Proceedings of The Institute of Acoustics

A SPEECH TRAINING AID FOR DEAF CHILDREN EMPLOYING TRACKING FILTERS FOR FORMANT EXTRACTION

J A S Angus(1) and M D Kendall(2)

(1) Electronics Department, University of York, York
(2) Racal Research Ltd., Reading

INTRODUCTION

The deaf can find integrating into society difficult because of the problems they have with speech communications. These problems include such aspects as poor speech quality, which can often be interpreted (unfairly) as indicating mental disability, and lack of response to normal speech protocols, which is often misinterpreted as standoffish by normal hearing people.

In short deafness strikes at the heart of communication and although sign language can assist in breaking down this barrier it only allows communication between a subset of all people.

Because deaf people have no acoustic feedback they find controlling speech sounds difficult as they are effectively operating "open loop" in control theory terms.

There have been attempts to close the auditory feedback loop by communicating relevant speech parameters via tactile, visual and electrical means for both prosthetic and training purposes however these aids suffer generally from two disadvantages:

1) They tend to be expensive which often puts them out of the financial reach of specialist schools and NHS supported speech therapists let alone private individuals.

2) The design of the training aid often assumes a certain minimum mental ability and/or stage of language development which tends to restrict their use to older children and adults.

However a crucial time in language development is the age range 0 - 5 years (approx). During this time a child with normal hearing is doing two things (relevant to speech communication):

1) The child is making noises (which it can hear) and experimenting with its vocal apparatus. The child is also observing the effect these noises have on other people.

2) The child is listening to its acoustic environment and learning about different sounds and often trying to mimic them (car noises, birds, words). The child is also associating these sounds with particular objects and concepts.

A SPEECH TRAINING AID FOR DEAF CHILDREN

But a deaf child will in contrast:

a) Receive little stimulation or reward for making a noise (it can't hear them).

b) Be unable to listen to its acoustic environment.

These factors compound to make language acquisition difficult for a deaf child and although currently available training aid can assist when the child is older. The acquisition of language at that age is harder because the brain appears to be less able to learn language when the child is older.

Therefore there is a need for a speech training aid which will provide both stimulation and useful visual feedback to a young child. The rest of this paper describes a prototype speech training aid aimed at young children in particular it will discuss; The speech parameters which must be extracted and displayed, and then describe a cost effective implementation and its performance. The paper will conclude with suggestions for further work.

SPEECH PARAMETER EXTRACTION

Speech can be modeled as a waveform generated by the structure shown in Figure 1. In this model speech is considered to be the result of two types of input waveform, glottal fold and noise due to turbulent airflow, whose spectrum is modified by the time varying acoustic filters formed by the vocal tract (mouth, tongue, lips etc.).

From this model one can argue that relevant speech parameters are:

1) Pitch and Amplitude of the speech waveform

2) Voicing i.e. whether the excitation is periodic, noise or a mixture.

3) Frequency, Amplitude and Bandwidth of the major resonances of the vocal tract (formants)

In English the combination of the position of the major resonances in conjunction with voicing separate out the basic speech sounds for the most part whereas the pitch and amplitude variation with time affects the stress and mode of the utterance.

Both aspects of speech are important and ideally a training aid should be able to present information about rythm and intonation as well as formant and voicing parameters.

However we decided to concentrate our initial work on establishing the positions of the formants as a first step.

The mouth can be considered to be two independently controllable cavities so one has independent control over the two lowest major resonances F1 and F2 (see fig 2) however because it is a distributed system there will also be higher order resonances F3 F4 ....(see fig 2) which will also vary but are not independently controllable. However the higher order resonances can be used to give some information about the length of the vocal tract and so provide a degree of child/adult normalisation. In view of this we decided to extract F1, F2 and F3.

FORMANT EXTRACTION TECHNIQUES

Algorithms exist for extracting formant information in speech [Ref 1] but they require large amounts of computing power for their implementation and so where unsuitable for a cost effective system.

