



Language Manual

HQ and CO Russian

Language Manual: HQ and CO Russian

Published 4 February 2015

Copyright © 2013-2015 Acapela Group

All rights reserved

This document was produced by Acapela Group. We welcome and consider all comments and suggestions. Please, use the *Contact Us* link at our website:
<http://www.acapela-group.com>

Table of Contents

1	General	1
2	Letters in orthographic text.....	2
3	Punctuation characters.....	3
4	Other non alphanumeric characters	4
5	Number processing.....	7
6	How to change the pronunciation.....	18
7	Russian phonetic text	19
8	Non-Cyrillic input	24
9	Abbreviations.....	25
10	Web-addresses and email	26

1 General

This document discusses certain aspects of text-to-speech processing for the Russian text-to-speech system, in particular the different types of input characters and text that are allowed.

This version of the document corresponds to the High Quality (HQ) and Colibri (CO) Russian voices.

Please note that the *User's Guide*, mentioned several times in the manual, is called *Help* in some applications.

Note: For efficiency reasons, the processing described in this document has a different behaviour in some Acapela Group products. Those products are:

- Acapela TTS for Windows Mobile
- Acapela TTS for Linux Embedded
- Acapela TTS for iOS
- Acapela TTS for Android



For these products, the default processing of numbers, phone numbers, dates and times has been simplified for the low memory footprint (LF) voice formats. Developers have the possibility to change the default behaviour from *simplified* to *normal* preprocessing by setting corresponding parameters in the configuration file of the voice. Please see the documentation of these products for more information. In the following chapters, each simplification will be described by the indication *[not SP]* following the description of the standard behaviour. The *SP* in the indication stands for *Simplified Processing*.

2 Letters in orthographic text

Russian Cyrillic and Latin characters from A-Я, а-я, А-Z and а-z may form a word. Certain other characters are also treated as letters, notably those used as letters in other Slavic languages using Cyrillic alphabet “с, j, ђ, ѣ, ѱ, э, r, к, ѿ, ѹ, ѻ, Ѽ, ѽ, i, ї, є”, but they are accordingly mapped onto close in pronunciation Russian sounds “с, ў, ч, ѳ, ж, з, з, к, л, н, у, и, о, э”, when occurring in a word.

Words written in Roman characters are treated as English input, but are pronounced with Russian sounds.

Characters outside of these ranges, i.e. numbers, punctuation characters and other non-alphanumeric characters are not considered as letters.

3 Punctuation characters

Punctuation marks appearing in a text affect both rhythm and intonation of a sentence. The following punctuation characters are permitted in the normal input text: , ; . ? ! () — “ ” « » { } []

3.1 Comma, colon, semicolon and em dash

Comma ',', colon ':', semicolon ';' and em dash '—' cause a brief pause to occur in a sentence.

3.2 Quotation marks

The quotes '“”' and '«»' and '""' appearing around a single word or a group of words cause a brief pause after the quoted text. The opening quotes are not read out.

3.3 Full stop

A full stop '.' is a sentence-final punctuation mark, which causes the end-of-sentence intonation pattern and is accompanied by a somewhat longer pause. A full stop may also be used as a decimal marker in a number (see chapter *Number processing*) and in abbreviations (see chapter *Abbreviations*).

3.4 Question mark

A question mark '?' ends a sentence and causes a rising yes/no question-intonation.

3.5 Exclamation mark

The exclamation mark '!' behaves in a similar manner to the full stop, causing a falling intonation pattern followed by a pause.

3.6 Parentheses

Parenthesis '()', brackets '[]', and braces '{}' appearing around a single word or a group of words cause a brief pause before and after the bracketed text.

4 Other non alphanumeric characters

4.1 Non-punctuation characters

The characters listed below are processed as non-letter, non-punctuation characters. Some are pronounced at all times (see the table below) and others are only pronounced in certain contexts, which are described in the following sections of this chapter.

Table: Non-punctuation symbols with fixed pronunciation

Symbol	Reading
/	слеш
\	обратный слеш
+	плюс
±	плюс минус
€	евро
™	торговая марка
§	параграф
©	знак авторского права
®	знак регистрации
‰	промилле
№	номер
<	меньше чем
>	больше чем
^	циркумфлекс
	вертикальный слеш
~	тильда
—	подчёркивание
@	собачка
=	равно
-	see below
*	see below

4.2 Symbols whose pronunciation varies depending on the context

4.2.1 Hyphen

A hyphen '-' is pronounced *минус* in two cases:

1. if followed by a digit and no other digit is found in front of the hyphen.
2. if followed by a digit and an equals sign '='.

