-The Stability of the Thai Three-Way Voicing Distinction in Conversation

Arthur S. Abramson

The University of Connecticut and Haskins Laboratories

1. Background

1.1 *Laryngeal timing in voicing distinctions*

For Standard Thai, earlier acoustic (Lisker and Abramson1964, Abramson 1989) and perceptual (Lisker and Abramson 1970, Abramson and Lisker 1970) studies revealed the importance of the temporal control of the laryngeal aperture (the glottis) in differentiating the categories of stop consonants traditionally called voiced, voiceless unaspirated, and voiceless aspirated.

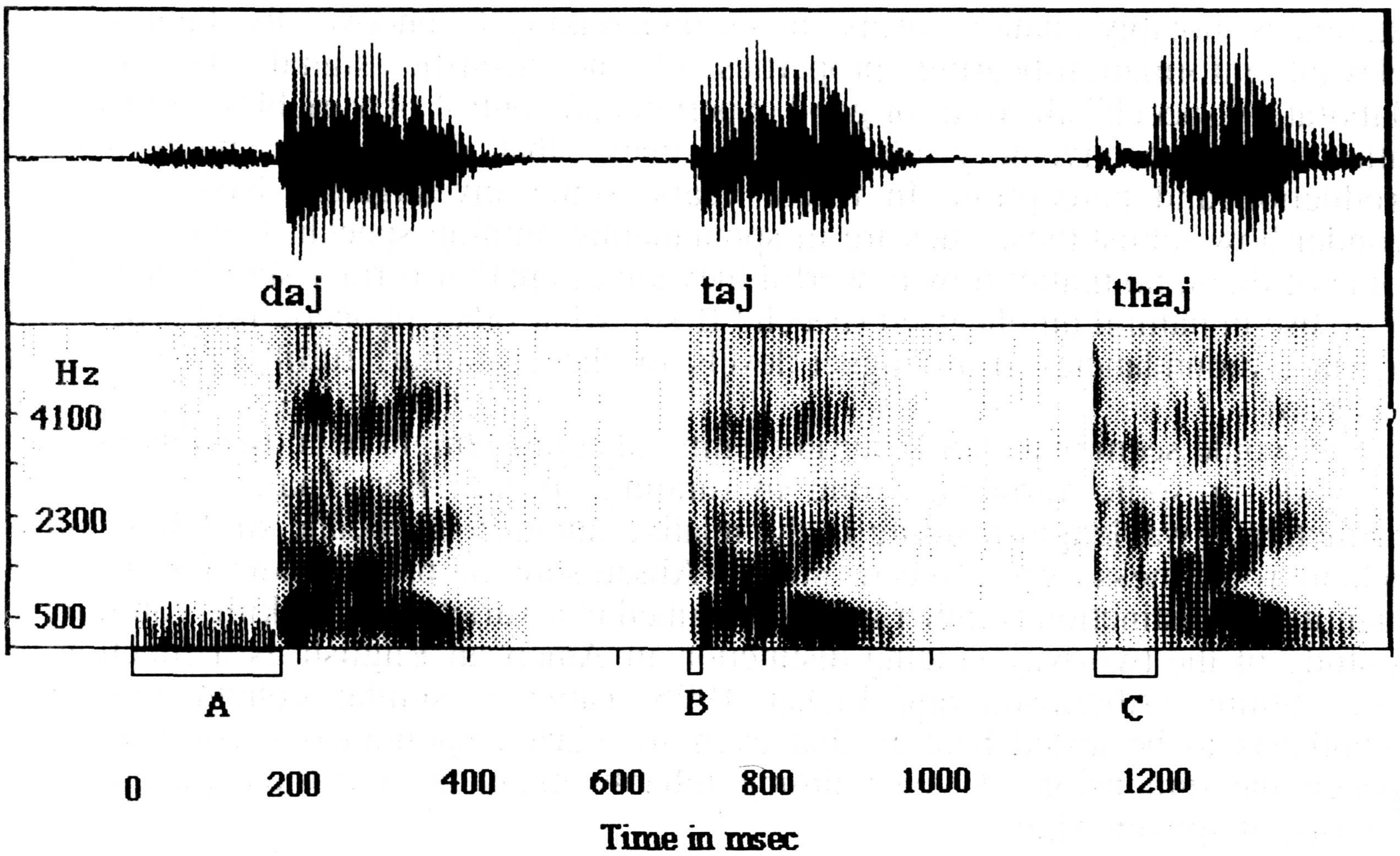

Figure 1. Waveforms above and spectrograms below of a minimal triplet of Thai words differentiated by voice timing. A = voicing lead; B = short lag; C = long lag.

The citation forms of three Thai words in Figure 1 illustrate the situation for utterance-initial position. The common convention is to measure voice onset time (VOT) with reference to the moment of release of the consonantal occlusion, which is assigned the value of 0 on the time-line. Thus, in Figure 1, the word /daj/ 'which' has "voicing lead" with a negative value in temporal units, here milliseconds, relative to zero. The voicing lead is a low-frequency resonance during much or all of the stop closure. This is seen in A in the spectrogram and in the corresponding portion of the waveform above. As is typical of voiceless unaspirated stop consonants, the word /taj/ 'kidney' has a very short "lag" after the release before the onset of glottal pulsing; this is shown in B, where we can see the vertical striation for the release and almost immediately thereafter the onset of voicing. The initial aspirate of /thaj/ 'Thai' has a long voicing lag, as shown in C. It is easy to see that the formant structure of the lag, especially for the higher frequencies, is acoustically excited by turbulence coming from the open glottis; this is the physical correlate of the phonetic property of aspiration. The turbulence is also visible in the beginning portion of the waveform above.

