Ann.Bull.RILP(1984)No.18, 5-11

INTERACTIVE LPC ANALYSIS SYNTHESIS PROGRAM ON VAX WITH AN ATTACHED ARRAY PROCESSOR

Hiroshi Imagawa, Shigeru Kiritani,
Sootaroo Sekimoto and Shuzo Saito

Last year, a new computer facility, VAX 11/780 with an attached array processor, was installed at the data processing center, faculty of medicine, University of Tokyo. By exploiting the efficient signal pocessing of the system, we have implemented LPC analysis synthesis programs which permits quick, intercative processing of speech data. The program is primarily intended for naive users and to provide a basic, standard means of LPC analysis synthesis of speech signal, i.e.,

1. an LPC analysis of the speech spectrum, formant and pitch frequencies, and
2. an modification of the LPC data and synthesis of the modified speech signal.

An outline of the system's characteristics will be given below.

Hardware system configuration

The central computer is the VAX 11/780 with 4M bytes of core memory. The system has an attached array processor FPS-100 (Floatiing Point Systems Inc.), and its nominal processing speed is 8M flops. A basic routine of the LPC analysis is supplied as part of the signal processing library on FPS-100. Input and output of the speech signal is performed through A/D and D/A converters which are attached to the intelligent Direct Memory Access controller (LPA11-K, DEC). The small system overhead of the present operating system together with the DMA capability of

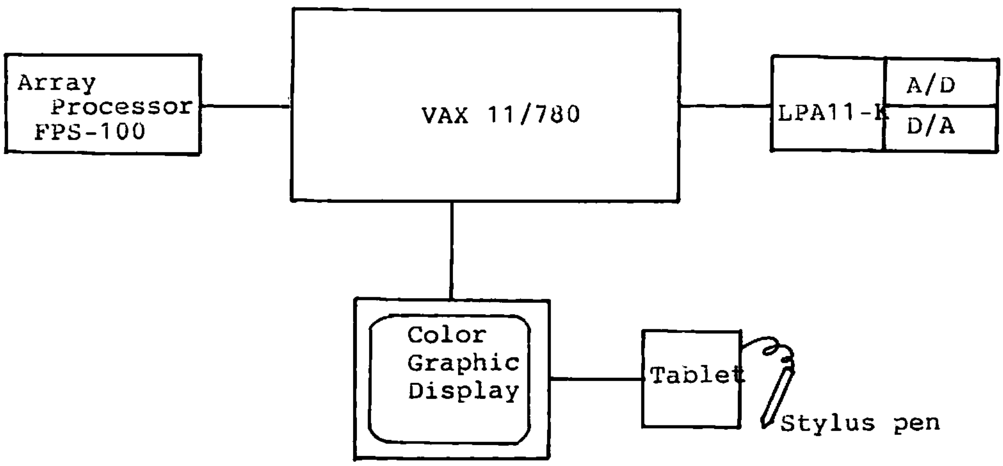

Fig. 1 System configuration.

the LPA11-K enables input/output of speech signals with sufficiently quick timing control under normal TSS operation of the VAX system.

For data display, an intelligent color graphic display terminal(NWX 235, JRC) is used. The resolution of the display unit is 1024x1024 points and 16 colors. The terminal has a 512k-byte segment buffer for local manipulation of the picture segments. For the purpose of intercative operations, a tablet board with stylus pen is attached to the display unit. A cross hair cursor is generated on the display screen, its position being controlled by the stylus pen.

In the following programs, almost all of the operations are controlled by using the stylus pen (and by the movement of the cross hair cursor). Marking of a selected time frame and specification of the parameter values are carried out through cursor positioning on the data display. Selection of the program operation is generally performed by moving the cursor to one of the operations listed in a menu displayed at the bottom of the screen.

