

# How to enrich speech applications using IPA & ToBI

Caroline Henton

Founder, CTO Talknowledgy

[contact@talknowledgy.com](mailto:contact@talknowledgy.com)



# The IPA has it all

- The International Phonetic Alphabet (IPA) was created by the [International Phonetic Association](#) in the 1880s to transcribe the sounds of all spoken languages. It is based (mostly) on Latin letters and uses a large number of diacritics.
- Full Unicode support

# The IPA chart: consonants

## THE INTERNATIONAL PHONETIC ALPHABET (revised to 2005)

CONSONANTS (PULMONIC)

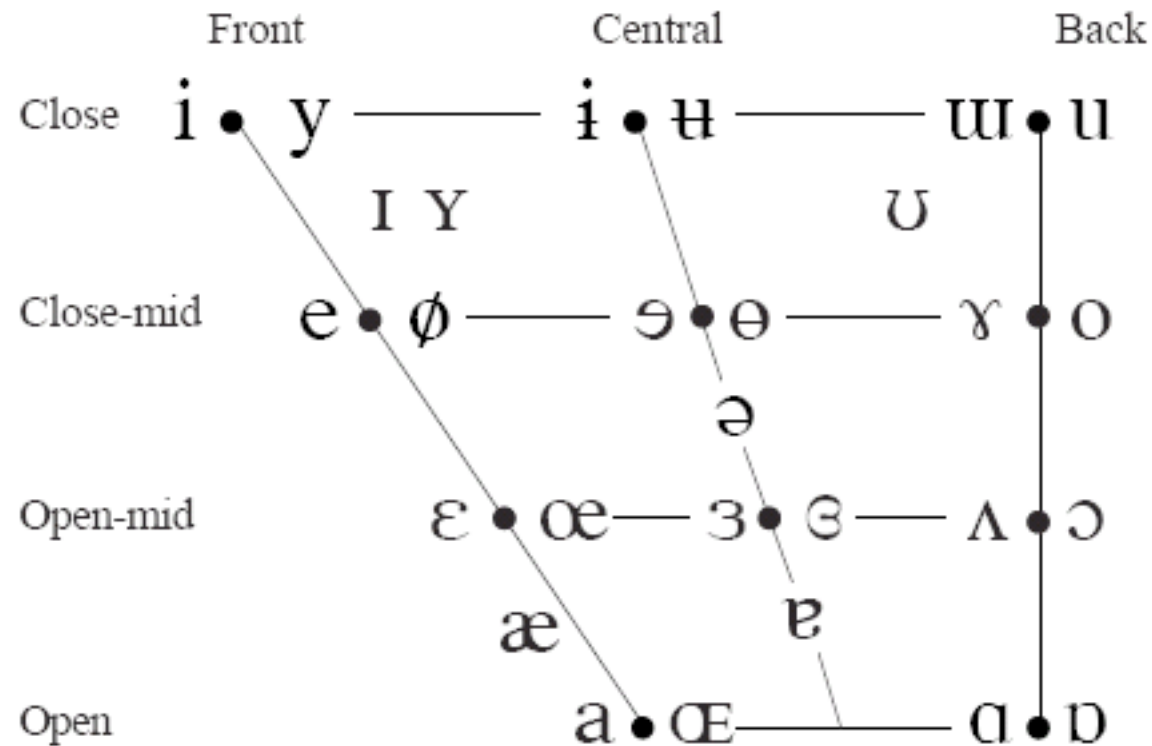
© 2005 IPA

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

# The IPA chart: vowels

## VOWELS



# The IPA chart: diacritics

DIACRITICS Diacritics may be placed above a symbol with a descender, e.g.  $\overset{\circ}{\underset{\cdot}{j}}$

