



HAL
open science

Database description: Russian fricatives recorded in 198 real speech sentences from 59 speakers

Natalja Ulrich

► To cite this version:

Natalja Ulrich. Database description: Russian fricatives recorded in 198 real speech sentences from 59 speakers. *Data in Brief*, 2023, 48, pp.109205. <10.1016/j.dib.2023.109205>. <hal-04573358>

HAL Id: hal-04573358

<https://hal.science/hal-04573358v1>

Submitted on 13 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License



Data Article

Database description: Russian fricatives recorded in 198 real speech sentences from 59 speakers

Natalja Ulrich

Lab Dynamics of Language UMR 5596, CNRS and University Lyon 2, France

ARTICLE INFO

Article history:

Received 7 March 2023

Revised 1 May 2023

Accepted 2 May 2023

Available online 11 May 2023

Dataset link: [Natalja Ulrich: Russian Fricatives \[Dataset\]](#), Université Lumière Lyon 2. Distributed by SWISSUbase, Lausanne, 2022. (Original data)

Keywords:

Russian fricatives

Large read speech dataset

Machine learning

Speaker identification

ABSTRACT

This speech dataset is primarily designed to investigate linguistic and speaker information in fricative sounds in Russian.

Acoustic recordings were obtained from 59 students (30 females and 29 males) between 18 and 30 years. Eighteen participants were recorded in a second session. The participants were born and lived since their early childhood in St. Petersburg. The participants did not report any speech or hearing impairment. The recording sessions were conducted at the phonetic laboratory of the Phonetic Institute in St. Petersburg, in an audiometric booth using the recording program Speech-Recorder version 3.28.0 at a sample rate of 44.1 kHz (16-bit encoding). During the recordings, a clip-on microphone (Sennheiser MKE 2-P) was placed at a distance of 15cm from the speakers' mouth and connected through an audio interface (Zoom U-22) to a laptop computer.

The participants were instructed to read 198 randomized sentences from a computer screen. The fricatives [f], [s], [ʃ], [x], [v], [z], [ʒ], [sʲ], [ɕ], [vʲ], [zʲ] were embedded into those sentences. Two sentence structures were designed to obtain each real-word lexemes produced in three different contexts. The first type of sentence is a so-called carrier sentence with the structure of "She said "X" and not "Y" ". Minimal pairs of real words, containing one of the 11 tested fricatives were placed in both "X" and "Y" positions. The second type of pre-designed sentence was a natural language sentence including each of the lexemes.

E-mail address: ulrichnatalja@gmail.com

<https://doi.org/10.1016/j.dib.2023.109205>

2352-3409/© 2023 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

All raw audio files were first automatically pre-processed by applying the online tool Munich Automatic Segmentation system. Then, the files of the first recording session were filtered below 80 and above 20050 Hz, and the boundaries were manually corrected using Praat.

The dataset consists of 22,561 fricative tokens. The number of observations per sound differs across categories, because of their natural distribution. The dataset is made available as a collection of audio files in *wav* format along with companion Praat *TextGrid* files for each sentence. Target fricatives are furthermore available as individual *wav* files.

The whole dataset can be accessed with the DOI <https://doi.org/10.48656/4q9c-gz16>.

Additionally, the experimental design allows the investigation of other sound categories. The number of speakers recorded gives further possibilities for phonetic-oriented speaker identification studies.

© 2023 The Author(s). Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Specifications Table

Subject	Signal processing
Specific subject area	Acoustic dataset containing recordings of 59 speakers. The full dataset includes 198 sentences per speaker and a fricative dataset of 22,561 tokens extracted from these sentences.
Type of data	Automatically forced-alignment (MAUS) audio files Automatically forced-alignment (MAUS, manual corrected (Praat) and Filtered (pass Hann filter 80 Hz- 22500Hz) audio files TextGrid with three Tiers Sentence list Metadata Table Praat scripts
How the data were acquired	In an audiometric booth at the phonetic laboratory of the Phonetic Institute in St. Petersburg 59 Russian native speakers were recorded reading 198 sentences. The recording program <i>Speech-Recorder</i> version 3.28.0 was used. A clip-on microphone (Sennheiser MKE 2-P) was placed at a distance of 15cm from the speakers' mouth and connected through an audio interface (Zoom U-22) to a laptop computer.
Data format	Raw Analyzed
Description of data collection	Participants were recorded while reading 198 sentences presented to them by the experimenter, one by one on a computer screen.
Data source location	Phonetic Institute in St. Petersburg, Russia
Data accessibility	Repository name: Swissubase (www.swissubase.ch) Data identification number: 2183 The database is available at (13:italic) www.swissubase.ch/(13:italic) through the DOI: https://doi.org/10.48656/4q9c-gz16 . To access the data, a SWITCH edu-ID at https://eduid.ch needs to be created. This ID allows the user to log in to <i>SWISSUbase</i> and to download the data along with the Supplementary materials. The data has a closed license and is only available for teaching and research purposes. The following two links provide additional information on data policy and usage guide.

