ForumAcusticum
2 0 0 5

Multilanguage Phonetic Database Formation (Comparative Analysis of Vibrant Sounds in Russian, English & German Speech)

Elena Loseva
Moscow State Linguistic University, Department of Applied and Experimental Linguistics, 119992 Moscow, Ostozhenka 38, Russia,
elenaloseva@yahoo.com, eloseva@pt.comcor.ru

The paper presents preliminary results of the comparative theoretical and experimental acoustic analysis of vibrant sounds in Russian, English and German languages. The undertaken research aims at giving description of vibrant allophone systems in the above-mentioned languages. One of the final purposes of the research is to find some common and distinctive features of vibrant systems in Russian, English and German. Such a database may be applied, e.g., in interactive expert systems for speaker identification, computer-mediated language learning systems, for speech synthesis. The results of the experiments can be used for further development of the theory of corpora linguistics, comparativistics, forensic linguistics. The research is being carried out under the guidance of the Head of the Department of Applied and Experimental Linguistics of the Moscow State Linguistic University D.Sc. of Philology, Sc.Degree of Professor, Academician of International Informatization Academy Potapova R.K.

1 Introduction

In the modern world of rapidly developing new information technologies the problem of speaker identification automatization is still at the top of the agenda, one of its main aspects is creation of new and development of existing speaker identification systems, databases.

The formation of representative phonetic databases is one of the main conditions for successful speech recognition, speech synthesis, speaker identification. For many languages such databases have been formed for many years and are available now, but still the number of standard databases meeting all requirements is not very large. The situation with the Russian language was even worse: no database of any satisfactory quality existed until 2001 when Cognitive Technologies wrote a Russian database for Intel Corporation [21].

Information of the segmental level as well as information of the suprasegmental level plays an important role in the process of speaker identification. Precise acoustic and articulatory properties of sounds are necessary for successful speech recognition and synthesis. There are still some uncertainties about different sound properties.

2 Brief description of Russian, English and German vibrants

The most widely-spread /r/-sounds in different languages are apical [r] (like the Russian one), retroflex approximant [□] (typical for American English), the so-called "tap", or "flap" [™], produced with a delicate touch of the tip of the tongue against the alveolar ridge, uvular [r] (like in French).

Russian palatalized and nonpalatalized /r/-sounds are sonorant vibrants by the manner of noise production and the type of closure, as for the active speech organ they are front cacuminal, alveolar as for the place of articulation, voiced regarding the involvement of voice, oral taking into consideration the position of the velum (the configuration of the vocal tract for /r/ and /r'/ is shown in Figure 1 [13]). The open phase of nonpaltalized /r/ has the following formant structure: F_1 ~ 400-600Hz, F_2 ~ 1300-1600Hz, F_3 ~ 1800-2300Hz, in the close phase the values of F_1, F_2 are decreasing (F_3 can hardly be traced in most cases), as well as the intensity of formants F_1, F_2, F_3 by ~ 2-15dB, 7-22dB and 6-8dB relatively. For /r'/ F_2 and F_3 are higher than for the main variant of nonpalatalized /r/ (F_2 ~ 1500-1900Hz, F_3 ~ 2300-2500Hz) [2, 13]. Russian vibrants can become voiceless, nasalized, they can be realized as flaps.

