



Language Manual

HQ and CO Italian

Language Manual: HQ and CO Italian

Published 5 February 2015

Copyright © 2008-2015 Acapela Group.

All rights reserved

This document was produced by Acapela Group. We welcome and consider all comments and suggestions. Please, use the *Contact* 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	14
7	ITALIAN PHONETIC TEXT	15
8	ABBREVIATIONS AND SYMBOLS.....	18
9	WEB-ADDRESSES AND EMAIL.....	21

1 General

This document discusses certain aspects of text-to-speech processing for the Italian 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) Italian 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

Characters from A-Z, a-z, as well as á, é, í, ó, ú, è, ì, à, ò may constitute a word. Certain other characters are also considered as letters, notably those used as letters in other European languages, i.e. ñ, Ñ, ü, ö, å.

Characters used in other languages å, ö, ç, are mapped into readable characters, for instance å is read as a.

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 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 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 similar manner to the 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
/	barra
	barra verticale
+	più
\$	dollaro/i
£	sterlina/e
€	euro
¥	yen
<	minore di
>	maggiore di
%	per cento
^	accento circonflesso
~	tilde
@	chiocciola
=	uguale
²	See below
³	See below
-	See below
*	See below

\$ alone is read as *dollaro*; £ alone is read as *sterlina*.

4.2 The ² and ³ signs

The reading of expressions with ² and ³ is:

Expression	Reading
mm ²	millimetro/i quadrato/i
cm ²	centimetro/i quadrato/i
m ²	metro/i quadrato/i

Expression	Reading
km ²	chilometro/i quadrato/i
mm ³	millimetro/i cubico/i
cm ³	centimetro/i cubico/i
m ³	metro/i cubico/i
km ³	chilometro/i cubico/i

All the units of measure such as *mm*, *mm²*, *cm*, *cm²* are read in the plural form when alone.

4.3 Symbols whose pronunciation varies depending on the context

4.3.1 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, i.e. as in the pattern -X but not in X-Y or X-Z where X, Y, and Z are numbers.
2. if followed by a digit and an equals sign '=', i.e. as in the pattern X-Y=Z. Space is allowed between digits, hyphen and equals sign.

If there is no equals sign, as in X-Y or X-Z, the hyphen is pronounced *trattino*.

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

Expression	Reading	
-3	meno tre	
44-3	quarantaquattro trattino tre	
44-3=41	quarantaquattro meno tre uguale quarantuno	
44 - 3 = 41	quarantaquattro meno tre uguale quarantuno	
feb 12-14	febbraio dodici a quattordici	[not SP]
febbraio 6-10	febbraio sei a dieci	[not SP]
2001-2004	duemilauno a duemilaquattro	[not SP]
02-02-2002	due febbraio duemiladue	
ex-ministro	ex ministro	

4.3.2 Asterisk

Asterisk '*' is pronounced *per* if enclosed by digits that are followed by '='. In other cases it is pronounced *asterisco*.

Expression	Reading
2*3=6	due per tre uguale sei
2*3	due asterisco tre
*bc	asterisco BC

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 familiarise the user with the various types of formatted and non-formatted strings of digits that are recognised by the system, we provide below a brief description of the basic number processing 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
Telephone 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 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 cardinal number read is 999999999999 (12 digits). Larger numbers are read as separate digits.

Number	Reading
1103	mille centotre
1.103	mille centotre
1 103	mille centotre
128141	centoventottomila centoquarantuno
128 141	centoventottomila centoquarantuno
128.141	centoventottomila centoquarantuno
2128141	due milioni centoventottomila centoquarantuno
2 128 141	due milioni centoventottomila centoquarantuno
2.128.141	due milioni centoventottomila centoquarantuno

5.2 Leading zero

Numbers that begin with 0 (zero) are read as a zero followed by the number read as a whole.

Number	Reading
09253	zero nove due cinque tre
020	zero venti

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*. If the decimals (the part after comma or full stop) are more than three, the decimal part is read as separate digits.

Note: A number followed by a decimal and exactly three digits is read as a full number, following the rules in the section *Full number pronunciation*.

