THE DEVELOPMENT OF A GENERAL PURPOSE SPEECH EDITOR

H.S.J.Purvis

Speech, Hearing, and Language Research Centre
Macquarie University.

ABSTRACT - This paper describes a general purpose speech editing and analysis program being developed at Macquarie University by the author. A brief history of the project is given. The design philosophy is explained. The general structure of the data and programs is examined.

INTRODUCTION.

The Speech, Hearing, and Language Research Centre decided to develop a general purpose speech editing and analysis facility because the alternative systems available at the time were either beyond the budget of the research centre or were not sufficiently versatile to meet its research needs. For example they did not permit multiple data channels with different sample rates such as the requirement to record one or two speech channels sampled at 20 KHZ and up to 16 physiology data channels sampled at 200 HZ.

HISTORY

The first computer program was used to construct stimulus tapes for listening tests. The traditional method of recording these tapes is to use two tape recorders and to copy the items from the master tape to the working tape in the required random sequence. The computer enables this procedure to be automated. The sound is sampled at 20 KHZ and recorded on a disc. The sounds could then be played back and recorded on an audio tape recorder in any required sequence.

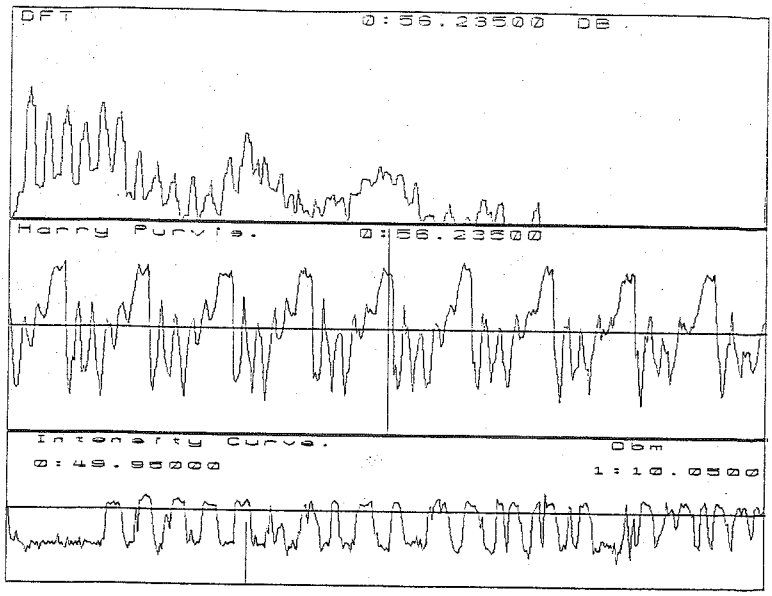
The program was expanded to include editing functions and analysis functions. The editing functions include the ability to measure the amplitude of a sound, to adjust the amplitude, to segment sounds and to combine sounds or portions of sounds. The analysis functions include Fourier transforms, Linear Prediction, and Pitch measurement. Another useful function is to enable the creation of complex tones with specified harmonic content.

In addition to this program, a program was developed that enabled the recording and analysis of physiology and speech data. This did not provide any of the editing functions of the first program but enabled the recording of up to ten minutes of continuous speech and for the analysis of air flow and EMG data as well as the speech data.

NEW PROGRAMS

These programs were limited by the operating system of the computer and the fact that they had been expanded as required to meet the increasing needs of the research centre rather than having been designed as a uniform whole.

When a VAX 750 computer was purchased it was decided that a much better program could be written for it based of our past experiences and making use of the greater power and flexibility of the new computer.



Example of Display Produced by the Speech Editor.

USER INTERFACE

The first area to be considered was the user interface. This is the section of a program that is often ignored or tacked on afterwards as it does not perform any 'useful' functions. However it is most important to the user, in particular the new user.

