A Seven-Tone Dialect in Southern Thai with Super-High: Pakphanang Tonal Acoustics and Physiological Inferences

Phil Rose

Australian National University

INTRODUCTION

Some varieties of Southern Thai are notable for the large number of surface tonal contrasts they exhibit. As one progresses south in this 500 kilometer long dialect continuum, the number of tones increases from five to six to seven (Diller, 1979b, p. 119). The latter varieties are thus members of the small number—15% — of the world's tone languages with more than six tones (Hombert, 1977, p. 21). This paper describes part of the tonal acoustics in one of these seven-tone varieties: Pakphanang (PPhN). This variety is tonally interesting not only for its large number of surface contrasts, but also for its super-high tone. It is an honour to offer this contribution to Dr. Vichin Panupong, and to be part, however modestly, of the tradition of Southern Thai phonetic studies she founded.

TONOLOGICAL DESCRIPTION

As in most varieties of Tai, the number of surface tonal contrasts on citation monosyllables in PPhN depends on the structure of the Rhyme, specifically the absence or presence of a syllable-final stop [p t k] in the Coda. Like the neighbouring dialect of Nakhon Sithammarat (Haas, 1958), PPhN contrasts seven tones on non-stopped syllables with phonologically at least two sonorants in the Rhyme. Examples of minimal and subminimal contrasts, from the field notes of A. Diller, are given in Table 1 in a conventional transcription for Tai varieties, with the long vowels [aː] and [uː] represented as a sequence.

As in Nakhon dialect (Diller, 1987, p. 148), PPhN tones 1, 2, 3, and 4 are in complementary distribution with respect to syllable-initial obstruents; tones 1 and 2 co-occur with voiceless aspirated stops and fricatives, and tones 3 and 4 with voiceless unaspirated stops. But as the examples in Table 1 show, there are minimal and sub-minimal pairs on syllables with the initial palatal glide /y/ to demonstrate a tonemic contrast between tones 1 and 3, and 2 and 4.

The auditory characteristics of the seven unstopped Pakphanang citation tones are as follows. These are based on my transcriptions of recorded utterances of one male and one female speaker. *Tone 1* is the super-high tone. It has convex pitch very high in the speaker's pitch range, with the falling portion a little more salient than the rising. The voice quality is often falsetto, which is possibly a consequence of the very high pitch target. This tone is auditorily very salient in running speech, and its extreme pitch characteristics are used by speakers of other, more distant Southern dialects to stereotype PPhN (A. Diller, p.c.). *Tone 2* has level pitch in the upper mid pitch range. *Tone 3* has convex pitch in the lower half of the pitch range, with the rising portion longer and more salient. In a second recording of the same speaker several years later, this tone had only a low rising pitch, which suggests that the falling part is optional. *Tone 4* has level pitch in mid pitch range. *Tone 5* has a relatively short level component in mid pitch range followed by a fall. *Tone 6* has a fairly low dipping pitch rising just into the mid pitch range. *Tone 7* has falling pitch in the low pitch range, and sounds longer than the other tones.

Table 1. *Examples of Tonal Contrast in Pakphanang Unstopped Syllables. (Bold face indicates forms acoustically analysed in this paper; the second row shows historical tonal category and consonant class).*

Tone	1	2	3	4	5	6	7
	*A H *B H	*C H	*A M *B M	*C M	*A L	*B L	*C L
	khaa *leg*	**khaa** *kill*	**kaa** *crow*	**kaa** *mark*	**khaa** *thatch grass*	**khaa** *value*	**khaa** *trade*
	khuu *threaten*	**phuu** *male*	**kuu** *I*	**kuu** *borrow money*	**khuu** *ditch*	**phuu** *wasp*	**ruu** *know*
	naa *thick*	naa *face*	paa *jungle*	paa *parent's older sister*	naa *rice field*	naa *worth doing*	naa *mother's younger sibling*
	yaa *fish sp.*	yaa *grass*	yaa *medicine*	yaaŋ *roast*	yaa *don't*	yaa *father's mother*	

The above description suggests that the PPhN pitch range is neither optimally nor neatly exploited. Thus the convex pitches of tones 1 and 3 lie more or less at extremes of the pitch range, but the level pitches of tones 2 and 4 are concentrated in the mid pitch range, and the two falling pitches of tones 5 and 7 lie in the lower half of the pitch range. A reasonable characterisation of the PPhN tonal system described above would be that it comprises four distinctive pitch contours—convex, level, falling, and concave—the first three of which each have a contrasting higher and lower value. Phonetic effects other than pitch, e.g., falsetto voice with tone 1, longer duration with tone 7, reinforce the pitch contrasts.