As is well known. among the consonants of Thai voicing is phonologically relevant only for the plosives. In addition, the three-way distinction is found only in the labial and dental stops. That is, the paradigm might be said to be defective for the dorso-velar stops /k kh/ and the alveolo-palatal affricates /c ch/, since neither of the latter two series has a voiced member (Tingsabadh and Abramson 1993). Attention will be given in the section on Procedure to the aligning of the affricates with the velar stops with respect to VOT.

1.2 *The robustness of acoustic cues*

For a long time, the field of experimental phonetics has concentrated on short utterances, notably citation forms, in its undertaking to uncover the linguistically relevant information-bearing properties of the acoustic signal. The use of "laboratory speech" allowed for good experimental control of variables and led to a rather elaborate inventory of "acoustic cues" that differentiate phonemes in production and perception. In recent years, some investigators have begun to wonder how robust these cues are in spontaneous running speech. Could it be that some of them, no matter how powerful they are in citation forms, simply have little or no use in natural unrehearsed speech? If so, what other phonetic properties, and not just redundancies from higher levels of the grammar, fulfill their apparent function?

It would seem that cues requiring temporal control might be most vulnerable to the wide prosodic variation found in running speech. For Thai, for example, distinctive vowel length residing in the relative duration of vocalic stretches, might lack in robustness, yet a recent study (Abramson, in press) indicates that this quantitative distinction is rather well maintained in a variety of speaking conditions. A study of the two-way voicing distinction in American English as a function of voice timing (Abramson and Lisker 1995) came to similar conclusions. The hypothesis to be tested here is that even in relaxed spontaneous conversational speech the dimension of voice timing reliably maintains the three-way voicing contrast of spoken Thai.

2. Procedure

2.1 *Recording*

Recording long sessions of natural conversation on tape is hard to arrange. If you do it unbeknownst to the speakers, ethical questions of privacy arise. Indeed, in some places and circumstances it may even be illegal. For the present work I was fortunate in having four members of the staff of Chulalongkorn University, two men (ST and SA) and two women (PM and TL), who were quite used to microphones and knew each other well enough to relax in a recording booth, one couple at a time, and carry on a very pleasant chat about personal and professional topics of current interest. Each couple consisted of a man and a woman.

Each conversation lasted ten or fifteen minutes. The speech was very natural, varying greatly in tempo. emphasis and clarity. Parts of the conversations, not surprisingly, were unusable because of overlapping utterances, laughter, and other distortions.

Dividing each person's share in a conversation into spans of ten seconds or so, I digitized the speech at a sampling rate of 12,100 Hz for analysis in the Signalyze™ computer program. I identified for analysis all morphemes that I could recognize, in context, to be in the Thai lexicon with particular stops or affricates. For some uncertain instances I had the help of a native speaker of Thai. Items that I could not identify, with or without help, were not analyzed. Anything after a pause was taken to be utterance-initial.

2.2 *Stop consonants*

My procedure was to measure the voice timing of all stops with reference to the closure and its release. The simplest case was that of utterance-initial position, where it was only necessary to do a conventional measurement of VOT, as in Figure 1. Thus A in Figure 2 shows long voicing lag for utterance-initial /kh/. In utterance-medial position I determined whether there was a voicing break, as in C in Figure 2 and A and B in Figure 3.

