

# Indian Language Speech sound Label set (ILSL12)

## Version 2.1.6



Indian Language Speech sound Label set (ILSL12), 2012  
developed by Indian Language TTS Consortium & ASR Consortium

Copyright (c) 2012 Indian Language TTS Consortium & ASR Consortium  
Dr Samudravijaya, Tata Institute of Fundamental Research, Mumbai  
([chief@tifr.res.in](mailto:chief@tifr.res.in), [samudravijaya@gmail.com](mailto:samudravijaya@gmail.com))  
Dr Hema A Murthy, Indian Institute of Technology, Madras  
([hema@cse.iitm.ac.in](mailto:hema@cse.iitm.ac.in), [hema.arunachalam@gmail.com](mailto:hema.arunachalam@gmail.com))

This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/4.0/>.

---

This document specifies a standard set of labels (in Roman script) for speech sounds commonly used in Indian languages. This document lists the label set for 13 languages (currently being processed by ASR/TTS consortia of TDIL, DIT, GoI). These labels are to be used for computer processing of spoken Indian languages.

- (1) **Similar** sounds in different languages are given a single label.
- (2) The IPA symbol refers to an exemplar (Hindi/Tamil/other) language.
- (3) This is not an IPA chart of sounds of Indian languages.
- (4) The label set is designed such that the native script is **largely** recoverable from the transliteration.

A label may consist of a sequence of alphanumeric characters of the Roman alphabet; they will not contain any special character such as quote, hyphen etc. All labels are in lower case even though the labels are case insensitive. Since the number of speech sounds are larger than the Roman alphabet, a system of suffixes as well as letter combinations are used for labels.

### (A) Notes on suffixes:

1. **Aspiration**: Use suffix **h** to denote aspiration: **k** (क) versus **kh** (ख).
2. **Retroflex consonants**: Use suffix **x** to denote retroflex place of articulation: **t** (ट) versus **tx** (ट्र).
3. **Nukta / bindu** : Use suffix **q** to denote a nukta/bindu: **dx** (ड) versus **dxq** (ड़). Nukta (a dot below the glyph) may denote a flap/tap or a fricative variant of the consonant. Bindu (a dot

above an [vowel] glyph) denotes a nasal after the vowel; the place of articulation of the nasal will be the same as that of the following consonant. If there is no consonant after the bindu, the vowel is nasalized.

4. **Nasalized vowel:** Use suffix **n** to denote nasalization of a vowel: k a h **aa** (कहा) versus k a h **aan** (कहाँ).
5. **Geminated sounds:** The label for a geminated consonant is the label of the corresponding single consonant with the first letter of the label repeated. Example: p a **k** aa (पका) versus p a **kk** aa (पक्का) in Hindi; a **ddx** aa (अड़्डा) in Hindi; a **ll** a m (ginger in Tamil).
6. **Other special cases:** Use suffix **x** to denote certain special cases: reduced vowel (**axx**) in various languages; “a” of Bangla; apical affricates of Marathi; special **r** of Dravidian languages etc.
7. **Priority of suffixes:** Some symbols may have multiple suffixes. In such cases the following is the priority (in decreasing order): **x h q n**

## (B) Notes on Matras, Diphthongs and Halant:

1. The label for an vowel *matra* is the same as that of the vowel.
2. The label of a diphthong is generated as a concatenation of the labels of the corresponding vowel. The exceptions to this rule are “ae”, “ea” and “eu”; these are monophthongs.
3. The halant in Indian scripts denotes the absence of the implicit “a” in Indian consonant characters. It is not a sound and hence there is no label for halant. The morphological analyser of the language shall delete the implicit “a” when a halant is present in the script.
4. Punctuation marks: The 'transliteration' module will retain the punctuation marks (exception: '|' and “||” will be replaced by fullstop); these are useful for prosody generation. The morphological analyser will remove the punctuation marks while generating the word/phone level transcription.

(C) **Language specific notes\*** : This section has notes on sounds (and labels) specific to a subset of the languages. Whenever possible, minimal pairs for language specific phonemes are provided. In other cases, exemplar words containing common phones in that language are written.

### 1. Hindi

(i) The glyph ञ is pronounced as a sequence of two phones: "g y". The morphological analyser of the language will do this translation from glyph to the phone sequence.

