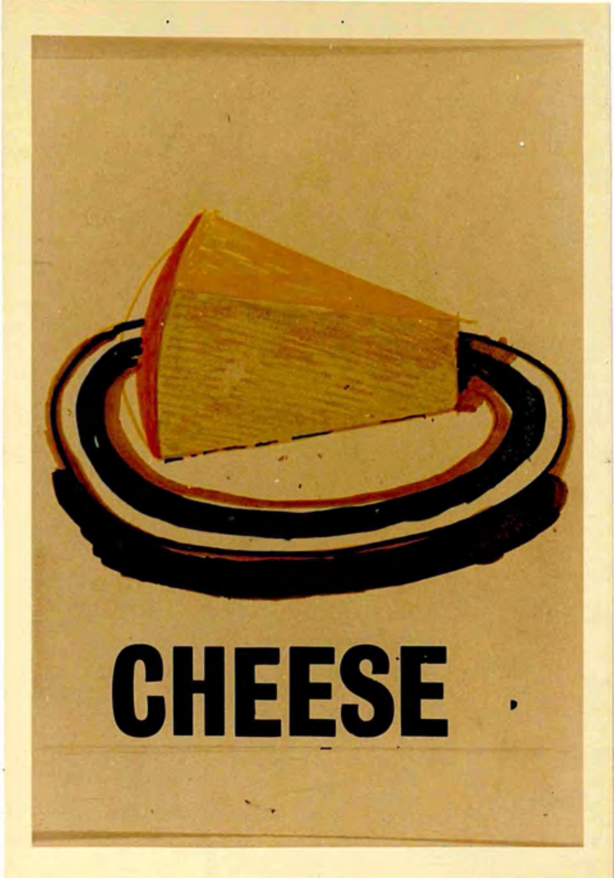

TAB

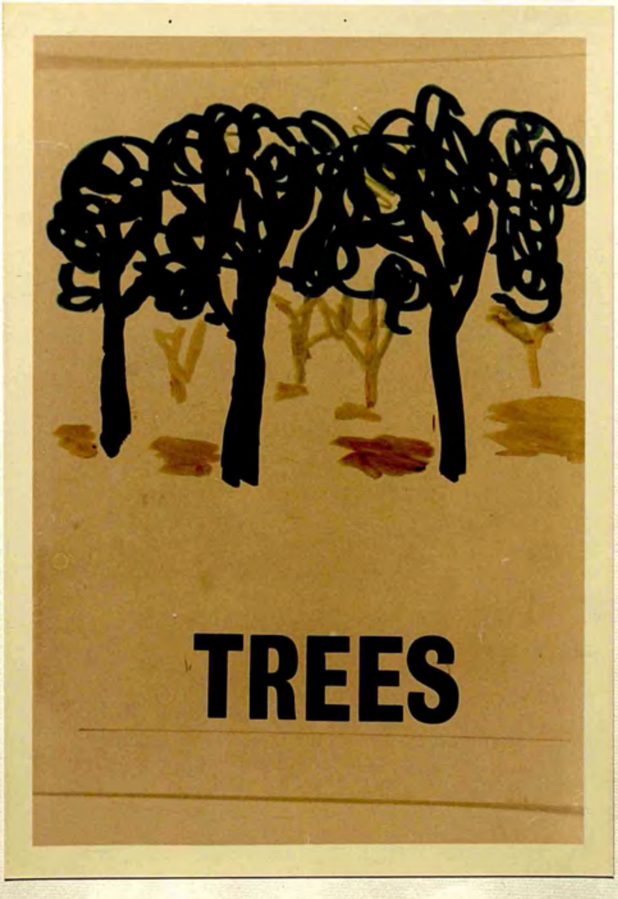

CAB

Plates 8 and 9. Training stimuli used in articulation training series 2, under "recoding" conditions for group 1 subjects, and "no-recoding" conditions for group 2 subjects.


SHES (SKIS)


SEAS


CHEESE


TREES


TAR


CAR

SOUND

Sound, 1971, 5, 58–61

INFORMATION THEORY MEASURES OF THE EAR/BRAIN SYSTEM—SOME DOUBTS

Max Velmans

Goldsmiths College, University of London

Abstract

In the quest to provide precise functional models and objective measures for various human functions, a specification of those functions in terms of information theory has from time to time been made.

In this paper the analogy of the ear/brain system considered as a communication channel is analysed. In particular, the utility of specifying a transmission rate in 'bits'/sec for the ear/brain system is evaluated.

Information theory as devised by Claude Shannon (1948) was initially applied with great success to the evaluation of the properties of information flow in telecommunication systems, where what constituted 'information' could be defined in a precise and objectively measureable way.

The analogy between the functions of various electronic data handling systems, and human functions proved to be very compelling from the 1950s onwards. It is now common for instance for the central nervous system to be referred to as a 'central processor', primary memory as a 'working store', the ego as fulfilling the functions of an 'executive program', etc. Our attention will be directed in this paper to that analogy where the ear/brain system (hence referred to as the EBS) is treated as a communication channel, with a specifiable 'channel capacity', measurable in terms of 'bits' of information.

It is worth keeping in mind that such analogies have provided useful constructs around which hypotheses for research could be framed. There is, however, a misleading tendency to take such analogies as identity statements (e.g. 'the CNS *is* the central processor') in spite of the fact that analogy constitutes neither description nor explanation.

The applicability of information theory to the evaluation of the EBS as an 'information channel' for instance, rests solely on the *assumption* that the behaviour of the EBS in speech perception is functionally equivalent in certain ways to that of a telecommunication system, as opposed to simply analagous.

There are a number of reasons for the desirability of such an assumption. If information theory can be used, then the 'information handling ability' of the EBS can be measured precisely, and according to Roworth (1970), be a 'physical concept' . . . 'independent of the semantics of the message' (p. 95).

It will be argued here however: (a) That a measure of the 'information transmitted' (although itself free from semantics), when applied to speech perception, does not refer to properties which are 'physical', or 'independent of the semantics of the message'. (b) That even if such requirements are dropped, in the light of our present knowledge (and ignorance) of the mechanisms underlying speech perception, 'information' measures are likely to have very little utility.

Let us consider the meaningfulness of the statement: 'There are about 40 different speech sounds in English, and at the most we can speak about 15 per second; the statistics of this situation require a transmission rate of only 80 bits/sec on the assumption that the phonemes are completely disconnected and unrelated' (Roworth, 1970, p. 96).

The conclusion is derived as follows: each speech sound contains an amount of information H_i (measured in 'bits'), which for equiprobable stimuli can be calculated from the formula

$$H_i = \log_2 n$$

—where n is the size of the set from which the speech sound is drawn, i.e. n is the number of possible speech sounds in the English language.

$$\therefore H_i = \log_2 40 \text{ bits.}$$

Further, if the maximum rate at which speech sounds can be spoken is 15 per sec, then the maximum information input rate H_{max} is given by:

$$H_{max} = 15 \times \log_2 40 \triangleq 80 \text{ bits/sec.}$$

If the information input is equal to the information transmitted, then the information transmitted $\triangleq$ 80 bits/sec.

The procedure normally adopted in calculating the information transmitted through the EBS, is to regard the latter as a 'black box', where the experimenter records only the 'input' (e.g. the spoken stimulus), the 'output' (e.g. the spoken identification response made by the subject), and the information contained in the input and output respectively. By observing the covariance of the input and the output a calculation of the transmitted information can be made.

Now it is important to note that 'transmitted information calculated in this way, refers to the preservation (or loss) of the sequence information contained in the stimulus set only. It is irrelevant to the measure, whether the responses correctly identify the stimuli (see Attneave, 1959). For example, consider a situation where a child is asked to point to a picture of a cat, bat, or rat, when the corresponding word is spoken. If he consistently points to bat when cat is spoken, rat when bat is spoken, and cat when rat is spoken, the calculated 'information transmitted' will be the same as if he had identified the pictures correctly. (The information measure is itself independent of meaning.)

What can be obtained from this procedure is a measure of the child's ability to *discriminate* between stimuli, e.g. we may infer that he perceives cat to *be different from* rat.

However in most situations where a specification of the EBS is required (e.g. the evaluation of the sensorineural EBS with one type of hearing aid, as opposed to another), we are interested also in the subject's ability to identify the stimulus, to imitate it, and with connected speech, to understand the 'meaning'. We note in passing therefore, that while an information measure can be obtained in this way, its utility is strictly limited.

Roworth's measure becomes even more suspect when we consider that it is not possible to state the transmission rate in 'bits' of information, unless n, the size of the input stimulus set, is clearly defined.

It can be shown that in actual fact, the size of the set of English 'speech sounds' is not clearly defined. Consider the following:

The specification of n = 40 (in Roworth's measure) reveals the *assumption* implicit in the argument, that the unit of analysis in perception is the phoneme (as there are ~ 40 phonemes in the English language). However a phoneme is *not* a specific speech sound; it is a linguistic *abstraction,* specified by observing the discriminations required, to separate the functionally different messages in a given language community. As Jakobson and Halle (1956) point out, 'Phonemes denote nothing but otherness'.

Each phoneme maps on to a class of speech sounds, whose members are functionally interchangeable (termed allophones).

The size of any given allophone set is however, not clearly defined, as the spectral pattern can be shown to vary (a) with different speakers, (b) with the same speaker in different verbal contexts, (c) with the same speaker in the same verbal context, on different occasions of utterance. *A fortiori,* the size of the total speech sound set is not clearly defined.

We note therefore, that Roworth's information measure is not (and could not be) based on the set of *physical* stimuli produced in the speech of English language users (as n is not clearly defined).

If the measure is intended to relate to speech perception, it can only be sustained if H_{max} refers to transmission rate at a point inside the EBS, i.e. at that hypothetical decoding stage in the EBS, where the total speech sound set is mapped onto the set of phonemes.

But the functional differences between messages which determine the phoneme set, are themselves defined largely *in terms of* semantics, e.g. the procedure by which one infers that |Z| and |S| are different phonemes for English language speakers, is by observing for instance, that a different meaning is attached to 'You have nice eyes', and 'You have nice ice'. If a Spaniard reacts to both statements in the same way, it is merely implied that the |Z|, |S| distinction does not exist in his language.

Therefore, if the phoneme be accepted as the basis of the information measure, the first conclusion (A) follows: that the requirement that the measure refer to a 'physical concept' . . . 'independent of the semantics of the message', must be dropped (with an attendant loss of precision and ease of measurement), as the phoneme is (a) abstract, and (b) semantically dependent.

The above conclusion does not in itself make an information measure at the level of phonemes inapplicable, as sufficient conditions for applying such a measure are (a) that the size of the stimulus set be clearly defined, and (b) that the probabilities of occurrence of members of the set are specifiable.

