SIGNAL RECOGNITION

John R. Naylor

Department of Applied Acoustics, University of Salford,
Salford, M5 4WT

INTRODUCTION

Speech Recognition has received a great deal of attention in the past. Research in this area has primarily been concerned with recognising specific utterances, usually for a limited group of speakers. The subject of signal recognition, by contrast, is relatively untouched. It is concerned with recognising signal types. This paper describes work done as part of a final year project to develop a system by FORTRAN simulation which would recognise a signal as either speech or "noise" where noise refers to anything which is not speech. It is a continuation of the work done by J. Parker [1] who developed a squelch system that relied on recognising signals as speech or noise for its operation. The recognition was based on the autocorrelation function of the signals which was realised by analogue techniques.

Modern digital signal processors (DSP's) have such facilities as 16 bit word size, hardware multiply/accumulate and fast instruction rates enabling many traditionally analogue operations such as filtering and correlation to be implemented digitally.

Were a signal recognition system required to be built it would, almost certainly, find a DSP based realization. The first step in the development would be to investigate the techniques available by simulation - an approach which has many advantages.

- Algorithms are easier to develop in high level languages.
- If something does not work in simulation it will never work in practice thereby saving effort in the wrong direction.
- The approach can be changed quickly as one is not tied to specific hardware.
- Most of the bugs can be removed in simulation leaving fewer expensive iterations to arrive at a working hardware.
- Simulations only cost as much as the computer time used.

The work discussed concerns a digital implementation of the autocorrelation function and the properties of this which make it a good tool for signal recognition.

THEORY

Autocorrelation

The autocorrelation function is defined as:

$$R_x(\tau) = \lim_{T \to \infty} \frac{1}{T} \int_{t=0}^{T} x(t)\, x(t+\tau)\, dt \qquad (1)$$

A direct conversion of this into the discrete domain yields:

$$R_x(k) = \frac{1}{N} \sum_{\substack{n=o \\ N \to \infty}}^{N-1} x(n)\ x(n+k) \qquad (2)$$

This is clearly not realizable as N must be finite and preferably small (<100) if a practical implementation is to be achieved. The minimum value of N complete is given by

$$N = \text{integer part of } \left(\frac{T\ell}{Ts} + 1\right) \qquad (3)$$

where Ts is the sample period and Tℓ is the period of the lowest frequency of interest. At 10kHz sample rate and only considering the telecommunications Bandwidth (300Hz - 3.3kHz) N = 34.

Limiting N in this manner causes a bias to be present in the autocorrelation function. This may be removed by using one of the following equations.

$$R_x(k) = \frac{1}{N-k} \sum_{n=o}^{N-1} x(n)\ x(n+k) \quad k < K \qquad (4)$$

$$R_x(k) = \frac{1}{N} \sum_{n=o}^{K} x(n)\ x(n+k) \quad \begin{matrix} k \leq K \\ N = 2K \end{matrix} \qquad (5)$$

where K is the largest delay of interest.

Equation (4) is arrived at by consideration of the expression for the standard deviation of a population and is taken from [2]. Equation (5) is an expression of the Modified Autocorrelation Function [3]. It was equation (5) that was implemented as it avoided the division by a variable, a time consuming operation.

The properties of the autocorrelation function pertinent to this application are:

(i) The positioning of the peaks (and hence zero crossings) depends on the fundamental frequency of the signal.

(ii) The autocorrelation of random noise decays with increasing delay. Thus that of a signal in noise is close to that of the signal alone at longer delays. The noise has been filtered out.

Signals Considered - Basis for Recognition

It was found that speech could be recognised as different from random noise and FSK data waveforms by the frequency behaviour of the signals.

The fundamental pitch of speech wanders around, apparently aimlessly whereas FSK takes one of two frequencies. This is illustrated in figures 1 and 2.

The frequency behaviour of signals was observed by the number and location of the zero crossings in successive autocorrelation functions spaced 60ms apart. The autocorrelation function of some selected signals are illustrated in figures 3, 4 and 5. Other spacings were used but 60ms was found to be the best [4]. From the following tests on the two autocorrelation functions a very accurate recognition algorithm was written.

Recognition Algorithm

(i) Frequency test - if the number of zero crossings in both autocorrelations is greater than 6 the signal is not speech. Otherwise apply (ii).

(ii) Pitch shift test 1 - if the number of zero crossings are the same in each autocorrelation function the signal could be a low frequency sinusoid. Apply test (iii). If the number of zero crossings are different then the pitch has changed. The signal is speech.

(iii) Pitch shift test 2 - check if the zero crossings have moved. If they have the signal is speech otherwise it is not speech.

EXPERIMENT

The algorithm was coded as a subroutine. It was presented with autocorrelations calculated using equation 5 to test its response to various signals as follows:

- 1.4 s of the author's speech sampled at 10kHz and stored as a file.
- FSK waveforms at different baud rates. The data used was a pseudo random binary sequence [5].
- Random noise lowpass filtered at 3.3kHz.
- The first two at various signal to noise ratios.

