
Language Manual

# Finnish

Sanna

Language Manual
Finnish
Sanna
July 2007

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

Acapela Group
Box 1328
SE-171 26 Solna
Sweden

Phone +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 Colombier
ZAC de l'Hers
F-31670 Labege
France

Tel: +33 (0)5 62 24 71 00
Fax: +33 (0)5 62 24 71 01

www.acapela-group.com

# List of contents

# 1 General

This document discusses certain aspects of text-to-speech processing for the Finnish 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 voice Sanna.

# 2 Letters in orthographic text

Characters from **A-Ö** and **a-ö** may constitute a word, along with charcters **Š,š,Ž** and **ž**. Certain other characters are also considered as letters, notably those used as letters in other European languages, i.e. “ñ, ç, é”. These letters are not pronounced as in their native languages though, they are pronounced as regular “n, c, e” etc.

Characters outside of these ranges, i.e. numbers, punctuation characters and other non-alphanumeric characters are not considered as letters.

In most cases, the reading of orthographic words is case-insensitive, in other words small and capital letters are read identically. For instance, “Matkusta Helsinkiin.” and “maTKUsta helsinKIIn.” are read identically.

## *2.1 Ü and Scandinavian letters*

The special letters of Scandinavian languages and the letter “ü” are read as indicated in Table 1.

| Character | Pronunciation in word | Spelling |
|---|---|---|
| å | *o* | ruotsalainen o |
| Å | *o* | ruotsalainen o |
| æ | ä | tanskalainen ä |
| Æ | ä | tanskalainen ä |
| ø | ö | tanskalainen ö |
| Ø | ö | tanskalainen ö |
| ü | y | saksalainen y |
| Ű | y | saksalainen y |

**Table 1 Scandinavian letters and letter ü in the Finnish system**

# 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.
Colon can also be used to attach case endings to numbers, abbreviations and symbols.

## *3.2 Quotation marks*

Quotes < **“ ”** > or < **" "** >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 5) and in abbreviations (see chapter 8).

## *3.4 Question mark*

A question mark <**?**> ends a sentence and causes a “question intonation”, which is most often a falling intonation pattern, followed by a pause.

## *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 < **( )** > 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.

| Symbol | Reading |
|---|---|
| / | vinoviiva |
| \ | kenoviiva |
| + | plus |
| $ | dollari |
| ¢ | sentti |
| £ | punta |
| € | euro |
| ¥ | jeni |
| < | pienempi kuin |
| > | suurempi kuin |
| % | prosentti |
| ^ | sirkumfleksi |
| \| | pystyviiva |
| ~ | tilde |
| @ | ät-merkki |
| = | yhtäsuuruusmerkki |
| ² | yläindeksi kaksi (and see below |
| ³ | yläindeksi kolme (and see below) |
| - | see below |
| * | see below |

**Table 2 Non-punctuation characters**

## 4.2 The ² and ³ signs

The reading of expressions with ² and ³ is:

| Expression | Reading | Reading with number |
|---|---|---|
| mm² | neliömillimetri | neliömillimetriä |
| cm² | neliösenttimetri | neliösenttimetriä |
| m² | neliömetri | neliömetriä |
| km² | neliökilometri | neliökilometriä |
| | | |
| mm³ | kuutiomillimetri | kuutiomillimetriä |
| cm³ | kuutiosenttimetri | kuutiosenttimetriä |
| m³ | kuutiometri | kuutiometriä |
| km³ | kuutiokilometri | kuutiokilometriä |

Case endings can be attached to these abbreviations with a colon.

| Expression | Reading |
|---|---|
| 2 m³:n | kahden kuutiometrin |

## *4.3 Symbols whose pronunciation varies depending on the context*

### 4.3.1 Hyphen

A hyphen < **-** > is pronounced “miinus” in two cases:

a) if followed by a digit and no other digit is found in front of the hyphen

b) if followed by a digit and an equals sign. If there is no equals sign, it is pronounced “viiva”.