The 'meaningfulness' of the measure when applied to a point inside the EBS, can however be questioned on the following grounds.

Firstly, it cannot be assumed in the present stage of knowledge that the initial stage of decoding involves the segmentation of the message into phonemes. A variety of alternative 'basic units' of analysis have been suggested, ranging from the microstructure underlying 'distinctive features' (Delattre, 1965), up to phrase structure constituents, or even sentences (Miller, 1962). The evidence indicates that the unit of analysis may in fact vary with the requirements of the perceptual task. It is unlikely for instance that the same processing strategy is involved in a speech training situation, requiring a child to say 'sip', as opposed to 'ship', as is involved in decoding the meaningfulness of continuous speech. Neisser (1967) is forced to conclude that '. . . there seems to be no unit of fixed size on which speech perception depends' (p. 189).

The information measure is, however, dependent on the segment size, for the reasons that both the number of possible values of the segment, and their probabilities of occurrence will vary.

It follows therefore, that if the information measure is intended to apply to the initial stage of speech decoding, we require a prior solution of the segmentation problem.

Secondly, there is controversy over whether speech decoding proceeds in sequential fashion, starting with identification of all lowest order units (e.g. phonemes), concatenating these into higher order units (e.g. morphemes), then into words, sentences, etc., or whether the EBS operates in parallel fashion with decisions being made at all levels simultaneously, with higher level semantic decisions affecting lower level phonemic ones (as well as vice versa). (See e.g. Fry, 1970; Miller and Isard, 1963.)

If semantic rules are used by the EBS in the process of decoding, e.g. the phoneme string, then we have even less reason (than in argument A) to believe that an information measure describes some property which is, 'independent of the semantics of the message'.

Thirdly, there is controversy over whether speech perception is a largely 'passive' process, involving successive stages of analysis and classification of the input stimulus, until a best match' is made with some stored representation of the sentence (referred to by Licklider (1952) as the 'filter' model), or whether the emphasis should be placed on an 'active' mechanism, capable of generating hypotheses as to the nature of the input stimulus, and storing not the correlates of already perceived sentences, but the rules by which such sentences are generated (as in the analysis-by-synthesis model of Halle and Stevens (1962)).

It follows from the three arguments above, that the complex interactions of the EBS (and our poor understanding of them), make it difficult to isolate an objectively demonstrable property of the mechanism, to which an information measure can be made to correspond.

This is not the case for instance, in a computer system, where meaningful measures can be made of the information handling capacity of individual links (e.g. a tape reader), as well as of the system as a whole.

The second conclusion (B) follows therefore, that in the present state of knowledge of the EBS, information measures are likely to have very little utility.

References

Attneave, F. (1959). *Applications of information theory to psychology.* New York: Holt, Rinehart and Winston.

Delattre, P. (1965). *Comparing the phonetic features of English, French, German and Spanish: an interim report.* George G. Harrap & Co. Ltd.

Fry, D. B. (1970). Speech reception and perception. In Lyons, J. (ed.), *New horizons in linguistics.* Pelican Books, 29-52.

Halle, M. and Stevens, K. (1962). Speech recognition: a model and a program for research. *IRE Transactions on information theory,* IT-8, 155-159.

Jakobson, R. and Halle, M. (1956). *Fundamentals of language.* The Hague: Mouton.

Licklider, J. C. R. (1952). On the process of speech perception. *J. acoust. Soc. Amer.*, 24, 590-594.
Miller, G. A. (1962). Decision units in the perception of speech. *IRE Transactions on information theory*, IT-2, 129-137.
Miller, G. A. and Isard, S. (1963). Some perceptual consequences of linguistic rules. *J. verb. Learn. verb. Behav.*, **2**, 217-228.
Neisser, U. (1967). *Cognitive psychology.* New York: Appleton-Century Crofts.
Roworth, D. A. A. (1970). The re-coding of speech. *Sound*, Vol. 4, No. 4, 95-105.
Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. tech. J.*, **27**, 379-423, 623-656.

Comment on 'Information Theory Measures of the Ear/Brain System—Some Doubts'

Surely Dr. Velmans is not suggesting that the brain depends upon mysterious metaphysical processes? If the workings of the ear/brain system obey natural laws, then they can be described (and probed) by the application of information theory, which is simply a statement of the statistical nature of information transmission in the physical world. The accuracy of the description depends upon the depth of our knowledge, but we cannot afford the luxury of waiting until our knowledge is complete before making use of it. Admittedly the EBS is neither time- nor signal-invariant, in a way which we cannot precisely define, but this does not reduce the validity of the approach—merely its accuracy.

It is irrelevant to the argument that the set of English speech sounds is not clearly defined in number: what matters is that there *is* a set. It is also irrelevant what units of analysis comprise the set, but the fact that phonemes do indeed 'denote nothing but otherness' surely makes them ideal candidates.

The approximations cannot be denied in what Velmans gratifyingly calls 'Roworth's information measure' (though it is not mine at all, being found in several standard works). The requisite bit-rate for perception clearly does depend on the inter-phonemic constraints imposed by the semantics of the language. However, the effect of these constraints is to reduce the requisite information rate still further (1), merely strengthening the conclusion that the transmission capacity available in the profoundly deaf ear (as determined by discrimination measures (2)) is greater than that needed for continuous speech. Thus it ought to be possible to recode speech satisfactorily for the profoundly deaf.

Phonemes are more than mere linguistic abstractions, which is why it turns out to be such a peculiarly difficult task to train Velmans' Spaniard to discriminate between |Z| and |S|. Within his brain there is a mechanism which has been set during the plastic period of his early childhood to react most efficiently to the phonemic structure of Spanish. This is attested to by perceptual discrimination measures using synthetic speech stimuli, which reveal a very distinct sharpening of feature discrimination at the consonantal phoneme boundaries, even in the absence of any semantic clues (3). The mechanism involved loses its plasticity early in life, which is why it is essential to fit deaf children with a hearing aid at as early an age as possible. Without adequate stimulation during the period of maturation, significant perceptual sharpening does not occur, and the discrimination ability shows a permanent deficit which cannot be restored by an aid later in life. The form of stimulation must be one which can transmit essential speech features through the damaged ear, so that in the profoundly deaf case a recoding aid is indicated.

As yet, no-one has a successful recoding aid to offer, but it is to be hoped that a more unified attack on the problem will begin to produce results. The main aim of my paper was to help this by suggesting a set of criteria which such an aid must fulfil (4), and these conclusions are not affected by any qualms one may have about the precise validity of information theory in this context.

D. A. A. Roworth,
ITT Components Group Europe,
Standard Telephone and Cables Ltd.,
Manufacturing Services Division,
Cefndy Road, Rhyl, Flintshire.

References

(1) Roworth, D. A. A. (1970). Page 96, The recoding of speech. *Sound*, 4, 4, 95-105.
(2) For example: Mazeas, R. (1968). Hearing capacity, its measurements and calculation. *American Annals of the Deaf*, **113**, 2, 268-274.
(3) Liberman, A. M., Harris, S. K., Hoffman, H. S., and Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *J. Exptl. Psych.*, 54, 5, 358-368.
(4) *Ibid.*, Table 2.

PATENT SPECIFICATION

(11) 1340 105

DRAWINGS ATTACHED

(19)

(21) Application No. 19476/71 (22) Filed 8 June 1971

(23) Complete Specification filed 6 June 1972

(44) Complete Specification published 12 Dec. 1973

(51) International Classification H04R 3/00

(52) Index at acceptance

H4J 24B1A 24C1

H4R 22A

(71) Inventor MAX LEOPOLD VELMANS

(54) AIDS FOR DEAF PERSONS

(71) We, NATIONAL RESEARCH DEVELOPMENT CORPORATION, a British corporation established by statute, of Kingsgate House, 66/74 Victoria Street, London, SW1E 6SL, do hereby declare the invention, for which we pray that a patent may be granted to us, and the method by which it is to be performed, to be particularly described in and by the following statement:—

This invention relates to hearing or speech training aids for deaf persons having some residual hearing.

Deafness is conventionally classified as being either of the "conductive" kind, in general involving middle ear malfunctions, or of the "perceptive" kind, involving inner ear or auditory nerve malfunctions. When such malfunctions result in a loss of sensitivity to a part of the speech spectrum, e.g. to high frequency components of the speech spectrum, speech perception, acquisition and production may be impaired.

Conventional hearing aids can compensate for the reduced high frequency sensitivity in conductive deafness by selectively amplifying the high frequency speech components. In perceptive (sensory-neural) deafness, however, such techniques are frequently ineffective because the neural circuits mediating frequencies above a particular level, e.g. above 1000 Hz., are completely inoperative (as opposed to lowered in sensitivity), thus rendering inaccessible to the deaf person sections of the speech spectrum, e.g. the fricatives, essential for intelligibility.

To make better use of any residual hearing (usually low frequency residual hearing) which the perceptively deaf person may possess, various means for producing low frequency correlates of the high frequency sections of the speech spectrums have been proposed. Such prior proposals have involved the transposition of the whole or part of the speech spectrum into the region of residual low frequency hearing, e.g. the region below 1000 Hz., using frequency compression or time domain compression systems. The signals produced by such means, although detectable by the perceptively deaf, are not "speech-like" in character and have accordingly met with only very limited success in assisting the perceptively deaf in the understanding, acquisition and production of speech.

The present invention has as its object to enable an aid to be provided for a person with perceptive deafness but with some residual hearing which will enable normal speech to be reproduced within a restricted frequency range so as to have a more speech-like character than is possible with the prior proposals above referred to.

The present invention provides a hearing aid or speech training aid for a deaf person, comprising means for changing electrical signals having a first range of electrical frequencies corresponding to a first range of audio frequency signals used in normal speech into electrical signals having a second range of electrical frequencies different from said first range of electrical frequencies in such a manner that the frequency differences of said first range of electrical frequencies are substantially maintained in said second range of electrical frequencies and for superimposing the changed electrical signals onto electrical signals having frequencies within said second range of electrical frequencies and which correspond to a second range of audio frequency signals used in normal speech.

The invention also provides a method of recording audio frequencies used in normal speech in a hearing or speech training aid for the deaf, the method comprising translating the audio frequencies used in normal speech into electrical signals, changing those electrical signals having a first range of electrical frequencies corresponding to a first range of said audio frequencies into electrical signals having a second range of electrical frequencies different from said first range of electrical frequencies in such a manner that the frequency differences of said first range of electrical frequencies are substantially maintained in said second range of electrical frequencies, and superimposing the changed electrical signals onto electrical signals having frequencies within said second range of electrical frequencies and which correspond to a second range of audio frequency signals used in normal speech.

