



# *Language Manual*

*HQ and CO Polish*

---

Language Manual: HQ and CO Polish

Published 5 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.....	6
6	How to change pronunciation .....	17
7	Polish phonetic text .....	18
8	Abbreviations.....	22
9	Web-addresses and email .....	25

## 1 General

This manual describes various aspects relevant to the processing of the written Polish language that can help users to achieve the desired pronunciation by the Polish Text-to-Speech system of Acapela. In particular, the manual sums up the different types of characters: such as punctuation, signs, numbers and textual formats, which are allowed as input, and read out in specific ways.

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

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***

Characters from A-Ź and a-ź may constitute a word. Certain other characters are also considered as letters, notably those used as letters in other European languages, i.e. é, ç. These letters are pronounced as regular Polish letters close to their native language pronunciation.

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 string: , ; “ ” . ? ! ( ) { } [ ]

### **3.1 Comma, colon and semicolon**

Comma ',', colon ':' and semicolon ';' cause a brief pause in a sentence. The pause is accompanied by a small rising intonation pattern just prior to the punctuation character.

### **3.2 Quotation marks**

The quotes ' " ' and ' " ' appearing around a single word or a group of words cause a brief pause after the quoted text.

### **3.3 Full stop**

A full stop '.' is a sentence-final punctuation mark that causes a falling end-of-sentence intonation pattern. It 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 question-intonation, first rising and then falling.

### **3.5 Exclamation mark**

The exclamation mark '!' is treated in a manner similar to that of full stop, causing a falling intonation pattern followed by a pause.

### **3.6 Parentheses, brackets and braces**

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 and others are only pronounced in certain contexts, which are described in the following sections of this chapter.

Table: Non-punctuation characters

Symbol	Reading
/	ukośnik
+	plus
\	ukośnik wsteczny
\$	dolar
€	euro
<	mniejszy niż
>	większy niż
%	procent
‰	promil
^	daszek
	pionowa linia
~	tylda
—	podkreślnik
¨	umlaut
&	ampersand
§	akapit
@	małpa
÷	dzielone przez
×	mnożone przez
±	plus minus
=	pówna się
®	ochrona znaku towarowego
©	znak praw autorskich
™	znak towarowy
-	(see below)
*	(see below)

## 4.2 Symbols whose pronunciation varies depending on the context

---

### 4.2.1 Hyphen

A hyphen '-' is pronounced minus 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. If there is no equals sign '=', it is pronounced *łącznik*.

In certain date formats, in between days or years, the hyphen is pronounced "do". In other cases the hyphen is never pronounced.

Number	Reading	
-3	minus trzy	
44-3	czterdzieści cztery łącznik trzy	
44-3=41	czterdzieści cztery minus trzy równa się czterdzieści jeden	
1998-2004	tysiąc dziewięćset dziewięćdziesiąty ósmy do dwa tysiące czwarty	[not SP]
7-8.02.2007	siódmy do ósmego lutego dwa tysiące siedem	[not SP]
2002-2-2	drugi lutego dwa tysiące dwa	[not SP]
loga-sowiński	loga sowiński	
społeczno-politycznych	społeczno politycznych	

### 4.2.2 Asterisk

Asterisk '\*' is only pronounced as *mnożone przez* if enclosed by digits and followed by the equals sign. In other cases it is pronounced as *gwiazdka*.

Number	Reading
2*3	dwa gwiazdka trzy
2*3=6	dwa mnożone przez trzy równa się sześć
*bc	gwiazdka be ce



## 5 *Number processing*

Strings of digits are processed by the text-to-speech converter in different ways, depending on the format of the string of digits and its immediate surrounding context. To familiarize the user with the various types of formatted and non-formatted strings of digits recognized by the system, this section provides a brief description of the basic number processing with examples. Number processing is subdivided into the following categories:

Full number pronunciation  
Leading zero  
Decimal numbers  
Currency amounts  
Ordinal numbers  
Arithmetic operators  
Mixed digits and letters  
Time of day  
Year  
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
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 or full stop (not comma). 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 from 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 999999999999 (twelve digits). Numbers higher than this are read as separate digits.

