A Speech Analyser based on the IBM Personal Computer

P R Alderson, G Kaye, S G C Lawrence, D A Sinclair

Speech Research, IBM UK Scientific Centre,
Athelstan House, St Clement Street, WINCHESTER, SO23 9DR

Introduction

This paper describes an experimental tool suitable for use in a wide range of speech research and signal processing applications. It was designed as part of the very-high-quality speech synthesis project at the IBM UKSC. In building such a system detailed properties of human speech must be determined, and this involves extensive analysis of human utterances. Additionally it is necessary to analyse the synthesised speech to enable those properties that contribute to intelligibility and acceptability to be quantified. Another requirement is for a convenient method of preparation of material for experiments in speech perception. The following system has been designed to provide comprehensive speech analysis and signal processing capabilities, whilst at the same time being accessible by linguists, phoneticians, psychologists and speech therapists with no computational expertise, and the minimum of training.

SAY - Speech Analyser Functional Description

The UKSC Speech Analyser (SAY) is a laboratory system for the:

Digitisation
Editing
Analysis
Storing
Replaying

of speech signals. Since all of its facilities, except for the actual speech acquisition and replay, can be applied to any one dimensional digital signal the system could usefully be used for a wide range of non-speech applications eg seismic processing, time domain reflectometry, vibrational analysis.

To accommodate the varied user group, the user interface balances the flexibility needed by the computer professional with the ease of access appropriate to a linguist or other person with less computational experience. This was achieved by adopting a largely menu driven system and by the provision of a special control panel containing dedicated controls for cursor functions, waveform scrolling and speech editing and playback. The more experienced user may expand the basic functions by the entry of single commands, or command lists using a *full screen* command mode, while still retaining the dedicated control panel key functions where required. Many of the more sophisticated facilities are built up from basic operations using this command interface. These operations are provided by a library of system functions which can be easily extended by the user. In implementing these functions extensive use has been made of the host resident signal processing functions provided by IAX [1,2].

System Function and Components

The SAY speech analyser consists of the following functional components:

Analog audio recording equipment
Speech digitiser and playback
Waveform display and editor
Speech and waveform analysis system
Spectrum display
Speech synthesisers

A Speech Analyser based on the IBM Personal Computer

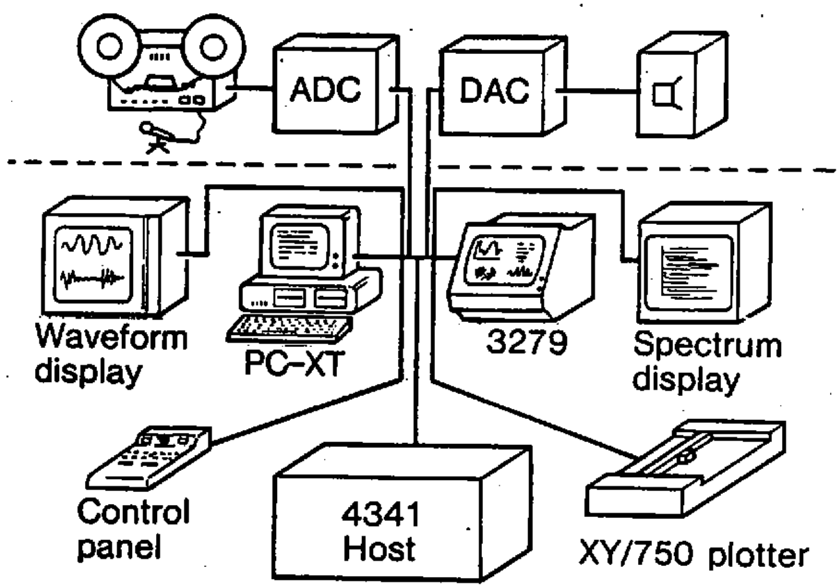

Figure 1. System Components

The system is controlled by an IBM PC-XT Personal Computer with Expansion Unit (figure 1). The component functions are:

PC-XT: This controls all digitisation, editing, wave storage and playback functions. It is also used for the entry of all system commands and for the display of guidance information as described below. It is connected to an IBM 370 host computer via a high-speed link.

Conversational monitor: This is the PC-XT monochrome display. It is used to display system menus and status information, such as the number of waves being edited, cursor positions, time of day. It also shows the system help information when requested, and provides guidance on the function key and control panel switch definitions.

Speech Digitisation: The speech sources are:

Microphone
Pre-recorded tape
Output from a synthesiser

The sampling rate and filter can be selected from the PC under the control of the user. The digitisation system uses DMA techniques which permit the recording and replay of speech samples with concurrent hard disk activity. This enables virtually unlimited lengths of speech to be acquired, processed and reviewed.

Playback is through an appropriate reconstruction filter selected by the system according to the original sample rate.

A Speech Analyser based on the IBM Personal Computer

Waveform Display: This is a high resolution X-Y directed beam vector display capable of displaying one or two 1024 point windows on to a speech waveform. It is used for the display and manipulation of waveforms. All the functions associated with its control are implemented in the PC. The viewing window may be scrolled along the waveform by the operation of a joystick on the control panel. The display has five levels of brightness (off, low, medium, high and blinking). These brightness attributes are useful for delineating sections of the waveform that are being processed (eg when they are deleted or copied).