Although the aid of the present invention is primarily intended to assist those persons suffering from perceptive deafness but having some residual hearing, it is envisaged that the aid might also provide more comfortable hearing as compared with a conventional aid using selective amplification in cases of conductive deafness wherein there is a loss of sensitivity to a part, e.g. the high frequency part, of the speech spectrum.

Various combinations of changed and unchanged signals may be used according to the residual hearing of particular deaf persons. For example, the electrical signals corresponding to an upper audio frequency range, e.g. the electrical signals having frequencies of 4000 Hz. and above and which correspond to that range of audio frequencies which includes the fricatives, may be changed into electrical signals within a lower frequency range and superimposed on the electrical signals within said lower frequency range and which correspond to a lower range of audio frequencies, e.g. the electrical signals having frequencies of from 100 Hz. and above and which correspond to that range of audio frequencies which includes the low frequency vowel sounds. In this way, a 'speech-like' recoding of fricative information can be obtained without interfering with the low frequency vowel information.

The changing means may comprise modulating means for modulating an input signal, a single side band filter for filtering the output signal from the modulating means, and demodulating means for demodulating the output signal from the single side band filter. The output signal from the changing means may be superimposed upon the signal within conventional amplifier means. The amplifier means may comprise a pre-amplifier and an amplifier and the output from the changing means may be superimposed upon the signal within the amplifi -. Means for providing a choice of outputs within desired frequency ranges may be associated with the amplifier means for rectifying sensitivity losses within the residual hearing range of the deaf person. The means for providing a choice of outputs may comprise a plurality of filters having outputs within different frequency ranges and switch means whereby any one of said filters can be selectively connected with the amplifier means. If desired amplitude compression means may be provided so as to better utilize the dynamic range of the deaf ear.

An aid according to the present invention comprises either a hearing aid or a speech training aid. Where the aid comprises a speech training aid it may comprise means whereby both recorded and direct speech can be reproduced within a restricted audio frequency range and/or visual, tactile or like means for aiding the person receiving speech training. Where visual and/or tactile means are provided for aiding the deaf person, the arrangement may be such that the changed electrical signals or the output signals derived when the changed electrical signals are superimposed upon the electrical signals corresponding to said second range of audio frequency signals are used to produce visual and/or tactile information or to control means producing visual and/or tactile information.

In order that the invention may be the more readily understood reference will hereinafter be made by way of example to the accompanying diagrammatic drawings, in which:—

Fig. 1 is a diagram illustrating the changing, in an aid according to the present invention, of electrical signals having a first range of electrical frequencies into electrical signals having a second range of electrical frequencies different from the first range and the superimposing of the changed electrical signals onto electrical signals having frequencies within said second range, and

Fig. 2 is an example of a circuit diagram for a hearing or speech training aid according to the present invention.

Referring to Fig. 1 it will be seen that the high frequency components I and J between four kilohertz and five kilohertz of an input signal have been changed so as to have frequencies of from 0 to 1 kilohertz and have been imposed respectively upon the low frequency components A and B having frequencies below one kilohertz so that a resultant output signal comprises simply $A+I$ and $B+J$ in the low frequency range below one kilohertz.

Turning now to Fig. 2 it will be seen that the circuit illustrated comprises an input 1 for an electrical input signal fs which corresponds to an audio signal in the normal range of audio frequencies used in speech and parallel circuit paths 2, 3 for the input signal. The electrical input signal fs is obtained from suitable converter means, e.g. a microphone, which converts an audio input into an electrical output. The circuit path 2 includes a frequency modulator 4 whereby the input signal fs is modulated by a modulating signal fc to give sum and difference signals $fc+fs$ and $fc-fs$. These sum and difference signals are fed to a single side band filter 5 which filters out signals outside a required frequency range, and passes only a selected range of modulated electrical frequencies. The output signal (fs') from the single side band filter 5 is then passed to demodulator 6 where the output signal from the filter 5 is demodulated by a demodulating signal fc_1

to give an output signal $fs'-\Delta$, wherein $fc_1-fc=\Delta$. The circuit path 3 comprises a pre-amplifier 7 and an amplifier 8 for the input signal fs. The recoded signal $fs'-\Delta$ passes to the amplifier 8 by way of resistor 9 and is superimposed upon the pre-amplified input signal fs within the amplifier 8. The amplified signal $fs+(fs'-\Delta)$ from the amplifier 8 then passes by way of filter means 10 and an amplitude compressor 11 to an output 12. The filtered signal at the output 12 is then fed to a converter, e.g. a loudspeaker or an earphone which converts the electrical signal to an audio signal. The filter means 10 could comprise simply a single filter adapted to provide a predetermined frequency response. However, in the circuit illustrated, the filter means 10 comprises three filters 10*a*, 10*b* 10*c* adapted to provide different predetermined frequency responses and switch means 10*d* whereby any one of said three filters can be selected. This enables the frequency response of the output at 12 to be selected according to the requirements of a deaf person according to the range and extent of residual hearing of the person concerned. It will be understood that whilst the filter means 10 has been shown as comprising three filters, it could if desired comprise more or less filters to give a greater or lesser degree of selection. Thus, where the circuit is for a hearing aid, probably only one filter would be provided at 10 which would be selected according to the requirements of the person for whom the aid is intended. Where on the other hand, the circuit is for a speech-training aid which is likely to be used by different persons, then it is clearly preferable that the filter means 10 comprise a plurality of filters as shown.

A conventional deaf aid, such as a conventional group hearing aid or group speech training aid, having an auxiliary input connector can readily be converted into an aid according to the present invention by connecting to said auxiliary input connector a separate module comprising a circuit as shown in Fig. 2.

WHAT I CLAIM IS:—

1. A hearing aid or speech training aid for a deaf person, comprising means for changing electrical signals having a first range of electrical frequencies corresponding to a first range of audio frequency signals used in normal speech into electrical signals having a second range of electrical frequencies different from said first range of electrical frequencies in such a manner that the frequency differences of said first range of electrical frequencies are substantially maintained in said second range of electrical frequencies and for superimposing the changed electrical signals onto electrical signals having frequencies within said second range of electrical frequencies and which correspond to a second range of audio frequency signals used in normal speech.

2. An aid according to claim 1, wherein electrical signals corresponding to signals in an upper part of the audio frequency range are changed and superimposed upon electrical signals corresponding to audio signals in a lower part of the audio frequency range.

3. An aid according to claim 2, wherein electrical signals having frequencies of from 4000 Hz. and above and corresponding to signals in an upper part of the audio frequency range are changed so as to have frequencies of from 100 Hz. and above, and wherein the changed electrical signals are superimposed upon electrical signals having frequencies of from 100 Hz. and above and corresponding to audio signals in a lower part of the audio frequency range.

4. An aid according to claim 1, 2 or 3, wherein the changing means comprises frequency modulating means for modulating an input signal, a single side band filter for filtering the output signal from the modulating means, and demodulating means for demodulating the output signal from the single side band filter.

5. An aid according to claim 4, wherein amplifier means is provided in parallel with a circuit providing the changing means and wherein the output from the changing means is superimposed upon the signal within the amplifier means.

6. An aid according to claim 5 wherein the amplifier means comprises a pre-amplifier and an amplifier and wherein the output from the changing means is superimposed upon the signal within the amplifier.

7. An aid according to claim 5 or 6, wherein means is associated with said amplifier means for providing a choice of outputs within desired frequency ranges.

8. An aid according to claim 7, wherein the means for providing a choice of outputs comprises a plurality of filters having outputs within different frequency ranges and switch means whereby any one of said filters can be selectively connected with the amplifier means.

9. An aid according to any preceding claim, comprising visual and/or tactile means for aiding a deaf person.

10. An aid according to claim 9, wherein the changed electrical signals or the output signals derived when the changed electrical signals are superimposed upon the electrical signals corresponding to said second range of audio frequency signals are used to produce visual and/or tactile information or to control means producing visual and/or tactile information.

11. A method of recoding audio frequencies used in normal speech in a hearing or speech training aid for the deaf, the method comprising translating the audio frequencies used

in normal speech into electrical signals, changing those electrical signals having a first range of electrical frequencies corresponding to a first range of said audio frequencies into electrical signals having a second range of electrical frequencies different from said first range of electrical frequencies in such a manner that the frequency differences of said first range of electrical frequencies are substantially maintained in said second range of electrical frequencies, and superimposing the changed electrical signals onto electrical signals having frequencies within said second range of electrical frequencies and which correspond to a second range of audio frequency signals used in normal speech.

12. A hearing aid or speech training aid for a deaf person, the aid being substantially as herein described with reference to and as illustrated in the accompanying drawings.

13. A method of recoding audio frequencies used in normal speech in a hearing or speech training aid for the deaf, substantially as herein described with reference to the accompanying drawings.

ERIC POTTER & CLARKSON,
Chartered Patent Agents,
Kingsway House, Kingsway,
London, W.C.2.

Printed for Her Majesty's Stationery Office, by the Courier Press, Leamington Spa, 1973.
Published by The Patent Office, 25 Southampton Buildings, London, WC2A 1AY, from which copies may be obtained.

1,340,105 COMPLETE SPECIFICATION
1 SHEET This drawing is a reproduction of the Original on a reduced scale.

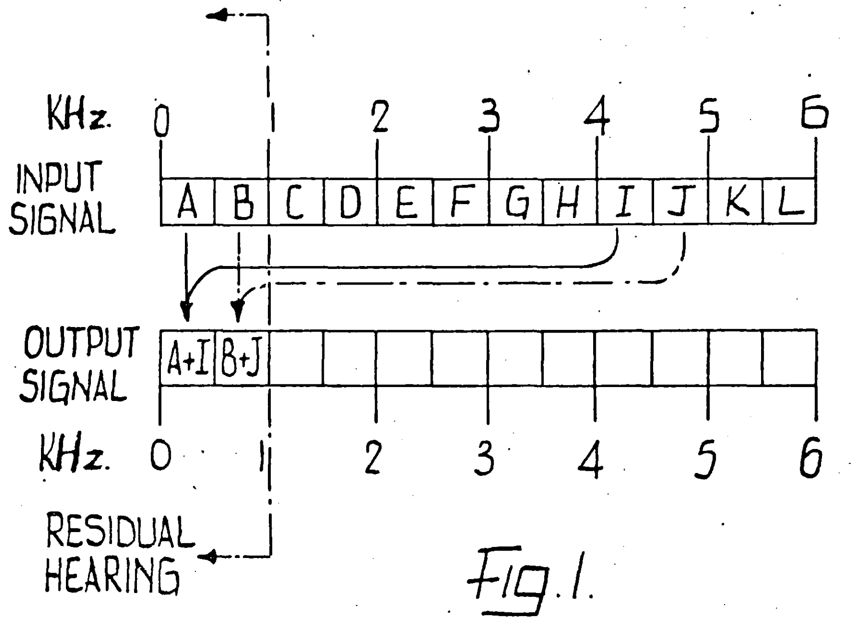


Fig. 1.

