A Gestural Account of Minor Syllables: Evidence from Khmer

Becky Butler
Cornell University

1 Introduction
Khmer, an Austro-Asiatic language spoken in Cambodia, has a subset of words – sesquisyllables – consisting of a major (final) syllable and a minor (initial) syllable. The minor syllable has been described as reduced syllable with a schwa-like nucleus. This paper investigates the nature of these schwas through the lens of a gesture-based framework. Results from an acoustic investigation suggest that these schwas are the result of gestural coordinations involving gestural separation and that in fact consonant gestures are spread apart even when no schwa-like material is present. This further suggests that sesquisyllables should be classed with monosyllabic words and not disyllabic words since they have only one full nucleus. In addition, the sonority of the consonants surrounding the schwa-like material plays a significant role in determining the presence of schwa and the duration of the defective syllable. Finally, by being set in a gesture-based framework, sesquisyllables are shown to be related to various cross-linguistic phenomena.

2 Background
2.1 Gestural timing and intrusive schwa
Over the past 20 years, a theoretical framework – Articulatory Phonology – has been developed (Browman and Goldstein 1986, 1989, 1992; Saltzman and Kelso 1987; inter alia) in which speech is formalized as a set of coordinated dynamical gestures, which can be described by how the gestures are phased with one another. These types of models differ from traditional segment-based frameworks in that they can encode temporal information, so that speech sounds can be understood as overlapping, and acoustic outputs can be understood as the result of varying degrees of target attainment within the vocal tract. In addition to overlapping, gestures can also separate from one another, resulting in underlap, which is often notable for the appearance of schwa-like material between consonant gestures.

One striking consequence of this model is the C-center effect, which results from differences in phasing relations between gestures. Onset consonants are timed with vowel gestures, such that they are in a 0°, or in-phase, relation with each other. Consonants, however, repel one another such that they are in a 180°, or anti-phase, relation. When multiple consonants occur in an onset cluster and

*Many thanks to Abby Cohn, Sam Tilsen, members of the Cornell Phonetics Lab and the audience at CLS 48 for their comments. Special thanks to all the Khmer participants.
both of them are phased to the following vowel, these phasing relations compete, yielding the C-center effect (Goldstein et al. 2007). Languages like English (Marin & Pouplier 2008), French (Kühnert et al. 2006) and, for some speakers, Georgian (Goldstein et al. 2007) exhibit the C-center effect, by which the middle point of the consonant gestures is timed in-phase with the vowel gesture. Other languages, however, have been documented to lack C-center effects and exhibit simplex timing instead. Goldstein et al. (2007) show that multiple word-initial pre-vocalic consonants in Tashyyhiyt Berber do not function as onsets and therefore multiple consonants in word-initial position have no effect on the timing relation between the rightmost consonant and the following vowel. Hermes et al. (2008) also show that non-sibilant clusters in Italian display the C-center effect while sibilant-initial clusters do not.

In simplex onset timed languages, word-initial consonant sequences are often interrupted by schwa-like material (hereafter [ə] or intrusive schwa), which is not necessarily phonological (Gafos 2002, Hall 2004). To test the phonological status of intrusive schwa in Tashyyhiyt Berber, Ridouane and Fougeron (2011) conduct a series of articulatory and acoustic experiments to determine the phonetic conditioning necessary for the appearance of intrusive schwa as well as the durational differences across CC and CoC sequences to determine if [ə] has a timing slot and is therefore phonological in nature. They find that intrusive schwa occurs in about 55% of their data and is most common in CC sequences in which C2 is voiced. They also find that CoC sequences are not significantly longer than CC sequences, suggesting that [ə] does not have its own timing slot and therefore is not phonological. In the experiment presented here, I use a similar duration-based analysis to determine the status of intrusive schwa in Khmer and to extend the analysis to other similar languages.

2.2 Sesquisyllables

A sesquisyllable (1) is a word type comprising one right-aligned “major” syllable, which is usually heavy, bears stress or tone, and allows a more complex syllable shape and a larger segmental inventory. The major syllable is preceded by a phonologically reduced “minor” syllable, which is usually light and has a less complex syllable shape and fewer possible segments than the major syllable. Sesquisyllables are commonly (though not exclusively) found in mainland Southeast Asian languages, such as Burmese (Green 2005), Bunong (Butler, forthcoming), Moken (Pittayaporn 2005), and many others.

