Language Manual

HQ Polish

Language Manual – Polish: Ania

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

Acapela Group
Box 1328
SE - 171 26 Solna
Sweden
Tel. +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 Lac
BP 98153
31681 Labège Cedex
Tel. +33 (0)5 62 24 71 00
Fax. +33 (0)5 62 24 71 01

http://www.acapela-group.com
Published March 15, 2011

Table of Contents

Table of Contents....3
List of Tables....4
Chapter 1 General....5
Chapter 2 Letters in orthographic text....6
Chapter 3 Punctuation characters....7
3.1 Comma, colon and semicolon....7
3.2 Quotation marks....7
3.3 Full stop....7
3.4 Question mark....7
3.5 Exclamation mark....7
3.6 Parentheses....7
Chapter 4 Other non-alphanumeric characters....8
4.1 Non-punctuation characters....8
4.2 Symbols with pronunciation variable depending on the context....9
Chapter 5 Number Processing....10
5.1 Full number pronunciation....10
5.2 Leading zero....11
5.3 Decimal numbers....11
5.4 Currency amounts....11
5.5 Ordinal numbers....12
5.6 Arithmetical operators....13
5.7 Mixed digits and letters....13
5.8 Time of day....13
5.9 Dates....14
5.10 Phone numbers....15
Chapter 6 How to change pronunciation....18
Chapter 7 Polish Phonetic Text....19
7.1 Consonants....19
7.2 Vowels....20
7.3 Lexical stress....20
7.4 Word boundary assimilation....21
7.5 Glottal stops....21
7.6 Pause....21
Chapter 8 Abbreviations....22

List of Tables

Table 4-1 Non-punctuation characters....8
Table 7-2 Symbols for the Polish consonants....19
Table 7-3 Symbols for the Polish vowels....20
Table 8-4 Polish Abbreviations....22

Chapter 1 General

This document discusses certain aspects of text-to-speech processing for the Polish 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) voice Ania.

Please note that the *User's Guide*, mentioned several times in the manual, is called Help in some applications.
Note: This language manual is general and applies to all Acapela Group HQ Italian voices specified above. One or more of the voices may be included in a certain Acapela Group product.

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 Symbian

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*.

Chapter 2 Letters in orthographic text

Characters from A-Ż and a-ż 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 e, s 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.

Chapter 3 Punctuation characters

Punctuation marks appearing in a text affect both the rhythm and the 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 to occur in a sentence, accompanied by a small rising intonation pattern just prior to the character.

3.2 Quotation marks

Quotes ‘“”’ 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 terminal punctuation mark which causes a falling end-of-sentence intonation pattern and is accompanied by a somewhat longer pause. A full stop may also be used as a delimiter 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 question intonation, first rising and then falling.

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

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

.

Chapter 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 4-1 Non-punctuation characters

Symbol	Reading
/	ukośnik
\	ukośnik wsteczny
+	plus
$	dolar
€	euro
<	mniejszy niż
>	większy niż
%	procent
^	daszek
|	pionowa linia
~	tylda
@	małpa
=	równa się
-	(see below)
*	(see below)

4.2 Symbols with pronunciation variable 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.

Expression	**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 an equals sign. In other cases it is pronounced as *gwiazdka.*

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

Chapter 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 below, along 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

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 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 *999999999999* (twelve digits). Numbers higher than this are read as separate digits.

Number	**Reading**
2580	dwa tysiące pięćset osiemdziesiąt
2 580	“
2.580	“
25800	dwadzieścia pięć tysięcy osiemset

25 800	“
25.800	“
2580350	dwa miliony pięćset osiemdziesiąt tysięcy trzysta pięćdziesiąt
2 580 350	“
2.580.350	“
1000000000	jeden miliard
1234567890123	jeden dwa trzy cztery pięć sześć siedem osiem dziewięć zero jeden dwa trzy
234 567 890 123	dwieście trzydzieści cztery miliardy pięćset sześćdziesiąt siedem milionów osiemset dziewięćdziesiąt tysięcy sto dwadzieścia trzy