In certain date formats, in between days or years, the hyphen is pronounced “viiva”.
In compounds or between words, the hyphen is not pronounced. Examples: ala-aste, La Plata-joki

| **Expression** | **Reading** |
|---|---|
| -3 | miinus kolme |
| 44-3 | neljäkymmentäneljä viiva kolme |
| 44-3=41 | neljäkymmentäneljä miinus kolme on yhtä kuin neljäkymmentäyksi |
| 1998-2004 | tuhatyhdeksänsataayhdeksänkymmentäkahdeksan viiva kaksituhattaneljä |
| 2.2.2002 - 3.2.2002 | toinen toista kaksituhattakaksi viiva kolmas toista kaksituhattakaksi |
| 2-2-2002 | toinen toista kaksituhattakaksi |

### 4.3.2 Asterisk

Asterisk < ***** > is only pronounced as “kertaa” if enclosed by digits and followed by an equals sign. In other cases it is pronounced as “asteriski”.

| **Expression** | **Reading** |
|---|---|
| 2*3 | kaksi asteriski kolme |
| 2*3=6 | kaksi kertaa kolme on yhtä kuin kuusi |
| *bc | asteriski 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, preceded by a section concerning the inflection of numbers and abbreviations linked to numbers:

Full number pronunciation
Leading zero
Decimal numbers
Currency amounts
Ordinal numbers
Arithmetic operators
Mixed digits and letters
Time of day
Year
Dates
Phone numbers

## *5.1 The Inflection of numbers and abbreviations*

The Finnish text-to-speech system includes a facility for inflecting numbers and abbreviations linked to numbers and currencies. The case ending must be separated from the number or the abbreviation by a colon. The system inflects the whole combination according to the case indicated by the ending. The system supports both cardinal and ordinal numbers in all singular and plural cases except for the instructive case.

The ending recognized for the various cases are:

| **Case** | **Endings** | **Example** | **Reading** |
|---|---|---|---|
| **Singular:** | | | |
| Genitive | :n | 6:n | kuuden |
| Inessive | :ssa/:ssa | 6:ssa | kuudessa |
| Elative | :sta/:stä | 6:sta | kuudesta |
| Illative | :een | 6:een | kuuteen |
| Adessive | :lla/:llä | 6:lla | kuudella |
| Ablative | :lta/:ltä | 6:lta | kuudelta |
| Allative | :lle | 6:lle | kuudelle |
| Partitive | :a | 6:a | kuutta |
| Essive | :na/:nä | 6:na | kuutena |
| Translative | :ksi | 6:ksi | kuudeksi |
| Comitative | :ne/:ine | 6:ne | kuusine |
| Abessive | :tta/:ttä | 6:tta | kuudetta |
| | | | |
| **Plural:** | | | |
| Genitive | :ien/:den/:tten | 6:ien/6:den | kuusien |
| Inessive | :issa/:issa | 6:issa | kuusissa |
| Elative | :ista/:istä | 6:ista | kuusista |
| Illative | :iin | 6:iin | kuusiin |
| Adessive | :illa/:illä | 6:illa | kuusilla |
| Ablative | :ilta/:iltä | 6:ilta | kuusilta |
| Allative | :ille | 6:ille | kuusille |
| Partitive | :ia | 6:ia | kuusia |
| Essive | :ina/:inä | 6:ina | kuusina |
| Translative | :iksi | 6:iksi | kuusiksi |
| Comitative | :ne/:ine | 6:ine | kuusine |
| Abessive | :itta/:ittä | 6:itta | kuusitta |

**Examples**

| **Combination** | **Reading** |
|---|---|
| 6:ssa | kuudessa |
| 25:sta | kahdestakymmenestäviidestä |
| 125:ssä | sadassakahdessakymmenessäviidessä |
| 3 000:ssa | kolmessatuhannessa |
| 15:issä | viisissätoista |
| 4:nnessä | neljännessä |
| 12:nsia | kahdensiatoista |
| 10 €:lla | kymmenellä eurolla |
| 7 %:sta | seitsemästä prosentista |

