1127

Proceedings of The XIVth International Congress of Phonetic Sciences

San Francisco
1-7 August 1999

Sponsored by: University of California, Berkeley
Co-sponsored by: Acoustical Society of America, Australian Speech Science & Technology Association, European Speech Communication Association, International Phonetic Association, Linguistic Society of America

Edited by: John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville & Ashlee C. Bailey

FUNDAMENTAL FREQUENCY AS A CUE TO WORD-INITIAL CONSONANT LENGTH: PATTANI MALAY

Arthur S. Abramson
Haskins Laboratories and the University of Connecticut

ABSTRACT

All word-initial Pattani Malay consonants are distinctively short or long. The dominant perceptual cue is consonant-closure duration. In voiceless plosives, however, closure-duration is audible only utterance-medially after a vowel, yet native speakers identify such words in isolation well and so must be using other cues. Disyllabic words beginning with long consonants seem to have greater salience on the first syllable. Measurements revealed significant differences in both peak amplitude and fundamental frequency (F0). Relative amplitude had a significant perceptual effect but was not a sufficient cue by itself. In this study increments of F0 were imposed on words with short initial consonants and decrements on words with long initial consonants. The stimuli were played to 30 native speakers for identification. The effect for voiceless stops was highly significant; nevertheless, F0 too is not a sufficient cue by itself. The next step is to combine the two properties, amplitude and F0.

1. BACKGROUND

1.1. Distinctive length

Pattani Malay is an Austronesian language spoken by about a million people in the southernmost provinces of Thailand along the border with Malaysia [1]. It is very unusual in having the feature of distinctive length in word-initial position for all phonetic classes of consonants.[1] This characteristic also has the consequence of letting such words appear in utterance-initial position.

For such consonantal categories as nasals, laterals, and fricatives, such an unusual characteristic ought to present no problem in the production and perception of the contrast. Whether the closure or constriction of any of them is relatively short or long, it can easily be pronounced with accompanying acoustic excitation even at the beginning of an utterance. In addition, such excitation, whether voiced or voiceless, is likely to be loud enough to signal relative closure durations to the listener.

The matter of stop consonants is more troublesome. The voiced stops of Pattani Malay have glottal pulsing during their closures. Long voiced stops would presumably require enough compensatory maneuvering to provide for the continuation of the transglottal air pressure drop needed for a significantly longer voiced closure than that of the short voiced stop. Also, the fairly low amplitude of this pre-release voicing, limited to the low end of the spectrum, might not be as audible as the voiced closures of the other consonant categories. Finally, we have to deal with the voiceless unaspirated stops of the language with no acoustic excitation during the closures. In a word-pair like /tawa/ 'bland' vs. /t:awa/ 'to show wares,' there can be no direct signal of the relative durations of the stop closures in utterance-initial position, although there would be in utterance-medial position after a vowel through the relative durations of the silent gaps.

1.2. Previous work

Measurements of the acoustic manifestations of consonantal closures in the language [3] revealed highly significant differences in duration for both utterance-initial and utterance-medial positions.[2] Of course, without the means to make direct measurements of articulation, voiceless stops could be examined only in utterance-medial position.

Other experiments [4] demonstrated the perceptual efficacy of relative duration as a sufficient and powerful cue to the distinction. The stimuli for these experiments were made by lengthening the closures of short consonants and shortening the closures of long consonants in a series of 10-ms steps in disyllabic words. The predominant form of the word in the language is apparently disyllabic.

Since control tests with unaltered isolated words revealed a high level of identifiability, even for utterance-initial voiceless stops [4], the question remained as to how native speakers could so reliably label the latter without any acoustic signal during the closures. Also, in those tests in which both members of a manipulated pair had voiceless initial stops embedded in utterance-medial position, the perceptual crossover points along the time axis in going from short to long and long to short categories differed significantly from each other, suggesting that other acoustic cues were influencing the placement of the category boundary.

In close listening to the productions of several speakers, prompted by the foregoing results, the first syllable struck the ear as more salient relative to the second syllable for words beginning with a long consonant but not for those beginning with a short consonant.[3] These observations led to the proposing of a prosodic hypothesis, namely that the differentiation of short and long utterance-initial voiceless stops is a function of properties of the acoustic signal, concomitant with closure duration, that give greater salience to the first syllable of words beginning with long stops. While this may also be true to a certain extent for affricates and voiced stops, it is not likely to be relevant for the rest of the consonant categories.

