

[Next](#)
[Up](#)
[Previous](#)
[Contents](#)
[Index](#)

Next: [The EAGLET term database](#) Up: [Bibliographical references](#) Previous: [References](#)

## SAMPA and X-SAMPA phonetic symbols

The SAMPA alphabet was developed in the late 1980s by John Wells, in consultation with a wide range of colleagues, to meet a need for a simple machine-readable encoding of phonetic transcriptions with symbols of the International Phonetic Alphabet (IPA) for file interchange purposes. At that time, standardisation of symbol codes and IPA fonts was not highly developed. The underlying principle of SAMPA was to select those IPA symbols which were conventionally used to represent phonemes in the major languages of the European Union, and to assign a 7-bit ASCII code number (below 128) to each. One of the secondary criteria was the visual similarity of the IPA symbol and the letter representing the ASCII code.

Since that time, the standardisation of IPA encoding has progressed, with the system developed by John Esling (the 'Esling codes'), and, more recently, Unicode representations. For practical purposes, however, little has changed at the time of writing, and there is still a need for a straightforward machine-readable encoding.

In the meantime, SAMPA is widely used, and extensions of SAMPA have now been developed for many other languages. In order to aid the development of such extensions, the extended code-set X-SAMPA was devised by John Wells, and encompasses the complete set of IPA conventions. For a number of symbols, human readability had to be sacrificed in favour of simple, unambiguous machine-readability, owing to the restricted number of ASCII codes. The present collation of SAMPA and X-SAMPA is by Inge Mertins.

For further details, consult [ [Gibbon et al.\(1997\)Gibbon, Moore Winski](#) ] and the relevant IPA and SAMPA Internet sites, including project sites with working versions of SAMPA for specific languages.

For prosodic annotation, a number of systems are available. A number of these are discussed in Chapter [1](#). The most widely used in extensive corpus annotation, computational linguistics and speech technology is currently ToBI (Tones and Break Indices); the SAMPROSA system [GibbonMooreWinski:1997](#) contains additional symbols which are suitable for more detailed dialogue transcription.

Readers should be aware that there is still considerable need for standardisation with respect to the use of IPA codes and fonts in consumer software such as word processors and Internet browsers.

### VOWELS

DESCRIPTION	IPA	SAMPA/ X-SAMPA	ASCII/ ANSI
close front unrounded	i	i	105
close front rounded	y	y	121
close central unrounded	ɨ	1	49
close central rounded	ɥ	}	125
close back unrounded	ɯ	M	77
close back rounded	u	u	117
near-close front unrounded (lax i)	ɪ	I	73
near-close front rounded (lax y)	ʏ	Y	89
near-close back rounded (lax u)	ʊ	U	85
close-mid front unrounded	e	e	101
close-mid front rounded	ø	2	50
close-mid central unrounded	ɘ	@\	64, 92
close-mid central rounded	ɚ	8	56
close-mid back unrounded	ɘ	7	55
close-mid back rounded	o	o	111
mid central unrounded (schwa)	ə	@	64
open-mid front unrounded	ɛ	E	69
open-mid front rounded	œ	9	57

open-mid central unrounded	ɜ	3	51
open-mid central rounded	ɜ	3\	51, 92
open-mid back unrounded	ʌ	V	86
open-mid back rounded	ɔ	O	79
near-open front unrounded	æ	{	123
near-open central unrounded	ɛ	6	54
open front unrounded	a	a	97
open front rounded	œ	&	38
open back unrounded	ɑ	A	65
open back rounded	ɒ	Q	81

