Acoustic transitions in Khmer word-initial clusters

James P. Kirby

Linguistics and English Language, University of Edinburgh, Scotland, UK

j.kirby@ed.ac.uk

Abstract

Onset clusters in Khmer (Cambodian) often appear with an acoustic transition between consonants, but the phonological status of these elements is indeterminate. If transitions result from gestural separation, they may disappear in fast speech. Acoustic analysis of data from 10 speakers shows that vocalic transitions in Khmer are found in largely predictable set of consonantal contexts. While their presence is modulated by speech rate, they never disappear completely, in some cases becoming more rather than less frequent in fast speech. Clusters containing transitions are generally longer in duration than those that do not, and are also longer than monosyllables containing a lexical schwas, but the transitions do not show any spectral evidence of a distinct gestural target. The possible interpretations of these findings are discussed in the context of the range of articulatory variation known to occur in the implementation of speech rate.

Keywords: Khmer, intrusive vowel, speech rate, onset cluster

1. Introduction

The presence of excrescent, intrusive or otherwise 'transitional' elements in consonant clusters is common in many of the world's languages. However, research suggests that not all of these elements have the same status: in some cases they are argued to be epenthetic, involving the insertion of some type of phonological unit (e.g., Dutch: Warner et al., 2001) while in other languages they are claimed to be fundamentally phonetic effects, e.g. byproducts of gestural retiming (e.g., German: Jannedy et al., 2008). Teasing these two possibilities apart is not always straightforward: for some languages, evidence in favour of both positions has been advanced (e.g., Tashlhiyt Berber: Coleman, 2001; Ridouane & Fougeron, 2011).

Khmer, the national language of Cambodia, has a rich inventory of no less than 70 onset clusters (Table 1). Huffman (1972) proposed that Khmer onsets can be separated into three classes based on the type of transition they contain: (a) 'Class 3' clusters such as /kd/ and /lp/, which are separated by a brief voiced transition (hereafter abbreviated ∃) ; (b) 'Class 2' clusters such as /pl/ and /tk/, which are separated by a brief voiceless transition (i.e., aspiration); and (c) 'Class 1' clusters such as /pr/ and /ks/, which are articulated without audible transitions. This suggests that the set of conditions giving rise to intrusive vowels in Khmer may differ from that documented for languages such as Tashlhiyt Berber (Ridouane & Fougeron, 2011).

Recently, Butler (2012, 2014) conducted an acoustic study of 20 Khmer onset clusters. Her data suggest that the presence of an acoustic transition can vary with cluster type: for some clusters (e.g., /lb/) ∃ is consistently present, while others (e.g., /mt/) may occur without an audible transition. Butler also compared spectral and durational properties of transitional vocoids

Table 1: *Onset clusters in Khmer (after Huffman, 1972). Class 1 are dark, Class 2 are light, and Class 3 are not shaded.*

C_1	C_2														
↓	p	t	tʃ	k	ʔ	b	d	m	n	ɲ	ŋ	l	r	s	h
p		pt	ptʃ	pk	pʔ		pd		pn	pɲ	pŋ	pl	pr	ps	ph
t	tp			tk	tʔ	tb		tm	tn		tŋ	tl	tr		th
tʃ	tʃp			tʃk	tʃʔ	tʃb	tʃd	tʃm	tʃn		tʃŋ	tʃl			tʃh
k	kp	kt	ktʃ		kʔ	kb	kd	km	kn	kɲ	kŋ	kl	kr	ks	kh
s	sp	st		sk	sʔ	sb	sd	sm	sn	sɲ	sŋ	sl	sr		
m		mt	mtʃ		mʔ		md		mn	mɲ		ml	mr	ms	mh
l	lp			lk	lʔ	lb		lm			lŋ				lh

in words like ម្ទេស /mteːh/ > [mˀteːh] 'pepper' to short lexical vowels in words like មីត /mɨt/ 'dash away'. She reports that clusters containing transitional vocoids were significantly shorter than monosyllables with lexical short vowels, and the formant structure of the vocoids was more variable than that of corresponding lexical vowels. Butler interprets her findings in the context of Articulatory Phonology (Browman & Goldstein, 1992) as indicating that transitional elements in Khmer onset clusters lack an associated articulatory gesture and are simply surface-level effects of the phasing of the consonant gestures, as has been argued for languages like Tashlhiyt or German.

