A voiced onset time analysis of word-initial stops in Thai

Jack Gandour
Purdue University

Voice onset time (VOT) is defined as the temporal relation between the onset of glottal pulsing and the release of the initial stop consonant (Lisker & Abramson, 1964, 1967). Voicing detected before the release is called voicing lead, while voicing starting after the release is called voicing lag. Lisker & Abramson (1964) found that VOT is a good index to laryngeal timing for stop-consonant distinctions in word-initial position across a variety of languages. Although VOT generally refers to the timing relation between the burst release and onset of glottal pulsing, it, in fact, represents a constellation of acoustic attributes including the intensity and duration of the burst release, cutback and starting frequency of the first formant, and the presence or absence of frication noise upon consonantal release (Lisker & Abramson, 1964; Abramson, 1977).

In word-initial stops, Thai exhibits a three-way contrast in voicing at the bilabial (voiced unaspirated, /b/; voiceless unaspirated, /p/; voiceless aspirated, /p^h/) and alveolar (voiced unaspirated, /d/; voiceless unaspirated, /t/; voiceless aspirated, /t^h/) places of articulation, and a two-way contrast (voiceless unaspirated, /k/; voiceless aspirated, /k^h/) at the velar place of articulation. Only three previous studies have investigated the perceptual relevance of VOT in distinguishing these stop categories (Abramson & Lisker, 1970; Lisker & Abramson, 1970; Gandour, 1982). Up to the present, the only VOT production data available on Thai word-initial stops can be found in Lisker & Abramson (1964). Their study included three Thai speakers only. It has been almost 20 years since the publication of these VOT data on Thai word-initial stops. In view of the linguistic and psychological importance of the VOT dimension, it is imperative that we expand our data base on VOT production in Thai. Accordingly, the aim of this paper is to present additional measurements of VOT in Thai word-initial stops.

Method

Five graduate-level university students participated in the experiment. All were native speakers of Thai. None had any known speech or hearing impairment. The test stimuli consisted of eight monosyllabic words: two minimal triplets contrasting bilabial and alveolar stops, respectively, and one minimal pair contrasting velar stops (Table 1). All words contained an initial stop consonant followed by a low back unrounded vowel and a final nasal consonant. Tone (mid) was held constant across all eight words.

Table 1

Stimulus Set

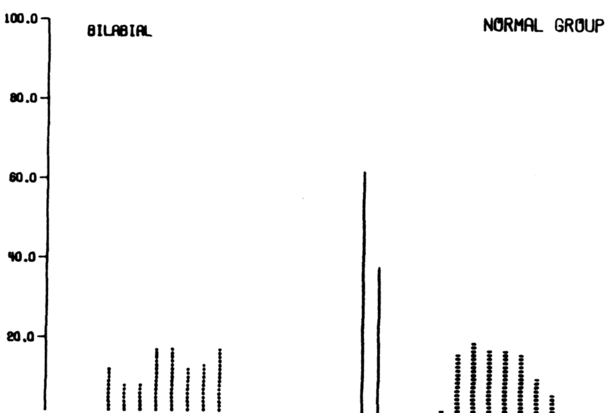
Bilabial	Alveolar	Velar
/baan/ 'to bloom'	/dam/ 'black'	
/paan/ 'birthmark'	/tam/ 'to pound'	/kan/ 'to prevent'
/p^{h}aan/ 'ignorant'	/t^{h}am/ 'to do'	/k^{h}an/ 'to itch'

All words were written in large Thai letters on 3 x 5 cards for presentation to the subject. A different random order of presentation was used for each subject. Subjects were instructed to produce the words in the carrier frame /nîi ____/ 'This (is) _____.' at a comfortable speaking rate. Subjects' responses were tape-recorded using a Sony ECM-150 microphone and a Nagra IV-S tape recorder. Each subject was tested individually in a sound-proofed booth in a single session. A minimum of 80 utterances (8 words x 10 repetitions) were recorded for each subject.