One or two digits followed by a decimal and by two digits are interpreted as hour format, for example: 2.50 is read as *due e cinquanta*.

Number	Reading
16,224	sedici virgola duecento ventiquattro
3,1415	3 virgola 1 4 1 5
1103,04	mille centotre vigola zero quattro

Number	Reading
1.103,04	mille centotre vigola zero quattro
2,50	2 virgola 50
2.50	2 e 50
3.141	tremila cento quarantuno

5.4 Currency amounts

The following principles are followed for currency amounts:

- Numbers with zero, one or two decimals preceded or followed by either the currency symbols £, \$, ¥, €, L, L. or € or [not SP] the abbreviations Lit., DM, are read as monetary amounts.
- [not SP] Numbers with zero, one or two decimals followed by the words *sterlina, dollaro, yen, lira, deutch mark* or *euro* (singular or plural) are read as monetary amounts.
- Accepted decimal markers are comma ',' and full stop '.'.
- The decimal part (consisting of two digits) in monetary amounts is read as *i nn pence* and *i nn centesimi*.
- If the decimal part is 00 it will not be read.

Example	Reading	
\$15.00	quindici dollari	
15.00£	quindici sterline	
15.00 euro	quindici euro	[not SP]
€ 200.50	duecento euro e cinquanta centesimi	
1.000.000 ¥	un milione yen	

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

\$ 1 milione	un milione dollari
--------------	--------------------

5.5 Ordinal numbers

Numbers are read as ordinals when followed by *o, a, i, e*.

Expression	Reading
2o	secondo
3a	terza
4e	quarte
5i	quinti

5.6 Arithmetic operators

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

Expression	Reading
-12	meno dodici
14-2	quattordici trattino due
14-2=12	quattordici meno due uguale dodici
+24	più ventiquattro
2+3	due più tre
2+3=5	due più tre uguale cinque
2*3=6	due per tre uguale sei
2/3	due terzi
6/2=3	sei diviso due uguale tre
25%	25 per cento
3,4%	tre virgola quattro per cento

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. One, two or three digits are pronounced as a normal numbers, four digits are pronounced as two groups of two digits and more than four digits are spelled out.

Expression	Reading
77B84Z3	settantasette bi ottantaquattro zeta tre
0092B87-B	zero zero novantadue bi ottantasette bi
FT2592B87Z	effe ti venticinque novanta due bi ottantasette zeta
TN12345L5	ti enne uno due tre quattro cinque elle cinque

5.8 Time of day

The colon is used to separate hours, minutes and seconds. Possible time formats are:

- hh:mm* or *h:mm*
- hh:mm:ss* or *h:mm:ss*
- hh* or *hh-hh*

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

In pattern a:

If the *mm*-part is equal to *00*, this part will not be read. This pattern can be preceded or followed by time indications such as *A.M.*, *AM*, *a.m.*, *P.M.*, *PM*, *p.m.*, *h*, or *h..*

In pattern b:

An *i* will be inserted before the *ss*-part, and *secondi* will be added after it. If the *ss*-part is equal to *00*, this part will not be read. This pattern can be preceded or followed by time indications such as *A.M.*, *AM*, *a.m.*, *P.M.*, *PM*, *p.m.*, or *PM*.

[not SP] In pattern c:

The hours can appear alone, but must be preceded by time indications, such as: *A.M.*, *AM*, *a.m.*, *P.M.*, *PM*, *p.m.*, or *PM*. They can also appear in a time range and must then be followed by a time indication.

The character *H* in a time expression such as *2H20* is read as *e*.

