



# *Language Manual*

*HQ, CO, and HD American English*

---

Language Manual: HQ, CO, and HD American English

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 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.....                  | 7  |
| 6 | How to change the pronunciation.....    | 16 |
| 7 | American English phonetic text.....     | 17 |
| 8 | Abbreviations.....                      | 20 |
| 9 | Web-addresses and email .....           | 23 |

## 1 General

This document discusses certain aspects of text-to-speech processing for the American English 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), Colibri (CO), and High Density (HD) American English voices. Note that there are some differences between the voices when it comes to text processing.

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*.

Note2 : Some versions have a disabled preprocessing: preprocessing is reduced to a strict minimum. In each section describing the preprocessing, a note will appear explaining how the format is handled in this version.

## **2 Letters in orthographic text**

Characters from A-Z and a-z may constitute a word. Certain other characters are also considered as letters, notably those used as letters in other European languages, i.e. ñ, ò, å, ç, é. These letters are not pronounced as in their native languages though, they are pronounced as regular *n, o, a, c, e* when occurring in a word

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**

Closing quotes '""' after a group of words (not after a single word between quotes) cause a brief pause 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 [4.1] Non-punctuation characters**

| Symbol       | Reading      |
|--------------|--------------|
| /            | slash        |
| +            | plus         |
| \$           | dollar       |
| £            | pound        |
| €            | euro         |
| ¥            | yen          |
| <            | less than    |
| >            | greater than |
| %            | percent      |
| ^            | circumflex   |
|              | pipe         |
| ~            | tilde        |
| @            | at           |
| =            | equals       |
| <sup>2</sup> | (see below)  |
| <sup>3</sup> | (see below)  |
| -            | (see below)  |
| *            | (see below)  |

### 4.2 The <sup>2</sup> and <sup>3</sup> signs

The reading of expressions with <sup>2</sup> and <sup>3</sup> is:

| Expression      | Reading            |
|-----------------|--------------------|
| mm <sup>2</sup> | square millimeters |
| cm <sup>2</sup> | square centimeters |
| m <sup>2</sup>  | square meters      |
| km <sup>2</sup> | square kilometers  |
| mm <sup>3</sup> | cubic millimeters  |
| cm <sup>3</sup> | cubic centimeters  |

| Expression      | Reading         |
|-----------------|-----------------|
| m <sup>3</sup>  | cubic meters    |
| km <sup>3</sup> | cubic kilometer |

In the disabled preprocessing, each of these is spelled out character by character.

### 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 *dash*.

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

In the disabled preprocessing, the hyphen is pronounced *dash* in all cases.

| Expression       | Reading                                        |          |
|------------------|------------------------------------------------|----------|
| -3               | minus three                                    |          |
| 44-3             | <i>forty-four dash three</i>                   |          |
| 44-3=41          | forty-four minus three equals forty-one        |          |
| 44 - 3 = 41      | forty-four minus three equals forty-one        |          |
| 15-20<br>October | the fifteenth to twentieth of October          | [not SP] |
| 6-10 Nov         | the sixth to tenth of November                 | [not SP] |
| 1998-2004        | nineteen ninety-eight to two thousand and four | [not SP] |
| 02-02-2002       | February second two thousand and two           |          |
| low-income       | low income                                     |          |
| mother-in-law    | mother in law                                  |          |

#### 4.3.2 Asterisk

Asterisk '\*' is pronounced *times* if enclosed by digits and followed by equals sign '='. In other cases it is pronounced *asterisk*.

In the disabled preprocessing, the asterisk is pronounced *asterisk* in all cases.

**Expression**

2\*3

2\*3=6

\*bc

**Reading**

two asterisk three

two times three equals six

asterisk b c

## 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  
Years  
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 comma (not full stop). 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.

