The complex acoustic output of a single articulatory gesture: Pattani Malay word-initial consonant length[1]

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(ː)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:

makɛ	'to eat'	mːakɛ	'to be eaten'
labɔ	'to profit'	lːabɔ	'spider'
siku	'elbow'	sːiku	'hand-tool'
dʑalɛ	'way'	dʑːalɛ	'to walk'
butɔ	'blind'	bːutɔ	'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'	pːagi	'early morning'
paka	'to use'	pːaka	'usable'
tawa	'bland'	tːawa	'to show wares'
katoʔ	'to strike'	kːatoʔ	'frog'
kamɛŋ	'goat'	kːamɛŋ	'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 distinction in question were relevant ONLY for voiced stops, the language could be seen as having an otherwise unattested use of the VOT dimension for a three-way voicing distinction (Lisker & Abramson 1964): long voicing lead, short voicing lead, and short lag, i.e., early prevoicing, late prevoicing, and voiceless unaspirated.)

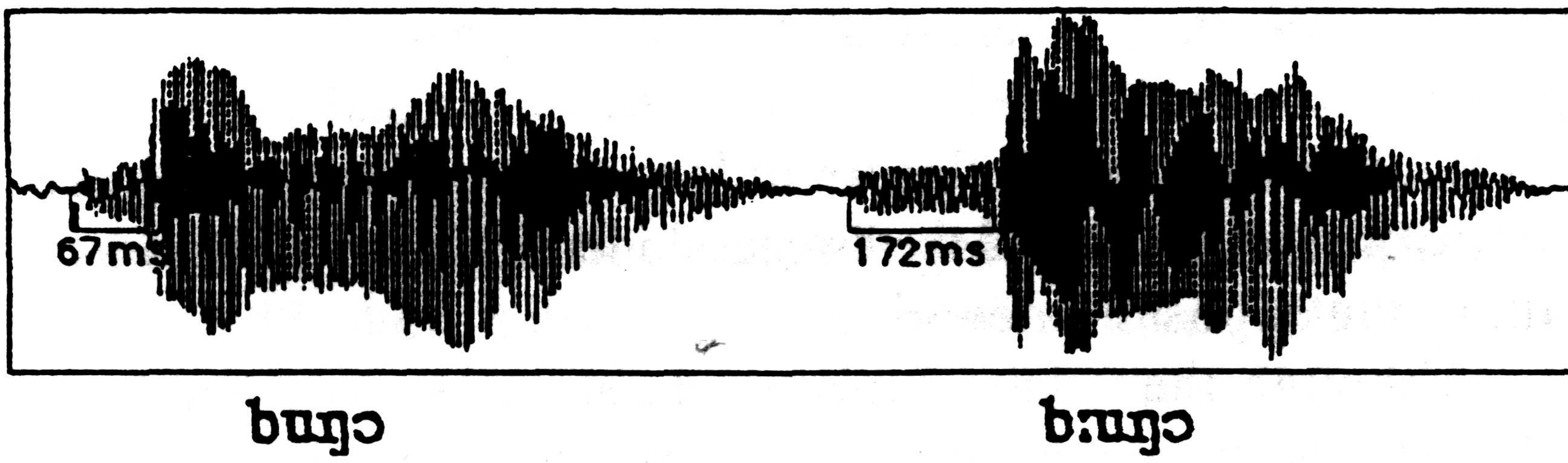

FIGURE 1. Waveforms of /buŋɔ/ 'flower' and /bːuŋɔ/ 'to bloom,' showing the durations of the voiced closures.

In the pair shown in Figure 1, the closure of the long stop, as measured acoustically, is 2.6 times longer than the closure of the short stop. In fact, analysis of a large number of productions (Abramson 1987) showed the closures of long consonants, in both initial and intervocalic positions, to be on average three times longer than those of short consonants. Of course, for voiceless stops and affricates there was no direct acoustic sign of the closure in utterance-initial position, so all durational data for them were obtained in intervocalic position, i.e., utterance medially after words ending in vowels.

The next stage of the work, although my report on it appeared earlier (Abramson 1986), studied the efficacy of closure duration as a perceptual cue. First, to assess the robustness of the distinction, control tests were run with isolated members of 16 short-long pairs that were recorded by two native speakers, randomized, and played through headphones to 21 other native speakers for identification as words. Words with initial nasals, laterals, and fricatives (henceforth to be grouped as CONTINUANTS) were labeled with an accuracy of 90% or better. Labels of voiced and voiceless stops were slightly less accurate at 80% or better. The affricates, especially the voiceless ones, did not fare well, with identifications not much better than chance. These results justified further perceptual experimentation but with the exclusion of the affricates for the time being.