<b>Number</b>	<b>Reading</b>
2580	dwa tysiące pięćset osiemdziesiąt
2 580	dwa tysiące pięćset osiemdziesiąt
2.580	dwa tysiące pięćset osiemdziesiąt
25800	dwadzieścia pięć tysięcy osiemset
25 800	dwadzieścia pięć tysięcy osiemset

Number	Reading
25.800	dwadzieścia pięć tysięcy osiemset
2580350	dwa miliony pięćset osiemdziesiąt tysięcy trzysta pięćdziesiąt
2 580 350	dwa miliony pięćset osiemdziesiąt tysięcy trzysta pięćdziesiąt
2.580.350	dwa miliony pięćset osiemdziesiąt tysięcy trzysta pięćdziesiąt
1000000000	jeden miliard
234567890123	dwieście trzydzieści cztery miliardy pięćset sześćdziesiąt siedem milionów osiemset dziewięćdziesiąt tysięcy sto dwadzieścia trzy
1234567890123	jeden dwa trzy cztery pięć sześć siedem osiem dziewięć zero jeden dwa trzy

The abbreviations such as “tys.”, “mln.” and “mld.” (accordingly “thousands”, “millions” and “billions”) are expanded into full numbers in agreement with the preceding number.

Number	Reading	
23 tys.	dwadzieścia trzy tysiące	[not SP]
1 tys.	jeden tysiąc	[not SP]
22 mld.	dwadzieścia dwa miliardów	[not SP]

## 5.2 Leading zero

---

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

Number	Reading
09253	zero dziewięć dwa pięć trzy
020	zero dwa zero

## 5.3 Decimal numbers

---

Comma may be used to write decimal numbers.

The full number part of the decimal number (the part before comma) is read according to the rules in the section *Full number pronunciation*. The decimals (the part after comma) are read as separate digits if there are more than 3 digits after the comma, otherwise, it is read out as a full number.

Number	Reading
3,1411	trzy przecinek jeden cztery jeden jeden
1251,04	tysiąc dwieście pięćdziesiąt jeden przecinek zero cztery

Number	Reading
1.251,04	tysiąc dwieście pięćdziesiąt jeden przecinek zero cztery
2,50	dwa przecinek pięćdziesiąt
3,141	trzy przecinek sto czterdzieści jeden

## 5.4 Currency amounts

---

The following principles are followed for currency amounts:

- Numbers with zero up to two decimal places preceded or followed by the currency markers \$, €, PLN, eur, zł., zl, dol., USD, GBP, CZK, DKK, NOK, SEK are read as currency amounts.
- Numbers with zero or two decimal places preceded or followed by the words *złoty, euro, dolar, funt* (singular or plural) are read as currency amounts.
- Accepted decimal markers are comma ',' and full stop '.'.
- The decimal part (consisting of two digits) in monetary amounts is read as *i xx centów, i xx groszy, i xx hależy, i xx pensów* or *i xx ore*.
- If the decimal part is 00 it will not be read.

Example	Reading	
\$15,00	piętnaście dolarów	
15,00 \$	piętnaście dolarów	
15,00 euros	piętnaście euro	[not SP]
€ 200,50	dwieście euro i pięćdziesiąt centów	[not SP]
1.000.000 GBP	jeden milion funtów	[not SP]
15,50 zł.	piętnaście złotych i pięćdziesiąt groszy	

There is also a possibility to write large amounts as follows:

\$ 23 tys.	dwadzieścia trzy tysiące dolarów	[not SP]
\$ 10 mln.	dziesięć milionów dolarów	[not SP]
\$ 1 miliard	jeden miliard dolar	[not SP]
\$ 1 mrd.	jeden miliard dolar	[not SP]

## 5.5 Ordinal numbers

---

Numbers are read as ordinals in the following cases:

- A full stop is attached to the number, followed by a space and another digit, a lower case letter or a punctuation mark.
- 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.
- The number represents a day in a date format.
- The number is in a range of days.