◌ <sup>◌</sup>	Voiceless	$\overset{\circ}{n}$ $\overset{\circ}{d}$	◌ <sup>◌◌</sup>	Breathy voiced	$\overset{\circ}{b}$ $\overset{\circ}{a}$	◌ <sub>◌</sub>	Dental	$\overset{\circ}{t}_{\text{̪}}$ $\overset{\circ}{d}_{\text{̪}}$
◌ <sub>◌</sub>	Voiced	$\underset{\cdot}{s}$ $\underset{\cdot}{t}$	◌ <sub>◌</sub>	Creaky voiced	$\underset{\cdot}{b}$ $\underset{\cdot}{a}$	◌ <sub>◌</sub>	Apical	$\overset{\circ}{t}_{\text{̟}}$ $\overset{\circ}{d}_{\text{̟}}$
◌ <sup>h</sup>	Aspirated	$t^h$ $d^h$	◌ <sub>◌</sub>	Linguolabial	$\overset{\circ}{t}_{\text{̟̞}}$ $\overset{\circ}{d}_{\text{̟̞}}$	◌ <sub>◌</sub>	Laminal	$\overset{\circ}{t}_{\text{̟̞̞}}$ $\overset{\circ}{d}_{\text{̟̞̞}}$
◌ <sub>◌</sub>	More rounded	$\text{ɔ̞}$	◌ <sup>w</sup>	Labialized	$t^w$ $d^w$	◌ <sub>◌</sub>	Nasalized	$\tilde{e}$
◌ <sub>◌</sub>	Less rounded	$\text{ɔ̟}$	◌ <sup>j</sup>	Palatalized	$t^j$ $d^j$	◌ <sup>n</sup>	Nasal release	$d^n$
◌ <sub>◌</sub> <sup>+</sup>	Advanced	$\underset{+}{u}$	◌ <sup>ɤ</sup>	Velarized	$t^ɤ$ $d^ɤ$	◌ <sup>l</sup>	Lateral release	$d^l$
◌ <sub>◌</sub> <sup>-</sup>	Retracted	$\underset{-}{e}$	◌ <sup>ʕ</sup>	Pharyngealized	$t^ʕ$ $d^ʕ$	◌ <sup>̚</sup>	No audible release	$d^{\text{̚}}$
◌ <sup>◌◌</sup>	Centralized	$\ddot{e}$	◌ <sub>◌</sub>	Velarized or pharyngealized	$\text{ɫ}$			
◌ <sup>×</sup>	Mid-centralized	$\overset{\times}{e}$	◌ <sub>◌</sub> <sup>◌</sup>	Raised	$\overset{\text{̟}}{e}$ ( $\underset{\text{̟}}{\text{ɹ}}$ = voiced alveolar fricative)			
◌ <sub>◌</sub>	Syllabic	$\underset{\text{̟}}{n}$	◌ <sub>◌</sub> <sup>◌</sup>	Lowered	$\overset{\text{̟}}{e}$ ( $\underset{\text{̟}}{\beta}$ = voiced bilabial approximant)			
◌ <sub>◌</sub>	Non-syllabic	$\overset{\circ}{e}$	◌ <sub>◌</sub> <sup>◌</sup>	Advanced Tongue Root	$\overset{\text{̟}}{e}$			
◌ <sub>◌</sub>	Rhoticity	$\overset{\circ}{e}^r$ $\overset{\circ}{a}^r$	◌ <sub>◌</sub> <sup>◌</sup>	Retracted Tongue Root	$\overset{\text{̟}}{e}$			

# IPA in Unicode

## SIL IPA93 Doulos Font

a ɐ ɑ ɒ æ b ɓ ɋ β c ɔ ɘ ɥ d  
ɔ ɖ ɔ̃ e ə ə ɛ ɜ ɞ f g ɡ ɠ ɢ  
h <sup>h</sup> ħ ħ ħ ɥ ɨ i ɪ i ɨ ɪ j ɟ  
j ʝ ɰ ʃ k l <sup>l</sup> ɭ ɮ ɬ L m <sup>m</sup> ɱ  
ɰ ɱ n <sup>n</sup> ŋ <sup>ŋ</sup> ŋ ɲ ɳ <sup>ɳ</sup> N o ø ɵ ϕ  
θ œ œ ⊙ p q r ɹ ɺ ɻ ɼ R Ɔ  
s ʂ ʃ t ɹ u ɥ u u v ʌ ʏ ʘ ɣ  
w <sup>w</sup> ʘ x ɣ y ʎ ʏ z ʐ z ʑ ʒ  
ʒ ʕ ʕ ɿ ɿ

# IPA review and beyond

- Consonants, clicks, vowels, diacritics, suprasegmental marks, tones, syllable and word boundaries, and much more.
- BUT user-customization is always needed.....
- Customize pronunciation (with punctuation and phonetic spelling).
- Customize intonation and affect (using embedded commands).

