PHONETIC DATABASE OF THE RUSSIAN SPEECH VARIABILITY

Vladimir I. Kuznetsov, Tatiana Y. Sherstinova
St.Petersburg State University, Russia

ABSTRACT

Database of contemporary Russian speech is being created for the Computer Fund of the Russian Language. Database consists of the corpus of sound files and their descriptions. Database sound material is formed by recordings of the Phonetically Representative Texts pronounced by 1) standard Russian speakers, 2) regional Russian speakers, 3) speakers from the former USSR republics to which Russian is a second language, and 4) foreigners from European, Asian and American countries. The detailed description is being made for each syllable of the database. It comprises in particular the "real" transcription, "ideal" phonemic and phonetic transcriptions, attributes of the sound, duration of the signal segments, acoustic features, phonetic and paralinguistic comments. Database may be used in phonetic and related studies, for building of speech algorithms, and evaluation of speech resources, speech technologies and products.

1. INTRODUCTION

The importance of multi-level variability evaluation of speech follows from the fact that there is not such reality as ideal speech. Variability models of real speech should be taken into account for optimisation of many speech technologies and products in addition to accepted ideal rules. High variability of Russian speech is a well-known phenomenon which deserves special investigation. Thus, processing of normative (standard) Russian speech on the level of real transcription shows that there do exist much difference between individual normative speakers. The variety of regional or not-standard pronunciation is evidently higher. That is why in addition to the normative speech corpora it was decided to create database of the speech variability which will also include the recordings of. regional Russian speakers, speakers from the former USSR republics to which Russian is a second language, and foreigners from European, Asian and American countries. The main functions of the database are the following: preservation of contemporary Russian speech, it's detailed description, phonetic and related studies, building of speech algorithms, and evaluation of speech resources, speech technologies and products.

2. DATABASE STRUCTURE

Phonetic database of the Russian speech variability is designed as the new component of the Phonetic Fund of the Russian Language. The other components of the Phonetic Fund are the following ([1], [4]):

1. *Speech corpora* including the recordings:
- all possible *Consonant+Vowel* syllables;
- lists of selected words (comprising "orthoepically difficult");
- phonetically representative text and dialog;
- Russian frequency dictionaries.

Speech corpora has been recorded by standard Russian speakers, digitized and processed.

2. *Morphemic information* contains statistical data on Russian morpheme structure and peculiarities of its pronouncing.

3. *Word-form information* contains statistics on Russian word structure.

4. *Russian Automatic Transcriber* allows to receive detailed phonetic transcription of any written text reflecting either standard or regional pronouncing [2], [10]. Prescribed by norm transcription is called "ideal" and describes the complete-style pronunciation [5].

The structure of data presentation in the Russian Speech variability database is borrowed to the considerable extent from that of the database of the Normative Russian speech FON2 [11]. It consists of the corpora of sound files, their descriptions (relational database) and user-friendly interface. The data presented in relational database may be processed by means of DBMS utilities and specially created programs. The structure of descriptions was specially developed so that to present in details real characteristics of Russian speech. Sound corpora is formed by recordings of the Phonetically Representative Texts (PRT) pronounced by 1) standard Russian speakers, 2) regional Russian speakers, 3) speakers from the former USSR republics to which Russian is a second language, and 4) foreigners from European, Asian and American countries.

PRT was specially composed so that to comprise in different positions the most frequent Russian syllables which cover 75% of any text. Text contains words in different phrase positions, which may considerably modify their phonation. A pronunciation of the PRT by a speaker may be thus considered as his own personal model of Russian pronunciation.

The recordings have been made in specially equipped sound-treated chamber. The speakers read out the text, with which they had been acquainted beforehand. The digitised speech was segmented into open syllables (consonant(s) + vowel) and terminal consonants. Each syllable was transferred into separate file and segmented into allophones.

Syllable Number	Orthography	Ideal Phonemic Transcription	Ideal Phonetic Transcription	Real Transcription		
				Speaker 1	Speaker 2	Speaker 3
1	бы	bˠы	b°ы	b°ы	bɨ	b°ы
2	л ти	l t'i	ɫ t'i	ɫt'i	ɫt'i	ɫt'i
3	хи	x'i	x'ɯ	x'i	x'i	x'i
4	й се	j s'e	j s'e	⊥s'i̯e	⊥s'⊥	s'i̯ɛ
5	ры	ɾы	ɾы	ɾы	ɾыⁱ	ɾыi̯
6	й ве	j v'eⁱ	j v'eⁱ	⊥v'i	⊥v'i	v'i̯⊥
7	че	čɯ	č⊥	če	če	č ə
8	р	ɾ	ɾ	ɾ		

Phrase 1. «Был тихий, серый вечер.»