Results

Conditions	% Correct Decisions
100 tests made at separate points in the speech file.	100
100 tests made at separate points in band limited random noise	91
100 tests made on noise free FSK data at: 75 baud	100
150 baud	100
300 baud	100
600 baud	100
1200 baud	100

The results for speech and FSK data against the signal to noise ratio are given in figure 6.

CONCLUSIONS

The speech signal is reliably recognised by its autocorrelation function even at poor signal to noise ratios.

FSK data is invariably recognised as not speech regardless of baud rate and SNR.

One time in ten random noise would be mistaken speech.

DISCUSSION

It is felt that if the system were implemented in hardware and the same experiments made, there would be a high correlation of measured and predicted performance. This is because a digital system (PRIME computer) is simulating the operation of another digital system (DSP). The calculations made would be identical. Any differences to arise would be due to the limited precision (integer arithmetic) of the DSP.

Ommissions made necessary by the lack of time available were experiments based on different data formats, ASK and PSK for example, although these were taken into consideration when the algorithm was written and experiments based on different speakers, female particularly.

REFERENCES

1. J. P. Parker. "HF SSB Squelch".
   Master of Science Thesis, Birmingham University, May 1982.
2. Schwartz and Shaw. "Signal Processing, Discrete Spectral Analysis, Detection and Estimation".
   McGraw Hill, 1976.
3. Rabiner L. and Shafer. "Digital Processing of Speech Signals".
   Prentice Hall.
4. J. Naylor. "Signal Recognition".
   Batchelor of Science report, Department of Applied Acoustics, University of Salford.
5. H.J. Beker. "Communications Security - A Survey of Cryptography".
   Proc. IEE Vol. 129, pt A. No. 6, August 1982.

Fig 1 Pitch Variation of Speech

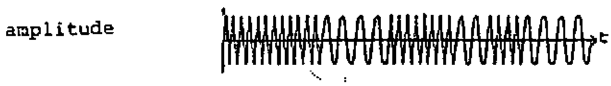

Fig 2 Pitch variation of FSK Data

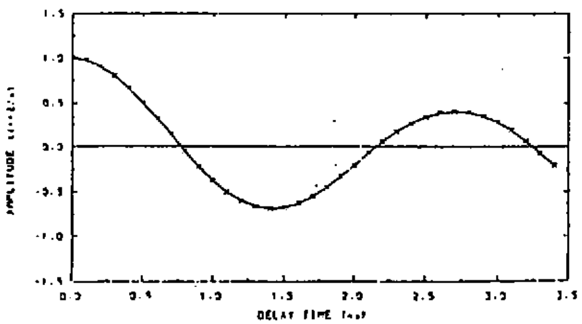

Fig 3 Autocorrelation of Voiced Speech

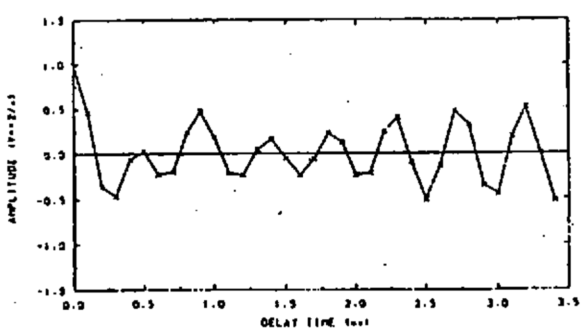

Fig 4 Autocorrelation function of FSK Data

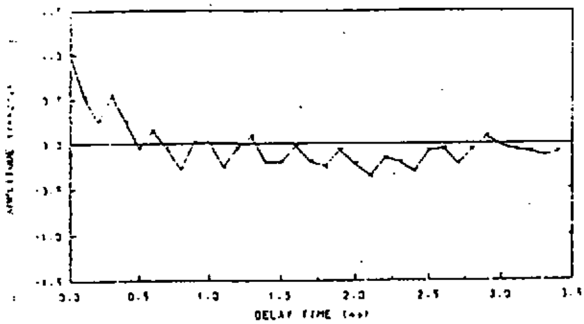

Fig 5. Autocorrelation Function of Random Noise

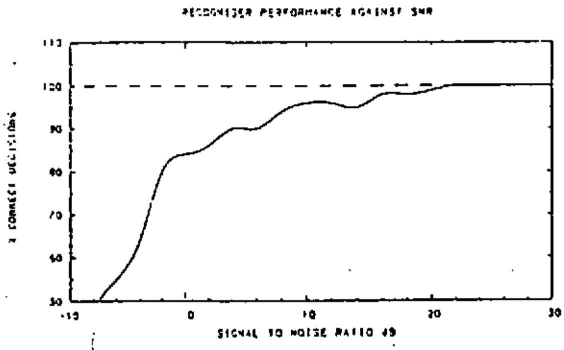

Fig 6