## CONSONANTS (PULMONIC)

DESCRIPTION	IPA	SAMPA/ X-SAMPA	ASCII/ ANSI
voiceless bilabial plosive	p	p	112
voiced bilabial plosive	b	b	98
voiceless dental/alveolar plosive	t	t	116
voiced dental/alveolar plosive	d	d	100
voiceless retroflex plosive	ʈ	t`	116, 96
voiced retroflex plosive	ɖ	d`	100, 96
voiceless palatal plosive	c	c	99
voiced palatal plosive	ɟ	J\	74, 92
voiceless velar plosive	k	k	107
voiced velar plosive	g	g	103
voiceless uvular plosive	q	q	113
voiced uvular plosive	ɢ	G\	71, 92
glottal stop	ʔ	?	63
bilabial nasal	m	m	109
labiodental nasal	ɱ	F	70
dental/alveolar nasal	n	n	110
retroflex nasal	ɳ	n`	110, 96
palatal nasal	ɲ	J	74
velar nasal	ŋ	N	78
uvular nasal	ɴ	N\	78, 92
bilabial trill	β	B\	66, 92
alveolar trill	r	r	114
uvular trill	ʀ	R\	82, 92
alveolar tap	ɾ	4	52
retroflex flap	ɽ	r`	114, 96
voiceless bilabial fricative	ɸ	p\	112, 92
voiced bilabial fricative	β	B	66
voiceless labiodental fricative	f	f	102
voiced labiodental fricative	v	v	118
voiceless dental fricative	θ	T	84
voiced dental fricative	ð	D	68
voiceless alveolar fricative	s	s	115
voiced alveolar fricative	z	z	122
voiceless postalveolar fricative	ʃ	S	83
voiced postalveolar fricative	ʒ	Z	90
voiceless retroflex fricative	ɬ	s`	115, 96
voiced retroflex fricative	ɮ	z`	122, 96

## CONSONANTS (PULMONIC), CONTINUED

DESCRIPTION	IPA SAMPA/ X-SAMPA	ASCII/ ANSI
voiceless palatal fricative	ç	C 67
voiced palatal fricative	ʝ	ʝ\ 106, 92
voiceless velar fricative	x	x 120
voiced velar fricative	ɣ	G 71
voiceless uvular fricative	χ	X 88
voiced uvular fricative	ʁ	R 82
voiceless pharyngeal fricative	ħ	X\ 88, 92
voiced pharyngeal fricative	ʕ	?\ 63, 92
voiceless glottal fricative	h	h 104
voiced glottal fricative	ɦ	h\ 104, 92
voiceless alveolar lateral fricative	ɬ	K 75
voiced alveolar lateral fricative	ɮ	K\ 75, 92
labiodental approximant	ʋ	P (or v) 80 (118, 92)
alveolar approximant	ɹ	r\ 114, 92
retroflex approximant	ɻ	r\` 114, 92, 96
palatal approximant	j	j 106
velar approximant	ɰ	M\ 77, 92
dental/alveolar lateral approximant	l	l 108
retroflex lateral approximant	ɭ	l\` 108, 96
palatal lateral approximant	ʎ	L 76
velar lateral approximant	ʟ	L\ 76, 92

## CLICKS

DESCRIPTION	IPA SAMPA/ X-SAMPA	ASCII/ ANSI
bilabial	ɸ	O\ 79, 92 (capital O)
dental	ǀ	\ 124, 92
(post)alveolar	ǃ	\ 33, 92
palatoalveolar	ǃ̸	=\ 61, 92
alveolar lateral	ǁ	N\ 124, 92, 124, 92

## EJECTIVES, IMPLOSIVES

DESCRIPTION	IPA SAMPA/ X-SAMPA	ASCII/ ANSI
bilabial ejective	p̰	p_> 112, 95, 62
dental/alveolar ejective	t̰	t_> 116, 95, 62
velar ejective	k̰	k_> 107, 95, 62
alveolar fricative ejective	s̰	s_> 115, 95, 62
voiced bilabial implosive	ɓ	b_< 98, 95, 60
voiced dental/alveolar implosive	ɗ	d_< 100, 95, 60
voiced palatal implosive	ɠ	ʝ_< 74, 92, 95, 60
voiced velar implosive	ɠ	g_< 103, 95, 60
voiced uvular implosive	ɠ	G_< 71, 92, 95, 60
The following were withdrawn from the IPA in 1993:		
voiceless bilabial implosive	ɸ̰	p_< 112, 95, 60
voiceless dental/alveolar implosive	ɸ̰	t_< 116, 95, 60
voiceless palatal implosive	ɸ̰	c_< 99, 95, 60
voiceless velar implosive	ɸ̰	k_< 107, 95, 60
voiceless uvular implosive	ɸ̰	q_< 113, 95, 60