In dates, between days or years, the hyphen is not pronounced.

In compound words, word-internally, or in words ending with numbers, the hyphen is not pronounced. Examples: *кто-нибудь*, *mother-in-law*, *ТУ-104*. In all other contexts, it is pronounced as *дефис*.

Expression	Reading
-3	минус три
44-3	сорок четыре дефис три
44-3=41	сорок четыре минус три равно сорок один
2-2-2002	второе февраля две тысячи второго года

4.2.2 Asterisk

Asterisk '*' is pronounced as *умножить на* if enclosed by digits and followed by the equals sign '='. In other cases it is pronounced as *звёздочка*.

Expression	Reading
2*3	два звёздочка три
2*3=6	два умножить на три равно шесть
*bc	звёздочка би си

4.2.3 Percent

[not SP] When preceded by a digit string (with or without a blank space in between), the percent sign '%' is pronounced in different grammatical cases/numbers (Genitive Singular or Plural) depending on the preceding digits. In all other contexts, it is pronounced as Nominative Singular *процент*.

Expression	Reading
21%	двадцать один процент
2%	два процента
24%	двадцать четыре процента
46%	сорок шесть процентов

4.2.4 Degree

[not SP] When immediately preceded by a digit string, the degree sign '°' is pronounced in different grammatical cases/numbers (Genitive Singular or Plural) depending on the preceding digits. In all other contexts, it is pronounced as Nominative Singular *градус*.

Expression	Reading
21°	двадцать один градус
-2°	минус два градуса
24°	двадцать четыре градуса
46°	сорок шесть градусов

4.2.5 Currency sign \$

The currency sign '\$' is expanded as *доллар* declined in different grammatical case/number (Genitive Singular or Plural) depending on the preceding digits. Their treatment is described in section *Currency amounts* dealing with currency processing.

5 *Number processing*

Strings of digits that are sent to the text-to-speech converter are processed in several different ways, depending on the format of the string of digits and the immediately surrounding punctuation or non-numeric characters. To familiarize the user with the various types of formatted and non-formatted strings of digits that are recognized by the system, a brief description of the basic number processing is provided, along with examples. The number processing is subdivided into the following categories which are described in the subsections below:

Full number pronunciation
Leading zero
Decimal numbers
Currency amounts
Abbreviations of quantity
Ordinal numbers
Arithmetic operators
Mixed digits and letters
Time of day
Dates
Phone numbers

5.1 *Full number pronunciation*

Full number pronunciation is given for the whole number part of the digit string.

Example

2425 full number

2 425 full number

24,25 24 is a full number, 25 is the decimal part

Numbers denoting thousands, millions and billions (numbers larger than 999) may be grouped using space. The comma or full stop will trigger the pronunciation of digits as a decimal number (see section *Decimal Numbers*). In order to achieve the right pronunciation the grouping must be done correctly.

The rules for grouping of numbers are the following:

- Numbers are grouped in groups of three starting at the end.
- The first group in a number may consist of one, two, or three digits.
- If a group, other than the first, does not contain exactly three digits, the sequence of digits is not interpreted as a full number.
- The highest number read is 9999999999 (eleven digits). Numbers higher than this are read as separate digits.

Number	Reading
2580	две тысячи пятьсот восемьдесят
2 580	"
25800	двадцать пять тысяч восемьсот
25 800	"
2580350	два миллиона восемьдесят тысяч триста пятьдесят
2 580 350	"
1000000000	один миллиард
2000000000	два миллиарда
2000	две тысячи
1234567890123	один два три четыре пять шесть семь восемь девять ноль один два три
25 456 789 012	двадцать пять миллиардов четыреста пятьдесят шесть миллионов семьсот восемьдесят девять тысяч двенадцать

N.B. Full numbers are spelled out in most common Nominative/Accusative grammatical cases. For the correct pronunciation of numbers in other grammatical cases, they should be spelled out literally, in full words. Unless stated differently, this also counts for the numbers in occurring in currencies, dates and metric units.