## *5.2 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 | kaksituhattaviisisataakahdeksankymmentä |
| 2 580 | --"-- |
| 2.580 | --"-- |
| 25800 | kaksikymmentäviisituhattakahdeksansataa |
| 25 800 | --"-- |
| 25.800 | --"-- |
| 2580350 | kaksimiljoonaa viisisataakahdeksankymmentätuhatta kolmesataaviisikymmentä |
| 2 580 350 | --"-- |
| 2.580.350 | --"-- |
| 1000000000 | miljardi |
| 123456789012 | satakaksikymmentäkolmemiljardia neljäsataaviisikymmentäkuusimiljoonaa seitsemänsataakahdeksankymmentäyhdeksäntuhatta kaksitoista |
| 23 456 789 012 | kaksikymmentäkolmemiljardia neljäsataaviisikymmentäkuusimiljoonaa seitsemänsataakahdeksankymmentäyhdeksäntuhatta kaksitoista |

## *5.3 Leading zero*

Numbers that begin with 0 (zero) are read as separate digits.

| **Number** | **Reading** |
|---|---|
| 09253 | nolla yhdeksän kaksi viisi kolme |
| 020 | nolla kaksi nolla |

## *5.4 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 5.1. If the decimal (the part after comma or full stop) has more than three places, the decimal part is read as separate digits, otherwise it is read as a full number. Note: A number containing a full stop followed by exactly three digits is not read as a decimal number but as a full number, following the rules in 5.1.

| **Number** | **Reading** |
|---|---|
| 2,50 | kaksi pilkku viisikymmentä |
| 2.50 | kaksi piste viisikymmentä |
| 3,141 | kolme pilkku sataneljäkymmentäyksi |
| 3,1411 | kolme pilkku yksi neljä yksi yksi |

## *5.5 Monetary amounts*

The following principles are followed for monetary amounts:

- Numbers with zero or two decimal places preceded or followed by the currency markers **£, $, ¥**, **e, E** or **€** are read as monetary amounts.
- Numbers with zero or two decimal places preceded or followed by the words “puntaa”, “dollaria”, “euroa”, “jeni”, “markkaa”, “GBP”, “NOK”, “SEK”, “USD”, “FIM”, “mk”, “kr”, “EKK” or “DKK” (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 “ja xx senttiä”, “ja xx äyriä” or “ja xx penniä”.
- If the full number part is “0” then it will not be read
- If the decimal part is “00” it will not be read.

| **Example** | **Reading** |
|---|---|
| $15.00 | viisitoista dollaria |
| 15 euroa | viisitoista euroa |
| 15,50 € | viisitoista euroa ja viisikymmentä senttiä |
| euroa 15,50 | viisitoista euroa ja viisikymmentä senttiä |
| € 200.50 | kaksisataa euroa ja viisikymmentä senttiä |
| 15,00 €:sta | viidestätoista eurosta |
| 0,50 $ | viisikymmentä senttiä |
| 1.000.000 € | miljoona euroa |
| 1 000 000 E:sta | miljoonasta eurosta |

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

| | |
|---|---|
| € 1 miljoona | miljoona euroa |

## *5.6 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 appears with a suffix for ordinal numbers followed, in inflected cases, by the appropriate case ending. The suffix for ordinal number is **:s**, **:nen**, **:nne** or **:nte** as appropriate. Examples: 6:s (“kuudes”), 6:sissa (“kuusissa”), 6:nnessa (“kuudennessa”), 8:ntena (“kahdeksantena”), 2:nen or 2:s (“toinen”), 1:nen or 1:s (“ensimmäinen”), 1:sestä (“ensimmäisestä”).
- The number is followed by a full stop and a month name or month name abbreviation, and the number is smaller or equal to 31. The number may be preceded by a day or an abbreviation for a day. Examples:

  tiistai, 10. tammikuuta 2002
  ti., 10. tammik. 2002

- The number represents a day in a date format. Example: 15.3.2007
- The number is in a range of days. Example: 14. – 15.2.2005

**Valid abbreviations for months:** tammik., helmik., maalisk., huhtik., toukok., kesäk., heinäk., elok., syysk., lokak., marrask., jouluk., tam., hel., maa., huh., tou., kes., hei., elo., syy., lok., mar., jou.

