A SYLLABLE APPROACH TO THE SPEECH INFORMATICS

A.KNIPPER

Institute for the Problems of Information
Transmission Academy of Sciences
101447, Moscow, GSP-4, USSR

ABSTRACT

A problem devoted to trends of syllable development for the usage in speech informatic systems is under consideration. It is noted that simple open CV syllables (C is a consonant, V is a vowel) are the most stable discrete phonetic units of continuous speech with respect to the context , speaker variability and noise. Problems of that type syllable classification and statistics for the Russian speech and their relations with letter records of the speech information are discussed. Some experiments on compilation syllable synthesis of the Russian speech of free contents and on analysis of the speech signal using CV fragments are briefed.

INTRODUCTION

The main problem of the speech informatics is development of man-machine communication systems on the base continuous speech. In that case speech communication between a user and a system is ensured with the best conditions. In continuous speech recognition / understanding the most promising approach is representation of the speech flow with the help of symbol sequences similar to the speech transcription with afterwards decoding at the word or phrase lewel / 1, 2, 3 /. The main requirment of that approach is transformation of a continuous signal into a discrete sequence of speech elements, phonetically stable to speaker variability, context , noise and other facts which influence the speech signal features. I that case in the process of speech recognition / understanding system operation and its new vocabulary training the conviniency for users is ensured / 4 /.

In the process of synthesis of any piece of speech information an iverse problem is solved, i.e. a letter sequence is transcribed by phonetic symbols and then is transformed into the corresponding acoustical signal, and besides for comfortable usage it is desirable to synthesise any voice and any speaker sterotype according to a user choice.

The choice of a phoneme as a phonetical symbol for a speech communication system is the most reasonable and convinient, as it permits relatively easily to pass to the conventional letter representation of any data, accessible and intelligible by broad circles of users. However, numerous researches on phonetics and speech informatics show that there is no direct relationship between speech segments and phonemes. The same sounds match speech segments with essentially different spectral and temporal characteristics, that is determined by context, positional and speavariability of the speech. Simple open CV syllables have more stable characteristics especially those that are cut off from left and right from a transition line, named CV fragments / 5, 6 /. It is considered that CV syllables are base speech elements for Russian, Italian, Japanese and other languages / 7, 8, 9, 10, 11, 12/ and is more widely used in different speech informatic system. In the following sections it is shown that usage of CV syllables as base units of Russian permits to perform a rather distinct classification of context depended pairs of sounds and to choose the minimal alphabet of discrete phonetic elements which describe the continuous speech.

BASE ELEMENTS OF THE RUSSIAN SPEECH

It is considered that open syllable is a speech universal unit for the majority of languages / 8, 9 /. In the speech informatics open syllables may be also prefered, due to the fact that there is a distinct transition from the corresponding consonant to the vowel in the interval of that open syllable that makes easier to label the continuous speech visually and with the help of technical means /5, 6, 13 /. The number of open syllables for Russian is great, i.e. about 2500 / 14 /, however any open syllables may be represented as a concatenation of the base CV syllable and separate consonants and vowels. Thus compound open syllables such as CCV, CCCV, CVV may be expressed by C+CV, CC+CV, CV+V correspondingly those constructions the most strong coarticulation can be observed in CV combinations / 8, 10 /,

that determines the necessity to examine that speech element as a whole. In that case CV syllables cover about 80% of any text of the Russian speech. The stressed and the first prestressed syllables with stabile phonetic quality have frequency occurrence in the text equal to 50% /6/.

For the needs of analysis and synthesis it is useful to represent the base CV syllables in the table form. In that table consonants and vowels are written not in the phonetic symbols, but in traditional Russian letters, that allows to transform a written sequence of letter symbols into to the corresponding syllable one. Twenty consonants are written in the vertical column of the table according to the manner of production, and in the horizontal rous according to the plase of articulation (labels L, D, A, P, Vl,V, Lq and N correspond to labial, dental, alveolar, palatal, voiceless, voiced, liquid and nasal consonants). Ten vowels are divided into two groups which form hard or soft variants of