5.2 Leading zero

Numbers beginning with 0 (zero) are read out digit by digit.

Number	Reading
09253	ноль девять два пять три
020	ноль два ноль

5.3 Decimal numbers

Comma or full stop may be used when writing decimal numbers.

The full number part of the decimal number (the part before comma or full stop) is read according to the rules in the section *Full number pronunciation*. The decimals (the part after comma or full stop) are read as a full number without leading zero's with one, two or three numbers strings corresponding to *десятых* (tenth), *сотых* (hundredth) and *тысячных* (thousands). Decimal parts containing digit strings of longer than three characters are spelled out digit by digit.

Expression	Reading
16,234 and 16.234	шестнадцать целых двести тридцать четыре тысячных
3,1415 and 3.1415	три точка один четыре один пять
1251,04 and 1251.04	тысяча двести пятьдесят одна целая четыре сотых
1,251,04 and 1.251.04	тысяча двести пятьдесят одна целая четыре сотых
2,50 and 2.50	две целых пятьдесят сотых

5.4 Currency amounts

[not SP] The following principles are followed for currency amounts:

- Numbers with zero or two decimal places preceded or followed by either the currency markers \$ or € are read as currency amounts.
- Numbers with zero or two decimal places preceded or followed by the words *рубль*, *копейка* or *доллар* (in Singular or Plural) are read as currency amounts in Nominative case.
- Accepted decimal markers are comma ',' and full stop '.'.
- No spaces are allowed in the number.
- If the decimal part is 00 it will not be read.
- Single spaces between different words are required for correct pronunciation.

Expression	Reading
\$15,00	пятнадцать долларов
15,00 GBP	пятнадцать фунтов стерлинга
15.00 €	пятнадцать евро
15.00 евро	пятнадцать евро
15.30 руб.	пятнадцать рублей тридцать копеек

There is also the possibility of writing large amounts with the abbreviations *тыс.*, *млн.* or *млрд.* as follows:

\$ 5 млн.	пять миллионов долларов
-----------	-------------------------

Other currencies (such as e.g. *руб.* and *коп.*), which are spelled as abbreviations are also expanded similarly to the above rules.

The supported abbreviations of currencies include:

\$, €, *czk, gbp, usd, руб., коп., euro, евро, евр., дол., рубл(-ь, -я, -ей), копейк(-а, -и, -ек), доллар(-, -а, -ов), фунт(-, -а, -ов).*

5.5 Abbreviations of metric units

[not SP] When preceded by a digit string, some abbreviations of quantity have to be pronounced in different grammatical cases/numbers (Genitive Singular or Plural) depending on the preceding digits. The preceding number is pronounced in the Nominative case. In all other contexts, the abbreviations are expanded in Genitive plural case or into words in ambiguous cases.

Number	Reading
21 км	двадцать один километр
2 км	два километра
24 км	двадцать четыре километра
46 км	сорок шесть километров
21 т.	двадцать одна тонна
2 т.	две тонны
24 т.	двадцать четыре тонны
46 т.	сорок шесть тонн

The supported abbreviations of metric units include:

м, см, км, дм, мм, м2, м3, см2, см3, км2, км3, дм2, дм3, мм2, мм3, г, мг, кг, сг, °C, °F, км/ч, м/с, мл, мин, сек, час, ч, с., сл, дл, л, га, т, cal, kcal, кал, ккал, кВ.

The following exceptions / rules apply:

- There must be a blank space between the digit and the following abbreviation.
- The ambiguous abbreviation “z” is expanded into *год* in dates, or *грам* if preceded by digits in non-date formats. It can be expanded as *город* in front of some frequent city names. All other lower-case entries as well as and the upper-case character ‘Г’ are expanded as a letter. For correct rendition in other contexts, we advice to use full words.
- The ambiguous words *час* and *сек* are not treated as abbreviations if not preceded by digits.
- Single-letter abbreviations *м(.), ч(.), с(.), л(.), м(.)* are only pronounced as metric units when written in lower-case characters. The corresponding uppercase entries are pronounced as letters.
- Abbreviations ambiguous with frequent single-letter prepositions such as e.g. “c” must be followed by a full-stop.