A simpler approach is to examine the time domain waveform of speech as shown in Fig 3. From this we can observe that Average Zero crossing rate of the waveform is a function of the formant frequencies but it is affected by the presence of more than one formant in the signal. If one can separate the signal into single formants then on the waveform of a single formant is a simple damped sinusoid shown in Fig 4 whose Zero crossing rate is a function of the formant frequency.

Therefore if one could separate at these formants then a simple frequency measurement could extract the relevant formant frequencies. However if one examines the typical range of formant frequencies (see Fig 5) one sees that fixed filters cannot separate the formant frequencies because the frequencies overlap.

Watanabe [Ref 2] suggested a cross coupled connection of tracking notch filters to track the formant (Fig 6) . This system works because each channel notches out the interfering formant frequencies thus leaving only one per channel. Thus the formant frequency in each channel can be estimated and used to control the notch filters in the other two channels. However this system suffers from two problems, firstly any error in formant estimation results in considerable interference and secondly the system can "turn turtle" as there is no way of enforcing the allocation of formant frequencies to a particular channel.

We have implemented a similar idea but instead have used high pass and low pass notch filters (see Fig 7) so that the system cannot "turn turtle" and has some resiliance to the (inevitable) occasional mistracking of formants.

The system also uses the fact that switched capacitor filters (SCF) have their cut-off frequency controlled by their clock frequency to realise a simple, and potentially integratable tracking filter system.

A SPEECH TRAINING AID FOR DEAF CHILDREN

SYSTEM IMPLEMENTATION

A block diagram of the system is shown in Fig 8. The tracking filters were realised using the MF10 (SCF) and the CD4049 phase locked loops were used to extract the formant frequencies. The control voltages of the VCO's were fed into audio Bargraph display drivers which then drove an LED matrix.

The voltages of the VCO's were held when the loops were unlocked so that the display was consistent across stops. Normalization was simply achieved by using the F3 VCO control voltage as one of the reference voltages of the Bargraph display chips.

RESULTS

The system could track formants in real time but produced a jittery output however with some smoothing a pleasing single dot display was produced. The internal waveforms of the formant tracker are shown in Figure 9 and from this one can see that F1 and F3 are extracted reliably despite the fact that F2 has some F1 interference on it. F2 is extracted less reliably due to the fact that it only used notch filters to get rid of F1 and F3.

Children have used the system and find it fascinating despite its simplicity.

FURTHER WORK

The tracking performance could be improved by using frequency locked loops, to remove jitter caused by pitch pulses and better filtering in the F2 channel. Ideally the display needs a greater resolution. The facility to display a track and an organisation which relates in a better way to the position of the articulators, e.g. the vocal trapezium.

It is also necessary to address the problems of rythm and intonation, all these aspects are currently under investigation.

CONCLUSION

It is possible to develop cost effective formant tracking hardware using analogue techniques. This provides the possibilities of a speech training and stimulation aid for deaf young children which is both affordable and exciting to use.