Expression	Reading	
2:40 a.m.	2 e 40 di mattina	
2:40 p.m.	2 e 40 del pomeriggio	
08:22 or 8:22	otto e ventidue	
08:22:33 or 8:22:33	otto ventidue minuti e trentatré secondi	
08-09 a.m.	otto alle nove di mattina	[not SP]
08-09 p.m.	otto alle nove di sera	[not SP]
2H45	due e quarantacinque	[not SP]

5.9 Dates

The valid date formats are:

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

Examples of valid formats and their readings:

Type 1:

10-02-2003 or 10-2-2003	dieci febbraio duemila tre
10.02.2003 or 10.2.2003	"
10/02/2003 or 10/2/2003	"

Type 2:

10-02-03 or 10-2-03	dieci febbraio duemila tre
10.02.03 or 10.2.03	"
10/02/03 or 10/2/03	"

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

Expression	Reading
1998-1999	1998 a 1999
1968-74	1968 a 74
14-15 febbraio	14 a 15 febbraio
febbraio 14-15	febbraio 14 a 15

[not SP] Other possible date formats include:

Expression	Reading
Martedì, 30 apr. 1999	martedì trenta aprile millenovecentonovantanove
5 ago. 2003	cinque agosto duemilatre
05 08 2003	"
05 08 03	"
2003-08-05	"
2003/08/05	"

[not SP] Abbreviations of months and days in date formats:

Months:

*gen., genn., feb., febbr., mar., apr., magg., mag., giu., lug., ago.,
ag., sett., ott., nov., dic.*

Days:

lun., mar., mer., gio., ven., sab., dom.

5.10 Phone numbers

In this section the patterns of digits that are currently recognized as telephone numbers are illustrated. [not SP] Mobile phone numbers to be recognized have to be written with a space between the first three digits and the rest of the digits.

- xxx xxxxxx

- xxx xx xx xx
- xxx xxx xxxx
- xxx xxx xx xx

Mobile numbers.

- xxx xxxxxx

International phone numbers have the following pattern:

International prefix + Country code (with or without parenthesis) + space + Local number.

Expression	Reading	
081640421	zero otto uno sei quattro zero quattro due uno	
081 640421	zero otto uno sei quattro zero quattro due uno	
081 64 04 21	zero ottantuno sessantaquattro zero quattro due uno	
003981640421	zero zero tre nove otto uno sei quattro zero quattro ventuno	
0039 81 64 04 21	zero zero trentanove ottantuno sessantaquattro zero quattro ventuno	
347 456123	tre quattro sette quattro cinque sei uno due tre	[not SP]
0039 347 456123	zero zero tre nove tre quattro sette quattro cinque sei uno due tre	[not SP]

6 *How to change the 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 *Italian phonetic text*). Phonetic transcriptions can also be entered directly in the text, using a *PRN* tag (see *User's guide*).

7 Italian phonetic text

The Italian text-to-speech system uses the Italian subset of 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

The table below lists the phonetic symbols used for the Italian consonants along with example words and their transcriptions.

Table: Italian consonants

Symbol	Word	Phonetic text
b	bimba	b i1 m b a
bb	babbo	b a1 bb o
tS	lancia	l a1 n tS a
ttS	caccia	k a1 ttS a
d	dado	d a1 d o
dd	freddo	f r e1 dd o
dz	manzo	m a1 n dz o
ddz	mezzo	m E1 ddz o
f	stufa	s t u1 f a
ff	stoffa	s t O1 ff a
g	lago	l a1 g o
gg	leggo	l E1 gg o
dZ	agio	a1 dZ o
ddZ	maggio	m a1 ddZ o
L	dirgli	d i1 r L i
LL	aglio	a1 LL o
J	gnocchi	J O1 kk i
JJ	giugno	dZ u1 JJ o
j	pioggia	p j O1 ddZ a
k	cuoco	k w O1 k o
kk	pacco	p a1 kk o
l	pala	p a1 l a
ll	palla	p a1 ll a
m	lama	l a1 m a

Symbol	Word	Phonetic text
mm	mamma	m a1 mm a
n	nona	n O1 n a
nn	nonna	n O1 nn a
p	Papa	p a1 p a
pp	pappa	p a1 pp a
r	faro	f a1 r o
rr	ferro	f E1 rr o
s	sosta	s O1 s t a
ss	rosso	r o1 ss o
S	scienza	S E1 n ts a
SS	ascia	a1 SS a
t	tuta	t u1 t a
tt	tutta	t u1 tt a
ts	forza	f O1 r ts a
tts	pizza	p i1 tts a
v	viva	v i1 v a
vv	avvocato	a vv o k a1 t o
w	uomo	w O1 m o
z	rosa	r O1 z a
N	tronco	t r O1 N k o
M	invito	i M v i1 t o
R	charcot	S a R k o1
RR	Blair	b l e1 i RR

7.2 Vowels

The table below lists the phonetic symbols used for the Italian vowels along with example words and their transcriptions.