Hombert (1977, p. 29, f.n. 9) notes that languages with more than six tones usually recruit an additional distinctive phonetic parameter like phonation type into their tone system. Typologically, then, PPhN appears unusual. However, genetically it is not atypical, since the *distinctive* use of such parameters is not common in Tai. One documented occurrence is in the N.W. Tai dialect of Thai-Phake, which has tonally contrastive creak (see Rose (1990) for acoustic and aerodynamic data on this contrast).

PROCEDURE

In this paper, the acoustic properties of the seven unstopped tones will be described. Apart from fundamental frequency (F0), which is the major acoustical dimension in which tonal contrasts are cued perceptually, radiated amplitude (Ar), and Formant-pattern (F-pattern) were also examined for correlation with tone.

Due to well-known assumed intrinsic effects from concomitant segmental articulation, tones can never be acoustically observed independent of their segmental realisation. Because these effects are differential, it is not possible to infer a representative tonal shape from segmentally invariant data: to look at the F0 shape of a particular tone on a [u] vowel, for example, and assume this is also the shape representative of the tone. A corpus was therefore chosen using two vowels—[a] and [u]—which might be expected to elicit maximum difference in intrinsic effect of vowel quality on F0, Ar, and duration (D). It was further assumed that a better idea of the PPhN tonal F0 could be achieved by calculation of the mean value of these two extremes. This assumption needs to be tested, of course; one cannot be sure that the intrinsic effect of an [a] represents the same degree of divergence from the underlying, presumably invariant tone command as that of [u], or if perhaps the [a] interferes less with the realisation of the tone than the [u]. The corpus thus comprised the seven PPhN tones demonstrated on unstopped syllables with [a:] and [u:] vowels. The words used are given in bold face in Table 1. Note that the corpus reflects the above-mentioned complementary distribution of tone with respect to aspiration on obstruents. The words were elicited using Standard Thai orthography. Diller (1987, p. 15) notes that Thai orthography is capable of indicating Southern segmentals directly; that it represents either the Southern or Central tonal system with about equivalent degrees of fit; and that a text written in Thai can be given a seven-tone Southern Thai reading quite effortlessly. The words were recorded under laboratory conditions by a 46 year-old female native speaker in 1981. The speaker paused after each item. The [a:] forms were read first. Three repeats of the corpus were recorded (the informant did not know how many repeats were going to be requested). The aspiration on the obstruents was noticeably weak (this was also noted with another speaker's utterances); the unaspirated /k/ in tone 4 was noticeably fortis (but not ejected); and all examples of tone four were observed to register a much higher amplitude peak on the recorder's Vu meter. Since the informant was also typically fluent in Standard Thai, she was also recorded speaking corresponding Standard Thai forms in order to investigate how the Standard tones relate to those of the Southern varieties. Some initial results of this comparison were reported in Rose (1985).

The procedure used to extract F0, Ar, and D was essentially that described in Rose (1982, pp. 7–10). The duration of the Rhyme was first determined from wide band spectrograms with their good time-domain resolution. F0 was then sampled at percentage points of this duration from aligned expanded narrow-band spectrograms. Slightly different duration bases were used for tones 3 and 6; these are described below. This method of F0 measurement has an accuracy of +/- 4Hz at the 90% confidence level. Ar was sampled from the oscillographic trace from an F/J intensity meter, using 20 ms integration time and flat response. Alignment with the F0 curve was achieved by an audio wave on the oscillogram. This method is accurate to +/- 0.5 dB at the 90% confidence level. The high sampling frequency (mean value ca. 17 Hz, or once every 6 csec.) was chosen to permit detailed resolution of both F0 and Ar contours. \Measurements of possible F-pattern correlations of the tones were carried out as follows. Tokens were processed digitally, using the API and SGM commands of the ILS signal processing package (API did not reliably resolve two clear poles in the expected frequency range for F1 and F2 in the [u:] tokens, so these were not further analysed). The first three formants of the [a:] tokens were identified and sampled at three points each at approximately 25%, 50%, and 75% of the duration of the Rhyme. This gave nine tokens per formant per tone, over which means and standard deviations were then calculated.