(continued on next page)

	Data Policy: https://resources.swissubase.ch/wp-content/uploads/2022/06/Linguistics_Data-Policy_en-1.pdf
	User guide: https://resources.swissubase.ch/wp-content/uploads/2022/06/Linguistics_User-Guide_en-2.pdf
	Please cite the database as follows: Natalja Ulrich: Russian Fricatives [Dataset]. Université Lumière Lyon 2. Distributed by SWISSUbase, Lausanne, 2022. https://doi.org/10.48656/4q9c-gz16 .
Related research article	Ulrich, N., Allassonnière-Tang, M., Pellegrino, F., & Dediu, D. (2021). Identifying the Russian voiceless non-palatalized fricatives /f/, /s/, and /ʃ/ from acoustic cues using machine learning. <i>The Journal of the Acoustical Society of America</i> , 150(3), 1806-1820.

Value of the Data

- The dataset represents a large collection of recordings of Russian sentences produced by native speakers (59), with a focus on Russian fricatives.
- The data was gathered under highly controlled and systematic conditions.
- All recordings are automatically forced aligned and the boundaries of the target fricatives were all manually inspected and corrected.
- The dataset can be used for a number of teaching and research purposes for instance for phonetic research, forensic speaker comparison, and/or ASR research in fricatives and other phoneme classes. For instance, acoustic correlates of contrastive fricatives can be investigated. Furthermore, speaker or gender identification tasks can be performed using isolated fricatives, full sentences, or words.

1. Objective

The construction of the dataset was primarily motivated by the intention to investigate linguistic and speaker-specific information of complex sounds such as fricatives. Even though fricative sounds have been extensively studied in the past, accessible datasets designed for investigations of fricative sounds in a format suitable for reuse are rare or inaccessible. In terms of data size, most studies on fricatives have been based on data from a few speakers. In general, the data and research available are very limited, especially for understudied languages such as Russian.

2. Data Description

The data consists of acoustic recordings, a sentence list used for the recordings, a metadata table on the target fricatives containing a number of information, and two Praat scripts used to filter the *wav* files and to extract the target fricatives.

2.1. Acoustic Recordings

The entire dataset consists of recordings from 59 people, and 18 participants were recorded a second time on the same stimuli.

The data results in 77 recordings. Each of the 77 recordings contains the same list of sentences, read by the participants. Target fricatives (293) were extracted from the sentence of the first recording session and built the fricative dataset.

All 77 recordings are stored in two folders: *Recording_1* and *Recording_2*.

The folder *Recording_1* contains 59 subfolders with the recordings of each speaker and one subfolder *all_sound_session_1*.

Each of the 59 subfolders contains all sentences as *wav* and *TextGrid* files (N=396).

Subfolder *all_sound_session_1* contains extracted fricatives from the first recording session (N= 17287).

The folder *Recording_2* contains 18 subfolders with the recordings of the second session.

2.2. Data Coding Wav and TextGrid Files

To avoid showing the speakers' identities, the recordings were anonymized.

The names of the speaker's subfolders (N=77) e.g. *01_F_22_1* are coded as follows: The first two digits *01* refer to speaker ID (01-59). The characters *F* and *M* refer to the speaker's gender. The following two digits *22* indicate the age (18-30) of the participant at the time of the recording (2018). The last number *1* indicates the recording session (1,2).

Each speaker's subfolder contains the recordings of 198 sentences as *wav* and *TextGrid* files.

The filenames of the sentences e.g., *01_F_22_1_001* represent the same information as the folder names and additionally the sentence number in the last three digits.