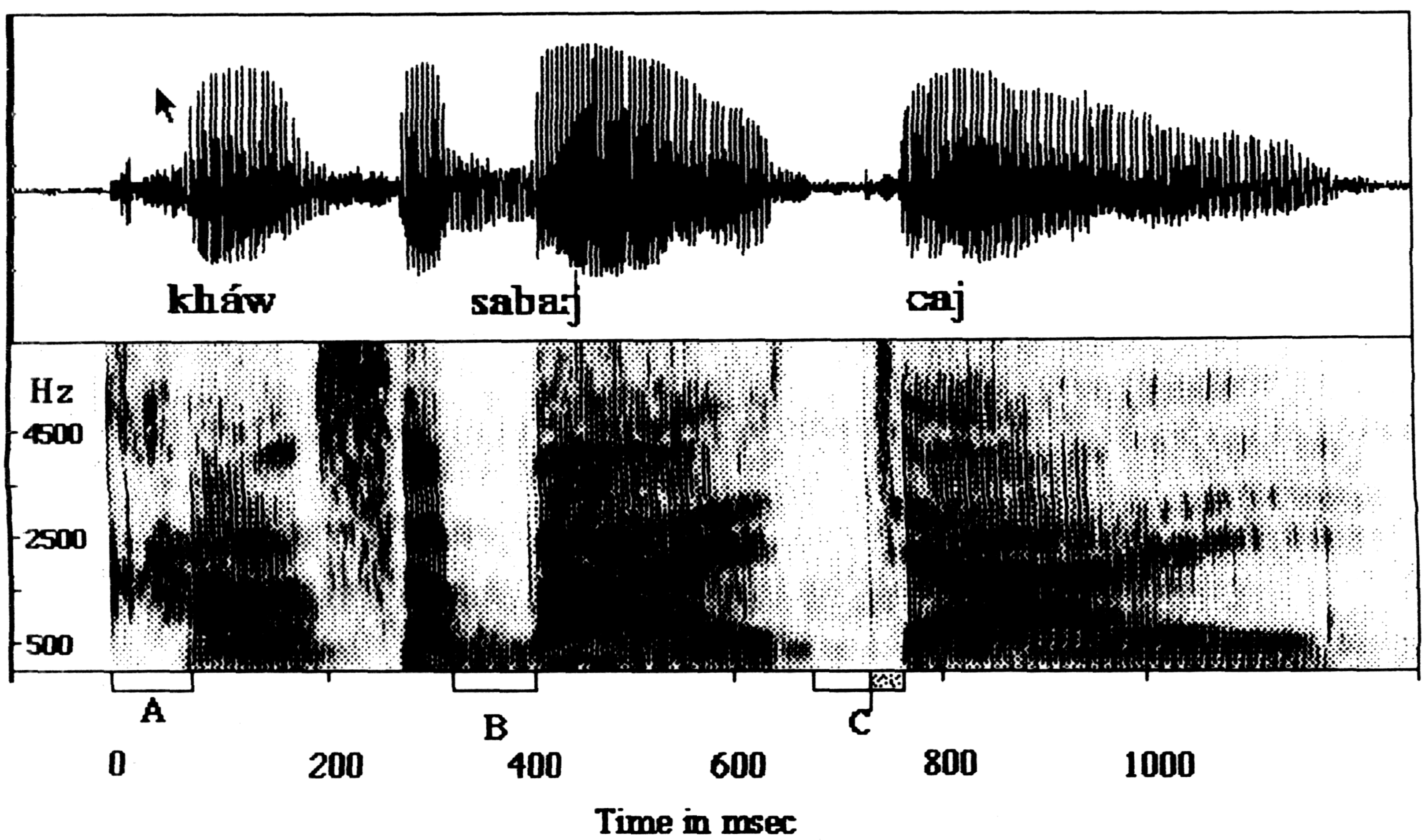

Figure 2. Waveform and spectrogram of an excerpt from the speech of PM. A = long voicing lag for /kh/; B = unbroken voicing for intervocalic /b/; C = full voicing break for intervocalic /c/, with short lag during the frication to the right of the vertical line.

The treatment of the medial occurrences is a bit more complicated. My procedure was to measure the duration of each voicing break in the region of the consonant and to note how much of that break was found before the release and how much after it. Any break after the release is, of course, the same as VOT. This condition is found in A of Figure 3, where the stippled right-hand portion of the break indicates post-release voicing lag. (The affricates in the figures will be discussed separately.) If a nominally voiceless consonant showed glottal pulsing from a preceding context going though its closure or constriction, I could record only a measure of post-release voicing break, i.e., voicing lag. Most of the medial voiced stops were in environments of preceding voicing and showed no voicing break; these were tabulated as having unbroken pulsing, as in B in Figure 2.