- The abbreviations of time *ч(.)*, *с(.)* are only expanded in fixed time formats (see section *Decimal Numbers*).
- Other abbreviations of metric units will be expanded whether or not followed by a full stop, and they are case-insensitive.

5.6 Ordinal numbers

[not SP] Numbers are read as ordinals in the following cases:

- The number is followed by a month name or one of the month name abbreviations and the number is smaller or equal to 31. The number may be preceded by a day or an abbreviation for a day.

Example:

3 января, 3 мар, пн 3 янв

- The number consists of a day interval followed by a month name/abbreviation.

Example:

15-16 января

- The number is immediately followed by a hyphen followed by *й, я, ая, го, х, му*.

Example:

1-й, 1-я, 3-го, 23-му

The valid abbreviations for months are: *янв, февр, мар, апр, май, июн, июл, авг, сент, окт, ноя* and *дек*.

The valid abbreviations for days are: *пн, вт, ср, чт, пт, сб*, and *вс*.

The abbreviations above are only expanded to names of months and days when appearing in correct date contexts.

Expression	Reading
30 апреля 1999	тридцатое апреля тысяча девятьсот девяносто девятого года
апрель 30, 1999	тридцатое апреля тысяча девятьсот девяносто девятого года
май 1953 г.	май тысяча девятьсот пятьдесят третьего года
3 мая	третье мая
1999-го	тысяча девятьсот девяносто девятого

5.7 Arithmetic operators

Numbers together with arithmetical operators are read according to the examples below.

Expression	Reading
-12	минус двенадцать
+24	плюс двадцать четыре
2*3	два умножить на три
2*3=6	два умножить на три равно шесть
2/3	две третьих

5.8 Mixed digits and letters

If one or more upper-case letters appear within an alphanumeric sequence, the letters are read one by one. The strings of digits are split up into numbers according to the examples below.

Expression	Reading
77B184Z3	семьдесят семь би сто восемьдесят четыре зэд три
0092B87-B	ноль ноль девяносто два би восемьдесят семь би
FT2892B87Z	эф ти двадцать восемь девяносто два би восемьдесят семь зэд
TN12345L5	ти эн один два три четыре пять эл шесть
СПБ17ВА	эс пэ бэ семнадцать вэ а

5.9 Time of day

[not SP] The colon is used to separate hours, minutes and seconds. Abbreviations such as *час*, *мин* and *сек*, *ч.*, *м.* and *с.* may follow the time, with a space inserted between the time and the abbreviation. The use of the short abbreviations *ч.*, *м.* and *с.* is only permitted in the contexts where hours, minutes and seconds co-occur altogether (such as e.g. *1 ч 20 м 2 с*). The time abbreviations are expanded as *час*, *минута*, *секунда*, and are declined in different grammatical case/number (Genitive Singular or Plural) depending on the preceding digits. The numbers are expanded as whole numbers. Trailing zero's are not pronounced.

Possible patterns are:

1. *hh:mm* or *h:mm*
2. *hh:mm:ss* or *h:mm:ss*
3. *hh:mm:ss''* or *h:mm:ss''*

Example: 12:30'45''

h = hour, *m* = minute, *s* = second.

In pattern a:

If the *mm*-part is equal to *00*, this part will not be read. Instead, *ровно* will be added to the hours.

Examples:

9:00	девять часов ровно
13:00	тринадцать часов ровно

In pattern b:

If the seconds-part is equal to *00*, it will not be read out.

In pattern c:

Pattern (c) follows the rules for pattern (b).

Expression	Reading
9 час.	девять часов
9 час 20 мин	девять часов двадцать минут
2 час 3 мин	два часа три минуты
1 час. 20 мин. 2 сек.	один час двадцать минут две секунды
1 ч. 20 м. 2 с.	один час двадцать минут две секунды
9:20	девять часов двадцать минут
2:03	два часа три минуты
1:20'02"	один час двадцать минут две секунды
12:00	двенадцать часов ровно
0:00	ноль ровно
00:00	ноль ровно

5.10 Dates

The valid formats for dates are:

- *dd-mm-yyyy*, *dd.mm.yyyy*, and *dd/mm/yyyy*
- *dd-mm-yy*, *dd.mm.yy*, and *dd/mm/yy*

yyyy is a year written in four digits (e.g. 2007), *yy* is a year written in two digits (e.g. 07 for 2007), *mm* is a month number between 1 and 12 and *dd* a day number between 1 and 31. Hyphen, full stop, and slash may be used as delimiters. In all formats, one or two digits may be used in the *mm* and *dd* part. Zeros may be used in

front of numbers below 10. Months and days from 1 to 9 may consist of one digit or can include trailing zeros (e.g. *01* or *1* for *January*).