REFERENCES

1. Markel, J. G., and Gray, A. H. 1976. "Linear Prediction of Speech". (New York: Springer Verlag).

2. Watanabe, A. 1979. "A Real Time Formant Tracker Using Inverse Filters". STL-QPSR, 3-4/1979 p1 - 30.

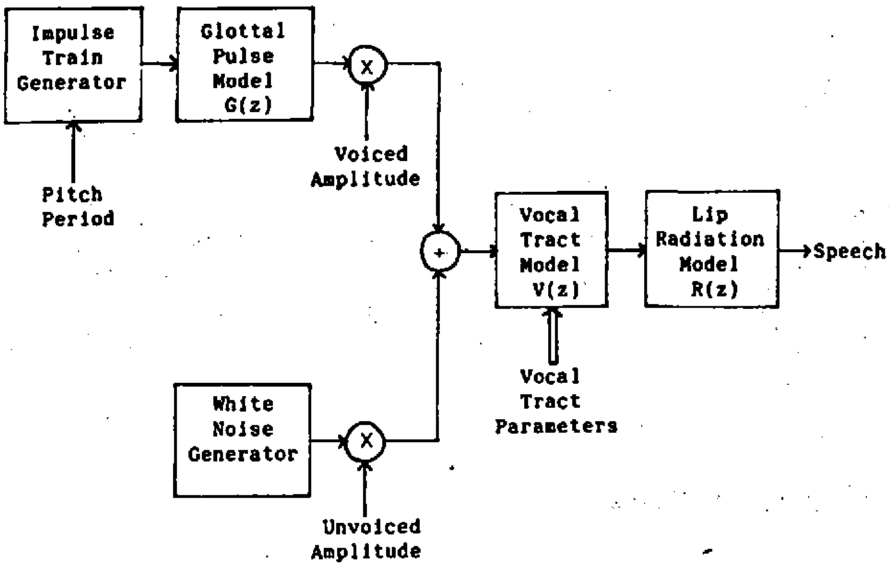

Fig. 1 Filter Based Speech Model