However, the fact that transitional vocoids differ acoustically from lexical short vowels does not necessarily mean that these elements do not involve a phonological target of some kind (gesture or segment), only that this target is not the same as that of a nuclear vowel in a lexical monosyllable. Many Khmer monosyllables with CC- onset clusters are known to have developed from the loss of the initial syllable of disyllabic iambs, e.g. ប្ដី /pdəj/ 'husband' < Sanskrit *pati*, ឆ្កែ /ckae/ 'dog' < Old Khmer *cake* (Huffman, 1972, p. 61) as part of a historical process common to many Southeast Asian languages (Matisoff, 2003). It is thus conceivable that these elements are still part of the phonological specification of lexical items where they (predictably) appear, or that they are historically intrusive but have become lexicalized with their own unique gestural target(s).

Here, I look for additional evidence bearing on the phonological status of acoustic transitions in Khmer onset clusters by considering acoustic data from 35 cluster types in the context of a speech rate manipulation. Disappearance of intrusive or excrescent vowels in fast speech has been observed in a number of languages (see Hall, 2006 for a review), arguably as the result of an increase in the relative overlap of extant articulatory gestures (Gay, 1981; Munhall & Löfqvist, 1992; Byrd & Tan, 1996). If acoustic transitions in Khmer onset clusters are found to disappear in fast speech, this would lend further support to the idea that they are phonetically excrescent and do not involve phonological specification of a segment or gesture.

2. Method

2.1. Materials and subjects

10 native Khmer speakers (3 female) of the Phnom Penh dialect were recorded reading 79 lexical items 3 times each at two self selected speech rates. Participants were instructed to produce the first set of items slowly and carefully, and the second set as rapidly as possible. All items in both conditions were embedded in the frame sentence /kɲom thaː ___ tiət/ 'I say ___ again', and randomized and counterbalanced across blocks and participants. Here, we focus on the 61 items that include one of 35 onset clusters /pd ph pk pl pn pr ps pt tb th tk tl tn tŋ tp tr kd kh kl kn kŋ kp kr ks kt lb lh lp md mh ml mn mr ms mt/. Rime qualities varied but all items were of the shape $C_1C_2V(V)(C)$.[1]

2.2. Segmentation

Acoustic data were segmented and annotated using Praat 5.3.57 (Boersma & Weenink, 2013) to determine total syllable duration along with durations of C_1 release, C_2 closure and release, transitional vocoid (if present), and syllable rime. A combination of spectral and acoustic cues was used to determine presence of a transitional vocoid, including presence of a periodic waveform, increase in signal energy at C_1 release, and/or a region of formant structure (see Figure 1).

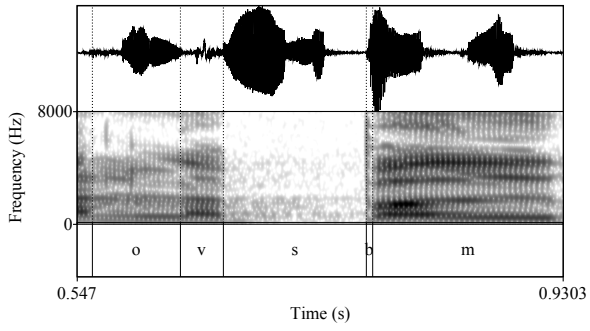

Figure 1: *Segmentation of token* ល្ពៅ */lpiv/ 'pumpkin', speaker 6, slow condition, repetition 3 showing C_1 (o)nset, (v)ocoid, C_2 clo(s)ure and release (b)urst, and syllable ri(m)e.*

2.3. Analysis

Results were analysed with hierarchical ('mixed-effects') logistic or linear regressions as appropriate, with random intercepts for subjects and items. Use of random slopes, where appropriate (and where models would converge), is noted below.[2]

3. Results

3.1. Rate and distribution of voiced transitions

The rate of appearance of voiced transitions by speech rate is given in Table 2. In general, either of the members of the cluster being voiced was sufficient to condition ∃ (cf. Butler, 2014:94). There was just one instance in 3,640 tokens where a transitional vocoid appeared in a completely voiceless cluster, a fast speech token of ក្រើង /tkaəŋ/ 'prosperous'. Overall rates of vocoid appearance were relatively stable across subjects and conditions (slow: $\mu = 0.39, \sigma = 0.03$; fast: $\mu = 0.27, \sigma = 0.08$).