Arithmetical mean and standard deviation values for F0, Ar, and D for the seven tones with [a:] and [u:] vowels, as well as VOT of initial obstruents and F-pattern for [a] vowels, are given in Table 2. The percentage points at which F0 and Ar were sampled are also shown under "SP." So, for example, in tone 2 words with [a:] vowels, the mean F0 at the 50% sampling point was 199 Hz, with a standard deviation of 8 Hz. The 50 % sampling point occurred at (56.8 csec * 0.5=) csec 28.4. For the low convex tone 3, visual inspection of the between-token variation in F0, and the above-mentioned free variation with low rising pitch, suggested that it would be better to treat F0 and Ar as splines, and sample them as functions of duration to F0 peak, and duration from F0 peak to phonation offset. Thus the sampling point at 60% of duration to F0 peak in tone 3 words with [u:] vowels occurs at (31.8 csec.* 0.6 =) csec. 19.1, and the 60% of duration to F0 offset at ((31.8 csec + (17.1 csec * 0.6))=) csec 42.1. In the low dipping tone 6, F0 and Ar were sampled as functions of duration to F0 peak, and also at phonation offset ("F0off").

INTRINSIC EFFECTS

In all except the convex tones 1 and 3 the [u:] allotype has, as expected, a higher F0. In the second half of the convex tones however (Figure 1), the intrinsic relationship is either reversed (tone 1) or cancelled (tone 3). It is easy to see that, because of the expected intrinsic differences in duration between [a:] and [u:] allotypes (the duration of the [a:] allotone is greater than the [u:] for all tones except 5, and some of the differences are quite big (e.g., 7 and 5 csec for tones 1 and 2) the nature of the intrinsic relationship on tones with appreciable F0 movement depends crucially on what F0 points are considered comparable between the two allotypes. If F0 values are compared as functions of absolute duration, then in the convex tones the intrinsic relationships are as just described. However, if F0 values are compared as functions of equalised duration (as will be done below), then for both convex tones the [u:] allotype has an intrinsically higher F0 for most of its duration. Moreover, for all tones the magnitude of an intrinsic difference thus established appears to correlate directly with its position in the speaker's F0 range. This relationship was quantified by linearly regressing intrinsic F0 differences at a given sampling point on their mean at that point.

Table 2. *Means and Standard Deviations (x, sd) for Fundamental Frequency (F0), Radiated Amplitude (Ar), Duration (D), Voice Onset Time (VOT), and F-pattern (F1,F2,F3) in Pakphanang Unstopped Tones on Syllables with* [a:] *and* [u:]. SP = sampling point; F0off = F0 at phonation offset. n = 9 (F-pattern), otherwise n = 3.

	TONE 1				TONE 2			
	[kha:]		[khu:]		[kha:]		[phu:]	
SP	F0	Ar	F0	Ar	F0	Ar	F0	Ar
0	- -	12.7,1.5	- -	12.0,1.4	- -	13.5,1.5	- -	11.9,2.8
5	230,22	15.9,1.4	254,5	18.0,0.8	224,13	18.0,1.0	226,7	20.0,2.9
10	230,13	18.6,3.0	257,10	20.2,1.8	221,11	21.0,2.9	231,3	21.4,3.4
20	243,8	18.9,3.3	275,6	19.6,0.5	211,10	21.1,3.3	225,2	20.5,3.1
30	254,3	16.9,2.3	285,5	18.3,1.1	203,8	19.3,3.0	220,5	20.3,3.9
40	264,2	17.4,1.8	285,9	17.3,0.6	202,9	18.1,2.5	216,3	19.7,3.1
50	259,7	17.0,1.2	271,12	17.2,0.2	199,8	16.0,2.3	210,4	18.8,2.6
60	235,13	14.8,0.8	249,21	16.8,0.2	193,8	14.7,1.5	207,8	18.4,2.9
70	200,6	13.6,1.6	217,30	16.3,1.0	192,9	13.6,1.1	205,3	15.7,3.5
80	169,5	14.2,1.5	185,27	14.0,1.0	193,10	11.5,1.2	208,1	12.1,2.9
90	172,7	11.3,2.1	161,17	9.0,0.9	198,9	8.7,1.2	208,4	7.1,3.7
95	- -	- -	- -	- -	194,11	- -	206,7	- -
100	169,10	4.7,1.9	145,7	4.0,0.0	- -	3.7,0.3	- -	2.7,1.0
F1	1127, 94				1140, 106			
F2	1723, 67				1656, 74			
F3	2675, 124				2809, 175			
D	47.3, 2.1		40.0, 5.2		56.8, 2.0		54.2, 6.5	
VOT	6.8, 1.9		7.8, 0.2		5.8, 1.0		4.8, 0.8	

