Acoustic Cues to Word-Initial Stop Length in Pattani Malay

Arthur S. Abramson

Haskins Laboratories, New Haven, CT, U.S.A. and University of Connecticut, Storrs, CT, U.S.A.

E-mail: abramson@uconnvm.uconn.edu

ABSTRACT

Previous work showed that the Pattani Malay length distinction for all word-initial consonants is audible for all categories, even voiceless plosives, and that relative closure-duration is a significant differentiator. Listening suggested greater salience on the first syllable of words with a long plosive, and higher amplitude and fundamental frequency were found on such first syllables. Amplitude variants yielded significant effects but not complete perceptual shifts. Changes in F0 on both syllables gave the same results. This study tests the combining of relative amplitude and F0. For every word-pair the experiment included the original words and stimuli derived from each member of the pair by changing the peaks of F0 and amplitude on both syllables. Combinations of values of F0 and amplitude provide sufficient and powerful cues to the length of initial voiceless stops but not for voiced stops. For nasals and fricatives the effect was clear.

1. INTRODUCTION

The existence in southern Thailand of a language, Pattani Malay [1], with phonologically distinctive length in all word-initial consonants [2] raises interesting questions about the production and perception of speech. If relative duration of the consonantal hold is the sole defining articulatory mechanism, ought we to find that audible length of the interval of closure or constriction is the sole acoustic cue for perception, or might duration alone yield other cues? If, however, more cues are found than can be explained simply by the duration of the hold, might we be confronted with a unit undergoing transition?

In previous work control tests [3] showed that citation forms of words beginning with consonants of most of the categories of the language were correctly identified by native speakers virtually 100% of the time. The identification of voiced and voiceless stops was less good but still at a level over 80%. Affricates did even less well but better than chance. It is reasonable to suppose that the length difference as such can be heard in all contexts only when audible acoustic excitation is present during the closure or constriction. In voiceless stops and affricates, of course, the closure is silent, and its length can be heard as longer or shorter only when preceded by a vowel; consequently, the question arises as to how listeners do so well in labeling them in initial position. Another study [4] showed that the closures were indeed significantly longer for the "long" consonants. Naturally, acoustic measurements of the closures of voiceless stops could be made only in an intervocalic context. It was also found that in the Pattani Malay lexicon, which is largely disyllabic, the amplitude is higher on the first syllable of words with long consonants.

Further work [3] tested the efficacy of closure duration versus that of any possible concomitant cues by incrementally lengthening the closures of short consonants and shortening those of long consonants. In every experiment relative duration was dominant. In yet another perceptual study [5] the interaction of amplitude with closure duration was significant, but relative amplitude by itself while effective was not powerful enough to achieve a full change in category-identification.

At this point the growing suspicion that the distinction was undergoing a change to an accentual distinction led to a change in strategy. Minimal pairs were analyzed acoustically for relative values of amplitude, amplitude slope, vowel duration, and fundamental frequency (F0) across the two syllables of each word [6]. Also obtained were duration-ratios of the first vowel and medial closure. All the factors were significant, but the most striking ones, especially for the continuants, were the ratios of F0 and amplitude. A follow-up study [7] focused on the perceptual efficacy of the ratios of F0. Increments of F0 rise were imposed upon the first syllables of words with short consonants and F0 lowering upon words with long consonants. The effects both ways were highly significant; nevertheless, F0 by itself was not a sufficiently powerful cue to cause a crossover in perception.

2. NEW PERCEPTUAL EXPERIMENTS

The present work continues to study the possible link between accentual salience and the length distinction with stimuli in which F0 and amplitude co-vary in an elaborate way. From each word of each pair in Table 1, 35 stimuli were made, the original word plus 34 variants.

pagi	'morning'	pːagi	'early morning'
tido	'to sleep'	tːido	'put to sleep'
kitɔ	'we, us'	kːitɔ	'among us'
dapo	'kitchen'	dːapo	'in the kitchen'
siku	'elbow'	sːiku	'hand-tool'
make	'to eat'	mːake	'to be eaten'

Table 1: Sources of the stimuli (male speaker).

