



Language Manual

HQ Canadian French

Language Manual: HQ Canadian French

Published 16 March 2011

Copyright © 2008-2011 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
3.1. Comma, colon and semicolon	3
3.2. Quotation marks	3
3.3. Full stop	3
3.4. Question mark	3
3.5. Exclamation mark	3
3.6. Parentheses, brackets and braces	3
4. Other non alphanumeric characters	4
4.1. Non-punctuation characters	4
4.2. The ² and ³ signs	4
4.3. Symbols whose pronunciation varies depending on the context	5
5. Number Processing	6
5.1. Full number pronunciation	6
5.2. Leading zero	7
5.3. Decimal numbers	7
5.4. Currency amounts	7
5.5. Ordinal numbers	8
5.6. Arithmetic operators	8
5.7. Mixed digits and letters	9
5.8. Time of day	9
5.9. Dates	10
5.10. Phone numbers	10
6. How to change the pronunciation	12
7. Canadian French Phonetic Text	13
7.1. Consonants	13
7.2. Vowels	13
7.3. Pause	14
8. Abbreviations	15
9. Web-addresses and email	17

List of Tables

4.1. Non-punctuation characters	4
7.1. Symbols for the Canadian French consonants	13
7.2. Symbols for the Canadian French vowels	13
8.1. Abbreviations	15

Chapter 1. General

This document discusses certain aspects of text-to-speech processing for the Canadian French text-to-speech systems, 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 Louise.

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 and HD American English voices. 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-Z and a-z, as well as ç and vowels with a diaeresis, grave, acute or circumflex accent, such as é, è, ê, ô, â, î, and ë, may constitute a word. Certain other characters are also considered as letters, notably those used as letters in other European languages, i.e. ñ, ð, å. These letters are not pronounced as in their native languages though, they are pronounced as regular *n*, *o*, *a* etc.

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 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.

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
/	barre oblique
+	plus
\$	dollar
£	livre sterling
€	euro
¥	yen
<	plus petit que
>	plus grand que
%	pour cent
^	accent circonflexe
	barre
~	tilde
@	arobas
²	au carré
³	au cube
=	égal
-	(see below)
*	(see below)

4.2. The ² and ³ signs

The reading of expressions with ² and ³ is:

Expression	Reading
mm ²	millimètres carrés
cm ²	centimètres carrés
m ²	mètres carrés
km ²	kilomètres carrés
mm ³	millimètres cubes
cm ³	centimètres cubes
m ³	mètres cubes
km ³	kilomètres cubes

4.3. Symbols whose pronunciation varies depending on the context

4.3.1. Hyphen

A hyphen '-' is pronounced *moins* 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 *tiret*.

[not SP] In certain date formats, in between days or years, the hyphen is pronounced *au*. In between years it is pronounced *à*. In other cases the hyphen is never pronounced.

Expression	Reading	
-3	moins trois	
44-3	quarante-quatre tiret trois	
44-3=41	quarante-quatre moins trois égal quarante et un	
44 - 3 = 41	quarante-quatre moins trois égal quarante et un	
15-20 octobre	quinze au vingt octobre	[not SP]
6-10 nov	six au dix novembre	[not SP]
1998-2004	mille neuf cent quatre-vingt-dix-huit à deux mille quatre	[not SP]
02-02-2002	deux février deux mille deux	
arc-en-ciel	arc en ciel	

4.3.2. Asterisk

Asterisk '*' is pronounced *fois* if enclosed by digits and followed by equals sign '='. In other cases it is pronounced *astérisque*.

Expression	Reading
2*3	deux astérisque trois
2*3=6	deux fois trois égal six
*bc	astérisque b c

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 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
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 or full stop (not comma). In order to achieve the correct 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.

Number	Reading
2580	deux mille cinq cent quatre-vingts
2 580	"
2.580	"
25800	vingt-cinq mille huit cents
25 800	"
25.800	"
2580350	deux millions cinq cent quatre-vingt mille trois cent cinquante

Number	Reading
2 580 350	“
2.580.350	“
1000000000	un milliard
23 456 789 012	vingt-trois milliards quatre cent cinquante-six millions sept cent quatre-vingt-neuf mille douze
1234567890123	un deux trois quatre cinq six sept huit neuf zéro un deux trois

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	zéro neuf mille deux cent cinquante-trois
020	zéro vingt

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 containing full stop followed by exactly three digits is not read as a decimal number but as a full number, following the rules in the section *Full number pronunciation*.