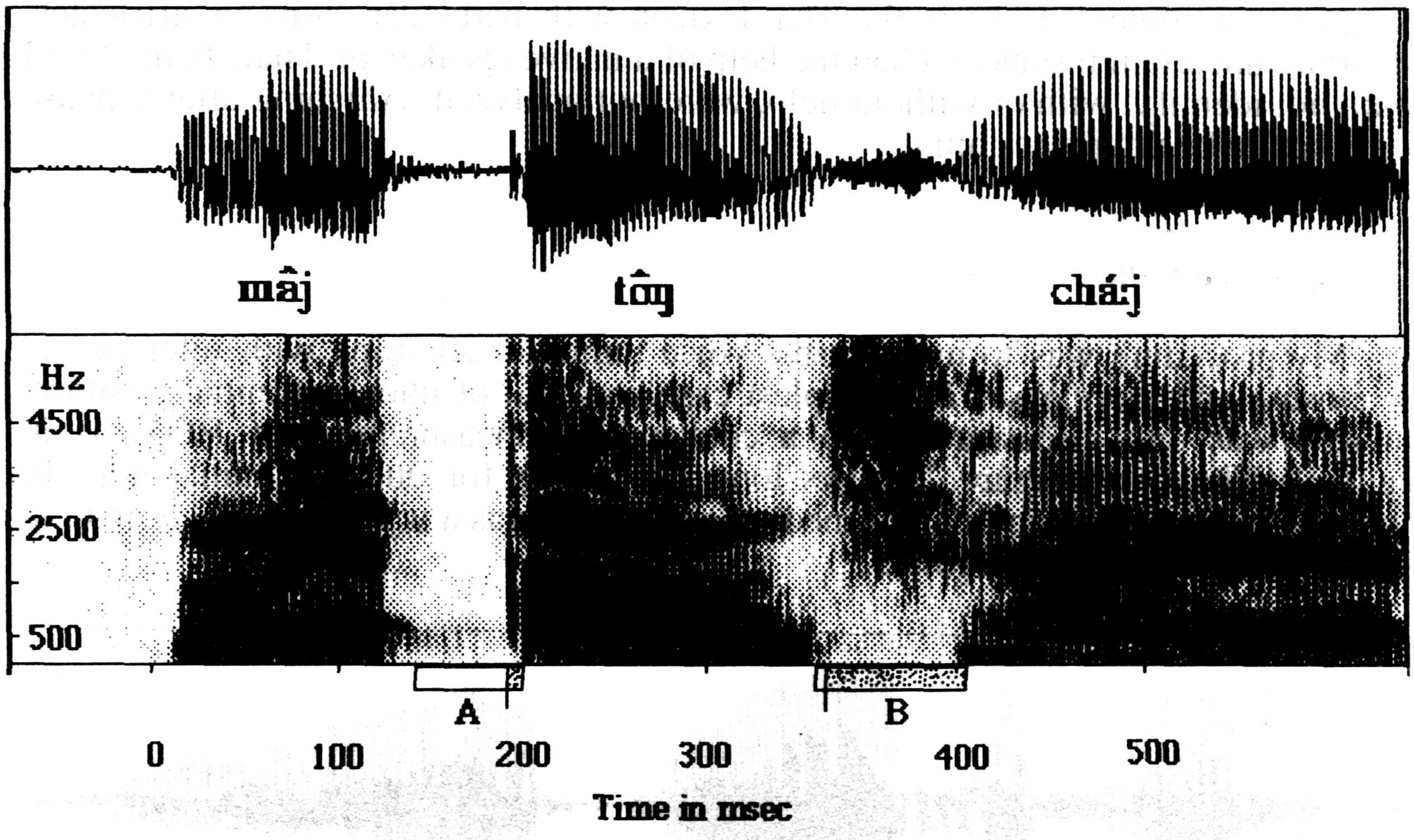

Figure 3. Waveform and spectrogram of an excerpt from the speech of PM. A = total voicing break for the intervocalic /t/, but only the short stippled portion to the right represents voicing lag; B = total voicing break for /ch/, with most of it stippled to stand for frication with open glottis.

2.3 *Affricates*

Conventionally, probably because it seems to yield phonological symmetry, the affricates /c ch/ are aligned respectively with the inaspirates and aspirates of the stop system. Actually, the frication of the "aspirated" affricate (Fig. 3), as well as the very brief frication of the unaspirated affricate (Fig.2), is locally generated in the anterior part of the mouth where the tongue-blade is constricted close to the alveolo-palatal surface. In other words, these plosives are formed by closing the vocal tract as for a stop and releasing the closure into a shorter or longer fricative. Since it is true that the glottis remains open during the frication, it will be convenient here to use the same criteria for the affricates as for the stops.

In Figure 2, then, C shows the total voicing break for /c/, with its frication-duration or short voicing lag indicated by the stippling to the right. An instance of /ch/ is shown in B of Figure 3 with most of its voicing break taken up with frication or long voicing lag. Note, by the way, that although voicing begins in the lower formants at the end of the rectangle marked B, in the upper region of the spectrum, from about 300 Hz upward, there is still much turbulence mixed with the voice pulsing, suggesting that the vocal folds have not yet fully approximated for normal voicing. In such instances, the perceptually relevant exact moment of voice onset time might be a matter of the relative intensities of the components (Repp 1979).

2.4 *Word-final stops*

A Thai syllable ending in a stop is most likely to be a word or morpheme. In this position there is no contrast of voicing or aspiration. Such final stops in citation forms or fairly slow, careful speech hardly ever exhibit glottal pulsing during the closure (Abramson 1972). Perhaps, however, in casual running speech there would be more variation. I noted all final stops as to their occurrence in utterance-final position, before a voiced segment, or before a voiceless segment, and as to whether glottal pulsing was present in the closures.

3. Results

3.1 Stop consonants

At the present stage of this research, I have exhaustively analyzed the productions of just one of the speakers,ST, but only half or fewer of the productions of the other three speakers. Doing preliminary statistics on the voicing breaks of ST, I found that he had a highly significant difference between the total breaks of the aspirates and those of the inaspirates, while his differences in the breaks before consonantal release were not significant. Clearly, then, the responsibility for the significant difference in the total breaks lay in the voicing lags. Inspection of the data for the other three speakers gave the same outcome; therefore, all data here will be only for the post-release voice onset times (VOT) for the voiceless stops. For the voiced stops, VOT will be given in negative values of voicing lead.