ISBN 1-876346-48-5 © 2003 UAB

PARAMETER SETTINGS									
From a Short Consonant (C)					From a Long Consonant (CC)				
Stimuli	**Syllable 1**		**Syllable 2**		**Stimuli**	**Syllable 1**		**Syllable 2**	
	ST	**dB**	**ST**	**dB**		**ST**	**dB**	**ST**	**dB**
C1	None	None	None	None	**CC1**	None	None	None	None
C2	None	None	-0.75	None	**CC2**	None	None	+0.75	None
C3	None	None	-1.5	None	**CC3**	None	None	+1.5	None
C4	None	None	None	-3	**CC4**	None	None	None	+1.5
C5	None	None	None	-6	**CC5**	None	None	None	+3
C6	None	None	-0.75	-3	**CC6**	None	None	+0.75	+1.5
C7	None	None	-0.75	-6	**CC7**	None	None	+0.75	+3
C8	None	None	-1.5	-3	**CC8**	None	None	+1.5	+1.5
C9	None	None	-1.5	-6	**CC9**	None	None	+1.5	+3
C10	+2	None	None	None	**CC10**	-2	None	None	None
C11	+2	None	-0.75	None	**CC11**	-2	None	+0.75	None
C12	+2	None	-1.5	None	**CC12**	-2	None	+1.5	None
C13	+3	None	None	None	**CC13**	-3	None	None	None
C14	+3	None	-0.75	None	**CC14**	-3	None	+0.75	None
C15	+3	None	-1.5	None	**CC15**	-3	None	+1.5	None
C16	None	+3	None	None	**CC16**	None	-3	None	None
C17	None	+3	None	-3	**CC17**	None	-3	None	+1.5
C18	None	+3	None	-6	**CC18**	None	-3	None	+3
C19	None	+6	None	None	**CC19**	None	-6	None	None
C20	None	+6	None	-3	**CC20**	None	-6	None	+1.5
C21	None	+6	None	-6	**CC21**	None	-6	None	+3
C22	+2	+3	None	None	**CC22**	-2	-3	None	None
C23	+2	+3	-0.75	None	**CC23**	-2	-3	+0.75	None
C24	+2	+3	-1.5	None	**CC24**	-2	-3	+1.5	None
C25	+2	+3	None	-3	**CC25**	-2	-3	None	+1.5
C26	+2	+3	None	-6	**CC26**	-2	-3	None	+3
C27	+2	+3	-0.75	-3	**CC27**	-2	-3	+0.75	+1.5
C28	+2	+3	-1.5	-6	**CC28**	-2	-3	+1.5	+3
C29	+3	+6	None	None	**CC29**	-3	-6	None	None
C30	+3	+6	-0.75	None	**CC30**	-3	-6	+0.75	None
C31	+3	+6	-1.5	None	**CC31**	-3	-6	+1.5	None
C32	+3	+6	None	-3	**CC32**	-3	-6	None	+1.5
C33	+3	+6	None	-6	**CC33**	-3	-6	None	+3
C34	+3	+6	-0.75	-3	**CC34**	-3	-6	+0.75	+1.5
C35	+3	+6	-1.5	-6	**CC35**	-3	-6	+1.5	+3

Table 2. Parameter values applied to the original utterances to create stimuli. ST = semitones.

The design of this set of experiments involved incrementally altering the salience of the syllables of minimal pairs of words to test whether this would change the percepts of native speakers, even though all other properties of the original utterances, especially closure duration, were left unchanged. For words with short consonants, the salience of the first syllable was raised, while that of the second syllable was reduced. Contrariwise, for words with long consonants, the salience of the first syllable was lowered, while that of the second syllable was raised. Utterances of the words in Table 1 were recorded by an educated adult male speaker of Pattani Malay and examined for ranges of peak values of F0 and amplitude, the two most promising acoustic properties [6], on the first and second syllable. With rather short ranges determined by observational data, two values along each range were used as parameter settings in the Peak™ program, as shown in Table 2. The program enables one to alter various prosodic properties with minimal or no disturbance to other aspects of the speech signal. Thus we see in the table that C1 and CC1, representing any original pair of words differentiated by the length of the initial consonants, have not been changed at all. *None* is entered in all the boxes. In C2 and CC2 only F0 on the second syllable has been changed, 0.75 semitones downward on C2 and 0.75 semitones upward on CC2. The greatest changes in the stimuli are on line 35. The 70 stimuli for each word-pair were randomized into six separate tests, one for each pair, and played through headphones for identification as words by native speakers. Given occasional absence or losing of one's place, the number of usable subjects ranged from 20 to 30.