**Valid abbreviations for days:** maan., tiist., kesk., torst., perj., lauant., sunn., ma., ti., ke., to., pe., la., su.

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

## *5.7 Arithmetic operators*

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

| **Expression** | **Reading** |
|---|---|
| -12 | miinus kaksitoista |
| +24 | plus kaksikymmentäneljä |
| 2*3 | kaksi asteriski kolme |
| 2*3=6 | kaksi kertaa kolme on yhtä kuin kuusi |
| 2/3 | kaksi kolmasosaa |
| 25% | kaksikymmentäviisi prosenttia |
| 3.4% | kolme piste neljä prosenttia |
| 3,4% | kolme pilkku neljä prosenttia |

## *5.8 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.

Examples:

| **Expression** | **Reading** |
|---|---|
| 77B84Z3 | seitsmänkymmentä seitsemän B kahdeksankymmentäneljä Z kolme |
| 0092B87-B | nolla nolla yhdeksän kaksi B kahdeksankymmentäseitsemän B |

## *5.9 Time of day*

The colon and full stop are used to separate hours, minutes and seconds. When the full stop is used, the format must be preceded by “klo”, “kello” or “kello on” to be recognized as a time format.

Examples: klo 15:00 kello viisitoista nolla nolla

klo 15.00 kello viisitoista nolla nolla

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 ss s** **or** **hh t mm min ss s**

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

| | | |
|---|---|---|
| Example: | klo 9.30 | kello yhdeksän kolmekymmentä |
| | klo 9.30.20 | kello yhdeksän kolmekymmentä ja kaksikymmentä |
| | 9:00 | yhdeksän nolla nolla |
| | 13:00 | kolmetoista nolla nolla |
| | 12 h 30 min 50 s | kaksitoista tuntia kolmekymmentä minuuttia ja viisikymmentä sekuntia |
| | 12 t 30 min 50 s | kaksitoista tuntia kolmekymmentä minuuttia ja viisikymmentä sekuntia |

In pattern b): "ja" will be inserted before the "ss"-part.

In pattern c): "ja" will be inserted before the "ss"-part, and the words "tuntia", "minuuttia" and "sekuntia" will be added in the appropriate places.

## *5.10 Dates*

The valid formats for dates are:

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

"**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: | dd-mm-yyyy, dd.mm.yyyy, and dd/mm/yyyy | | |
| 10-02-2003 | or | 10-2-2003 | kymmenes toista kaksituhattakolme |
| 10.02.2003 | or | 10.2.2003 | " |
| 10/02/2003 | or | 10/2/2003 | " |
| | | | |
| Type 2: | dd-mm-yy, dd.mm.yy, and dd/mm/yy | | |
| 10-02-03 | or | 10-2-03 | kymmenes toista nolla kolme |
| 10.02.03 | or | 10.2.03 | " |
| 10/02/03 | or | 10/2/03 | " |

Ranges of years and dates are also supported.

**Examples:**

| | |
|---|---|
| 1998-1999 | tuhatyhdeksänsataayhdeksänkymmentäkahdeksan viiva tuhatyhdeksänsataayhdeksänkymmentäyhdeksän |

| | |
|---|---|
| 1939-45 | tuhatyhdeksänsataakolmekymmentäyhdeksän viiva neljäkymmentäviisi |
| 14.2.2005 - 15.3.2005 | neljästoistatoista kaksituhattaviisi viiva viidestoistakolmatta kaksituhattaviisi |
| 2005-2-14/15 | neljästoista viiva viidestoistatoista kaksituhattaviisi |
| 2005-2-14/2-15 | neljästoistatoista viiva viidestoistatoista kaksituhattaviisi |
| 2005-2-14/2005-2-15 | neljästoistatoista kaksituhattaviisi viiva viidestoistatoista kaksituhattaviisi |
| 14. – 15.2.2005 | neljästoista viiva viidestoistatoista kaksituhattaviisi |