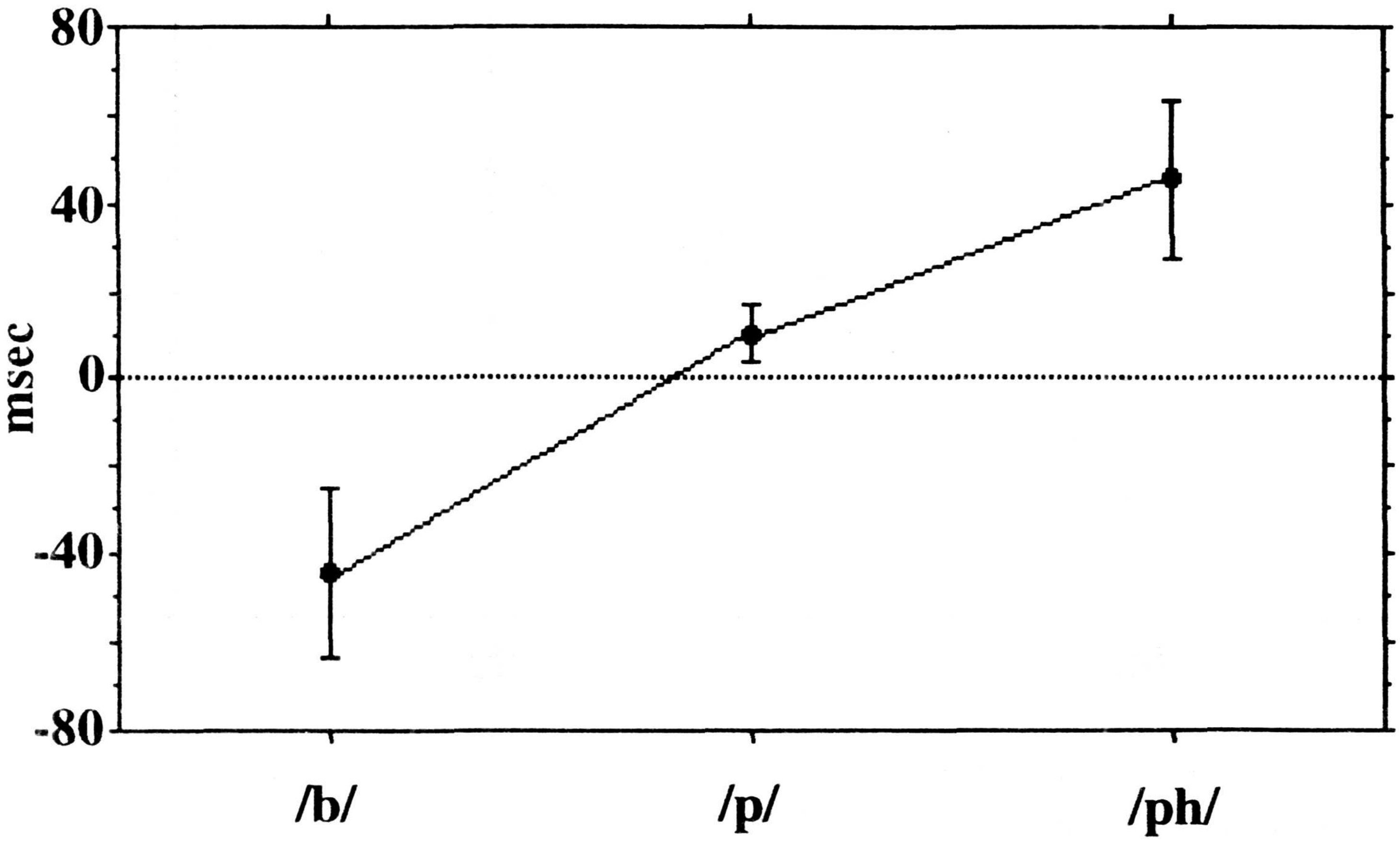

Figure 4. Labial VOT: Means and standard deviations for the pooled data of all four speakers. Number of items: /b/ = 4, /p/ = 38, /ph/ = 36.

The pooled VOT data for the four speake s are given in the graphs of Figures 4, 5, and 6 by place of articulation. The very small numbers of voiced stops displayed must be understood in light of the fact that most occurrences of /b d/ exhibited unbroken pulsing through their closures. For /b/ there were an additional 20 unbroken items (83.3%) and for /d/ and additional 32 items (82.1%^).