Manipulation of closure duration was the method used (Abramson 1986). Shortening the closures of long utterance-initial /gː/ and long utterance-initial /lː/ in a series of small steps made listeners shift their labels to the short category. For voiceless stops, the tests had to be limited to utterance-medial intervocalic position; shortening the closure of long /pː/ and lengthening the closure of short /p/ caused the listeners' identification functions to shift to the opposite categories.

Although the value of relative duration as a cue had been demonstrated, there remained the question of how subjects could do so well with isolated words beginning with voiceless unaspirated stops, which have no DIRECT acoustic sign of closure duration. Although it is true that in intervocalic position relative duration is a sufficient cue for the voiceless stops, such that any concomitant cues that might be present in the original utterances are overridden by lengthening original short closures and shortening original long ones, it is also true that the crossover points of the identification curves in the two conditions differ significantly from each other. This suggests that even in the presence of the powerful duration cue something else must be contributing to the auditory distinction. Given the fact that words in this language are typically disyllabic, it seemed likely that such an additional differentiating property might reside in intersyllabic relations. Auditory impressions and a quick sampling of acoustic analyses suggested relative amplitude as the most promising feature for investigation.

Root mean square amplitudes of both syllables of a large number of Pattani Malay disyllabic words were obtained for the speech of one man by means of a computer program (Abramson 1987). The difference between his short and long voiceless plosives with respect to amplitude ratio was highly significant. The voiced long plosives also showed a higher ratio than the short ones but with only moderate significance. Although the continuants showed a tendency in the same direction, the effect was not statistically significant. Given the complex effects on speech signals of variation in timing (Lisker 1974), I wondered whether greater duration of consonantal closure entailed a higher buildup of oral air pressure, which, upon release, might yield a higher amplitude peak on the syllable. If so, perhaps this would serve as a cue in the absence of audible closure duration.

The final stage of my previous research was to test the perceptual relevance of the findings on relative amplitude (Abramson 1991). Three pairs of words with short and long initial voiceless stops were recorded at the end of a carrier sentence ending in a vowel. The amplitude peak of the first syllable of words with short initial stops was raised in five 2-dB steps; the peak of the first syllable of words with long initial stops was lowered in five 2-dB steps. These variants were pitted against variants in closure duration. In addition, stimuli with the same amplitude variation, but necessarily without changes in closure duration, were prepared with a pair of isolated words.

The experiments with amplitude pitted against duration revealed that duration is not only sufficient, as before, but also dominant, although relative amplitude turned out to have SOME cue-value in that the perceptual crossover point was earlier for higher amplitude peaks. In the stimuli derived from isolated words, amplitude influenced the identification functions somewhat but was insufficient as a cue to the categories.

NEW WORK

2.1. HYPOTHESIS. Although the ratio of amplitude peaks in disyllabic words physically differentiates short and long voiceless stops, it is by itself a very weak cue for distinguishing the two categories perceptually; nevertheless, the two categories are identified rather well even in isolated words. The hypothesis proposed for testing in the present study holds that some several properties, among them relative amplitude, work together to give greater acoustic salience to the first syllable of words beginning with long stops. In addition, relative salience could also help with the differentiation of short and long voiced stops and perhaps even some of the continuants. In this part of the research, the hypothesis is being tested for validity in speech production. Should it turn out to be tenable, it will be necessary to validate it for perception.

One might ask as to whether a feature of accent or stress is arising in the language in conjunction with the consonant-length distinction. Indeed, in response to my stated plan to look at relative amplitude and perhaps other properties, Christopher Court of Monash University wrote to me on 23 May 1976, 'I have a feeling that what we are dealing with here is something like the Swedish-Norwegian "accentual" system, or the Serbo-Croatian or Japanese one, which is only realized, or at least BEST realized, over the span of two syllables, and that the picture will become clearer when you compare the two syllables.' Looking back at this old letter from a scholar who had done much work with the language, I felt encouraged to return to an idea that I had neglected.

2.2. PROCEDURE. Lists of words forming minimal pairs with respect to consonantal length were recorded in isolation and embedded in carrier sentences by four young adults, all native speakers of Pattani Malay. The utterances of two of them, both women, were given to me by Jimmy G. Harris who, together with Christopher Court, first called my attention to the language. The other two informants, both men, were recorded by me. Since these words were recorded at different times for different purposes, only the lists of the first two speakers are identical. This will be reflected in the statistical treatment of the results. For this stage of the new work, only the isolated words were analyzed.

