

Testing the learnability of writing systems

SHARON INKELAS, KEITH JOHNSON, CHARLES LEE, EMIL MINAS,
GEORGE MULCAIRE, GEK YONG KENG, TOMOMI YUASA
University of California, Berkeley

1. Introduction

The world's sound-based writing systems differ according to the size of the typical speech chunk which is mapped to a symbol: the phone, in so-called alphabetic writing systems, and the mora, demisyllable or syllable, in so-called syllabaries. This paper reports the results of an artificial learning study designed to test whether the acoustic stability of the speech chunks mapped to symbols is a factor in subjects' ability to learn a novel writing system.

2. Background

Sound-based writing systems in the world's languages are typically classified as syllabaries or alphabetic systems (see e.g. Sampson 1990, Daniels & Bright 1996, Coulmas 2003). In true syllabaries, symbols represent possible syllables (V, CV, VC, CVC) and are atomic, not analyzable into combinations of segment-sized symbols. Japanese kana and Cree orthography, for example, use distinct symbols for speech chunks like [ka], [ki], [ku], [ke], [ko], [ta], [ti], [tu], [te], [to], etc. (see e.g. Smith 1996:211, Nichols 1996:600). In a true alphabetic system, each symbol would correspond to a single phone. Spanish is a relatively good example of an alphabetic system, though, like most alphabetic systems of any vintage, departs from canonicity (e.g. the silent "h", digraphs such as "ch", "qu").

Syllabaries are far more common, among independently evolved writing systems, than alphabets. However, most syllabaries are impure, either incorporating alphabetic symbols in some cases or providing too few symbols to accurately transcribe all of the possible words in the language.

Cherokee is an example of an impure syllabary. Symbols in the Cherokee writing system generally correspond either to V or to CV, which are possible syllables in the language. There is also a symbol for [s], used to write syllables beginning with sC clusters. However, Cherokee has syllables beginning with

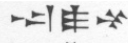
other consonant clusters than cannot be accurately transcribed using the symbols in the inventory, below (Scancarelli 1996):

D	a	R	e	T	i	ᵔ	o	ᵕ	u	i	v
ᶑ	ga	ᶑ	ge	ᶑ	gi	A	go	J	gu	E	gv
ᵕ	ka										
ᵕ	ha	ᶑ	he	ᶑ	hi	ᶑ	ho	ᶑ	hu	ᶑ	hv
W	la	ᵔ	le	ᶑ	li	ᶑ	lo	M	lu	ᶑ	lv
ᶑ	ma	ᶑ	me	H	mi	ᶑ	mo	ᶑ	mu		
ᵕ	na	ᶑ	ne	ᶑ	ni	Z	no	ᶑ	nu	ᵕ	nv
ᶑ	hna										
G	nah										
ᶑ	qua	ᶑ	que	ᶑ	qui	ᶑ	quo	ᶑ	quu	ᶑ	quv
ᵕ	s	ᶑ	se	ᶑ	si	ᶑ	so	ᶑ	su	R	sv
U	sa										
ᶑ	da	ᶑ	de	ᶑ	di	V	do	S	du	ᶑ	dv
W	ta	ᶑ	te	ᶑ	ti						
ᶑ	dla	L	tle	C	tli	ᶑ	tlo	ᶑ	tlu	P	tlv
ᶑ	tla										
G	tsa	ᶑ	tse	ᶑ	tsi	K	tso	ᶑ	tsu	ᶑ	tsv
G	wa	ᶑ	we	ᵕ	wi	ᵕ	wo	ᶑ	wu	ᶑ	wv
ᶑ	ya	ᶑ	ye	ᶑ	yi	ᶑ	yo	ᶑ	yu	B	yv

Figure 1: Cherokee syllabary

Consonant clusters not involving [s] are written with CV symbols one of whose vowels is not pronounced. This is a common strategy in syllabaries. An example is given in (1a), from Scancarelli 1996:520, in which [li]+[s] = [ls]:

- (1) a. W-ᶑ-ᵕ-A-ᶑ
 ta-li-s-go-hi
 [t^haʔlsko:hi]
 ‘twenty’

- b. 
 ba-aš-mu
 bašmu
 ‘pertains’

Another example of how syllabaries accommodate complex syllables comes from Akkadian cuneiform, in which symbols represent V, VC and CV, and a subset of the possible CVC syllables in the language. As shown in (1b), above, from Cooper (1996:56), CV_i and V_iC could combine into a CV_iC syllable: [ba] + [aš] = [baš].