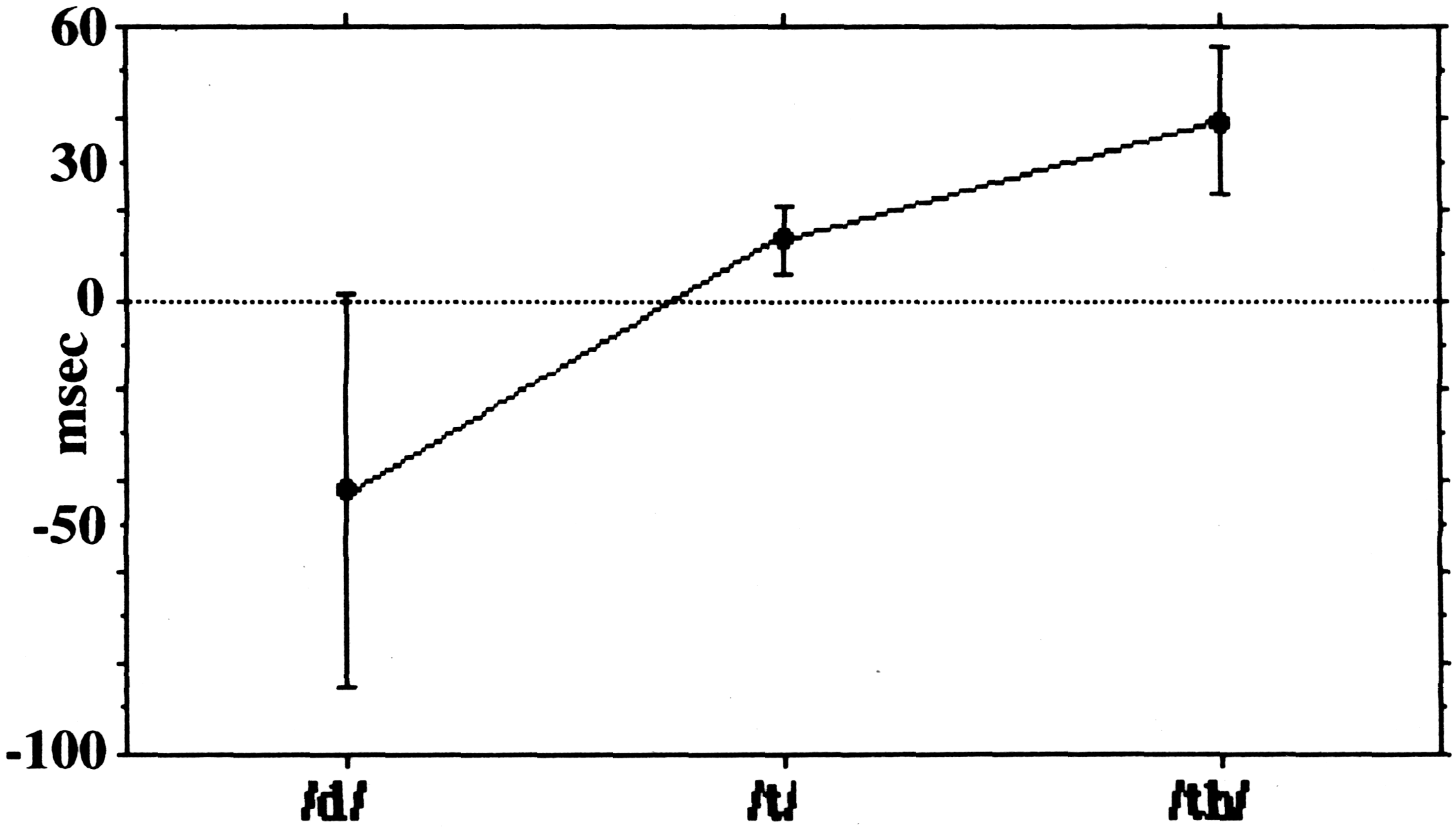

Figure 5. Dental VOT: Means and standard deviations for the pooled data of all four speakers. Number of items: /d/ = 7, /t/ = 41, /th/ = 41.

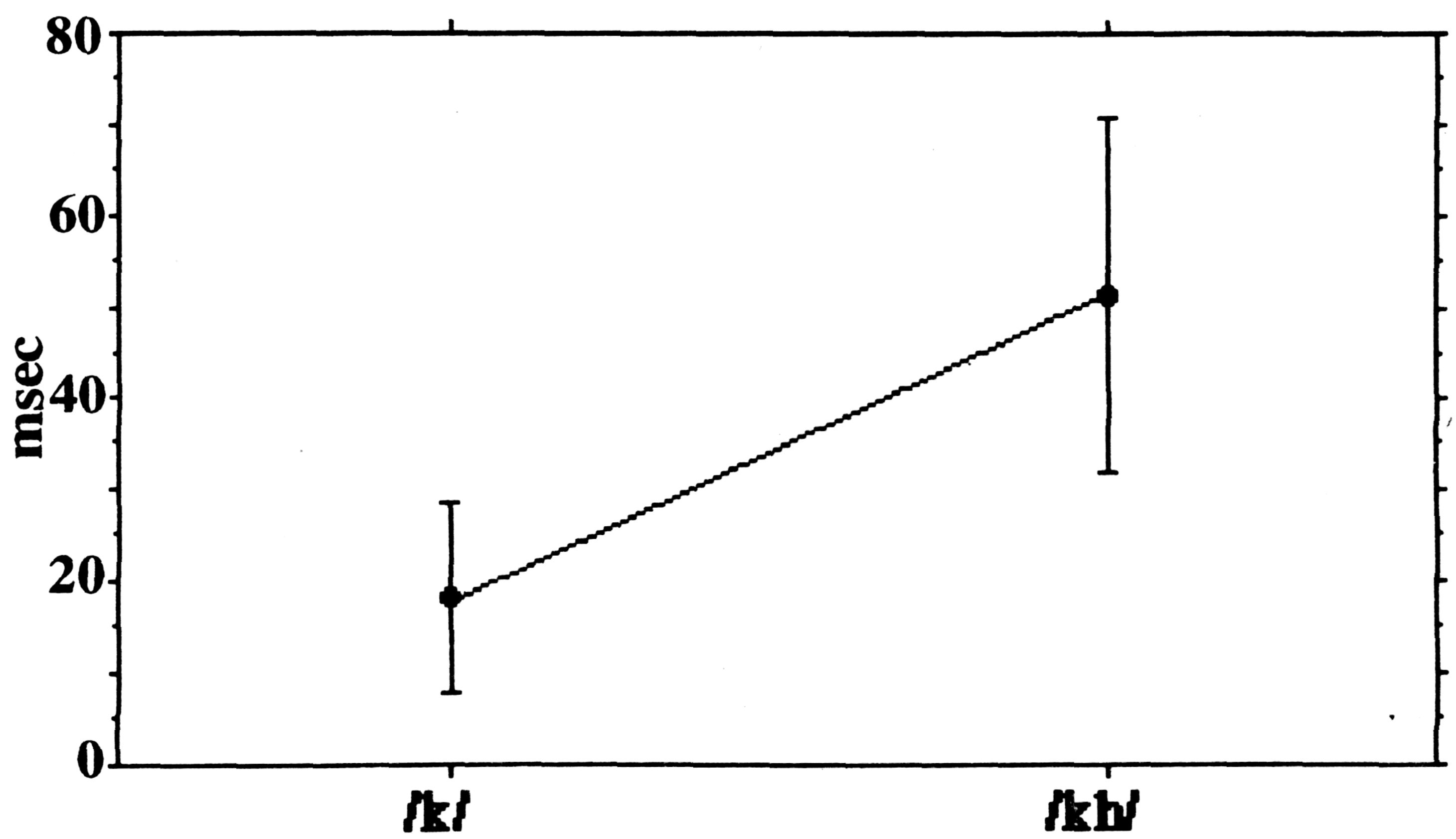

Figure 6. Velar VOT: Means and standard deviations for the pooled data of all four speakers. Number of items: /k/ = 43, /kh/ = 87.

In Figures 4, 5, and 6, the consonants are rather well separated by VOT. That is, merely separating the data according to place of articulation ignores individual-speaker differences and variation in such contextual factors as tempo, tone, sentence intonation, and neighboring segments; nevertheless, the VOT dimension does a good job.