In the disabled preprocessing, only digits in a single group are considered as full numbers. Groups of digits separated with space or comma will be seen as so many individual numbers.

| <b>Number</b>  | <b>Reading</b>                                                                                    |
|----------------|---------------------------------------------------------------------------------------------------|
| 2580           | two thousand five hundred and eighty                                                              |
| 2 580          | "                                                                                                 |
| 2,580          | "                                                                                                 |
| 25800          | twenty-five thousand eight hundred                                                                |
| 25 800         | "                                                                                                 |
| 25,800         | "                                                                                                 |
| 2580350        | two million five hundred and eighty thousand three hundred and fifty                              |
| 2 580 350      | "                                                                                                 |
| 2,580,350      | "                                                                                                 |
| 1000000000     | one billion                                                                                       |
| 23 456 789 012 | twenty-three billion four hundred fifty-six million seven hundred eighty-nine thousand and twelve |
| 1234567890123  | one two three four five six seven eight nine zero one two three                                   |

## 5.2 *Leading zero*

---

Numbers that begin with 0 (zero) are read digit by digit.

| <b>Number</b> | <b>Reading</b>           |
|---------------|--------------------------|
| 09253         | zero nine two five three |
| 020           | zero two zero            |

## 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*. The decimals (the part after comma or full stop) are read as separate digits. Note: A number containing a comma followed by exactly three digits is not read as a decimal number but as a full number, following the rules in the section *Full number pronunciation*.

In the disabled preprocessing, decimal numbers are not recognized. When a group of digits is found containing a period or a decimal, the whole group (digits and punctuation mark) will be spelled out.

| <b>Number</b> | <b>Reading</b>                |
|---------------|-------------------------------|
| 16.234        | sixteen point two three four  |
| 3.1415        | three point one four one five |

| Number   | Reading                                                |
|----------|--------------------------------------------------------|
| 1251.04  | one thousand two hundred and fifty-one point zero four |
| 1,251.04 | one thousand two hundred and fifty-one point zero four |
| 2.50     | two point five zero                                    |
| 2,50     | two comma five zero                                    |
| 3,141    | three thousand one hundred and forty-one               |

## 5.4 Currency amounts

---

The following principles are followed for currency amounts:

- Numbers with zero, one or two decimals preceded or followed by the currency markers \$, £, €, ¥, \$CDN, \$CAN, \$C, \$US are read as currency amounts.
- [not SP] Numbers with zero, one or two decimals preceded or followed by the currency markers Eur, DM, USD, AUD, CAD, HKD, SGD, NZD, (CAD), C\$, US\$, (US\$), FF, kr, QAR, QR are read as currency amounts.
- [not SP] Numbers with zero, one or two decimals followed by the words *pound(s)*, *dollar(s)*, *yen* or *euro(s)* (singular or plural) are read as currency amounts.
- Accepted decimal markers are comma ',' and full stop '.'.
- The decimal part (consisting of one or two digits) in currency amounts is read as *and nn pence* and *and nn cents*.
- If the decimal part is 00 it will not be read.

In the disabled preprocessing, currency amounts are not recognized. When a group of digits is found containing a period or a decimal and a currency amount, the whole group (digits, punctuation mark and currency) will be spelled out.

| Example     | Reading                           |          |
|-------------|-----------------------------------|----------|
| \$15.00     | fifteen dollars                   |          |
| 15.00£      | fifteen pounds                    |          |
| 15.00 euros | fifteen euros                     | [not SP] |
| € 200.50    | two hundred euros and fifty cents |          |
| € 30.3      | thirty euros and thirty cents     |          |
| 1,000,000 ¥ | one million yen                   |          |

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

|              |                     |
|--------------|---------------------|
| \$ 1 million | one million dollars |
|--------------|---------------------|

## 5.5 Ordinal numbers

---

Numbers are read as ordinals in the following cases:

- [not SP] 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.
- [not SP] The number consists of a day interval followed by a month name/abbreviation.
- The number is followed by *st*, *nd*, *rd*, *th*, *d*.
- Examples: 1st, 2nd, 3rd, 4th, 23d.

[not SP] The valid abbreviations for months are: *Jan*, *Feb*, *Mar*, *Apr*, *Jun*, *Jul*, *Aug*, *Sept*, *Oct*, *Nov*, and *Dec*.

[not SP] The valid abbreviations for days are: *Mon*, *Tue*, *Wed*, *Thu*, *Thurs*, *Fri*, *Sat*, and *Sun*.

[not SP] The abbreviations above are only expanded to names of months and days when appearing in correct date contexts.