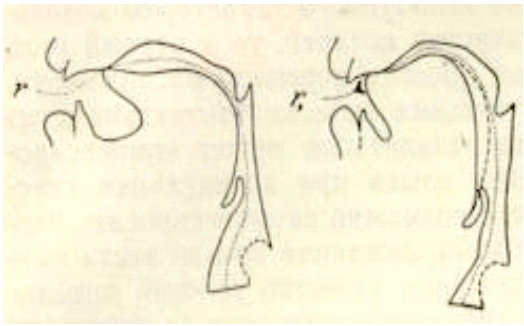

Figure 1: Configuration of the vocal tract for Russian /r/, /r'/

English /r/ is a voiced oral post-alveolar consonant which exhibits considerable labialization so that the following vowel can be labialized. The configuration of the vocal tract is rather complicated for this sound: there are three constrictions while articulating it: between the tip of the tongue and the alveopalatal zone, between the dorsum and the pharyngeal wall and in the area of the lips (as it is labialized). The typical F-picture of this consonant is very similar to that of the retroflex vowel [ə˞] (that is why /r/ is frequently called a semivowel) and can be characterized by the following frequencies: F_1 – 300 Hz, F_2 – 1000 Hz, F_3 – 1600 Hz (a very low F_3 can be regarded as a peculiarity of this English sound) [2]. The analysis of different classifications of English /r/-sounds and preliminary experimental analysis allow to speak about the following allophones of the English phoneme /r/: approximant, rolled, flap, voiceless, fricative, post-alveolar /r/. After plosives /t/, /d/ vibrants usually turn into fricatives, after voiceless sounds they tend to be wholly or partially devoiced.

Linguists usually single out such allophones of *German /r/* as postdorsal-uvular [R] and [ʁ], apical-alveolar vibrant [r], postdorsal- prevelar [ɐ], laryngeal /r/, vocalic allophone [□] [8, 23, 25, 29]. In the initial position of a syllable consonant allophones are more likely to occur, whereas in the final syllabic position after vowels vibrants are either vocalised or assimilated. In multisyllabic words and after the long [a:] complete assimilation has much greater probability. Vibrant allophones are usually devoiced after voiceless obstruents [8]. Mean formant frequencies for the allophones of German /r/ are shown in Table 1.

Table 1: Mean formant frequencies for [ʁ], [R], [r] between vowels and in the context ['bYRgeR]

Allophone	*Context*	$\overline{F}_1$	$\overline{F}_2$	$\overline{F}_3$	$\overline{F}_4$
[ʁ]	/V-V/	460	1113	2408	3347
	/Y-g/	468	1031	2240	3149
[R]	/V-V/	480	1205	2291	3095
	/Y-g/	486	1082	2328	2966
[r]	/V-V/	521	1265	1959	2408
	/Y-g/	478	1104	1981	2396

3 Experimental investigation of vibrant sounds

The experimental research is dedicated to the examination of acoustic properties of Russian, English and German vibrants. There is made an attempt to single out the most typical vibrant allophones. The singled out allophones were described in acoustical, articulatory and contextual terms. Such factors as the position of the neighbouring vowels regarding wordstress, the quality of the neighbouring consonants, word as well as phrase boundaries were taken into account.

3.1 Method

In the experiment involving *Russian speech* 510 phonetically representative phrases of 3-5 words read by male and female speakers served as speech material. The phrases included 623 vibrant variants, 72% of them were nonpalatalized variants, 28% – palatalized.

English data base is based on *Corpus IViE* formed by Oxford Phonetic Laboratory and Cambridge Speech Center, containing 36 hours of speech data obtained from adolescent speakers (six male, six female) from 9 regions of England. Taking into account the large size of the database for the present study some dialects have been chosen and analysed (Cambridge, Cardiff, Dublin). Five speaking styles were recorded:

- 22 phonetically controlled sentences with a range of grammatical structures,
- a read text, the fairy tale 'Cinderella',
- a retold version of the text,
- a map task (single sex pairs),
- discussion on the topic 'smoking'.

So different types of speech activity were involved in the experiment: prepared and unprepared reading (the sentences and the fairy tale), spontaneous and quasispontaneous speaking (retelling of the fairy tale, the map task, dialogues).

German speech material was formed from the read and spoken texts (male and female speakers of different age) from CDs and Internet.

The experimental research included measuring of the following parameters:

– frequencies of the first three vibrant formants;

– amplitudes of the first three vibrant formants;

– temporal characteristics: phrase duration, vibrant duration and duration of vibrant components.

3.2 Hypothesis

As it is well-known the quality and duration of consonants depend on the rhythmic structure of the word and the position in the phrase.

It was supposed that the following factors would influence the realization of vibrants:

– Context

It may be assumed to be a working hypothesis that the quality of neighbouring sounds as well as the position in the word and in the phrase will influence the quality of the vibrant. For example, voiceless consonants can devoice vibrants.

– Types of speech activity

It was supposed that the quicker rate of speaking in comparison with unprepared reading will lead to a greater variety of allophones and duration shortening of vibrants and their components.

