WORKSTATION FOR SPEECH ANALYSIS

S. Andreyev and V. Chuchupal

Computer Centre of the USSR Academy of Sciences, Moscow

ABSTRACT

SLIRE-3 (Speech and Language Interactive Research Environment) is the workstation software for IBM PC compatible computers fitted for research in such areas as speech recognition and synthesis, phonetics, criminology and so on. The system was developed in Computer Centre of the USSR Academy of Science during the last 3 years. The main features of this system is its openess and the possibility for phonetists and linguists to organize data bases on it.

1. INTRODUCTION

Several years of work in the area of speech analysis gave us an understanding of the features of the workstation which we need for our investigations [1,2]:

- visualization of the main speech signal forms; the multy window system with windows easily modified for specific applications;
- realization of the main speech signal analysis algorithms;
- system openess; new algorithms can be easily added to the system;
- possibility to organize data bases;
- realization of statistical analysis procedures;
- system friendliness and comfortable interface.

We realized our project on IBM PC compatible computer. We don't think that it is a very suitable for speech analysis computer and we know that there'll be some requirements which we'll not have a chance to fulfil on such a computer, but today it is the most popular computer in phonetic and linguistic laboratories of USSR.

The most valuable requirement which, we think, can't be realized on IBM PC compatible computers is the real time performance. The most prolonged procedure in speech analysis is spectra calculation and everything based on it, for example, sonogram calculation. For spectra calculation we use different FFT algorithms: Vinograd algorithm (slightly modified by ourselves) and Walch transformation. Certainly, they can't be used for real-time analysis, but they are suitable for ordinary speech analysis with data base.

Though our system can be used in different areas it is oriented on phonetic and linguistic analysis, especially in the part connected with data bases.

The most part of speech analysis is impossible without the data base, because:

- a lot of files (signals) are used;
- users may need the segmentation of those signals;
- different projects may use different segmentation of the same files;
- a lot of researchers use the statistical analysis, which can be done only on stored data;
- users may need an instrument for quick search through a lot of accessible data.

2. INTERFACE

Specialists in speech analysis rarely are good computer specialists that's why the workstation which is difficult to use will not be used at all. The same result will be with the workstation easy to use but primitive in availability. Our system is menu driven, sends different messages, includes brief and full HELPs. The user will open new possibilities during work sessions and he must not remember a lot of information at the beginning of his work.

3. PROGRAMMABILITY

We understand perfectly that though we tried our best to fulfil the maximum of users requirements nearly every researcher wants to add his own algorithm to the system. The only possibility for researcher to add his own algorithm to the system is to become a programmer for some time and to prepare only his algorithm. The visualization of the results will be done automatically.

4. SIGNAL FORMS

SLIRE-3 can display speech in traditional forms: different waveforms, spectrograms, sonogram and diagrams of parameters (zero-crossing, energy and so on) together with timing and segment marks (from database). Users can easily switch analysis format to choose the most suitable for their task. User can work with the signals of any duration (from very short to very long); different types of analysis filters can be selected for better time and frequency resolution. Any part of any signal can be cut out and deleted or added to another signal and in such a way the new artificial signals can be obtained.

Figures 1, 2 show examples of different SLIRE-3 layouts, composed by different users for their specific
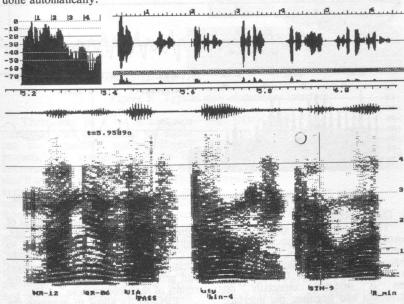

Figure 1: A layout with entire signal and sonogram of the signal part.

research needs. Any part of a signal can be shown in most suitable view from extension to high degree compression and in big amplitude scale diapason. Each waveform type is shown in different window.