(1) a) [rəˈbin] ‘gourd’ Bunong (Butler, in progress)
b) [θəˌjè] ‘saliva’ Burmese (Green 2005)
c) [pəˌnuk] ‘full’ Moken (Pittayaporn 2005)
d) [kəˌbal] ‘head’ Khmer (Huffman 1972)

Whereas in many languages like Bunong or Moken, words are maximally sesquisyllabic in length, Khmer is described as having several different types of
word shapes, including CVC monosyllables (2) as well as disyllabic and longer words (3). In addition, it contains a set of words variably realized as either monosyllables with complex onsets or as sesquisyllables (4). Indeed, Khmer is claimed to be a sesquisyllabic language based on this set of words (Henderson 1952), although because of the longer word types it may not be considered sesquisyllabic in the strictest sense.

(2)  ['mʌt] ‘dash away’ មឹត
(3)  [mʌt.'pot] ‘stretch one’s back’ មឹតពត់
(4)  a-1)  [mteh] ‘pepper’ ម្មេស
       a-2)  [mə'teh]  
       b-1)  [ptʃoap] ‘attach’ ប៊ុតូ
       b-2)  [pətʃoap]  

Khmer has a large number of possible word-initial consonant sequences, as seen in Table 1 below. Following Huffman’s (1972) analysis, the clusters are divided into three groups. First, the consonant sequences with no shading are described as having a “weak intruded vocalism of a mid-central quality” (represented here by [ǝ]). A second set of consonant sequences, which are lightly shaded, are claimed to have “slight aspiration” between C1 and C2 (represented by [ə]). Finally, the darkly shaded consonant sequences are not claimed to have any material intervening between the consonants and are described as having “a relatively close transition” from C1 to C2 (p. 55). The clusters in double-outlined boxes are reported on in this study.

<table>
<thead>
<tr>
<th>Cl1</th>
<th>C2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>s</td>
</tr>
<tr>
<td></td>
<td>p</td>
</tr>
<tr>
<td></td>
<td>t</td>
</tr>
<tr>
<td></td>
<td>tf</td>
</tr>
<tr>
<td></td>
<td>k</td>
</tr>
<tr>
<td></td>
<td>s</td>
</tr>
<tr>
<td></td>
<td>m</td>
</tr>
<tr>
<td></td>
<td>l</td>
</tr>
</tbody>
</table>

**Table 1:** Khmer word-initial consonant sequences (Based on Huffman 1972).

Those clusters with intrusive schwa are traditionally considered to be sesquisyllables. In what follows, I report the results of an acoustic experiment in which I seek to understand the nature of minor syllable schwa in Khmer. There are several possibilities as to its phonological status. First, it might be present in the lexical form of the word, identical to an unstressed syllable in a disyllabic
word. Second, it might be phonologically epenthetic, such that its environment is predictable and its presence consistent, and it has a phonological target. These first two possibilities are impossible to distinguish in Khmer since there is no morphological alternation in which a non-schwa form would be possible. Finally, intrusive schwa might be excrescent, such that it has no target but is simply the result of gestural underlap.

3 Hypotheses

In order to determine the phonetic and phonological nature of minor syllable schwa in Khmer, this study investigates the duration of word-initial consonant sequences. Based on previous descriptions of Khmer consonant clusters, the presence of intrusive schwa is likely largely predictable. Although this may indicate that it is not lexical, it tells us very little about whether intrusive schwa is epenthetic or excrescent. However, because sesquisyllables have traditionally been described as different than disyllabic words, in the sense that they do not comprise two full syllables, minor syllable schwas are likely substantively different from underlying unstressed schwas. This means that they are likely the result of gestural underlap, and therefore word-initial consonant sequences with intrusive schwa should be durationally identical to word-initial consonant sequences without schwa but shorter than unstressed syllables in disyllabic words.