	TONE 4				TONE 5			
	[ka:]		[ku:]		[kha:]		[khu:]	
SP	F0	Ar	F0	Ar	F0	Ar	F0	Ar
0	185,9	20.3,1.6	189,8	21.0,1.5	- -	17.0,1.2	188,9	15.8,2.9
5	182,5	27.8,2.7	197,3	29.0,2.3	190,3	19.2,1.6	208,4	23.1,2.5
10	178,4	26.2,3.0	196,2	29.8,2.6	188,4	21.0,2.3	215,4	23.7,2.8
20	171,3	24.5,2.0	192,2	26.0,2.6	186,4	22.0,2.5	208,5	24.1,2.8
30	172,3	21.5,0.9	185,5	23.0,3.0	183,5	21.2,2.0	201,4	24.9,2.8
40	169,2	19.9,0.8	189,4	22.9,2.1	177,6	20.5,2.3	196,5	25.5,2.3
50	167,2	18.9,0.6	186,6	20.7,3.1	171,7	20.0,2.0	183,6	23.2,1.3
60	166,3	17.2,0.3	185,4	18.4,2.4	158,7	18.4,2.4	161,5	21.9,0.7
70	168,4	15.1,0.9	188,4	15.5,3.7	141,6	17.2,1.9	146,4	17.0,1.9
80	166,4	12.9,0.8	189,5	12.2,2.4	130,4	14.3,4.8	137,6	12.6,0.5
90	175,12	10.2,1.2	190,6	11.4,3.6	135,10	10.8,3.3	142,10	13.0,2.1
95	182,7	- -	189,5	- -	129,3	- -	137,2	- -
100	- -	3.9,0.8	- -	3.7,1.2	- -	7.1,3.3	- -	4.8,1.0
F1	1072, 49				1121, 41			
F2	1549, 41				1610, 43			
F3	2815, 137				2854, 64			
D	60.7, 2.6		57.7, 1.9		44.6, 3.6		44.7, 2.5	
VOT	2.2, 0.3		2.3, 0.7		7.9, 0.6		9.4, 0.4	

Table continues.

TONE 3

Duration to F0 peak (100%)

SP	[ka:] F0	[ka:] Ar	[ku:] F0	[ku:] Ar
0	- -	17.6,3.5	- -	15.6,4.3
10	173,5	25.2,1.3	179,1	24.6,2.9
20	170,3	23.7,2.5	177,1	25.2,2.8
40	168,2	22.1,3.1	180,9	23.6,3.3
60	181,5	19.9,1.7	195,9	20.9,3.3
80	195,2	17.8,2.9	203,1	19.8,3.5
F0 peak	201,2	15.7,2.8	212,5	17.0,3.0
F1	1112, 76			
F3	2826, 100			
D	32.6, 2.1		31.8, 0.7	
VOT	3.5, 1.6		2.2,1.1	

Duration from F0 peak to F0 offset (100%)

SP	[ka:] F0	[ka:] Ar	[ku:] F0	[ku:] Ar
20	195,1	13.8,2.2	207,3	16.4,3.6
40	189,2	12.5,0.8	191,17	14.0,3.8
60	173,7	12.6,0.6	179,20	12.7,5.3
80	169,7	8.8,1.1	173,20	8.3,2.7
90	166,5	3.3,0.3	168,11	3.7,0.3
F2	1589, 72			
D	19.0, 2.8		17.1, 4.0	

TONE 6 / TONE 7

SP	Tone 6 [kha:] F0	Tone 6 [kha:] Ar	Tone 6 [khu:] F0	Tone 6 [khu:] Ar	Tone 7 [kha:] F0	Tone 7 [kha:] Ar	Tone 7 [ru:] F0	Tone 7 [ru:] Ar
0	- -	14.2,1.5	- -	15.7,0.6	184,7	15.0,1.5	166,5	16.7,4.7
5	171,4	19.7,1.1	176,2	23.1,2.0	181,4	21.6,1.1	182,1	25.0,4.3
10	165,5	22.8,1.4	173,3	24.8,2.3	172,3	23.0,1.3	181,2	26.9,4.0
20	155,2	23.0,2.5	165,5	24.2,3.4	161,3	24.0,1.1	171,0	26.8,3.3
30	151,2	22.5,1.4	163,6	24.4,3.3	154,5	22.0,1.0	163,0	26.2,2.6
40	153,3	21.3,2.7	164,7	24.9,3.1	147,4	20.7,0.8	156,1	21.7,2.1
50	152,2	19.8,2.4	167,8	23.9,3.0	142,2	19.1,1.6	151,0	19.7,2.3
60	154,3	18.1,1.8	168,8	22.0,3.1	136,2	17.0,1.2	144,1	18.3,2.8
70	158,2	16.9,1.6	170,11	18.8,5.0	131,2	13.4,1.5	143,5	15.1,4.6
80	165,4	14.2,1.3	177,10	15.1,5.8	132,5	11.2,1.8	144,2	12.6,3.2
90	179,8	10.6,0.5	199,6	12.3,4.6	136,6	8.1,2.3	148,1	9.6,2.3
100	200,9	5.3,1.5	223,8	4.8,2.5	141,1	3.9,1.4	147,5	3.0,1.0
F0off	200,9		216,14					
F1	1070, 79				1085, 64			
F2	1563, 88				1568, 62			
F3	2799, 78				2818, 110			
D	50.1, 2.5		46.4, 1.9		63.1, 2.5		60.9, 2.8	
Doff	55.0, 3.0		49.0,3.5					
VOT	6.6, 0.2		6.8, 1.4		7.6, 1.1			

