Language Manual

Russian

Alyona

1 Swedish Phonetic Text
Russian
Alyona
13/12/2007

This document was produced by Acapela Group. We welcome and consider all comments and suggestions. Please send them to:

Acapela Group
Box 1328
SE-171 26 Solna
Sweden

Phone +46 (0) 8 799 86 00
Fax + 46 (0) 8 799 86 01

Acapela Group
33, Boulevard Dolez
7000 Mons
Belgium

Tel: +32 (0)65 37 42 75
Fax: +32 (0)65 37 42 76

Acapela Group
Immeuble Les Erables
Rue du Colombier
ZAC de l'Hers
F-31670 Labege
France

Tel: +33 (0)5 62 24 71 00
Fax: +33 (0)5 62 24 71 01

www.acapela-group.com

List of contents

1 General

This document discribes some important aspects of linguistic processing supported by the Russian text-to-speech system. In particular, it describes the different types of characters and formats that are allowed in the input text. The description is based on the characters supported by the ANSI standard codepage 1251 for Cyrillic alphabets.

This version of the document corresponds to the High Quality Russian voice “Alyona”.

2 Letters in orthographic text

Russian Cyrillic and Latin characters from А-Я, а-я, A-Z and a-z may form a word. Certain other characters are also treated as letters, notably those used as letters in other Slavic languages using Cyrillic codepage 1251 “s, j, ћ, ђ, ц, ґ, ґ, ќ, љ, њ, ў, і, ї, є”, but they are accordingly mapped onto close in pronunciation Russian sounds “с, й, ч, щ, ж, г, г, к, л, н, у, и, и, 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 marks 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 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

Quotes < **“ ”** > and < « » > appearing around a single word or a group of words cause a brief pause before and after the quoted text.

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 5) and in abbreviations (see chapter 8).

3.4 Question mark

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

3.5 Exclamation mark

Exclamation mark <!> behaves in a similar manner to the full stop, causing a falling intonation pattern followed by a pause.

3.6 Parentheses

Parentheses < **( )** > 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 differently to alphanumerical characters and punctuation. Some are pronounced at all times in the same way (see Table 1) and others are only pronounced in certain contexts, which are described in the following subsections of this chapter.

Symbol	Reading
/	слеш
\	обратный слеш
{	левая фигурная скобка
}	правая фигурная скобка
[	левая квадратная скобка
]	правая квадратная скобка
‘	апостроф
`	апостроф
+	плюс
±	плюс минус
€	евро
™	торговая марка
§	параграф
©	знак авторского права
®	знак регистрации
‰	промилле
№	номер
<	меньше чем
>	больше чем
^	циркумфлекс
|	вертикальный слеш
~	тильда
_	подчёркивание (see also chapter 9)
@	собачка
=	равно
-	see below
*	see below

Table 1 Non-punctuation symbols with fixed pronunciation

4.2 Symbols with the pronunciation variable depending on the context

4.2.1 Hyphen

A hyphen < **-** > is pronounced as “минус” in two cases:

a) if followed by a digit and no other digit is found in front of the hyphen

b) if followed by a digit and an equals sign.

In dates, inbetween days or years, the hyphen is not pronounced.
In compounds or between words, the hyphen is not pronounced. Examples: low-income, mother-in-law, pre- and post-war Europe. 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 only pronounced as “умножить на” if enclosed by digits and followed by the equals sign. In all other contexts, 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

4.2.2 Percent

When preceded by a digit string (with or without a blank space inbetween), the percent sign < **%** > is pronounced in different grammatical cases/numbers (Genitive Singular or Plural) depending on the preceeding 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

4.2.2 Degree

When immediately preceded by a digit string, the degree sign < **°** > is pronounced in different grammatical cases/numbers (Genitive Singular or Plural) depending on the preceeding 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

4.2.2 Currency sign $

The currency sign < $ > is expanded as “доллар” declined in different grammatical case/number (Genitive Singular or Plural) depending on the preceeding digits. Their treatment is described in section 5.4 dealing with currency processing.