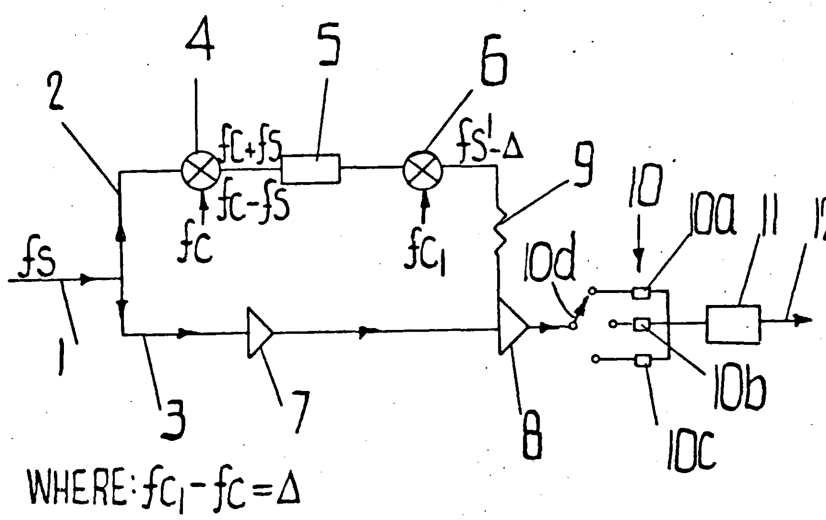


Fig. 2.

SPEECH IMITATION IN SIMULATED DEAFNESS, USING VISUAL CUES AND 'RECORDED' AUDITORY INFORMATION

MAX VELMANS
Bedford College, University of London

Reprinted from
LANGUAGE AND SPEECH
Vol. 16, Part 3, July-Sept 1973, pp. 224-236

SPEECH IMITATION IN SIMULATED DEAFNESS, USING VISUAL CUES AND 'RECODED' AUDITORY INFORMATION

MAX VELMANS*
Bedford College, University of London

A technique for the simulation of deafness, and its alleviation by a new " frequency transposition " device are described. The effects of visual cues (provided by the experimenter's articulatory movements), and of 'recoded' auditory information provided by the " frequency transposer "), on subjects' ability to make a 'judged match' imitation response to spoken CVC nonsense syllables, were evaluated. The improved imitation when visual cues were provided, of both " manner " of articulation ($p < 0.01$), and " place " of articulation ($p < 0.001$) of certain fricative, sibilant and stop consonants, together with a highly significant improvement in overall imitation scores ($p << 0.0005$), supported the conclusion that normal hearing subjects, although untrained in lipreading, had a well developed ability to 'integrate' auditory and visual speech information. Further, the improved imitation of both " manner " of articulation ($p < 0.025$), and " place " of articulation ($p < 0.025$), of certain fricative, sibilant and stop consonants brought about by 'recoding' the speech signal (without any experimentally guided articulation or discrimination training) supported the hypothesis that the recoding technique would produce a signal which was sufficiently 'speech-like' to be utilised with effect by the ear-brain system, and may therefore be of use in the rehabilitation of the " perceptively " deaf.

In the normal hearing child, the acquisition of speech involves at least three interrelated processes:
(a) auditory information, and to a lesser extent visual cues (arising from the articulatory movements of the speaker), permit the identification and discrimination of the speech sounds of the child's language community,
(b) auditory and tactual-kinaesthetic feedback allow the child to monitor his articulatory attempts to imitate those sounds, and
(c) the link thus established by auditory feedback between self produced sounds and the articulatory movements required to produce those sounds itself facilitates learning to discriminate functional segments in the speech of others, e.g. phonemes, syllables, etc. (as suggested from differing points of view by Liberman, Harris, Eimas, Lisker and Bastian, 1961; Mecham and Arbess, 1970; Whetnall and Fry, 1964).

In the deaf child, however, the reduction of auditory input commonly leads to:
(a) the dominance of visual information in monitoring the speech of others,
(b) the dominance of tactual-kinaesthetic information in monitoring his own articulations, and
(c) the virtual elimination of the facilitatory interaction between speech learning by

* *Now at Goldsmiths College, University of London.*

imitation and learning to discriminate the speech of others, for the reason that the sense modalities used to monitor these are now largely different.

Rehabilitation techniques, therefore, either train the child to make better use of the now dominant modality (e.g. teaching him to discriminate speech solely by lipreading), or, in the " multisensory " approach, to make simultaneous use of information available to different sense modalities (e.g. by teaching a joint reliance on lipreading and any low frequency residual hearing).

Such retraining can be supplemented by electronic devices, designed to recode the speech signal into a form capable of being more efficiently utilised by the remaining input sensors. Such a recoding potentially re-establishes an imitation-discrimination interaction, as information regarding the speech of others is transformed in the same way as that relating to self-produced speech.

It is argued here, therefore, that any test situation designed to evaluate the utility of a given recoding technique must provide opportunity for both imitation learning and discrimination learning to take place, thereby allowing any facilitatory imitation-discrimination interaction to occur. This will apply whether the recoding be to a visual display (e.g. Børrild, 1967; Hudgins, 1935; Martony, 1967; Risberg, 1967) to a pattern of vibro-tactile signals (e.g. Geldard, 1957, 1960; Kringlebotn, 1967), or into some acoustic form intended to make greater use of any residual hearing (the approach followed in the present paper).

Further, Krug (1960) points out that auditory information and visual information interact in ways that may be non-additive, depending on the nature of the hearing impairment, and Siegenthaler and Gruber (1969) review evidence that in sensorineural deafness, the effectiveness of each sense modality is in fact heightened by the use of the other. It is further argued therefore that any test situation designed to evaluate the utility of a given auditory recoding technique should include a visual condition, to permit any facilitatory, visual-auditory interaction to occur.

The suitability of any given auditory recoding technique is dependent on the nature of the deafness. The selective amplification of specified frequency ranges used to correct " conductive " deafness for instance, may be largely inappropriate in a case of sensorineural (" perceptive ") deafness, where the neural circuits mediating frequencies above 1000 c.p.s. are totally inoperative (as opposed to lowered in sensitivity).

To make better use of the residual hearing below 1000 c.p.s., in such instances, various " frequency transposition " devices have been built, which map either a part, or the whole of the inaccessible speech spectrum on to a set of low frequency correlates within the residual hearing range (e.g. from 100 c.p.s. to 1000 c.p.s.), thereby rendering accessible information from classes of speech sound essential to intelligibility (e.g. the fricatives).

The output from such devices have varying degrees of resemblance to ' normal ' speech, (see Risberg, 1969, for a review), the general assumption being that acquisition of any given set of speech sounds will be assisted, if for each such sound a unique, discriminable, low-frequency correlate is provided, irrespective of whether that correlate has a ' perceptual resemblance ' to the original sound.

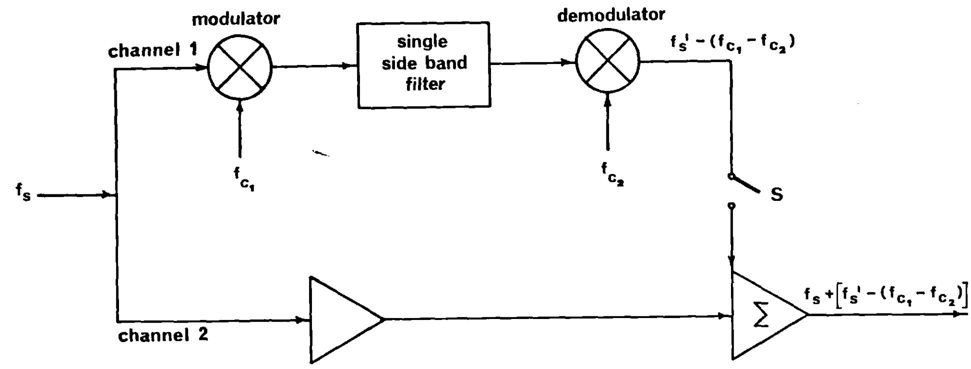


Fig. 1. Block diagram of the transposer[1], where fs is the input signal, fs' is a selected high frequency band of the input signal, fc_1 is the modulating frequency, fc_2 is the demodulating requency, $(fc_1 - fc_2)$ is the frequency " shift " performed on each frequency component of fs', and $fs + [fs' - (fc_1 - fc_2)]$ is the output signal.

The findings of the few controlled studies comparing speech acquisition with and without frequency transposition are largely inconclusive, however, indicating in the main " the need for basic studies on the nature of the cues profundly deaf subjects can use in making auditory discriminations . . . . " (Ling, 1968).

Further, there is considerable evidence (a) that there are strong hereditary predispositions to speech acquisition (see reviews by Lenneberg, 1964, 1967), and (b) that the ear-brain system is predisposed to distinguish between sounds that are ' speech-like ', and those that are not (Broadbent and Gregory, 1964; Liberman, Cooper, Harris, MacNeilage and Studdert-Kennedy, 1967; Kimura, 1961; Webster and Chaney, 1967).

It is argued here therefore, that in order to activate those mechanisms predisposed to the acquisition of speech, the recoded signals must be, not only discriminable, but also in some sense, ' speech-like '.

Accordingly, experiments are reported below with a transposer using the " frequency shift " principle, designed to maintain the speech-like nature of the input signal, by minimising the distortions performed on it in the recoding process (see Figure 1).

The output from the device consists of the amplified but otherwise unaltered input signal (on channel 2), with the recoded high frequency information (on channel 1), superimposed on the low frequency band. The only spectral properties to be altered in the " shifted " high frequency band, are the " frequency ratios " $\left(\frac{fs'_a}{fs'_b}\right)$ of the components. All other spectral interrelations, e.g. the " frequency differences " $(fs'_a - fs'_b)$ remain undisturbed (ignoring minor distortions produced by any such device).

It follows therefore, to the extent that:

(a) the frequency ratios are not an essential cue for perception, and

(b) interference between the shifted signal (on channel 1) and the unaltered signal (on channel 2) is avoided, the output will retain the 'speech-like' nature of the input.