Speech input/output program

In many data collection experiments, a subject is asked to read a list of, typically, several tens of test utterances. The program SPI01 is designed to sample and store in the computer such speech data by simple operations. When the program is running, it continuously samples input speech signals and stores the sampled data in a cyclic buffer. When a keyboard key is hit, the program switches the cyclic buffer to aonther cyclic buffer. At the same time, the program starts the transfer of the data in the previous cyclic buffer to a disk file (speech wave data file). Thus, in order to store a sequence of test utterances, the experimenter runs an audio tape recorder, leading speech signal to the A/D converter, monitors the speech sound and hits a keyboard key at the end of each utterance which has to be stored in the computer.

In such speech data, there is generally a pause of at least 0.5sec between utterances. This pause is sufficient for the program to respond to the keyboard , i.e, to locate a new disk file and start data transfer to the disk.

Another speech input program, SPI02, is used for storing speech signals of longer duration. Transfer of the sampled data to the disk file is performed parallel with sampling of the input speech signal. The program uses two data buffers. When one of the buffers is filled with the sampled data, data storage is switched to another buffer and the content of the previous buffer is transferred to the disk file. The two buffers are used alternately. Sampling is started by hitting a key, and sampling continues for a pre-specifeid time interval. There is virtually no limitation to the duration of the speech signal that can be stored.

Speech wave data editing program

The amplitude envelope of specified speech wave data is displayed and, on the display, portions of the speech signal can be selected interactively (Fig. 2). Selected portions of the speech signal can be monitored through an A/D converter or can be stored in a new disk file.

Selection of portions of the speech signal and also the selection of the program operations are performed using a cross hair cursor in the following way. (These operations are basically common to all of the subsequent programs.)

First, a selected moment is marked by positioning a cursor on the amplitude display and is memorized as the "current frame" by the program. The "current frame" can be defined as the "head" or "tail" of the selected time period by pointing to the HEAD or TAIL in the menu of the program operation displayed at the bottom of the screen. LISTEN and STORE in the menu specifies sound monitoring and data transfer to the disk file, respectively.

Formant analysis program

This program performs an LPC analysis of specified speech wave data and plots the formant and pitch curves.
Step 1 Amplitude envelope is displayed and the portion of the speech signal for analysis is specified using the cursor as in the editing program described above.
Step 2 Time functions of the formant and pitch frequencies are displayed together with the amplitude envelope and voiced/unvoiced parameter. Optionally, the sequence of the LPC spectrum envelope is plotted (Fig. 3).

The display in this program is intended to serve as a kind of digital sound spectrogram. A hard copy of the display is obtained through a laser beam plotter connected to the display unit. Numerical data of the formant frequencies, etc. at selected time frames are printed or stored in a specified formant data file for later processing.

The program provides several means of manual correction of the analyzed data. Among the calculated poles of the LPC spectrum, only those poles which are selected as formant poles are plotted with thick dots, and the remaining non-formant poles with thin dots. This formant/non-formant disticntion can be reversed by pointing to individual dots with the cursor (thus changing the formant/non-formant judgment about the corresponding pole). Extraction of the formant poles in the present program is based on the following relationship between the formant frequency and the band width,

$$Bi<Fi^{0.27}/0.016$$

where Fi and Bi are the pole frequency and the bandwidth of i-th formant, respectively.

In order to reduce errors in the pitch extraction, a re-calculation of the pitch frequency can be carried out after specifying the minimum and maximun values of the range of the pitch frequency through a visual inspection of the displayed pitch curve.

LPC analysis program

LPC analysis of specified speech wave data is performed and the LPC data are stored in a LPC data file.

Step 1 Amplitude envelope of the speech signal is displayed. On the display, a portion of the speech signal is selected and the LPC analysis is performed.

Step 2 The time functions of the calculated LPC parameters (pitch, residual wave power and the voiced/unvoiced parameter) are diplayed (Fig. 4).