(i) The compound glyph ज्ञ= “ज + ञ्” is pronounced as a sequence of two phones: "g y". The morphological analyser of the language will do this translation from glyph to the phone sequence.

### 2. Marathi

Label	IPA	Glyph	Minimal pair	Exemplar words/ Comment
nxh	/ɳ <sup>h</sup> /	ण्ह		
nh	/n <sup>h</sup> /	न्ह	नाणे (coin) न्हाणे (take bath)	
mh	/m <sup>h</sup> /	म्ह	म्हण (say) मण ()	
rh	/r <sup>h</sup> /	ऱ्ह	रास (heap)	

Label	IPA	Glyph	Minimal pair	Exemplar words/ Comment
			ऱ्हास (decay)	
lh	/l <sup>h</sup> /	ल्ह		उल्हास, केल्हाय
wh	/w <sup>h</sup> /	व्ह	वाळ (a kind of cereal) व्हाळ (will become)	

\* content not complete

(1) This is a common Label Set (in Roman script) for the purpose of computer processing of spoken Indian languages.

(2) **Similar** sounds in different languages are given a single label.

(3) The IPA label refers to an exemplar (Hindi/Tamil/other) language.

(4) This is NOT an IPA chart of sounds of Indian languages.

Sl.No.	Label	IPA	Hindi	Marathi	Rajasthani	Gujarati	Odia
1	a	a	अ	अ	अ	અ	-
2	ax	ɔ	-	औ	-	ઓ	ଅ
3	aa	a:	आ	आ	आ	આ	ଆ
4	axx	ə	-	-	-	-	-
5	i	I, i	इ	इ	इ	ઈ	ଇ, ଇ
6	ii	i:	ई	ई	ई	ઈ	-
7	u	u, ʊ	उ	उ	उ	ઉ	ଉ, ଊ
8	eu	ɯ	-	-	-	-	-
9	uu	u:	ऊ	ऊ	ऊ	ઊ	-
10	rq	-	ऋ, ॠ	ऋ, ॠ	ऋ	-	ଋ
11	e	e	-	-	-	-	ଏ
12	ee	e:	ए	ए, ऐ	ए	એ	-
13	ea	ɛ	-	-	-	-	-
14	ei	ɛ:	ऐ	-	ऐ	ઐ	-
15	ai	aɪ	-	ऐ	-	-	ଐ
16	oi	oɪ	-	-	-	-	-
17	o	o	ओ	ओ, औ	ओ	ઓ	ଓ
18	oo	o:	-	-	-	-	-
19	ae	ae	-	ँ	-	ઁ	-

20	au	aʊ	-	औ	-	-	औ
21	ou	oʊ	औ	-	औ	औ	-
22	k	k	क	क	क	क	क
23	kh	k <sup>h</sup>	ख	ख	ख	ਖ	ਖ
24	g	g	ग	ग	ग	ਗ	ਗ
25	gh	g <sup>h</sup>	घ	घ	घ	ਘ	ਘ
26	ng	ŋ	ङ	ङ	ङ	ਙ	ਙ
27	c	tʃ	च	च	च	ਚ	ਚ
28	ch	tʃ <sup>h</sup>	छ	छ	छ	ਚ਼	ਚ਼
29	cx	tʃ	-	च	-	-	-
30	j	dʒ	ज	ज	ज	ਜ	ਜ,ਯ
31	jh	dʒ <sup>h</sup>	झ	झ	झ	ਝ	ਝ
32	jx	dʒ	-	ज	-	-	-
33	nj	ɲ	ञ	ञ	-	ਅ	ਞ
34	tx	t	ट	ट	ट	ਟ	ਟ
35	txh	t <sup>h</sup>	ठ	ठ	ठ	ਠ	ਠ
36	dx	d	ड	ड	ड	ਡ	ਡ
37	dxh	d <sup>h</sup>	ढ	ढ	ढ	ਢ	ਢ
38	nx	ɳ	ण	ण	ण	ਨ	ਨ
39	t	t̪	त	त	त	ਤ	ਤ
40	th	t̪ <sup>h</sup>	थ	थ	थ	ਥ	ਥ
41	d	d̪	द	द	द	ਦ	ਦ
42	dh	d̪ <sup>h</sup>	ध	ध	ध	ਧ	ਧ
43	n	n	न,न	न,न	न	ਨ	ਨ