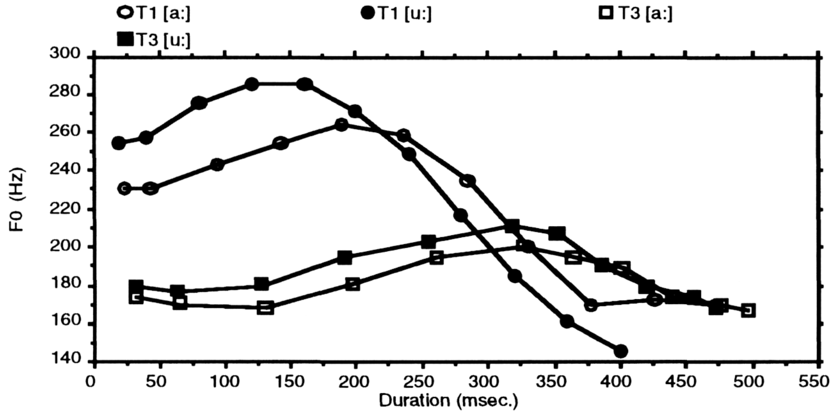


Figure 1. Intrinsic F0 Differences on Convex Tones 1 and 3.

For the speaker's combined Standard and Pakphanang tones (96 df), this gave a slope of 0.115, and an intercept of -5.5, with a very highly significant Pearson's *r* of 0.517. Thus the intrinsic difference associated with [a:] and [u:] vowels will be about 10 Hz at the lower end of the speaker's F0 range (ca.140 Hz), and more than twice that–about 26 Hz–at the upper (ca. 270 Hz). This difference in F0 correlating with vowel height is not enough to swamp the difference in F0 correlating with the tonal distinction between the level tones 2 and 4. There is no overlap between the F0 on the [a:] allotone of the higher of the two level tones (tone 2) and the [u:] allotone of the lower of the two level tones (tone 4): the tone 4 [u:] allotone still lies at least about 15 Hz below the tone 2 [a:] allotone.

Apart from the intrinsic F0 effects observed, differences can also be found in the tonal Ar. Of interest is that, contrary to expectations, the PPhN [u:] tones have about 1 to 2 dB higher overall Ar than the [a:], at least for most of their duration (i.e., from about the 10% sampling point to 80%). There appears to be a small differential effect, with the relatively higher pitched tones 1 to 4 showing a difference of 1.0 to 1.5 dB, and tones 5 6 and 7 showing a difference of between 2.1 and 2.5 dB. The same difference is also found in the speaker's Standard Thai forms, but here it is even bigger, with the [u:] allotypes being about 5 dB higher than the [a:] allotypes. In these data then, the variation in Ar associated with vowel height tends to mirror the intrinsic F0 differences.

TONAL EFFECTS

In order to obtain acoustic values that better reflect the PPhN tones, two things were done. Many of the F0 and Ar perturbations observable before phonation offset—for example the rise and fall in F0 at the end of tone 5, or the abrupt drop in Ar in all tones—are very similar to those which that over the last 15 to 25 csec of phonation in Tai-Phake (Rose, 1990, pp.395–396). Concomitant airflow data indicate the Tai-Phake perturbations to be caused by abduction of the vocal cords in anticipation of the voicelessness of the following pause. These perturbations do not therefore constitute part of the individual tones' acoustical properties, and so they were discarded by ignoring F0 and Ar values after 80% of duration (tones 1,2,4,5,7); after F0 peak (tone 6); and, for tone 3, after 60% of duration from F0 peak (these values gave the best agreement with the Tai-Phake data). Secondly, intrinsic differences associated with the [a:] and [u:] allotypes were also factored out by calculating the arithmetical means and standard deviations of all six tokens of a given tone. The resulting values are graphed as functions of absolute duration in Figures 2 (F0) and 3 (Ar).