The single fricative files, stored as *wav* files in the subfolder *all_sound_session_1*, follow the same coding patterns. For instance, the filename *01_F_1_001_11_v* contains speaker ID, gender, recording session, sentence number, information on the interval number (obtained from the *TextGrid*) of the third *Tier*, and the fricative label in SAMPA.

2.3. Sentencelist

The *csv* file *sentence_list* represents a table listing all recorded sentences. The first column *sentenceNo* corresponds to the sentence number in the file name. The second column presents the sentence in Russian and the third gives a transcription in IPA. A fourth column lists separately the target words from which the fricatives were extracted in IPA. The last column *comments* indicate which sentences were excluded from the fricative extraction.

More specifically, some of the sentences turned out to be miss-constructed and were excluded from the fricative dataset. A few sentences were modified in the middle of the experiment and were therefore different between the participants (N=3). These sentences are marked in the sentence list. Some of the natural sentences turned out to be useless for the defined research purposes and were also excluded from the dataset (N=11). To fill the gaps of missing fricative examples from words embedded in natural sentences, these fricatives were extracted from other words.

2.4. Fricative Metadata

The *csv* file *metadata* table gives all kinds of information on the target fricatives.

This table was used to extract the target fricatives from the first recording session and can be used to extract the same fricatives from the second session.

The columns *SentenceNo* and *IntervalNo* indicate the sentence and interval number by which the fricatives can be identified.

SentenceTyp refers to one of the two types of sentences: *CS* stands for carrier sentence and *NC* for non-carrier (natural) sentence (for the explanation please refer to the next section).

The column *Position* specifies the location of the target fricative in a carrier sentence. The annotation *1* stands for "X" position and *2* for "Y" position.

The column *FricativePosition* shows the word position of the fricative. The annotation *B* stands for a word-initial position, *M* for word-medial, and *E* for word-final.

The columns *SAMPA* and *IPA* indicate the transcription format.

The columns *PrecedingSound* and *FollowingSound* contain, as the column names indicate, the preceding and following sounds of the target fricatives.

The columns *Voicing* and *Palatalization* show whether the target fricative is voiced or palatalized.

The sentence number, interval number and fricative label of each fricative filename correspond to the columns *SentenceNo*, *IntervalNo* and *SAMPA* in the metadata table.

2.5. Praat Scripts

Furthermore, two Praat scripts are available. The Praat codes are used to filter the sound files (*1_filter_the_noise*) and to extract the target fricatives (*2_fricative_extraction*).

3. Experimental Design, Materials and Methods

3.1. Participants

The participants were 59 students (30 female) between 18 and 30 years old, studying at different departments of St. Petersburg University in Russia. All participants were born or lived in St. Petersburg since early childhood and still lived there at the time of the recording.

No participants reported any speech or hearing impairment.

3.2. The Stimuli

To obtain recordings of the target fricatives, a list of 94 real words containing minimal pairs (words that vary by only a single sound contrasted by place of articulation, voicing, and palatalization) of the target fricatives was collected.

The fricatives [ʃ], [ʒ] or [x] were not recorded because there are only a very few minimal pairs in Russian with these sounds.

Each word contains one of the target fricatives in a) word-initial position as a CV syllable (e.g. *fara*), b) word-medial position in inter-vocal context (e.g. *maSa*), and in c) word-final position in a VC syllable (e.g. *ves*). The stimuli consist of a list of 198 Russian sentences, which include 293 target fricatives. All sentences are listed in the sentence list.

Each word appears three times in the stimuli list and is embedded in two different sentence structures: i) carrier sentence with the structure of “She said “X” and not “Y.” Minimal pairs of real words containing one of the 11 tested fricatives were placed in both “X” and “Y” positions, ii) natural language sentences including each of the lexemes.

Target fricatives in carrier sentences always occur in inter-vocal positions, except for fricatives at the end of the word or at the end of the sentence.

In natural sentences, the vowel context was not controlled. Therefore, fricatives in stimuli of type a) (CV words) can appear with a preceding consonant.

It is noted that the number of tokens per sound differs strongly between the categories due to the naturally uneven distribution found in Russian as shown in [Table 1](#).

As an illustration, the fricative [x] only occurs in a few minimal pairs contrasted by place of articulation. This fricative was therefore infrequently recorded.

Fricatives [z] and [v] also appear very rarely in words and are therefore included in only a few minimal pairs contrasted by place of articulation or palatalization. Due to these reasons, these three sounds contain a very small number of examples per speaker.