It was hypothesized that condition (a) would be satisfied for those high frequency fricative, sibilant and stop consonant components in the region 4000 c.p.s. and above (henceforth termed HF-components), and further that if this region (having minimal vowel energy), were shifted down to the region 0 c.p.s. and above, that HF-component information in the latter region would be increased without interfering with the predominantly vowel information, in the low frequency band, (thus satisfying condition (b)).

The experiment reported below is intended to provide a test of these hypotheses, and in addition, an initial evaluation of (a) the utility of such an auditory recoding, and (b) the role played by visual cues, in assisting untrained normal hearing subjects under simulated deafness conditions, to imitate the articulations of an experimenter, in the hope that the technique may eventually be of use to the perceptively deaf.

The experiment was designed to simulate both "perceptive" deafness, and a 'spontaneous' speech imitation situation as closely as possible, giving opportunity for

(a) any imitation-discrimination interaction, and

(b) any visual-auditory interaction, to occur.

Specifically, it was predicted,

(a) that the simulated deafness conditions (by removing the HF-components) would make phonemes or phoneme clusters containing a major HF-component, more difficult to imitate than those containing no major HF-component.

(b) that imitation of phonemes or clusters containing a major HF-component would be improved, by 'shifting' the HF-component information down to the low frequency region (where it would avoid elimination by the simulated deafness conditions), and

(c) that imitation of phonemes and clusters both with and without HF-components would be improved, if the experimenter's face were visible to the subject (by virtue of the visual cues provided by the experimenter's articulatory movements).

METHOD

Apparatus

The simulation of deafness was accomplishea as illustrated in Fig. 2 (a).

The subject was seated facing the experimenter across a table 0.6 metres wide. The speech of both experimenter and subject were passed through B & K 4132 condenser microphones, mixed and amplified, before being processed by the transposer.

The processed speech signal was then passed through a Mullard GSS 001/01, variable "low pass" filter, with cut-off frequency set at 900 c.p.s. (to simulate the puretone hearing loss of a sensorineural, partially-hearing child). The filtered signal was then mixed with pre-processed white noise (at $\frac{S}{N} \doteqdot 15$ db.), before being fed to

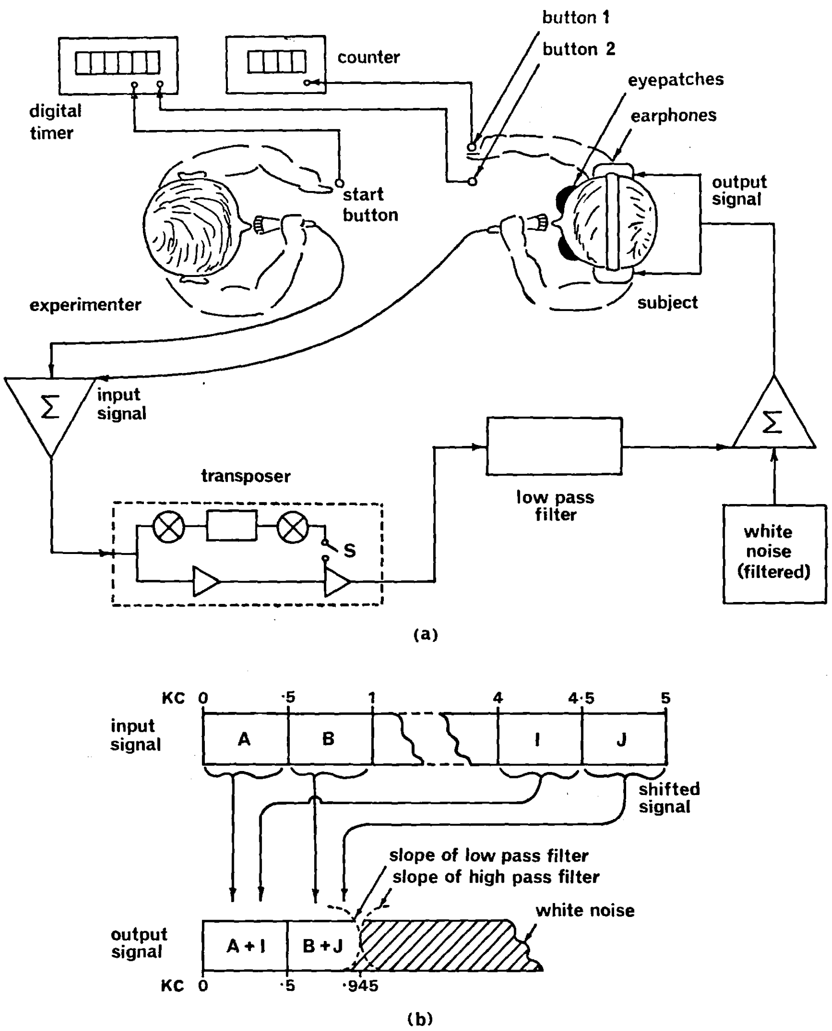


Fig. 2. (a) Schematic diagram of the simulated deafness situation.
(b) Spectrographic representation of recoding, filtering, and white noise masking operations, performed on the input speech signal.

a pair of Sharpe HA - 10A earphones, worn by the subject.

The white noise had been previously passed through a Mullard GSS 001/02, variable high pass filter, with cut-off frequency set at 990 c.p.s., and had the function of masking any high frequency components of either the experimenter's speech or of the subject's own speech, which was:

(a) not eliminated by the low pass filter,

(b) transmitted by air conduction (through the sealed earphone cups), or

(c) transmitted by bone conduction.

The white noise components below 990 c.p.s. had been removed to avoid interference with both residual low frequency information, and the recoded HF-component information, passed by the low-pass filter. The various operations performed on the speech signal are shown in Fig. 2 (b).

Test Material

The test material had to satisfy the contrary requirements of being both representative of English speech (the utility of the device being assessed for English language speakers), and being meaningless, to orient subjects away from guessing behaviour, and towards a dependence on the auditory and visual input in guiding their articulatory imitations (as would be the case with a deaf child acquiring speech).

Accordingly a list of the most frequently occurring CVC nonsense syllables (drawn from a sample of English telephone conversations), was generated from the " logatom " tables of Richards and Berry (1952), by

(a) selecting the 21 initial consonants or consonant clusters occurring with relative frequency > 1% (which accounted for 90.8% of all initial consonants),

(b) selecting the vowel most frequently combined with each respective initial consonant, and

(c) selecting the final consonant or consonant cluster, most frequently combined with each respective vowel (or, if the resultant CVC combination were meaningful, the next most frequently combined final consonant, etc.).

Subjects and Design

16 normally hearing psychology undergraduates and staff were randomly divided into two groups of eight subjects each.

Both groups were placed in the simulated deafness situation, and required to perform an imitation task (described in the *Procedure*), under both visual and non-visual conditions, in a balanced design. The experimental group however, received recoded HF-component information under all conditions, whereas the control group received no recorded HF-component information. The 21 CVC stimuli were presented in a different random order, for each subject, under each condition.

Procedure

Each subject was informed that wearing the earphones would produce simulated deafness, and that he should familiarize himself with how words ' sounded ' under this condition by reading out the words from the phonetically balanced list supplied into his microphone. Each subject was also encouraged to alter the gain control on the recoding channel of the transposer, to " note any effects ", in order to direct his attention to the recoded HF-components (although with the control group, the gain control had no effect). After familiarization, both channels on the transposer were set to the same ' comfortable hearing ' level.

Each subject was then given the following written instructions:

" In the following task you must try to imitate the *nonsense* syllables I am going to say to you. *You may say each syllable to yourself as often as you like* (into your microphone).

If you wish *me* to repeat the syllable press BUTTON 1[2].

When you consider that the sound you are making is the same as the sound I produce, press BUTTON 2[2].

I will then say the syllable again, after which you must repeat it for the last time. The first trial will be a practice trial. Indicate when you have understood the instructions and are ready to begin. "

Then, on each trial, the experimenter pressed a Venner timer " start " button[2], before presenting the CVC syllable. The pressing of BUTTON 2 by the subject, stopped the timer, thereby giving the total ' decision time ' to make a ' judged match ' response, to each CVC stimulus.

The pressing of BUTTON 1 triggered a digital counter, which therefore recorded the number of repetitions required by the subject, of each CVC stimulus.

When the subject had decided that his own articulations were the same as those of the experimenter (by pressing BUTTON 2), the latter then, (a) tape recorded both his own final presentation of the stimulus and the subject's final ' judged match ' response, (b) made an immediate phonetic transcription of the subject's response, thereby taking account of the visual cues provided by the subject's articulatory movements (to be cross-checked against the taped response), and (c) recorded the ' decision time ', and the ' required repetitions ' of each stimulus, (to obtain a measure of the ' subjective difficulty ' of obtaining a ' judged match ').

The visual condition differed from the non-visual condition, only in that in the latter, the subject wore eyepatches.[2]

The recoded HF-component condition differed from the non-recoded condition, only in that in the former, switch S (on the transposer)[2], was closed.

It was important:

(a) to the simulation of a ' spontaneous ' speech imitation situation, that the imitation procedure was largely under the subject's own control, allowing him to adapt the auditory and/or visual information to his own learning strategies, and

[2] *See Fig. 2.*

TABLE 1

Percentage correct phoneme (or phoneme cluster) imitation scores, for all subjects in all conditions.

CONDITION	NON-VISUAL HF-component	NON-VISUAL Non-HF-component	VISUAL HF-component	VISUAL Non-HF-component
No Recoding	32.9	51.9	66.5	86.3
Recoding	52.3	59.5	79.6	89.0

(b) to the evaluation of the 'speech-like' nature of the recoded signal, that no information regarding the correctness or incorrectness of an imitation attempt, was supplied by the experimenter (as discrimination training might be expected to produce some improvement even with signals which were not 'speech-like').

RESULTS

Imitation Scores

Each final imitation response (constituting the subject's 'judged match' to the CVC stimulus) was scored for the correctness of initial consonant (or consonant cluster), middle vowel, and final consonant (or consonant cluster) respectively.

To assess the predicted effects of (a) the simulated deafness conditions, and (b) the recoded high frequency information, on the imitation of phonemes (or clusters) containing major HF-components, the imitation scores for these (identified from tables provided by Fletcher, 1956), were separated from the imitation scores for phonemes (or clusters) containing no major HF-Components. The percentage of correct imitations (for all subjects) of HF-component phonemes (or clusters), and of non HF-component phonemes (or clusters) are given in Table 1.

