A LARGE BULGARIAN CENTRAL ALLOPHONES DATA BASE

PHILIP CHRISTOV

Voice Man-Machine Communication Lab.
Mechanics & Biomechanics Institute
Sofia, Bulgaria, 1090, P. O. Box 373

ABSTRACT

In this paper a large data base of Bulgarian central allophones is considered consisting of about 5000 utterings from 60 professional speakers (30 male and 30 female). They are imbeded in words each pronounced at the end of a standard carrier sentence. Professional analog (motion picture magnetic and optical) recordings are available of the carrier sentences together with the digital recordings (IBM compatible) of the vowel and consonant segments, manually extracted from the carrier. Each such record is labeled with a digital code which allows computer selecting, sorting and merging the data, according to the particular research purpose. The data base is verifyed by a group of 20 listeners by means of a semiautomatic aural identification procedure (CHRISTOV, Ph., Acustica, 29, 347-349, 1973).

INTRODUCTION

Contemporary computing technology offers vast prospects in processing large amounts of speech material in sensible stretches of time. The results are free of individual interpretations of the experimental data (as is the case, for example, by visual reading of spectrograms). If by the preparation of the machine speech input suitable knowledge is used, the machine output should be comparable not only within the limits of a given language, but also within the much broader limits of its language group.

SPEECH INPUT

A principal requirement in performing meaningfull acoustic measurements upon the phonetic units of speech /1/ is to use as experimental stuf PHONES (See REMARK) of comparable ALLOPHONES imbeded in words, uttered in equal phrases with equal intonation.

In agreement with this requirement the composite parts of the data base are phones of the central allophones of the Bulgarian speech, namely:

/C-V̀-C/ - stressed vowels in relevant environment of two bilabial consonants

/a-C-à/ - consonants in relevant environment of a preceeding unstressed /a/ and a following stressed /à/ /2/.

Each phone, together with its relevant environment is put into a major stress fraction of a word, taken from Bulgarian unilingual dictionaries and inserted at the end of a carrier sentence.

The carrier sentence consists of four phrases: The first phrase carries the label of the individual speaker, the second-the label of the phonemic group, the third-the label of the particular allophone and the last-the carrier word.

As shown in the TABLE except of the six Bulgarian vowels from the central allophones /bV̀b/ and /pV̀p/, there are vowels from two more allophones which are not central. The first of them, /tV̀t/ is chosen because common words could be found for all Bulgarian vowels. The second, /bVb/, is the unstressed opposition of the central allophone /bV̀b/. All consonant phones of the central allophones /aCà/ are contained in the same environment speech fraction /naCà/ which is imbeded in the initial part of the carrier word.

The consonant phones in initial and final position in words are presented by the allophones /Càn.../ and /...nàC/. The latter may as well be considered as central for the consonants of these particular phonetic groups. Judjing from the similar phonetic context there is little doubt about their comparability with the central allophones /naCà/.

BUILD UP PROCEDURE

The data base build up procedure is similar to that employed in the classical study of Potter and Steinberg /3/ the difference being in its intensification by application of computing technology (Fig.1).

The test utterings are extracted from two groups of speakers:

30 bass-bariton males
30 mezzo-soprano females

They all are professionals with distinct pronunciation selected among the radio and television speakers and the actors from the theaters of the Bulgarian capital. Their voices are energetic and belong to people between 25 and 50 years old. From them has been demanded to pronounce each phrase of the carrier sentence indifferently, in slow style and with falling intonation.

The working language in front of the microphone was Standard Bulgarian, i.e., the language the speakers are practizing during their public performances.

The speakers read the test material two times from randomly mixed cards to avoid the practice effect.

AUDITORY VERIFICATION

The auditory verification of the data base /4/ is carried on by a group of 20 lay listeners all native Bulgarians from different parts of the country and with some technological education.

The listeners have been presented two times with sound recordings of the carrier sentences. Having at their disposal listeners cards, containing labeled carrier words, they reacted by filling in the empty spaces in the carrier words with the letters of the phones they heard.

The listeners output was punched on machine cards and processed by a computer program which excludes the responses of the false listeners and punches on cards the labels of uncorrectly pronounced phones.

ANALOGUE DATA

During the sound recording session the speakers read the input cards with moderate voice effort and kept a sound level between 60 and 80 dB. The microphone was placed in the middle of a highly damped camera for acoustic measurements.

The audio recording was carried on by a professional sound recording staff which used studio equipment. The frequency response of the sound recording equipment via microphone and magnetic tape was flat between 100 and 15000 Hz (± 2 dB) and the noise level via magnetic tape was -64 dB. During the recordings the voltage and the frequency of the altrnating current network remained inside their standard limits: 220 V (+4.5%, -0%) and 50 Hz (±0%). The original speech data are recorded on 6.35 mm magnetic tape with tape speed of 760 mm/sec. Working copies have been prepared on standard 35mm motion picture perforated magnetic and photographic tapes including high quality oscillograms and

trivial variable-area photographic motion picture sound recordings. They can be displayed simultaneously and synchronously on motion picture sound editting equipment thus offering the technological prerequisite for simultaneous audio-visual inspection of the analog sound recordings.

SEGMENTATION

The "segmentation" of the analog data was carried on manually on a sound reading bench after carreful audio-visual inspection of the sound records and their control oscillograms.
It consists of marking the beginning and the end of each magnetic tape segment, carrying a labeled phone, with a perforation and a strong magnetic pulse (Fig. 2).
The places of the markers belonging to each such magnetic tape segment have been determined after:

1. The content of the segment has been HEARD as coextensive with the labeled-phone quality

2. The visual duplicate of the same segment has been OBSERVED on the control oscillogram or motion picture sound record:

a) By the VOWEL-phones: As a mighty tone burst between two weak noisy signals

b) By the CONSONANT-phones: As a moderate or weak noise-like signal between two mighty tone bursts or between a pause and a mighty tone burst.