5 Number processing

5.1 Introduction

Strings of digits that are sent to the text-to-speech converter are processed in 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 5.3). 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 99999999999 (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
123456789012	один два три четыре пять шесть семь восемь девять ноль один два
25 456 789 012	двадцать пять миллиардов четыреста пятьдесят шесть миллионов семьсот восемьдесят тысяч двенадцать

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

5.2 Leading zero

Numbers that begin with 0 (zero) are read as a whole number, with a zero preceding it.

Number	**Reading**
09253	ноль девять тысяч двести пятьдесят три
020	ноль двадцать

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) is read according to the rules in 5.1. 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). Note: A number containing a comma followed by exactly three digits is not read as a decimal number but as a full number, following the rules in 5.1. Decimal parts containing digit strings of longer than three characters are spelled out digit by digit.

Number	**Reading**
16,234	шестнадцать целых двести тридцать четыре тысячных
3,1415	три точка один четыре один пять
1251,04	тысяча двести пятьдесят одна целая четыре сотых
1,251,04	тысяча двести пятьдесят одна целая четыре сотых
2,50	две целых пятьдесят сотых

5.4 Monetary amounts

The following principles are followed for monetary amounts:

- Numbers with zero or two decimal places preceded or followed by the currency markers **$** or **€** are read as monetary amounts.
- Numbers with zero or two decimal places preceded or followed by the words "евро", "рубль", "копейка" or "доллар" (in Singular or Plural) are read as monetary 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 word are required for correct pronunciation.

Example	**Reading**
$15,00.	пятнадцать долларов
15,00£.	пятнадцать фунтов стерлинга
15.00 евро.	пятнадцать евро
15.30 рублей	пятнадцать рублей тридцать копеек

There is also the possibility of writing large amounts as follows:

$ 5 млн.	пять миллионов долларов

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

The supported abbreviations of currencies/abbreviations of currenties include: $, €, czk, gbp, usd, руб., коп., euro, евро, дол., рубл(-ь, -я, -ей), копейк(-а, -и, -ек), доллар(- ,-а, -ов), фунт(- ,-а, -ов).

5.5 Abbreviations of metric units

When preceded by a digit string (with a blank space inbetween), abbreviations of quantity are pronounced in different grammatical cases/numbers (Genitive Singular or Plural) depending on the preceeding digits. The preceding number is pronounced in the Nomnative case. In all other contexts, the abbreviations are expanded in Genitive plural case or into words in ambiguous cases.

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

The following exceptions / rules apply:

- The abbrevation “г” is expanded into “год” in dates , or “грам” if preceded by digits in non-dates, or “город” in other contexts;
- The abbreviations “час” and “сек” are not expanded if not preceded by digits;
- The abbreviations will be expaneded whether or not followed by a full stop
- Case-sensitive abbreviations in the above list are those containing uppercase characters. Other abbreviations are case insensitive.

5.6 Ordinal numbers

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. For example: 3 января, 3 мар, пн 3 янв
- The number consists of a day interval followed by a month name/abbreviation. For example: 15-16 января
- The number is immediately followed by a hyphen followed by “й, ая, е, го, х, му”. For example: 1-й, 1-е, 3-го, 23-му.

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

Valid abbreviations for months: янв, фев, мар, апр, май, июн, июл, авг, сен, окт, ноя and дек.

Valid abbreviations for days: пон, 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.

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 a letter appears within a sequence of digits, the groups of digits will be read as whole numbers according to the rules above. The letter marks the boundary between the numbers. The letters will be spelled out.

Examples:

Expression	**Reading**
77B84Z3	семьдесят семь би восемьдесят четыре зэд три
0092B87-B	ноль ноль девяносто два би восемьдесят семь би

5.9 Time of day

The colon is used to separate hours, minutes and seconds. Abbreviations such as “час” , “мин” and “сек” (along with the possible variant “ч”) may follow the time, with a space inserted between the time and the abbreviation. The time abbreviations are expanded as “час”, “минута”, “секунда” are declined in different grammatical case/number (Genitive Singular or Plural) depending on the preceeding digits. The numbers are expanded as whole numbers. Trailing zero’s are not pronounced.