For a syllabary to be able to transcribe all possible words in a language with complex syllables, the number of symbols would have to be very large. Compromises like those in Cherokee and Akkadian cuneiform avoid this explosion, yet render the system imperfect. Why, given this frequent need to incorporate alphabetic symbols or compromise on accuracy, are syllabaries so common among independently evolved writing systems? Could there be a phonetic or phonological reason for this?

Our hypothesis, based on earlier observations (see e.g. Daniels 2009), is that symbols mapping to CV or VC speech chunks have a clear advantage over symbols mapping to phones. CV and VC speech chunks, in which each segment is cued by the other, have greater motor-acoustic stability than individual phones. The instability of single segments in psychoacoustic space has also been recognized by specialists in literacy, some of whom have advocated for teaching writing syllable-sized chunks:

“[Two] critical cognitive problems normally are confounded in reading instruction. The first is learning that English orthography directly maps sound rather than meaning, and the second is learning that orthographic units correspond to highly abstract and inaccessible phonological segments that must be blended to form words... The barrier to acquisition of alphabetic units appears to be purely psychoacoustic: the child has difficulty in segmenting the sound-stream into phonemic chunks and therefore cannot map the discrete alphabetic units onto equivalently discrete speech units.” (Gleitman & Rozin (1973:479)

Gleitman & Rozin used the above observation as the jumping-off point for a study in which young children who were struggling to learn to read in school were successfully taught writing systems in which symbols corresponded to syllables, as used in (2) to write the phrase “Candy for Andy”:

(2)

☺-D 4 AND-D

3. Hypotheses

Our study compares the learnability of four different possible writing systems:

- (3) Segmental: one symbol per consonant, vowel
 Mora: one symbol for each CV, one for each (C) coda
 Onset-Rime: one symbol for each (C) onset, one for each (VC) rime
 Demisyllable: one symbol for each CV, VC

Each of these systems is phonologically motivated. The role of the segment in phonological patterns is uncontested. The role of the mora is strongly supported in languages like Japanese, where the mora governs syllable size, accent placement, and a myriad of prosodic morphology constructions (Hyman 1985, Hayes 1986). The role of the onset and rime is clearly supported by cross-linguistic patterns of poetic rhyme (Kiparsky 1981, Selkirk 1984).

The demisyllable is the only system that does not correspond to syllable-internal constituency commonly assumed by phonologists, though see Fujimura 1989, Ito & Mester 1995. However, the use of diphones is widespread in speech recognition and synthesis, and phoneticians have long recognized the key role that CV and VC transitions play in the production and perception of speech.

These possible systems are also all attested in the world's writing systems. English, with its digraphs and its many-to-one mappings between symbol and sound, is a highly imperfect example of an alphabetic system.

Japanese kana is an example of a moraic system. In Japanese, the maximal syllable is CVX, where X stands for a moraic nasal, the first half of a geminate consonant, the second half of a diphthong, or the second half of a long vowel. (C)V is one mora; X is another. There is a separate symbol for each CV chunk. The 'syllabic' nasal has its own symbol. Consonant gemination and vowel length are indicated with symbols of their own. (See e.g. Smith 1996 for an overview.)

Bopomofo, used to phonetically render Chinese characters in Taiwan, is one of the few examples of an onset-rime system. In Bopomofo, one set of symbols represent initial consonants, while another set represents possible syllable rimes (see e.g. Mair 2006:204-205).

Akkadian cuneiform exemplifies what a demisyllabic writing system might look like, though it also contains symbols for segments and CVC syllables.

Clearly, all four systems are possible ways for a human to segment and transcribe a word. But which is best? Our study tests the Acoustic Stability hypothesis, stated in (4), which ranks the four systems as indicated below:

- (4) ACOUSTIC STABILITY: symbols mapping to acoustically stable speech chunks, like VC or CV, will be learned more accurately than those mapping to less stable speech chunks, like V or C.