Table 2: *Rate of appearance of voiced transitions (∃). Dark is Class 1, light is Class 2, white is Class 3. For each cluster the top row gives the rate in slow speech, bottom row in fast speech.*

C_1	C_2										
↓	p	t	k	b	d	n	ŋ	l	r	s	h
p		0	0		0.83	0.13		0.05	1.00	0	0
		0	0		0.82	0.19		0.02	0.28	0	0
t	0		0	0.83		0.07	0.07	0.10	1.00		0
	0		0.02	0.70		0.27	0.13	0.27	0.25		0
k	0	0			0.93	0.07	1.00	0.02	1.00	0	0
	0	0			1.00	0.24	0.62	0.06	0.32	0	0
m		1.00			0.97	1.00		0.93	1.00	0.90	0.90
		0.43			0.87	0.63		0.27	0.41	0.43	0.50
l	1.00			1.00							0.97
	0.60			0.73							0.66

A hierarchical logistic regression predicting presence of an intrusive element from covariates C_1, C_2 and RATE confirms the main effect of RATE ($\beta = -1.53, SE = 0.12, p < 0.001$), but while the appearance of ∃ is modulated by rate, it never disappears entirely. For Class 3 items, ∃ occurred less frequently in fast speech ($\beta = -2.26, SE = 0.21, p < 0.001$), although this effect was greater for some clusters than others (compare e.g. /md/ vs. /ml/). For Class 2 (C + nasal) sequences, on the other hand, ∃ actually occurred *more* frequently in fast speech ($\beta = 1.24, SE = 0.25, p < 0.001$).[3]

3.2. Duration of voiced transitions

If transitional elements are part of the phonological specification of the syllable, we might expect to observe an increase in the duration of onset clusters containing such elements. Figure 2 shows the duration of C_1C_2 sequences by RATE according to the presence or absence of ∃ (defined as when duration of ∃ > 0 ms). Clusters containing ∃ were significantly longer than those that did not in slow speech ($\beta = 13.5, SE = 2.8, p < 0.001$) and marginally so in fast speech ($\beta = 5.98, SE = 3.13, p = 0.057$, with random by-subject and by-item slopes for RATE).

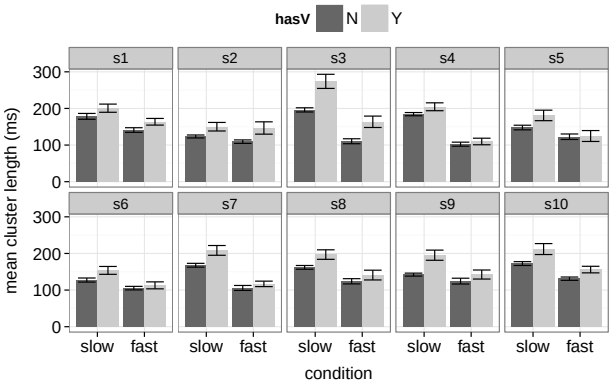

Figure 2: *Duration (in ms) of C_1C_2 with and without ∃ by subject and speech rate. Error bars show 95% confidence intervals.*

[1] The treatment of aspirated stops as sequences (clusters) of plain stops + /h/ is usually argued for on morphological grounds (see e.g. Henderson, 1952), but cf. Section 3.5.

[2] Data and code may be found on the author's website at `http://www.lel.ed.ac.uk/~jkirby/khmer/`.

[3] The rate of ∃ in rapid speech /Cr/ sequences may also be impacted by an independent process of casual speech reduction affecting this particular cluster in the Phnom Penh dialect (see Kirby, 2014).