To assess the physical plausibility of this hypothesis, measurements were made in the utterances of four speakers of several acoustic properties that could contribute to an auditory impression of salience (accentual prominence) on the first syllable of disyllabic words [3, 5]. Two highly significant factors that stood out were higher amplitude and higher fundamental frequency (F0) on the first syllable of words beginning with long

Category	Pairs	Short C	Gloss	Long C	Gloss
Nasal	2	makɛ	to eat	m:akɛ	to be eaten
Lateral	1	lama?	late	l:ama?	cause to be late
Voiced fricative	1	yatɔ	comprehensive	y:atɔ	to spread out
Voiceless fricative	3	sɛpa?	to kick	s:ɛpa?	to be kicked
Voiceless affricate	1	cabɛ	branch	c:abɛ	side road
Voiced stops	6	buŋɔ	flower	b:uŋɔ	to bloom
Voiceless stops	7	kukoh	to be stable	k:ukoh	to render stable

Table 1. Phonetic categories in the control tests, the number of minimal pairs used for each, and a sample pair for each category.

consonants, especially voiceless stops. Other somewhat less reliable factors for the first syllables of words with long consonants were faster amplitude rise-time, greater vowel duration, and higher ratio of the vowel duration to the duration of the word-medial consonant closure.

For the perceptual validation of the prosodic hypothesis, it seemed prudent to begin by testing each of the two principal factors one at a time. In the one study completed and published up to now [6], two experiments with relative amplitude were conducted with voiceless stops. In the first, amplitude was pitted against duration that changed in 20-ms steps, increasing for original short closures and decreasing for original long closures. The first syllable of each durational variant of an original short stop was increased in amplitude in five 2-dB steps; for original long consonants, the first syllable of each variant was *decreased* in five 2-dB steps. The responses of 30 native speakers made it clear, as before, that when it is available to the ear, the cue of relative duration is too powerful to be overcome; nevertheless, the effects of amplitude on the perceptual boundaries between the length categories were highly significant.

In the second experiment reported in the same paper [6], the efficacy of relative amplitude was tested in the absence of any acoustic cue to closure duration by playing the amplitude variants of the original short and long voiceless stops of one of the word pairs as isolated words to the same 30 native speakers. Here too, amplitude definitely influenced perception in that the identification curve for each category was gradually shifted toward the opposite category, approaching the 50% crossover point but never crossing it. That is, it is a fairly weak cue, insufficient by itself for unambiguous assignment by listeners to a length category.

2. PROCEDURE AND RESULTS

2.1. Introduction

Against the background of the foregoing findings, the purpose of my new research was to investigate the efficacy of the other major acoustic property found to be concomitant with relative duration, namely F0, as a cue to the perception of the length difference in Pattani Malay word-initial consonants, especially stops. As before, all my experiments were run in a language laboratory on the campus of the Prince of Songkhla University in the Province of Pattani in southern Thailand.

2.2. Control tests

Because of my concern over the possibility of dialectal variation among my potential subjects, who, as students, come to this campus from various sections of the four provinces in which the language is spoken, and because of occasional rumors about the weakening of the distinction among the young, I always administer control tests to be sure that my new subjects do indeed have the distinction and to establish a baseline with which to compare the effects of my manipulations. Table 1 shows the phonetic categories chosen for the test, the number of word pairs used for each category, and a word pair to exemplify each category. Using the recorded speech of an adult male, I prepared seven tests, each one containing a randomization of six tokens of each member of three of the 21 pairs. Each word presented was repeated immediately to give the subjects time to identify it before the paired presentation of the next word. My 30 undergraduate subjects (17 women, 13 men), all native speakers of Pattani Malay, gave written responses to all seven tests.

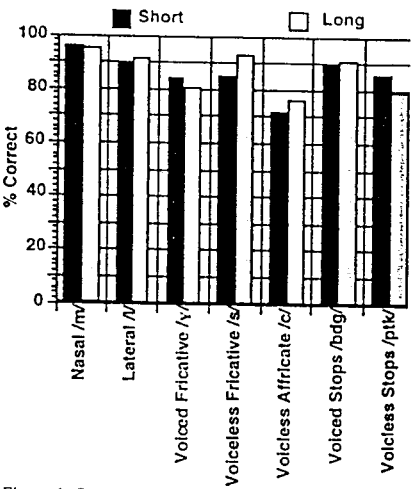

Figure 1. Control tests: Identification of isolated words.