A total of 420 utterances were analyzed by means of a Voiceprint Model 700 sound spectrograph using a wide-band pattern and amplitude display. Vertical lines were drawn on the spectrograms at the release of the stop consonant. Stop release was identified by the onset of a burst of frication noise following the closure interval and concomitant abrupt rise in amplitude. In the case of voicing lag, the "...sudden onset of vertical striations in the second and higher formants" (Klatt, 1975, p. 687) was taken to represent the onset of voicing following the stop release; in the case of voicing lead, the sudden onset of low energy vertical striations in the absence of acoustic energy in the formant frequency range was taken to represent the onset of voicing preceding the stop release (Lisker & Abramson, 1964, p. 389). Adopting the convention of assigning a timing value of zero to the moment of stop release, measurements of VOT after the stop release (voicing lag) are stated as positive numbers, while measurements of VOT before the stop release (voicing lead) are stated as negative numbers (Lisker & Abramson, 164, p. 389). Each measurement was rounded to the nearest five msec. VOT measurements obtained from the spectrograms were grouped into 10-msec intervals and frequency distributions were plotted for both the group and individual subjects.

Results and Discussion

Figure 1 presents the distribution of VOT productions at all three places of articulation pooled across the five normal adults. No overlap is evident in the distribution of VOT productions for any of the stop consonants. They are all distributed into relatively discrete areas of VOT. Inspection of the graph also shows that the range of distribution is narrower for voiceless unaspirated stops than for either voiced unaspirated or voiceless aspirated stops. Finally, it is noted that the intercategory interval between the voiced and voiceless unaspirated stops is wider than that between the voiceless unaspirated and voiceless aspirated stops.



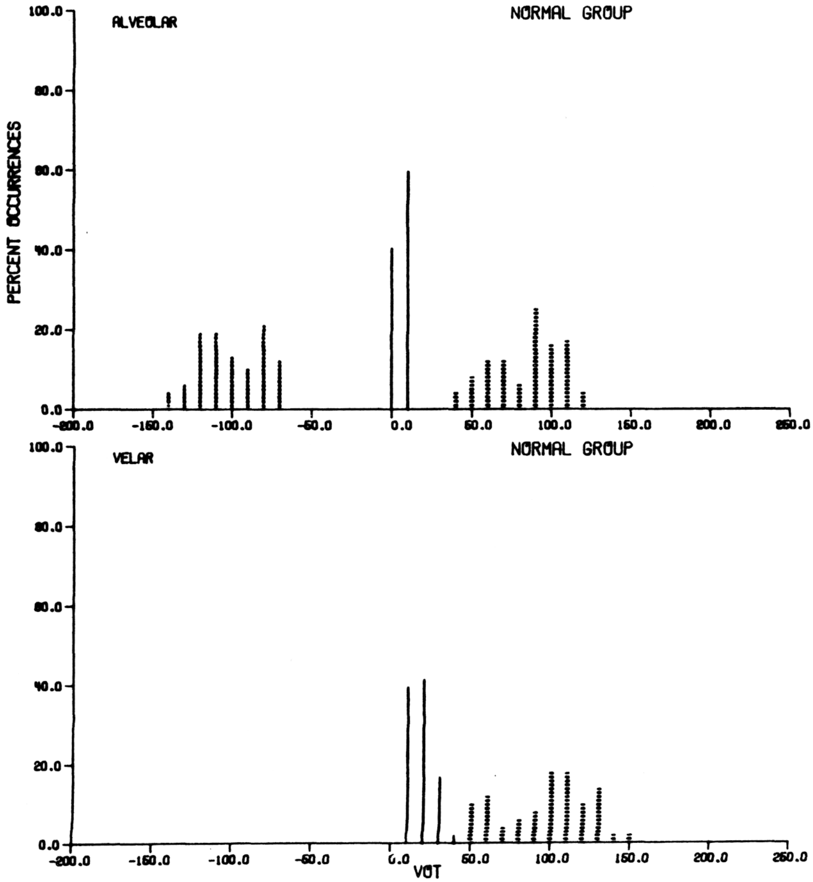


Figure 1. Distribution of VOT productions for bilabial, alveolar, and velar stop consonants in word-initial position for a group of five normal Thai adults. The abscissa of each graph represents VOT values in 10-msec. intervals and the ordinate the percentage of utterances falling within the specified VOT interval. Voiced stops, [b] and [d], are indicated by the dotted lines; voiceless unaspirated stops, [p], [t], and [k], by the solid lines; and voiceless aspirated stops, [p^h], [t^h], and [k^h], by the striped lines.