“*Tys.*”, the abbreviation for “thousand”, is expanded in agreement with the preceding number. [not SP]

Number	**Reading**	
23 tys.	dwadzieścia trzy tysiące	[not SP]
1 tys.	jeden tysiąc	[not SP]

5.2 Leading zero

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

Number	**Reading**
09253	zero dziewięć tysięcy dwieście pięćdziesiąt trzy
020	zero dwadzieścia

5.3 Decimal numbers

Comma is used when writing 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.* If the decimal (the part after comma) has more than three places, the decimal part is read as separate digits, otherwise it is read as a full number.

Number	**Reading**
2,50	dwa przecinek pięćdziesiąt
3,141	trzy przecinek sto czterdzieści jeden
3,1411	trzy przecinek jeden cztery jeden jeden
1251,04	tysiąc dwieście pięćdziesiąt jeden przecinek zero cztery
1.251,04	tysiąc dwieście pięćdziesiąt jeden przecinek zero cztery

5.4 Currency amounts

The following principles are followed for currency amounts:

- Numbers with zero or 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 euro	piętnaście euro	[not SP]
15,50 zł.	piętnaście złotych i pięćdziesiąt groszy	[not SP]
€ 200.50	dwieście euro i pięćdziesiąt centów	
1.000.000 GBP	jeden milion funtów	[not SP]

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

€ 1 milion	jeden milion euro	[not SP]

In this format, abbreviations used for the large numbers such as thousands, millions and billions are expanded in agreement with the preceding number*: tys., mln, mld*

Number	**Reading**	
$ 23 tys.	dwadzieścia trzy tysiące dolarów	[not SP]
$ 10 mln	dziesięć milionów dolarów	[not SP]
$ 1 mld	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.

Expression	**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	
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:

10-cio milionowy	dziesięcio-milionowy	[not SP]

5.6 Arithmetical operators

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

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

5.7 Mixed digits and letters

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

Expression	Reading
77B84D3	siedemdziesiąt siedem be osiemdziesiąt cztery de trzy
0092B87-B	zero zero dziewięćdziesiąt dwa be osiemdziesiąt siedem be

5.8 Time of day

The colon and full stop are 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.

The 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 dziewiąta trzydzieści	[not SP]
9:30.20	godzina dziewiąta trzydzieści i dwadzieścia sekund	[not SP]
9:00	godzina dziewiąta zero zero	
g. 9:00	godzina dziewiąta zero zero	[not SP]
12 h 30 min	dwanaście godzin i trzydzieści minut	[not SP]
1 godz. 45 min	jedna godzina i czterdzieści pięć minut	[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* 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. Zeros may be used in front of numbers below 10.

Examples of valid formats and their readings:

Type 1:			
10.02.2003	or	10.2.2003	dziesiąty lutego dwa tysiące trzy
10/02/2003	or	10/2/2003	“
Type 2:			
10.02.03	or	10.2.03	dziesiąty lutego dwa tysiące trzy
10/02/03	or	10/2/03	“
Type 3:			[not SP]
2003.02.10	or	2003.2.10	dziesiąty lutego dwa tysiące trzy

2003-02-10	or	2003-2-10	“
Type 4:			[not SP]
10 II 2003		dziesiąty lutego dwa tysiące trzy	
10.II.2003			“
Type 5:			[not SP]
10 II 03		dziesiąty lutego dwa tysiące trzy	
10.II.03			“

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

Expression	**Reading**
1998-1999	tysiąc dziewięćset dziewięćdziesiąty ósmy do tysiąc dziewięćset dziewięćdziesiąty dziewiąty
1939-45	tysiąc dziewięćset trzydziesty dziewiąty do czterdziesty piąty
7–8.02.2007	siódmy do ósmego lutego dwa tysiące siedem
od 5 do 13 maja	od piątego do trzynastego maja

Other possible formats include: [not SP]

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 trzecego roku

Valid abbreviations for months: *stycz., mar., marz., kwie., czerw., czer., lip., sierp., sier., wrze., paźdz., listop., list., grudz.,* and *grud*. [not SP]

Valid abbreviations for days: *pn., pon., wt., śr., cz., czw., pt., so., sob., nd.* and *niedz*. [not SP]

5.10 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 with a pause between the groups.