Furthermore, the preceding and following vowels vary between the fricatives.

Table 1

Overview of the fricatives in the dataset and their segmentation status.

FRICATIVE	[f]	[s]	[ʃ]	[x]	[v]	[z]	[ʒ]	[sʲ]	[ç]	[vʲ]	[zʲ]
Frequency by speaker	36	67	55	11	29	27	24	15	15	11	3
Total automatically and manually aligned	2301	3951	3245	649	1711	1593	1416	885	885	649	177
Total automatically aligned	648	1206	990	198	522	486	432	270	270	198	54

3.3. Recording Procedure

The recording sessions were conducted in an audiometric booth at the phonetic laboratory of the Phonetic Institute in St. Petersburg.

Demographic data, such as sex and age, were registered before the experiment started.

Eighteen participants agreed to a second recording session.

The recording program Speech-Recorder version 3.28.0 <https://www.bas.uni-muenchen.de/Bas/software/speechrecorder/> was used at a sample rate of 44.1 kHz (16-bit encoding).

A clip-on microphone (Sennheiser MKE 2-P) was placed at a distance of 15cm from the speakers' mouths. The microphone was connected through an audio interface (Zoom U-22) to a laptop computer.

The participants were then instructed to read the 198 sentences aloud from a computer screen. The sentences were presented one by one in random order. The participants could repeat a sentence in the case of a production error.

3.4. Forced-Alignment of the Raw Data

All raw audio files were automatically pre-processed using the Munich Automatic Segmentation System, MAUS [3,4] available online at <https://www.bas.uni-muenchen.de/Bas/BasMAUS.html>. During this process, *TextGrid* annotation files were generated. The *TextGrid* output contains three *Tiers* and is shown in Fig. 1. The first *Tier* shows the sentence segmented by word in Russian. In the second *Tier*, the sentence was transcribed in SAMPA [2]. The third *Tier* provides time-aligned word segmentation at a phonemic level.

3.5. Manual Segmentation

In *Tier* 3, the boundaries of the target fricatives were manually corrected using *Praat* version 3.9 [1]. See Fig. 1 for an example.

It is to be noted that only the recordings of the first session were manually corrected, and the 18 recordings from the second session were only automatically reprocessed.

In total, there are 22561 fricatives in the dataset, of which 17287 (all target fricatives of the first recording session) were extracted and are available as individual *wav* files (Table 1).

In order to define the onset and offset of the full consonant, the broadband spectrogram was considered more relevant than the start of an aperiodic waveform with rising zero crossing rates. In intervocalic fricatives, the presence of formant columns is defined as the onset and offset of the fricative (following [5]).

Some speakers ended their voiceless fricatives with a somehow long post-aspiration in intervocalic positions and/or when the fricative appeared at the end of the word and sentence.

In these cases, the fricatives were segmented according to the changes in high-energy events and the post-aspiration part was not considered.

The voiced fricatives represented also a segmentation challenge, because the waveform and spectrogram may be insufficiently informative to define the onset and offset. The boundaries of

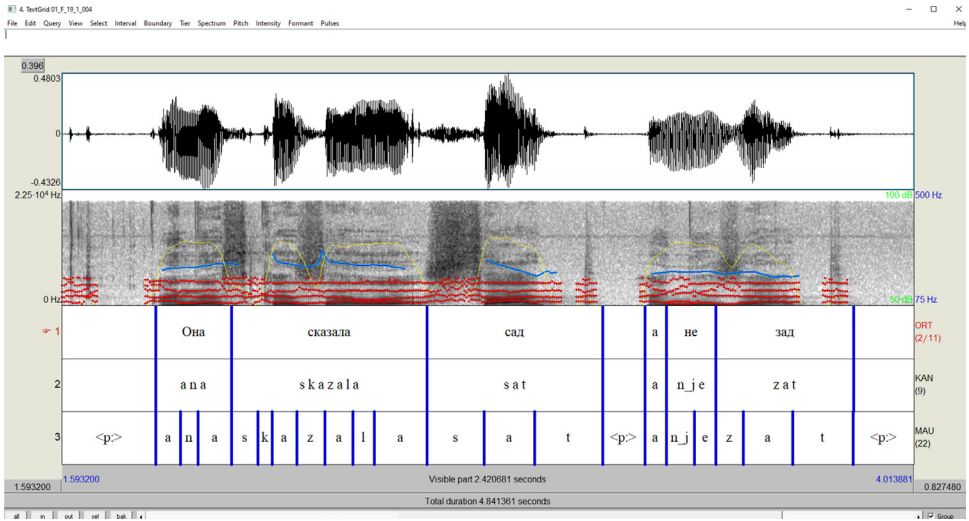


Fig. 1. The Praat window presents an example of a sentence. It shows the waveform, the spectrogram and three Tiers.

these sounds were identified according to perceptual judgments. In general, it should be noted that it is hardly possible to standardize segmentation criteria across all fricatives and speakers.