Prediction: Demisyllable > Mora, Onset-Rime > Segment

Given that our subjects are all familiar with the alphabetic system of English, however, it is possible that they will exhibit a bias for the systems whose symbols are most similar in their sound-to-symbol mapping to English. Thus we also test

the alternative hypothesis, in (5), which predicts a different ranking among the four systems being compared:

- (5) ENGLISH ORTHOGRAPHY BIAS: subjects will be biased to learn new systems which are similar to the orthographies they already know. In the case of English speakers (those who participated in this study), that bias is for an alphabetic system.

Prediction: Segment > Mora, Onset-Rime > Demisyllable

4. Study procedures

In this computer-based study, subjects learned a set of symbols through exposure and assessment. Subjects were then trained, with feedback, on how to combine symbols into CVC words. Finally, subjects were tested learners on their ability to read aloud new CVC words written with the symbols they learned.

4.1 Subjects

Subjects were recruited from undergraduate classes at UC Berkeley. All were literate in English; some were also literate in other languages, including some with syllabaries (Japanese) or logographic (Chinese) writing systems. Subjects sat at a computer station wearing a head-mounted mike for the procedure, which took between 30 and 45 minutes.

80 subjects completed the study, distributed over the conditions as follows:

- (6) Segment: 20 Mora: 22 Onset-Rime: 19 Demisyllable: 19

4.2 Symbols and symbol learning

All subjects were taught the same 20 symbols, shown in 2:

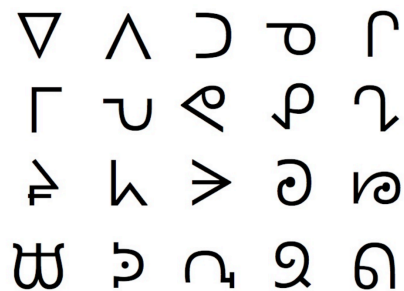


Figure 2: Symbol set used in study

These symbols come from the (considerably larger) Cree syllabary. They were selected on the basis of a pilot study with several English-speaking subjects, who rated the full set of Cree symbols on their confusability with one another and on their resemblance to letters in the Latin alphabet. The symbols selected for this study were optimally dispersed in terms of graphical similarity and did not strongly evoke letters in English orthography.

Symbols were associated with the speech chunks shown in Figure 3. (It was not possible to keep the number of symbols and the number of phones constant across the three conditions. Priority was given to controlling the number of symbols that subjects had to learn.)

Segment			Onset-Rime		Mora		Demisyllable	
C1	V	C2	C	VC	CV	C	CV	VC
g	i	k	g	ik	gi	k	gi	ik
b	a	m	b	itʃ	bi	tʃ	bi	itʃ
z	o	tʃ	z	im	zi	m	zi	im
h	ε	s	h	am	ha	n	ha	am
ɕ		t	ɕ	as	ɕa	s	ɕa	as
w		n	w	ak	wa	t	wa	ak
d		p	d	il	ɕi	f	ɕi	il
v			v	is	wi	l	wi	is
ɹ			j	atʃ	ga	ʃ	ga	atʃ
			ɹ	ap	za	p	za	ap

Figure 3: Speech chunks associated with symbols in the four conditions

Speech chunks were recorded in a sound-protected booth by a female speaker of American English. All final consonants except fricatives were audibly released into a very short (around 10 ms.) schwa-like vowel.

A randomizer was used to determine the mapping between speech chunk (V, C, CV, or VC) and symbol. For each condition, ten different randomizations were used. The purpose of randomization was to minimize the effects of any sound-symbol association bias on the part of individual subjects.

Subjects were introduced to individual symbols by being shown a symbol on the computer screen while hearing its corresponding sound value. Subjects were presented with symbols in incremental blocks. The first block contained four symbols. In the next block, four more were added. Successive blocks added three symbols each. The sequence of the symbols shown within a block was

randomized. Sampling (within a block) was done with replacement, so that the same symbol might appear multiple times within one block.