Making use of a computer program requires the user to answer many questions. Some questions ask the user to choose a command or command option and some are used to enter data values. It was decided that a uniform procedure would be used for all questions and answers and that the user should be able to choose one of the following options in answer to any question.

1. Enter the required data or command name.

This is the normal response when the default value is not what is wanted.

2. Obtain help on the question answering procedure.

This explains the general procedure for answering any question.

3. Obtain help on the current question.

This option helps to answer the current question by calling up the help file reading subroutine with the appropriate key word.

4. Obtain help on any topic related to the program.

This is a variation of option 3 to enable help on any part of the speech editor to be obtained.

5. Accept a default value as the answer to this question.

In some cases a series of questions must be answered, such as when designing a new display layout. Those values that do not need to be changed can be answered by pressing return, and nothing will be changed.

6. Return to a previous question when answering a series of questions.

This option does not 'undo' any changes, it gives the user the option of retracing his/her steps without the need no repeat the whole of the procedure. Each command is treated as being independent and it is not possible to go back from one command to a previous command only to a previous question in the current command.

7. Abort the current command.

The abort option always returns to the main command selection part of the program, unless it is used in that part and then the speech editor will stop.

8. Answer the current question and a number of following questions by typing all the answers in one go.

A number of answers may be typed in sequence separated by '¶'. If one of these contains an error, then all remaining answers will be discarded.

9. Type an answer that requires more than one line.

A continuation character may be used to type answers that require more than a single line.

10. Direct the program to read answers from a disc file.

When the answers in the disc file have been read, the speech editor returns to its normal mode of expecting answers from the terminal. If an error is found the remaining answers are ignored and the user is prompted for the correct replies.

The VMS operating system used on the VAX 750 enables programs to incorporate a help system that is the same as the system help procedure. This made the provision of on line help much easier, and enables access to help on any topic from any section in the speech editor. Two commands are used, '?' produces help on the question answering procedure, and 'HELP' produces help on the particular question. The 'HELP' command can be used to obtain help on any part of the speech editor by using the return key to go to a more general part of the help file and choosing another key word.

Where applicable each question has a default answer, and this answer is accepted by pressing the return key. The previous answer to a question becomes the default answer next time to simplify the repeated use of commands.

The method used to implement these options is that all questions are asked and all answers obtained by a set of subroutines known as the Input Management Package.

A method of responding to typing errors in commands has been devised. The command reading subroutine is given a list of legal command words. If the reply is such that it matches only one of these commands then this command is selected for execution. This means that only sufficient characters need to be typed to make the word unambiguous. If the command is ambiguous then the program will select the 'nearest commands' and ask the user to choose one of them by typing a number or by typing zero if none are acceptable. It was considered that this would be an improvement over the system that requires the user to retype every command that is not perfectly correct.

COMMANDS

The program provides two types of commands, Editing commands and Analysis commands.

Editing commands are used to segment data, to make adjustments to the data, and to combine segments of data in various sequences.

It is anticipated that data will be recorded on the computer disc in a large file containing several minutes of speech and physiology information. This will inevitably contain some unwanted sections such as pauses, repeats, and coughs and other noises. The wanted sections can be marked out and transferred to one or more smaller files for analysis.

The examination of data and the marking of data sections is performed by using an interactive display. A the present time this display makes use of a DEC VT240 terminal. The amount of data displayed can be set to any value up to the complete data set, however if a large amount of data is displayed then the time resolution will be poor as the number of data points in the display is fixed at 800. The display consists of a number of windows, each window shows data from one channel. The number of windows, the location of the windows on the terminal, their size, and the data to be displayed can all be changed at any time. The display definition may be stored in a disc file to simplify the selection of a number of commonly used formats. The windows can be divided into groups, each group has the same time values and is from the same data set. It is possible to display two or more sections of a data set on the screen at the same time, or to display sections of different data sets.

Any display on the terminal can be printed on the system line printer. A printed report containing data values and simple graphs can be produced from any section of data on the display.