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

Example	Reading	
3. 1	trzeci jeden	[not SP]
3. a	trzeci a	[not SP]
3. !	trzeci	[not SP]
12 sierpnia 2003	dwunastego sierpnia dwa tysiące trzy	[not SP]
sobota, 12 sierpnia 2003	sobota dwunastego sierpnia dwa tysiące trzy	[not SP]
15.3.2007	piętnasty marca dwa tysiące siedem	[not SP]
7-8.02.2007	siódmy do ósmego lutego dwa tysiące siedem	[not SP]

Compound adjectives with a number in the first part are also expanded:

Example	Reading	
10-cio milionowy	dziesięcio-milionowy	[not SP]

## 5.6 Arithmetic operators

---

Numbers together with arithmetical operators are read as in the examples below:

Expression	Reading	
-12	minus dwanaście	
44-3	czterdzieści cztery łącznik trzy	
44-3=41	czterdzieści cztery minus trzy równa się czterdzieści jeden	
+24	plus dwadzieścia cztery	
2+3	dwa plus trzy	
2+3=5	dwa plus trzy równa się pięć	
2*3	dwa gwiazdka trzy	
2*3=6	dwa mnożone przez trzy równa się sześć	
2/3	dwietrzecie	
6/2=3	sześć dzielone przez trzy równa się dwa	
25%	dwadzieścia pięć procent	
3,4%	trzy przecinek cztery procenta	[not SP]
3.4%	trzy kropka cztery procenta	[not SP]

## 5.7 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	77 B 184 Z 3
0092B87-B	zero zero 92 B 87 B
FT2892B87Z	F T 28 92 B 87 Z
TN12345L5	T N 1 2 3 4 5 L 6
E17	E 17

## 5.8 Time of day

---

The colon and full stop can be used to separate hours, minutes and seconds. The word *godzina*, *godz.* or *g.* can precede the time, but will be pronounced in any case, whether it is there or not.

Possible patterns are:

- |                          |    |                       |
|--------------------------|----|-----------------------|
| a) hh:mm (or h:mm)       | or | hh.mm (or h.mm)       |
| b) hh:mm:ss (or h:mm:ss) | or | hh:mm.ss (or h:mm.ss) |
| c) hh h mm min           |    |                       |
| d) hh godz. mm min       |    |                       |
| e) hh h mm               |    |                       |

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

### In patterns a, b and e:

These formats are recognized for expressing the time of the day. The word *godzina* will be inserted before the *hh*-part, the word *i* will be inserted before the *ss*-part, and *sekunda*, *sekundy* or *sekund* will be inserted at the end.

Format 'a' (hh.mm) will only be recognized if it is preceded by the words *godzina*, *godziny* or *godzin*, or one of their abbreviations. Otherwise this format will be recognized as a decimal number.

### In patterns c and d:

These formats are recognized for expressing duration. The words *godzina i*, *godziny i* or *godzin i* will be inserted between the *hh*-part and *mm*-part, and *minuta*, *minuty* or *minut* will be inserted at the end.

Expression	Reading
godz. 9.30	godzina dziewięta trzydzieści [not SP]
9:00	godzina dziewięta zero zero

Expression	Reading	
g. 9:00	godzina dziewiąta zero zero	[not SP]
9:30.20	godzina dziewiąta trzydzieści i dwadzieścia sekund	[not SP]
1 godz. 45 min	jedna godzina i czterdzieści pięć minut	[not SP]
12 h 30 min	dwanaście godzin i trzydzieści minut	[not SP]
09 h 30	godzina dziewiąta trzydzieści	[not SP]

Time intervals will also be recognized correctly, as well as time formats in prepositional phrases.