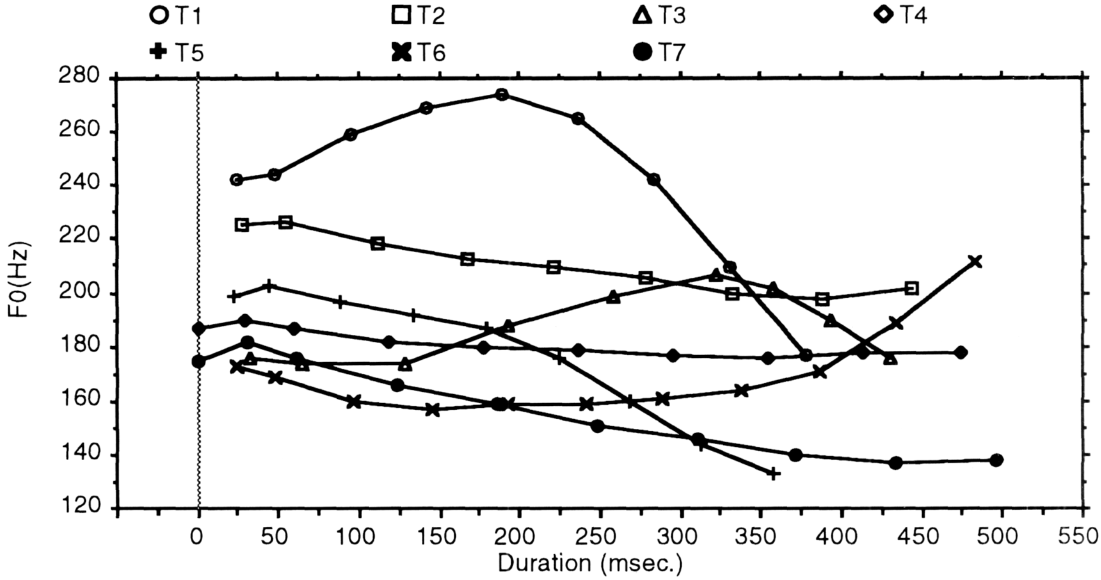


Figure 2. Mean fundamental frequency of Pakphanang unstopped tones.

Fundamental Frequency

The most conspicuous point in Figure 2 is the unevenness of the distribution of the F0 configuration. If F0 range is taken to be the distance between the highest and lowest targets (i.e., the highest point in the super-high tone 1 (ca. 275 Hz) and the lowest point in the falling tones 7 and 5 (ca. 135 Hz)), it can be seen that the F0 curves of this speaker's seven tones are located in a range of 140 Hz. However, five out of the seven tones are squashed wholly into the lower half of this range, and even two thirds of tone 2 lies within the lower half. Only the first two thirds of tone 1 and the first third of tone 2 occupy the upper half of this range. It is likely that the clear positive skew of the PPhN F0 configuration is due largely to the high value of the upper range-defining point of the super-high tone 1. This is indicated by a comparison with the speaker's Standard Thai tonal F0 recorded in the same session, where her upper range point is defined by the peak target value in the (high) Falling tone. Unlike the super-high tone, which sounds as if it lies well above normal pitch range, the speaker's Standard Thai Falling tone sounds as if it falls from the top of her normal pitch range (Rose, 1985). This tone has a peak target value (at 5% of duration) of 242 Hz, which is the same as the 5% F0 value in the PPhN super-high tone. With this value as the upper defining point of normal F0 range (the lower values are uncontroversially the same in both Standard and PPhN), mid range lies at about 190 Hz. Defining normal F0 range in this way results in a better characterisation of the speaker's tonal F0 shapes, because it corresponds better to the pitch impression of the tones. It is still crude, because it does not take into account the obvious declination of the data. Note, also, that the speaker's upper range-defining point is not recoverable from the PPhN data alone. Of the two convex pitch tones, the first half of the super-high tone 1 is located above the normal range, and its second half falls just into the upper part of the lower F0 range. The lower convex tone 3 starts in the upper part of the lower F0 range, and is fairly evenly distributed around the mid range point, rising into the lower part of the upper F0 range, and falling back into the upper part of the lower F0 range. The relative perceptual salience of the rising and falling pitches on these tones correlates nicely with the duration of their rising and falling F0 components. The two level tones lie on either side of the mid range point. Tone 2 is located mostly in the lower part of the upper F0 range, and tone 4 nearer the mid-range point, in the upper part of the lower half of the range. Both have mildly concave F0 shapes, with tone 2 showing a slightly steeper overall slope. Both falling tones 5 and 7 traverse the lower half of the pitch range. Almost the first third of the F0 time course of tone 5 has a considerably less steep slope than the remainder, and this probably corresponds to the initial level pitch noted for this tone. Although tone 5 starts just in the lower part of the upper half of the F0 range, it lies predominantly in the lower F0 range, with most of the initial flatter part distributed around the mid-range point. Tone 7, which lies wholly within the lower half of the F0 range, shows the opposite behaviour in contour: it falls for the first two thirds of its duration, and then flattens out. Note that, in addition to the difference in contour, tone 7 has a considerably longer duration than tone 5, and so their tonemic contrast in (pitch) height corresponds to an acoustic contrast in both duration and contour. It should be pointed out that, despite the similarity in F0 shape between tones 2 and 7—both have parallel falling F0 shapes in contrast to the more level shape of tone 4 which lies between them—tones 2 and 7 still differ in their (level vs. falling) pitch contour. Finally, the F0 of the low concave tone 6 is located mostly in the upper part of the lower F0 range, rising finally to a point in the middle of the upper F0 range.

