The Stability of Distinctive Vowel Length in Thai

Arthur S. Abramson

Haskins Laboratories and The University of Connecticut

INTRODUCTION

Many languages have phonological distinctions of quantity in consonants or vowels or both. Among them, Italian is known for its word-medial intervocalic short and long consonants, while Pattani Malay (Abramson, 1987) is unusual in having word-initial prevocalic short and long consonants. Swedish, some dialects of German, and Thai have short and long vowels. Finnish has a length distinction for both consonants and vowels. Such distinctive length in segments is, of course, different from other communicatively relevant roles of timing in speech, e.g., in stress and intonation.

The obvious physical correlate of the length distinction in phonetic segments is relative duration. That is, in the simplest case the articulatory configuration is held longer for the "long" segment than for the "short" one. Limiting our attention here to vowels, we note an important observation made by Daniel Jones (1950, p. 28): "In languages where vowel length is significant it very often happens that the quality of a long vowel is not quite the same as that of the corresponding short vowel." Ilse Lehiste (1970, pp. 30–33) amplifies the point by commenting that in "quantity" languages some differences in the phonetic quality of short and long vowels can be observed, although such languages differ somewhat in the amount of correlation between length and quality. To the extent that relative duration *is* the primary differentiator of the two classes of vowels, some linguists may prefer to handle the timing difference phonologically as one of gemination rather than distinctive length. Gemination means that what I have been calling a long segment is in fact a sequence of two instances of the same speech sound. This implies rearticulation at the onset of the second occurrence of the segment. Auditory impressions and acoustic observations suggest strongly that such rearticulation is highly unlikely, especially within a single morpheme; nevertheless, whether or not such an argument is tenable phonetically is not a likely outcome of the data to be presented in this paper.[1]

The language of concern here is Standard Thai, the official language of Thailand. It is the standard variety of Central Thai, the regional dialect of Bangkok and a sizable area around it. Traditional Thai grammar posits nine short vowels and nine long counterparts, as well as various diphthongs and vowel clusters. Linguists working on the language, both Thai and foreign, generally accept this view, although some may prefer to transcribe the long vowels as geminates (Tingsabadh & Abramson, 1999).

In my own early experimental phonetic approach to Thai (Abramson, 1962, chap. 2; cf. also Abramson, 1974), I examined the vowel-length contrast in isolated vowels, word-pairs in carrier sentences, and a small sampling of running speech. The resulting acoustic data clearly supported relative duration as the major differentiator of the two

[1]Formal linguistic criteria may make it convenient to posit gemination, even when no phonetic evidence supports this analysis. An example is the presence of a morpheme or word boundary within the long segment. See Dunn (1993) for data supporting the probability of "unitary" geminates (long consonants) in Finnish but the probability of overlapping articulatory gestures in Italian.

classes of vowels. The average ratio of long vowels to short vowels was 2.9 for isolated vowels, 2.5 for the pairs in carriers, and 2.5 for running speech. In addition, experiments on perception demonstrated that for native speakers of the language relative duration provides a sufficient auditory "cue" for this phonemic distinction. At that time, the stimuli for the listening tests were made by shortening original long vowels in minimal pairs of words to values within the ranges of their short counterparts. More recently (Abramson & Ren, 1990), computer-manipulation allowed us also to lengthen original short vowels incrementally. Work by other investigators (Sittachit, 1972; Saravari & Imai, 1983; Gandour, 1984; Gandour & Dardarananda, 1984; Gandour, Weinberg, Petty, & Dardarananda, 1987; Svastikula, 1986) confirms the role of relative duration.

THE ROBUSTNESS OF ACOUSTIC CUES

The work being presented here is part of a larger endeavor, one that seeks to investigate the stability of acoustic cues to phonemic distinctions in a range of styles of speech. The term *acoustic cue* or just *cue* was coined by the Haskins Laboratories group in the early fifties.[2] Acoustic analysis of utterances in a language should yield certain properties that differentiate one class of phonemes from all other classes in the system; furthermore, a more detailed breakdown of each such class should reveal subcategories of such properties that serve to differentiate the phonemes within the class. Experiments may show that these properties not only separate phonemes in speech production but are also sufficient to distinguish them in perception. The latter does not automatically follow from the former, since a phonemic distinction could rest on several properties with varying amounts of power as information-bearing elements for perception. A property with such power in speech perception is called an acoustic cue. Examples are shifts upward and downward in frequency of beginnings and endings of formants (resonances of the vocal tract) for the place of articulation of stop consonants, relative frequency-heights of formants for vowels, spectral location—higher or lower in frequency—and extent of frication energy for fricatives, and, for our purposes here, the relative durations of vocalic stretches for the contrast between short and long vowels.

To this day, most of what we know about the acoustic properties of speech signals and their value as cues, as well as the underlying motor behavior controlled by various physiological mechanisms, comes from the study of short utterances carefully recorded in the laboratory. Such utterances are likely to be isolated words, short expressions, or key words embedded in a carrier sentence. For perception testing, such utterances may be manipulated on the computer along certain dimensions, although most experimental work on perception has used synthetic speech. In perceptual experiments, the listeners' choice of responses may be words or even nonsense syllables that are phonologically "legal" within the language.