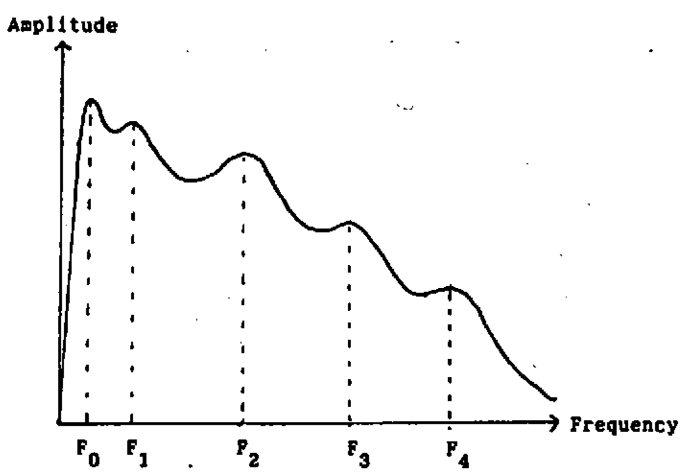

Fig. 2 Formant Frequencies

A SPEECH TRAINING AID FOR DEAF CHILDREN

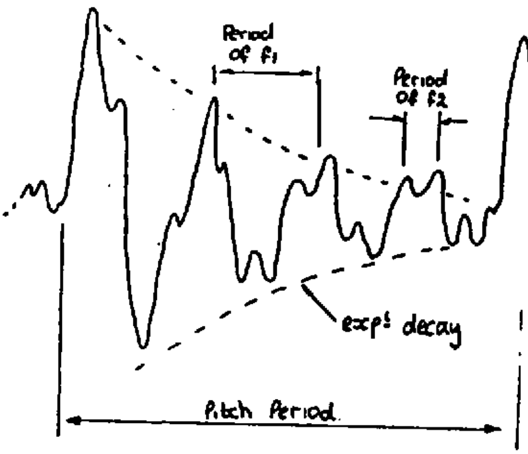

Fig. 3 A Typical Speech Waveform

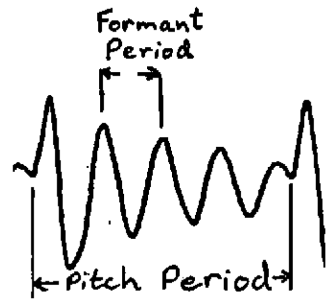

Fig. 4 A Single Formant Waveform