In the disabled preprocessing, ordinal numbers are not recognized. The number will be pronounced, and the ordinal part will be spelled out.

| Expression    | Reading                           |
|---------------|-----------------------------------|
| 3 January     | third of January                  |
| 3 Jan         | third of January                  |
| Tuesday 3 Jan | Tuesday third of January          |
| 15-16 January | fifteenth to sixteenth of January |
| 2nd May       | second May                        |
| 4th Jun 2007  | fourth June 2007                  |

## 5.6 Arithmetic operators

---

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

In the disabled preprocessing, arithmetic operators are not recognized. The numbers will be pronounced, and symbols will be spelled out.

| Expression | Reading                          |
|------------|----------------------------------|
| -12        | minus twelve                     |
| 14-2       | fourteen dash two                |
| 14-2=12    | fourteen minus two equals twelve |
| +24        | plus twenty-four                 |

**Expression**

2+3  
 2+3=5  
 2\*3  
 2\*3=6  
 2/3  
 6/2=3  
 25%  
 3.4%

**Reading**

two plus three  
 two plus three equals five  
 two asterisk three  
 two times three equals six  
 two thirds  
 six divided by two equals three  
 twenty-five percent  
 three point four percent

## 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 numbers are read according to the examples below.

**Expression**

77B184Z3  
 0092B87-B  
 FT2892B87Z  
 TN12345L5

**Reading**

seventy-seven B one hundred and eighty-four  
 Z three  
 zero zero ninety-two B eighty-seven B  
 F T twenty-eight ninety-two B eighty-seven Z  
 T N one two three four five L five

## 5.8 Time of day

---

The colon is used to separate hours, minutes and seconds.

[not SP] Abbreviations such as *A.M.* and *P.M.* (possible variants: *a.m.*, *am*, *AM*, *p.m.*, *pm*, *PM*) may follow or precede the time, with a space inserted between the time and the abbreviation. (Exception: when an hour is alone, the abbreviation must follow it. Example: 3 *AM* and not *AM* 3)

In the disabled preprocessing, time of day is not recognized. The numbers will be pronounced, and symbols will be spelled out.

**Example**

9 A.M.  
 4 pm

Possible patterns are:

- a. hh:mm or h:mm
- b. hh:mm:ss or h:mm:ss
- c. [not SP] hh:mm:ss" or h:mm:ss"

Example: 12:30'45"

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

In pattern a:

If the *mm*-part is equal to *00*, this part will not be read. Instead, *o'clock* will be added if the hours are less than 13, or *hundred hours* will be added if the hours are greater than or equal to 13.

| Expression | Reading                |
|------------|------------------------|
| 9:00       | nine o'clock           |
| 13:00      | thirteen hundred hours |
| 12:00      | noon                   |
| 0:00       | midnight               |

In pattern b:

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

| Expression | Reading                            |
|------------|------------------------------------|
| 10:24:00   | ten twenty-four                    |
| 10:24:20   | ten twenty-four and twenty seconds |

[not SP] In pattern c:

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

## 5.9 Years

---

Numbers between 1000 and 1999 are always read as hundreds (year reading) with the exception of numbers containing decimals. Years (2 or 4 digits) can also be followed by 's to indicate decades.

In the disabled preprocessing, years are handled in the same way, but for the hyphen which is pronounced *dash*.

| Expression     | Reading                                                    |          |
|----------------|------------------------------------------------------------|----------|
| 1988           | nineteen eighty-eight                                      |          |
| 1939-45        | nineteen thirty-nine to forty-five                         | [not SP] |
| 1088           | one thousand eighty-eight                                  |          |
| 1988.0         | one thousand nine hundred and eighty-eight point zero      |          |
| 1988.32        | one thousand nine hundred and eighty-eight point three two |          |
| September 1999 | September nineteen ninety-nine                             |          |
| 70's           | seventies                                                  |          |
| 1980's         | nineteen eighties                                          |          |

## 5.10 Dates

---

The valid formats for dates are:

1. mm-dd-yyyy, mm.dd.yyyy, and mm/dd/yyyy
2. mm-dd-yy, mm.dd.yy, and mm/dd/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:                 | Reading                           |
|-------------------------|-----------------------------------|
| 10-02-2003 or 10-2-2003 | October second two thousand three |
| 10.02.2003 or 10.2.2003 | "                                 |
| 10/02/2003 or 10/2/2003 | "                                 |

| Type 2:             | Reading                           |
|---------------------|-----------------------------------|
| 10-02-03 or 10-2-03 | October second two thousand three |
| 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     | nineteen ninety-eight to nineteen ninety-nine |
| 1939-45       | nineteen thirty-nine to forty-five            |
| 2002/3        | two thousand two to three                     |
| 14-15 January | fourteenth to fifteenth of January            |
| October 19-20 | October nineteenth to twentieth               |

[not SP] Other possible formats include:

| Expression         | Comment                                 |
|--------------------|-----------------------------------------|
| Monday, 15 January | Monday fifteenth of January             |
| Monday 15 January  | Monday fifteenth of January             |
| Mon, January 15    | Monday January fifteenth                |
| Mon January 15     | Monday January fifteenth                |
| 30 April 1999      | thirtieth of April nineteen ninety-nine |
| April 30 1999      | April thirtieth nineteen ninety-nine    |
| May 1953           | May nineteen fifty-three                |
| 3 May              | third of May                            |

In the disabled preprocessing, date formats are not recognized. The numbers will be pronounced, words will be read as they are written.

Hyphens, dots and slashes will be read as normal characters, comas are considered as punctuation marks.

## ***5.11 Phone numbers***

---

In this section the patterns of digits that are recognized as phone numbers are described. In the pronunciation of phone numbers, all numbers are read out digit by digit with pauses between groups of numbers.

No phone numbers are recognized in the Simplified Processing voices.

### ***5.11.1 Ordinary phone numbers***

Sequences of digits in the following formats are treated as phone numbers.

The following sequences of digits can be separated by a space, a period, or a hyphen:

#### **Format**

xxx xx xx xx  
xxx xxxx  
xx (xx) xxx xx xx  
(xx) xx xx xx xx xx  
xx (x)x xx xx xx xx  
xx (x) x xx xx xx xx  
xx x xx xx xx xx

The following sequences can only appear in these formats:

#### **Format**

(xx)-xxxx-xxx-xxx  
(xx).xxxx.xxx.xxx  
xx xxx xx xx  
x-xxx-xxx-xxxx

Other formats are preceded by an area code that can consist of 1-3 numbers, either surrounded by parentheses or not. The groups of digits can be separated by a space, slash, hyphen or period. If the area code is surrounded by parentheses, it can be grouped with the other digits.

#### **Format**

(area code) xxx xxxx  
(area code) xxx xxx

In the disabled preprocessing, phone numbers are not recognized. The numbers will be pronounced, and symbols will be spelled out.

### ***5.11.2 International phone numbers***

International phone numbers follow the pattern below:

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

|                       |             |
|-----------------------|-------------|
| International prefix: | 00 or +     |
| Country code:         | 1-3 digits  |
| Local number:         | 6-12 digits |

All formats included above can be preceded by an international prefix and a country code.

**Examples**

001-12-456-7894

001-12 578 21 56

0032 (71).4521.521.843

In the disabled preprocessing, international phone numbers are not recognized. The numbers will be pronounced, and symbols will be spelled out.

## **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 *American English phonetic text*). Phonetic transcriptions can also be entered directly in the text, using the *PRN* tag (see *User's guide*).

## 7 American English phonetic text

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 American English text-to-speech system uses the American English subset of the SAMPA phonetic alphabet (*Speech Assessment Methods Phonetic Alphabet*), with a few exceptions. The symbol /o/ was replaced by /@U/ (ex. *nose*), /e/ was replaced by /EI/ (ex. *hate*), /aI/ was replaced by /AI/ (ex. *light*), and /3`/ and /@`/ were replaced by /r=/ (ex. *furs*, *corner*). The symbol /4/ was also introduced to represent a flapped *t* (ex. *better*). The symbols are written with a space between each phoneme.