The choice of acoustic properties for examination was based on the literature on distinctive stress in other languages (e.g., Fry 1955). That is, if one is to test a hypothesis of relative salience or accent accompanying the length distinction, the following aspects of the speech signal are most promising. It was decided to obtain relative values of amplitude, amplitude slope, vowel duration, and fundamental frequency (F_0) across the two syllables of each word. In addition, duration-ratios of the first vowel and the medial closure would be included.

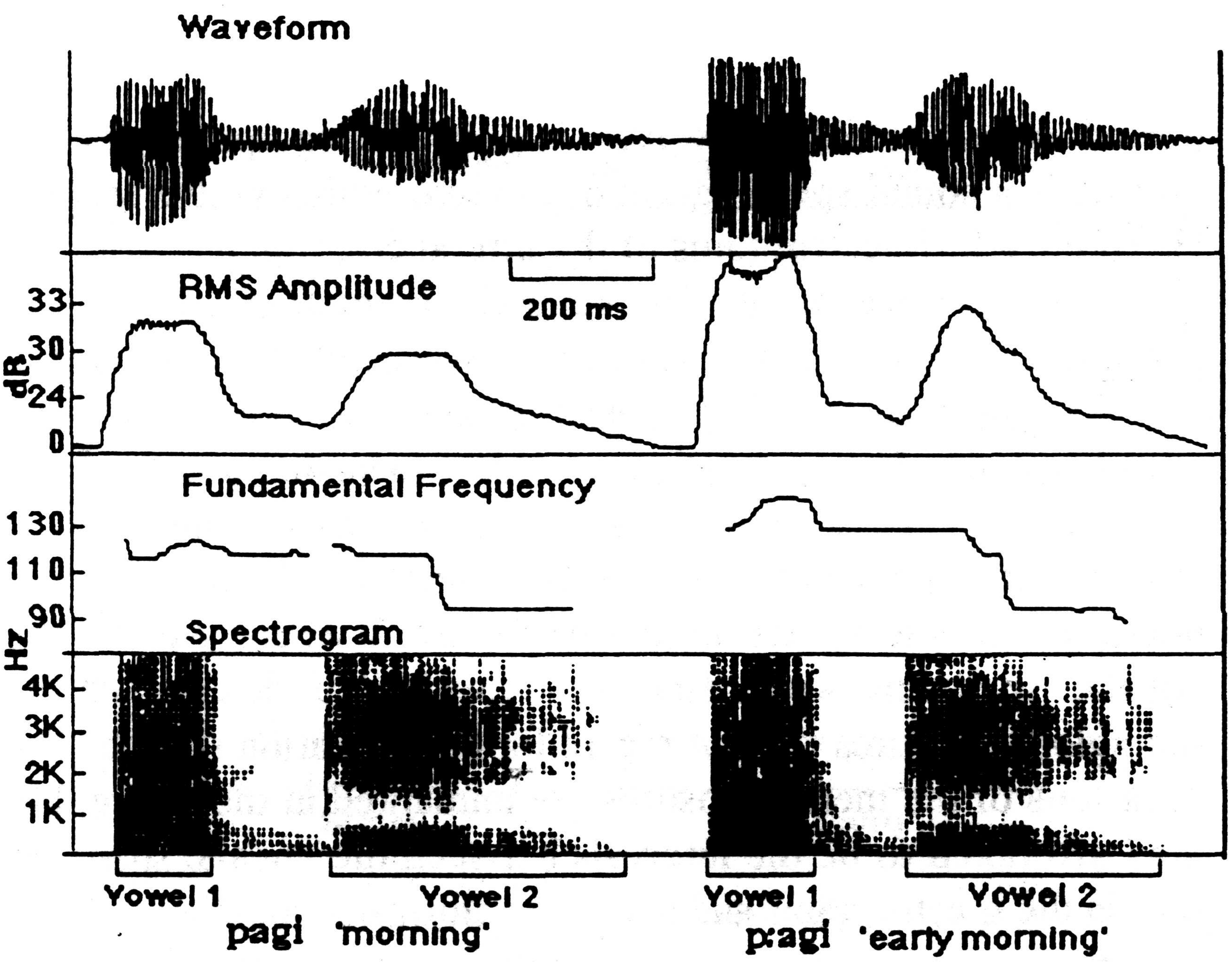

FIGURE 2. Acoustic analysis. The space between Vowel 1 and Vowel 2 in each of the spectrograms is the voiced /g/ closure.