Expression	Reading	
12:00-14:00	godzina dwunasta zero zero do czternastej zero zero	[not SP]
od godz. 12:00 do 14:00	od godziny dwunastej zero zero do godziny czternastej zero zero	[not SP]
po godz. 12.00	po godzinie dwunastej zero zero	[not SP]
przed godz. 12.00	przed godziną dwunastą zero zero	[not SP]
w godz.12.00-14.00	w godzinach od dwunastej zero zero do czternastej zero zero	[not SP]

## 5.9 Dates

---

The valid formats for dates are:

1. *dd.mm.yyyy* and *dd/mm/yyyy*
2. *dd.mm.yy* and *dd/mm/yy*
3. *yyyy.mm.dd* and *yyyy-mm-dd*
4. *dd MM yyyy* and *dd.MM.yyyy*
5. *dd MM yy* and *dd.MM.yy*

*yyyy* is a four-digit number, *yy* is a two-digit number, *mm* is a month number between 1 and 12, *MM* is a month number between 1 and 12 in roman numerals, and *dd* is a day number between 1 and 31. Full stop and slash (and space in the last two formats) may be used as delimiters. In all formats, one or two digits may be used in the *mm* and *dd* part. Zero's may be used in front of numbers below 10.

Below, some examples of valid formats and their readings:

Type 1:	Reading
10-02-2003 or 10-2-2003	dziesiąty lutego dwa tysiące trzy
10.02.2003 or 10.2.2003	dziesiąty lutego dwa tysiące trzy
10/02/2003 or 10/2/2003	dziesiąty lutego dwa tysiące trzy
Type 2:	Reading
10-02-03 or 10-2-03	dziesiąty lutego dwa tysiące trzy
10.02.03 or 10.2.03	dziesiąty lutego dwa tysiące trzy
10/02/03 or 10/2/03	dziesiąty lutego dwa tysiące trzy
Type 3:	Reading
2003.02.10 or 2003.2.10	dziesiąty lutego dwa tysiące trzy [not SP]
2003-02-10 or 2003-2-10	dziesiąty lutego dwa tysiące trzy [not SP]
Type 4:	Reading
10 II 2003	dziesiąty lutego dwa tysiące trzy [not SP]
10.II.2003	dziesiąty lutego dwa tysiące trzy [not SP]
Type 5:	Reading
10 II 03	dziesiąty lutego dwa tysiące trzy [not SP]
10.II.03	dziesiąty lutego dwa tysiące trzy [not SP]

[not SP] Ranges of days and years are also supported.

Expression	Reading
1998-1999	tysiąc dziewięćset dziewięćdziesiąty ósmy do tysiąc dziewięćset dziewięćdziesiąty dziewiąty
1939-45	tysiąc dziewięćset trzydziesty dziewiąty do czterdziesty piąty

Expression	Reading
7–8.02.2007	siódmy do ósmego lutego dwa tysiące siedem
od 5 do 13 maja	od piątego do trzynastego maja

[not SP] Other possible date formats include:

Expression	Reading
rok 2003	rok dwa tysiące trzeci
od 1984 roku	od tysiąc dziewięćset osiemdziesiątego czwartego roku
do 1984 roku	do tysiąc dziewięćset osiemdziesiątego czwartego roku
po stycz. 2003	po styczniu dwa tysiące trzy
przed sob., 12 stycz. 2003	przed sobotą dwunastego stycznia dwa tysiące trzy
sobota, 12 lutego 2003	sobota dwunastego lutego dwa tysiące trzy
w sob., 12 lutego 2003	w sobotę dwunastego lutego dwa tysiące trzy
12 lutego 2003	dwunastego lutego dwa tysiące trzy
12 lutego 2003 r.	dwunastego lutego dwa tysiące trzeciego roku

For months and days, the following abbreviations can be used in the above formats:  
[not SP]

- Months: *stycz., mar., marz., kwie., czerw., czer., lip., sierp., sier., wrze., paźdz., listop., list., grudz., and grud.*
- Days of the week: *pn., pon., wt., śr., cz., czw., pt., so., sob., nd.* and *niedz.*



## 5.10 Phone numbers

---

[not SP] Some digit patterns are recognized as phone numbers, and pronounced accordingly. If a pattern is recognized as a phone number, the strings longer than three digits are read out as groups of two or three full cardinal numbers with pauses in between. Leading zero's are pronounced separately digit by digit. Valid phone formats containing sequences of digits longer than 4 characters (without spaces or hyphens) are read out digit by digit.