Other possible formats include:

| | |
|---|---|
| maanantai, 15. tammikuuta | maanantai, viidestoista tammikuuta (with or without the comma) |
| ma, 15. tammik. | maanantai, viidestoista tammikuuta |
| 30. huhtikuuta 1999 | kolmaskymmenes huhtikuuta tuhatyhdeksänsataayhdeksänkymmentäyhdeksän |
| huhtikuun 30. 1999 | huhtikuun kolmaskymmenes tuhatyhdeksänsataayhdeksänkymmentäyhdeksän |
| 3. toukok. | kolmas toukokuuta |
| 3. toukok. 2007 | kolmas toukokuuta kaksituhattaseitsemän |

## *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 a pause between the groups.

### 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, hyphen or slash (/) :

- (area) xxxx (09) 4568, 016 4568, 0400 4568
- (area) xx xxx 09 45 688, 016 45 688, 0400 45 688
- (area) xxx xx 09 456 88, 016 456 88, 0400 456 88
- (area) xxx xxxx 09 456 8899, 016 456 8899, 0400 456 8899
- (area) xxxx xxx 09 4563 889, 016 4563 889, 0400 4563 889
- (area) xxxx xxxx 09 4563 8891, 016 4563 8891, 0400 4563 8891
- (area) xx xxxx 09 63 8891, 016 63 8891, 0400 63 8891
- (area) xxxxx 09 21235, 016 21235, 0400 21235
- (area) xxxxxx 09 212358, 016 212358, 0400 212358
- (area) xxxxxxx 09 2123586, 016 2123586, 0400 2123586
- (area) xxx xxx (09) 123 456, (016) 123 456, (0400) 123 456 (here parentheses are
- obligatory)

(The area code consists of a “0” followed by 1 to 3 digits. The area code can be written inside parentheses.)

The following sequences can only appear in these formats:

| | |
|---|---|
| xx xx xx xx | 09-22 83 21 |
| xx xxx xx xxx | 09-681 78 450 |
| xx xxx xxx xx | 09-131 552 22 |
| xx xxx xxxxx | 09 191 57745 |
| xxxx xxx | 0204 911 |

The abbreviations “p.” and “puh.” (*puhelin*), “f.” (*faksi*), “matkap.” and “matkapuh.” (*matkapuhelin*), “k.” (*koti*), “kotipuh.” (*kotipuhelin*), “t.” (*työ*) and “työpuh.” (*työpuhelin*) are expanded if they occur in front of a phone number.

| | | |
|---|---|---|
| e.g. | puh. 016 231 645 | puhelin 016 231 645 |
| | f. 016 231 645 | faksi 016 231 645 |

### 5.11.2 International phone numbers

All the preceding formats can be recognised if preceded by an international prefix, with the “0” of the area code removed:

| Format | Example |
|---|---|
| +x | +9 9 45 688 |
| +xx | +47 16 45 688 |
| +xxx | +358 9 45 688 |

# 6 How to change pronunciation errors

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 7). Phonetic transcriptions can also be entered directly in the text, using a PRN-tag (see User's guide).

# 7 Finnish Phonetic Text

The Finnish 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 SAMPA 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*