The numbers are expanded as ordinal numbers, either in Nominative or in Genitive grammatical case depending on the context.

Examples of valid formats:

10-02-2003, 10-2-2003, 10.02.2003, 10.2.2003, 10/02/2003, 10/2/2003, 10-02-03, 10-2-03, 10.02.03, 10.2.03, 10/02/03, 10/2/03

All these examples of dates are read out as *десятое февраля две тысячи третьего года*.

[not SP] Some ranges of days and years are also supported when used with prepositions *с, по, от, в, к*.

Expression	Reading
с 1998-1999 г.	с тысяча девятьсот девяносто восьмого по тысячу девятьсот девяносто девятый год
с 1939-45 г.	с тысяча девятьсот тридцать девятого по сорок пятый год
с 14-15 февраля	с четырнадцатого по пятнадцатое февраля
с 14-15 февраля 1999 г.	с четырнадцатого по пятнадцатое февраля тысяча девятьсот девяносто девятого года
с 14 февраля 1999 г.	с четырнадцатого февраля тысяча девятьсот девяносто девятого года
от 14 февраля 1999 г.	от четырнадцатого февраля тысяча девятьсот девяносто девятого года
с 14 по 15 февраля	с четырнадцатого по пятнадцатое февраля
с 14 по 15 февраля 1999 г.	с четырнадцатого по пятнадцатое февраля тысяча девятьсот девяносто девятого года
с 1999 г.	с тысячи девятьсот девяносто девятого года
в 1999 г.	в тысяче девятьсот девяносто девятом году
к 1999 г.	к тысяче девятьсот девяносто девятому году

[not SP] Other possible formats include:

Expression	Reading
30 апреля 1999	тридцатое апреля тысяча девятьсот девяносто девятого года
май 1953 г.	май тысяча девятьсот пятьдесят третьего года
3 мая	третье мая

5.11 Phone numbers

[not SP] In this section the patterns of digits that are recognized as phone numbers are described. All numbers are read out as whole numbers formed by the groups of digits separated by space, forward slash, period or hyphen, with a pause between the groups. Leading zeros are pronounced as zero (*ноль*). Groups of four digits and longer are pronounced digit by digit.

5.11.1 Ordinary phone numbers

Sequences of digits in the following formats are treated as phone numbers.

The following sequences of digits can be separated by a space or a hyphen:

Format

xxxxx xxxxxx

xxxxx xxx xxx

xxxxx xxxxx

xxxx xxxxxxx

xxxx xxx xxxx

xxxx xxxxxx

xxx xxxx

xx xx xx

xxx xxxx xxxx

(area code) xxxx xxxx

(area code) xxxxxxx

(area code) xxxxxx

(area code) xxxxx

(area code) xxx xxxx

(area code)-xxx-xxxx

The *area code* is a sequence of 0 followed by 2 to 7 digits.

The following sequences can only appear in these formats:

Format

xxx/xxx-xxxx

xxx/x-xx-xx

xxx/xxx-xxx

xxx/xx-xx-xx

xx/xxx-xx-xx

Format

xxx-xxx-xxx
(x)-xxx-xxx
(xx)-xxx-xxx
(xxx)-xxx-xxx
(x).xxxx.xxx.xxx
(x)-xxxx-xxx-xxx
(xx).xxxx.xxx.xxx
(xx)-xxxx-xxx-xxx
(xxx).xxxx.xxx.xxx
(xx) xxx-xx-xx
(xx) xxx.xx.xx
(xxxx) xx.xx.xx
(xxxx) xx-xx-xx
(xxxx) x.xx.xx
(xxxx) x-xx-xx
(xxx) xxx.xx.xx
(xxx) xxx-xx-xx

The sequence *xxx-xxx* is recognized as a phone format only if preceded by *tel*, *mob*, *tel*; *телефон*, *тел*; *тел*; *моб. т*; *моб. тел*..