The Signalyze™ program was used on the Macintosh computer for the acoustic analysis. Figure 2 can serve as a guide to the analytic procedures. At the top we see waveforms of a minimal pair of words taken from a randomized set of recordings and put together for this display. This is the starting point for the following analyses. A 200-ms span under the waveforms serves for temporal orientation for all the displays of Figure 2. Just below is a root-mean-square amplitude trace with a decibel scale to the left. For each disyllabic word the ratio of the amplitude peak of the first syllable to that of the second syllable was obtained. In addition, the average rise time of the amplitude slope of the first syllable of each word was obtained by dividing the peak amplitude by the duration of the rise from the baseline to the peak. Next down is an F_0 trace with a scale in Hertz (Hz) to the left. For each disyllabic word the ratio of the F_0 peak of the first syllable to that of the second syllable was obtained. Although the program, with appropriate adjustments, does a good job of extracting F_0, the accuracy of a sampling of the analyses was checked by period-by-period measurements of waveforms on the computer. At the bottom is a sound spectrogram of the two words with a scale in Hz to the left. Spectrograms of the utterances, together with the waveforms, were used for general orientation and, more particularly, for measurements of the vowels and medial closures.

The physical criteria for measuring vowel duration require some comment. Vowel duration was taken to be the span from the moment of release of the first consonant to the moment of achievement of closure for the final consonant. The procedure is illustrated in the spectrograms of Figure 2 where we see the vocalic spans marked. That is, vowels are determined by supraglottal gestures without regard for the excitation source. The durations of the medial closures are unmarked in the figure, but they are taken to be the intervals between the vocalic spans, as seen in the spectrograms and in the waveforms. Acoustic evidence of glottal pulsing at the low end of the spectrum, appropriate to the voiced consonant /g/ is seen in these intervals. For each disyllabic word, then, the duration ratios of the first and second vowel and the first vowel and the medial closure were obtained.

2.3. RESULTS. The amplitude ratios are best shown in numerical tables; however, as a possible aid to the reader, a graphic display of one set of measurements is given in Figure 3. The large dots give the mean amplitude ratios of the first syllable to the second syllable for words beginning with short and long voiceless stop consonants. The vertical bars show one standard deviation about the mean. Note that values below the horizontal line are negative; this indicates a lower amplitude in the first syllable than in the second. (Obviously, some of the values are positive; the amplitude display for /pagi/ with a short initial stop in Figure 2 happens to be such a case.) The values for the mean and the full range of the standard deviation for the long stop are positive, indicating a higher amplitude in the first syllable.