It will be recalled that the upper and lower members of the convex and level pairs of tonemes are almost in complementary distribution with respect to aspiration/continuance on the syllable-initial consonant. It was also pointed out that their tonemic status is based on minimal pairs with sonorant onsets. Figure 2 shows that the segmental origin of the tonal split giving rise to this distribution is still reflected in similarity of F0 contour and relative F0 height. The shape of the high convex tone 1 occurring after aspirated consonants has a similar convex contour to that of the low convex tone 3 after unaspirated consonants, with the F0 of tones after aspirated consonants being overall higher. The same applies *mutatis mutandis* to tones 2 (after aspirated) and 4 (after unaspirated) consonants. The presence of a tonemic contrast precludes regarding the upper and lower pairs of the convex and level tones as conditioned by aspiration on the initial syllable and identifying the differences as correlates of the difference in aspiration on the initial consonant. There is also some comparative phonetic evidence from the absolute F0 values of the upper and lower members of each pair that clearly points to the extrinsic nature of the synchronic contrast. Data showing intrinsic differences in tonal F0 conditioned by a contrast between syllable-initial [x] and [k] (Rose, 1990, pp. 395–398) indicate that the magnitude of difference expected from conditioning by a syllable-initial aspiration/continuance contrast in a female high level and high falling tone would be about 30 Hz at phonation onset, but the effect would disappear within 20 csec. The F0 difference of about 70 Hz between the PPhN upper and lower convex tones is therefore far too big, and lasts too long, to be the result of intrinsic conditioning. And the F0 difference of about 30 Hz between the upper and lower level tones, although of comparable magnitude with intrinsically induced effects, differs from them in its perseveration through the duration of the tones.

F-Pattern

Analysis of variance (ANOVA) on the raw formant values followed by Fisher's PLSD showed that the super-high tone has a higher F2 and a lower F3 than the other tones. Mean differences (134 Hz, F2; 145 Hz, F3) are significant at least at the 0.05 level. This appears to be part of a general correlation between F0 and F-pattern also present in the other tones, where linear regression of formant values at the three different sampling points in the syllable with corresponding F0 values showed that F2 (and F1) correlate positively with F0, and F3 negatively. However, the effect seems exaggerated in tone 1, because, as shown by the ANOVA results, it has F2 values somewhat higher, and F3 values lower, than predicted from the correlations from the other tones. It is well known that larynx height is positively correlated with F0, but it is difficult to reconcile the observed PPhN F-pattern correlations with larynx raising, which produces either overall increases in all formant values (Sundberg & Nordstroem, 1976), or an increase in F1, but decrease in F2 and F3 (Nolan, 1983, p.182). According to the nomographic data in Fant (1960, pp. 83, 84), which are admittedly for male vocal tract dimensions, the convergence of F2 and F3 for back vowels under conditions of constant, high F1 for an open vowel is consistent with a shift towards the larynx in the location of the point of maximum constriction in the area function. This may therefore be an additional productional feature of the super high tone. (F2 and F3 will also converge as a function of decrease in the cross-sectional area at the maximum constriction point, but this is accompanied by a substantial decrease in F1, which is not observed in the data.) Because of the general correlation between F-pattern and F0, tonal height is also reflected in F1 and F2: High tones (1, 2, and 5) have higher F1 and F2 than Low tones (3, 4, 6, and 7).

Amplitude

The most interesting thing about the Ar shapes of the seven PPhN tones is that they look so similar. Figure 3 shows that they all share a contour involving a quick initial rise of about 8 dB to peak, followed by a more gradual decay to offset. This shape and lack of differentiation also characterises the Ar of Thai Phake tones, irrespective of F0 shape (Rose, 1990, p.396). Note, however, that this does not appear to be the case with Standard Thai tonal Ar, in which the tones have Ar contours that generally follow their F0 shape (Abramson, 1975, p.7).