Figures 2 through 4 present the VOT data for each of the five individual subjects at the bilabial, alveolar, and velar places of articulation, respectively. Despite some interspeaker variability in the size of the ranges of VOT productions for a given stop consonant, all five speakers exhibit clear-cut nonoverlapping VOT distribution patterns.

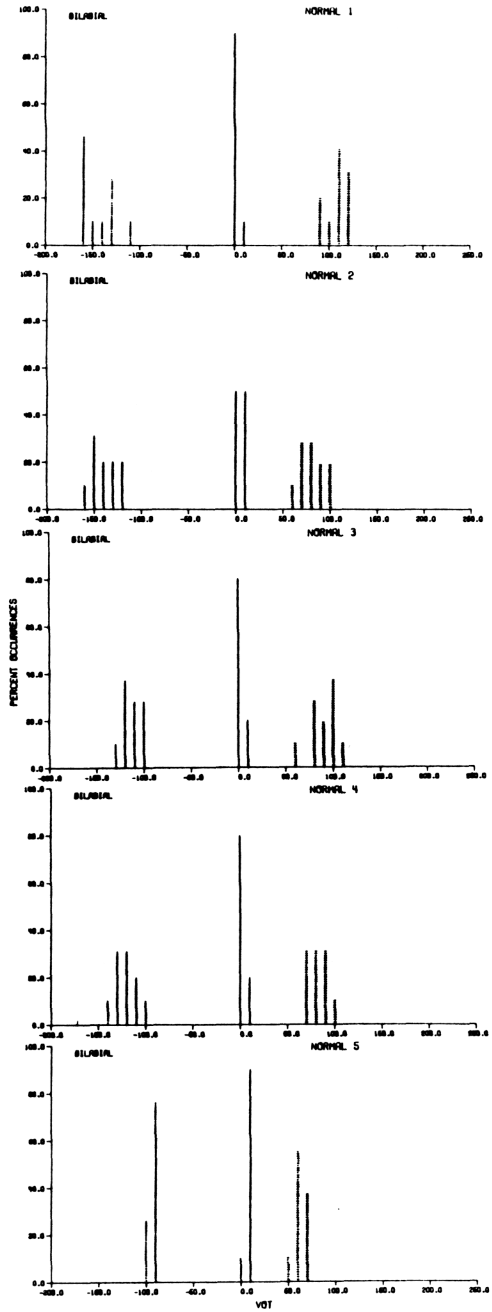


Figure 2. Distribution of VOT productions for bilabial stop consonants, [b], [p], and [pʰ], in word-initial position for each of the five speakers. See also caption for Figure 1.