### 5.10.1 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	Example	Reading
x xxxx xxxx	2 3456 4567	2 34 56 45 67
x-xxxx-xxxx	2-3456-4567	"
x xx xx xx xx	5 45 67 67 67	5 45 67 67 67
x-xx-xx-xx-xx	5-45-67-67-67	"
xx xxx xxxx	22 394 5867	22 394 58 67
xx-xxx-xxxx	22394-5867	"
xx xxxx xxx	22 3406 888	22 34 0 6 888
xx-xxxx-xxx	22-3406-888	"
xxx xx xx xx	532 45 46 71	532 45 46 71
xxx-xx-xx-xx	532-45-46-71	"
xx xxx xx xx	12 326 71 01	12 326 71 0 1
xx-xxx-xx-xx	12-326-71-01	"
xxx xxx xx xx	728 651 12 00	728 651 12 0 0
xxx-xxx-xx-xx	728-651-12-00	"
xx xx xxx xxx	31 77 546 666	31 77 546 666
xx-xx-xxx-xxx	31-77-546-666	"
(xx) xxx xx xx	(22) 494 00 34	22 494 0 0 34
(xx) xxx-xx-xx	(22) 494-00-34	"
(xxx) xxx xx xx	(788) 516 41 41	788 516 41 41
(xxx) xxx-xx-xx	(788) 516-41-41	"
(xxx) xxxxxx	(315) 5253707	315 5 2 5 3 7 0 7
(xxx)xxxxxxx	(315) 5253707	"
0-xx xxx xx xx	0-22 494 00 34	0 22 494 0 0 34
0xxx xxx xxx	0214 940 034	0 2 14 940 0 34
0xxx xx xx xx	0899 51 16 43	0 8 99 51 16 43

The abbreviations *tel.* (telefon), *kom.* (komórka), *tel.kom.* or *tel. kom.* (telefon komórkowy), *tel/fax* and *tel./fax* (telefon i fax) are expanded if they occur in front of a valid phone number. The words *telefon*, *komórka*, and *fax* can also be found in front of a phone number.

Some additional formats are also recognized as phone numbers if one of the above words or abbreviations, or an international prefix, or a combination, appear in front of the number:

The additional format are:

xxx xxx xxx  
 xxxxxxxxx  
 x xxx xxx xxx

Example	Reading
Tel. 123 456 789	telefon 123 456 789
tel. (22) 494 00 34	telefon 22 494 0 0 34
tel.kom. 016231645	telefon komórkowy 016 231 645
tel/fax 016231645	telefon i fax 016 231 645

### 5.10.2 International phone numbers

International phone numbers follow the pattern below:

*International prefix + Country code + space or hyphen + Local number.*

International prefix:	+ or 00
Country code:	1-3 digits

All the local phone number formats described in the previous section will be recognised as an international phone number is preceded by an international prefix.

Example	Reading
+48 31-77-546-666	plus 48 31 77 546 666
0048 (22) 494 00 34	0 0 48 22 494 0 0 34
+31 (09) 494-00-34	plus 31 0 9 494 0 0 34
004 (788) 516 41 41	0 0 4 788 516 41 41
0048 (31) 5253707	0 0 48 31 5 2 5 3 7 0 7

Some additional formats below are recognized as a phone number if preceded by an international prefix [not SP]:

Format	Example	Reading
xx xxxx xxxx	+42 5674 3445	plus 42 56 74 34 45
xxx xx xxxx	00342 45 5678	0 0 342 45 56 78
xxx xxx xxxx	+456 456 4436	plus 456 456 44 36
xxx xxxx xxx	+787 450 30 00	plus 787 450 30 0 0
xxxx xx xx xx	004567 22 13 16	0 0 45 67 22 13 16
xxx xxxx xxxx	00317 4501 0001	0 0 317 45 0 1 0 0 0 1
xxxx xxx xxxx	+2345 200 4545	plus 23 45 200 45 45
xxxx xxxx xxx	+4532 3030 4232	plus 45 32 30 30 42 32
xxxxx xx xx xx	0034567 45 00 45 43	0 0 3 4 5 6 7 45 0 0 45 43
xxxx xxxx xxxx	+0423 3454 4567	plus 0 4 23 34 54 45 67
xxxxx xxx xxxx	+34500 345 4545	plus 34 500 345 45 45
xxxxx xxxx xxx	0034500 4567 560	0 0 3 4 5 0 0 45 67 560
xxxxxx xx xx xx	+600434 45 45 66	plus 60 0 4 34 45 45 66

## **6   *How to change pronunciation***

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 *Polish phonetic text*). Phonetic transcriptions can also be entered directly into the text, using *PRN* tags (see *User's guide*).