### 7.1 Consonants

**Table [7.1] Symbols for the American English consonants**

| Symbol | Word     | Phonetic text  | Comment                                     |
|--------|----------|----------------|---------------------------------------------|
| b      | bad      | b {1 d         |                                             |
| t      | tsunami  | t s u n A1 m i |                                             |
| t_h    | tab      | t_h {1 b       | aspirated t                                 |
| 4      | better   | b E1 4 r=      | flapped t                                   |
| ?      | fountain | f aU1 n ? n=   | unreleased t<br>(followed by /@/<br>or /n=) |
| p      | stop     | s t O1 p       |                                             |
| p_h    | pine     | p_h AI1 n      | aspirated p                                 |
| d      | date     | d EI1 t        |                                             |
| k      | scare    | s k E1 r       |                                             |
| k_h    | cone     | k_h @U1 n      | aspirated k                                 |
| g      | gag      | g {1 g         |                                             |
| m      | man      | m {1 n         |                                             |
| n      | nose     | n @U1 z        |                                             |
| r      | rose     | r @U1 z        |                                             |
| l      | let      | l E1 t         |                                             |
| N      | ring     | r I1 N         |                                             |
| f      | fat      | f {1 t         |                                             |
| v      | vote     | v @U1 t        |                                             |
| s      | sat      | s {1 t         |                                             |
| z      | zoo      | z u1           |                                             |
| S      | shin     | S I1 n         |                                             |
| tS     | chin     | tS I1 n        |                                             |
| Z      | measure  | m E1 Z r=      |                                             |

| Symbol | Word  | Phonetic text | Comment |
|--------|-------|---------------|---------|
| dʒ     | gin   | dʒ ɪ n        |         |
| D      | this  | D ɪ s         |         |
| T      | thin  | T ɪ n         |         |
| w      | wait  | w eɪ t        | glide   |
| j      | yacht | j Aɪ t        | glide   |
| h      | hit   | h ɪ t         |         |

## 7.2 Vowels

**Table [7.2] Symbols for the American English vowels**

| Symbol | Word              | Phonetic text               | Comment |
|--------|-------------------|-----------------------------|---------|
| r=     | corner            | k_h Oɪ r n r=               |         |
| A      | father            | f Aɪ D r=                   |         |
| O      | thought           | T Oɪ t                      |         |
| I      | lit               | l ɪ t                       |         |
| i      | neat              | n i t                       |         |
| u      | zoo               | z u                         |         |
| V      | hut               | h Vɪ t                      |         |
| U      | put               | p_h Uɪ t                    |         |
| {      | pat               | p_h {ɪ t                    |         |
| E      | net               | n Eɪ t                      |         |
| @      | allow             | @ l aUɪ                     |         |
| Eɪ     | main              | m Eɪ n                      |         |
| Aɪ     | high              | h Aɪ                        |         |
| Oɪ     | boy               | b Oɪ                        |         |
| @U     | nose              | n @Uɪ z                     |         |
| aU     | pout              | p_h aUɪ t                   |         |
| l=     | saddle<br>crystal | s {ɪ d l=<br>k_h r ɪ s t l= |         |
| n=     | sudden<br>potent  | s Vɪ d n=<br>p_h @Uɪ t n= t |         |

## 7.3 Lexical stress

In words with more than one syllable, one (and normally only 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 the symbol marking the word stress.

In the phonetic transcriptions the word stress is indicated by the symbol /1/ placed directly after the stressed vowel (with no space between the vowel symbol and the stress symbol).

#### **7.4 Glottal stop**

---

A glottal stop, represented by the phonetic symbol /ʔ/, is a small sound which is often used to separate two words when the second word starts with a stressed vowel. This sound can be inserted in a phonetic transcription in order to improve the pronunciation.

#### **7.5 Pause**

---

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

## 8 Abbreviations

In the current version of the American English 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 “\*”) and 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 chapter *Ordinal numbers*).

In the disabled preprocessing, most abbreviations are no longer recognized. Only the most frequent, non-ambiguous ones are recognized, such as “mr, mrs, ms, blvd, gov”. Metric units and state abbreviations are no longer recognized.