Table. Classification of the base elements of the Russian speech

Consonants			L	D	A	P	Vowels — Hard					Vowels — Soft				
Manner of production							A	O	У	Ы	Э	Я	Ё	Ю	И	Е
							1	2	3	4	5	6	7	8	9	10
I	Fricatives	Vl				Х		/	///	///	///	///	/	/	/	
2					Ш			/		/	X	X	X	X	X
3					Щ		X	X	X	X	X	/	/	/		/
4			С				/		/	/	/	/	
5		Ф				/	/	/	/	///	///	/	///	/	/
6				Ж		/	/	/		/	X	X	X	X	X
7		V	З				/	/	/	/		/	/	/
8		В					/	/	///
9	Affricates				Ч		X	X	X	X	X	/	/		/
10			Ц				/	/		X	X	X	X	X
II	Plosives	Vl				К		/	///	///	///	///	/	/	/
12		Т					/	/	/	/	/
13		П					/	/	/	/	/
14				Г		///	///		/	/
15		V	Д				/	/	/	/	/
16		Б				/		/	/	/	/	/	/
17	Sonants	Lq	Л				/	/		/	/
18				Р		/		/	/	/	/	/
19		N	Н				/	/	/	/	/
20		М					/	/	/

consonants in the base CV syllables. Diphthongs Я,Е , Ю belong to the soft vowel as well. Vowels pairs А - Я ,О - Ё ,У - Ю , Ы - И ,Э - Е have similar properties in ranking according to their typical duration and positional variability, however , they essentially differ in spectral and temporal characteristics of the transition segment of sounds. At the same time the place of articulation of a consonant influence on the transition segment of a vowel, therefore CV syllables including consonants with the same place of articulation have similary characteristics for the initial part of each vowel. Consonants of CV syllables have the colour of the following wovel due to the effect of coarticulation and that effect is more associated with the place of articulation than the manner of production of consonants. Thus the characteristics of consonants and vowels determinates their context (allophonic) variability. The table is made for the strssed syllables and besides in the cells, which are formed at the intersection of consonant rows and vowel columns, the rough frequency of occurence of the base syllables from / 6, 14 / is given. From 200 possible CV combinations of Russian 25 are not used and 14 combinations occur vary seldom, those syllables are labeled by X and /// in the table. Seventy syllables corresponding to the table empty cells are used more often and cover about 50% any Russian speech and 91 syllables marked by / occur less frequetly. Thus the common number of Russian base elements is relatively not great. For unstressed CV syllables all consonants have realisations with rather good phonetic quality, and the number of vowels decreases up to three, including only sounds А , У and И / 8, 9 /. The duration of unstressed CV syllables shortens 1.5 or 2 times as much as the duration of the stressed syllables both at the expense of consonants and vowels.

SYNTHESIS OF ANY SPEECH ON THE
BASE OF CV SYLLABLES

In the process of synthesis on the base of CV syllables two aims were pursued first, a practic one, to develop a speech synthesator which could synthesize any Russian text any speakers voice, including a female one. The second aim was to make clear if it is possible to synthesize a speech signal perceived as qualitative continuous iformation, concatenated from a minimal alphabet of discretely pronounced speech elements. For that purpose a group of speakers pronounced (in according with the table) 175 syllables and 10 vowels afterwards stored in the computer "Eklipse - 330". For each CV syllable places of transitions from consonants to vawels were marked using a graphic display

with the following audition and correction, those data and syllable duration data were recorded into the computer memory. Unstressed, reduced syllables and separate consonants can be formed due to shortening of vowel duration of any stressed CV syllable. The perception of hardness or softness of Russian consonants was achieved due to the maximum vowel reduction of any syllable. The effect of coarticulation between cosonants in compound open syllables was produced by concatenating syllables reduced to minimum, those syllables having the same vowel as the base CV syllable. Coarticulation in the words consisted of concatenations of open syllables with different vowels was simulated with the help of addition of a short segment of the succeeding vowel to the end of the preceeding one. The duration of that segment depended on the contrast F-picture of the adjacent vowels and was increasing proportionally to that contrast increase. A more detailed description of CV syllable compilation speech synthesis is given in / 15, 16, 17 /. Algorithms and computer programmes of the syllable synthesis including phonetic transcription of any Russian text were developed by I.Orlov /18/, using the syllable interpretation of a letter record in accordance with the table. Concatenation of speech elements into a continuous piece of information was produced without any additional transformations exept the preliminary amplitude compression of the sygnal. The speech compiled of discretely pronounced syllables sounded as continuous and rather naturally with high percentage of word intelligibility equal to 97 - 99%. That experiment besides having practical significance proves that CV syllables are really the base elements of the speech.

CV ANALYSIS OF CONTINUOUS SPEECH

The syllable analysis of continuous speech pursueing an aim of automatic transcription of a speech signal is much more complicated than the problem of the speech synthesis. Difficulties of the speech analysis mainly depend on the variability of a speech signal and were briefed in Introduction. However, the choice of an analysis unit is of great importance since in addition the number and the type of the base speech elements are determined and their spectral and temporal characteristics become preliminary known as well. The continuous speech analysis as well as the speech synthesis is reasonable to carry out on the base of CV syllables. That approach is discussed in details in / 5, 6, 19, 20 /. That is why we brief here only some conclusions.