## 7 Polish phonetic text

The Polish text-to-speech system primarily uses the Polish subset of the SAMPA phonetic alphabet (*Speech Assessment Methods Phonetic Alphabet*). The symbols are written with a space after each phoneme.

Only the symbols listed here may be used in phonetic transcriptions. Symbols not listed here are not valid in phonetic transcriptions and will be ignored if included in the user lexicon or in a *PRN* tag. The symbol "1" immediately following vowels in the tables below indicates lexical stress on prominent syllables (for further details c.f. section *Lexical stress*).

### 7.1 Consonants

Table: Symbols for the Polish consonants

Symbol	Word	Phonetic text	Comment
b	bach	b a1 x	
t	taka	t a1 k a	
p	pac	p a1 ts	
d	dalej	d a1 l e j	
k	klap	k l a1 p	
g	gazu	g a1 z u	
m	mach	m a1 x	
n	nami	n a~1 m' i	
r	rano	r a~1 n o	
l	lada	l a1 d a	
N	ongi	o~1 N g' i	
f	forma	f o1 r m a	
v	wami	v a~1 m' i	
s	sam	s a~ m	
S	szast	S a1 s t	
z	zatem	z a1 t e~ m	
Z	żaden	Z a1 d e~ n	
x	chodnik	x o1 d n' i k	
j	ja	j a	
w	łup	w u1 p	
ts	cup	ts u1 p	
tS	czasem	tS a1 s e~ m	
dz	dzyń	dz l~1 n'	
dZ	john	dZ o~1 n	
j~	państwa	p a~1 j~ s t f a	
w~	czyją	tS l1 j o~ w~	

Symbol	Word	Phonetic text	Comment
dd	oddaj	o1 dd a j	Geminate*
nn	inna	i~1 nn a	Geminate
ll	millar	m' i1 ll e r	Geminate
ss	bezsens	b e1 ss e~ w~ s	Geminate
SS	bliższa	b l' i1 SS a	Geminate
zz	rozzłość	r o1 zz w o s' ts'	Geminate
ww	mełło	m e1 ww o	Geminate
b'	biada	b' j a1 d a	Palatalized consonant
t'	tirli	t' i1 r l' i	Palatalized consonant
p'	pik	p' i1 k	Palatalized consonant
d'	dialekt	d' j a1 l e k t	Palatalized consonant
k'	kilka	k' i1 l k a	Palatalized consonant
g'	gigant	g' i1 g a~ n t	Palatalized consonant
m'	million	m' i1 l' j o~ n	Palatalized consonant
n'	niemiec	n' e~1 m' j e ts	Palatalized consonant
r'	austria	a1 w~ s t r' j a	Palatalized consonant
l'	lipiec	l' i1 p' j e ts	Palatalized consonant
f'	fik	f' i1 k	Palatalized consonant
v'	wiele	v' j e1 l e	Palatalized consonant
s'	siedem	s' e1 d e~ m	Palatalized consonant
z'	zima	z' i~1 m a	Palatalized consonant
x'	chi	x' i1	Palatalized consonant
ts'	ciebie	ts' e1 b' j e	Palatalized consonant
dz'	dzięki	dz' e~1 N k' i	Palatalized consonant
nn'	inni	i~1 nn' i	Palatalized geminate
ss'	passie	p a1 ss' e	Palatalized geminate
zz'	rozziew	r o1 zz' e f	Palatalized geminate

\* Note on geminates: any other "double" graphemes can be represented by a sequence of two non-geminate symbols (e.g. "kk" /k k /).