Each block of symbols was presented twice, followed by an assessment to determine how accurately subject had learned the symbols. Assessment consisted of subjects choosing, from a grid of all 20 symbols, the symbol corresponding to the sound they heard through the speaker. The location of symbols in the grid changed with each assessment so that subjects were being tested on their memory of the symbol, not its location. A subject who scored 80% or higher on an assessment was permitted to move to the next block. Otherwise, the subject repeated the block and was assessed again. In order to stay within the one-hour time constraint, a subject was moved to the next block after 4 assessments, even with a score below 80%. The minimum number of assessment blocks an individual subject could experience was 6; the maximum was 24.

4.3 Combination training

Following the symbol training phase, subjects were trained on how to combine symbols into CVC “words”. In order not to expose subjects during combination training to any of the CVC combinations used later in the testing phase, the training phase combined each of the ‘test’ sounds with a ‘filler’ sound (Figure 4):

	Segment			Onset-Rime		Mora		Demisyllable	
	C1	V	C2	C	VC	CV	C	CV	VC
Test sounds	g	i	k	g	ik	gi	k	gi	ik
	b	a	m	b	itʃ	bi	tʃ	bi	itʃ
	z		tʃ	z	im	zi	m	zi	im
	h		s	h	am	ha	n	ha	am
	ɕ			ɕ	as	ɕa	s	ɕa	as
	w			w	ak	wa	t	wa	ak
Filler sounds	d	o	t	d	il	ɕi	f	ɕi	il
	v	ε	n	v	is	wi	l	wi	is
	ɹ		p	j	atʃ	ga	ʃ	ga	atʃ
				ɹ	ap	za	p	za	ap

Figure 4: Speech chunks associated with symbols in the four conditions

The training items are presented in Figure 5. The set of training items varies across the conditions as a consequence of the need to introduce each test symbol in combination with a filler symbol (as in Figure 4):

Segment			
got	zet	ɕop	dit
gek	zok	ɕotʃ	dap
gom	zos	ɕes	vat
bən	hon	wən	vin
betʃ	hem	wek	ran
bos	hetʃ	wom	rip

Onset-Rime			
gil	gis	dik	vik
bil	bis	ditʃ	vitʃ
zil	zis	dim	vim
haʃ	hap	jam	ram
ɕaʃ	ɕap	jas	ras
waʃ	wap	jak	rak

Mora			
gif	gil	ɕik	wik
bif	bil	ɕitʃ	witʃ
zif	zil	ɕin	win
haʃ	hap	gam	zam
ɕaʃ	ɕap	gas	zas
waʃ	wap	gat	zat

Demisyllable			
gil	gis	ɕik	wik
bil	bis	ɕitʃ	witʃ
zil	zis	ɕim	wim
haʃ	hap	gam	zam
ɕaʃ	ɕap	gas	zas
waʃ	wap	gak	zak

Figure 5: Training items

Subjects had five seconds from the moment of exposure to read the CVC word aloud. The correct recording of the word was subsequently played aloud for the subject as guidance, providing feedback and reinforcement of symbol training.

4.4 Testing

After symbol learning and combination training, subjects were tested on their ability to read aloud novel CVC words constructed from the symbols they had learned. In the testing phase, subjects were shown 18 novel symbol combinations whose phonetic transcriptions are given in Figure 6:

gik	gitʃ	gim	ham	has	hak
bik	bitʃ	bim	ɕam	ɕas	ɕak
zik	zitʃ	zim	wam	was	wak

Figure 6: test items

Subjects were instructed to read the word in the five-second recording time frame, but were not given guidance on their level of correctness. The information from these final recordings was coded for accuracy and reaction time.

5. Results

The accuracy data support the Acoustic stability hypothesis over the Alphabetic familiarity hypothesis. Figure 7 illustrates performance on the four conditions, measured by overall correctness on the 18 test items. Participants in the Demisyllable condition were most accurate (86%) than those in the Segment condition (73%), with Mora (76%) and Onset-Rime (81%) in between. The difference between Demisyllable and Segment did not, however, reach statistical significance ($p < .096$ on a 2-tailed t-test).