Expression

Reading

(09) 345-46-71	ноль девять, триста сорок пять, сорок шесть, семьдесят один
тел. (09) 345-46-71	телефон, ноль девять, триста сорок пять, сорок шесть, семьдесят один
(093) 45-46-71	ноль девяносто три, сорок пять, сорок шесть, семьдесят один
093/45-46-71	ноль девяносто три, сорок пять, сорок шесть, семьдесят один
093/400-071	ноль девяносто три, четыреста, ноль семьдесят один
093/400071	ноль девяносто три, четыре, ноль, ноль, ноль, семь, один

5.11.2 International phone numbers

[not SP] All preceding formats can be recognized if preceded by international prefix and a space:

00x	+x	00(x)	+(x)
00xx	+xx	00(xx)	+(xx)
00xxx	+xxx	00(xxx)	+(xxx)

Expression

+ 32 (09) 345-46-71

тел. + 32 (09) 345-46-71

007 (09) 345-46-71

Reading

плюс тридцать два, ноль девять, триста сорок пять, сорок шесть, семьдесят один

телефон, плюс тридцать два, ноль девять, триста сорок пять, сорок шесть, семьдесят один

ноль ноль семь, ноль девять, триста сорок пять, сорок шесть

6 *How to change the pronunciation*

The user can affect the way that the words are pronounced by using the user lexicon (see *User's guide*). It can be done in order to specify an alternative pronunciation of a certain word or to correct an erroneous pronunciation.

The pronunciation can be specified in two ways in the user lexicon. One method is to modify the spelling of the word and another is to write a phonetic transcription of the word (see chapter *Russian phonetic text*). Phonetic transcription can also be entered directly in the text, using the PRN-tag (see *User's guide*).

7 Russian phonetic text

The Russian text-to-speech system from Acapela uses a subset of the SAM phonetic alphabet (*Speech Assessment Methods Phonetic Alphabet*) or SAMPA. Only SAMPA may be used in phonetic transcriptions. Symbols not listed in the tables below are not valid and will be ignored if included in the user lexicon or in a PRN tag. The SAMPA symbols are written with a space after each phoneme. The numbers / 1 / and / 2 / after the vowels indicate primary and secondary lexical stress (also called *word stress*, see section *Lexical stress*).

7.1 Consonants

The table below lists the phonetic symbols that can be used for Russian consonants.

Consonant symbols ending with the single quote (') are all palatalized. Consonant symbols, which do not end with the quote, are all non-palatalized (with the exception of 'j', which is inherently palatalized). Palatalized and non-palatalized consonants usually co-occur with a mutually exclusive set of stressed and unstressed vowels explained in *Table: Symbols for the Russian vowels*.

Table: Symbols for the Russian consonants

Symbol	Word	Phonetic transcription	Comment
p	пара	p A1 r @	
p'	пять	p' {1 t'	
b	бар	b A1 r	
b'	бег	b' e1 g	
t	так	t A1 k	
t'	тяга	t' {1 g @	
d	дать	d A1 t'	
d'	дядя	d' {1 d' \$=	
k	кот	k o1 t	
k'	кит	k' i1 t	
g	год	g o1 d	
g'	гид	g' i1 d	
m	мать	m A1 t'	
m'	мять	m' {1 t'	
n	нот	n o1 t	
n'	нет	n' e1 t	
l	лук	l u1 k	
l'	люк	l' }1 k	