Number	Reading
16,234	seize virgule deux cent trente-quatre
3,1415	trois virgule un quatre un cinq
1251,04	mille deux cent cinquante et un virgule zéro quatre
1.251,04	mille deux cent cinquante et un virgule zéro quatre
2,50	deux virgule cinquante
2.50	deux point cinquante
3.141	trois mille cent quarante et un

5.4. Currency amounts

The following principles are followed for currency amounts:

- Numbers with zero or two decimals preceded or followed by the currency markers £, \$, ¥ or € are read as currency amounts.
- [not SP] Numbers with zero or two decimals followed by the words *livre*, *dollar*, *yen* or *euro* (singular or plural) are read as currency amounts.
- Accepted decimal markers are comma ',' and full stop '.'.
- The decimal part (consisting of two digits) in currency amounts is read as *et nn pence* and *et nn centimes*.
- If the decimal part is 00 it will not be read.

Example	Reading	
\$15.00	quinze dollars	
15.00£	quinze livres	
15.00 euros	quinze euros	[not SP]
€ 200.50	deux cents euros et cinquante centimes	
1.000.000 ¥	un million de yens	

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

\$ 1 million	un million de dollars
--------------	-----------------------

5.5. Ordinal numbers

Numbers are read as ordinals in the following cases:

- [not SP] The number '1' is followed by a month name or one of the month name abbreviations. The number may be preceded by a day or an abbreviation for a day.

Examples: 1 janvier, 1 jan, mardi 1 jan.

- The number is *1er*, *1ère*, *2nd*, *2nde*.
- The number is followed by *ème*, *ème*, *e*, *è*.

Examples: *5e*, *6ème*, *3eme*, *7è*.

[not SP] Valid abbreviations for months: *jan*, *févr*, *fév*, *avr*, *juil*, *sep*, *sept*, *oct*, *nov* and *déc*.

[not SP] Valid abbreviations for days: *lun*, *mar*, *mer*, *jeu*, *ven*, *sam* and *dim*.

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

5.6. Arithmetic operators

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

Expression	Reading
-12	moins douze
14-2	quatorze tiret deux
14-2=12	quatorze moins deux égal douze
+24	plus vingt-quatre
2+3	deux plus trois
2+3=5	deux plus trois égal cinq
2*3	deux astérisque trois
2*3=6	deux fois trois égal six
2/3	deux tiers
6/2=3	six divisé par deux égal trois
25%	vingt-cinq pour cent
3,4%	trois virgule quatre pour cent

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
77B84Z3	soixante-dix-sept B quatre-vingt-quatre Z trois
0092B87-B	zéro zéro quatre-vingt-douze B quatre-vingt-sept B

5.8. Time of day

The colon is used to separate hours, minutes and seconds. When there are no seconds, *H* or *h* can be used to separate hours and minutes.

[not SP] Abbreviations such as *A.M.* and *P.M.* may follow or precede the time.

Possible patterns are:

- a. *hh:mm* or *h:mm*
- b. *hh:mm:ss* or *h:mm:ss*
- c. [not SP] *hhHmm* or *hHmm*

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

In pattern a:

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

Expression	Reading	
9:00	neuf heures	
9:30 P.M.	neuf heures trente du soir	[not SP]
13:00	treize heures	
12:00	midi	
0:00	minuit	

In pattern b:

An *et* will be inserted before the *ss*-part, and *secondes* will be added after it. If the *ss*-part is equal to *00*, this part will not be read.

Expression	Reading	
10:24:00	dix heures vingt-quatre	
10:24:00 A.M.	dix heures vingt-quatre du matin	[not SP]
10:24:20	dix heures vingt-quatre et vingt secondes	

[not SP] In pattern c:

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

Expression	Reading
12H30	douze heures trente
3h00	trois heures

Format	Reading	Example
xxx-xxx-xxxx	xxx tiret xxx tiret xxxx	800-666-7362
xxx xxx xxxx	xxx xxx xxxx	416 412 5999

5.10.2. International phone numbers

International phone numbers follow the pattern below:

International prefix + Country code + Regional number + Local number.