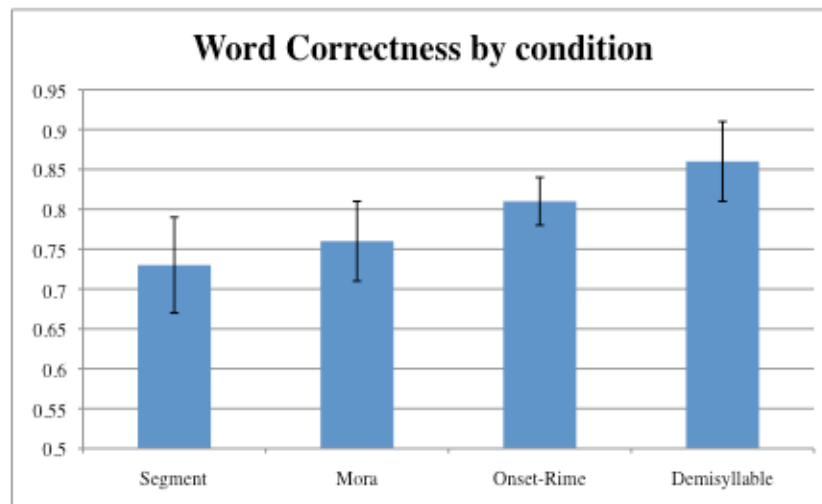


Figure 7

Examination of accuracy at the segmental level revealed a similar distribution. Subjects in the Demisyllable condition were the most accurate, overall, on each individual segment in the test words (92%) vs. subjects in the Segment condition (87%) (Figure 8):

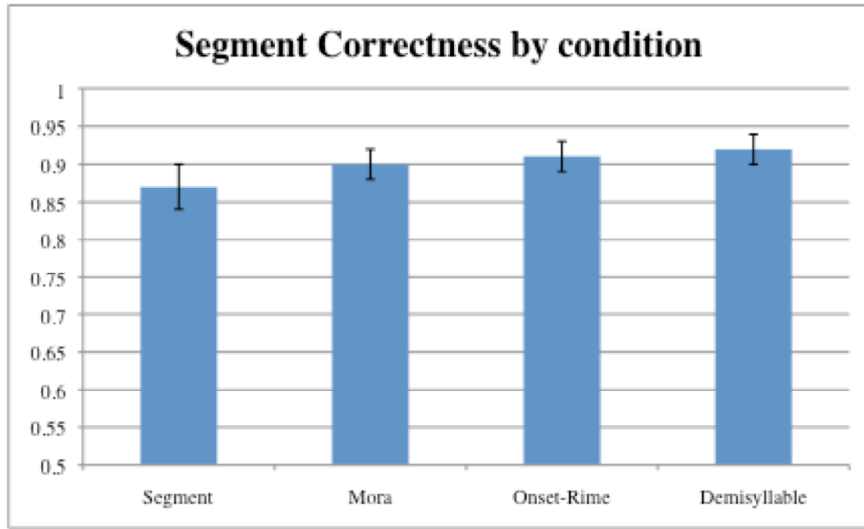


Figure 8

Overall segment accuracy differences across conditions did not reach statistical significance. However, significant differences did appear when vowel and consonant accuracy were examined separately. Vowel accuracy was 94% in the Demisyllable condition vs. only 81% in the Segment condition ($p < .05$ on a 2-tailed t-test). Consonant accuracy did not differ across the four conditions:

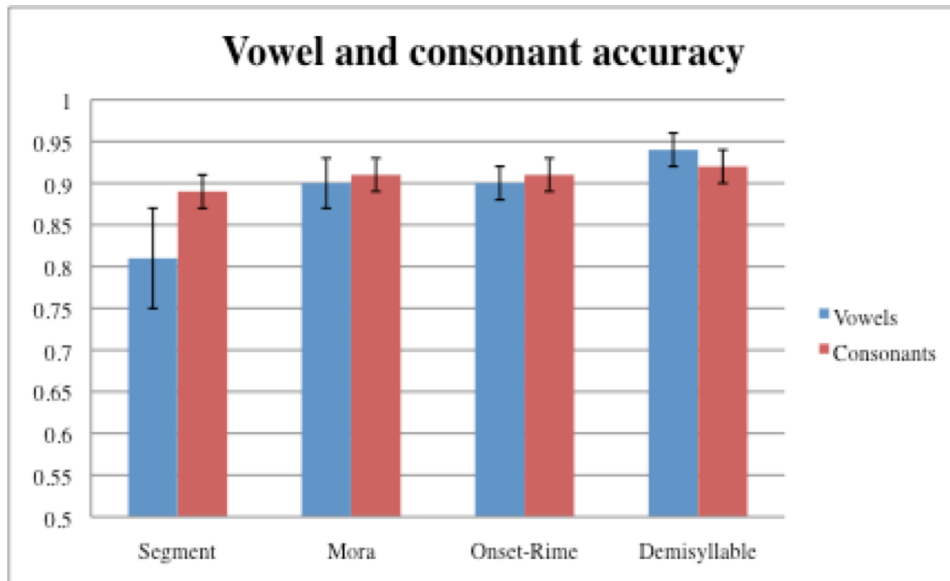


Figure 9

Learning the mapping between a vowel phone and a symbol appears to be significantly enhanced when the vowel is part of a VC or CV chunk in which the transitions provide cues to vowel quality. This is consistent with the Acoustic Stability hypothesis. (Recall that individual consonants, other than fricatives, were produced with a vocalic release in the audio stimuli; this may explain why consonant accuracy did not vary across the conditions. Perhaps all consonants derived sufficient benefit from vocalic cues.) In conclusion, the accuracy results support the Acoustic Stability hypothesis over the Alphabetic Familiarity hypothesis.

6. Discussion

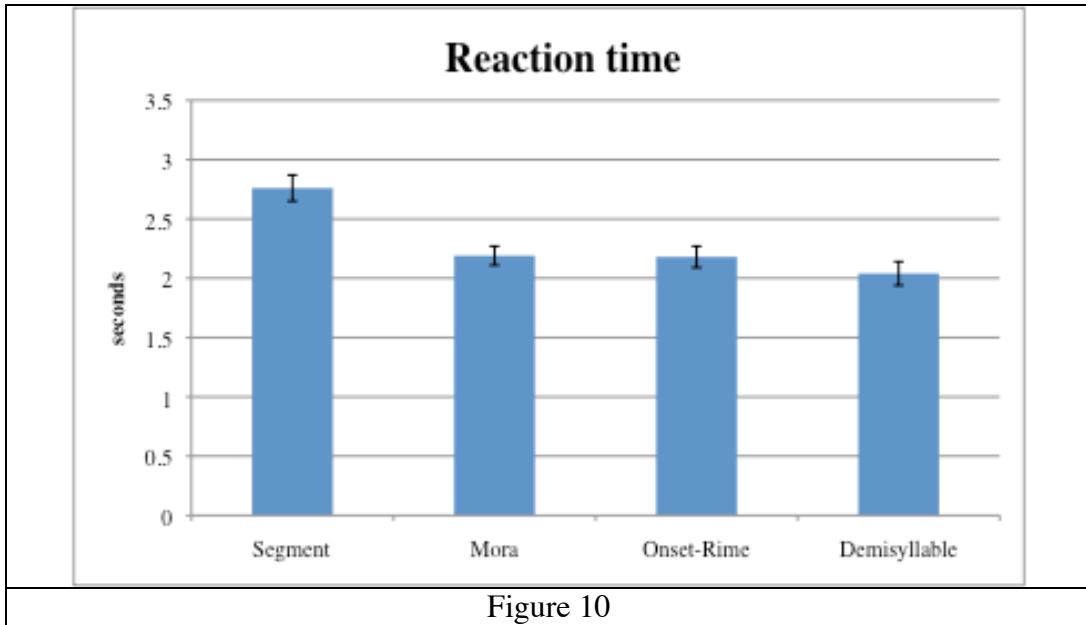
Two confounding factors must be addressed before accepting the conclusion that the Acoustic Stability hypothesis has been supported: symbol count and phonetic space. Each predicts that the Segment condition will be more difficult for subjects and that the other three will be equally difficult.