# Computer phonemes ≠ IPA phonemes (US English)

Vowels	r-coloured vowels	Nasals, glides syllabic Cs	Fricatives	Stops etc.
IY	AR	m	h	p
IX	ER	n	f	b
IH	IR	NG	v	t
EH	OR	l	TH	d
AE	UR	LX	DH	k
AA		r	s	g
AH		w	z	DX
AO		y	SH	KX
UH		EL EM EN	ZH	PX
AX		RX	CH	TX
EY			JH	QX
AY				DD
OY				P-
AW			SIL	T-
OW				K-
UW				

# English phonemes $\neq$ phonemes of other languages

## Travelocity's needs:

**Airlines:** Aer Lingus, Lufthansa, Quantas

**Destinations:** Blenheim, Brisbane, Caribbean, Edinburgh, Madras, Monticello, Reims, Norwich, Norfolk, Taj Mahal, Versailles, Xian

**Hotel/Restaurants:** Fointainebleau (FL), L'Ecluse, Le Quai St. Pierre (London), Les Jardins du Marais

**Addresses:** 42 Rue des Bernardins, St. Germain des Près, Beauchamp Place, Berkeley Sq.

**Names:** Marlborough, Van Gogh, Tutankhamen

# Travelocity Worldwide

Travelocity's International Sites & Partnerships:

Australia (E)\*

Canada (E,F)\*

Denmark, Norway, Sweden (D,N,S)

Germany (G)

Hong Kong (E,C)

Korea (K)

Singapore (E)\*

Taiwan (C)

U.K.(E)\*

\*Beware the *faux amis*

# The IPA Suprasegmentals

## SUPRASEGMENTALS

- ' Primary stress
- ˈ Secondary stress
- ˌ *fəʊnəˈtɪʃən*
- ː Long *eː*
- ˑ Half-long *eˑ*
- ◌ Extra-short *ĕ*
- | Minor (foot) group
- || Major (intonation) group
- Syllable break *li.ækt*
- ◌ Linking (absence of a break)

# The IPA Tones and Word Boundaries

TONES AND WORD ACCENTS			
LEVEL		CONTOUR	
◌̎ or ◌̏	Extra high	◌̎ or ◌̏	Rising
◌̍	High	◌̍	Falling
◌̎̄	Mid	◌̎̄	High rising
◌̎̌	Low	◌̎̌	Low rising
◌̎̎̄	Extra low	◌̎̎̄	Rising-falling
◌̎̎̎̄	Downstep	↗	Global rise
↗	Upstep	↘	Global fall

# Fine-tuning in Mac OS X

## Embedded commands

- [[emph + | -]]
- [[rate wpm]]
- [[volm n.n]]
- [[emph +; rate 230; volm 0.6]]
- Input modes TEXT | PHON | TUNE

Further tuning: pbas; pmod; slnc; rset

Lexical stress; syllable breaks; normal and  
destressed words

# ToBI: Tones and Break Indices

ToBI = framework for developing conventions for transcribing the intonation and prosodic structure of spoken utterances in a language variety.

ToBI transcriptions have two important tiers:

1. a tone tier
2. a break-index tier

Note: ToBI is **not** an International Phonetic Alphabet for prosody. Because intonation and prosodic organization differ from language to language, and often from dialect to dialect within a language, there are many different ToBI systems, each one specific to a language variety.

## ToBI Phrase Accents & Final Boundary Tones

L- H- For non-final intermediate phrases.

Intonational phrase boundaries, often at punctuation:

L-L% Low final endpoint, like at most periods.

L-H% Final rise from a low value, often at a comma.

L-% Continuation-like L-H% that's missing the rise.

H-H% Rise from a mid to high value, often used in questions.

H-L% High level, often used in lists.

# ToBI intonation phrases

4 typical intonation phrases:

**L-L%** The default DECLARATIVE phrase.

**L-H%** The LIST ITEM intonation (nonfinal items only).

E.g. "I need food L-H%, shelter L-H%, and comfort L-L%."

"You said you would run home this afternoon L-H%, grab your golf clubs L-H%, jump in the car L-H% and race to the club L-L%."

**H-H%** YES-NO QUESTION. Eg, "Are you going L\* today?" H-H%

"So then are you going L\* to the store this afternoon? H-H%?"  
(where pitch rises right after the L\* and stays high til the end).

**H-L%** The PLATEAU. A previous H\* or complex accent 'upsteps' the final L% to an intermediate level. "I just TOLD you why" L+H\* !H-L%

# ToBI Pitch Accents

## Pitch Accents

Associated only with accented (prominent) syllables:

$L+H^*$  Low immediately preceding a steep rise.

