

The complex acoustic output of a single articulatory gesture: Pattani Malay word-initial consonant length¹

Arthur S. Abramson

The University of Connecticut and Haskins Laboratories

INTRODUCTION

1.1. MOTIVATION. A linguistic investigation of Pattani Malay need be justified only as a contribution to Southeast Asian linguistics. My interest in this language, which is spoken by ethnic Malays in southern Thailand, is further motivated by its possibly unique characteristic of having phonemically distinctive word-initial consonant length for all classes of consonants.

The disyllabic pattern C(:)VCV(C) is the usual structure of the Pattani word. Thus, the initial short or long consonant can appear in three contexts: (1) Utterance-initial, (2) intervocalic, and (3) post-consonantal. The first context requires no comment except, of course, to say that syntactic rules might constrain what classes of words may appear at the beginning of a sentence. The second context is found when a word-final vowel occurs without a following pause before the word in question. Likewise, a word-final consonant without a pause before the word-initial consonant provides the third context.

If the terms 'short' and 'long' are to be taken seriously, the articulatory closure or constriction—henceforth to be called simply CLOSURE—of a long consonant is held significantly longer than that of its short counterpart. For the experimental phonetician this word-initial contrast raises some interesting questions about the limits of human performance in the production and perception of speech: How much information is carried only by the relative durations of the closures? Does the longer articulatory hold have any concomitant perceptually relevant effects on the speech signal? Is the control of relative duration accompanied by a separately controlled mechanism that has its own perceptually relevant acoustic effects?²

1.2. ARTICULATORY GESTURES AND THE ACOUSTIC SIGNAL. We can normally recognize an acoustic disturbance, even one embedded in noise or badly distorted, as a speech signal. To do this we need neither understand the linguistic message nor know the language in which it was uttered. Presumably this is so because the acoustic signal sounds like the possible output of a human vocal tract. Indeed, the synthesis of speech or, if you will, speechlike sequences, is feasible only if the parameters of the synthesizer are set well enough to simulate the acoustic effects of states and movements, i.e., GESTURES, of the articulators.³ Thus, the listener may realize that the 'utterance' has come from a robot or some other sort of machine but still accept it as speech, albeit synthetic speech.⁴

The linguist's concern with phonologically relevant properties of speech, called by some scholars distinctive features, brings us to the question of the links between these fairly abstract properties and phonetic reality. For the work being presented here it is desirable to limit our attention to phonological properties that are defined in terms of actions of articulators or physiological mechanisms (Browman & Goldstein 1986).

The simplest case would be that of a single gesture with a single audible acoustic effect. Perhaps a good example is the movement of the tongue into and out of the position for an apical constriction suitable for the turbulence appropriate to the sound[s].⁵

A more complicated but probably phonologically tolerable case would be that of a single articulatory gesture with multiple acoustic consequences each of which is audible. This is seen in voicing distinctions in initial stop consonants for which the timing of the laryngeal gesture relative to supraglottal gestures causes the valvular action of the glottis to yield a variety of acoustic effects (Lisker & Abramson 1964, Abramson 1977, House & Fairbanks 1953). These include differences along at least three dimensions: the occurrence of glottal pulsing, noise-excitation of formants upon release, i.e., aspiration and fundamental frequency, each of which is detectable by ear. Variation along these acoustic dimensions, even the fundamental frequency of the voice upon release of the stop according to current research (Löfqvist et al.

1989), is apparently a function of the timing of the laryngeal gesture.

Finally, let us consider a phonological distinction involving separately controlled gestures with multiple acoustic consequences. A good example might be the voicing distinction in English word-final consonants. The aforementioned laryngeal gesture can handle the matter of whether or not glottal pulsing persists in the consonant closure. This is audible. At the same time, in some contexts there is a significant correlation between the duration of the preceding vowel and the voicing state of the consonant (Peterson & Lehiste 1960), which is perceptually relevant (Denes 1955, Raphael 1981). The latter property turns out not to be just one more output of the laryngeal gesture but rather part of the separately controlled articulation of the vowel (Raphael 1974). Phonologists of certain schools of thought seem to get around this irregularity by describing it as the application of a 'vowel-lengthening rule' before voiced consonants. The logic may be impeccable, but the phonetic motivation for such an 'explanation' is not at all obvious.

It seems to me that this last situation is phonetically very interesting with important implications for models of speech perception, especially if, as has been argued (e.g., Liberman & Mattingly 1985), the perception of speech directly entails articulatory gestures. It also presents a challenge to those models of phonology that try to be phonetically realistic by incorporating specifications of gestures (e.g., Browman & Goldstein 1986). As suggested earlier (§1.1), the word-initial length distinction in Pattani Malay may offer fertile ground for research into this topic.

1.3. PREVIOUS PHONETIC RESEARCH ON PATTANI MALAY. Within a class of unstressed morphemes the Pattani Malay language has developed distinctive consonant length diachronically through reduction and eventual loss of the vowel and assimilation of the final consonant to the initial consonant of the following morpheme (Chaiyanara 1983).⁶ Although many words with long initial consonants can still be seen to be sitting astraddle a morpheme boundary, it is also true that such a boundary is no longer obviously present in a good number of other words. I am aware of no statistical treatment of the matter.

Of course, the continued presence of a morpheme boundary in much of the lexicon would lead some phonologists to posit gemination rather than distinctive length. Taken seriously as a phonetic statement rather than a phonological abstraction, this would mean a sequence of two instances of the same speech sound, implying rearticulation at the beginning of the second segment. In the absence of a pause during the sustained hold, such a phenomenon seems highly unlikely and is not evident in the available data, so I cling to the concept of distinctive length even while conceding that in this language higher-level grammatical considerations might make it convenient for the phonologist to prefer analyzing phonetic length as a string of two consonants.⁷

Here are some word-pairs with the contrast:

make	'to eat'	m ^h ake	'to be eaten'
labɔ	'to profit'	l ^h abɔ	'spider'
siku	'elbow'	s ^h iku	'hand-tool'
dʒale	'way'	dʒ ^h ale	'to walk'
butɔ	'blind'	b ^h utɔ	'kind of tree'

The foregoing words, which are transcribed with symbols of the International Phonetic Association, all have acoustic excitation in the closures of the initial consonants, making it reasonable to suppose that the onsets on the right could be heard as longer than those on the left. In words beginning with voiceless unaspirated stops, however, there is no closure-excitation that might help to hear a difference in length. Here are some examples:

pagi	'morning'	p ^h agi	'early morning'
paka	'to use'	p ^h aka	'usable'
tawa	'bland'	t ^h awa	'to show wares'
katoɿ	'to strike'	k ^h atoɿ	'frog'
kamɛŋ	'goat'	k ^h amɛŋ	'goatlike'

In Figure 1 are seen the waveforms of a minimal pair of words differentiated by the relative durations of the voice-excited initial labial closures. The words were not recorded together but were put together for the display. (Curiously enough, if the