WIKIPEDIA

ARPABET

ARPABET (also spelled **ARPAbet**) is a set of phonetic transcription codes developed by Advanced Research Projects Agency (ARPA) as a part of their Speech Understanding Research project in the 1970s. It represents phonemes and allophones of General American English with distinct sequences of ASCII characters. Two systems, one representing each segment with one character (alternating upper- and lower-case letters) and the other with two or more (case-insensitive), were devised, the latter being far more widely adopted.[1]

ARPABET has been used in several speech synthesizers, including Computalker for the S-100 system, SAM for the Commodore 64, SAY for the Amiga, TextAssist for the PC and Speakeasy from Intelligent Artefacts which used the Votrax SC-01 speech synthesiser IC. It is also used in the CMU Pronouncing Dictionary. A revised version of ARPABET is used in the TIMIT corpus.[1][2]

Contents

Symbols

Stress is indicated by a digit immediately following a vowel. Auxiliary symbols are identical in 1- and 2-letter codes. In 2-letter notation, segments are separated by a space.

Vowels[3]

ARPABET		IPA	Example(s)
1-letter	2-letter		
a	AA	ɑ	b**a**lm, b**o**t
@	AE	æ	b**a**t
A	AH	ʌ	b**u**tt
c	AO	ɔ	st**o**ry
W	AW	aʊ	b**ou**t
x	AX	ə	comm**a**
N/A	AXR[4]	ɚ	lett**er**
Y	AY	aɪ	b**i**te
E	EH	ɛ	b**e**t
R	ER	ɝ	b**ir**d
e	EY	eɪ	b**ai**t
I	IH	ɪ	b**i**t
X	IX	ɨ	ros**e**s, rabb**i**t
i	IY	i	b**ea**t
o	OW	oʊ	b**oa**t
O	OY	ɔɪ	b**oy**
U	UH	ʊ	b**oo**k
u	UW	u	b**oo**t
N/A	UX[4]	ʉ	d**u**de

Consonants[3]

ARPABET		IPA	Example
1-letter	2-letter		
b	B	b	**b**uy
C	CH	ʧ	**Ch**ina
d	D	d	**d**ie
D	DH	ð	**th**y
F	DX	ɾ	bu**tt**er
L	EL	l̩	bott**le**
M	EM	m̩	rhyth**m**
N	EN	n̩	butt**on**
f	F	f	**f**ight
g	G	g	**g**uy
h	HH *or* H[4]	h	**h**igh
J	JH	ʤ	**j**ive
k	K	k	**k**ite
l	L	l	**l**ie
m	M	m	**m**y
n	N	n	**n**igh
G	NX *or* NG[4]	ŋ	si**ng**
N/A	NX[4]	ɾ̃	wi**nn**er
p	P	p	**p**ie
Q	Q	ʔ	uh**-**oh
r	R	ɹ	**r**ye
s	S	s	**s**igh
S	SH	ʃ	**sh**y
t	T	t	**t**ie
T	TH	θ	**th**igh
v	V	v	**v**ie
w	W	w	**w**ise
H	WH	ʍ	**wh**y
y	Y	j	**y**acht
z	Z	z	**z**oo
Z	ZH	ʒ	plea**s**ure

Stress and auxiliary symbols[3]

AB	Description
0	No stress
1	Primary stress
2	Secondary stress
3...	Tertiary and further stress
-	Silence
!	Non-speech segment
+	Morpheme boundary
/	Word boundary
#	Utterance boundary
:	Tone group boundary
:1 *or* .	Falling or declining juncture
:2 *or* ?	Rising or internal juncture
:3 *or* .	Fall-rise or non-terminal juncture

TIMIT

In TIMIT, the following symbols are used in addition to the ones listed above:[2][5]

Symbol	IPA	Example	Description
AX-H	ə̥	s**u**spect	Devoiced /ə/
BCL	b̆	o**b**tain	[b] closure
DCL	d̆	wi**d**th	[d] closure
ENG	ŋ̇	Wash**ing**ton	Syllabic [ŋ]
GCL	ğ	do**g**tooth	[g] closure
HV	ɦ	a**h**ead	Voiced /h/
KCL	k̆	do**c**tor	[k] closure
PCL	p̆	acce**p**t	[p] closure
TCL	t̆	ca**t**nip	[t] closure
PAU	N/A	N/A	Pause
EPI	N/A	N/A	Epenthetic silence
H#	N/A	N/A	Begin/end marker

See also

- Other ASCII phonetic codes:
  - SAMPA, language-specific

- X-SAMPA, encoding the whole International Phonetic Alphabet
- Worldbet
- Pronunciation respelling for English

References

1. Klautau, Aldebaro (2001). "ARPABET and the TIMIT alphabet" (https://web.archive.org/web/20160603180727/http://www.laps.ufpa.br/aldebaro/papers/ak_arpabet01.pdf) (PDF). Archived from the original (http://www.laps.ufpa.br/aldebaro/papers/ak_arpabet01.pdf) (PDF) on June 3, 2016. Retrieved September 8, 2017.
2. Lopes, Carla; Perdigão, Ferdinando (2011). "Phone Recognition on the TIMIT Database" (http://www.intechopen.com/books/speech-technologies/phoneme-recognition-on-the-timit-database). In Ipsic, Ivo (ed.). *Speech Technologies*. InTech. ISBN 978-953-307-996-7.
3. Rice, Lloyd (April 1976). "Hardware & software for speech synthesis" (https://archive.org/stream/dr_dobbs_journal_vol_01#page/n93). *Dr. Dobb's Journal of Computer Calisthenics & Orthodontia*. **1** (4): 6–8.
4. Jurafsky, Daniel; Martin, James H. (2000). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall. pp. 94–5. ISBN 0-1309-5069-6.
5. "Table of all the phonemic and phonetic symbols used in the TIMIT lexicon" (https://catalog.ldc.upenn.edu/docs/LDC93S1/PHONCODE.TXT). Linguistic Data Consortium. October 12, 1990. Retrieved September 8, 2017.

External links

- The CMU Pronouncing Dictionary (http://www.speech.cs.cmu.edu/cgi-bin/cmudict)