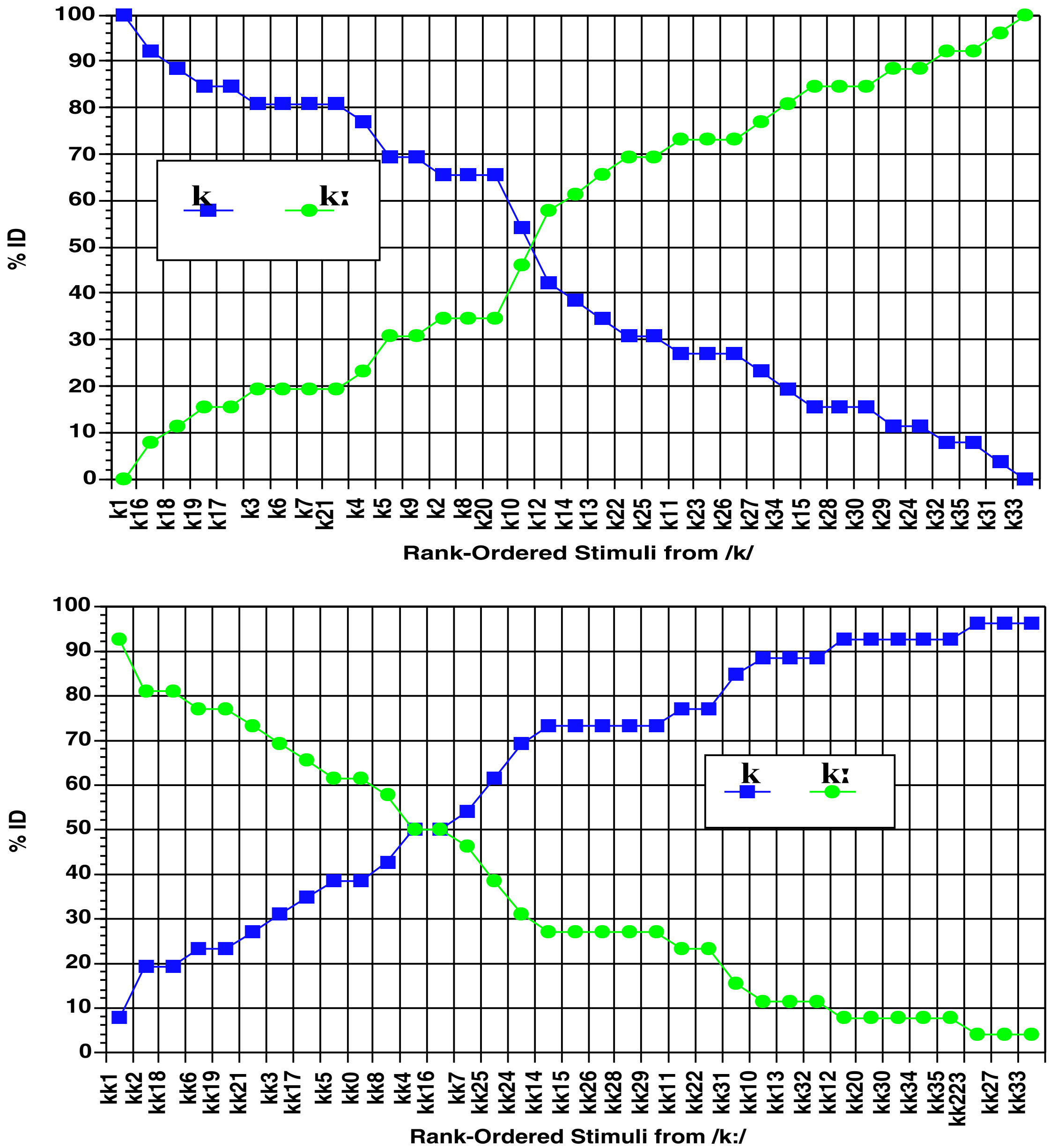

Figure 1. Responses by 26 subjects to the stimuli made from /kitɔ/ and /kːitɔ/..

3. RESULTS

It is clear from Table 2 that there is no single physical dimension underlying the 70 stimuli of each experiment. Such a complicated design came about from the hope that finally, after five previous studies, the vexing questions would be answered. As a result, the data are displayed in Figure 1 for the dorso-velar voiceless stops as a rank-ordering of the perceptual effects. In the upper graph we see that stimulus K1, equivalent to C1 in Table 2, i.e., the original word /kitɔ/, was heard 100% of the time as short /k/. At the far right, the stimulus least often heard as /k/ but 100% as long /kː/ is k33. In the lower graph we see that kk1, the original word /kːitɔ/, is heard most often as /kː/, while it at the far right it is kk33 that is heard most often as short /k/.

Unfortunately, there is not enough space to display legible graphs of the rest of the data. The results for the other pairs of voiceless stops, /p pː/ and /t tː/, are similar but not as robust for the shift from long to short consonant. The one pair of voiced stops tested, /d dː/, showed the same trend but no complete crossover in identification for either series of stimuli. For the nasals /m mː/ and the fricatives /s sː/ there were perceptual crossovers but a bit weaker than those of Figure 1.

With the lack of a single underlying physical dimension, it was seemingly impossible to apply the usual statistical tests for significance. Instead, a correlation coefficient was run on the rank-orderings to see how well they matched each other, as shown in the matrix of Table 3.