The measurable voiced stops in Figures 4 and 5, with their clearly negative VOT values shown as means and standard deviations below the zero-line, it seems to me, are so clearly in a special category, especially since so many more of them have unbroken pulsing before the release, that there is no need to oppose them statistically to the voiceless categories. Consequently, I have limited my tests for significance to the contrast between the inaspirates and aspirates in the following tables. In addition, the paucity of data at this point for three of the speakers makes it undesirable to treat each place of articulation separately, except for the affricates. ST, on the other hand, has plenty of data, so his stops will appear by place of articulation in a separate table. The statistical analysis was done by means of unpaired two-tailed *t*-tests for significant differences between the means. In the few instances of matched numbers in the pairs, a paired *t*-test was used. This is indicated in Table 3.

Table 1 gives the data by aspiration category for each speaker in the group of three. The levels of significance are high for all of them.

Table 1. *Positive voicing breaks (VOT) in the unaspirated and aspirated voiceless groups for three of the speakers: Means, standard deviations, and significance levels from unpaired* t-*tests.*

Speaker:	PM	TL	SA
/p t k/			
M	15.1	13.0	15.4
SD	7.5	6.4	9.7
n	20	16	17
/ph th kh/			
M	42.1	54.6	50.9
SD	12.6	26.8	14.2
n	29	23	24
df	47	37	39
t	-8.6	-6.1	-8.9
p	.0001	.0001	.0001

Table 2 is arranged a little differently. The pairs of opposed consonants from the conversational material of ST are displayed by place of articulation from left to right. Here too the differences are highly significant.

Table 2. *Positive voicing breaks (VOT) in the voiceless stop consonants of Speaker ST.*

	/p/	/ph/	/t/	/th/	/k/	/kh/
M	9	44.9	14.1	37.3	18.3	49.6
SD	7.1	17.7	7.0	17.5	11.1	18.9
n	23	16	26	22	24	50
df		37		46		72
t		-8.8		-6.2		-7.5
p		.0001		.0001		.0001

3.2 Affricates

The VOT data for all four speakers are given in Figure 7. The numerical data for each speaker are given in Table 3. The levels of significance are quite high although less so than for the stops.