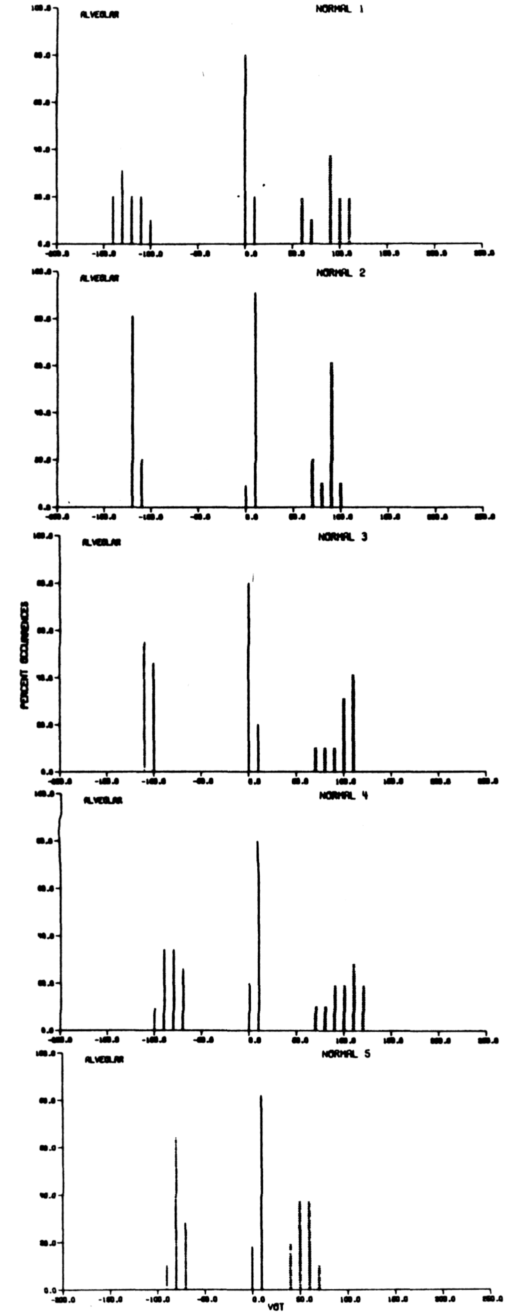


Figure 3. Distribution of VOT productions for alveolar stop consonants, [d], [t], and [t^h], in word-initial position for each of the five speakers. See also caption for Figure 1.

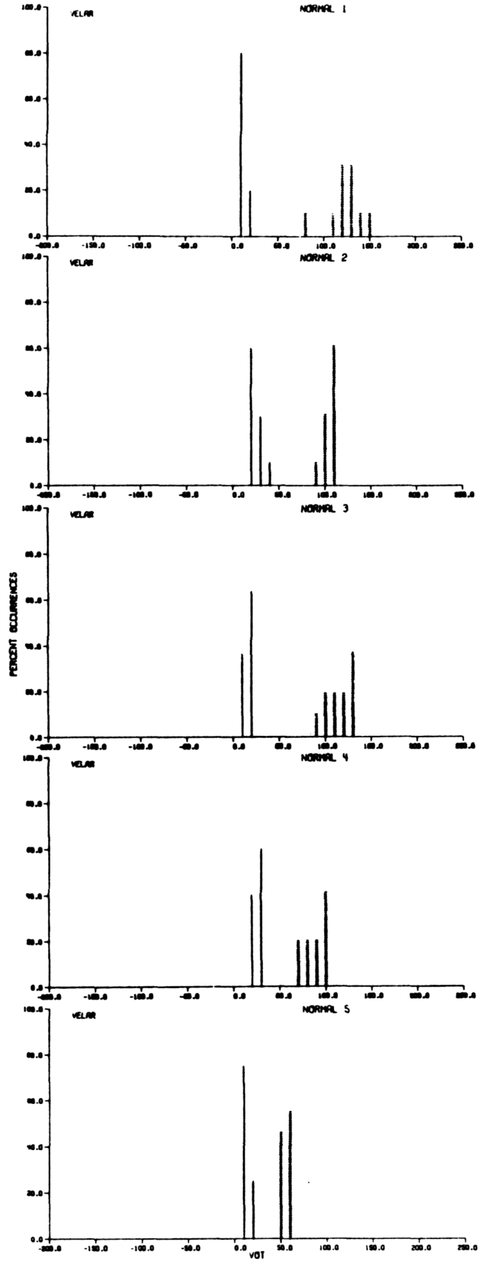


Figure 4. Distribution of VOT productions for velar stop consonants, [k] and [k^h], in word-initial position for each of the five speakers. See also the caption for Figure 1.

Descriptive statistics for the distribution of VOT productions for each speaker are given in Table 2. As can be seen in this table, individual speakers differed only slightly in their VOT productions of voiceless unaspirated stops, /p/, /t/, and /k/. However, there was some interspeaker variability in the production of voiced and voiceless aspirated stops. For the voiced stops, /b/ and /d/, speakers who exhibited wider ranges of VOT distribution tended to exhibit lower average VOT values; and conversely, speakers who exhibited narrower ranges of VOT distribution tended to exhibit higher average VOT values. Speaker 1, for example, shows the widest range and lowest average VOT for both /b/ and /d/. Speaker 5, on the other hand, shows the narrowest range and highest average VOT for /b/. Thus, given any compression of the VOT continuum, a speaker must necessarily compress the ranges of VOT distribution if he is to maintain relatively discrete areas of VOT for the different stop consonants. For speaker 5, this adjustment in the ranges of VOT distribution is borne out in the standard deviation values. He tends to show the least amount of variability in VOT productions among the five speakers. This same inverse relationship between the range and average VOT values is also evident in the production of the voiceless aspirated stops, /p^h/, /t^h/, and /k^h/. The shorter VOT continuum for speaker 5 may be attributed to a faster speaking rate.