A split plot, 3-way analysis of variance was carried out on the results. As predicted:—

(a) the elimination of high frequency information by the simulated deafness conditions made the imitation of phonemes or clusters with major HF-components more difficult than those with no major HF-components ($F = 20.466$, d.f. $= 1,14$, $p < 0.0005$), (b) the re-insertion of HF-component information into the low frequency band (by the "transposer"), improved overall imitation score ($F = 5.791$, d.f. $= 1,14$, $p < 0.025$), with a significantly greater improvement taking place for HF-component

TABLE 2

Percentage of HF-component phoneme imitations, with correct "manner" and/or "place" of articulation.

CONDITION	NON-VISUAL Manner	NON-VISUAL Place	VISUAL Manner	VISUAL Place
No Recoding	50.5	41.3	71.2	71.7
Recoding	69.0	62.5	89.7	89.1

phonemes, than for non HF-component phonemes (for the interaction, $F = 3.2$, d.f. $= 1{,}14$, $p < 0.05$), and

(c) allowing subjects to see the experimenter's face, brought about a highly significant improvement in imitation performance ($F = 137.788$, d.f. $= 1{,}14$, $p << 0.0005$).

Apart from the interaction between the effects of recoding, and whether a phoneme (or cluster) had a major HF-component (mentioned in (b), above), interaction effects were not significant.

"*Distinctive feature*" *analysis*

To obtain more precise data on the nature of the speech information conveyed by (a) the recoded auditory signal, and (b) the visual cues, those phoneme stimuli with major HF-components, and their corresponding 'judged match' responses, were grouped into confusion matrices. Making use of the "phone parameter" table, provided by Peterson (1970, p.168), the matrices were then examined for how well information regarding "manner" of articulation (sibilant, fricative, stop etc.)., and "place" of articulation (unilabial, alveolar, palatal etc.) had been transmitted (see Table 2).

A split plot two-way analysis of variance was carried out for "manner" of articulation scores and "place" of articulation scores respectively. The recoding of HF-components was found to produce a significant improvement in both the imitation of "manner" of articulation ($F = 8.609$, d.f. $= 1{,}14$, $p < 0.025$), and the "place" of articulation ($F = 8.566$, d.f. $=1{,}14$, $p < 0.025$), for HF-component phonemes.

Visual cues provided by the experimenter's articulatory movements further improved both imitation of "manner" of articulation ($F = 14.237$, d.f. $= 1{,}14$ $p < 0.01$) and "place" of articulation ($F = 41.42$, d.f. $= 1{,}14$, $p < 0.001$), for HF-component phonemes.

There were no significant interactions between improvements brought about by recoding and improvements brought about by visual cues.

Decision times and required repetitions

The decision times, and required repetitions could only be assessed (under the

TABLE 3

Mean decision time, and required repetitions, per CVC stimulus.

CONDITION	NON-VISUAL		VISUAL	
	Decision Time (seconds)	Required Repetitions	Decision Time (seconds)	Required Repetitions
No Recoding	12.0	1.8	9.1	1.4
Recoding	19.2	3.3	11.4	1.8

procedure used), for each CVC stimulus as a whole, although it is plausible to attribute any differences occurring between the recoding and no recoding conditions mainly to perceived differences in the HF-components of the respective CVC syllables.

The mean decision time per stimulus, and the mean number of required repetitions per stimulus are given in Table 3.

A split plot two-way analysis of variance was carried out on the decision times and " required repetitions " respectively. When stimuli were recoded, the times required for subjects to make a " judged match " imitation response was significantly greater ($F = 5.609$, d.f. = 1,14, $p < 0.05$), but the number of requests for the stimuli to be repeated, while greater, was not significantly so ($F = 3.54$, d.f. = 1,14, $p > 0.05$).

When the experimenter's face was visible, a significant decrease occurred in both the decision times ($F = 15.301$, d.f. = 1,14, $p < 0.01$), and in the required repetitions ($F = 12.237$, d.f. = 1,14, $p < 0.01$).

DISCUSSION

The Effects of Recoding

The recoding of HF-components brought about a significant improvement in imitation performance, particularly for those phonemes (or clusters) containing major HF-components (e.g. /s/, /ʃ/, /z/, /f/, /v/, /t/, /st/, etc.), bringing their percentage correct imitation score up from 32.9% to 52.3% under non-visual conditions, and from 66.5% to 79.6% under visual conditions (see Table 1).

These findings may best be evaluated in the light of a similar experiment carried out by Risberg (1965), investigating the effects of recoding as provided by a " Johansson transposer " (Johansson, 1961), on the recognition of CV syllables, by untrained, normally hearing subjects, under simulated deafness conditions.

Risberg (1965) found that the " Johansson transposer " improved subjects' ability to recognize " manner " of articulation of fricative and stop consonants, but had no effect on the recognition of " place " of articulation. In the present experiment however, the confusion matrix analysis indicated a significant improvement in both " manner " of articulation ($p < 0.025$), and " place " of articulation ($p < 0.025$), in the recoded condition (see Table 2).

As Risberg (1965) points out, for fricative and stop consonants, the information regarding " manner " of articulation is likely to be transmitted by the duration and timing of the noise components (the frequency spectrum being unimportant), whereas the " place " of articulation, is related to the frequency spectrum of both the noise components and the formant transitions.

Now the major difference between the transposition technique used by Johansson and the present technique is that the former utilizes a " frequency compression " (see Johansson, 1961), which allows a wider band of HF-component information to be " folded " into the residual hearing range, but destroys spectral interrelationships such as the " frequency differences ", and the " frequency ratios ", whereas the present technique merely " shifts " the signal (without compression), thereby preserving all spectral interrelationships (apart from the " frequency ratios ").

It is argued here therefore, that the improved perception of " place " of articulation in the present experiment may plausibly be attributed to this preservation of spectral interrelationships in this " shifted " signal, (although other important differences between the two experiments also exist, e.g. in the experimental task, and in the simulation of deafness technique), whereas the improved perception of " manner " of articulation may be attributed to the preservation of the temporal attributes of the " shifted " signal (as in the case of the " Johansson transposer ").

The mean time required to make a 'judged match' response (in the present experiment) was significantly greater however ($p < 0.05$) for recoded stimuli (see Table 3), indicating a need for subjects to familiarize themselves with the HF-component information, in its recoded form.

Nonetheless, the improved imitation performance, obtained without any experimentally guided articulation or discrimination training, leads to the conclusion that although the recoding process alters the HF-component information, to the extent of requiring some familiarization, the recoded signal is still sufficiently 'speech-like' to be effectively utilized by the ear-brain system in the imitation task (and *may* therefore be of use in the rehabilitation of the " perceptively " deaf).

The effects of visual cues

The facilitatory visual-auditory interaction effects described by Siegenthaler and Gruber (1969), were not significant in the present experiment, where the improvements brought about by recoded auditory information and visual cues were largely additive (see Tables 1 and 2).

This may have been due to the fact that in the experiments reported by Siegenthaler and Gruber, the sum of speech intelligibility scores (a) with " voice " information only, and (b) with " lipreading " information only, was compared with the intelligibility score when subjects received simultaneous " voice " and " lipreading" information, whereas in the present experiment, only the interaction of an *increment* in "voice" information (recoded HF-components), with the addition of visual cues (provided by the experimenter's articulatory movements), was investigated.

The provision of visual cues brought the percentage correct imitation score (for all phonemes and clusters), up from 42.4% to 76.4% under no-recoding conditions and from 55.9% up to 84.3% under recoding conditions (see Table 1), with a significant improvement occurring in imitation of both "manner" of articulation ($p < 0.01$) and "place" of articulation ($p < 0.001$) of the HF-component phonemes.

As indicated by the highly significant value of F ($F = 137.788$, d.f. $= 1,14$ $p << 0.0005$), the overall improvement (on both HF-component and non HF-component phonemes), occurred with great consistency for individual subjects, under the various conditions of the experiment.

Further, when visual cues were permitted, a significant decrease took place in both the time required to make a 'judged match' response ($p < 0.01$), and the number of times the subject required the CVC stimulus to be repeated ($p < 0.01$), indicating that although subjects had received no lipreading training, the integration of auditory information and visual cues required no 'familiarization', but on the contrary, simplified the task of making a 'judged match' response.

These findings (based on 'judged match' imitation responses), considered together with the findings of Erber (1969), Neeley (1956), O'Neill (1959), Sanders (1967) and Sumby and Pollock (1954) (using "recognition" responses), provide strong grounds for concluding that for CVC syllables, normal hearing subjects have a well developed (although untrained) ability, to 'integrate' auditory and visual speech information, under difficult hearing conditions.

References

Børrild, K. (1967). Experience with the design and use of technical aids for the training of deaf and hard of hearing children. In Pickett, J. M. (ed.), *Proc. Conf. Speech-Analysing Aids for the Deaf* (Gallaudet College, Washington, D.C.), 168.

Broadbent, D. E. and Gregory, M. (1964). Accuracy of recognition for speech presented to the right and left ears. *Quart. exp. Psychol.*, 16, 359.

Erber, N. P. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. *J. Speech Hear. Res.* 12, 423.

Fletcher, H. (1958). *Speech and Hearing in Communication* (New York).

Geldard, F. (1957). Adventures in tactile literacy. *Amer. Psychol.* 12, 115.

Geldard, F. (1960). Some neglected possibilities of communication. *Science*, 131, 1583.

Hudgins, C. V. (1935). Visual aids in correction of speech. *Volta Rev.*, 37, 637 and 703.

Johansson, B. (1961). A new coding amplifier system for the severely hard of hearing. In *Proc. Third Int. Congr. Acoust.* (New York), 655.

Kimura, D. (1961). Cerebral dominance and the perception of verbal stimuli. *Can. J. Psychol.*, 15, 166.

Kringlebotn, M. (1967). Experiments with some visual and vibrotactile aids for the deaf. In Pickett, J. M. (ed.), *Proc. Conf. Speech-Analysing Aids for the Deaf* (Gallaudet College, Washington, D.C.), 311.

KRUG, R. F. (1960). Effects and interactions of visual and auditory cues in oral communication. Ph.D dissertation, University of Oklahoma, L.C. card No. Mic 60 - 5189.

LENNEBERG, E. H. (1964). A biological perspective of language. In Lenneberg, E. H. (ed.), *New Directions in the Study of Language* (Cambridge, Mass.), 65.

LENNEBERG, E. H. (1967). *Biological Foundations of Language* (New York).

LIBERMAN, A. M., COOPER, F. S., HARRIS, K. S., MACNEILAGE, P. F. and STUDDERT-KENNEDY, M. (1967). Some observations on a model for speech perception. In Wathen-Dunn, W. (ed.), *Models for the Perception of Speech and Visual Form* (Cambridge, Mass.), 68.