6.1 Symbol count

While the study controlled the number of symbols that subjects had to learn, symbol count in combination training and testing items differed across the three conditions. The Segment condition required subjects to process strings of three symbols, while the other conditions used only two symbols to write CVC words. If symbol count is a factor, then subjects would be predicted to perform less well in the Segment condition than in the other three conditions.

As seen, subjects did perform less accurately in the Segment condition. However, we also observed differences in performance across Mora, Onset-Rime and Demisyllable conditions that cannot be attributed to symbol count.

What does correlate particularly well with symbol count is reaction time (RT). We measured the time each subject took to start speaking in the testing phase. (Sometimes subjects provided two responses — an original and a correction — in the time allotted; in such cases we did not include RT in the averages reported below.) RT for the segment condition, at 2.76 seconds, was considerably higher than RT in the other three conditions (2.19 seconds for the Mora condition, 2.18 for the Onset-Rime condition, and 2.04 for the Demisyllable condition). Pairwise comparisons in RT between Segment and each of the other three conditions were significant ($p < .01$, according to a 2-tailed t-test).



It is of course not surprising that subjects would take longer to compute and read aloud a string of 3 symbols, vs. a string of 2 symbols.

6.2 Phonetic space

The four conditions differed in the phonetic space defined by the set of phones used in the speech chunks mapping to symbols. While the set of phones employed in testing was identical (because the test items were identical) across conditions, the conditions differed in the number and type of filler symbols needed to introduce all of the symbols in combination training while avoiding any combinations used in testing. Figure 13 lists the phone sets required:

	C1							V				C2												
	g	b	z	h	ʒ	w	d	v	r	j	a	i	o	ε	k	tʃ	m	s	t	p	n	l	f	ʃ
Demi	█											█				█								
Mora	█											█				█								
Onset-Rime	█											█				█								
Segment	█											█				█								

Figure 11

The size of the phonetic space that a set of symbols could be used to describe might be expected to correlate positively with difficulty of learning the system. Phonetic space can be computed either additively or multiplicatively. Either

method produces a scale in which the Demisyllable condition has the smallest phonetic space and the Segment condition has the largest.

(7) Phonetic space

	Additive (# of phones)	Multiplicative (# of CVC words constructable from phone set)
Demi	14	$6*2*6 = 72$
Mora	18	$6*2*10 = 120$
Onset-Rime	18	$10*2*6 = 120$
Segment	20	$9*4*7 = 252$

The resulting ordering is consistent with the ordering predicted by the Acoustic Stability hypothesis, creating doubt as to which factor is the most explanatory: Acoustic Stability, phonetic space, or (as discussed just above) symbol count?

We can address this question by focusing on vowels. As observed earlier, vowel accuracy is the dimension on which subjects differed most sharply across the conditions. The figure below plots vowel accuracy by (additive) vowel phonetic space. Subjects in the Segment condition, with 4 vowel phones, displayed 94% phonetic accuracy on vowels in testing, while subjects in the other three conditions, all of which had 2 vowel phones, were accurate on 89% of their vowels. This difference is significant ($p < .01$).