There are three possible realizations of the word-initial consonant sequences. First, they may be produced with an intrusive schwa, i.e. [ǝ], between C1 and C2 (Figure 1, left). This is characterized by a voiced period with formant structure between the two consonants. This is voiced underlap case. Second, they may be realized with non-harmonic material, i.e. [ʂ], between C1 and C2 (Figure 1, right). This is likely what Huffman (1972) refers to as aspiration and what I will call voiceless underlap. Finally, they could be produced such that no underlap is discernible (Figure 2). In this latter case, an absence of visible underlap may indicate either that there is no separation between consonants, i.e. no underlap, or that underlap is present but is being obscured by frication noise from C1.
These possibilities lead to several other questions. First, is there a difference between [ǝ] and [ǝ̯] in Khmer? In addition, is the no visible underlap case also different? What are the conditioning factors for their realizations and distributions, and what role do manner and place of articulation play? Next, how does underlap, i.e. what is traditionally considered a minor syllable vowel in Khmer, compare with unstressed lexical schwa in disyllabic words?

4 Method
Eighteen native speakers of Khmer between the ages of 18 and 44 (μ = 27) were recorded. Although recordings were made in Phnom Penh, many participants were from other provinces of Cambodia. Stimuli were randomized and presented to the participants one word at a time. Participants were instructed to read each aloud three times in the frame sentence [nîjî̞̂̊̆̈̋̆̇̈̋̆̇̈̋̆̇̈̋̆̇̈̋̆̇̈̋̆̇̈̋̆̇̈̋̆̇̈̋̆̇̈̋̆̇̈̋̆̇̈̋̆̇̈~mdɔŋ tiǝt\] (‘Say ____ one more time.’). Twenty words of type CCVC/CǝCVC/Cǝ̯CVC were recorded (Table 2), along with 4 disyllabic CAC.’CVC words and 13 monosyllabic CAC words, as controls. In monosyllabic words and in the unstressed syllables of disyllabic words, all vowels were phonologically short lexical schwas.
Table 2: Sesquisyllabic stimuli.

<table>
<thead>
<tr>
<th>IPA</th>
<th>gloss</th>
<th>orthography</th>
<th>IPA</th>
<th>gloss</th>
<th>orthography</th>
</tr>
</thead>
<tbody>
<tr>
<td>p[foap</td>
<td>attach</td>
<td>ប្រូប</td>
<td>stŋ</td>
<td>river</td>
<td>ស្វែង</td>
</tr>
<tr>
<td>pka</td>
<td>flower</td>
<td>ដង</td>
<td>skoal</td>
<td>acquainted w/</td>
<td>ស្គាល់</td>
</tr>
<tr>
<td>pnom</td>
<td>mountain</td>
<td>ភ្នំ</td>
<td>smæǝ</td>
<td>grass, hay</td>
<td>ស្រស់</td>
</tr>
<tr>
<td>pŋut</td>
<td>bathe</td>
<td>ផ្លូត</td>
<td>sŋiom</td>
<td>quiet</td>
<td>ស្រស់</td>
</tr>
<tr>
<td>tŋaj</td>
<td>day, sun</td>
<td>ថ្ងង</td>
<td>mteh</td>
<td>pepper</td>
<td>ម្សងៀម</td>
</tr>
<tr>
<td>præǝ</td>
<td>to use</td>
<td>ប្រូប</td>
<td>mnoah</td>
<td>pineapple</td>
<td>ផ្លូត</td>
</tr>
<tr>
<td>tŋha</td>
<td>to fry</td>
<td>ឆា</td>
<td>lkak</td>
<td>noisily</td>
<td>ស្រស់</td>
</tr>
<tr>
<td>tŋkaj</td>
<td>dog</td>
<td>ឆ្កក</td>
<td>lbaijŋ</td>
<td>game</td>
<td>ស្រស់</td>
</tr>
<tr>
<td>tŋbah</td>
<td>be clear</td>
<td>ប្រូប</td>
<td>lmom</td>
<td>sufficient</td>
<td>ស្រស់</td>
</tr>
<tr>
<td>tŋŋaj</td>
<td>distant</td>
<td>ដង</td>
<td>lnajf</td>
<td>afternoon</td>
<td>ស្រស់</td>
</tr>
</tbody>
</table>

All measurements were done in PRAAT (Boersma and Weenink 2012). Segmentation was completed by eye using spectrograms and waveforms. Beginnings and endings of vowels were considered to be the onset and offset of the second formant, where possible. Nasal to vowel and vowel to nasal transitions were demarcated at sudden rises or drops in the second formant.