3.3 Results

As a result of the Russian speech material research the following most typical allophones were singled out with the help of two criteria:

– 'quantity' of vibrant components *(for nonpalatalized /r/:* without any vibrations, with one, two, three, four vibrations; *for palatalized /r'/:* without any vibrations, with one, two vibrations);

– 'quality' of vibrant components (*for /r/:* the main variant, fricative, devoiced, nasal, approximant, creaky; *for /r'/:* the main variant, approximant, devoiced, creaky).

It is noteworthy that the singled out allophones were influenced greatly by the context. Thus, for example, fricative allophones were pronounced mainly before voiced and voiceless fricative sounds, devoiced vibrants appeared before and after voiceless plosives, before a pause, in the phrase-final position.

According to the analysis of the English speech material it has been found out that English /r/-sounds are mainly realised as approximants, less frequently as rolled /r/, flaps. Women tend to stay away from pronouncing the linking /r/ in spontaneous speech unlike men, which coincides with what Laurie Bauer claims [24]. As far as the quality of linking /r/ is concerned, it was mainly pronounced as an approximant which coincides with the statement made by some linguists, for example [24]. As we know, some phoneticians insist on its being a flap [17], and it was actually sometimes the case. Very rarely one could come across vibrants in this position. Fig. 2, 3 illustrate the spectrograms of linking vibrants as an approximant and a vibrant.