Analysis commands include DFT, LPC, Pitch and Formant measurement. Analysis results may be included in the display on the terminal.

DATA STRUCTURE

The data structure is intended to be as flexible as possible. The term 'data set' has been used rather than 'data file' as it consists of four files :-

1. The header file contains all 'permanent' information on the data set, such as sample rate, block sizes, number of speech and physiology channels, number of computed data channels, titles, measurement units, and scaling factors. The header file can be edited by the system editor but it is not possible to change such basic variables as sample rates or block sizes by this procedure.

2. The data file contains the speech and physiology data. The file is unformatted to save disc space. The records can be read and written in any sequence and records may be deleted.

3. The computed data file contains data derived from the speech or physiology data such as the 'intensity curve' or the 'pitch curve'.

4. The description file contains the information used to define sections (words sentences etc) in the data and computed data files. It also contains comments on points in the data.

A point in the data is defined in the description file by an entry containing the following information :-

Level. This is a letter from 'A' to 'Z'. Two entries with level 'A' define the start and finish of the data set and no other level 'A' entries are permitted. The other levels define sections of the data set and may use many entries. Different levels are used to define different sets of sections and in most cases only levels 'A' and 'B' will be required.

Number. Each level can have up to 10,000 entries.

Kind. This specifies if the point is the beginning of a section, the end of a section or a comment.

Time. Each point has a time value in minutes and seconds.

Comment. A comment about the point.

A typical example would be a data set consisting of several sentences. The description file contents could be as follows :-

```
A 1 B 0: 0.00000  * Start of data
B 1 B 0: 1.45000  * Start of first sentence
C 1 B 0: 1.45000  * Start of word
C 1 E 0: 1.55000  * End   of word
C 2 B 0: 1.60000  * Start of word
C 2 * 0: 1.65000  * Start of voicing
C 2 E 0: 1.76300  * End   of word
B 1 E 0: 1.98500  * End   of first sentence
B 2 B 0: 3.12570  * Start of second sentence
B 2 E 0: 4.50000  * End   of second sentence
A 1 E 1:10.00000  * End   of data
```

Only two level 'A' entries are permitted, these indicate the total length

of the data set.

Overlap of sections is permitted :-

```
C 1 B 0: 1.45000  * Start of word
C 2 B 0: 1.60000  * Start of word
C 1 E 0: 1.76300  * End   of word
C 2 E 0: 1.98320  * End   of word
```

When a data set is connected to the speech editor, the data in the description file is read into memory. When the data set is disconnected, the description file will be rewritten if and changes have been made. The description file is maintained in time sequence, it is a normal sequential text file and may be edited using the system editor, but if this is done the time sequence must be retained.

PROGRAM STRUCTURE.

The program structure has been designed to make the programs as flexible as possible. The programs were written in Ratfor and this has imposed some limitations, for example the size of all data arrays must be determined when the programs are written and can only be changed by compiling and loading the complete editor. At the present time advantage has been taken of the virtual memory management on the VAX 750 to combine all commands and options into one program. At a later stage it may be beneficial to divide it into several smaller programs.

The display, analysis, and report formats are defined by groups of variables in common blocks that are available to all subroutines. This greatly reduces the number of constants and should limit the need to revise the programs as needs change in the future.

Before a data set can be examined or analysed it must be 'connected'. This command reads the header file and description file and sets up the data into internal arrays so that it is readily accesible. Because of the flexibility of the program every command needs to use this data.

The 'show' command is used to obtain information about the display, reports, analysis, and connected data sets.

The 'set' command is used to change data for the display, reports, and analysis. This data can be written into a disc file using the 'save' command and read by the 'set' command.

The 'create' command is used to create a new data set, it asks many questions as all parameter values must be established. The new data set consists of empty data files, a complete header file, and a description file with two entries.

The 'duplicate' command creates a new data set with the same parameter values as an existing data set. It does not ask any questions and is much quicker to use than the 'create' command.