Syllable Number	Orthography	Ideal Phonemic Transcription	Ideal Phonetic Transcription	Real Transcription		
				Speaker 1	Speaker 2	Speaker 3
11	ду	du	du	dɤ	dᵚu	dʊ
12	л ве	l v'e	ɫ v'eⁱ	ɫv'i̯e	ɫv'i	ɫv'i⊥
13	те	t'i	t'⊥	t'e	t'ə	t'
14	р,	ɾ	ɾ	ɾ		
15	/					
16	сла	slɑ	sɫɑ	sɫɑ	sɫɑ	sl°ɑ°
17	бы	bы	b°ы	bы	bыⁱ	b°ы
18	й и	j i	j ⊥	⊥		
19	тё	t'o	t'ⁱɔ	t'i̯ø	t'ᵚo	t'ᵞø
20	плы	pɫы	pɫы	pɫɛ	pɫɨ	pɫы
21	й	j	j	⊥̥		

Phrase 2. «Дул ветер, слабый и теплый.»

Tables 1-2. Fragments of relational database (standard speakers).

(*Speaker 1* - female voice, St.Petersburg standard, *Speaker 2* - male, St.Petersburg standard, *Speaker 3* - male, Moscow standard)

A detailed description was made for each syllable of the database. It comprises in particular:

- ideal phonemic transcription;
- ideal phonetic transcription;
- stressed/unstressed position;
- "real" phonetic transcription (estimated by the expert-phonetician with the use of software tools);
- duration of the signal segments,
- acoustic features,
- phonetic and paralinguistic comments.

Real transcription of each syllable has been determined by the expert-phonetician by means of auditory and acoustic analysis. It takes into account as well its sonogram and oscillogram. Such method of transcribing to a certain extent is inevitably subjective. Nevertheless, it seems convenient to call the obtained transcription "real", especially in contrast to "ideal" one, made by Automatic Transcriber.

Parameters of descriptions were organized into relational database. Tables 1-2 demonstrate some fields of the database referring to the initial two phrases of the PRT pronounced by

three standard speakers (one woman and two men). Database structure for the interfered and regional speech is similar to it, though may contain additional comments.

NATURAL SPEECH AND IDEAL RULES
It is well-known, that in Russian speech full-style pronunciation never occurs consistently [5]. Because of that natural Russian speech always differs to some extend from the ideal one, based on traditionally prescribed rules.

Tables 1-2, reflecting samples of speech corpora recordings, allow to compare real transcription of syllables with the ideal phonetic and phonemic transcriptions, which has been got by means of the Russian Automatic Transcriber.

Evidently, the greatest difference was observed for the pronunciation of vowels. In fact, normative speakers may pronounce instead of theoretically predicted (or "near") vowels a variety of different sounds, even surprisingly unsimilar to the prescribed ones. On the other hand individual speakers may show an inclination to the use in speech of some or other particular sounds either generally or in concrete cases. Most of such "peculiarities" of standard speech are not usually noticeable in the process of speech communication and may be revealed only by means of computer analysis.

General conclusions concerning the comparison of real and ideal transcriptions are the following: Strong phrase position (under syntagmatic, logical or emphatic stress) assists the phonation striving to an ideal one (both phonetically and phonemically). Week phrase position on the contrary causes considerable modifications of the phonemes up to their complete falling out. In the result a number of vowels lose their own relevant features and even gain the characteristics of some other phonemes [9].

The detailed analysis of the actual vowel pronouncing, revealing complicated relations between real phonetic transcription and ideal phonetic transcription, sometimes induces to review traditional transcription rules used for the systems of connected speech synthesis and recognition (see [9]).

SPEECH EVALUATION IN THE DATABASE
Statistic database processing allows to evaluate actual variability of the Russian speech. Thus, the "distance" from any natural pronunciation and it's "ideal" variant in any phonetic aspect (both segmental and prosodic) may be statistically evaluated. The smallest distance will be observed for standard speakers (such as presented in Tables 1-2), the distance will increase for colloquial normative speech, not-normative and dialect speech, Russian speech by foreigners. We see our task to elaborate the procedures of speech estimation based on its comparison with the ideal pronunciation rules; first - on the material of normative speech, and later - not-normative and accent speech. The recordings of PRT seem to be the best material for such kind of research because: 1) they represent the models of Russian pronunciation for the concrete speakers, and 2) textual identity of recordings justify and simplify the comparison procedures.