DIGITAL DATA

The segmented analogue data were fed to the input of an analog to digital converter set "on" by each "Start" marker and "off" by the imediately following "Stop" marker. In the analog to digital converter the nonturbulent segments (vowels and resonants) were sampled with a rate of 20kHz and the turbulent ones with 40kHz to ensure no lose of audio frequency information below 10kHz, for the sonant-like, and below 20kHz for the turbulent sounds.
The output digital data files were stored in an IBM-compatible magnetic tape memory with recording density of 1600 bit/inch.
After the analog to digital conversion the data were processed by a servise program which does three things:

1. Records meaningfull 9-symbol labels at the head of the first block of each file. The simbols are decimal numbers carrying information about the origin, the hystory, the kind and the inventory of the file. The last symbol in the label is a key which lockes the file if ordered.

2. Performs correction of the segmentation.

3. Lockes unfinished files at the end of the tape or files for which no agreement was reached between the speaker and the group of reliable listeners.

CONCLUSION

The data base here considered was created for a research aimed at the build up of the phonetic fulcrum of the Bulgarian language for the purposes of computer recognition of continuous speech. It is available in 14 volumes, each stored on a 1/2 inch, 2600 feet magnetic tape together with a subroutine in FORTRAN IV for sequential or selected reading and/or rewriting of the labeled files. Its output can be easily reshaped by the use of standard sort/merge programs because in the labels are present explicite symbols except for the phonemic category but also for each individual speaker and its sex. This way the data base can be used for research purposes not only in the field of trivial and comparative acoustic phonetics but also in studying the prosodic features in speech connected with the personality and the sex of the speakers.

REMARK

From the different definitons of PHONEMES existing in the linguistic litterature the definition of B. Bloch /5/ is adopted here because it suits best the technological aspects of speech:

PHONEME - class of allophones

ALLOPHONE - class of phones in the same relevant environment

PHONE - any continuous fraction of a phrase that is heard as coextensive with a given quality

PHRASE - an utterance or part of an utterance bounded by successive pauses

REFERENCES

/1/ Twaddell, W.F., "Phonemes and Allophones in Speech Analysis", JASA, 24, 607-611(1952)

/2/ Stoikov, S., "Introduction into the Bulgarian Language Phonetics", 3-th rev.ed., Nauka i izkustvo, Sofia (1966) (In Bulgarian)

/3/ Potter, K.R., Steinberg, J.C., "Towards the Specification of Speech", JASA, 22, 807-820(1950)

/4/ Christov, Ph., "A Semiautomatic Speech Sounds Aural Identification Procedure with Its Application to Speech Analysis", Acustica, 29, 347-349 (1973)

/5/ Bloch, B., "Studies in Colloquial Japanese", Language, 26, 86 (1950)

TABLE of involved allophones:

Test uttering position in the carrier word	Initial	Medial	Final
Stressed vowels	/bV̀b/ /pV̀p/ /tV̀t/	/ / /	/ / /
Unstressed vowels	/bVb/	/	/
Resonants	/Càn/	/naCà/	/nàC/
Voiced turbulent	/Càn/	/naCà/	/
Voiceless turbul.	/Càn/	/naCà/	/nàC/

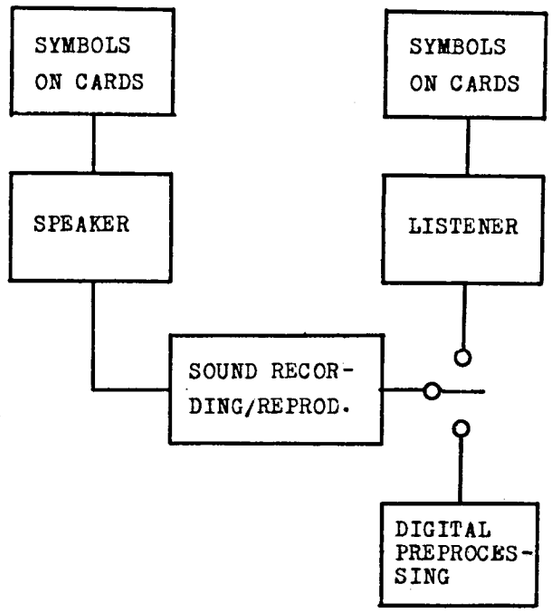

Fig. 1. Flow chart of the data base build up procedure

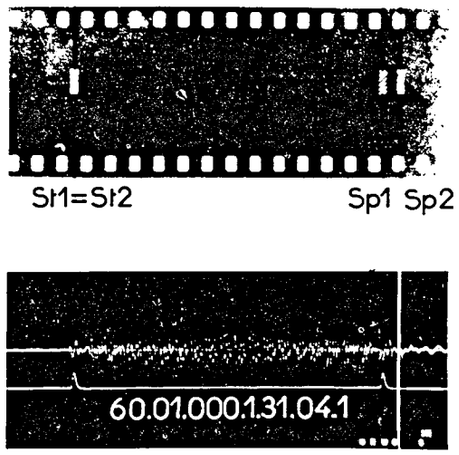

Fig. 2. (Top) Segment of magnetic tape record carrying the vowel /à/ from the allophone /bàb/ imbeded in the word /bàba/ uttered by speaker N.D.(male); (Bottom) Control oscillogram of the same segment

St1, Sp1 markers of primary segmentation

St2, Sp2 corrected segmentation