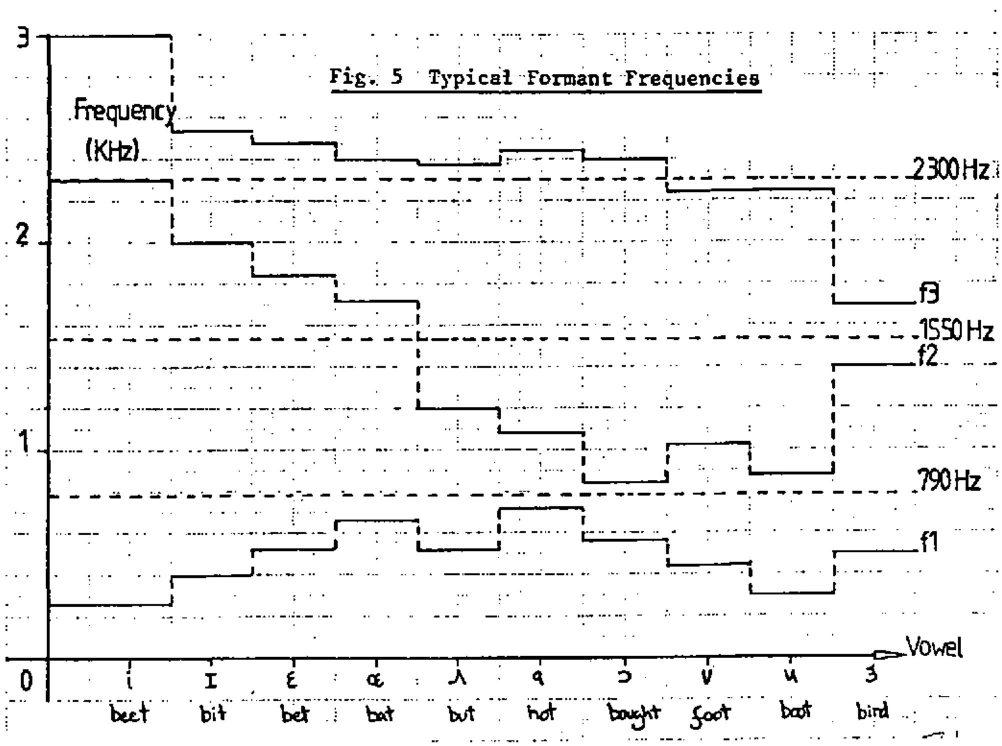

Fig. 5 Typical Formant Frequencies

A SPEECH TRAINING AID FOR DEAF CHILDREN

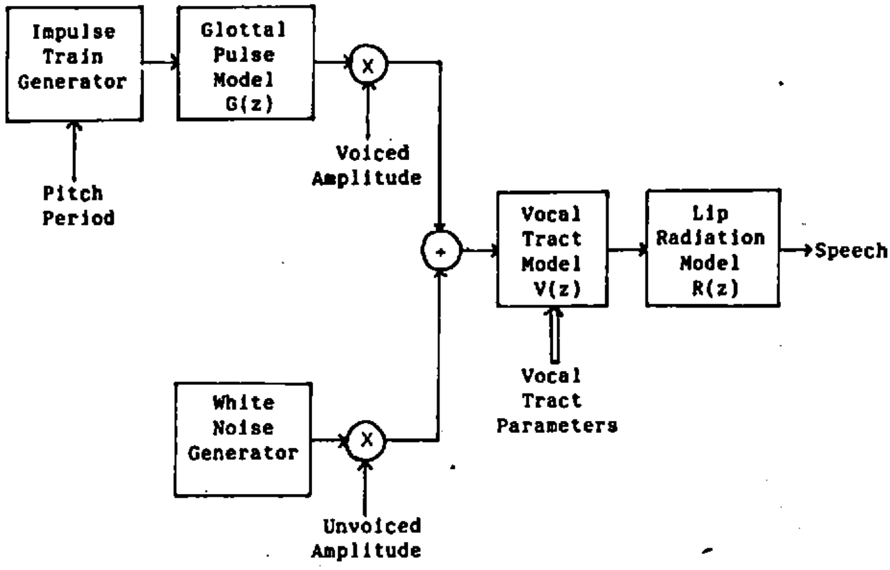

Fig. 1 Filter Based Speech Model

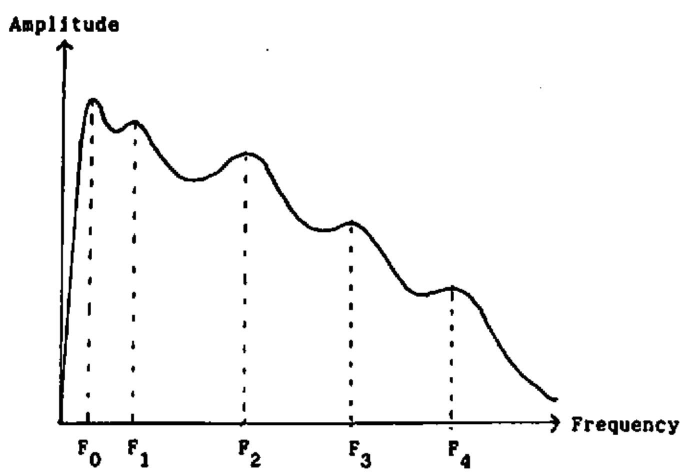

Fig. 2 Formant Frequencies

A SPEECH TRAINING AID FOR DEAF CHILDREN