3.3. Comparison with lexical vowels

Butler (2014) observes that the duration of the transitional vocoid in C∃C syllables was shorter than that of corresponding (short) lexical vowels. This was tested with the present data by comparing durations of C∃C clusters to the durations of eight CəC and CəN monosyllables (e.g. តឹក /tək/ 'granary', ប៊ិន /pən/ 'clever') also recorded as part of the production task. C∃C onset clusters (where C_1 is voiceless and C_2 is voiced, the primary environment where ∃ occurs) were longer than CəC monosyllables at slow ($\beta = 32.94, SE = 5.1, p < 0.001$) but not fast ($\beta = 6.35, SE = 13.25, p = 0.65$) speech rates. However, C∃N onset clusters were shorter than CəN monosyllables at both speech rates (slow: $\beta = 73.77, SE = 9.61$; fast: $\beta = 82.11, SE = 10.76$; both $p < 0.001$).

3.4. Evidence from formant transitions

The formant transitions of the transitional vocoids were also examined. If these elements have an articulatory target, one might expect to see this reflected in their height relative to the surrounding consonants. This effect would be most obvious in cases where ∃ is flanked by two consonants sharing the same place of articulation. In such cases, if ∃ is truly targetless, the difference between F2 at the midpoint and F2 at onset or offset should not differ significantly from zero. Indeed, this is what we find: even when including subject- and item-specific slopes, neither $F2_{ons} - F2_{mid}$ nor $F2_{mid} - F2_{off}$ is significantly different from zero ($F2_{ons} - F2_{mid}$: $\beta = 16, SE = 13, p = 0.31$; $F2_{mid} - F2_{off}$: $\beta = 68.6, SE = 38.3, p = 0.12$; see Figure 3).

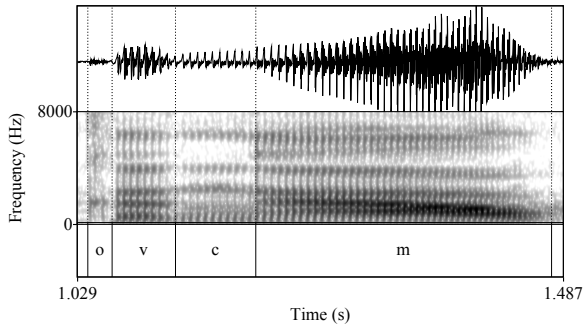

Figure 3: *Segmentation of token* ក្អែក */kɲaok/ 'peacock', speaker 10, slow condition, repetition 1 showing flat formant transitions throughout the transition (labeled v).*

3.5. Voiceless transitions

Huffman's class system suggests that the durational properties of the C_1 release may vary predictable with the manner of C_2. Figure 4 summarizes the durations of voiceless C_1 obstruents (/p t k/) by class and manner of C_2. The release duration of C_1 in (Class 1) C+/h/ clusters (i.e., aspirated stops) was not distinguishable from that of (Class 2) C + sonorant clusters at either slow or fast speech rates; durations of C_1 release when C_2 was a sibilant, obstruent, or rhotic, or when the cluster was /kŋ/, were similarly indistinct. In both conditions, however, the duration of the first group was significantly longer than that of the second (slow: $\beta = -35.33, SE = 2.1$; fast: $\beta = -21.86, SE = 2.19$), and the presence of ∃ reduced the duration of the C_1 release (slow: $\beta = -19.1, SE = 1.65$; fast: $\beta = -14.65, SE = 1.29$; all $p < 0.001$).

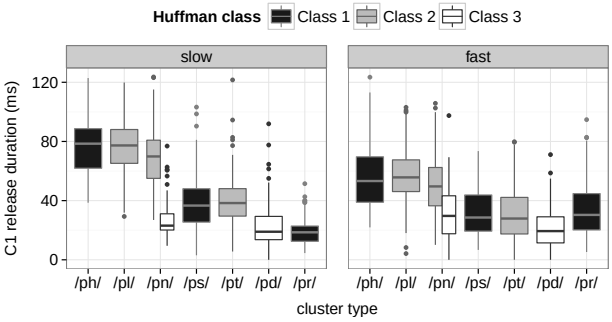

Figure 4: *Duration of C_1 release by cluster type and Huffman class. Labels are examples, so /pn/ subsumes /tn/, /kŋ/, etc.*

3.6. Cluster length as proportion of syllable

It is also instructive to consider the duration of onset clusters as a proportion of total syllable length by class and cluster type (Figure 5). Clusters where C_1 = /p t k/ appear to fall into one of two classes: the total duration of aspirated stops (C+/h/ 'clusters') is not distinct from that of /Cr/ clusters at either speech rate, nor are the durations of the remaining clusters types from other another. Proportion of total syllable duration remains extremely stable across speech rates for all cluster types (including sonorant-initial clusters, not shown here).