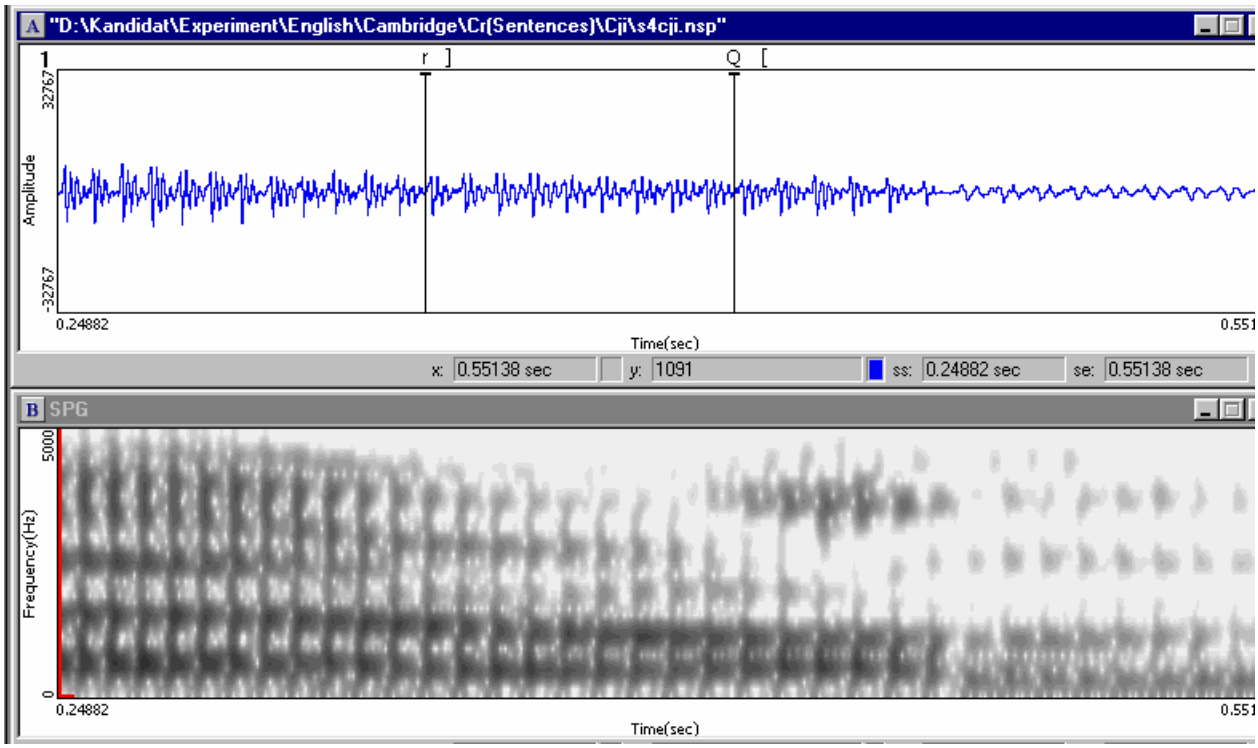

Figure 2: English linking /r/ as an approximant

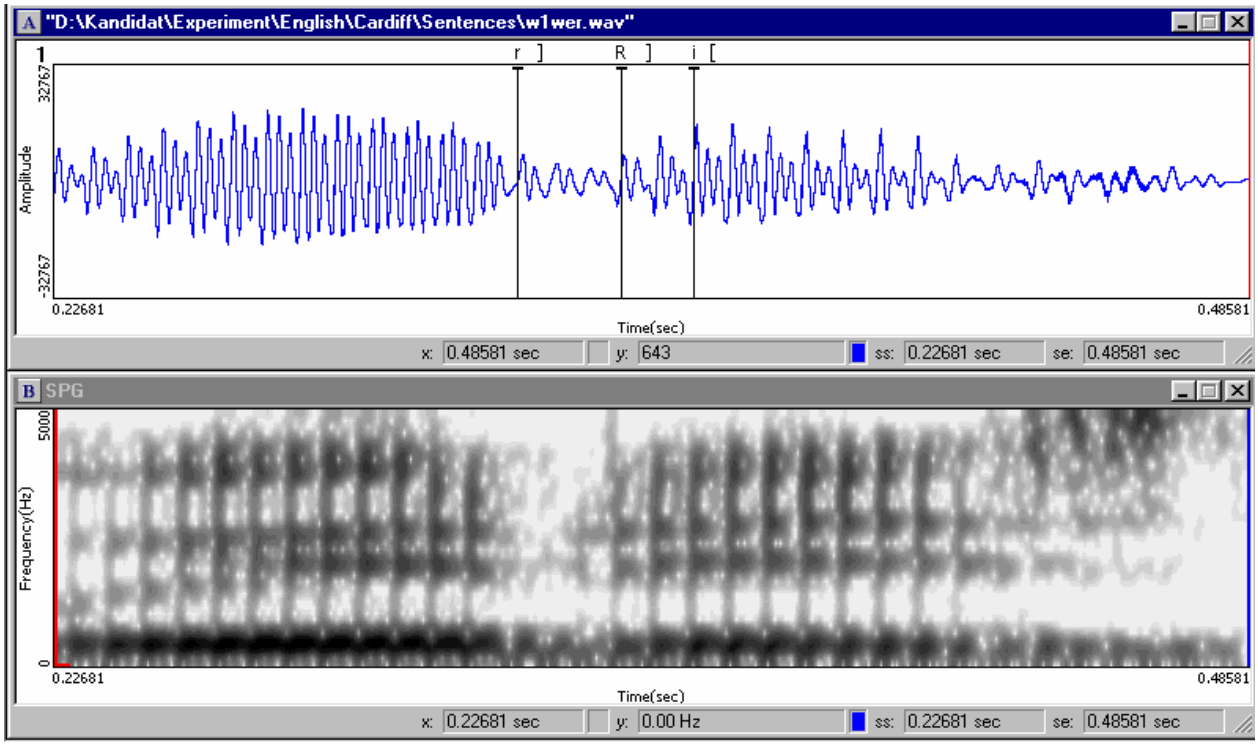

Figure 3: English linking /r/ as a vibrant

As for German vibrants as it was expected partially devoiced vibrants appeared after voiceless plosives and fricatives.

Speaking has really influenced the length of the sounds under consideration and resulted in shorter vibrants and their components.

4 Summary

Practically the obtained results can serve for enlarging of the existing database and developing of a new database for foreign languages of interactive expert systems for speaker identification, computer-mediated language learning systems. Hopefully the results of the undertaken research of vibrant systems in Russian, English and German can be used for further development of the theory of corpora linguistics, linguocriminalistics, comparativistics.