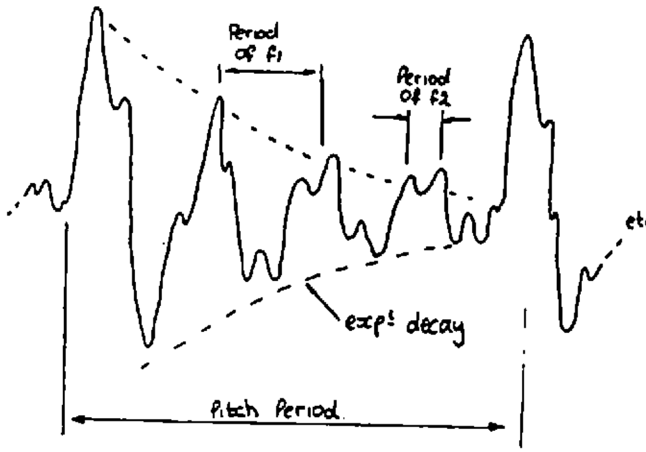

Fig. 3 A Typical Speech Waveform

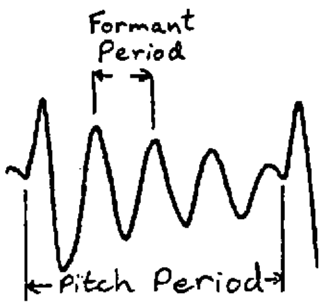

Fig. 4 A Single Formant Waveform

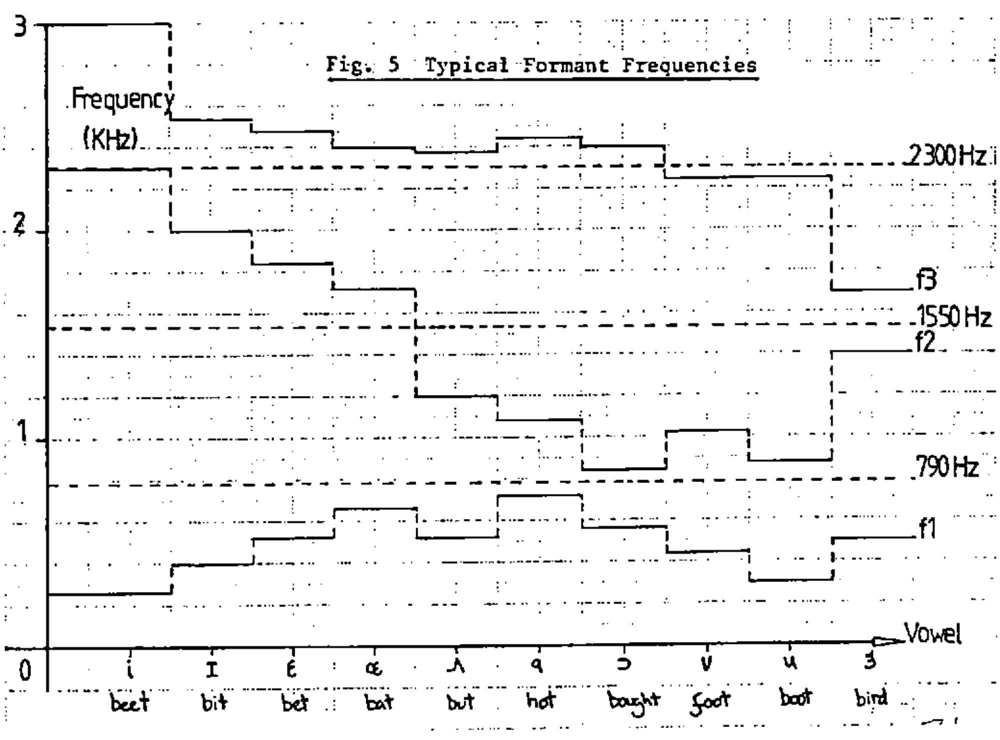

Fig. 5 Typical Formant Frequencies

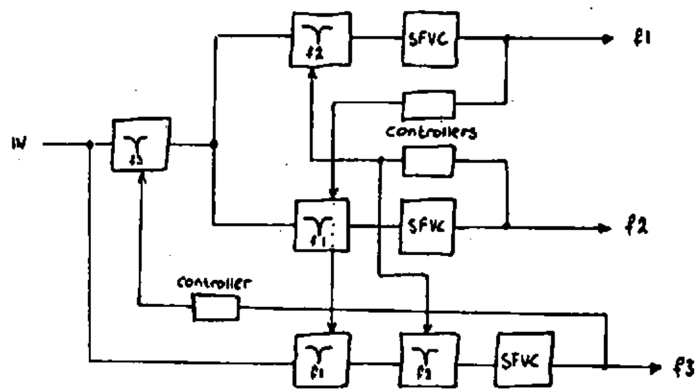