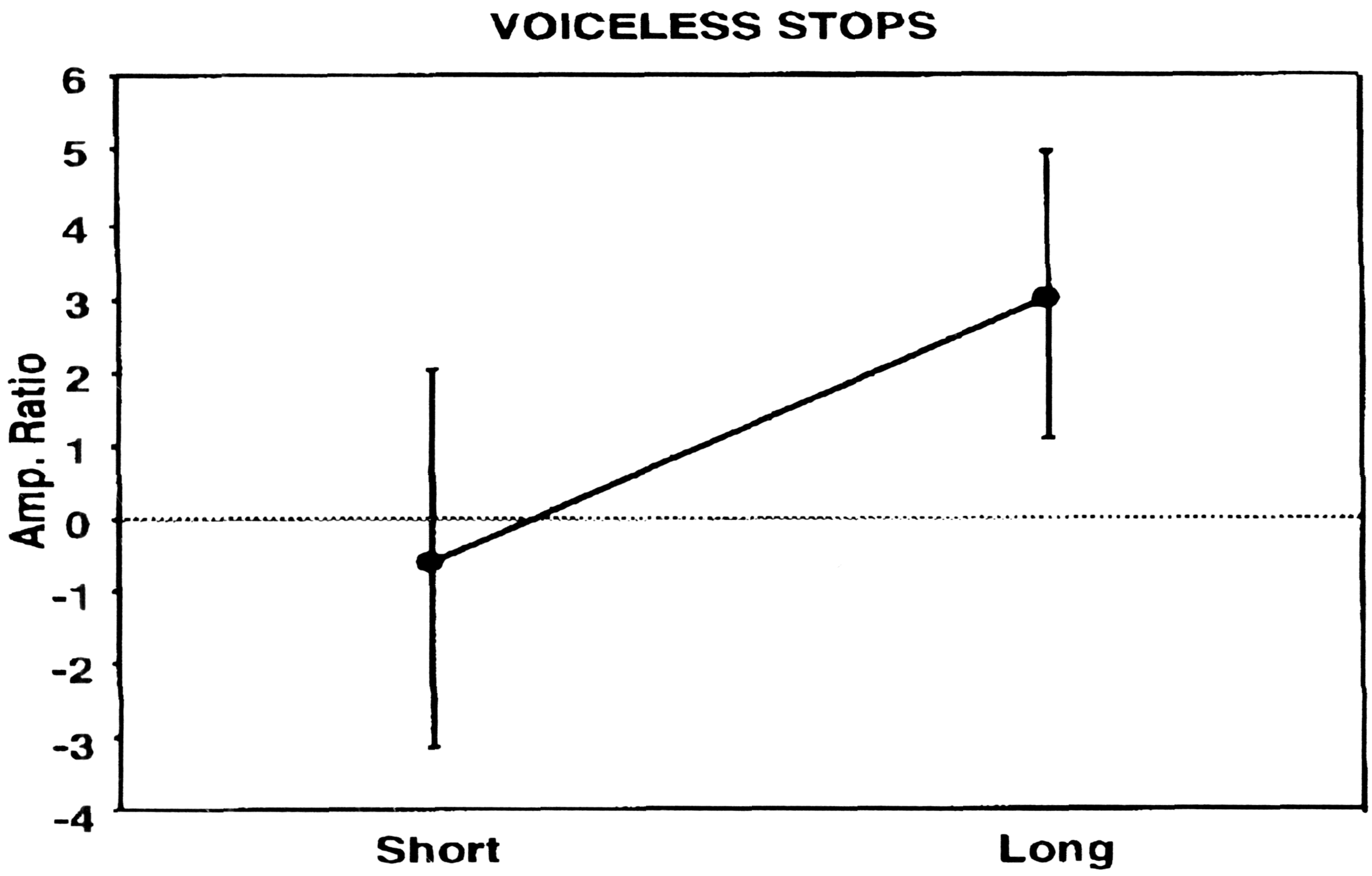

FIGURE 3. Means and standard deviations of the amplitude ratios of the first and second syllables of Pattani Malay words beginning with short and long voiceless stops: Four speakers.

The data underlying the graph for the voiceless stops in Figure 3 are given in Table 1 together with comparable amplitude ratios for the voiced stops and continuants. As stated in the first paragraph of Section 2.2, only two of the speakers, A and B, recorded exactly the same lists of words, so their data are put on one line for each phonetic category here and in the other tables. The data for speakers P and W are necessarily on separate lines. Column 2 shows the numbers of minimal pairs, columns 3 and 4 the means and standard deviations of the amplitude ratios for the short consonants, columns 5 and 6, the means and standard deviations for the long consonants. The three right-hand columns show the results of paired *t*-tests; column 7 gives the degrees of freedom, column 8, the *t*-values, and column 9, the levels of significance.

For the voiceless stops the amplitude ratios in Table 1 are highly significant for all the speakers. For the voiced stops A and B miss significance but not by much, P is highly significant, and W is moderately significant. For the continuants all are significant, although W is only moderately so.

TABLE 1. Amplitude Ratios

Spkrs	Pairs	Short		Long		Paired *t*-Test		
				Voiceless Stops				
	n	M	SD	M	SD	*df*	*t*	*p*
A+B	24	-.7	3.1	3.6	1.9	23	-7.3	<.001
P	14	-.6	2.2	1.7	1.7	13	-3.6	<.004
W	6	-.1	.8	3.6	1.2	5	-6.0	<.002
				Voiced Stops				
A+B	8	-.6	2.5	.9	1.7	7	-2.2	.07,ns
P	12	-.2	1.2	2.4	3.0	11	-4.0	<.003
W	6	2.0	3.0	4.8	1.5	5	-2.6	<.05
				Continuants				
A+B	20	.6	1.6	2.5	1.4	19	-4.5	<.001
P	10	-1.6	1.7	.9	1.7	9	-3.8	<.005
W	14	1.1	2.7	3.9	2.3	13	-2.7	<.02