Other possible patterns are:

a) **hh:mm (or h:mm)**

b) **hh:mm:ss (or h:mm:ss)**

c) **hh:mm’ss” (or h:mm’ss”)** eg. 12:30’45”

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

Expression	**Reading**
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
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 часов ровно
00:00	ноль часов ровно

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.

Example: 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.

Pattern (c) follows the same rules as in the pattern (b).

5.10 Dates

The valid formats for dates are:

Type 1: dd-mm-yyyy, dd.mm.yyyy, and dd/mm/yyyy
Type 2: 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”.

Ranges of days and years are also supported when used with preposition “с”, “по”, “от”.

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.

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

In this section the patterns of digits that are recognized as phone numbers are described. In the pronunciation of phone numbers, 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:

- xxxxx xxxxxx
- xxxxx xxx xxx
- xxxxx xxxxx
- xxxx xxxxxxx
- xxxx xxx xxxx
- xxxx xxxxxx
- xxx xxxx
- xx xx xx
- xxx xxxx xxxx

- (area) xxxx xxxx
- (area) xxxxxxx
- (area) xxxxxx
- (area) xxxxx
- (area) xxx xxxx
- (area)-xxx-xxxx

(The area code is equal to 0 followed by 2 to 7 digits)

The following sequences can only appear in these formats:

- xxxx/xxx-xxxx
- xxxx/x-xx-xx
- xxx/xxx-xxx
- xxx/xx-xx-xx
- xx/xxx-xx-xx
- 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

All preceding formats can be recognised if preceded by international prefix:

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

Expression	**Reading**
+ 32 (09) 345-46-71	плюс тридцать два, ноль девять, триста сорок пять, сорок шесть, семьдесят один.
тел. + 32 (09) 345-46-71	телефон, плюс тридцать два, ноль девять, триста сорок пять, сорок шесть, семьдесят один.
007 (09) 345-46-71	ноль ноль семь, ноль девять, триста сорок пять, сорок шесть.

6 How to change pronunciation errors

Words that are not pronounced correctly by the text-to-speech converter can be entered in the user lexicon (see User's guide). In this lexicon, the user enters a phonetic transcription of the word (see chapter 7). Phonetic transcriptions can also be entered directly in the text, using a 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. A single quote "'" after the consonants indicates palatalisation. The numbers "1" and "2" after the vowels indicate primary and secondary lexical accent (sometimes also called "word-stress", see section 7.3).

7.1 Consonants

7.1.1 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	
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 1 l	can optionally be transcribed as S'
x	хор	x o1 r	
x'	хек	x' e1 k	
ts	цирк	ts 11 r k	
dz	спец-завод	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 @ | j 91 g @	the sound of й

dZ	джем	dZ E1 m	
tS	имидж	i m' i1 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 j $ 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' $	
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 I $	
SS	расширить	r V SS 11 r' $ t'	the sound of "шш"
ZZ	разжиреть	r V ZZ I r' e1 t'	the sound of "жж"
SS'	расщипать	r @ SS' $ p A1 t'	the sound of "сщ"
vv	ввод	vv o1 d	
vv'	ввел	vv' 91 l	
mm	лемма	l' e1 mm @	
mm'	лемме	l' e1 mm' $	