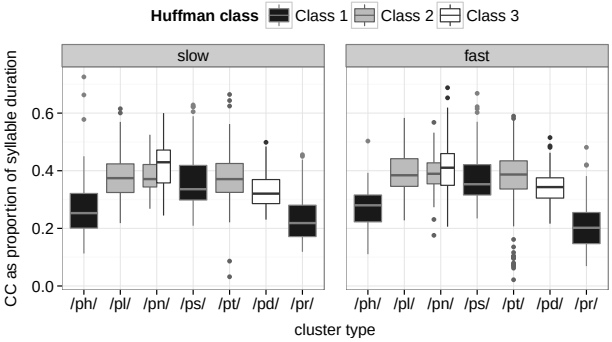

Figure 5: *Duration of onset cluster by cluster type and Huffman class. Labels are examples, so /pn/ subsumes /tn/, /kŋ/, etc.*

4. Discussion

The goal of the present study was to shed new light on the phonological status of transitional elements in Khmer through the use of a speech rate manipulation. The most unequivocal finding is that the presence of a single voiced element in a Khmer onset cluster is sufficient to license an intrusive vocoid. However, even when this basic condition is met, the appearance of ∃ is highly variable: it appears reliably when C_1 is a sonorant (/m/ and /l/) or C_2 is /r/, and remains frequent when C_2 is voiced (/b/ or /d/), but is notably less frequent when C_2 is a sonorant (/m n ŋ l/). The exception to this last generalisation is the cluster /kŋ/, which always occurs with ∃ in slow speech. This does not seem to merely be an effect of homorganicity (∃ is much less common in homorganic clusters like /tn/ or /tl/) but may instead reflect an articulatory strategy aimed at maximising perceptual salience, given that /ŋ/ and /n/ may be especially confusable in syllable onsets (Narayan, 2008).

The spectral characteristics of ∃ do not give any indication that it has an associated spatial target or tongue body gesture (consistent with the findings of Butler, 2014), and with few exceptions, overall cluster duration was found to be stable across speech rates for a given cluster type. However, for at least some speakers, ∃ appears to contribute significantly to total onset length (C∃C clusters were longer than CC clusters in slow speech, and marginally so in fast speech) and the resulting clusters are longer than lexical CəC (but not CəN) monosyllables (cf. e.g. Lebanese Arabic: Gouskova & Hall, 2009; Tashlhiyt Berber: Ridouane & Fougeron, 2011). These latter findings are consistent with an account on which ∃ is phonologically specified in some way.

One might be tempted to view the fact that ∃ generally fails to disappear in fast speech as additional evidence that it has a phonological target. However, this finding is consistent with a number of other interpretations. The use of the speech rate manipulation paradigm was motivated by the idea that if vowel intrusion in slow speech is a byproduct of gestural separation, it may be prone to disappear in fast speech as the relative overlap of the gestures increases. This is based on the assumption that the timing between gestures is pliable. Another possibility is that speakers shorten the duration of each consonantal gesture while maintaining the relative timing between them (Byrd & Tan, 1996); if two consonantal gestures are phased such that the onset of the second is initiated during the release phase of the first, this may result in an audible transition regardless of speech rate (Gafos, 2002). Moreover, individual speakers may differ in whether or not an increase in speech rate results in an increase in gestural overlap due to their employing different gestural implementations of the same cluster types (Tjaden & Weismer, 1998; Davidson, 2006). Such individual differences seem likely to underlie at least some of the variation observed in the present study. While fast speech impacted the appearance of ∃, it did not affect all cluster types in the same way. In particular, it is not clear why intrusion becomes more rather than less likely in fast speech for Class 2 clusters where C_2 = /n ŋ l/, nor why /kŋ/ should pattern with (Class 3) /pd/ and /tb/ rather than with (Class 2) /kn/ and /tŋ/ (though as noted above, there may be a perceptual explanation for the behaviour of this particular cluster). These findings raise the possibility that clusters in Khmer may be to some extent individually specified for gestural coordination and timing.