To determine the location of spectral energy, center of gravity was measured for affricate and fricative-initial consonant sequences. To do so, sounds were converted to spectra by a fast Fourier transform. Center of gravity measurements were taken over 20 equal windows for each sound. Because sounds were all of different lengths, window lengths varied by sound. However, windows for the majority of sounds were between 5ms and 10ms. For analysis purposes, the resulting measurements were aligned in real time.

5 Results
Of 358 CCVC/CǝCVC/CǝCVC words (2 were omitted due to error), 222 tokens (62%) have some form of underlap, and 136 tokens (38%) do not. For each consonant sequence tested, the percent of those pairs that have underlap are given in the corresponding cells in Table 3. Of the tokens with underlap, 123 (55%) have voiced underlap, and 99 (45%) have voiceless underlap. For those tokens with underlap, the percentage of those tokens with voiced underlap is given in the boxes in Table 4.
For sequences in which C1 is a non-sibilant obstruent, i.e. [p] or [t], when underlap is present (excluding [pr]), the underlap is voiceless. The voiced material in [pr] sequences is likely a result of the articulation of [r] which is realized as a tap or trill. For tokens in which C1 is voiced, i.e. [m] or [l], the underlap, if present, is always voiced. At first blush, for sequences with sibilant C1s, the presence and type of underlap seem more variable. However, underlap following sibilant C1s is consistently voiceless, except in about 28% of [tʃb] underlap tokens. Indeed, as will be shown in §5.2, center of gravity measurements suggest that underlap probably is present in many sibilant C1 tokens.

5.1 Non-sibilant C1s

Voiced underlap and voiceless underlap vary not only by their context, but also by their durational distributions. Figure 3 shows that voiceless underlap has a wider range of possible durations than does voiced underlap.
To further explore this distributional difference, the durations of CC sequences with voiced underlap were compared with the durations of CC sequences with voiceless underlap. However, because of the near-complementary relationship of the C1 contexts of voiced underlap and voiceless underlap, making meaningful comparisons across types, i.e. CəCVC and CǝCVC, is not possible. Because of inherent durational differences between voiceless stops ([p] and [t]) and sonorants ([l] and [m]), there are two possible sources of durational variation in CC sequences. In other words, there are both durational differences between consonant types and possible durational differences between underlap types. Therefore the variation of C1 and C2 types was removed by calculating the residuals of a regression of the total duration (or the underlap duration) with C1 and C2 types and Speaker as a random variable (5), and those residuals can be used to make meaningful comparisons.

(5) a) Total Duration = C1 Type + C2 Type + [Speaker] + εTotalDur
   b) Underlap Duration = C1 Type + C2 Type + [Speaker] + εUnderlapDur

Results show that once the variation in C1 and C2 type is removed via the residuals, underlap type, i.e. voiced underlap versus voiceless underlap, is not significantly correlated with the total duration ($p = 0.9934$) or with underlap duration ($p = 0.7565$). These results indicate that neither the total duration of the CC sequences nor the duration of underlap alone (whether voiced or voiceless) is correlated with underlap type, suggesting that voiced underlap and voiceless underlap are not durationally distinct. However, the total duration is correlated with the underlap duration ($p < 0.0001$), suggesting that the duration of the underlap contributes to the entire duration of the CC sequence.

Because there is no correlation between underlap type and duration of CC sequence or underlap, the large distributional difference in duration between voiced underlap and voiceless underlap as seen in Figure 3 remains unaccounted
for. Therefore the data were also analyzed according to sonority relationships. This is motivated by the frequency of cases of epenthesis which occur to resolve sonority violations (Fleischhacker 2001), as well as by the noted articulatory difficulties of producing clusters which match in sonority in addition to oral gesture (Gafos 2002).