Symbol	Word	Phonetic transcription	Comment
r	рад	r A1 d	
r'	ряд	r' {1 d	
f	фара	f A1 r @	
f'	финик	f' i1 n' \$ k	
v	ваза	v A1 z @	
v'	вяз	v' {1 z	
s	сад	s A1 d	
s'	сядь	s' {1 d'	
z	зад	z A1 d	
z'	зять	z' {1 t'	
Z	жар	Z A1 r	
S	шар	S A1 r	
S'	щит	S' i1 t	
Z'	товарищ был	t V v A1 r' \$ Z' b 11 l	can optionally be transcribed as S'
x	хор	x o1 r	
x'	хек	x' e1 k	
G	бог даст	b o1 G d A1 s t	can optionally be transcribed as x
ts	цирк	ts 11 r k	
dz	спецзавод	s p' e2 dz z V v o1 t	
tS'	чуть	tS' }1 t'	
dZ'	начдив	n A2 dZ' d' i1 f	
j	яма or йога	j {1 m @ or j 91 g @	the sound of й
dZ	джем	dZ E1 m	
tS	имидж	l1 m' i tS	can optionally be transcribed as dZ
ttS'	лётчик	l' 91 ttS' \$ k	the sound of “чч”
tts	учиться	U tS' i1 tts @	the sound of “цц”
tt	оттолкнуть	V tt V l k n u1 t'	
tt'	оттек	V tt' 91 k	
dd	оддал	V dd A1 l	
dd'	оддел	V dd' e1 l	
ss	рассадить	r @ ss V d' i1 t'	
ss'	рассеян	r V ss' e1 \$ n	
zz	раззадорить	r @ zz V d o1 r' \$ t'	
zz'	раззеваться	r @ zz' \$ v A1 tts @	
nn	ванная	v A1 nn @ \$=	
nn'	ванне	v A1 nn' \$	

Symbol	Word	Phonetic transcription	Comment
ll	вилла	v' i1 ll @	
ll'	вилле	v' i1 ll' \$	
rr	контрразведка	k o2 n t rr V z v' e t k @	
rr'	контрреволюция	k o2 n t rr' \$ v V l' }1 ts l \$=	
SS	расширить	r V SS 11 r' \$ t'	the sound of “шш”
ZZ	разжиреть	r @ ZZ l r' e1 t'	the sound of “жж”
SS'	расщипать	r @ SS' \$ p A1 t'	the sound of “щщ”
vv	ввод	vv o1 d	
vv'	ввел	vv' 91 l	
m m	лемма	l' e1 mm @	
m m'	лемме	l' e1 mm' \$	

7.2 Vowels

The table below lists the phonetic symbols, which can be used for Russian vowels. Note that if a vowel occurs only “under lexical stress” in comment must always be accompanied by primary or secondary stress marks (see section *Lexical stress*). Vowel symbols not occurring in stressed position cannot contain stress marks. Vowels said to occur after palatalized sounds, may not occur after non-palatalized sounds.

Table: Symbols for the Russian vowels

Symbol	Word	Phonetic transcription	Comment
A	альт	A1 l' t	under lexical stress
{	пять	p' {1 t'	stressed /A/ after palatalized sounds
V	завод	z V v o1 d	unstressed /A/ or /o/, in the 1st syllable before stress or word-initially
@	молоко	m @ l V k o1	unstressed /A/ or /o/ in all other syllables
u	угол	u1 g @ l	under lexical stress
}	люк	l' }1 k	under lexical stress
U	укол	U k o1 l	unstressed /u/
Y	новую	n o1 v U Y	unstressed /y/
o	он	o1 n	under lexical stress
9	нёс	n' 91 s	stressed /o/ after palatalized sounds

Symbol	Word	Phonetic transcription	Comment
E	этот	E1 t @ t	under lexical stress
e	есть	j e1 s' t'	stressed /E/ after palatalized sounds
l	данный	d A1 nn l j	unstressed /l/
l=	этаж	l= t A1 Z	unstressed /E/ before stressed syllables or word-finally
1	ыкать	11 k @ t'	under lexical stress
\$	теперь	t' \$ p' e1 r'	unstressed /i/ or /e/
\$=	Россия	r V ss' i1 \$=	unstressed /{/ before a pause
\$i	России	r V ss' i1 \$i	unstressed /i/ before a pause
i	идти	i tt' i1	under lexical stress or word initially

7.3 Lexical stress

Lexical stress indicates the level of prominence of a syllable in a word. In Russian, some words can be differentiated by the position of this lexical stress. For example, the word *замок* can either mean *a castle* or a *slot* depending on the position the lexical stress in the word (*a castle*: /z A1 m @ k / or *a keyhole*: /z V m o1 k /). Practically all words in Russian have lexical stress even if it does not always serve to differentiate between two different words. It is, therefore, important to include stress marks, when writing phonetic transcriptions.

In the phonetic transcriptions, primary stress is indicated by the symbol ' 1 ' placed directly after the stressed vowel without a space in between. Secondary stress is indicated by the symbol ' 2 '.