ISBN 1-876346-48-5 © 2003 UAB

From	/p/	/t/	/k/	/pː/	/tː/	/kː/
/p/	**1.000**	**.882**	**.847**	**-.809**	**-.641**	**-.729**
/t /	**.882**	**1.000**	**.763**	**-.787**	**-.679**	**-.706**
/k/	**.847**	**.763**	**1.000**	**-.778**	**-.728**	**-.819**
/pː/	**-.809**	**-.787**	**-.778**	**1.000**	**.712**	**.682**
/tː/	**-.641**	**-.679**	**-.728**	**.712**	**1.000**	**.651**
/kː/	**-.729**	**-.706**	**-.819**	**.682**	**.651**	**1.000**

Table 3: Correlation matrix of responses as short consonants to 35stimuli from each voiceless-stop source.

4. CONCLUSIONS

The questions motivating this study arose because of the rather good identifiability of utterance-initial voiceless unaspirated stops [3]. The perceptual crossovers of Figure 1, as well as those of the unseen graphs for the other stop-pairs, are well correlated with the differentiating effects of relative amplitude and F0 found in production data [5, 6]. Indeed, the speaker used in this study was one of the ones whose speech was analyzed earlier. At the same time, it must be recalled that when these voiceless stops are embedded in a postvocalic environment, closure-duration is a very powerful cue [3].

In each of the rest of the word-pairs tested the quite audible voiced excitation of the closures or constrictions, differing significantly in duration [4], remained unchanged. Thus, in the present experiments the cues of amplitude and F0 were pitted against the demonstrably relevant one of closure duration. In fact, the combined cues, while causing a trend in the expected direction, did not succeed in effecting proper perceptual crossovers with the voiced-stop stimuli. On the other hand, in the experiments with nasals and voiceless fricatives perceptual crossovers were achieved in spite of the presence of audible differences in the durations of closures or constrictions. At this time it is hard to think of a convincing explanation of the apparent paradox.

Given the complex nature of the stimuli and the consequent lack of a single underlying physical dimension, the only statistical treatment of the data here is the correlation matrix of Table 3. We observe there that for the stimuli derived from each original word, one with a short voiceless stop and the other with a long one, within each length group the correlations come more or less close to 1. This would seem to indicate, for the rank-ordered responses, a fairly significant level of agreement in the perceptual treatment of the items.

A plausible interpretation of the overall results seems to be that the length distinction as such is still present in Pattani Malay but is undergoing a transition to a system of accentual prominence on the first syllable of words with initial long consonants. The property of acoustic salience, which may have long been concomitant with closure duration, at least in the voiceless stops, is probably being generalized across all consonantal categories. Perhaps relative duration as such will disappear.

ACKNOWLEDGMENTS

This research was supported by NIH Grant DC-02717 to Haskins Laboratories. My presence in Thailand was made possible by the RGJ Ph.D. Program under the Thailand Research Fund. The excellent help given by the Dept. of Linguistics, Chulalongkorn University, Bangkok and the Dept. of Islamic Studies, Prince of Songkhla University, Pattani is much appreciated.

REFERENCES

[1] A. Thavisak, *Malay dialects in Thailand* [In Thai]. Nakhorn Pathom, Thailand: Institute of Language and Culture for Rural Development, Mahidol University, 1987.

[2] P.M. Chaiyanara, *Dialek Melayu Patani des bahasa Malaysia* [The Pattani Malay dialect of the Malay language]. Unpublished M.A. thesis, University of Malaya, Kuala Lumpur, 1985.

[3] A.S. Abramson,. "The perception of word-initial consonant length: Pattani Malay, *Journal of the International Phonetic Association*, vol. 16, pp. 8–16, 1986.

[4] A.S. Abramson,. "Word-initial consonant length in Pattani Malay," in *Proceedings of the XIth International Congress of Phonetic Sciences*, vol. 6, pp. 68–70. Tallinn: Academy of Sciences of the Estonian S.S.R., 1987.

[5] A.S. Abramson, "Amplitude as a cue to word-initial consonant length: Pattani Malay," in *Proceedings of the XIIth International Congress of Phonetic Sciences*, vol. 3, pp 98–101. Aix-en-Provence: Université de Provence, 1991.

[6] A.S. Abramson, "The complex acoustic output of a single articulatory gesture: Pattani Malay word-initial consonant length," in *Papers from the 4th Annual Meeting of the Southeast Asian Linguistics Society 1994*, pp. 1–20. Tempe, Arizona: Arizona State University Program for Southeast Asian Studies, Monograph Series Press, 1998.

[7] A.S. Abramson, "Fundamental frequency as a cue to word-initial consonant length: Pattani Malay," in *Proceedings of the 14th International Congress of Phonetic Sciences*, pp. 591–594. Berkeley, CA: University of California Press, 1999.