## 7.2 Vowels

Table: Symbols for the Polish vowels

Symbol	Word	Phonetic text	Comment
a	aby	a1 b l	Oral vowel
e	euro	e1 w r o	Oral vowel
i	ile	i1 l e	Oral vowel
l	pstryk	p s t r l1 k	Oral vowel
o	obok	o1 b o k	Oral vowel
u	ucisk	u1 t s' i s k	Oral vowel
a~	anglik	a~1 N g l' i k	Nasal vowel
e~	cement	t s e~1 m e~ n t	Nasal vowel
i~	intro	i~1 n t r o	Nasal vowel
l~	wymiar	v l~1 m' j a r	Nasal vowel
o~	bom	b o~1 m	Nasal vowel
u~	bum	b u~1 m	Nasal vowel

## 7.3 Lexical stress

In words with more than one syllable, one of the syllables can be perceived as more prominent than the others. This is referred to as *word stress*, or *lexical stress*. Words consisting of one syllable also bear primary word stress when spoken in isolation, although many may lose stress in certain contexts. For the correct pronunciation of a word, it is important to include the symbol marking word stress.

In phonetic transcriptions, word stress is indicated by the symbol /1/ immediately following the stressed vowel with no space between the vowel and the stress symbol. (See examples below.)

Secondary stress can also occur in Polish in long words. Secondary stress refers to full (i.e. very clear) pronunciation of a vowel, but there are no main intonation changes on such vowels (unlike with primary stress). The secondary stress is indicated by the symbol /2/ immediately following the stressed vowel in the same way as for primary stress.

### Examples

praliny

technologicznej

kurs

### Transcription

p r a l' i~1 n l

t e2 x n o l o g' i1 t s n e j

k u1 r s

## 7.4 Word-boundary changes in transcriptions

---

### 7.4.1 Word boundary palatalization

Word boundary palatalization is produced in cases where a word ends with a consonant and the next word begins with the letter *i* or *j*. In this case, the consonant is palatalized.

Examples	Transcription
Pan	p a~1 n
Jan	j a~1 n
pan Jan	p a~1 n' j a~1 n

### 7.4.2 Word boundary nasalization

Word boundary nasalization occurs when the last letter in a word is a vowel and the next word begins with a nasal consonant *n*, *m*, *w* or *j*. The vowel in that case is nasalized.

Examples	Transcription
Mała	m a1 w a
Miska	m' i1 s k a
mała miska	m a1 w a~ m' i1 s k a

### 7.4.3 Word boundary voicing

Word boundary voicing occurs when a word ends with a voiceless consonant (obstruent) and the next word begins with a voiced consonant (obstruent). In that case the voiceless consonant gets voiced.

Examples	Transcription
brat	b r a1 t
zosi	z o1 s' i
brat zosi	b r a1 d z o1 s' i

## 7.5 Glottal stop

---

A glottal stop is a small glottal sound, which is often produced in speech to separate two words (or compound parts of the word) when the second word/part starts with a vowel. In phonetic transcriptions, it is represented by the phonetic symbol /ʔ/. This sound is often produced in clear speech. Glottal stop can also optionally be inserted in transcriptions in order to improve pronunciation, as in the examples below.

Examples	Transcription
Bo mam ich, a oni mnie.	b o~ m a~1 m ʔ i1 x _ a ʔ o~1 n' i~ m n' e1
Reaktor	r e ʔ a1 k t o r

## 7.6 Pause

---

An underscore symbol /\_/ in phonetic transcriptions generates a small pause.



## 8 Abbreviations

In current version of the Polish text-to-speech system, abbreviations are recognized in all contexts. Examples of such abbreviations are given in the table below. Some of these abbreviations are case-sensitive: i.e. particularly those containing upper-case characters in the below table, while others are case-insensitive. Also note that some of the abbreviations require full stops in order to be recognized as an abbreviation.

As mentioned above, there are also abbreviations for the days of the week and the months (see section *Dates*), as well as in numbers (see chapter *Number processing*).