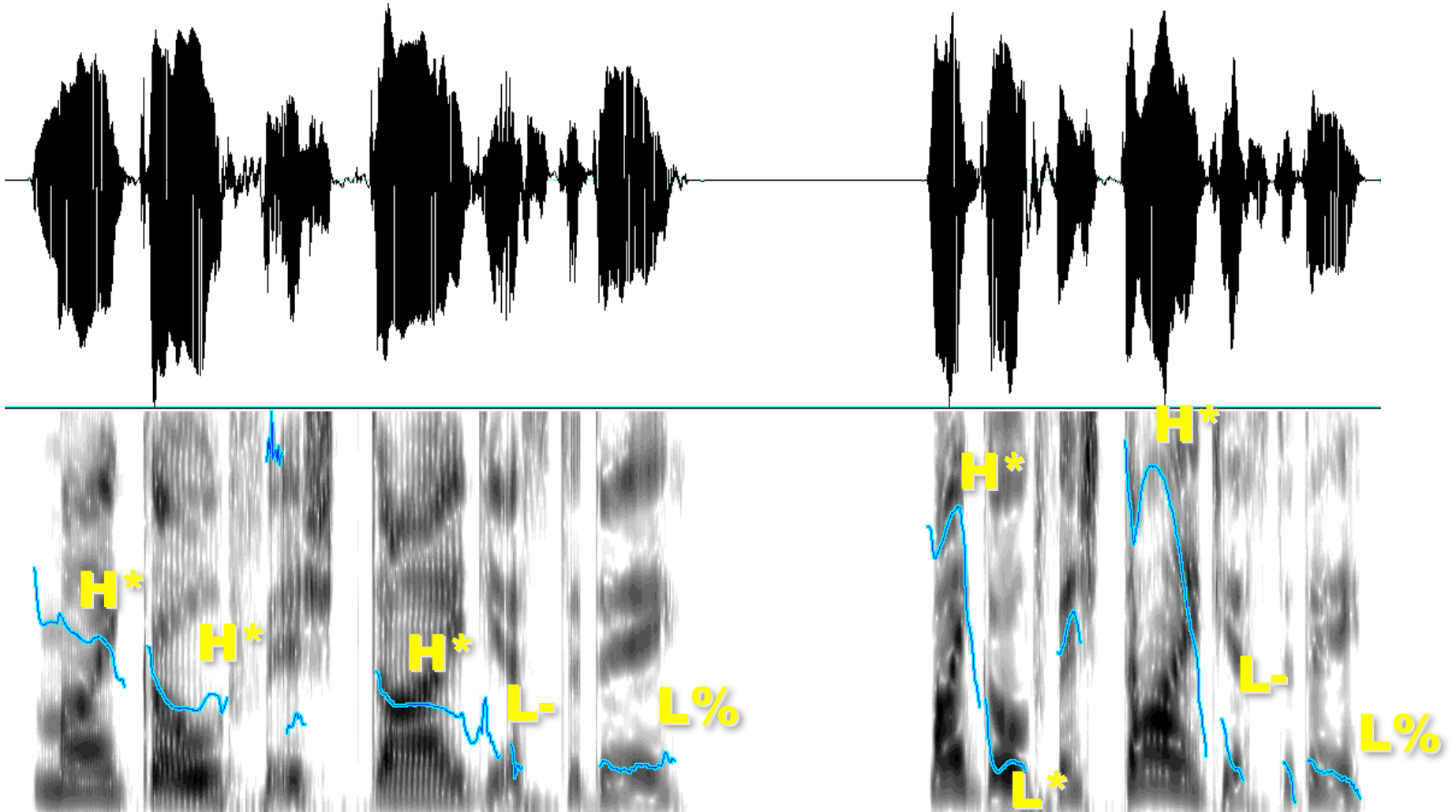
$H^*$  Local maximum or relatively high.

$L^*$ ,  $L^*+H$ ,  $H+L^*$ ,  $H+!H^*$  Less common pitch accents.

< A starred tone to the right of (after) the accent-bearing vowel.

> A starred tone to the left of (before) the accent-bearing vowel.

# ToBI Transcription Examples



# Issues in Prosodic Transcription

1. Difference between L- and L-%.
2. When is a word accented?
3. Where is the accent's F0 peak when there is a steep segmental perturbation from a preceding voiceless obstruent?
4. Missing L- phrase accents.
5. The difference between H\* and L+H\*. Look for a linear interpolation to an H\* peak from the preceding material.
6. Confusing low pitch due to glottalization with a low tone (L\*).
7. Noun compounds without accents in the tail should not get accents in the tail.
8. Difference between L-% and L-H%.
9. Some phrase accents end in H-, not L-.
10. Is a sentence-initial function word ever accented?
11. Polysyllabic words can contain two accents, e.g. "Tennessee".

# Continuing challenges for TTS

- Pronunciation sets & rules for other languages
- Duration rules
- Homophones, Homographs
- Diphone glue
- Constrained function words (phrasal verbs)
- Noun compounds (tuples)
- Dynamic weighting
- Pause insertions
- Intonation choice improvements: e.g. H\* instead of L+H\* in verbs following the subject in a sentence
- Non-Wh-questions
- Exclamation intonation!!!