Results show that consonant sequences which differ in sonority, i.e. whether obstruent or sonorant, are significantly longer than sequences that agree in sonority. In other words, Figure 4 shows that SǝO and OǝS sequences are significantly longer than SǝS and OǝO sequences ($p < 0.0001$). In addition, OS sequences with no underlap are longer than OO sequences with no underlap ($p < 0.0001$). Indeed, even OS sequences with no underlap are significantly longer than OǝO and SǝS sequences with underlap ($p < 0.0001$).

![Figure 4: CC durations by sonority type.](image)

The result that sequences differing in sonority are longer than sequences alike in sonority is somewhat unexpected. Longer sequences suggest that more gestural underlap is present, and longer underlap is usually motivated by principles of either perception or production. Therefore it is surprising that an obstruent-sonorant sequence (e.g. [pŋ]), which obeys the sonority sequencing principle (Clements 1990), is longer than an obstruent-obstruent sequence (e.g. [pk]). However, given the typological rarity of stop-nasal clusters, this increased duration could be the result of a more difficult articulation which would not be present in other obstruent-sonorant clusters like stop-[l]. Although this result warrants further investigation, it does demonstrate that tokens with underlap are not durationally distinct from tokens without underlap.

The sonority results prompt further investigation of two additional types. First, they merit a more detailed comparison of the sub-parts of CC sequences. In particular, is there any sort of directional trading relation between the duration of the consonants and the underlap? Second, what does it mean to say that CC sequences that differ in sonority are longer than CC sequences that are alike in sonority? Both of these types were compared to disyllabic words as a control.
First, regarding subparts of the CC sequence, because the duration results above suggest that underlap is the result of gestural spreading, we expect that there is no correlation between the duration of one or both of the consonants with the underlap. To test this, the data were separated by obstruent-sonorant type, and the underlap duration was compared to C1 and C2 for each of the four token types: SǝS, SǝO, OǝO and OǝS. None of the types showed a significant correlation between underlap duration and C1 or C2 duration, except for the OǝS sequence. However, as can be seen in Figure 5, the variance is so large (C1: $R^2 = 0.17$; C2: $R^2 = 0.09$), that the significance of the correlation is not meaningful. In other words, underlap duration and consonant duration are not inversely related.

![Figure 5: Relationship between underlap and consonants in OǝS sequences.](image)

Although these results show that there is no difference in how the internal parts of the CC sequence relate to each other, the result that sequences differing in sonority are longer than sequences similar in sonority is still striking. To suggest that speakers syllabify these forms differently (as monosyllables versus disyllables) according to their sonority relations is quite odd, as is the particularly long duration of the OS sequence. Therefore the durations of CC sequences with both types of underlap and without underlap were compared to CǝC sequences in unstressed syllables of disyllabic words (e.g. [mʌt,ˈpɒt]). Results show that the total durations of OǝO, OǝS and SǝO sequences in unstressed syllables are significantly longer ($p < 0.0001$) than the matching obstruct-sonorant sequences in word-initial CC sequences (No SǝS syllables were recorded, so a comparison was not possible). In addition, durations of [ʌ] in these environments was also tested. For OO and OS sequences, [ʌ] duration was significantly longer than underlap duration ($p < 0.0001$). However, for SO sequences, no significant difference was found ($p = 0.3899$). Nonetheless, overall these results suggest that monosyllables with complex onsets should not be grouped with disyllabic words, whether they have underlap or not.
5.2 Sibilant C1s

This lack of durational differences seen in Figure 4 again confirms the similarity in behavior of voiced underlap and voiceless underlap. However, it was not possible to directly test the durations of the consonants in the tokens with underlap against the tokens without visible underlap, again because of the near mutual exclusivity of their consonant types, i.e. tokens with underlap generally have stop or liquid C1s, and stops with no distinguishable underlap have sibilant C1s. However, when word type – underlap versus no underlap – is regressed by the residuals of the total duration, as in (5) above, results show that the presence or absence of underlap is not correlated with the total duration of the sequence ($p = 0.1263$), indicating that the presence of underlap, whether it is voiced or voiceless, does not add significantly to the duration of the sequence. This suggests that there is actually underlap in the affricate and fricative initial sequences, but it is obscured by the frication noise, such that there is no visible correlate of underlap on a spectrogram.