Table: Italian vowels

Symbol	Word	Phonetic text
a	sacco	s a1 kk o
e	sera	s e1 r a
E	miele	m j E1 l e
i	sito	s i1 t o
o	solo	s o1 l o
O	suolo	s w O1 l o
u	succo	s u1 kk o

Symbol	Word	Phonetic text
@	Beatles	b i1 t @ l s
e~	Etienne	e t j e~
a~	Mitterrand	m i t e1 R a~
o~	Manon	m a1 n o~

7.3 Lexical accent

A lexical accent is used to indicate the level of prominence (or emphasis) of a syllable in a word. In Italian, all words have a lexical accent, which may be indicated by an accent mark when it does not follow normal accentuation rules. It is therefore important to include stress marks when writing phonetic transcriptions.

In the phonetic transcriptions, the lexical accent is indicated by the symbol 1 placed directly after (no space) the accented vowel.

7.4 Pause

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

8 Abbreviations and Symbols

In the current version of the Italian text-to-speech system, the abbreviations and symbols in the table below are recognized in all contexts. These abbreviations and symbols are mostly case-insensitive and often require no full stop in order to be recognized as an abbreviation.

As previously mentioned, there are also abbreviations for the days of the week and the months (see section *Dates*).

Table: Abbreviations and Symbols

Abbreviation/Symbol	Reading
n°	numero
&	e commerciale
s.ta	santa
s.to	santo
art.	articolo
c.so	corso
cal.	calorie
cap.	capitolo
cfr	confronta
cm	centimetro
ref.	referenza
rep.	repubblica
dott.	dottor
dr	dottor
dr.essa	dottoressa
ecc.	eccetera
etc.	eccetera
ed.	edizione
egr.	egregio
es.	esempio
proff.	professori
fig.	figura
fr.	franchi
gent.ma	gentilissima
gent.mo	gentilissimo
pp.	pagine
prof.	professor
ibid.	ibidem

Abbreviation/Symbol	Reading
ing.	ingegner
Jr	giunior
kg	chilogrammo
km	chilometro
pagg.	pagine
pag.	pagina
p.za	piazza
OK	okei
mm	millimetro
min	minuto/i
Mr	mister
Mrs	missis
sec.	secolo
seg.	seguente
segg.	seguenti
sig.	signor
sig.a	signora
sig.ra	signora
sig.na	signorina
sigg.	signori
str.	strada
v.le	viale
vol.	volume
km/h	chilometri ora
m ³ (eq. 'm ³ ')	metri cubici
m ² (eq. 'm ² ')	metri quadrati
cm ² (eq. 'cm ² ')	centimetri quadrati
cm ³ (eq. 'cm ³ ')	centimetri cubici
km ² (eq. 'km ² ')	chilometri quadrati
Hz	hertz
GHz	gigahertz
dB	decibel
kW	chilowatt

Some symbols are recognized only if followed by digits, as for instance:

2 m = *due metri*

2 s = *due secondi*

2 l = *due litri*

20W = *venti Watt*

15 V= *quindici Volts*

1 q= *un quintale.*

9 Web-addresses and email

Web-addresses and e-mail-addresses are read as follows:

- *www* is read as three *w*'s spelled letter by letter.
- Full stops '.' are read as *punto*, slash '/' as *barra*, dash '-' as *trattino*.
- *us*, *uk*, *fr* and all the other abbreviations for countries are spelled out letter by letter.
- The @ is read *chiocciola*.
- 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.

String

www.acapela-group.com

http://www.acapela-group.com

vittorio@hotmail.it

Reading

W W W punto acapela trattino group punto com

H T T P due punti barra barra W W W punto acapela trattino group punto com

vittorio chiocciola hotmail punto I T