TABLE 2. Average Slope of Amplitude Rise

Spkrs	Pairs	Short		Long		Paired *t*-Test		
				Voiceless Stops				
	n	M	SD	M	SD	*df*	*t*	*p*
A+B	24	-.4	.1	.5	.1	23	-2.4	<.03
P	14	-.5	.3	.9	.6	13	-2.6	<.03
W	6	-.2	.1	.3	.1	5	-1.4	.23,ns
				Voiced Stops				
A+B	8	.1	.04	.3	.1	7	-10.8	<.001
P	12	.2	.1	.6	.3	11	-2.9	<.02
W	6	.2	.1	.4	.1	5	-1.0	.4,ns
				Continuants				
A+B	20	.1	.1	.2	.2	19	-1.8	.1,ns
P	10	.3	.2	.2	.1	9	1.6	.15,ns
W	14	1.1	.04	1.2	.2	13	-2.5	<.03

In Table 2 are shown the average slopes of amplitude rise. The results for three out of the four speakers are moderately significant. W is not significant. For the voiced stops the same three speakers are significant, with the addition that the results for A and B are highly significant, while W is not significant. As for the continuants, only one speaker, W, shows moderate significance. Curiously enough, he is the only one who does not have significant differences in slope for the other two categories.

Table 3 shows the fundamental-frequency ratios. For the voiceless stops all the speakers show highly significant differences between the short and long categories. For the voiced stops W is once again not significant, A and B are moderately significant, and P is highly significant.

TABLE 3. Fundamental-Frequency Ratios

Spkrs	Pairs	Short		Long		Paired *t*-Test		
		Voiceless Stops						
	n	M	SD	M	SD	*df*	*t*	*p*
A+B	24	.1	.1	1.1	.1	23	-6.1	<.001
P	14	.1	.02	1.1	.6	13	-6.6	<.001
W	6	.1	.1	1.1	.1	5	-4.5	<.007
		Voiced Stops						
A+B	8	1.1	.1	1.2	.1	7	-2.7	<.04
P	12	1.0	.04	1.1	.3	11	-3.6	<.005
W	6	1.0	.1	1.2	.1	5	-1.9	.12,ns
		Continuants						
A+B	20	1.0	.06	1.1	.2	19	-6.9	<.001
P	10	1.0	.04	1.1	.1	9	-3.8	<.005
W	14	1.1	.04	1.2	.2	13	-2.5	<.03

TABLE 4. Vowel-Duration Ratios

Spkrs	Pairs	Short		Long		Paired *t*-Test		
		Voiceless Stops						
	n	M	SD	M	SD	*df*	*t*	*p*
A+B	24	.6	.4	.8	.5	23	-3.8	<.001
P	14	.4	.1	.5	.1	13	-3.7	<.003
W	6	1.0	.5	1.0	.5	5	.8	.47,ns
		Voiced Stops						
A+B	8	.5	.2	.7	.3	7	-3.6	<.009
P	12	.5	.3	.6	.3	11	-2.7	<.02
W	6	.7	.4	.9	.3	5	-1.1	.3,ns
		Continuants						
A+B	20	.6	.2	.8	.3	19	-5.2	<.001
P	10	.5	.2	.6	.2	9	-5.2	<.001
W	14	.7	.3	.8	.3	13	-.7	.5,ns

TABLE 5. Duration Ratios of Vowel 1 and the Medial Closure

Spkrs	Pairs	Short		Long		Paired *t*-Test		
		Voiceless Stops						
	n	M	SD	M	SD	*df*	*t*	*p*
A+B	24	1.9	3.2	1.8	1.7	23	.1	.94,ns
P	14	1.0	.4	1.5	.5	13	-4.1	<.002
W	6	1.1	.1	1.2	.3	5	-1.3	.26,ns
		Voiced Stops						
A+B	8	1.6	.7	2.3	1.1	7	-3.3	<.02
P	12	1.1	.4	1.4	.4	11	-2.0	<.07,ns
W	6	.3	.7	2.3	1.5	5	-2.7	<.05
		Continuants						
A+B	20	1.5	.8	1.8	1.2	19	-2.9	<.001
P	10	.8	.2	1.3	.4	9	-6.0	<.001
W	14	.4	.1	1.4	.5	13	-.7	.52,ns

The duration-ratios of the first vowel to the second are given in Table 4. The findings for W's voiceless stops are not significant. The differences between the ratios for the other three speakers, however, are highly significant. For the voiced stops, we see that W is once again not significant, A and B are highly significant, and P is moderately significant. As for the continuants, W is not significant, but the others are highly significant.