In some kinds of phonetic research, for example, prosody, it has long been recognized that one must work with longer spans, usually sentences but maybe even a whole discourse. Much less has been done, however, in the study of vowels and consonants in running speech or even in other styles that are not citation forms. One

[2]For a brief summary of that early work, see Liberman (1957; reprinted with an introduction in Liberman, 1996, pp. 183–198). A good account of the evolution of the concept is to be found in Liberman and Cooper (1972).

expected characteristic of spontaneous speech is less articulatory precision than in citation forms; nevertheless, in the very same spontaneous style of speech a need, from time to time, to be very clear or emphatic may yield somewhat greater precision in the control of articulatory dynamics than in ordinary citation forms. In addition, in unrehearsed running speech, whether casual or deliberate, there is much top-down information from the phonological, morphological, syntactic, and pragmatic contexts. In the classical experiments on the cues, most of the top-down information was kept out of play through the use of isolated citation forms.[3] The work presented here is part of an effort to pursue implications in the literature (Barik, 1977; Levin, Schaffer, & Snow, 1982; Remez, Berns, Nutter, *et al.*, 1991; Laan, 1992) that acoustic differences between spontaneous and read speech are complex. The plan is to study how well phonemic distinctions, as they have been analyzed in citation-form speech in the past, are preserved phonetically in running speech. Furthermore, for the many phonemic distinctions that are no doubt well maintained, we ask whether the acoustic properties linked with the distinctions are easily derived from the cues found in traditional speech-perception research.

The foregoing matters are complicated by overlap between styles. Thus, speech read from written material includes both citation forms and the more or less fluent reading aloud of texts. (Of course, skilled actors can make read or memorized speech sound quite spontaneous.) Running speech includes both read speech and spontaneous talking. Somewhere between the last two is to be fitted the giving of a formal lecture, not from a written text but from an outline. Speakers apparently vary widely in the care with which they project bottom-up phonetic information across these styles. The phonetic precision and thus, perhaps, the perceptibility, of a word is often correlated with recent occurrence of the word in the discourse, familiarity of the topic to the listener, complexity of a task to be performed, surrounding noise level, and other such factors (Lieberman, 1963; Barik, 1977; Levin, Schaffer & Snow, 1982; Fowler & Housum, 1987; Fowler, 1988; Anderson, Bader, Bard, *et al.*, 1991; Remez *et al.*, 1991; Laan, 1992; Kohler, 1992).

My attention will be restricted here to the acoustic examination of the robustness of relative duration as a differentiator of phonemically short and long vowels in Thai. Inasmuch as vowels are notoriously vulnerable to expansion and compression in time as speakers vary their rates of articulation, their speaking styles, their focus on different parts of the discourse, the extent to which a vowel-length distinction is maintained through relative duration alone ought surely to be, in its simplicity, an excellent starting point for my investigation of the robustness of acoustic cues. Other factors, such as formant patterns, that might also serve as cues, even if secondary ones, to a vowel-length distinction (e.g., Straka, 1959; Hadding-Koch & Abramson, 1964; Bennett, 1968; Abramson & Ren, 1990) will not be treated here. Words embedded in short carrier sentences, short expressions, and spontaneous casual conversation will be examined. Although the data should have implications for perception, experiments testing perceptual hypotheses derived from the findings are planned for a sequel to the present study. These hypotheses could include the relevance of other phonetic characteristics in addition to duration.

---

[3]An apparent exception is the set of phonological constraints on syllable types within the language. Since one cannot utter a syllable without invoking such rules, we might argue that we are dealing here for all practical purposes with bottom-up information only.

PROCEDURE

Eight pairs of Thai words, each pair minimally distinguished by vowel length, were recorded in semantically appropriate carrier sentences by four educated native speakers of Standard Thai. The words and a sampling of the sentences are shown in Table 1.[4] The sentences were recorded in a random order. For the first reading, the speakers were asked to use a normal, comfortable rate. For the second reading, they were asked to read faster. Each list of sentences was recorded twice by each speaker. Although in such a procedure the speaking rates were likely to differ widely from speaker to speaker, it was felt that self-determination of normal and fast rates would make for more natural productions.

Table 1. *Minimal Pairs of Words in Sentences*

Words

tìp	'to sip'	tîp	'to flirt'
hèt	'mushroom'	hèt	'cause'
tàk	'to dip up'	tàk	'to dry'
cam	'to remember'	cam	'to sneeze'
khǎj	'to unlock'	khǎj	'to sell'
khùt	'to dig'	khùt	'to scrape'
thun	'fund'	thun	'to carry on the head'
sòt	'fresh'	sòt	'unmarried'

Sample of Sentences

phájajam hǎ hèt hâj khun	'I'm trying to find mushrooms for you.'
phájajam hǎ hè t hâj khun	'I'm trying to find reasons for you.'
jà khùt mâk kn paj	'Don't dig too much.'
jà khù t mâk kn paj	'Don't scrape too much.'
mâj sâp sòt ŕ plàw	'I don't know whether it's fresh.'
mâj sâp sò t ŕ plàw	'I don't know whether he's single.'

To obtain enough unrehearsed conversational speech, I found four members of the staff of Chulalongkorn University, two women and two men, who knew each other well, were quite used to microphones, and did not mind chatting informally about things of interest to them. Two at a time, a man and a woman, sat in a recording booth and talked to each other for ten to fifteen minutes about such topics as events on campus, plans for projects, and vacations. Their speech seemed very natural, varying widely in tempo, emphasis, and clarity. Some of it, not surprisingly, was unusable because of overlapping utterances, laughter, and other distortions.

---

[4]For seven out of the eight pairs, only words with mid and low tones were used, because they were meant originally for perceptual experiments in which the vowels were to be lengthened or shortened (Abramson & Ren, 1990). These tones are least susceptible to distortion in such an operation.