Table 2 Russian consonants

7.2 Vowels

7.2.1 Symbols for the Russian vowels

Symbol	Word	Phonetic transcription	Comment
A	альт	A1 l' t	under lexical accent
{	пять	p' {1 t'	a palatalised variant of /A/
V	завод	z V v o1 d	a reduced variant of /A/ and /o/ in the 1st prestressed syllable and word-initially
@	молоко	m @ l V k o1	a reduced variant of /A/ and /o/ in all other syllables
u	угол	u1 g @ l	under lexical accent
}	люк	l' }1 k	under lexical accent
U	укол	U k o1 l	unstressed variant of /u/
Y	новую	n o1 v U Y	unstressed variant of /}/
o	он	o1 n	under word-stress
9	нёс	n' 91 s	a palatalised variant of /o/ under lexical accent
E	этот	E1 t @ t	under lexical accent
e	есть	j e1 s' t'	a palatalised variant of /E/ lexical accent
I	этаж	I t A1 Z	unstressed variant of /1/ and /E/
1	ыкать	11 k @ t'	under lexical accent *
$	теперь	t' $ p' e1 r'	unstressed variant of /i/ and /e/
i	иди	$ d' i1	under lexical accent

Table 3 Russian vowels

7.3 Lexical accent

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

In the phonetic transcriptions, primary accent is indicated by the symbol "1" placed directly after (no space) the accented vowel. Secondary accent is indicated by the symbol "2". Some examples:

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

Note that the symbol "1" in Russian has two meanings: i.e. the SAMPA symbol for the vowel /1/ (letter "ы") and primary lexical accent. The symbol of the the primary lexical accent "1" always follows the vowel symbols without spaces in the second position and can be identified this way and be differentiated from the SAMPA sound [1]. For example /11/ in /v 11 t' / reflects the sound of "ы" under primary lexical accent "1".

7.5 Glottal stops

A glottal stop, represented by the phonetic symbol /? /, is a small optional sound produced at the glottis. It is often used to separate two words, when the second word starts with a stressed vowel. This sound can be inserted in a transcription in order to make the pronunciation more clear.
For example, the utterance "ну а он" can be pronounced fluently with no glottal stop
/n U V o1 n/, or more clearly with one or two glottal stops /n U ? V ? o1 n /.

7.6 Pause

An underscore < _ > in a phonetic transcription generates a small pause.

8 Non-cyrillic input

Words written Latin character set are pronounced according to the American English pronunciation rules, but with the 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 table 4 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 4.2.2. , 5.4 and 5.5.

Abbreviation	Reading
янв	j $ n v V r' {1
февр	f' $ v r V l' {1
мар	m A1 r t @
апр	V p r' e1 l' $
июн	$ j }1 n' $
июл	$ j }1 l' $
авг	A1 v g U s t @
сент	s' $ n t' $ b r {1
окт	V k t' $ b r {1
ноя	n @ $ b r' {1
дек	d' $ k V b r' {1
пн	p @ n' $ d' e1 l' n' $ k
вт	f t o1 t n' $ k
ср	s r' $ d A1
чт	tS' $ t v' e1 r g
пт	p' {1 t n' $ ts @
сб	s U b o1 t @
вс	v @ s k r' $ s' e1 n' j $
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 @
ru	r u1
англ	V n g l' i1 j s k' $ j
экз	I 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
мкад	m k A1 d
www	v E2 v E2 v E1
http	E1 j tS' t' i1 t' i1 p' i1
gmail	d Z 11 m E1 j l

hotmail	x V t m E1 j l
runet	r U n E1 t
com	k o1 m
ru	r u1
uk	j }1 k' e1 j
org	o1 r g
html	E1 j tS' t E1 ? E1 m ? E1 l
http	E1 j tS' t' i1 t' i1 p' i1

Table 4 Other abbreviations

9 Web-addresses and email

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

- “www” is read as three w’s spelled letter by letter.
- 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 “@” 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 American English) and in accordance with the lexicon.

String	**Reading**
www.acapela-group.com	w w w точка акапела дефис груп точка ком
http://www.acapela-group.com	h t t p двоеточие слеш w w w точка акапела дефис груп точка ком
smirnov@yahoo.ru	смирнов собачка yahoo точка ру
ivan_smirnov@yahoo.ru	иван подчеркивание смирнов собачка yahoo точка ру