**Table [8] Abbreviations**

| <b>Abbreviation</b> | <b>Reading</b>                                        |
|---------------------|-------------------------------------------------------|
| kg                  | kilograms                                             |
| °C                  | degrees Celsius                                       |
| °F                  | degrees Fahrenheit                                    |
| °K                  | degrees Kelvin                                        |
| asap                | A S A P                                               |
| b/f                 | before                                                |
| blvd                | boulevard                                             |
| cm                  | centimeters                                           |
| corp                | corporation                                           |
| dB*                 | decibel                                               |
| DM*                 | Deutschmark (only expanded after numbers)<br>[not SP] |
| eg                  | for example                                           |
| etc                 | et cetera                                             |
| ft                  | foot                                                  |
| gal                 | gallon (only expanded after numbers)                  |
| gov.                | governor                                              |
| hr                  | hour                                                  |
| hrs                 | hours                                                 |
| ie                  | that is                                               |
| jr                  | junior                                                |
| km                  | kilometers                                            |
| Km/h                | kilometers per hour                                   |
| mg                  | milligrams                                            |
| ml                  | milliliters                                           |
| mm                  | millimeters                                           |
| mph                 | miles per hour                                        |

| Abbreviation | Reading                                                        |
|--------------|----------------------------------------------------------------|
| mr           | mister                                                         |
| mrs          | missis                                                         |
| ms           | miss                                                           |
| mt           | mount                                                          |
| prof         | professor                                                      |
| sgt          | sergeant                                                       |
| sr           | senior                                                         |
| tsp          | teaspoons                                                      |
| vs           | versus                                                         |
| gen          | general                                                        |
| ltd          | limited                                                        |
| dept         | department                                                     |
| ct           | court                                                          |
| rd           | road                                                           |
| av           | avenue                                                         |
| ctrl         | control                                                        |
| lb           | pounds                                                         |
| sec          | seconds (Note: SEC is another abbreviation, it is spelled out) |

Some abbreviations are expanded differently depending on their position in the sentence. For example, *dr* and *st* are expanded into *drive* and *street* if they appear after a capitalized noun. They are expanded into *doctor* and *saint* when they appear before a capitalized noun.

| Example     | Reading       |
|-------------|---------------|
| Main st.    | Main street   |
| St John.    | Saint John    |
| Bayview dr. | Bayview drive |
| Dr. Jones.  | Doctor Jones  |

*m*, *g* and *in*. are expanded only when appearing after a number.

| Example | Reading            |                                        |
|---------|--------------------|----------------------------------------|
| 25 m    | twenty-five meters |                                        |
| 30 in.  | thirty inches      | note that the period is mandatory here |
| 45 g    | forty-five grams   |                                        |

The following state abbreviations are recognized in the American English system (because non-ambiguous):

|    |         |    |               |
|----|---------|----|---------------|
| AK | Alaska  | NE | Nebraska      |
| AL | Alabama | NH | New Hampshire |
| AZ | Arizona | NJ | New Jersey    |

|    |                |    |                |
|----|----------------|----|----------------|
| AR | Arkansas       | NM | New Mexico     |
| CA | California     | NY | New York       |
| CT | Connecticut    | PA | Pennsylvania   |
| DE | Delaware       | RI | Rhode Island   |
| FL | Florida        | SC | South Carolina |
| IL | Illinois       | TN | Tennessee      |
| KS | Kansas         | TX | Texas          |
| KY | Kentucky       | VT | Vermont        |
| MN | Minnesota      | WA | Washington     |
| MO | Missouri       | WI | Wisconsin      |
| MT | Montana        | WV | West Virginia  |
| NC | North Carolina | WY | Wyoming        |
| ND | North Dakota   |    |                |

Note that the other state abbreviations are ambiguous and are only read as states when found in addresses with postal codes.

Legitimate postal code formats are:

State abbreviation followed by 6 digits

Ex: MD 123456

State abbreviation followed by 5 digits

Ex: NM 12345

State abbreviation followed by 5 digits, a hyphen, and 4 digits

Ex: OR 12345-4587

## 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 *dot*, hyphens '-' as *dash*, underscore '\_' as *underscore*, slash '/' as *slash*.
- *us*, *uk*, *fr* and all the other abbreviations for countries are spelled out letter by letter.
- The @ is read *at*.
- 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

smith@yahoo.us

jane\_smith@yahoo.us

### Reading

w w w dot acapela dash group dot com

h t t p colon slash slash w w w dot acapela dash group dot com

smith at yahoo dot u s

jane underscore smith at yahoo dot u s