In order to evaluate the extracted LPC parameters, the speech signal of a selected portion is synthesized and listened to. The amplitude envelope of the synthetic speech signal and its waveform for the "current frame" are displayed for visual inspection of the characteristics of the synthetic speech. The amplitude envelope and the waveform of original speech can also be displayed for comparison.

Possible errors in the automatic extraction of the LPC parameters can be corrected manually. The value of the pitch frequency, residual wave power or voiced/unvoiced parameter for the selected time frame can be changed to a specified value by using the cursor. The position of the cursor on the data display specifies both the time frame and the parameter value for data correction. It is also possible to re-calculate the pitch frequency with given restrictions on the frequency range, as in the formant program.

LPC synthesis program

LPC parameters in the specified LPC data file can be modified. The speech signals for the modified LPC parameters are synthesized and stored in a speech wave data file. The program is primarily intended for manipulating and evaluating the prosodic characteristics (pitch and duration) of the speech signal.

As in the LPC analysis program, the time functions of the LPC parameters are displayed, parameter values can be modified manually and a selected portion of the speech signal can be synthesized.

In addtion to several manual correction operations in the LPC analysis program, the following operations are also possible in this program. The time function of the pitch frequency or residual wave power during a selected time period can be defined by specifying the parameter values at the "head" and "tail" frame of the selected period and interpolating these values over the entire period. It is also possible to delete selected time frames, to copy them to another place in time or to expand these time frames through duplication.