International prefix:	00 or +
Country code:	1-3 digits
Regional number:	1-3 digits with or without parentheses (see below for exact formats)
Local number:	6-8 digits

Examples:	Reading
(418) 656-2131	418 656 tiret 2131
+ 1-418-656-2131	plus 1 tiret 418 tiret 656 tiret 2131
+ 1 418 656 2131	plus 1 418 656 2131

Chapter 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 *Canadian French Phonetic Text*). Phonetic transcriptions can also be entered directly in the text, using the *PRN* tag (see *User's guide*).

Chapter 7. Canadian French Phonetic Text

The Canadian French text-to-speech system uses the French subset of the SAMPA phonetic alphabet (*Speech Assessment Methods Phonetic Alphabet*), with the exception of the symbol /J/ which was replaced by the sequence /n j/ (ex. *oignon*). 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 7.1. Symbols for the Canadian French consonants

Symbol	Word	Phonetic text	Comment
j	junior	Z y n j O R	glide
w	trois	t R w a	glide
H	huit	H i t	glide
p	papa	p a p a	
t	tante	t a~ t	
k	cacao	k a k a o	
b	bord	b O R	
d	dort	d O R	
g	galette	g a l E t	
f	femme	f a m	
s	sans	s a~	
S	chat	S a	
v	vol	v O l	
z	zéro	z e R o	
Z	jouet	Z u E	
l	long	l o~	
R	rat	R a	
m	mangue	m a~ g	
n	navette	n a v E t	
N	pudding	p u d i N	

7.2. Vowels

Table 7.2. Symbols for the Canadian French vowels

Symbol	Word	Phonetic text	Comment
i	ville	v i l	
e	et	e	
E	abaissais	a b E s E	
E:	abécédaire	a b e s e d E: R	Long E
a	chat	S a	

Symbol	Word	Phonetic text	Comment
O	nord	n O R	
o	gauche	g o S	
u	lourd	l u R	
y	but	b y t	
2	bleu	b l 2	
9	neuf	n 9 f	
@	demain	d @ m e~	
e~	main	m e~	Nasal
a~	grand	g R a~	Nasal
o~	rond	R o~	Nasal
9~	brun	b R 9~	Nasal

7.3. Pause

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

Chapter 8. Abbreviations

In the current version of the Canadian French text-to-speech system, the abbreviations in the table below are recognised in all contexts. These abbreviations are mostly case-insensitive (except for those indicated below by “*”) and require no full stop in order to be recognised as an abbreviation.

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

Table 8.1. Abbreviations

Abbreviation	Reading
kg	kilo
°C	degrés Celsius
°F	degrés Fahrenheit
°K	degrés Kelvin
bd	boulevard
bld	boulevard
bef	Franc belge
cie	compagnie
cm	centimètre
dB*	décibel
DM*	Deutschmark
dm	décimètre
dpt	département
dr	docteur
éd	editeur
etc	et cetera
ff	Franc français
gr	grams
jr	junior
km	kilomètres
kmh	kilomètres par heure
mgr	monseigneur
mlle	mademoiselle
MM*	messieurs
mm	millimètre
mme	madame
mr	monsieur
ms	millisecondes
n°	numéro
nb	nota bene
rte	route
ste	sainte
st	saint

Abbreviation	Reading
sts	saints
tél	téléphone
mt	mont
sr	sénior
BC*	avant Jésus-Christ
ml	millilitre
cl	centilitre
dl	décilitre

Chapter 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 *point*, hyphens '-' as *tiret*, underscores '_' as *underscore*, slashes '/' as *slash*.
- *be*, *uk*, *fr* and all the other abbreviations for countries are spelled out letter by letter.
- The @ is read *arobas*.
- 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

dubois@infonie.ca

jane_dubois@infonie.ca

Reading

w w w point acapela tiret group point com

h t t p deux points slash slash w w w point acapela tiret group point com

dubois arobas infonie point c a

jane souligné dubois arobas infonie point c a