## OTHER SYMBOLS

DESCRIPTION	IPA	SAMPA/ X-SAMPA ANSI	ASCII/ ANSI
voiceless labial-velar fricative	ʍ	w	87
voiced labial-velar approximant	w	w	119
voiced labial-palatal approximant	ɥ	H	72
voiceless epiglottal fricative	ħ	H\	72, 92
voiced epiglottal fricative	ʕ	<\	60, 92
epiglottal plosive	ʡ	>\	62, 92
voiceless alveolo-palatal fricative	ç	s\	115, 92
voiced alveolo-palatal fricative	ʝ	z\	122, 92
alveolar lateral flap	ɺ	l\	108, 92
simultaneous ʃ and x	ʃx	x\	120, 92
tie bar	͡	k p t s _	95

## DIACRITICS

DESCRIPTION	IPA SAMPA/ X-SAMPA	EXAMPLE IPA SAMPA
voiceless	_0 (0 = figure)	n_0
voiced	_v	s_v
aspirated	_h	t^h
more rounded	_O (letter O)	o_O
less rounded	_c	o_c
advanced	_+	u_+
retracted	_-	e_-
centralized	_"	e_"
mid-centralized	_x	e_x
syllabic	= (or _=)	n= (or n_=)
non-syllabic	_^	e_^
rhoticity	^	@^
breathy voiced	_t	b_t
creaky voiced	_k	e_k
linguolabial	_N	t_N
labialized	_w	t^w
palatalized	_j (or _j)	t^j (or t_j)
velarized	_G	t^V
pharyngealized	_ʔ	d^ʔ
velarized or pharyngealized	_e	l_e
velarized l, alternatively	5	
raised	_r	e_r
lowered	_o	e_o
advanced tongue root	_A	e_A
retracted tongue root	_q	e_q
dental	_d	t_d
apical	_a	d_a
laminal	_m	n_m
nasalized	_~ (or _~)	e~ (or e_~)
nasal release	_n	d^n
lateral release	_l	d^l
no audible release	_}	t_}

## SUPRASEGMENTALS

DESCRIPTION	IPA	SAMPA X-SAMPA ANSI	ASCII/
primary stress	"		34
secondary stress	%		37
long	:		58
half-long	ː		58, 92
extra-short	, eg ɛ̃	_X	95, 88
minor (foot) group			
major (intonation) group			124
syllable break	.	\$	36
linking mark	~	-\	45, 92

## TONES AND WORD ACCENTS

DESCRIPTION	IPA	SAMPA X-SAMPA	EXAMPLE IPA SAMPA
level extra high	˥ or ˩	_T or _1	e e_T or e_1
level high	˨ or ˧	_H or _2	é e_H or e_2
level mid	˦ or ˥	_M or _3	ē e_M or e_3
level low	˧ or ˦	_L or _4	è e_L or e_4
level extra low	˩ or ˥	_B or _5	e˩ e_B or e_5
downstep	↓	!	
upstep	↑	^	
contour, rising	˨˨ or ˩˩	_R or _J or _L_H	e˩ e_R, e_J, e_L_H
contour, falling	˨˨ or ˩˩	_F or _\ or _H_L	e˩ e_F, e_\ e_H_L
contour, high rising	˨˨ or ˩˩	_H_T	e˩ e_H_T
contour, low rising	˨˨ or ˩˩	_B_L	e˩ e_B_L
contour, rising-falling	˨˨ or ˩˩	_R_F or _J_\ or _M_H_L	e˩ e_R_F, e_J_\ e_M_H_L
global rise	↗	<R> or </>	
global fall	↘	<F> or <\>	

NB: Instead of being written as diacritics with ˩, ˥, all prosodic marks can alternatively be placed in a separate tier, set off by <>, as recommended for global rise and global fall.

## WIDELY USED BUT LESS STANDARDISED SYMBOLS

SAMPA ASCII	Comment
...	46,46,46 Silent pause
92	Phonetic case-shift (eg \*F might be used to signal a shift into French and would terminate the shift).
§	21 Phonological Phrase
#	35 Word Boundary
##	35,35 Absence of liaison
+	43 Morpheme boundary

[Next](#) [Up](#) [Previous](#) [Contents](#) [Index](#)

Next: [The EAGLET term database](#) Up: [Bibliographical references](#) Previous: [References](#)

Handbook of Multimodal and Spoken Dialogue Systems  
Resources, Terminology and Product Evaluation