### 7.1.1 Symbols for the Finnish consonants

| Symbol | Word | Phonetic text | Comment |
|---|---|---|---|
| b | bussi | b u1 ss i | |
| t | talo | t A1 l O | |
| p | peli | p e1 l i | |
| d | data | d A1 t A | |
| k | kala | k A1 l A | |
| g | gaala | g A:1 l A | |
| m | mato | m A1 t O | |
| n | nenä | n e1 n { | |
| r | retki | r e1 t k i | |
| l | latu | l A1 t u | |
| N | henki | h e1 N k i | |
| f | firma | f i1 r m A | |
| v | veli | v e1 l i | |
| s | sata | s A1 t A | |
| S | šakki | S A1 kk i | |
| z | azeri | A1 z e r i | |
| j | joki | j O1 k i | |
| h | hattu | h A1 tt u | |
| dZ | junk | dZ A1 n k | English phoneme |
| D | this | D I1 s | English phoneme |
| T | thin | T I1 n | English phoneme |
| w | sweet | s w i:1 t | English phoneme |
| bb | rabbi | r A1 bb i | |
| tt | hattu | h A1 tt u | |
| pp | keppi | k e1 pp i | |
| dd | addikti | A1 dd i k t i | |
| kk | kukka | k u1 kk A | |
| gg | suggestio | s u1 gg e s t i2 O | |
| mm | summa | s u1 mm A | |
| nn | kannas | k A1 nn A s | |
| rr | parras | p A1 rr A s | |
| ll | nolla | n O1 ll A | |
| NN | hengen | h e1 NN e n | |
| ff | leffa | l e1 ff A | |
| ss | kassa | k A1 ss A | |
| SS | pasha | p A1 SS A | |
| hh | hihhuli | h i1 hh u l i | very rare |
| vv | hernevati | h e1 r n e vv A2 t i | product of assimilation |
| jj | perhejoulu | p e1 r h e jj Ou2 l u | product of assimilation |
| x | ahdas | A1 x d A s | |
| C | yhden | y1 C d e n | |
| W | auvo | Au1 W O | |

**Table 3 Finnish consonants**

## *7.2 Vowels*

### 7.2.1 Symbols for the Finnish vowels

| Symbol | Word | Phonetic text | Comment |
|---|---|---|---|
| i | isä | i1 s { | Short vowel |
| y | tyttö | t y1 tt & | Short vowel |
| e | peli | p e1 l i | Short vowel |
| & | tölli | t &1 ll i | Short vowel |
| { | sänky | s {1 N k y | Short vowel |
| u | tuttu | t u1 tt u | Short vowel |
| O | orsi | O1 r s i | Short vowel |
| A | pato | p A1 t O | Short vowel |
| i: | siili | s i:1 l i | Long vowel |
| y: | tyyli | t y:1 l i | Long vowel |
| e: | geeni | g e:1 n i | Long vowel |
| &: | älkööt | {1 l k &: t | Long vowel |
| {: | jää | j {:1 | Long vowel |
| u: | paluu | p A1 l u: | Long vowel |
| O: | rooli | r O:1 l i | Long vowel |
| A: | laavu | l A:1 v u | Long vowel |
| Ai | laiva | l Ai1 v A | Diphthong |
| ei | leipä | l ei1 p { | Diphthong |
| Oi | voi | v Oi1 | Diphthong |
| ui | luiska | l ui1 s k A | Diphthong |
| {i | päivä | p {i1 v { | Diphthong |
| &i | söi | s &i1 | Diphthong |
| yi | lyijy | l yi1 j y | Diphthong |
| Au | taulu | t Au1 l u | Diphthong |
| Ou | koulu | k Ou1 l u | Diphthong |
| eu | neula | n eu1 l A | Diphthong |
| iu | kiulu | k iu1 l u | Diphthong |
| {y | käydä | K {y1 d { | Diphthong |
| &y | köysi | k &y1 s i | Diphthong |
| ey | yhteys | y1 C t ey s | Diphthong |
| iy | siistiytyä | s i:1 s t iy t y2 { | Diphthong |
| ie | kieli | k ie1 l i | Diphthong |
| uO | tuoli | t uO1 l i | Diphthong |
| y& | työ | t y&1 | Diphthong |

**Table 4 Finnish vowels**

## *7.3 Lexical accent*

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 the symbol marking the word stress.

Primary stress in SAMPA is denoted by a "**1**" placed immediately after the stressed vowel. Secondary stress, common in many longer words, is denoted by a "**2**".

Each purely Finnish word should have a primary stress mark on the first vowel. Many longer words have a secondary stress on one or more syllables later in the word.

Only vowels are stressed, i.e., a stress mark can only appear on a vowel, never on a consonant.