The digit strings will be read as follows:

- Groups of 1, 2 or 3 digits are read out as regular numbers.
- Groups of 4 digits are split into two groups of 2.
- Groups of 5 digits are split into a group of 2 and a group of 3.
- Groups of 6 digits are split into three groups of 2.
- Groups of 9 digits are split into three groups of 3.

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 hyphen, and preceded by *0* :

Format [not SP]

- x xxxx xxxx
- x xx xx xx xx
- xx xxx xxxx
- xx xxxx xxx
- xxx xx xx xx
- xx xxx xx xx
- xxx xxx xx xx
- (xx) xxx xx xx
- (xxx) xxx xx xx
- (xxx) xxxxxxx
- (xxx)xxxxxxx
- xx xx xxx xxx

The following sequences can only appear in these formats:

Format [not SP]

- 0-xx xxx xx xx
- 0xxx xxx xxx
- 0xxx xx xx xx

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 phone number. The words *telefon*, *komórka*, and *fax* can also be found in front of a phone number.

Expression	**Reading**	
tel. 123 456 789	telefon 123 456 789	[not SP]
tel.kom. 016231645	telefon komórkowy 016 231 645	[not SP]
tel/fax 016231645	telefon i fax 016 231 645	[not SP]

The following formats are only recognized as phone numbers if one of the above words or abbreviations, or an international prefix, or both, appear in front of the number:

Format [not SP]

- xxx xxx xxx
- xxxxxxxxx
- x xxx xxx xxx

5.10.2. International phone numbers

All the preceding formats can be recognised if preceded by an international prefix [not SP]:

Format	**Example**
+x	+1 (34) 567 89 05
+xx	+12 (34) 567 89 05
+xxx	+123 (34) 567 89 05
00x	001 123 456 789
00xx	0012 123 456 789
00xxx	00123 123 456 789

The following formats are specific international phone numbers, and will be preceded by 00 or +:

Format [not SP]

- xx xxxx xxxx

- xxx xx xxxx
- xxx xxx xxxx
- xxx xxxx xxx
- xxxx xx xx xx
- xxx xxxx xxxx
- xxxx xxx xxxx
- xxxx xxxx xxx
- xxxxx xx xx xx
- xxxx xxxx xxxx
- xxxxx xxx xxxx
- xxxxx xxxx xxx
- xxxxxx xx xx xx
- (xx xx) xxx xx xx

Chapter 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 in the text, using a PRN-tag (see *User's guide*).

Chapter 7 Polish Phonetic Text

The Polish text-to-speech system uses symbols based on the SAMPA phonetic alphabet (*Speech Assessment Methods Phonetic Alphabet*). The symbols are written with a space between 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.

7.1 Consonants

Table 7-2 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 I~1 n'	
dZ	jordan	dZ o1 r d a~ n	
j~	państwa	p a~1 j~ s t f a	
w~	czyją	tS I1 j o~ w~	
tt	nadto	n a1 tt o	Geminate
ttS	czczy	ttS I1	Geminate
dd	oddaj	o1 dd a j	Geminate
ddZ	dżdży	ddZ I1	Geminate
kk	miękko	m' j e~1 N kk o	Geminate
nn	inna	i~1 nn a	Geminate
ll	miller	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’	milion	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
kk’	lekkie	l e1 kk’ e	Palatalized geminate
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

7.2 Vowels

Table 7-3 Symbols for the Polish vowels

Symbol	Word	Phonetic text	Comment
a	aby	a1 b I	Short vowel
e	euro	e1 w r o	Short vowel
i	ile	i1 l e	Short vowel
I	pstryk	p s t r I1 k	Short vowel
o	obok	o1 b o k	Short vowel
u	ucisk	u1 ts’ i s k	Short vowel
a~	anglik	a~1 N g l’ i k	Nasal vowel
e~	cement	ts e~1 m e~ n t	Nasal vowel
i~	intro	i~1 n t r o	Nasal vowel
I~	wymiar	v I~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