Table 2

Descriptive Statistics of VOT for Word-Initial Stops

/b/

	Minimum	Maximum	Range	Mean	Standard Deviation
Normal 1	-160	-105	55	-142.3	18.1
Normal 2	-155	-115	40	-137.5	14.2
Normal 3	-130	- 95	35	-110.5	9.6
Normal 4	-135	-100	35	-118.5	10.6
Normal 5	- 95	- 85	10	- 90.4	3.3

/p/

	Minimum	Maximum	Range	Mean	Standard Deviation
Normal 1	0	10	10	4.5	2.8
Normal 2	5	10	5	7.5	2.6
Normal 3	0	10	10	3.5	4.1
Normal 4	0	10	10	4.0	3.9
Normal 5	5	15	10	10.5	2.8

/p^h/

	Minimum	Maximum	Range	Mean	Standard Deviation
Normal 1	90	120	30	108.0	11.4
Normal 2	65	100	35	82.7	11.9
Normal 3	65	110	45	90.5	13.1
Normal 4	70	100	30	82.0	10.3
Normal 5	50	70	20	64.1	6.3

/d/

	Minimum	Maximum	Range	Mean	Standard Deviation
Normal 1	-140	-100	40	-121.5	13.1
Normal 2	-120	-110	10	-115.0	3.3
Normal 3	-110	- 95	15	-103.2	4.6
Normal 4	-100	- 70	30	- 81.3	9.3
Normal 5	- 90	- 70	20	- 76.8	6.0

/t/

	Minimum	Maximum	Range	Mean	Standard Deviation
Normal 1	0	10	10	4.0	3.9
Normal 2	0	10	10	9.1	3.0
Normal 3	0	10	10	4.5	3.7
Normal 4	0	10	10	8.5	3.4
Normal 5	0	15	15	9.1	3.8

/t^h/

	Minimum	Maximum	Range	Mean	Standard Deviation
Normal 1	60	110	50	90.0	17.9
Normal 2	70	100	30	87.5	10.1
Normal 3	70	115	45	99.5	15.2
Normal 4	75	120	45	101.8	15.7
Normal 5	45	70	25	56.4	7.8

/k/

	Minimum	Maximum	Range	Mean	Standard Deviation
Normal 1	10	20	10	14.0	3.9
Normal 2	25	40	15	28.0	4.8
Normal 3	15	25	10	19.1	3.8
Normal 4	20	35	15	27.0	5.4
Normal 5	10	20	10	15.4	3.3

/k^h/

	Minimum	Maximum	Range	Mean	Standard Deviation
Normal 1	85	150	65	125.5	17.4
Normal 2	95	110	15	106.5	5.3
Normal 3	95	130	35	116.8	13.3
Normal 4	70	105	35	89.5	14.0
Normal 5	50	60	10	55.9	4.9

Note. Values are expressed in msec.

Normal 1 = male, 28 years old; Normal 2 = female, 33 years old; Normal 3 = female, 30 years old; Normal 4 = female, 29 years old; Normal 5 = male, 31 years old.

Pooling across subjects, the data in Table 2 indicate that the VOT for voiceless unaspirated and aspirated stops is approximately 5 msec less than average for bilabials and approximately 10 msec greater than average for velars. For individual subjects, the VOT generally increases from /p/ to /t/ to /k/ and from /p^h/ to /t^h/ to /k^h/. Lisker & Abramson 1964) found a similar relationship in their Thai data as well as data from other languages.

Table 3 presents a comparison of our findings with those obtained in Lisker & Abramson's (1964) study. In general, our findings are in agreement with Lisker & Abramson (1964). However, our average VOT values for the voiced stops, /b/ and /d/, and the voiceless aspirated stops, /p^h/ and /t^h/, are about 20 milliseconds lower and higher, respectively, than those reported in Lisker & Abramson (1964). Also, for the bilabial and alveolar places of articulation, the intervals between /b/ and /p/, /p/ and /p^h/, /d/ and /t/, and /t/ and /t^h/ are 30-45 milliseconds longer in the current study than in Lisker & Abramson (1964). These differences in average VOT and length of intervals between stop categories indicate both an expansion of the VOT continuum and an even clearer separation of the stop categories in the current study. Such discrepancies in the VOT distribution pattern may perhaps be attributed to the nature of the corpus, the carrier phrase, or speaking rate of individual subjects.

