AUTOMATIZED FORMING OF INDIVIDUAL SPEECH FILE FOR AUTOMATIC SPEECH RECOGNITION AND SYNTHESIS SYSTEM

Taras K. VINTSIUK
NAS Institute of Cybernetics, Kyiv, Ukraina

ABSTRACT

The automatized Individual Speech File (ISF) forming process is proposed. ISF is intended for introducing into computer to start automatic speech recognition and/or synthesis for a given person. For this the training procedures are used. They automatically calculate phoneme-threephone prototypes, temporal, prosodic, energetic and other parameters and characteristics of speaker voice which constitute so-called Individual Speaker File.

GENERAL PRESENTATIONS

ISF forming procedures appeal to Individual Speech Signal Archives (ISSA) and process all its signals.

ISSA is a set of original speech signals (e.g. speech realizations) expressed in amplitude-time domain. ISSA is considered as fulfilled then each speech realization is accompanied by such descriptions: 1) orthographical text; 2) phoneme (phonetical) transcription; 3) phoneme signal bounds (marks); 4) current pitch periods and bounds, and others.

ISSA forming is doing in laboratory conditions and controlled by an expert (e.g. researchers-phoneticians). The expert can correct the phoneme segment bounds using the mouse and by the way of hearing speech segments and regarding original and/or description speech signals (current autocorrelation, spectrum or cepstrum for example) through the computer audio-video-monitor. The marks accepted by the expert are then transferred automatically from one supported realization onto others ones. The expert can correct the realization transcription too. Each fixed speech segment (with fixed bounds) is considered as a realization of the phoneme-threephone, that is a base phoneme accompanied by both preceeded and followed phonemes accordingly with the phonetic context. The phoneme-threephone history that is word, sentence, realization number is conserved too. Further, fixed phoneme-threephone segments are used then as prototypes for automatic segmentation of speech realizations which correspond to other words and sentences. So, there are many procedures which copy expert actions and automatize the labelization process and ISSA forming.

ISSA and ISF for one supporting speaker are then used in ISSA fulfilling for a new speaker. The researcher-phonetician invasion into the process of forming of both Individual Speech Signal Archives and Individual Speech File expands the speech knowledge and improves the accuracy of individual automatic speech recognition as well as the quality of individual automatic speech synthesis.

ILLUSTRATIONS

As an example a speech signal realization of the Ukrainian word ОДИН (ONE in English) is presented in the Figure 1. The top graph is an amplitude-time speech signal. Then current autocorrelation, cepstr and spectr are shown in three lower graphs respectively. For these preprocessing presentations as well as for segment bounds the uniform discrete time with the step 15 *ms* was used. The phoneme-threephone bounds were arranged by the expert. Phoneme

segments are accompanied by phoneme symbols # (pause), О (О non-stressed), _Д_ (voiced stop phase of D), Д (plosive phase of D), И (I stressed), Н (N sonorant).

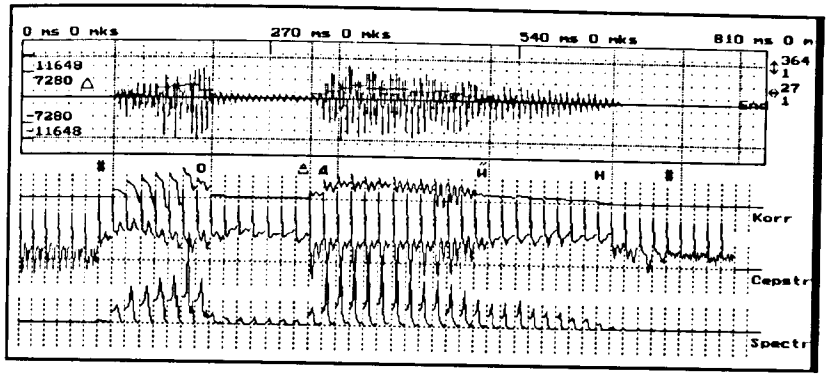

Figure 1. Monitor presentation of speech signal realization of the Ukrainian word ОДИН (ONE in English).

The result of automatic transferring of phoneme-threephone bounds accepted by the expert (see Figure 1) onto a new realization of the same word is shown in Figure 2.