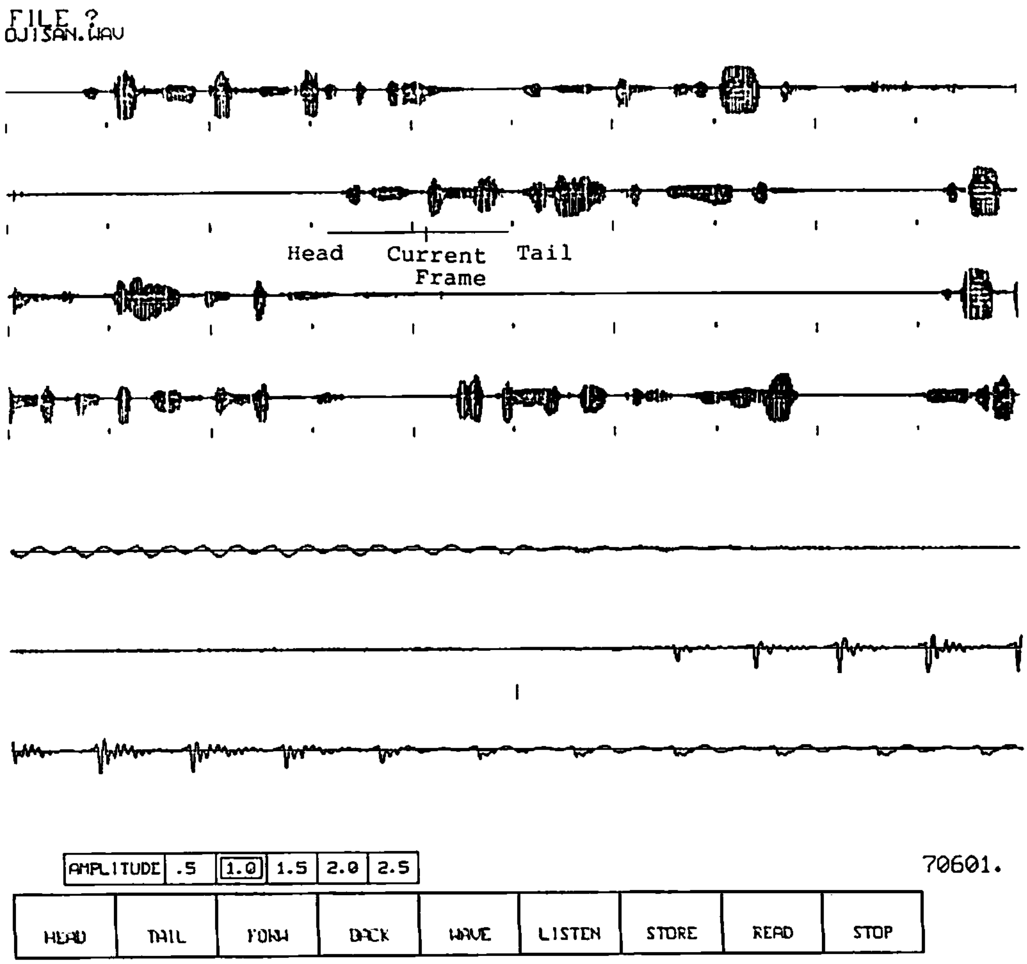

Fig. 2 Data display in the Speech Wave Data Editing Program.

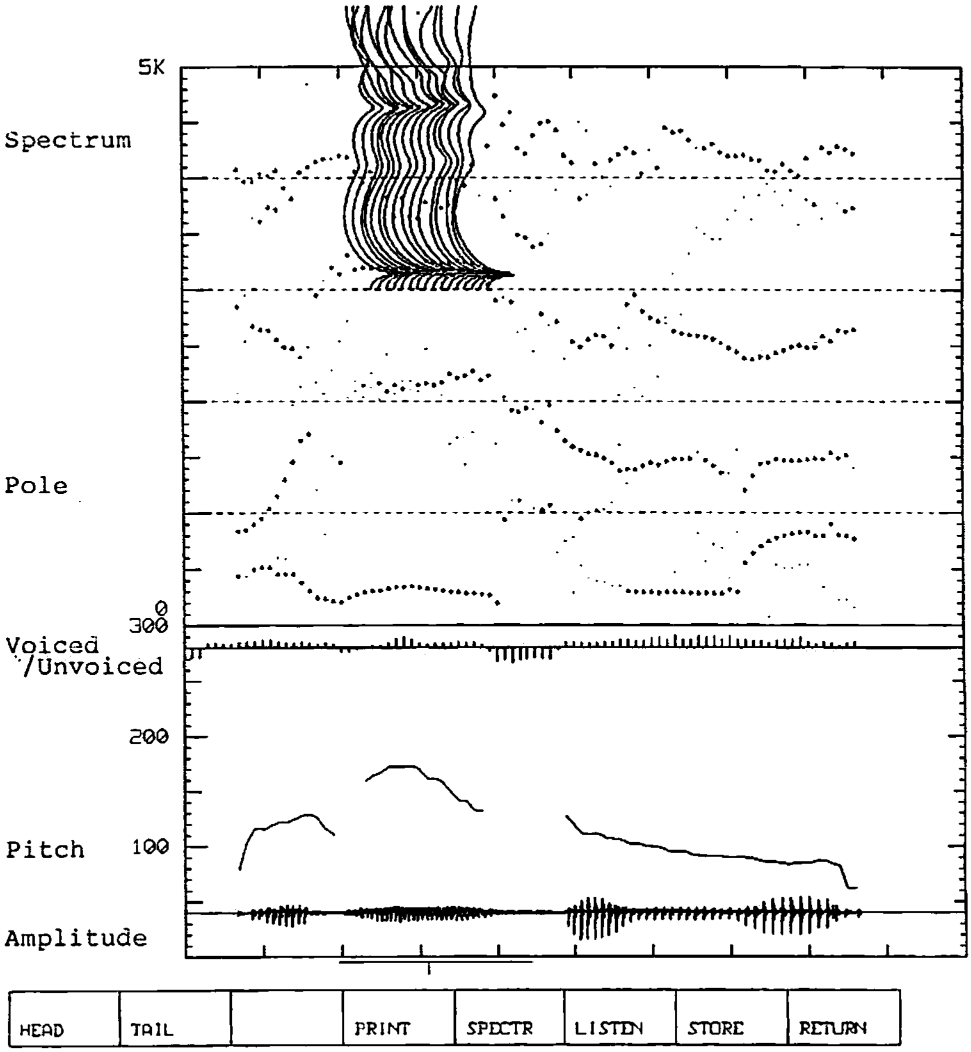

Fig. 3 Data display in the Formant Analysis Program.