The following method of speech estimation (that will be realized first in the segmental (transcriptional) aspect of normative speech) is proposed. It is planned to analyse the real transcription of database segments (allophones, groups of allophones and syllables) together with conditions of their realization (e.g., stress, phonemic identity, syntagmatic position, context, etc.).

All sound realizations may be divided then into the followings groups: 1) regular, which regularly coincide with ideal transcription (always or in determined cases); 2) not-regular, which may vary from ideal; 3) "inadequate", which practically never coincide with ideal; 4) personally specific, characterising the particular speaker. For the each group the sets of rules (conditions) should be defined, and the variational series for each parameter may be built. Generalization of the obtained material will produce the empirical model of Russian normative speech [12].

Using the same scheme it is planned to analyse further the variety of not-normative speech, and to determine the boundary values of estimation parameters that divide the pronunciation into "right" (normative) and "wrong" (not-normative). While analysing the groups of speakers (e.g., particular dialect speakers or native Americans studying Russian) the sets of common pronunciation features are to be defined. Only on this stage of research language pronunciation norm may be actually estimated and the adequate model of standard Russian speech may be created [12].

The described method may be used as well for evaluation of speech in other aspects, including prosody and temporal dynamics.

CONCLUSION
Variability models of natural speech should be taken into account for optimisation of many speech technologies and products. The described phonetic database will allow to solve a number of theoretical and practical tasks. In particular, actual data obtained in the result of the database processing allow to optimise transcribing rules and are being used for creation of the Automatic Flexible Transcriber of the Russian Speech [10]. Further development of the database is supposed to become the fruitful starting-point both for speech investigations and applied phonetic tasks solutions, such as Russian teaching, phonetic interference studies and forensic phonetics.

REFERENCES

[1] Bondarko, L.V. (1988) Phonetic Fund of Rusian Language. In: Bulletin of the Phonetic Fund of Russian Language. Bochum-Leningrad. N 1. P.15-22.

[2] Bondarko L.V., Bogdanova N.V., Ovcharenko E.B., Ryzhov I.V., Stepanova S.B. (1988) Automatic Transcriber of the Phonetic Fund of Rusian Language. In: Bulletin of the Phonetic Fund of Russian Language. Bochum-Leningrad. N 1. P.23-28.

[3] Bondarko L.V., Stepanova S.B. (1992) Text Module "Phonetically Representative Text". In: Bulletin of the Phonetic Fund of Russian Language. Bochum-Leningrad. N 4. P.132-134.

[4] Bondarko L.V., Svetozarova N.D., Skrelin P.A. (1992) Phonetic Fund of the Russian Language as the Research Program of the Department of Phonetics of the Leningrad State University. In: Bulletin of the Phonetic Fund of Russian Language. Bochum-Leningrad. N 4. P.5-16.

[5] Bondarko L.V., Verbitskaya L.A., Gordina M.V., Zinder L.R., Kasevitch V.B. (1974) Functional Styles and Types of Pronunciation. In: Voprosy jasykoznanija. N 2. P.64-70.

[6] Bondarko L.V., Zinder L.R., Stern A.S. (1977) Statistic Characteristics of Russian Speech. In: Hearing and Speech in Norm and Pathology. Leningrad

[7] Kuznetsov V.I. (1991) Vowels in Speech. Thes. of Ph,D. Dissertation. Leningrad State University.

[8] Kuznetsov V.I. (1997) Vowels in Russian Speech. St.Petersburg State University.

[9] Kuznetsov V.I., Sherstinova T.Y. Phonetic Transcription for the Systems of Speech Synthesis and Recognition. In: Proceedings of the International Workshop "Text and Dialog"'98. 22-26 September 1998, Brno.

[10] Shalonova K.B. (1997) Flexible transcriber for Russian Continuous Speech. In: Proceedings of the Second International Workshop "Speech and computer" SPECOM'97. 27-30 October 1997, Cluj-Napoka – Romania.

[11] Sherstinova T.Y. (1997) Database FON2 of Russian Normative Speech. In: Proceedings of the Second International Workshop "Speech and computer" SPECOM'97. 27-30 October 1997, Cluj-Napoka – Romania. P. 81-84.

[12] Sherstinova T.Y. (1998) Speech Evaluation in Russian Phonetic Database. In: Proceedings of the First International Conference on Language resources and Evaluation. Granada, Spain, 28-30 May 1998.