References

[1] The identification of speakers by means of the automatic system “Dialect”. Manuel for experts. // Popov N.F., Linkov A.N., Kurachenkova N.B., Bajcharov N.V., Fesenko A.V. 1996, 102 p. (in Russian)

[2] Kodzasov S.V., Krivnova O.F. General phonetics M., RSHU, 2001, 592 p. (in Russian)

[3] Loseva E.V. Formation of Phonetic Database for Experts. // Proceedings of 13-th International scientific conference 'Informatization and forensic information security' (Moscow, 25-26 May, 2004. – 528p.) p.368-372

[4] Matusevich M.I. Modern Russian language. Phonetics. M., Prosvetschenije, 1986, 288 p. (in Russian)

[5] Potapov V.V. On Language Contrastive-Comparative Analysis of English and Russian Phonetic Systems // Proceedings of the International workshop "Speech and Computer" SPECOM'2003. Moscow, 27-29 October, 2003 p.276-283

[6] Potapova R.K. Linguistic database of Electronic Encyclopaedia for forensic phonetics experts, MSR-Phono-Estra, M., 1999, 2000 CDROM (in Russian).

[7] Potapova R.K. What questions answers forensic expertise // journal “Russian Justice” №10, 2000 (in Russian)

[8] Potapova R.K. Articulatory variance of German speech // journal “Questions of Speech Science” №6, 2002 (in Russian) p.82-100

[9] Potapova R.K. Reconstruction of speaker portrait on the basis of cognitive-linguistic features // "Speech Science in Theory and Experiment " MSU – ISAA M., 2002 (in Russian)

[10] Potapova R.K. Speech: communication, information, cybernetics M.: 2-nd ed., completed, 2001 (in Russian)

[11] Potapova R.K., Popov N.F., Linkov A.N. et al. Criminalistic phonography M., 2000 (in Russian)

[12] Torsujev G. P. Constancy and variability in the phonetic system 1975 (in Russian)

[13] Fant G. Acoustic theory of speech production M., 1964, 284 p. (in Russian)

[14] Jacobson R., Fant G., Halle M. Introduction to speech analysis. Distinctive features and their correlates // New in Linguistics. M.,1962, ed.2 (in Russian)

[15] Augustin Simo Bobda, Hans-Georg Wolf, Lothar Peter. Identifying regional and national origin of English speaking Africans seeking asylum in Germany // Forensic Linguistics 6(2), 1999

[16] Dalston Rodger M. Acoustic characteristics of English /w, r, l/ spoken correctly by young children and adults // 1974, p.296-303

[17] Daniel Jones. An outline of English Phonetics (6^{th} ed.). − New York: Teubner, 1939, 326 p.

[18] Daniel Jones. The Pronunciation of English (4th ed.) Cambridge 1956

[19] Geoff Lindsey, Allen Hirson. Variable robustness of nonstandard /r/ in English: evidence from accent disguise // Forensic Linguistics 6(2), 1999

[20] Gimson A.C. An Introduction to the Pronunciation of English (4th ed.) London 1989

[21] Kibkalo A.A., Lotkov M.M. Choice of Phonetic Alphabet for Russian LVCSR System // Proceedings of the International Workshop "Speech and Computer" SPECOM'2003. Moscow 27-29 October 2003, p. 102 -105

[22] Kouznetsov V., Chuchupal V., Makovkin K. Design and Implementation of a Russian Telephone Speech Database // International workshop "Speech and Computer" SPECOM'99. Moscow, 4-7 October, 1999, p.179-181

[23] Krämer Wolfgang. Akustisch-phonetische Untersuchungen zum vokalischen /R/-Allophon des Deutschen. Hamburg: Buske 1979

[24] Laurie Bauer. Linking /r/ in RP: some factors // Journal of the International Phonetic Association 1984, vol.14, no.2, p.74-79

[25] Lindner Gerhard. Frequenz und Luftverbrauch beim Kehlkopf-R // Zs.f. Phonetik 11, 1958

[26] Olive Joseph P., Greenwood Alice, Coleman John. Acoustics of American English speech // AT&T Bell Laboratories, Springer-Verlag 1993

[27] Sabater Maria-Josep Solé. The phonetic basis of phonological structure: the role of aerodynamic factors // Universitat Autònoma de Barcelona, p.77-94

[28] Sivertsen Eva. Cockney Phonology 1960

[29] Wiese Richard. The Phonology of German 1996