The direct control over all display parameters of all windows is very simple and the disposition of the signals on the screen can be changed in zero time. Any window can be shown with (or without) the time scale and the user can choose the

number and type of characters to show in each window. The different windows are not time-synchronised; it allows to compare on the screen different parts of the signal. The synchronization is done at the window change moment.

The system disposes of interface convenient for the user, and all the analysis parameters as well as signal processing and parameters of visualization (near 160) including colour palette option for every window, can be adjusted by any user. All these parameters are easily changed by user during the session and automatically stored in configuration file after the end and used by system for next sessions. The system is bilingual (all messages are in Russian or English).
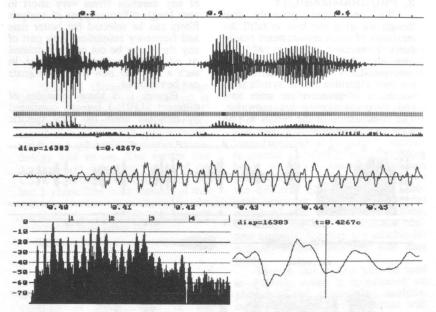

Figure 2: A layout of entire signal and different views of different parts.

5. DATABASE

The kernel of our system is the possibility for any researcher to organize speech data bases, to store and change data in those data bases and to analyze these data. Our workstation is oriented especially on phonetists and linguists; the interface of the system was organized in such a way that such specialists can easely work with it.

The users of our system can create new data bases and change any information which was included earlier. Data bases will contain information (objects) of 4 different types: DICTORs, PROJECTs, SIGNALs and SEGMENTs.

Object DICTOR contains a lot of signs (for example: birthplace, age, native language, education and so on). This information allows to make decisions about different languages and pronunciations and to obtain the results of their comparison.

PROJECTs allow to organize the different segmentations of the same signals by several researchers under the different rules and algorithms and those results will not interfere with one another during the statistical analysis.

Each object of SIGNAL type corresponds to the real signal (not file itself but only its special descriptor). It includes some information about signal (file's name and type, pronounced text, phonetic and linguistic notation and auxiliary information). The signals may be of two different types: initial (obtained by A/D) or artificial (obtained as the combination of parts of different initial signals). This sign is very important not to distort the results of statistical analysis. Every initial signal has a reference to the object DICTOR of this signal.

Object SEGMENT corresponds to every segment marked on the signal. All the statistical analysis is based on the segment's parameters (time, duration, type and an information to visualize with the mark of this segment). Each user can define his own types of segments or use common definitions for a group of users. There can be different segment types for sentences, words, phonemes, letters, sounds, types of sounds and so on. If the user wants to visualize the segmentation then each segment will be marked and auxiliary information connected with this mark will be seen too.
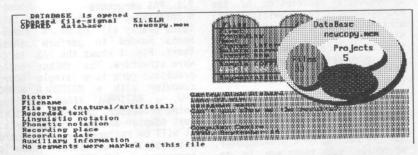

Figure 3: Typical view of database interface.

There may be a lot of segments connected with any signal; the signal can be segmented under different projects, the segments may be of different types and not all of them the user'll want to see simultaneously. The system has the simple interface to organize the visualization of only those groups of segments the user wants to see and in such a way he prefer to see them. Different types of segments can be shown on different screen levels one under another and the user only chooses the types and the order to show them. The order and the types of segments to show are choosed for each signal form independently so the user can organize it for example in such a way as to see words and accents under the full signal view, marks of vowels under sonogram and so on.

6. CONCLUSION

An interactive research environment was designed for speech signals analysis. It includes the possibility to organize data bases for phonetists and linguists.

7. REFERENCES

[1] ANDREYEV,S., CHUCHUPAL,V. (1990), "SLIRE-2 - an interactive system for speech signal analysis on IBM PC", Computer Centre of USSR Academy of Sciences, Moscow (in Russian)
[2] GONCHAROV,S., CHUCHUPAL,V. (1987), "An interactive software for speech signal analysis", XI-th ICPhS Proceedings, v.5, 63-67 (in Russian)