44	nd	-	-	-	-	-	-
45	p	p	प	प	प	प	ପ
46	ph	p <sup>h</sup>	फ	फ	फ	ଝ	ଫ
47	b	b	ब	ब	ब	ब	ବ
48	bh	b <sup>h</sup>	भ	भ	भ	भ	ଭ
49	m	m	म	म	म	म	ମ
50	y	j	य,य	य,य	य	य	ୟ
51	r	r	र,र	र,र	र	र	ର
52	l	l	ल	ल	ल	ल	ଲ
53	lx	l	-	ळ,ळ	ळ	ळ	ଳ
54	w	υ	व	व	व	व	ଝ,ଞ
55	sh	∫	श	श	श	श	-
56	sx	§	ष	ष	ष	ष	-
57	s	s	स	स	स	स	ସ, ଶ, ଷ
58	h	h	ह	ह	ह	ह	ହ
59	kq	q	क	क	-	-	-
60	khq	x	ख	ख	-	-	-
61	gq	Ƴ	ग	ग	-	-	-
62	z	z	ज	ज	-	-	-
63	jhq	3	झ	झ	-	-	-
64	dxq	ƭ	ड	ड	ड	-	ଡ
65	dxhq	ƭ <sup>h</sup>	ढ	ढ	ढ	-	ढ
66	dhq	-	-	-	ध	-	-
67	f	f	फ़	फ़	-	-	-

Sheet1

68	bq	-	-	-	ब	-	-
69	yq	-	-	-	-	-	-
70	nq	-	-	-	-	-	-
71	rx	ॠ	-	-	-	-	-
72	sq	-	-	-	स	-	-
73	zh	ॢ	-	-	-	-	-
74	nxh	ॣ <sup>h</sup>	-	ॣह	-	-	-
75	nh	n <sup>h</sup>	-	ॣह	-	-	-
76	mh	m <sup>h</sup>	-	ॣह	ॣह	-	-
77	rh	l <sup>h</sup>	-	ॣह	-	-	-
78	lh	l <sup>h</sup>	-	ॣह	-	-	-
79	wh	w <sup>h</sup>	-	ॣह	ॣह	-	-
80	q	-	ं	ं	ं	ं	ं°
81	hq	-	ः	ः	ः	ः	ः४
82	mq	-	ँ	ँ	ँ	ँ	ँ
83	x	x	-	-	-	-	-

- (1) This is a common Label Set (in Roman script) for the purpose of computer processing of spoken Indian languages.
- (2) **Similar** sounds in different languages are given a single label.
- (3) The IPA label refers to an exemplar (Hindi/Tamil/other) language.
- (4) This is NOT an IPA chart of sounds of Indian languages.

Sl.No.	Label	Bengali		Assamese		Manipuri		Bodo
		P	G	P	G	P	G	
1	a	-	-	-	-	অ	অ	অ
2	ax	অ	অ	অ	অ	-	-	-
3	aa	আ	আ	আ	আ	আ	আ	আ
4	axx	-	-	-	-	অ	অ	-
5	i	ই,ঐ	ই	ই,ঐ	ই	ই	ই	ই
6	ii	-	ঐ	-	ঐ	ঐ	ঐ	ঐ
7	u	উ,ঊ	উ	উ,ঊ	উ	উ	উ	উ
8	eu	-	-	-	-	-	-	-
9	uu	-	ঊ	-	ঊ	ঊ	ঊ	ঊ
10	rq	ঋ,ঌ	ঋ,ঌ	ঋ	ঋ	-	-	ঋ
11	e	এ	এ	এ	এ	এ	এ	এ
12	ee	-	-	-	-	-	-	-
13	ea	-	-	-	-	-	-	-
14	ei	-	-	-	-	-	-	ঐ
15	ai	-	-	-	-	ঐ	ঐ	-
16	oi	ঐ	ঐ	ঐ	ঐ	ওই	ওই	-
17	o	ও	ও	ও	ও	ও	ও	ও
18	oo	-	-	-	-	-	-	-
19	ae	-	অ্যা	-	-	-	-	-