A lexical accent is used to indicate the level of prominence (or emphasis) of a syllable in a word. In words with more than one syllable, one of the syllables is more prominent than the others. This is referred to as word stress, or lexical stress. Words of one syllable also have word stress when spoken in isolation, although many may lose the stress in certain contexts. For the correct pronunciation of a word, it is important to include stress marks when writing phonetic transcriptions. Polish words have primary stress mostly on the penultimate syllable. Many longer words have a secondary stress on the first syllable of the word.

In phonetic transcriptions, primary stress is indicated by the symbol 1 attached to the stressed vowel. Secondary stress, common in many longer words, is indicated by the symbol 2 attached to the stressed vowel.

7.4 Word boundary assimilation

7.4.1. Word boundary gemination

Word boundary gemination refers to a phenomenon which occurs at word boundaries. The Polish text-to-speech system can produce word boundary gemination in cases where a word ends, and the next word begins, with the same consonant. In this case, the two consonants are merged into a single geminate.

Example:

akurat / a k u1 r a t /	trzeba / t S e1 b a /
akurat trzeba	/ a k u1 r a tt S e1 b a/

7.4.2. 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.

Example:

pan /p a~1 n/	jan /j a~1 n/
pan jan	/p a~1 n' j a~1 n /

7.4.3. 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.

Example:

mała /m a1 w a/	miska /m' i1 s k a/
mała miska	/m a1 w a~ m' i1 s k a/

7.4.4. 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.

Example:

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

7.5 Glottal stops

A glottal stop, represented by the phonetic symbol /*?*/, is a small sound which is often used in front of words beginning with a vowel or at vowel meetings. This sound can be inserted in a transcription in order to improve the pronunciation.

7.6 Pause

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

Chapter 8 Abbreviations

In the current version of the Polish text-to-speech system, the abbreviations in the table below are recognized in all contexts. These abbreviations are mostly case-insensitive (except for those indicated below by '*'). The presence or absence of a full stop after the abbreviation should be observed.

As previously mentioned, there are also abbreviations for the days of the week and the months (see chapter Ordinal numbers).

Table 8-4 Polish Abbreviations

Abbreviation	Reading
ok.	około
n.e.	naszej ery
cm	centymetrów
km	kilometrów
dm	decymetrów
mm	milimetrów
km/godz	kilometrów na godzinę
m/s	metrów na sekundę
kg	kilogramów
°C	stopnie celsiusza
°F	stopnie fahrenheita
proc.	procent
dol.	dolarów
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
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
tzn.	to znaczy
tzw.	tak zwany
ul.	ulica
wg	według
woj.	województwo
wsch.	wschodni
zakł.	zakład
św.	święty
ang.	angielski
austral.	australijski
bryt.	brytyjski
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 [not SP].

Expression	**Reading**
25 m	wadzieścia pięć metrów
2 g	dwa gramy
1 s	jedna sekunda

The following abbreviations are also recognized, and vary according to the number preceding them [not SP]:

Expression	**Reading**
mm2	milimetr kwadratowy
cm2	centymetr kwadratowy
m2	metr kwadratowy
km2	kilometr kwadratowy
mm3	milimetr sześcienny
cm3	centymetr sześcienny
m3	metr sześcienny
km3	kilometr sześcienny

Chapter 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.
- The @ is read *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	**Reading**
www.acapela-group.com	vu vu vu kropka akapela łącznik group kropka kom
http://www.acapela-group.com	ha te te pe dwukropek ukośnik ukośnik vu vu vu kropka akapela łącznik group kropka kom
ewa@yahoo.pl	ewa małpa jahoo kropka pe el
ewa_bednarz@yahoo.pl	ewa podkreślnik bednarz małpa jahoo kropka pe el