ARPABET and the TIMIT alphabet

This short document is about phonetic alphabets. The most important set of symbols for phonetic transcriptions is the International Phonetic Alphabet (IPA). The IPA was developed in 1888 and revised several times. Roughly speaking, it tries to associate a symbol to each sound in all of the known languages in the world. More details about the IPA symbols can be found in [1] (page 29). The IPA is the result of much effort, but is not convenient for some applications.

The IPA makes extensive use of letters not usually available on computers. So, mappings from IPA to "computer-friendly" ASCII symbols as the ARPABET were proposed. The ARPABET (Fig. 1) is a selection of symbols used within the Advanced Research Projects Agency Speech Understanding Research (ARPA SUR) project [2]. It became popular among the speech recognition community. There are two representations in ARPABET, one adopts only one character and includes lower-case letters. The second uses only upper-case letters and is known as "2-characteres".

The alphabet used in the TIMIT database [3] (sometimes called TIMITBET) was inspired in ARPABET but is not the same. TIMITBET uses the 2-characteres representation of ARPABET for vowels. For example, when transcribing TIMIT, different labels were adopted for the closure and release parts of plosive sounds [3].

References

[1]   P. Ladefoged, *A Course in Phonetics*, 4 ed: Harcourt Brace, 2001.
[2]   J. E. Shoup, "Phonological Aspects of Speech Recognition," in *Trends in Speech Recognition*, W. A. LEA, Ed. Englewood Cliffs: Prentice Hall, 1980, pp. 125-138.
[3]   V. W. Zue and S. Seneff, "Transcription and alignment of the TIMIT database," presented at Proceedings of the Second Meeting on Advanced Man-Machine Interface through Spoken Language, 1988.

Phoneme	Computer Representation		Example	Phoneme	Computer Representation		Example
	1-Character	2-Characters			1-Character	2-Characters	
i	i	IY	beat	p	p	P	pet
I	I	IH	bit	t	t	T	ten
e	e	EY	bait	k	k	K	kit
ε	E	EH	bet	b	b	B	bet
æ	@	AE	bat	d	d	D	debt
ɑ	a	AA	Bob	g	g	G	get
Λ	A	AH	but	h	h	HH	hat
ɔ	c	AO	bought	f	f	F	fat
o	o	OW	boat	θ	T	TH	thing
U	U	UH	book	s	s	S	sat
u	u	UW	boot	š or ʃ	S	SH	shut
ə	x	AX	about	v	v	V	vat
ɨ	X	IX	roses	ð	D	DH	that
ɝ	R	ER	bird	z	z	Z	zoo
ɑU or ɑw	W	AW	down	ž or ʒ	Z	ZH	azure
ɑI or ɑy	Y	AY	buy	č	C	CH	church
ɔI or ɔy	O	OY	boy	ǰ	J	JH	judge
y	y	Y	you	ʍ	H	WH	which
w	w	W	wit	syl l, l̩	L	EL	battle ·
r	ɾ	R	rent	syl m, m̩	M	EM	bottom
l	l	L	let	syl n, n̩	N	EN	button
m	m	M	met	flapped t,ɾ	F	DX	batter
n	n	N	net	glottal stop, ʔ	Q	Q	
ŋ	G	NX	sing	Silence	–	–	
				non-speech Segment	!	!	laugh, etc.

AUXILIARY SYMBOLS (1- AND 2-CHARACTER CODES ARE IDENTICAL)			
Symbol	Meaning	Symbol	Meaning
+	Morpheme boundary	: 3 or .	Fall-rise or non-term juncture
/	Word boundary	* **	Comment (anything except * or **)
#	Utterance boundary	' '	Apos.-surround special symbol in comment
:	Tone group boundary	( )	Phoneme class information
:1 or .	Falling or decl. juncture	< >	Phonetic or allophonic escape
:2 or ?	Rising or inter. juncture		

STRESS REPRESENTATIONS (IF PRESENT, MUST IMMEDIATELY FOLLOW THE VOWEL)			
Value	Stress Assignment	Value	Stress Assignment
0	No stress	3	Tertiary stress
1	Primary stress	⋅	(Etc.)
2	Secondary Stress	⋮	

Fig. 1. ARPABET [2].