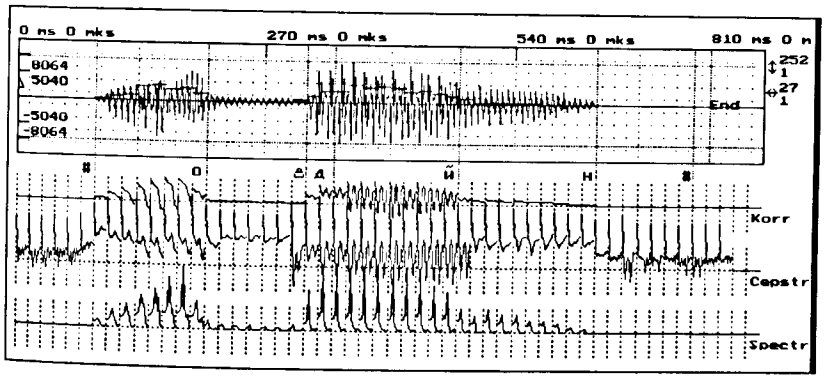

Figure 2. The result of automatic transferring of the phone bounds from the supported realization onto other one for the same word.

Figures 3, 4 and 5 present in detail the fragments of the speech signal realization shown in the Figure 1.

STATE OF THE ART

Now the Individual Speech Signal Archives and the Individual Speech File fulfiling technologies are being near completion. Mainly ISSA and ISF are made for Slavic languages especially Ukrainian and Russian.

The result of the study is used in the designing of Multilingual Speech Dialogue Systems [1], [2], Oral Dictation Machine, Oral Translation Machine.

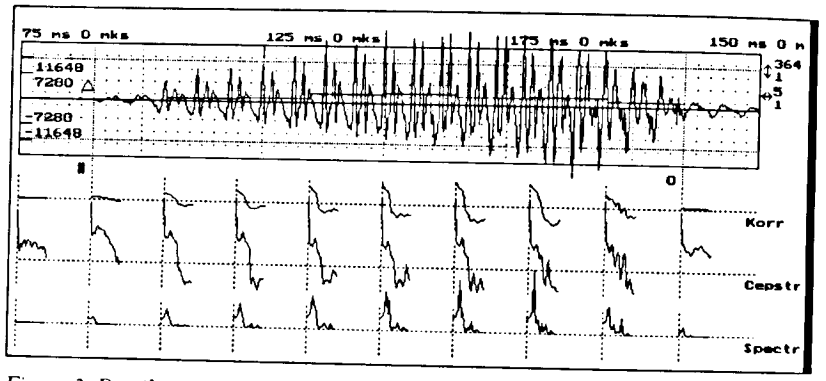

Figure 3. Detail monitor presentation of the supported speech signal realization of the Ukrainian word ОДИН (first fragment).

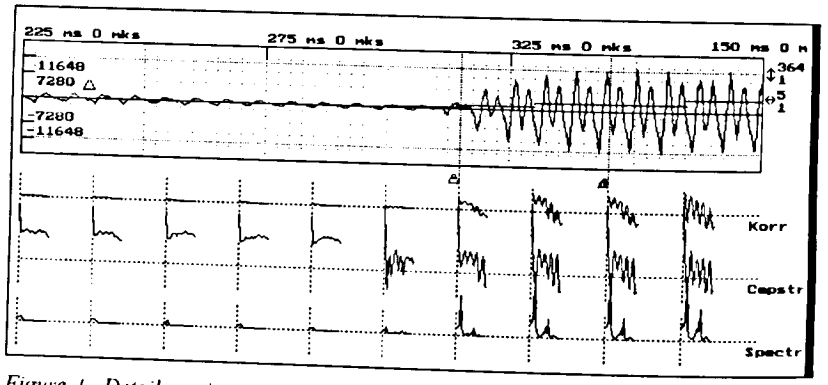

Figure 4. Detail monitor presentation of the supported speech signal realization of the Ukrainian word ОДИН (second fragment).

ACKNOWLEDGEMENT

This study have been inspired by the Research Programs of the Ukrainian State Committee on Science and Technology, and by the cooperation within ELSNet, and we want to express our gratitude to E. Klein and F. Neel and other people responsible for them.

REFERENCES:

[1] Final Report on the UNESCO Contract SC/RP 261060.8 "Development of Multilingual (including English and Russian Languages) Speech Dialogue System for a Microcomputer", Institute of Cybernetics of the AS of Ukraine, Kyiv, 1986, 97p.

[2] Final Report on the UNESCO Contract SC/RP 261377.9 "Advance of Multilingual Speech Dialogue System for a Microcomputer", Institute of Cybernetics of the AS of Ukraine, Kyiv, 1989, 33p.