LIBERMAN, A. M., HARRIS, K. S., EIMAS, P., LISKER, L. and BASTIAN, J. (1961). An effect of learning on speech perception : the discrimination of durations of silence with and without phonemic significance. *Language and Speech*, 4, 175.

LING, D. (1968). Auditory discrimination of coded speech by deaf children. Ph.D dissertation, Institute of Otolaryngology, McGill University.

MARTONY, J. (1967). On the correction of the voice pitch level for severely hard of hearing subjects. In Pickett, J. M. (ed.), *Proc. Conf. Speech-Analysing Aids for the Deaf* (Gallaudet College, Washington, D.C.), 195.

MECHAM, S. R. and ARBESS, J. (1970). The effect of syllable repetition on an auditory discrimination task in deaf children. Paper delivered to *Int. Congr. Educ. Deaf*, Stockholm, August 17-21, 1970.

NEELEY, K. (1956). Effect of visual factors on the intelligibility of speech. *J. acoust. Soc. Amer.*, 28, 1275.

O'NEILL, J. J. (1954). Contributions of the visual components of oral symbols to speech comprehension. *J. Speech Hear. Dis.*, 19., 429.

PETERSON, G. E. (1970). The speech communication process. In Malmberg, B. (ed.), *Manual of Phonetics* (Amsterdam), 153.

RICHARDS, D. L. and BERRY, J. (1952). Tables for the generation of logatoms based on conversational speech. Research Report No. 13596, Post Office Research Station, London.

RISBERG, A. (1966). The 'transposer' and a model of speech perception. In *Quart. Progr. & Status Rep.*, Speech Transmission Laboratory, Stockholm, STL-QPSR-4/1965.

RISBERG, A. (1967). Visual aids for speech correction. In Pickett, J. M. (ed.), *Proc. Conf. Speech-Analysing Aids for the Deaf* (Gallaudet College, Washington, D.C.), 178.

RISBERG, A. (1969). A critical review of work on speech analysing hearing aids. *IEEE Trans. Audio & Electro-acoust.*, Vol. AU-17, No. 4, 290.

SANDERS, D. A. (1967). The relative contribution of the visual and auditory components of speech to speech intelligibility. In *Int. Conf. Oral Educ. of the Deaf* (Northampton, Mass.), Vol. 1, 865.

SIEGENTHALER, B. M. and GRUBER, V. (1969). Combining vision and audition for speech reception. *J. Speech Hear. Dis.*, 34, 58.

SUMBY, W. and POLLOCK, I. (1954). Visual contributions to speech intelligibility. *J. acoust. Soc. Amer.* 26, 212.

WEBSTER, J. C. and CHANEY, R. B. (1967). Information and complex signal perception. In Wathen-Dunn, W. (ed.), *Models for the Perception of Speech and Visual Form* (Cambridge, Mass.), 246.

WHETNALL, E. and FRY, D. B. (1964). *The Deaf Child* (Springfield, Ill.).

British Journal of Audiology, 1974, **8**, 1-5

THE DESIGN OF SPEECH RECODING DEVICES FOR THE DEAF*

Max Velmans

Goldsmiths' College, University of London

Abstract
The paper reviews the present status of speech recoding (frequency transposition) devices and concludes that convincing evidence for the superiority of recoding devices over, for example, selective amplification, does not yet exist. A number of fundamental questions requiring answers are then outlined, upon which the design of some 'ideal' recoding device appears to be contingent. Some interim design principles are, however, proposed and a description is given of a recoding device designed with these principles in mind. Finally, some initial, encouraging results with the device are reported, and various questions relating to the utility of the device, requiring further investigation, are indicated.

The State-of-the-Art

In cases of sensory-neural hearing loss where the neural circuits mediating high frequencies are largely inoperative, it is commonly found that conventional amplifying hearing aids are of very limited value. In order to provide for better use of whatever low frequency hearing remains in these cases, a variety of techniques have been developed over the last twenty years to convert the inaccessible high frequency signals to a detectable low frequency form. This can be done with a variety of electronic techniques, e.g. frequency modulation, vocoding, frequency division or simply slowing down a tape recording (see Risberg, 1969, for a review).

The outputs produced by these devices have varying degrees of resemblance to normal speech, so it would appear that designers have in general assumed that articulation and perception of a given set of speech sounds may be assisted, if for each sound a discriminable correlate is provided within the low frequency residual hearing range (even if that correlate is a buzz or a pure tone).

Evidence that this is not a sufficient design criterion is provided by the fact that *controlled* validation studies with these devices have seldom reported a significant advantage for transposition over, for example, selective amplification.

In fact in studies with *deaf* subjects there appears to be only one reported positive result out of a host of studies with tape slowed material (Bennett and Byers, 1967) and with 'on-line' techniques only Guttman, Levitt and Bellefleur (1970) and Velmans (1975) have reported an advantage for transposition. Using a type of frequency division, Guttman, *et al.* (1970) found a slight advantage in the articulation of |s| and a 'substantial' advantage in the articulation of |ʃ| (compared to controls) after experimental subjects had been given approximately 70 sessions of individual training (20 minutes per session). On the basis of this study therefore, one might expect a slight advantage for the frequency division transposition technique used, to manifest itself after giving a child one 20 minute training session, five days a week, every week for three months.

Faster improvements with transposition have also been reported in a variety of relatively *uncontrolled* studies, e.g. in a number of studies using the 'Johansson transposer' (Johansson, 1966; Johansson and Sjögren, 1965; and Wedenberg, 1966). Typically these studies compare discrimination or articulation performance before and after a series of training sessions with a transposer. If improvement occurs, the learning effect is attributed to the transposition technique. The crucial objection to these studies is that intensive training of a well-motivated child is itself likely to produce learning (see e.g. Ling, 1968 and Velmans, 1975) so it is simply not possible to say at the end of such experiments whether the improvement which occurred is greater, for example, than one which would have occurred if only selective amplification had been used.

On the basis of the reported literature, therefore, the superiority of various modes of transposition over linear or selective amplification, in speech training, remains largely unproven (which is not to say, of course, that advantages for certain transposition techniques may not exist). Further, although a few trans-

---

** This paper was presented in a modified form to the conference on 'Communication Systems for the Hearing Impaired: Alternatives to Conventional Amplification', British Society of Audiology, April 1973.*

posers do exist in wearable form (e.g. the Oticon Tp 64) there appears to be no controlled study investigating the utility of any transposer when used as a hearing aid.

Various problems brought about by the use of inappropriate validation procedures are likely to be partly responsible for the generally inconclusive results (see Velmans, 1973, 1975, for a discussion). In addition to such factors, however, it will be argued here that the lack of success with the majority of transposing devices is likely to have been partially due to the recoded speech signals having been processed as 'non-speech' by the ear-brain system. Evidence for the corollary of this statement, i.e. that a successful transposer must produce an output that is 'speechlike', converges from three directions.

Firstly there is evidence that at some stage of the decoding process, the ear-brain system may separate those sounds which are speechlike from those which are not. For example, studies by Kimura (1961, 1964), Bryden (1965) and Curry (1967) indicate that under certain conditions, the dominant hemisphere is superior in the recognition of speech sounds, whereas the non-dominant hemisphere is superior in the recognition of non-speech sounds, e.g. melodies or environmental noises.

Secondly, the speed and orderliness of speech acquisition seems to require the postulation of powerful innate learning mechanisms, biologically programmed to extract the rules of speech (and language) from the body of speech to which the child is exposed (see e.g. Lenneberg, 1967, for a review of the evidence).

Thirdly, there is the fact that in the auditory modality, the speech code is far more efficient than any non-speech code yet devised.

For example, Liberman, *et al.* (1967) report a rather revealing parallel to the lack of success with 'recoded' speech for the deaf, in the attempt to construct reading machines for the blind, which operate by converting individual printed letters to a sound code (allowing one to 'read' with one's ears); in spite of around 50 years of experimentation with a variety of sound codes, the best auditory decoding speed achieved by practised subjects is little more than one-tenth the rate at which they can decode speech, in fact about the same speed achieved by a skilled morse code operator (a similar limitation may apply to 'recoded' speech which is not 'speechlike').

The implication of these three lines of evidence taken in conjunction, is that in order that recoded speech may achieve an efficiency approaching that of normal speech, the recoded signals must be, not only discriminable, but also accepted by the ear-brain system as being speechlike. Only if this is the case, will the recoded speech be able to activate the highly efficient learning and decoding mechanisms involved in normal speech perception (a similar conclusion is arrived at by Roworth, 1970).

Having roughly delineated the state-of-the-art (as reported in the literature) and suggested one major change of approach, I will now (a) outline briefly the sorts of theoretical questions to which it would be useful to have answers if one were to design some 'ideal' recoding system, (b) summarise some of the design principles for transposition devices which it seems reasonable to propose, on the basis of what is already known, and (c) give a brief description of the design and present state of evaluation of the Frequency REcoding Device (FRED) with which I have been experimenting (Velmans, 1971, 1973, 1975).

The Ideal Recoding Device

We might reasonably propose that an 'ideal' recoding of speech would provide all the 'essential' cues of normal speech in the low-frequency residual hearing range, in such a way that the ear-brain system could decipher the recoded cues with the same efficiency as it does normal speech. However, before we could begin to specify the design of such an 'ideal' device we would have to answer the following questions regarding the operation of speech decoding:

(i) What is the set of so called 'essential' speech cues (indeed does such a set exist)?

(ii) What properties of the speech code and/or the speech analysing mechanisms give the speech code its unique efficiency?

(iii) Precisely how do different conditions of deafness impose limitations on the speech analysing system?

(iv) Which recodings (or transforms) of speech would make the best possible use of whatever analysing capacity remains?

Unfortunately we are a long way from being able to answer these very general questions. There are, however, a number of principles which it seems reasonable to propose as provisional guidelines for design (given our present state of ignorance). These are briefly outlined below.

Some Interim Design Principles for Recoding Devices

1. In the absence of a defined set of 'essential' speech cues, partial solutions should be attempted using recoded forms of cues which are likely

to provide important information but which are otherwise inaccessible, e.g.

(i) Devices such as the Johansson transposer and the FRED concentrate on the high frequency components of certain fricative, sibilant and stop consonants which provide important cues but which are poorly perceived in a variety of deafness conditions.

(ii) Vowel information is usually more robust than consonant information, but in cases where even vowel perception is impaired, some recoded form of the second formant may eventually provide some improvement (initial experiments in this direction have, for example, been carried out by Thomas and Flavin, 1970).

2. In the light of evidence for a specialised speech decoder in man, when mapping a given set of cues onto the residual hearing range, transforms should be chosen which have some likelihood of preserving (in some sense) the 'speechlike' nature of the signals (one approach to this is described in the 'Design of the FRED', below).