Sheet1

20	au	-	-	-	-	ঔ	ঔ	-
21	ou	ঔ	ঔ	ঔ	ঔ	-	-	ঔ
22	k	ক	ক	ক	ক	ক	ক	ক
23	kh	খ	খ	খ	খ	খ	খ	খ
24	g	গ	গ	গ	গ	গ	গ	গ
25	gh	ঘ	ঘ	ঘ	ঘ	ঘ	ঘ	ঘ
26	ng	ঙ, ঙ	ঙ	ঙ, ঙ	ঙ	ঙ	ঙ	ঙ
27	c	চ	চ	-	-	চ	চ	চ
28	ch	ছ	ছ	-	-	ছ	ছ	ছ
29	cx	-	-	-	-	-	-	-
30	j	জ, য	জ	জ, য	জ, য	জ	জ	জ
31	jh	ঝ	ঝ	ঝ	ঝ	ঝ	ঝ	ঝ
32	jx	-	-	-	-	-	-	-
33	nj	-	-	ঞ	ঞ	ঞ	ঞ	ঞ
34	tx	ট	ট	-	ট	-	ট	ট
35	txh	ঠ	ঠ	-	ঠ	-	ঠ	ঠ
36	dx	ড	ড	-	ড	-	ড	ড
37	dxh	ঢ	ঢ	-	ঢ	-	ঢ	ঢ
38	nx	ণ	ণ	-	ণ	-	ণ	ণ
39	t	ত, ত্	ত, ত্	ট, ত, ত্	ত, ত্	ট, ত, ত্	ত, ত্	ত
40	th	থ	থ	ঠ, থ	থ	ঠ, থ	থ	থ
41	d	দ	দ	ড, দ	দ	ড, দ	দ	দ
42	dh	ধ	ধ	ঢ, ধ	ধ	ঢ, ধ	ধ	ধ
43	n	ন	ন	ণ, ন	ন	ণ, ন	ন	ন

## Sheet1

44	nd	-	-	-	-	-	-	-
45	p	প	প	প	প	প	প	প
46	ph	ফ	ফ	ফ	ফ	ফ	ফ	ফ
47	b	ব	ব	ব	ব	ব	ব	ব
48	bh	ভ	ভ	ভ	ভ	ভ	ভ	ভ
49	m	ম	ম	ম	ম	ম	ম	ম
50	y	য়	য	য়	য়	য়,য	য়,য	য়
51	r	র	র	ৰ,ড়	ৰ,ড়	র	র	র
52	l	ল	ল	ল	ল	ল	ল	ল
53	lx	-	-	-	-	-	-	-
54	w	-	ওয়	ৱ	ৱ	ৱ	ৱ	ৱ
55	sh	শ,ষ	শ	-	-	শ	শ	শ
56	sx	-	ষ	-	-	ষ	ষ	ষ
57	s	স	স	চ, ছ	চ, ছ	স	স	স
58	h	হ	হ	হ	হ	হ	হ	হ
59	kq	-	-	-	-	-	-	-
60	khq	-	-	-	-	-	-	-
61	gq	-	-	-	-	-	-	-
62	z	জ	-	-	-	-	-	-
63	jhq	-	-	-	-	-	-	ঝ
64	dxq	ড়	ড়	-	-	-	-	ড়
65	dxhq	ঢ়	ঢ়	ঢ়	ঢ়	-	-	ঢ়
66	dhq	-	-	-	-	-	-	-
67	f	-	ফ	-	-	-	-	-

Sheet1

68	bq	-	-	-	-	-	-	-
69	yq	-	য	-	-	-	-	-
70	nq	-	-	-	-	-	-	-
71	rx	-	-	-	-	-	-	-
72	sq	-	-	-	-	-	-	-
73	zh	-	-	-	-	-	-	-
74	nxh	-	-	-	-	-	-	-
75	nh	-	-	-	-	-	-	-
76	mh	-	-	-	-	-	-	-
77	rh	-	-	-	-	-	-	-
78	lh	-	-	-	-	-	-	-
79	wh	-	-	-	-	-	-	-
80	q	-	ং	-	ং	ং	ং	ং
81	hq	-	ং	ং	ং	-	-	ং
82	mq	ঁ	ঁ	ঁ	ঁ	ঁ	ঁ	-
83	x	-	-	শ,ষ,স	শ,ষ,স	-	-	-

- (1) This is a common Label Set (in Roman script) for the purpose of computer processing of spoken Indian languages.
- (2) **Similar** sounds in different languages are given a single label.
- (3) The IPA label refers to an exemplar (Hindi/Tamil/other) language.
- (4) This is **NOT** an IPA chart of sounds of Indian languages.