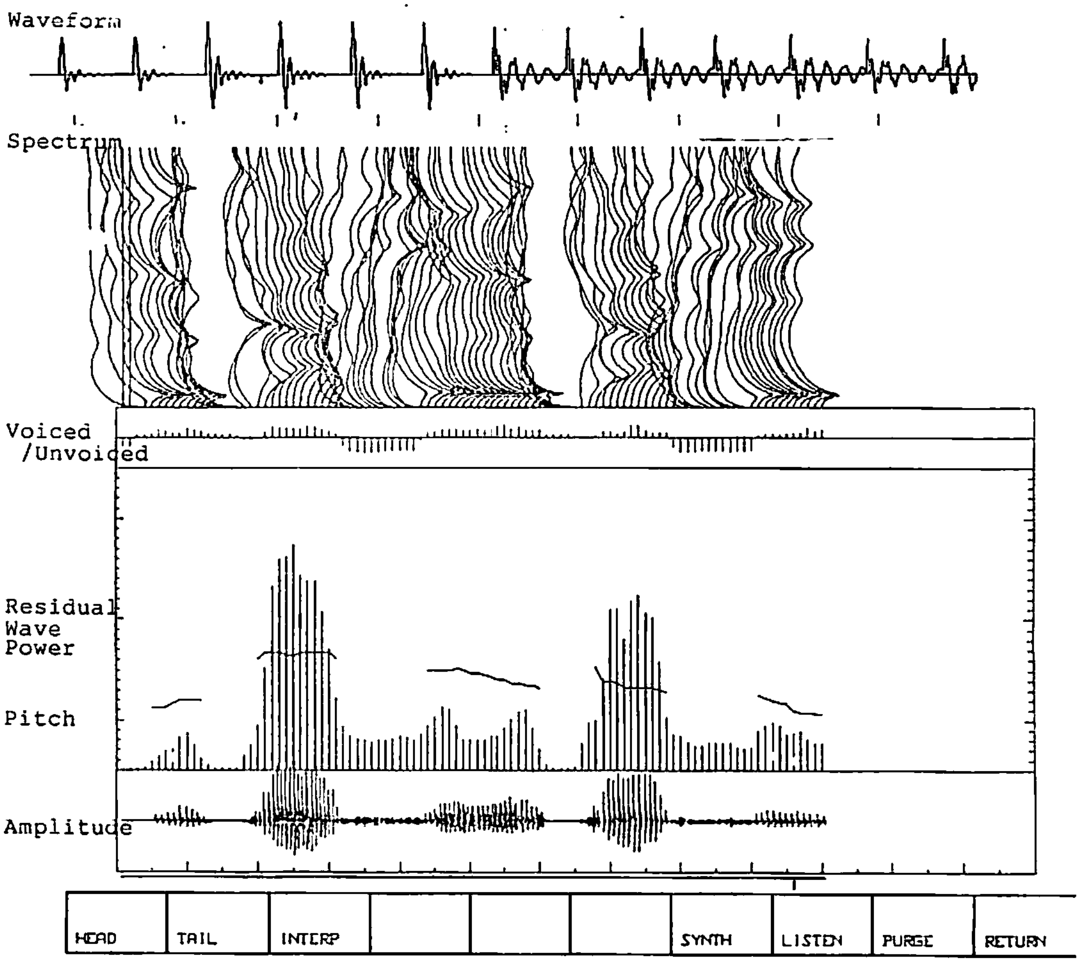

Fig. 4 Data dispay in the LPC Analysis Program.