In the Malay culture of southern Thailand it is not the practice to read and write in Pattani Malay, so there is no widely known form of writing for the language. Instead, literate people read and write Standard Malay and Standard Thai. As a result, all my answer sheets were prepared with Thai glosses for the Pattani Malay words. The subjects encircled their choices.

The results of the control tests are given as percentages correct in Figure 1. Even with identifications at levels somewhat better than chance, the voiceless affricates did not fare very well and so were not included in the next set of experiments. Given their very audible resonances, it is not surprising that the nasal consonants, with /m/ at 96% and /m:/ at 95%, were the best identified. Even the voiceless stops with their silent closures were rather well identified. In general, then, it can be said that the distinction is still rather robust for this generation of speakers.

2.3. F0 experiments

To create stimuli with controlled changes in F0 I used the Sound Designer II© program on the Macintosh computer. Starting with 0 semitones (the original word), I increased the F0 of the first syllable of each word beginning with a short consonant in six half-semitone steps up to 3 semitones, yielding seven stimuli. Likewise, I decreased the F0 of the first syllable of each word with a long consonant in six half-semitone steps, yielding seven more stimuli, 14 altogether for each minimal pair. These ranges bracketed those found in [5]. Working with the semitone as my unit, borrowed from the domain of music, was intended to give the impression of pitch changing in more or less equal intervals, although, of course, the subjects heard only randomizations in the listening tests.

For these tests I used a subset of the word-pairs in 2.2: one pair of nasals, one of voiceless fricatives, two of voiced stops, and four of voiceless stops. The voiceless stops, of course, form the most critical case for the present research. There was one test for each word-pair. With two tokens of each stimulus, there were 28 items, in paired presentations, in each test. The same 30 subjects as in 2.2 were told to label each item as one of two possible words printed on the answer sheet. If in doubt about an item, they were to guess and not leave the space blank on the answer sheet.

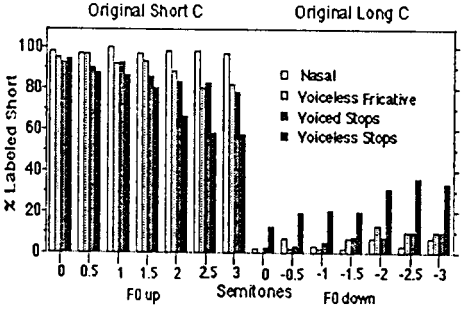

Figure 2. Identification of the F0 variants.

The labeling results are given in Figure 2. The treatment of the variants with original short consonants is shown by the tall bars along the left half of the horizontal axis, where F0 moves upward, while the treatment of the variants with original long consonants is shown by the short bars along the right half, where F0 moves downward, as indicated by the negative sign. Except, of course, for the unchanged items (0 semitones), there are no "correct" responses. Instead, the vertical axis shows the percentage of each phonetic category labeled short.[4]

Although the effects are not large, an analysis of variance, with the p-value obtained with the Huynh-Feldt adjustment, showed a highly significant interaction of phonetic class, direction of F0 change, and F0 height (F (18, 522) = 8.6, p = .0001). We can see in Figure 2 that the short voiceless stops are gradually shifted toward the long category, while the long voiceless stops are shifted somewhat spasmodically toward the short category. For the voiced stops there is a similar but smaller effect. The other classes are hardly affected, the nasals not at all.

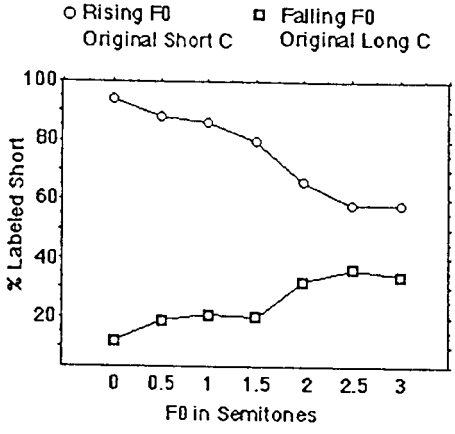

Figure 3. Identification of F0 variants of voiceless stops.

In Figure 3 the data for the class of voiceless stops have been taken from Figure 2 and plotted separately. The horizontal scale in semitones is to be taken as rising for the original short consonants in the upper curve and falling for the original long consonants in the lower curve. Again, the vertical axis gives the percentages labeled short. It is perhaps clearer here that the F0 shifts have a considerable effect; nevertheless, while the curves approach the 50% crossover point, they do not reach it, much less go beyond it.