Sl.No.	Label	Tamil	Malayalam	Telugu	Kannada
1	a	அ	അ	అ	ಅ
2	ax	-	-	-	-
3	aa	ஆ	ആ	ఆ	ಆ
4	axx	-	-	-	-
5	i	இ	ഇ	ఇ	ಇ
6	ii	ஈ	ഈ	ఈ	ಈ
7	u	உ	ഉ	ఉ	ಉ
8	eu	உ	ഊ	-	ಊ
9	uu	ஊ	ഊ	ఊ	ಊ
10	rq	-	ഋ	ఋ, ౠ	ಋ, ౠ
11	e	எ	എ	ఎ	ಎ
12	ee	ஏ	ഐ	ఏ	ಐ
13	ea	-	-	-	ಎ
14	ei	-	-	-	-
15	ai	ஐ	ഐ	ఐ	ಐ
16	oi	-	-	-	-
17	o	ஓ	ഓ	ఒ	ಓ
18	oo	ஔ	ഔ	ఔ	ಔ
19	ae	-	-	-	-

20	au	ഓ	ഔ	ഓ	ഔ
21	ou	-	-	-	-
22	k	ക	ക	ക	ക
23	kh	-	ഖ	ഖ	ഖ
24	g	ഗ	ഗ	ഗ	ഗ
25	gh	-	ഘ	ഘ	ഘ
26	ng	ങ	ങ	ങ	ങ
27	c	ച	ച	ച	ച
28	ch	-	ഛ	ഛ	ഛ
29	cx	-	-	-	-
30	j	ജ	ജ	ജ	ജ
31	jh	-	ജ	ജ	ജ
32	jx	-	-	-	-
33	nj	ഞ	ഞ	ഞ	ഞ
34	tx	ട	ട	ട	ട
35	txh	-	ഠ	ഠ	ഠ
36	dx	ഡ	ഡ	ഡ	ഡ
37	dxh	-	ഢ	ഢ	ഢ
38	nx	ണ	ണ	ണ	ണ
39	t	ത	ത	ത	ത
40	th	-	ഥ	ഥ	ഥ
41	d	ത	ദ	ദ	ദ
42	dh	-	ധ	ധ	ധ
43	n	ന	ന	ന	ന

Sheet1

44	nd	ந	-	-	।
45	p	ப	ப	ப	ப
46	ph	-	ஹ	ஹ	ஹ
47	b	பV	ബ	బ	బ
48	bh	-	ഭ	భ	భ
49	m	ம	മ	మ	మ
50	y	ய	య	య	య
51	r	ர	ര	ర	ర
52	l	ல	ല	ల	ల
53	lx	ள	ള	ళ	ళ
54	w	வ	വ	వ	వ
55	sh	-	ശ	ష	ష
56	sx	ஷ	ഷ	ష	ష
57	s	ஸ	സ	స	స
58	h	ஹ	ഹ	హ	హ
59	kq	-	-	-	-
60	khq	-	-	-	-
61	gq	-	-	-	-
62	z	-	-	-	-
63	jhq	-	-	-	-
64	dxq	-	ട	-	-
65	dxhq	-	-	-	-
66	dhq	-	-	-	-
67	f	ஃப	-	-	-

Sheet1

68	bq	-	-	-	-
69	yq	-	-	-	-
70	nq	-	ᄒ	-	-
71	rx	ᄒ	ᄒ	ᄒ	ᄒ
72	sq	-	-	-	-
73	zh	ᄒ	ᄒ	-	-
74	nxh	-	-	-	-
75	nh	-	-	-	-
76	mh	-	-	-	-
77	rh	-	-	-	-
78	lh	-	-	-	-
79	wh	-	-	-	-
80	q	-	ᄒᄒ	ᄒᄒ	ᄒᄒ
81	hq	-	ᄒᄒ	ᄒᄒ	ᄒᄒ
82	mq	-	-	-	-
83	x	-	-	-	-