## *7.4 Word boundary gemination*

Word boundary gemination refers to a phenomenon which occurs at word boundaries. The Finnish text-to-speech system can produce word boundary gemination in cases where a word ends in “-lle, -nne, -nsa, -nsä, -sti, -lti”, and the next word begins with a consonant that can be geminated.

Example:
“alle /A1 ll e/” and “koko /k O1 k O/”
“alle koko /A1 ll e kk O1 k O/”

A similar phenomenon appears in some compound words, for instance in most cases when the first part of the compound ends in “-e”. This is the result of old “assimilations”, see the examples under 7.1.1. The system automatically produces this gemination in some, but not all, such compounds.

## *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. 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.

# 8 Abbreviations

In the current version of the Finnish text-to-speech system, the abbreviations in table 5 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 5.6.

| Abbreviation | Reading |
|---|---|
| mm | millimetri |
| dam | dekametri |
| hm | hehtometri |
| ml | millilitra |
| cl | senttilitra |
| dl | desilitra |
| dal | dekalitra |
| hl | hehtolitra |
| kl | kilolitra |
| mg | milligramma |
| cg | senttigramma |
| dg | desigramma |
| dag | dekagramma |
| hg | hehtogramma |
| kg | kilo |
| ha | hehtaari |
| kJ* | kilojoule |
| MJ* | megajoule |
| mW* | megavatti |
| kW* | kilovatti |
| kWh* | kilovattitunti |
| kcal | kilokalori |
| vrk | vuorokautta |
| kpl | kappaletta |
| pv | päivä |
| vk | viikkoa |
| kk | kuukautta |
| aik. | aikaisempi |
| alv. | arvonlisävero |
| ao. | asianomainen |
| ark. | arkisin |
| eaa. | ennen ajanlaskun alkua |
| ed. | edellinen |
| eKr.* | ennen Kristusta |
| em. | edellä mainittu |
| ent. | entinen |
| esim. | esimerkiksi |
| hlö | henkilö |
| hra | herra |
| huom. | huomaa |
| jaa. | jälkeen ajanlaskun alun |
| jhk | johonkin |
| jKr.* | jälkeen Kristuksen |
| jne. | ja niin edelleen |
| jnk. | jonkun |
| joht. | johtaja |
| jssk | jossakin |
| jstk | jostakin |

| | |
|---|---|
| jtk. | jotakin |
| kko | kirkko |
| krs | kerros |
| lk. | luokka |
| lyh. | lyhenne |
| läh. | lähettäjä |
| ml. | mukaan luettuna |
| myöh. | myöhempi |
| nimim. | nimimerkki |
| nro | numero |
| n:o | numero |
| ns. | niin sanottu |
| nyk. | nykyinen |
| oy | osakeyhtiö |
| prof. | professori |
| rva | rouva |
| ry | rekisteröity-yhdistys |
| siht. | sihteeri |
| synt. | syntynyt |
| ts. | toisin sanoen |
| tv | televisio |
| vkl | viikonloppu |
| yht. | yhteensä |
| yo. | ylioppilas |

**Table 5 Abbreviations**

The following abbreviations for measurements, “m”, “g”, “a”, “A”, “l”, “s”, “V”, “W” and “J” are expanded only when appearing after a number.

| **Examples** | **Readings** |
|---|---|
| 25 m | kaksikymmentäviisi metriä |
| 2 g | kaksi grammaa |

# 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 “piste”, hyphens as “viiva”, underscore (“_”) as “alaviiva”, slash (“/”) as “vinoviiva”.
- “fl, us, fr” and all the other abbreviations for countries are spelled out letter by letter.
- The “@” is read “ät-merkki”.
- 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** | **Reading** |
|---|---|
| www.acapela.com | w w w piste acapela piste com |
| http://www.acapela.com | h t t p kaksoispiste vinoviiva vinoviiva w w w piste acapela piste com |
| matti@yahoo.fl | matti ät-merkki jahuu piste f l |
| matti_suomi@yahoo.fl | matti alaviiva suomi ät-merkki jahoo piste f l |