Some examples:

представитель	/p r' \$ ts t V v' i1 t' \$ l' /
спецпредставитель	/s p' e2 ts p r' \$ ts t V v' i1 t' \$ l' /

Note that the symbol ' 1 ' in Russian has two meanings: i.e. the SAMPA symbol for the vowel /ɨ/ (letter 'ы') and primary lexical stress. The symbol of the primary stress '1' always follows the vowel symbols without spaces in between, and its position differentiates its meaning from the SAMPA symbol 1. For example /11/ in /v 11 t' / reflects the sound of 'ы' under primary lexical accent '1'. A combination like /v 1 t' / would be illegal, cause vowel '1' cannot occur without the stress mark 1.

7.4 Glottal stop

A glottal stop, represented by the phonetic symbol /ʔ/, is a small optional sound produced with a brisk opening of the vocal folds at the onset of sounds.

In Russian, it usually occurs in front of vowels, in order to increase speech clarity. It is often used to separate two words, especially when the second one starts with a stressed vowel. This sound, thus, can optionally be inserted in a user transcription in order to make the pronunciation more clear. In Russian, glottal stop does not *replace* other sounds (e.g. consonants like t), but it can only be *added* in front of vowels, and can not occur in front of consonants.

For example,

a combination of words such as “город Омск” can be pronounced:

- Fluently without a glottal stop between the words
/g o1 r @ t o1 m s k /
- Clearly with a glottal stop in between
/g o1 r @ t ʔ o1 m s k /

In the former case, in very fluent speech the combination could even be confused with “город Томск”. In the latter case, the confusion is not possible due to the glottal stop.

7.5 Pause

An underscore /_ / in a phonetic transcription generates a small pause.

8 *Non-Cyrillic input*

Words written in Roman characters are pronounced according to the English pronunciation rules, but with Russian sounds (as if spoken by a Russian with a strong Russian accent).

9 Abbreviations

In the current version of the Russian text-to-speech system, the abbreviations in the table below are recognized in all contexts. These abbreviations are case-insensitive, and require no full stop in order to be recognized as an abbreviation. The abbreviations preceded by digits are not listed here, as they have already been handled in sections *Currency amounts* and *Abbreviations of metric units*.

Table 9.1. Abbreviations

Abbreviation	Reading
ссср	E2 s E2 s E2 s E1 r
эвм	E2 v E2 E1 m
снг	E2 s E2 n g E1
чп	tS' e2 p E1
др	d r U g' i1 \$
стр	s t r V n' i1 ts @
пр-т	p r V s p' e1 k t
пр	p r o1 tS' \$ \$
ст	s t A1 n ts l \$=
гл	g l V v A1
кв	k v V r t' i1 r @
фр	f r V n ts u1 s k' \$ j
тд	t A2 g d A1 l' \$ \$
табл	t V b l' i1 ts @
ул	u1 l' \$ ts @
ру	r u1
англ	V n g l' i1 j s k' \$ j
экз	l g z' \$ m p l' {1 r @ v
обл	o1 b l @ s' t'
ул	u1 l' \$ ts @
пл	p l o1 S' \$ t'
б-р	b U l' v A1 r
напр	n @ p r' \$ m' e1 r
мкд	m k A1 d

10 Web-addresses and email

Web-addresses and email-addresses are read as follows:

- *www* is read as “три дабл ю”.
- full stops '.' are read as *точка*, hyphens '-' as *дефис*, underscore '_' as *подчеркивание*, slash '/' as *слеш*.
- *us*, *uk*, *fr* and all the other abbreviations for countries (except for *ru*) are spelled out letter by letter.
- the @ sign is read as *собачка*.
- words/strings (including *org*, *com* and *edu*) are pronounced according to the normal rules of pronunciation in the system (Russian or English) and in accordance with the lexicon.

String

www.acapela-group.com

http://www.acapela-group.com

smirnov@yahoo.ru

ivan_smirnov@yahoo.ru

Reading

три дабл ю, точка, акапела, дефис груп, точка ком

h t t p, двоеточие, слеш слеш, три дабл ю, точка, акапела, дефис груп, точка ком

смирнов, собачка, yahoo, точка ру

иван, подчеркивание, смирнов, собачка, yahoo, точка ру