The last dimension, the ratio of the duration of the first vowel and that of the medial consonantal closure, is presented in Table 5. For the voiceless stops only P is significant and highly significant at that. As for the voiced stops, P is the only speaker whose data are not significant although not by much. The others are moderately significant. For the continuants, W is not significant, while the others are highly significant.

CONCLUSION

3.1. SUMMARY OF RESULTS. It seems clear that all the acoustic properties examined in this study tend more or less to help in the differentiation of disyllabic words beginning with short and long initial consonants.

Table 6 is an attempt at summarizing the results in a simple way. Cells are provided under each acoustic property at the intersections between every phonetic category and each speaker or, for A and B, pair of speakers. An entry 'Yes' in a cell means the achievement of a level of significance of at least $p < 0.05$. 'No' means failure to achieve that level; that is, it is not significant.

As for the voiceless stops, the speakers show positive results for amplitude ratio, amplitude slope, and fundamental-frequency ratio. The remaining two properties, vowel-duration ratio and the duration-ratio of the first vowel to the medial closure, are valid for all the speakers except W.

The voiced stops are positive in two out of three cells for every category. All the speakers have 'No' in one phonetic category or another, but speaker W has it more than anyone else.

The best factors for the continuants are the amplitude ratio and the fundamental-frequency ratio. All the rest have 'No' entered once or twice.

TABLE 6. Summary of the Measurements

	Amplitude Ratio		
Speakers	A+B	P	W
Voiceless Stops	Yes	Yes	Yes
Voiced Stops	No	Yes	Yes
Continuants	Yes	Yes	Yes
	Amplitude-Slope Ratio		
Speakers	A+B	P	W
Voiceless Stops	Yes	Yes	No
Voiced Stops	Yes	Yes	No
Continuants	No	No	Yes
	Fundamental-Frequency Ratio		
Speakers	A+B	P	W
Voiceless Stops	Yes	Yes	Yes
Voiced Stops	Yes	Yes	No
Continuants	Yes	Yes	Yes
	Vowel-Duration Ratio		
Speakers	A+B	P	W
Voiceless Stops	Yes	Yes	No
Voiced Stops	Yes	Yes	No
Continuants	Yes	Yes	No
	Duration-Ratio of Vowel 1 and Medial Closure		
Speakers	A+B	P	W
Voiceless Stops	No	Yes	No
Voiced Stops	Yes	No	Yes
Continuants	Yes	Yes	No

3.2. DISCUSSION. The speech-production data of this study support the hypothesis (§2.1) that in Pattani Malay several acoustic properties give more salience to the first syllable of disyllabic words beginning with a long voiceless stop consonant than to those beginning with a short voiceless stop. Indeed, this criterion works well for the voiced stops too and, less well, for the continuants. If we count the number of Yes entries, allowing two for each A+B cell, a rank order comes out of Table 6: (1) F_0 ratio, (2) amplitude ratio, (3) vowel-duration ratio, (4) amplitude-slope ratio and duration-ratio of vowel 1 and medial closure. The differences, however, are small.

It is to be recalled, of course, that 'long' consonants do have significantly longer closure durations than their 'short' counterparts, at least for all contexts in which they can be measured acoustically (Abramson 1987). Thus closure duration, the property originally thought to be the sole relevant one, appears to function in combination with a set of prosodic factors in differentiating the two categories in speech production.

What remains to be done is perceptual validation of these findings. My plan, with the earlier perceptual studies as a starting point (Abramson 1986, 1991), is to do perception testing in two contexts, utterance-initial and intervocalic. Stimuli will be prepared with variants along the dimensions measured in this study pitted against each other for presentation to native speakers of Pattani Malay for identification. If the perceptual efficacy of syllable prominence, as provided by combinations of these acoustic cues, is demonstrated, it will be desirable to do studies of motor control to determine whether it is the single factor of closure duration that underlies the several acoustic differences. A possible outcome is that for voiceless initial long stops, and maybe some other consonants, speakers of the language have come over time to enhance the distinction by adding gestures of a prosodic type. If so, it remains to be seen whether this will lead to the emergence of an accentual system.

NOTES

1. This work was supported by Grant HD–01994 from the National Institute of Child Health and Human Development to Haskins Laboratories, New Haven, Connecticut. I am grateful to members of the Department of Islamic Studies of The Prince of Songkhta University, Pattani, for their hospitality and help.