3. Except in very severe deafness cases, normal speech information provided by bone conduction and residual hearing should form an integral part of the recoding system, and should not be interfered with by recoded information. The reasons for this are, firstly, that if recoding were to supplant such information (e.g. if one were to provide a surrogate voice fundamental) it would be difficult to *mask* the information provided by bone conduction. Secondly, and more importantly, the utility of residual low frequency information is not in question, whereas the utility of various modes of recoding *is* in question. Finally, if recoded information does not interfere with existing low frequency information, then although we cannot be certain that recoding will improve intelligibility, it has at least reduced the likelihood that recoding will deteriorate intelligibility.

4. It should be possible for the normal hearing ear to discriminate amongst the recoded cues either immediately or after a short period of familiarisation (this can be established in simulated deafness studies); if this is not the case we have little reason to believe that the impaired ear will be able to learn such discriminations.

The operation of the FRED, which was designed with these criteria in mind, is outlined below.

Design of the FRED

Initially the recoding mode was designed to alleviate that condition of classical sensory-neural hearing loss, where only low frequency sensors, e.g. those mediating frequencies below 1kHz, are operating and important high frequency information, e.g. that relating to fricatives and sibilants, is totally lost. The function of the recoding process was to lower the frequency of cues relating to certain fricative, sibilant and stop consonants and to combine the recoded high frequency information with whatever information was already available in the low frequency region.

Interference between the recoded high frequency information and the low frequency (mainly vowel) information was avoided, by simply selecting the speech band 4000Hz and above, for recoding down to the low frequency region, the reason being that whereas this high frequency region is rich in fricative and sibilant energy, it contains negligible vowel energy; therefore, when this band is superimposed on the low frequency region, the high frequency information *slots into* the relatively empty consonant spaces around the vowel. This principle is demonstrated in Figure 1.

Of course, recoded consonants will be different to some extent to non-recoded consonants when perceived, e.g. by a normal hearing person. However, an attempt was made to maintain both the discriminability and the speechlike nature of the recoded signals by substantially preserving the spectral shape or pattern of the signals in the recoding process (where the spectral shape is defined by the 'difference frequencies' and relative intensities of the spectral components). This was done by simply subtracting a constant frequency of 4kHz from each frequency component in the region 4kHz and above, thereby mapping the high frequency region onto the region zero Hz and above (see Velmans, 1971, 1973 for details).

Whether the recoded components were in fact discriminable and in some sense speechlike, were questions investigated in a subsequent simulated deafness study (Velmans, 1973) and a validation study with deaf children (Velmans, 1975).

The simulation study assessed the effect of recoded high frequency information, on the ability of normal hearing subjects (under simulated deafness conditions) to imitate CVC nonsense syllables spoken by the experimenter, in a face-to-face, individual, articulation imitation task. It was found that recoding improved the imitation of both 'manner' and 'place' of articulation of consonants with high frequency components in the region 4kHz and above (e.g. |s|, |ʃ|, |tʃ| and |tr|) without deteriorating the imitation of phonemes with major energy com-

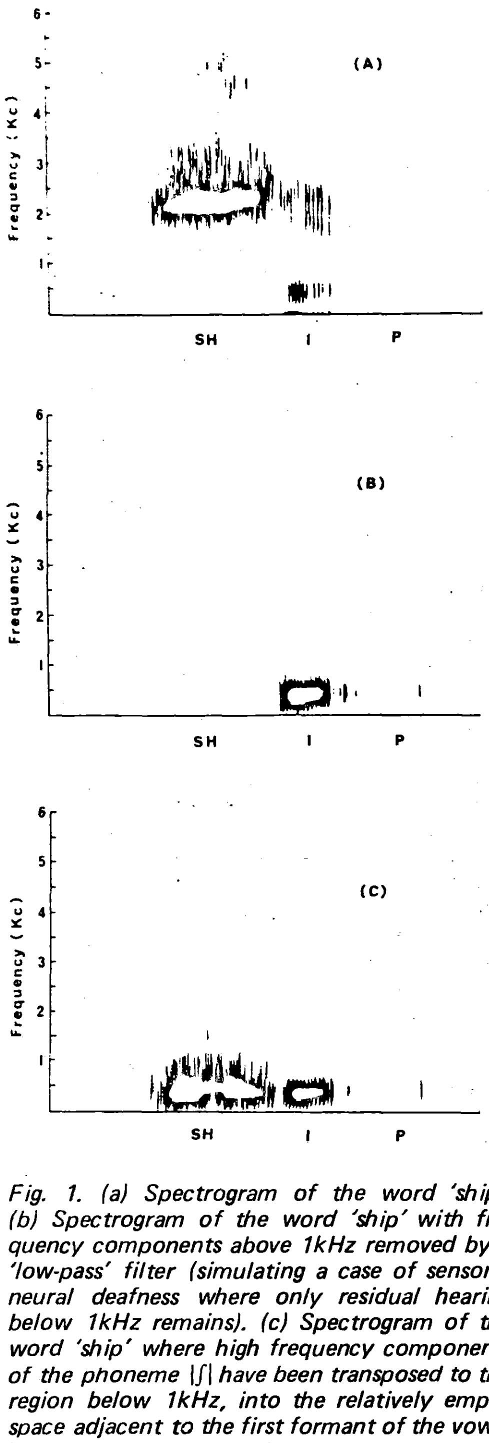


Fig. 1. (a) Spectrogram of the word 'ship'. (b) Spectrogram of the word 'ship' with frequency components above 1kHz removed by a 'low-pass' filter (simulating a case of sensory-neural deafness where only residual hearing below 1kHz remains). (c) Spectrogram of the word 'ship' where high frequency components of the phoneme |ʃ| have been transposed to the region below 1kHz, into the relatively empty space adjacent to the first formant of the vowel |i|, thereby avoiding elimination by the 'low-pass' filter, and now in principle detectable to the deaf ear with residual hearing below 1kHz only.

ponents in the low frequency region (e.g. the vowels). This improvement, occurring without formal discrimination or imitation training with the recoded stimuli, was taken to be evidence of the discriminability and the speechlike nature of the recoded stimuli, and the utility of the recoding to normal hearing subjects as sufficient grounds for carrying out further studies with the deaf.

Accordingly a validation study assessing the effects of recoding on the articulation learning of consonants with major high frequency components (i.e. |s|, |ʃ|, |tʃ|, |t| and |k|) was carried out with a group of six sensory-neural deaf children, placed in a face-to-face, individual, articulation training situation. It was found that recoding, combined with selective amplification and amplitude compression, led to significantly better articulation learning of the high frequency consonants used, than selective amplification and amplitude compression only, the improvement being manifest within only seven training sessions of five to ten minutes each (under recoding and no-recoding conditions respectively).

The improvements brought about by recoding in these two studies are encouraging; however, extensive further evaluation is required before an assessment can be made of the generality of utility of the recoding mode employed. In addition to replication studies, evaluation studies should be carried out with deaf adults as well as children, and hearing aids incorporating a FRED circuit need to be built before the effect of continuous exposure to the recoded speech can be examined. One surprising informal observation is that when fricative energy already exists in the low frequency band, e.g. when residual hearing extends up to and above 8kHz (as in the normal ear), the high frequency fricative energy, recoded by the FRED, combines with the existent fricative energy (i.e. the complete fricative spectrum) to produce a consonant that appears little changed from the original (such a blending of cues does not appear to occur for recoded and non-recoded vowel information). This effect requires formal investigation as it opens up the possibility that recoding high frequency consonant information to the low frequency region may also provide an alternative to those who can discriminate amplified fricatives, but only when the high frequencies are amplified to the point where discomfort occurs (e.g. as in certain recruitment cases).

Finally, in addition to the immediate questions outlined above, there are many questions of a more fundamental nature regarding the operation of normal and impaired speech decoding mechanisms to which we require answers (see e.g. 'The Ideal Recoding Device', above) as it is to these answers that general progress in the design of speech recoding devices may ultimately be linked.

References

Bennett, D. N. and Byers, V. W. (1967) Increased intelligibility in the hypacusic by slow-play frequency transposition. *J. Aud. Res.*, **7**, 107-119.

Bryden, M. P. (1965) Tachistoscopic recognition, handedness, and cerebral dominance. *Neuropsych.*, **3**, 1-8.

Curry, F. K. W. (1967) A comparison of left-handed and right-handed subjects on verbal and non-verbal dichotic listening tasks. *Cortex*, **3**, 343-352.

Guttman, N., Levitt, H. and Bellefleur, P. A. (1970) Articulatory training of the deaf using low-frequency surrogate fricatives. *J. Speech Hearing Res.*, **13**, 19-29.

Johansson, B. (1966) The use of the transposer for the management of the deaf child. *J. Internat. Audiol.*, **5**, 362-371.

Johansson, B. and Sjögren, H. (1965) The effect on normal hearing subjects and profoundly deaf children by coding voiceless consonants using frequency transposition. *Proceedings of the Fifth International Congress on Acoustics*. Liege, paper B69.

Kimura, D. (1961) Cerebral dominance and the perception of verbal stimuli. *Canad. J. Psychol.*, **15**, 156-165.

Kimura, D. (1964) Left-right differences in the perception of melodies. *Quart. J. Exp. Psychol.*, **16**, 355-358.

Lenneberg, E. H. (1967) *Biological Foundations of Language*. New York, Wiley.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P. and Studdert-Kennedy, M. (1967) Perception of the speech code. *Psych. Review*, **74** (6), 431-461.

Ling, D. (1968) Auditory discrimination of coded speech by deaf children. Ph.D. dissertation, Institute of Otolaryngology, McGill University, Montreal.

Risberg, A. (1969) A critical review of work on speech analysing hearing aids. *IEEE Transactions on Audio and Electro acoustics*, AU-17, (4), 290-297.

Roworth, D. A. A. (1970) The recoding of speech, *Sound*, **4** (4), 95-105.

Thomas, I. B. and Flavin, F. E. (1970) The intelligibility of speech transposed downward in frequency by one octave. *J. Audio. Eng. Soc.*, **18** (1), 56-62.

Velmans, M. (1971) Aids for deaf persons. British Patent Application No. 19476/71.

Velmans, M. (1973) Speech imitation in simulated deafness, using visual cues and 'recoded' auditory information. *Language and Speech* (In Press).

Velmans, M. (1975) Effects of frequency 'recoding' on the articulation learning of perceptively deaf children. *Language and Speech* (In Press).

Wedenberg, E. (1959) Auditory training of the severely hard of hearing using a coding amplifier. *Proceedings of the Third International Congress on Acoustics*. Stuttgart, **2**, 658-660.