Finally, while the disappearance of a transitional vocoid at fast speech rates may constitute good evidence that it is intrusive, the converse is not necessarily true (Hall, 2006). This means that the persistent appearance of intrusive vocoids at fast speech rates in Khmer does not permit us to conclude they are phonologically specified. Instrumental articulatory data may help build a more accurate picture of how Khmer onset clusters are implemented and whether or not there is a gestural or segmental target associated with transitional elements.

5. Summary

This study has examined the effect of consonantal context and speech rate on interconsonantal acoustic transitions in Khmer onset clusters. Both context and speech rate were found to have variable effects on the appearance of intrusive material. While the majority of the findings are consistent with an account on which transitions are fundamentally phonetic effects of gestural timing, additional study will be necessary to fully understand the status of acoustic transitions this language.

6. Acknowledgements

Special thanks to Mr Sor Sokny of the Buddhist Institute, Phnom Penh for his invaluable assistance with this project, and to Patrycja Strycharczuk for her insightful comments. This study was funded in part by the University of Edinburgh Hayter Fund and the Carnegie Trust for the Universities of Scotland.

7. References

Boersma, P. & Weenink, D. (2013). Praat: doing phonetics by computer [Computer program]. Version 5.3.57. http://www.praat.org.

Browman, C. P. & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49, 155–180.

Butler, B. A. (2012). Intrusive schwa in Khmer. Poster presented at LabPhon 13, Stuttgart, July 2012.

Butler, B. A. (2014). *Deconstructring the Southeast Asian sesquisyllable: A gestural account*. PhD thesis, Cornell University.

Byrd, D. & Tan, C. C. (1996). Saying consonant clusters quickly. *Journal of Phonetics*, 24, 263–282.

Coleman, J. (2001). The phonetics and phonology of Tashlhiyt Berber syllabic consonants. *Transactions of the Philological Society*, 99, 29–64.

Davidson, L. (2006). Schwa elision in fast speech: segmental deletion or gestural overlap? *Phonetica*, 63, 79–112.

Gafos, A. (2002). A grammar of gestural coordination. *Natural Language and Linguistic Theory*, 20, 269–337.

Gay, T. (1981). Mechanisms in the control of speech rate. *Phonetica*, 38, 148–158.

Gouskova, M. & Hall, N. (2009). Acoustics of epenthetic vowels in Lebanese Arabic. In S. Parker (Ed.), *Phonological argumentation: Essays on evidence and motivation* (pp. 203–225). London: Equinox.

Hall, N. (2006). Cross-linguistic patterns of vowel intrusion. *Phonology*, 23(3), 387–429.

Henderson, E. J. A. (1952). The main features of Cambodian pronunciation. *Bulletin of the School of Oriental and African Studies, University of London*, 14(1), 149–174.

Huffman, F. E. (1972). The boundary between the monosyllable and the disyllable in Cambodian. *Lingua*, 29, 54–66.

Jannedy, S., Fuchs, S., & Weirich, M. (2008). Schwa out of control? In R. Sock, S. Fuchs, & Y. Laprie (Eds.), *Proceedings of the 8th International Seminar on Speech Production* (pp. 129–132). Strasbourg.

Kirby, J. (2014). Incipient tonogenesis in Phnom Penh Khmer: Acoustic and perceptual studies. *Journal of Phonetics*, 43, 69–85.

Matisoff, J. A. (2003). *Handbook of Proto-Tibeto-Burman: System and Philosophy of Sino-Tibetan Reconstruction*. Berkeley: University of California Press.

Munhall, K. & Löfqvist, A. (1992). Gestural aggregation in speech: Laryngeal gestures. *Journal of Phonetics*, 20, 111–126.

Narayan, C. R. (2008). The acoustic-perceptual salience of nasal place contrasts. *Journal of Phonetics*, 36, 191–217.

Ridouane, R. & Fougeron, C. (2011). Schwa elements in Tashlhiyt word-initial clusters. *Laboratory Phonology*, 2, 275–300.

Tjaden, K. & Weismer, G. (1998). Speaking-rate-induced variability in F2 trajectories. *Journal of Speech, Language, and Hearing Research*, 41(5), 976–989.

Warner, N., Jongman, A., Cutler, A., & Mücke, D. (2001). The phonological status of Dutch epenthetic schwa. *Phonology*, 18, 387–420.