Table: Abbreviations

Abbreviation	Reading
ok.	około
n.e.	naszej ery
W-wa	Warszawa
W-w	Wrocław
Dz. U. *	dziennik ustaw
St.Zj. *	Stany Zjednoczone
St.Zjedn. *	Stany Zjednoczone
Sz. Pan *	Szanowny Pan
Sz. Pani *	Szanowna Pani
Sz. Państwo *	Szanowni Państwo
W. Pan *	Wielmożny Pan
W. Pani *	Wielmożna Pani
W. Państwo *	Wielmożni Państwo
Wlk. Bryt. *	Wielka Brytania
adm.	administracja
al.	aleja
bp	biskup
doc.	docent
dr	doktor
dr hab.	doktor habilitowany
ds.	do spraw
gen.	generał
z-ca	zastępca
im.	imienia
instyt.	instytucja
itd.	i tak dalej

Abbreviation	Reading
itp.	i tym podobne
k.k.	kodeks karny
kard.	kardynał
ks.	ksiądz
lek. med.	lekarz medycyny
lek. wet.	lekarz weterynarii
m.in.	między innymi
m. in.	między innymi
mgr	magister
n.p.m.	nad poziomem morza
np.	na przykład
nr	numer
p.n.e.	przed naszą erą
p.o.	pełniący obowiązki
p.p.m.	pod poziomem morza
pl.	plac
prof.	profesor
pw.	pod wezwaniem
płd.	południowy
płn.	północny
s-ka	spółka
sp.c.	spółka cywilna
sp.j.	spółka jawna
sp. kom.	spółka komandytowa
sp. p.	spółka partnerska
pod tyt.	pod tytułem
pod nr	pod numerem
tn.	to znaczy
tw.	tak zwane
ul.	ulica
wg	według
woj.	województwo
wsch.	wschodni
zakł.	zakład
św.	święty
ang.	angielski
austral.	australijski
bryt.	brytyjski

Abbreviation	Reading
europ.	europejski
franc.	francuski
germ.	germański
hiszp.	hiszpański
hol.	holenderski
max.	maksimum

The abbreviations of nouns are expanded to their default form: number SING, gender MASC, case NOM; in the same manner all the abbreviated adjectives are expanded to the default form SING MASC NOM.

The abbreviations for measurements m, g and s are only expanded when appearing after a number in gender/number agreement with the preceding number.

Expression	Reading	
25 m	dwadzieścia pięć metrów	[not SP]
1 s	jedna sekunda	[not SP]
2 g	dwa gramy	[not SP]

The following abbreviations are also recognized, and vary according to the number preceding them:

Expression	Reading	
mm <sup>2</sup>	milimetr kwadratowy	[not SP]
cm <sup>2</sup>	centymetr kwadratowy	[not SP]
m <sup>2</sup>	metr kwadratowy	[not SP]
km <sup>2</sup>	kilometr kwadratowy	[not SP]
mm <sup>3</sup>	milimetr sześcienny	[not SP]
cm <sup>3</sup>	centymetr sześcienny	[not SP]
km	kilometr	[not SP]
mm	milimetr	[not SP]
cm	centymetr	[not SP]
°C	stopień Celsjusza	[not SP]
°F	stopień Fahrenheita	[not SP]

## 9 Web-addresses and email

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

- *www* is read as three *v* following each other.
- Full stops (".") are read as *kropka*, hyphens ("-") as *łącznik*, underscore ("\_") as *podkreślnik*, slash ("/") as *ukośnik*.
- Country codes (e.g. *pl*, *us*, *fr*) are spelled out letter by letter.
- @ is read as *małpa*.
- Words/strings (including *org*, *com*, and *edu*) are pronounced according to the normal rules of pronunciation in the system and in accordance with the lexicon.

### Expression

*www.acapela-group.com*  
*http://www.acapela-group.com*  
  
*ewa@yahoo.pl*  
*ewa\_bednarz@yahoo.pl*

### Reading

*wu wu wu kropka akapela łącznik group kropka kom*  
*ha te te pe dwukropek ukośnik ukośnik wu wu wu*  
*kropka akapela łącznik group kropka kom*  
*ewa małpa jahoo kropka pe el*  
*ewa podkreślnik bednarz małpa jahoo kropka pe el*