DISCUSSION AND CONCLUSION

The role of relative fundamental frequency in the perceptual identification of short versus long consonants in Pattani Malay is apparently very similar to that of amplitude [6]. Its effect is highly significant, especially with regard to voiceless stops. At the same time, by itself it is far from being a sufficient cue for the perceptual separation of the two length categories.

This study, then, has assessed the role of one more acoustic property previously found [5] to be reliably concomitant with relative closure duration, F0, and found it to have a very significant effect, yet we have still not reached the goal of explaining how isolated words beginning with voiceless stops are so well identified as to length category. My view is that the prosodic hypothesis should certainly not be discarded yet; it requires more research. My plan is to design new stimuli in which

the properties of relative amplitude and relative fundamental frequency are carefully combined to act together in as natural a way as possible. To enhance the naturalness of the resulting stimuli it may be necessary also to make acoustic alterations of the second syllable as well.

In my desire to avoid overly complex stimuli, I plan to make a few series of them, starting with a relatively simple one in which amplitude and F0 co-vary in small steps. If that brings about an improvement over the results obtained so far but still does not yield perceptual categories as good as the natural ones, I can then complicate the stimuli by making changes in the second syllables and even include the additional properties found in [5], namely, relative amplitude rise time, vowel-duration ratios, and the ratio of the duration of the first vowel to the duration of the word-medial consonant closure. Not to be ignored is the possibility that I have missed yet another acoustic property of even greater importance.

Historically the length distinction in the language is reconstructed as having developed from phonetic changes at morpheme boundaries [1, 7]. At this time, there can be no doubt that, when present and audible, relative closure duration is a powerful and sufficient cue to the distinction. If, however, more experimental evidence emerges to support the prosodic hypothesis for at least voiceless initial stops, we may be witnessing a gradual phonological development leading to an accentual system for some contexts. If so, we might suppose that certain secondary characteristics entailed by the nature of the articulatory gestures needed for greater closure duration have become somewhat enhanced in the behavior of the speakers, who have learned to use them to help sharpen the phonological length distinction in both production and perception.

ACKNOWLEDGMENTS

This research was supported by NIH Grant DC-02717 to Haskins Laboratories. The cost of the author's travel to and from Thailand for the perriod 3 December 1997 to 3 March 1998 was paid by grants from the University of Connecticut Research Foundation. In Bangkok he was a guest of the Department of Linguistics, Faculty of Arts, Chulalongkorn University, where much help was given by the Head of the Department, Dr. Sudaporn Luksaneeyanawin, and her colleagues, as well as by the Dean of the Faculty, Dr. Kalaya Tingsabadh. The experiments were run at the Prince of Songkhla University, Pattani with the excellent help of Dr. Tananan Trongdee and Mr. Ruslan Uthai. The kind and generous hospitality of Mr. Wiboon Thansakun was greatly appreciated.

NOTES

1. It is unusual but not unique. For example, Taba, an Austronesian language of Indonesia, is rather similar [2].

2. Jimmy G. Harris (p.c.) has palatograms showing a wider articulatory contact for the long consonants.

3. Christopher Court of the University of Hawaii, who has had much experience with the language, has independently had the same impression (p.c.).

4. This is a matter of convenience in notation. The subjects, the reader will recall, actually decided for each item which of two possible words was being said.

REFERENCES

[1] Paramal, W. 1991. *Long consonants in Pattani Malay: The results of word and phrase shortening.* Unpublished M. A. thesis, Mahidol University, Bangkok.

[2] Bowden, J. and Hajek, J. 1996. Taba. *Journal of the International Phonetic Association*, 26, 55-57.

[3] Abramson, A. S. 1987. Word-initial consonant length in Pattani Malay. *Proceedings of the 11th International Congress of Phonetic Sciences, vol. 6.* Tallinn: Academy of Sciences of the Estonian S. S. R.

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

[5] Abramson, A. S. 1998. The complex acoustic output of a single articulatory gesture: Pattani Malay word-initial consonant length. In Warotamasikkhadit, U. and Panakul, T. (eds.), *Papers from the Fourth Annual Meeting of the Southeast Asian Linguistics Society, 1994.* Tempe, Arizona: Program for Southeast Asian Studies, Arizona State University.

[6] Abramson, A. S. 1991. Amplitude as a cue to word-initial consonant length: Pattani Malay. *Proceedings of the 12th International Congress of Phonetic Sciences, vol. 3.* Aix-en-Provence: Université de Provence.

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