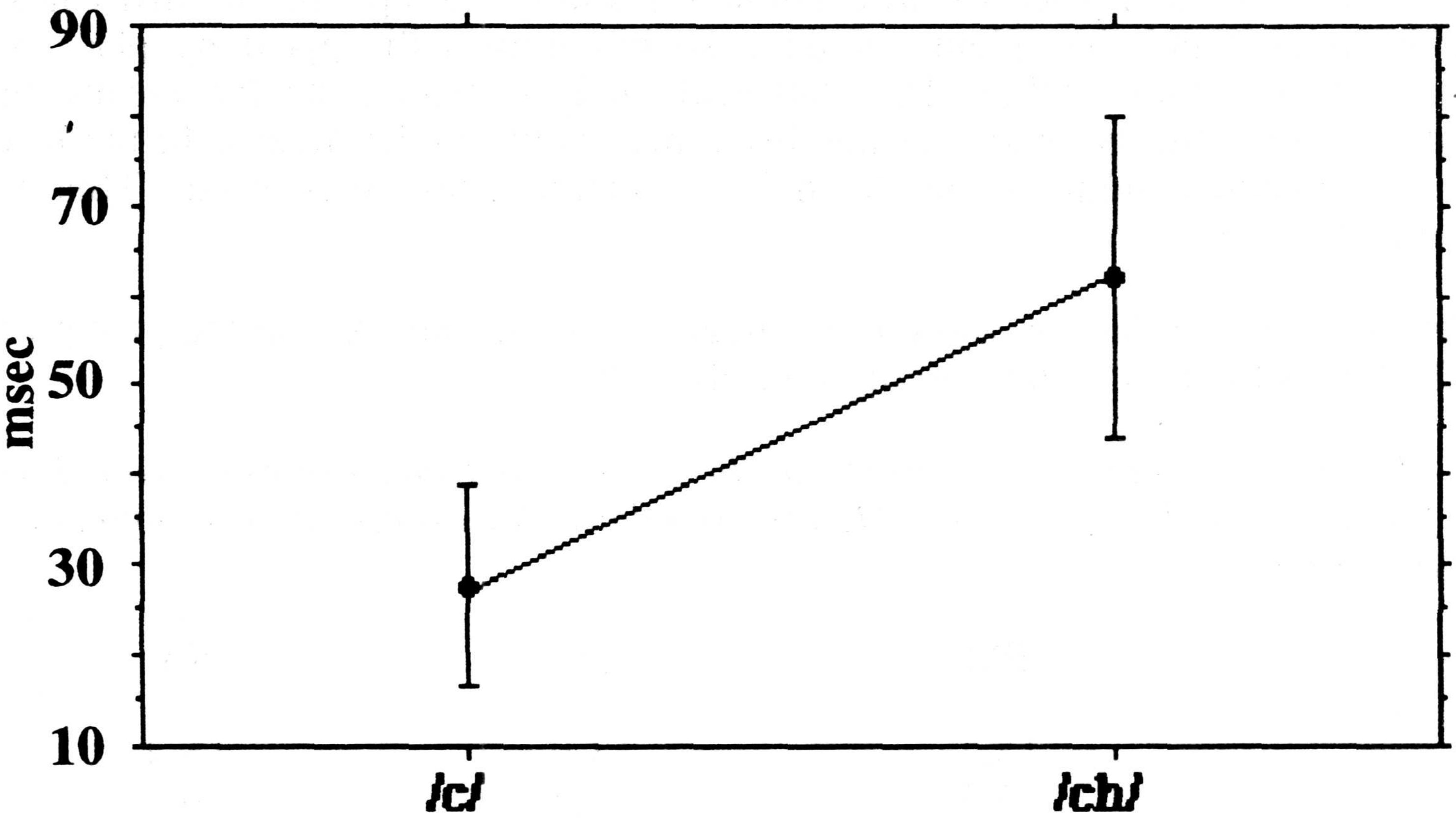

Figure 7. VOT in the affricates: Means and standard deviations for the pooled data of all four speakers. Number of items: /c/ = 56, /ch/ = 35.

Table 3. *Positive voicing breaks (VOT) in the affricates of all four speakers.*

Speaker:	PM	TL	SA	ST
/c/				
M	26.6	21.2	34.4	24.7
SD	9.7	9.7	9.5	11.2
n	5	5	18	28
/ch/				
M	71.8	69.1	55.3	61.7
SD	13.8	21.1	21.5	12.8
n	5	5	12	13
df	9	9	28	39
t	-5.8, paired	-4.7, paired	-3.6, unpaired	-9.4, unpaired
p	.0002	.0011	.0011	.0001

Given the lack of a voiced velar and a voiced palatal in the language, one might wonder whether the pressure of the phonological system is enough to prevent the inaspirates of these two categories from being assimilated to contextual voicing. After all, if it happens, there can be no communicative confusion. In fact, there is only a very slight tendency in this direction, even in my sample of casual running speech. Out of 46 tokens of initial /k/ 5 (10.9%) show unbroken pulsing. As for initial /c/, out of 58 tokens 8 (14%) have unbroken pulsing.

3.3 Final stops

The findings for the word-final oral stops are given in Table 4. The closures are predominantly silent. Only a small number show voiced closures, and most of them are before following voicing.

Table 4. The occurrence of glottal pulsing in the closures of word-final stops for the four speakers.

Final Oral Stop	Number	Number Voiced	% Voiced
/p/	23	3	13.0
/t/	20	3	15.0
/k/	22	1	4.5

4. Discussion and Conclusion

The findings of the present study evidently support the conclusion that even in spontaneous running speech characterized by much variability, as well as by the supplementing of phonetic properties with much top-down information, the dimension of voice timing, already found to be powerful in citation forms and short utterances, remains a very good differentiator of homorganic plosive categories. Although the pre-release voicing breaks in non-initial position do not serve to distinguish the inaspirates from the aspirates, they may still play a role in separating these two categories from the voiced stops.

The failure of most final stops, in spite of the lack of contrast, to be assimilated to neighboring voicing states may suggest that at least in final position the devoicing gesture may be one of glottal closure or constriction (Harris 1972:13-14). In initial position, however, the voiceless inaspirates are too likely to have instrumentally detectable turbulence during their very short voicing lag for this mechanism to be plausible.

Also noteworthy is the general preservation of the voicelessness of initial /k/ and /c/ in spite of the lack of a voiced counterpart. This kind of finding is pleasing to the linguist's belief in the pressure of the overall system.

Acknowledgment

This research was supported by NIH Grant HD-01994 to the Haskins Laboratories. I am grateful for their enthusiastic help to friends and colleagues in the Departments of Linguistics and English, as well as Audio-Visual Services, of the Faculty of Arts, Chulalongkorn University, Bangkok.

References

Abramson, Arthur S. 1972 Word-final stops in Thai. In Jimmy G. Harris and Richard B. Noss, eds., *Tai Phonetics and Phonology* (pp. 1-7). Bangkok: Central Institute of English Language.

Abramson, Arthur S. 1989 Laryngeal control in the plosives of Standard Thai. *Pasaa,* 19:85-93. *Proceedings of the 6th International Congress of Phonetic Sciences,* Abramson, Arthur S. In press The stability of distinctive vowel length in Thai. In Kalaya Tingsabadh and Arthur S. Abramson, eds., *Essays in Tai Linguistics*. Bangkok: Chulalongkorn University Press.

Abramson, Arthur S. and Lisker, Leigh 1970 Discriminability along the voicing continuum: Cross-language tests. *Proceedings of the 6th International Congress of Phonetic Sciences,* pp. 569-573.

Abramson, Arthur S. and Lisker, Leigh 1995 Voice timing for stop classification in conversational English. *Proceedings of the 13th International Congress of Phonetic Sciences,* vol. 4, pp. 128-131.

Harris, Jimmy G. 1972 Phonetic notes on some Siamese consonants. In Jimmy G. Harris and Richard B. Noss, eds. *Tai Phonetics and Phonology* (pp. 8-22). Bangkok: Central Institute of English Language.

Lisker, Leigh and Abramson, Arthur S. 1964 A cross-language study of voicing in initial stops: Acoustical measurements. *Word,* 20:384-422.

Lisker, Leigh and Abramson, Arthur S. 1970 The voicing dimension: Some experiments in comparative phonetics. *Proceedings of the 6th International Congress of Phonetic Sciences, pp. 563-567.*

Repp, Bruno H. 1979 Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech,* 27: 173-189.

Tingsabadh, Kalaya and Abramson, Arthur S. 1993 Thai. *Journal of the International Phonetic Association,* 23: 24-28.