Fig. 6 Watanabe's Formant Tracker

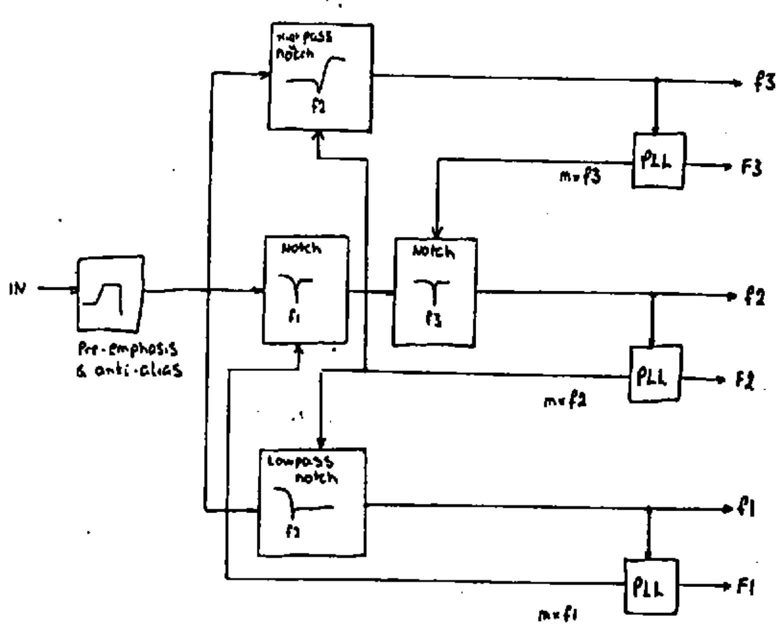

Fig. 7 The Author's Formant Tracker

A SPEECH TRAINING AID FOR DEAF CHILDREN

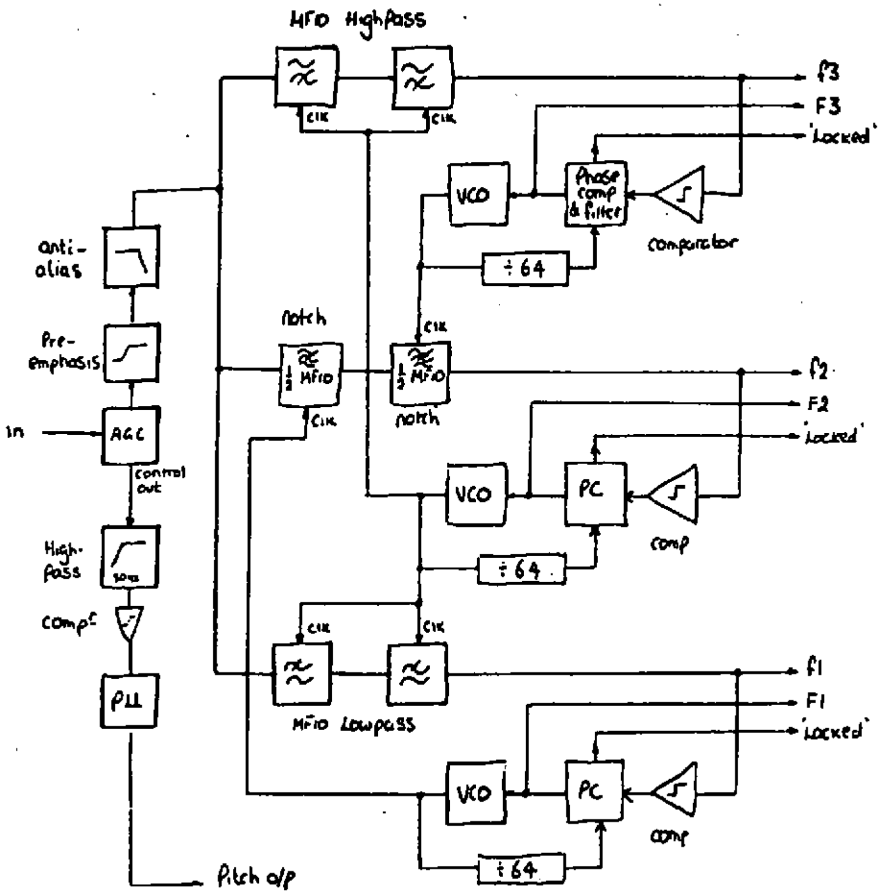

Fig. 8 The System Block Diagram

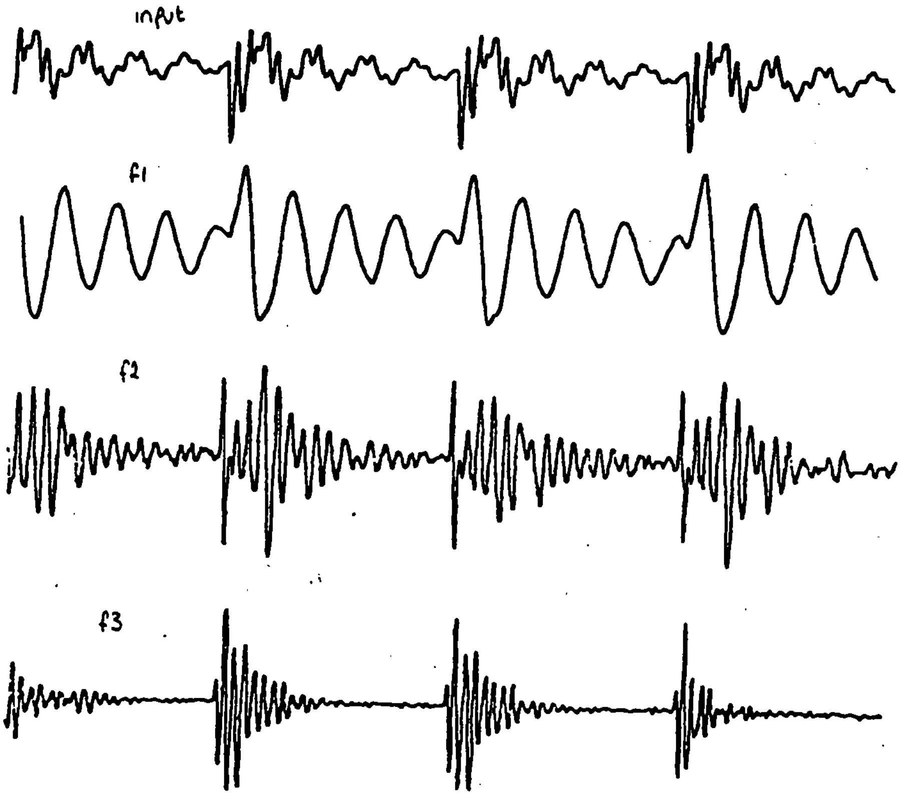

Fig. 9 Typical Output Waveforms