1. The number of base CV syllables as well as in the speech synthesis is equal to about 200.
2. A current analysis of the continuous

speech should be performed using fragments with duration of about 100-120 ms, in that case the dependence of spectral and temporal characteristics of CV syllables on the context and the position decreases and besides the analysed segment of a vowel shou ld be 2o-25 ms longer than of consonant. In addition to CV syllables it is necessary to extract separate consonants and vowels which form compound open syllables as CCV, CCCV, CVV etc. approximately at the same time window as CV segments. Naturally in that case very short sound · wouldn't be extracted, but their number in the Russian continuous speech is insignificant / 6 /.

3. It is useful to perform linear time normalisation of the CV fragment duration dependent on the speech rate typical for a definite speaker / 20 /.

4. A base problem in determination of rules for fragment extraction from continuous speech is parametrical representation of a signal. A lot of experiments show that the best speech representation is a formant one using pitch synchronisation / 6 , 19, 21 /.

CONCLUSION

The syllable approach has good prospects for usage in speech informatics, since it establishes sufficiently adequate correlation between physical and phonetic properties of continuous speech. However, that and higher levels of speech processing are specific for each national language and therefore they should be thoroughly studied for any language.

REFERENCES

/I/Кринов С.Н.,Савельев В.П.,Цемель Г.И. Классификация сегментов при распознавании устных команд.Труды АРСО-I3,ч.I, Новосибирск, 1984, IOI-IO3.

/2/ Зигангиров К.Ш., Сорокин В.Н. Применение последовательного декодирования к распознаванию слитной речи."Проблемы передачи информации",№ 4,1977, 8I-88.

/3/Кринов С.Н.,Слуцкер Г.С. Многоуровневая речевая диалоговая система "САПФИР". Труды АРСО-I4,ч.I, Каунас, 1986, 92-94.

/4/ Кринов С.Н.,Слуцкер Г.С. Автоматическое формирование звуковых эталонов слов по их орфографической записи.Труды АРСО-I4, ч.I, Каунас, 1986, стр.86.

/5/ Книппер А.В.,Мирошников В.С. Текущее выделение фрагментов в речи. Труды АРСО-6, Таллин, 1972, IO7-IIО.

/6/ Книппер А.В.,Махонин В.А. К описанию речевого сигнала."Речевое общение в автоматизированных системах". М.,"Наука", 1975, 46-59.

/7/ Чистович Л.А. и др. Речь.Артикуляция и восприятие. М.-Л.,"Наука", 1965.

/8/ Бондарко Л.В. Звуковой строй современного русского языка."Просвещение",1977.

/9/ Златоустова Л.В., Потапова Р.К., Трунин-Донской В.Н. Общая и прикладная фонетика. Изд. МГУ. М., 1986.

/IO/Francini G.L.,Debiasi G.B. and Spinabelli. Study of a system of minimal speech repr. units for Italian speech. JASA,v.43,№6,June 1968, 1282-1286.

/II/Tanaka A.,Togava F., et all. A study of the syllable oriented recognition of continuous speech. Speech Communication, 1983,v.2,N2-3, 207-210.

/I2/Потапова Р.К. Слоговая фонетика германских языков.М., "Высшая школа", 1986.

/I3/Книппер А.В. СГ-представление русской речи.Тезисы докл. конф."Просодия речи", (7-9 дек.), МПИИЯ,М.,1982, II7-II9.

/I4/Ёлкина В.Н., Юдина Л.С.,Статистика слогов русской речи.Сб. трудов ИМ СО АН СССР "Выч. системы",№10, 1964,58-78.

/I5/Вайншток А.П.,Книппер А.В.,Орлов И.А., Потапов В.Г. Фрагментная компиляция речи.Труды АРСО-I2,Киев-Одесса,1982, 387-389.

/I6/Вайншток А.П.,Книппер А.В.,Орлов И.А., Потапов В.Г. Способ слоговой компиляции речи. А/с №1075300,Б.И.№7,1984.

/I7/Книппер А.В.,Орлов И.А. Компиляция речи с учетом коартикуляции.Труды АРСО-I3, Новосибирск, 1984,150, 151.

/I8/Орлов И.А. СГ-слоговое преобразование текста для компиляционного синтеза речи. Труды АРСО-I4, ч.2,Каунас,1986, стр. 64.

/I9/ Knipper A. Acoustic events in CV syllables with liquid and nasal sounds. Signal Processing , N 3, 1981, 389-396.

/20/Книппер А.В. О нормировке СГ-фрагментов по темпу. Труды АРСО-I3,Новосибирск, ч.I, 1984, стр.IOO,IOI.

/21/Книппер А.В.,Махонин В.А.,Орлов И.А. Элементы формантного анализатора. В кн."Распознавание образов. Теория и приложения", М.,"Наука",1977,I2I-I25.