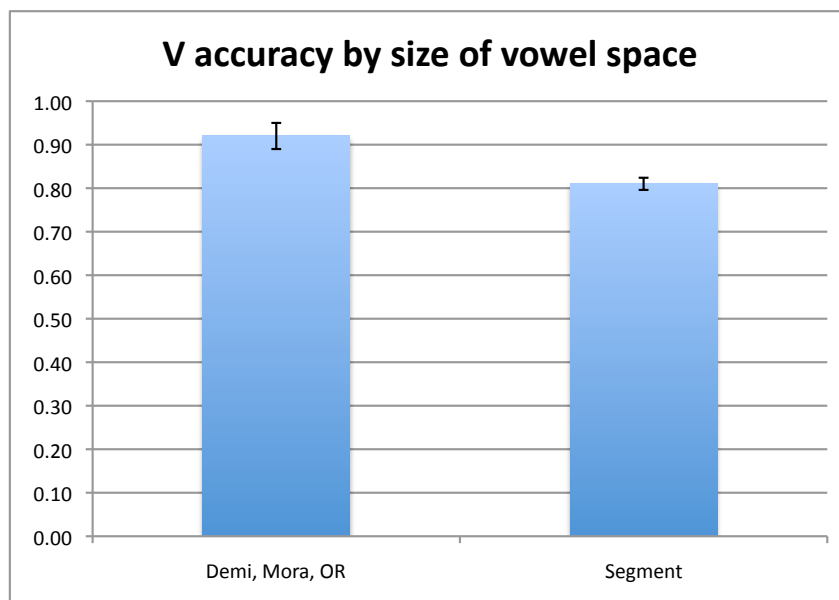


Figure 12

Vowel phone count correlates with symbol count (in training and testing items) in how it divides the conditions into groups. This makes it difficult to know whether

vowel phone count, symbol count or acoustic (in)stability is the cause of the observed poorer accuracy of subjects in the Segment conditions.

However, only the Acoustic Stability hypothesis predicts differences among Mora, Onset-Rime and Demisyllable conditions. Our results do show differences, but they are not statistically significant. Follow-up studies with more statistical power are needed to illuminate these suggestive findings.

7. Conclusions and implications

The findings from this study show that despite the inevitable bias in favor of alphabetic systems our subjects brought to the study, they performed better at learning nonalphabetic writing systems. Our results clearly show that subjects performed better in the Demisyllable condition than in the Segment condition. Whether this is due to acoustic stability of symbols (our hypothesis) or to the confounding factors of symbol count or phonetic space, the implications for literacy teaching are clear: people will learn a writing system better if the symbols are presented to them in speech chunks larger than the individual phone.

References

- Cooper, Jerrold S. 1996. Sumerian and Akkadian. In P. Daniels and W. Bright (eds.), 37-57.
- Coulmas, Florian. 2003. *Writing Systems*. Cambridge: Cambridge University Press.
- Daniels, Peter. 2009. Grammatology. In N. Torrance and D. Olson (eds.), *The Cambridge Handbook of Literacy*, 25-45, Cambridge University Press.
- Daniels, Peter and William Bright (eds.). 1996. *The World's Writing Systems*, New York: Oxford University Press.
- Fujimura, Osamu. 1989. Demisyllables as sets of features: comments on Clements's paper. In J. Kingston and M. Beckman (eds.), *Papers in Laboratory Phonology 1: Between the Grammar and the Physics of Speech*, 334-340, Cambridge University Press.
- Gleitman, Lila R. and Paul Rozin. 1973. Teaching reading by use of a syllabary. *Reading Research Quarterly* 8:447-483.
- Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20:253-306.

- Hyman, Larry M. 1985. *A theory of phonological weight*. Dordrecht: Foris.
- Itô, Junko and Armin Mester. 1995. Japanese phonology. In J. Goldsmith (ed.), *A handbook of phonological theory*, 817-838, Blackwell.
- Kiparsky, Paul. 1981. Remarks on the metrical structure of the syllable. In W. Dressler, O. Pfeiffer, and J. Rennison (eds.), *Phonologica 1980: Akten der Vierten Internationalen Phonologie-Tagung*, 245-256, Innsbruck: Institute für Sprachwissenschaft der Universität Innsbruck.
- Mair, Victor. 1996. Modern Chinese writing. In P. Daniels and W. Bright (eds.), 200-208.
- Nichols, John D. 1996. The Cree syllabary. In P. Daniels and W. Bright (eds.), 599-611.
- Sampson, Geoffrey. 1990. *Writing systems: a linguistic introduction*. Stanford: Stanford University Press.
- Scancarelli, Janine. 1996. Cherokee writing. In P. Daniels and W. Bright (eds.), 587-592.
- Selkirk, Elisabeth. 1984. On the major class features and syllable theory. In M. Aronoff and R.T. Oerhle, eds., *Language Sound Structure: Studies in Phonology Dedicated to Morris Halle by his Teacher and Students*. Cambridge: MIT Press, 107-113.
- Smith, Janet S. 1996. Japanese writing. In P. Daniels and W. Bright (eds.), 209-217.