If the user wishes to define the on and offset of the fricative boundaries differently, they can carry out a manual or automatic correction using their own criteria.

3.6. Filter

It should be noted that the recordings contain some background noise, which might influence the acoustics of different sounds. The noise is generated by the laptop and the room ventilation.

All recordings from session one (*Recording_1*) were filtered before the adjustment of the boundaries. To get rid of some of the noise, the audio files were filtered below 80 and above 20050Hz applying a *pass Hann band* filter in Praat.

3.7. Further Usage Notes

The provided data and materials give researchers interested in the target fricatives the possibility to perform analysis with no or minor adjustments.

Researchers intending investigations beyond these fricatives are highly recommended to carefully check the data for the following errors in the *TextGrid*: spelling errors in *Tier 1*; wrongly transcribed vowels by the automatic pre-processing in the *Tier 2* and 3; displaced boundaries in *Tier 2* and 3, the on and offset boundaries in both *Tiers* were only manually corrected for the target fricatives, all other segments need a manual correction.

Ethics Statements

At the time the recordings were taken, no ethical approval was needed.

Participants were briefly introduced to the purpose of the experiment, the expected duration, and the procedure.

To avoid biasing the productions, we waited until after the recordings were made to announce to each participant that the project was concerned with research on Russian fricative.

Participants who were invited to a second recording session were informed after the second recording of the purpose of the experiment.

The participants were told that they have the right to withdraw at any time during (and after) the experiment and they were provided with the contact details of a person that can answer all their questions concerning the research and their rights. The participants were compensated for their participation.

The recordings do not contain any personal information about the speaker. Furthermore, as explained in the previous section, all recordings were anonymized.

The dataset has a closed license and is available only for research and teaching purposes. To access the data, a SWITCH edu-ID at <https://eduid.ch> needs to be created. This ID allows the user to log in to SWISSUbase and to download the data along with the Supplementary materials. In case of a download of the data, the author receives a notification with information on the person, who downloads the data and the purpose of use.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Natalja Ulrich: Russian Fricatives [Dataset]. Université Lumière Lyon 2. Distributed by SWISSUbase, Lausanne, 2022. (Original data) (Swissubase).

CRediT Author Statement

Natalja Ulrich: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Project administration, Funding acquisition.

Acknowledgments

I wish to thank our participants, the Phonetic Lab in St. Petersburg (and the sound engineer Tatiana Chukaeva in particular), the University of Zürich for financial support, technical support, and help with the design of the experiment (to Volker Dellwo in particular).

NU was partly supported by a grant from the Doctoral Program of Linguistics of the Faculty of Arts and Social Sciences, University of Zürich, Switzerland, NU, was funded by the IDEXLyon Fellowship Grant 16-IDEX-0005 (2018–2021), and indirectly by the Labex ASLAN (ANR-10-LABX-0081) of the University of Lyon within the program Investissements d’Avenir (ANR-11-IDEX-0007) of the French National Research Agency (ANR).

References

- [1] P. Boersma, D. Weenink, Praat: Doing phonetics by Computer [Computer Program]. Version 6.2.14, 2022 <http://www.praat.org/>.
- [2] D. Gibbon, R. Moore, R. Winski, *Handbook of Standards and Resources for Spoken Language Systems*, Walter de Gruyter, 1997.
- [3] T. Kisler, U. Reichel, F. Schiel, Multilingual processing of speech via web services, *Comput. Speech Lang.* 45 (2017) 326–347, doi:10.1016/j.csl.2017.01.005.
- [4] F. Schiel, *Automatic Phonetic Transcription of Non-Prompted Speech*, 1999.
- [5] R. Skarnitzl, P. Machač, *Principles of Phonetic Segmentation*, 2013.