2. Thus, for Turkish voiceless stops, the distinction between 'geminate' and 'non-geminate' consonants includes significant differences between voice onset time (VOT) as well as closure duration; nevertheless, whatever mechanism may be responsible for the VOT difference, experimental

manipulations show that the single overriding cue in perception is closure duration (Lahiri & Hankamer 1988).

3. This description is correct for the traditional terminal-analogue synthesizer. In the newer articulatory synthesizers, the parameter-values are instructions to simulated articulators that determine the configuration of the modeled vocal tract, which when provided with a "voice," emits a speechlike signal.

4. Another possibility is to feed an extraneous source into the vocal tract of a human being who articulates speech with this unnatural carrier. In my childhood in the United States, a soap company, extolling the deodorant properties of its soap, warned against the dangers of body-odor (B.O.) by having a person articulate 'B.O.' with the sound of a foghorn in place of his own voice.

5. It does not pay here to fret that even a single gesture may be physiologically complex and that the turbulence of the fricative has its own spectral and temporal complexity.

6. For a comparison of the consonants of Malay regional dialects and those of Standard Malay see Thavisak (1987: 11–14).

7. In at least one instrumental study (Lehiste, Morton & Tatham 1973) electromyographic data are interpreted as supporting the conclusion that Estonian 'geminate' consonants are rearticulated.

REFERENCES

Abramson, Arthur S. 1977. Laryngeal timing in consonant distinctions. Phonetica 34.295–303.

———. 1986. The perception of word-initial consonant length: Pattani Malay. Journal of the International Phonetic Association 16.8–16.

———. 1987. Word-initial consonant length in Pattani Malay. Proceedings of the XIth International Congress of Phonetic Sciences, 6–70. Tallinn: Academy of Sciences of the Estonian S.S.R.

———. 1991. Amplitude as a cue to word-initial consonant length: Pattani Malay. Actes du XIIème congrès international des sciences phonétiques, Vol. 3, 98–101. Aix-en-Provence: Université de Provence.

Browman, Catherine P., and Louis Goldstein. 1986. Towards an articulatory phonology. Phonology Yearbook 3.219–252.

Chaiyanara, Phaythuun M. 1983. Dialek Melayu Patani des bahasa Malaysia [The Pattani Malay dialect of the Malay language]. Kuala Lumpur: University of Malaya master's thesis.

Denes, Peter B. 1955. Effect of duration on the perception of voicing. Journal of the Acoustical Society of America 27.761–764.

Fowler, Carol A., and Elliot Saltzman. 1993. Coordination and coarticulation in speech production. Language and Speech 36.171–195.

Fry, Dennis B. 1955. Duration and intensity as physical correlates of linguistic stress. Journal of the Acoustical Society of America 27.765–768.

House, Arthur S., and Gordon Fairbanks. 1953. The influence of consonant environment upon the secondary acoustical characteristics of vowels. Journal of the Acoustical Society of America 25.105–113.

Lahiri, Aditi, and J.E. Hankamer. 1988. The timing of geminate consonants. Journal of Phonetics 16.327–338.

Lehiste, Ilse; Katherine Morton; and Marcel A.A. Tatham. 1973. An instrumental study of consonant gemination. Journal of Phonetics 1.131–148.

Liberman, Alvin M., and Ignatius G. Mattingly. 1985. The motor theory of speech perception revised. Cognition 21.1–36.

Lisker, Leigh. 1974. On time and timing in speech. Current trends in linguistics, Vol. 12, ed. by T.A. Sebeok; A.S. Abramson; D. Hymes; H. Rubenstein; E. Stankiewicz; and B. Spolsky, 2387–2418. The Hague: Mouton.

Lisker, Leigh, and Arthur S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. Word 20.384–422.

Löfqvist, Anders; Thomas Baer; Nancy S. McGarr; and Robin Seider Story. 1989. The cricothyroid muscle in voicing control. Journal of the Acoustical Society of America 85.1314–1321.

Peterson, Gordon E., and Ilse Lehiste. 1960. Duration of syllable nuclei in English. Journal of the Acoustical Society of America 32.693–703.

Raphael, Lawrence J. 1974. The physiological control of durational differences between vowels preceding voiced and voiceless consonants in English. Journal of Phonetics 3.25–33.

———. 1981. Durations and contexts as cues to word-final cognate opposition in English. Phonetica 38.126–147.

Thavisak, Amon. 1987. ภาษามลายูถิ่นในประเทศไทย [Malay dialects in Thailand]. Nakhorn Pathom, Thailand: Institute of Language and Culture for Rural Development, Mahidol University.