Table 3

Comparison to Previous Research on VOT in Thai

	/b/		/p/		/p^h/	
	<u>Mean</u>	<u>Range</u>	<u>Mean</u>	<u>Range</u>	<u>Mean</u>	<u>Range</u>
G	-120	-160:-85	6	0:15	85	50:120
L&A	- 97	-165:-40	6	0:20	64	25:100
	/d/		**/t/**		**/t^h/**	
	<u>Mean</u>	<u>Range</u>	<u>Mean</u>	<u>Range</u>	<u>Mean</u>	<u>Range</u>
G	-100	-140:-70	7	0:15	87	45:120
L&A	- 78	-165:-40	9	0:25	65	25:125
			/k/		**/k^h/**	
			<u>Mean</u>	<u>Range</u>	<u>Mean</u>	<u>Range</u>
G			21	10:40	99	50:150
L&A			25	0:40	100	50:155

<u>Note</u>. Values are expressed in terms of msec.
G = Gandour, current study; 5 speakers

With the completion of the current study of VOT in Thai word-initial stops in a fixed segmental/suprasegmental environment, we are now in a position to examine various context effects on VOT (cf. Lisker & Abramson, 1967; Klatt, 1975; Weismer, 1979). Including the voiceless unaspirated and aspirated alveopalatal affricates, it is proposed that large-scale systematic studies be undertaken to determine the extent to which VOT is conditioned by speaking rate, stressed vs. unstressed syllables, tone, various consonant cluster environments, identity of the following vowel, voicing characteristic of a final consonant, word-initial vs. word-medial position, and sentential vs. citation material. Such detailed acoustic investigations would be of great interest to Tai historical-comparativists who are grappling with problems of voicing in Tai plosives from a diachronic perspective. These investigations, of course, would also be of interest to linguists, phoneticians, speech scientists, communications engineers, speech pathologists, and child language specialists. More generally, it is hoped that the present study represents a modest contribution toward the goal of detailed spectral and temporal information about Thai consonants and vowels.

Footnote

Research supported by NIH Grant No. NS-18116, NSF Grant No. BNS-7905854, and the National Research Council of Thailand. The data were collected during May-December 1980 while the author was affiliated with the Division of communication Disorders, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand. The hospitality of Poonpit Amatyakul, Director, and his staff of the Speech and Hearing Clinic at Ramathibodi Hospital is gratefully acknowledged. Special thanks to Soranee Holasuit Petty for making the spectrograms and producing the figures. Reprint requests should be sent to Jack Gandour, Department of Audiology and Speech Sciences, Heavilon Hall, Purdue University, West Lafayette, Indiana, U.S.A. 47907.

References

Abramson, A.S. Laryngeal timing in consonant distinctions. Phonetica, 1977, 34, 295-303.

Abramson, A.S. & Lisker, L. Discriminability along the voicing continuum: cross-language tests. In Proceedings of the Sixth International Congress of Phonetic Sciences. Prague: Academia Publishing House of the Czechoslovak Academy of Sciences, 1970.

Gandour, J. Voice onset time in aphasia: Thai. I. perception. Brain and Language, 1982, 17, 24-33.

Klatt, D.H. Voice-onset time, frication, and aspiration in word-initial consonant clusters. Journal of Speech and Hearing Research, 1975, 18, 686-706.

Lisker, L. & Abramson, A.S. A cross-language study of voicing in initial stops: acoustical measurements. Word, 1964, 20, 384-422.

Lisker, L. & Abramson, A.S. Some effects of context on voice onset time in English stops. Language and Speech, 1967, 10, 1-28.

Lisker, L. & Abramson, A.S. The voicing dimension: some experiments in comparative phonetics. In Proceedings of the Sixth International Congress of Phonetic Sciences. Prague: Academia Publishing House of the Czechoslovak Academy of Sciences, 1970.

Weismer, G. Sensitivity of voice-onset time (VOT) measures to certain segmental features in speech production. Journal of Phonetics, 1979, 7, 17-204.