Vertical zoom and horizontal zoom and compression are implemented using scaling factors of 2,4,8,16 and 32. It is very convenient to be able to display two waveforms simultaneously at different resolutions since this permits the user to have a global view of the speech waveform as well as a high resolution view of the particular area being worked with.

A pair of cursors is associated with the wave being edited. These may be manipulated from the control panel or under program control. They are used to mark certain sections of a wave for editing or replay purposes.

Control Panel: The control panel provides a means of scrolling waveforms and of moving cursors. It also provides single key access to the most commonly used editing, digitisation and replay functions. This means that the majority of editing activity can be performed under the control of one hand leaving the other free for taking notes, adjusting tape recorder controls etc. The panel consists of 16 keys, a one dimensional joystick and two tracker wheels with rotary encoders. The joystick is used for scrolling the active displayed waveform, the rate of scroll being proportional to the amount of joystick displacement. The tracker wheels are used to move the cursors horizontally on the waveform display.

Speech Editing: A full range of waveform editing functions is provided. Several waves can be involved in the editing process at any time. Sections may be freely copied and deleted from any wave.

The waveform editing facilities include:

- Delete a wave section
- Copy a section of a wave from one wave to another or within the wave
- Move sections of a wave within the wave
- Wave storage to hard disk
- Select a segment of a wave for subsequent processing e.g.
  - Compute the Fourier transform of the segment
  - Playback only the selected segments
  - Playback omitting the selected segments
  - Mark points in the wave for reference in subsequent processing

Spectrogram Computation and Display: Speech spectrograms can be computed on the host using IAX and then displayed on a raster graphics display or on the 3279 colour terminal. They can also be computed within the PC environment using dedicated Fourier transform hardware [3], and displayed on a high function raster display.

Graph Plotter: This is an A3 size, eight-pen IBM Instruments XY-750 digital plotter. Its support software runs on the PC-XT.

A Speech Analyser based on the IBM Personal Computer

Signal Processing: An extensive range of signal processing functions is available within IAX [1,2] at the host. The user can invoke these functions directly from the speech analyser. In the first version, the results are displayed on the 3279, while later versions display them on the wave and spectrogram displays attached to the PC. The command interface is via a modified form of the IBM PC Personal Editor. It provides both a command editor and a means of executing a single command or a command list. These commands may be executed locally within the PC or passed to the host processor.

IAX is a language for image and signal processing. An interpreter for the language has been implemented for IBM 370/303x/43xx machines. IAX expressions look similar to those of PL/I, Pascal or Fortran. However, the variables within these expressions can be images, vectors, numbers or strings. The IAX operators and functions operate on arrays as well as numbers. It is easy to write new functions for IAX either as macro functions or as compiled code. For example, there are IAX function primitives such as:

FT	Fourier Transform
IFT	Inverse Fourier Transform
LOG	Natural Logarithm
HAMM	Hamming Window
STIT	Unwrap phase
IMAG	Imaginary part (function and pseudo-variable)

Using such primitives a new function such as **Complex Cepstrum** (figure 2) can be written as a list of IAX commands. This new function is available as soon as the source file of IAX commands has been created. In fact the instructions may be executed interpretively as they are entered. Once filed the new function can be invoked directly by the user or as part of some higher level IAX function. The function can be changed by simply editing the source file which defines it.

Complex Cepstrum

```
Speech = Wave(Start:Start+255)
WindowedSpeech = HAMM(256)*Speech
FourierTransform = FT WindowedSpeech
LogFourierTransform = LOG FourierTransform
IMAG(LogFourierTransform) = STIT(IMAG(LogFourierTransform)
ComplexCepstrum = IFT LogFourierTransform
n = 21
Filter = ComplexCepstrum(0:n)||0(n+1:255-n)||ComplexCepstrum(256-n:255)
PowerSpectrum = LOG ABS FourierTransform(0:127)
VocalTractResponse = REAL(FT(Filter))(0:127)
```

Figure 2. IAX Function Example

In this way functions can be created to perform:

Cepstrum processing to extract the vocal tract response
Linear prediction to obtain vocal tract response
Spectrum generation
Pitch contour extraction

Help facilities: Help is provided for each function and facility in the system. It is displayed in either the lower or upper part of the conversational monitor as is most appropriate for the menu currently being displayed, and can be read whilst the user is completing the appropriate fields, or using the control panel. In general HELP is additive to what is currently displayed, and never overwrites a response field.

References

[1] P H Jackson, 'The IAX Image Processing Language', UK Science Centre Report no 113 Feb 1983

[2] P H Jackson, 'The IAX Image Processing Language - Functions and Commands', UK Science Centre Report no 114 May 1983

[3] A G Constantinides et al, 'VLSI Architecture for Signal Processing with Alternate Low-level Structures', ICASP (1984)

Acknowledgements

We would like to thank all our colleagues at the UKSC for their help and criticism. In particular we are grateful to Dr Briony Williams and Dr Gerry Wolff for their useful comments and Dr Paul Jackson for the substantial help he has given us with the IAX package. We would also like to thank Nick Allen and Dilbagh Singh for the significant contribution they have made in implementing the SAY code. The hardware Fourier transformer was developed in collaboration with Dr A G Constantinides of Imperial College, London.