Tsuchida (1994) shows that center of gravity is reliably used by Japanese speakers for the perception of voiceless vowels following sibilants, and although underlap does not have a gesture comparable to voiceless vowels, it should be reflected by a change in the distribution of spectral energy. Indeed, measurements show that at least in the case of the palatal affricate, underlap is actually also present in sibilant-initial consonant sequences, although the underlap is not visible on a spectrogram. Indeed, Euclidian distance from the peak intensity to the end of the fricative noise is significantly different for [tʰ] as C1 in a tʰCVC word versus as the simplex onset of a tʰVC monosyllable ($p <0.0001$). The much more gradual slope toward the end of the segment, as seen in Figure 6, indicates a transitory underlap period. However, this difference is not visible or statistically significant for [s] in the same contexts ($p = 0.7234$).

![Figure 6: Center of gravity measurements for [tʰ] and [s] in Khmer.](image)

6 Discussion

Khmer intrusive schwa appears in a highly predictable context. The above results have shown that voiced underlap and voiceless underlap are durationally indistinguishable. In addition, tokens with both types of underlap are durationally
equivalent to tokens without any visible underlap. Furthermore, the duration of word initial consonant sequences is significantly shorter than unstressed syllables with lexical schwa. These results combined indicate that sesquisyllables in Khmer are better understood as monosyllables with complex onsets that have gestural underlap, rather than a separate word type.

Traditionally, only those forms with voiced underlap have been considered sesquisyllables. Consonant sequences with voiceless underlap would not have been interpreted as minor syllables since they have no vowel-like material. However, durational measurements and the predictability of the voicing context suggest that voiced underlap and voiceless underlap are both the result of gestural spreading, where the voicing between C1 release and C2 constriction in most cases derives from that of C1. In addition, because even those tokens without any visible underlap are durationally indistinguishable to the underlap tokens, I suggest that they also have underlap which is obscured by the frication noise, a claim which can be tested by investigating the spectral distribution of the frication noise as well as through more artificatory means. Therefore understanding minor syllables in terms of gestures is a more accurate and informative way to conceptualize the sesquisyllable in Khmer and potentially in other languages as well.

In addition, a gestural account can help explain the variation in minor syllables over time in various languages. Matisoff (2003) suggests that there is a pervasive cycle of compounding and reduction in many East and Southeast Asian languages, part of which is a stage of sesquisyllabicity (Figure 7). Such a model provides an interesting example of language change in that it demonstrates how both mistiming of gestures and misinterpretation of gestural timing can affect the lexical representation of a word. In other words, underlap may be misinterpreted as lexical schwa or vice versa. However, the above results also suggest that this sort of model of language change might need to be reconceptualized, such that complex monosyllables and sesquisyllables are conflated into one category (Figure 8). Undoubtedly, however, data from several other sesquisyllabic languages should be studied in order to verify this claim.

![Figure 7: Compounding-Prefixation Cycle](Matisoff 2003) ![Figure 8: Revised Compounding-Prefixation Cycle]
7 Conclusion
Reconceptualizing sesquisyllables in terms of gestures rather than segment-based units allows what was once considered to be an areal feature to be cast within a broader framework. In important ways, minor syllables share properties with complex onsets in many other languages. They are largely an effect of constraints on articulation, namely the articulatory effects of consonant place and voicing. In addition, a dynamic model of language offers a direct account – instead of descriptive one – of the types of language change pervasive in Southeast Asian languages.

Nonetheless, many questions remain regarding this phenomenon. First, the unexpected correlations between sonority type and duration require more investigation to understand the perceptual and/or articulatory constraints on Khmer. Second, although comparing durations across underlap types and word types is informative, vowel quality measurements are also an integral part of determining the nature of intrusive schwa, namely whether or not it has a gestural target (cf. Butler, forthcoming). Finally, this account has suggested that intrusive schwa in Khmer is purely phonetic and does not have any phonological status. However, as suggested above, the historical record suggests that over time intrusive schw as can become phonologized, and lexical schwas can become de-phonologized. Understanding both of these processes in terms of gestural timing will be an integral part of future work.

References


Huffman, F. 1972. The boundary between the monosyllable and the disyllable in Cambodian. Lingua, 29. 54-66.


