Speech Communication 3 (1984) 165–166
North-Holland

MACHINE-READABLE PHONETIC ALPHABET FOR ENGLISH AND FRENCH

Matthew LENNIG* and Jean-Paul BRASSARD
INRS-Télécommunications, 3 Place du Commerce, Ile des Soeurs, Verdun Québec, H3E 1H6 Canada

The purpose of this paper is to present a one-character, computer-readable phonetic alphabet suitable for both English and French. Since our laboratory is engaged in computational phonetic research on both these languages, we have a need for a computer phonetic alphabet (CPA) suitable for both. It is hoped that this paper will also stimulate discussion leading to a standard, multilingual CPA, which would facilitate international communication between laboratories using computational methods to solve phonetic problems.

The problem of developing a standard, machine-readable phonetic alphabet was first addressed by the ARPA/SUR project in the early 1970s, resulting in the ARPABET (Shoup, 1980). Two versions of the ARPABET exist: an all upper case, two-character version in which each character pair represents one phone, and a one-character, upper and lower case version. The all upper case ARPABET, which is used widely in speech laboratories, is difficult to read due to its use of paired symbols to represent single phones. The one-character version is more readable, but is not in widespread use. Neither version is well suited to languages other than English.

In designing the bilingual CPA proposed here, the following principles were followed:

1) Phone symbols should resemble as closely as possible the symbols used in the International Phonetic Alphabet (IPA). Thus, [j] is used for the initial semivowel of "**y**es" and [y] is used for the high-front-rounded vowel of French "b**u**".

2) A set of diacritical symbols is used to qualify phone symbols. The set of diacritical symbols is disjunct from the set of phone symbols. Diacritics are placed after the phone symbol they modify. Thus, transcriptions do not require separating spaces to be uniquely decodable.

3) Only two diacritical symbols are provided in this paper. The user is free to invent additional ones suitable for his application provided that new diacritical symbols do not correspond to existing phone symbols. Generally, letters are used for phone symbols while nonalphabetic characters are used for diacritics. For example, the diacritic symbol [:] could be used after a segmental symbol to indicate length.

4) Wherever possible, systematicity is attained by using upper and lower case pairs for phones having similar paradigmatic relationships. Thus, lower case vowel symbols represent closer vowels while upper case symbols represent more open vowels: high-mid [e] contrasts with low-mid [E], high-mid [o] with low-mid [O], tense [i] contrasts with lax and somewhat lower [I], tense [u] with lax and somewhat lower [U], etc.

Table 1 lists the CPA symbols, their IPA representations, and keywords illustrating their use in French and English. Keywords marked (Que) refer to Quebec French pronunciations of the indicated words.

With minor additions and modifications, CPA is general enough to apply to a wide range of languages. The authors invite suggestions from scientists working with other languages on the design of an international CPA using a standard set of symbols.

* M. Lennig is also with Bell-Northern Research, Ltd.

Table 1
Computer phonetic alphabet (CPA) compared with the international phonetic alphabet (IPA), including keywords for english and french

CPA	English keyword	French keyword	IPA
[i]	cr**ea**m	cr**i**	[i]
[I]	b**i**t	fi**ch**e (Que)	[ɪ]
[e]	b**ai**t	f**é**e	[e]
[E]	b**e**t	f**ai**te	[ɛ]
[3]	b**ir**d (r-less variant)		[ɜ]
[@]	b**a**t		[æ]
[a]		p**a**tte	[a]
[A]	f**a**ther	p**â**te	[ɑ]
[^]	b**u**t		[ʌ]
[u]	b**oo**t	c**ou**p	[u]
[U]	f**oo**t	t**ou**te (Que)	[ʊ]
[o]	b**oa**t	b**eau**	[o]
[O]	c**au**ght	n**o**te, f**o**rt	[ɔ]
[y]		v**u**	[y]
[Y]		b**u**tte (Que)	[ʏ]
[x]		f**eu**	[ø]
[X]		b**œu**f, fl**eu**r	[œ]
[*]	synth**e**size	quatr**e**	[ə]
[aj]	b**y**		[ɑj]
[aw]	c**ow**		[ɑw]
[Oj]	b**oy**		[ɔj]
[e~]		v**in** (Que)	[ẽ]
[E~]		v**in**	[ɛ̃]
[a~]		v**en**t (Que)	[ã]
[A~]		v**en**t	[ɑ̃]
[o~]		p**on**t	[õ]
[X~]		br**un**	[œ̃]
[j]	**y**ank	mai**ll**ot	[j]
[H]		h**ui**t	[ɥ]
[w]	**w**ick	**ou**i	[w]
[hw]	**wh**ich		[hw]
[l]	**l**ap	**l**it	[l]
[r]	**r**ap (retroflex)	**r**ond (apical)	[r]
[R]		**r**ond (uvular)	[R]
[m]	**m**ap	**m**ont	[m]
[n]	**n**ip	**n**id	[n]
[l´]	bott**le** (syllabic)		[l̩]
[r´]	b**ir**d, heat**er** (syllabic)		[ɚ]
[m´]	bott**om** (syllabic)		[m̩]
[n´]	butt**on** (syllabic)		[n̩]
[G]		oi**gn**on	[ɲ]
[g~]	si**ng**	campi**ng**	[ŋ]
[f]	**f**oe	**f**ait	[f]
[v]	**v**ery	**v**ie	[v]
[T]	**th**in		[θ]
[D]	**th**ey		[ð]
[s]	**s**it	**s**ou	[s]
[z]	**z**ip	bi**s**e	[z]
[S]	**ch**ute	**ch**amps	[ʃ]
[Z]	vi**s**ion	**j**e	[ʒ]
[h]	**h**it	**h**a! **h**a!	[h]
[p]	**p**it	**p**ont	[p]
[b]	**b**ond	**b**on	[b]
[t]	**t**ea	**t**on	[t]
[d]	**d**ip	**d**onc	[d]
[k]	**c**ake	**c**ape	[k]
[g]	**g**ive	**g**ant	[g]
[tS]	**ch**eek		[tʃ]
[dZ]	**j**eep		[dʒ]
[?]	(glottal stop)		[ʔ]
[-]	(silence)		

References

[1] *Principles of the International Phonetic Association*, 1949. International Phonetic Association, University College, Gower Street, London WC1E 6BT, England.

[2] Shoup, June E. 1980. Phonological aspects of speech recognition. Wayne A. Lea, ed., *Trends in Speech Recognition*, Englewood Cliffs, NJ: Prentice-Hall, p. 127.