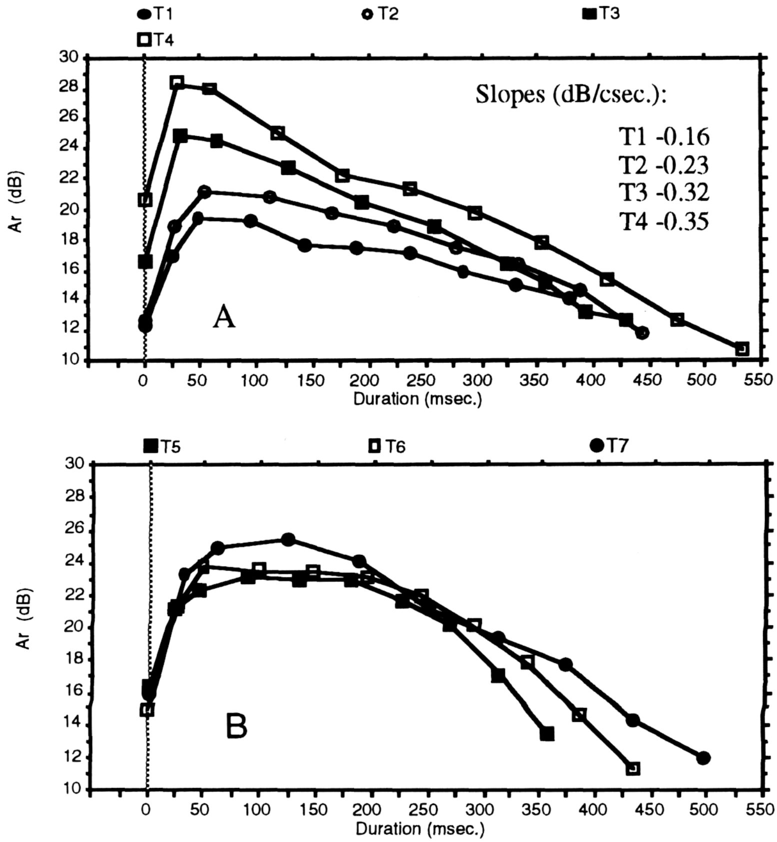


Figure 3. Mean Ar shapes for Pakphanang unstopped tones.

Despite their overall similarity, systematic features of the PPhN Ar shapes can be found that differentiate six of the seven tones. The first is the Ar profile after about the first 5 centiseconds. For most of their duration after this point, the Ar of tones 1, 2, 3, and 4 decays in an effectively linear fashion (Figure 3a); the other tones have more convex Ar shapes (Figure 3b). Tones 1, 2, 3, and 4, with straight profile, can be divided into two groups on the basis of the slope of the post-peak Ar. Tones 3 and 4 (the low convex and level tones) have relatively steep slopes; tones 1 and 2 (the high convex and level tones) have relatively flatter slopes. It will not have escaped attention that this slope feature also correlates with aspiration: amongst tones with the straight Ar shapes, those with unaspirated consonants have the steeper post peak profiles. However, unlike the Ar details at onset which correlate highly with aspiration as reflected in VOT values (tones with aspirated initial consonants have more curved Ar onset profiles, with delayed peaks occurring from 4 to 12 csec. after onset, and less steep slopes of from 1.7 to 0.8 dB/csec.), the slope feature is realised over effectively the rest of the syllable, which would be the expected domain of realisation of tonal, as opposed to (initial) segmental features. Ar height serves to distinguish the Ar of the convex tones 1 and 3 from the level tones 2 and 4: level tones have higher Ar values than convex tones (Figure 3a). This feature reflects a general inverse correlation between Ar and F0: the convex tones lie overall higher in F0 than the corresponding level tones, but have lower Ar. As far as the tones with convex Ar profile are concerned (tones 5, 6 and 7), Ar height also serves to distinguish tone 7 from tones 5 and 6. Note that for the two falling F0 tones 5 and 7, as with the straight profile tones, Ar height correlates inversely with F0. I cannot see any feature in the Ar shape of tones 6 and 7 which can plausibly differentiate them. The fact that they have effectively the same Ar shape is interesting considering that their F0 shapes—low dipping and low falling—are so different.

Because of their lack of contrastivity, the PPhN Ar contours almost certainly do not have any distinctive perceptual function. This is not to say that they may be perceptually irrelevant. It has been demonstrated for Standard Thai, for example, that although tones cannot be identified from their characteristic Ar shapes in the absence of F0, addition of realistic Ar contours enhances tonal identification (Abramson, 1975).

These findings are significant in the light of data on tonal Ar shapes in varieties of Chinese. In Zhenhai dialect for example, citation tone Ar shapes are just as contrastive as F0 shapes and tones can be visually identified on the basis of their amplitude shape (Rose, 1982, pp. 35–37). In Peking dialect, 71% of tones in unedited running speech can be statistically identified from their Ar contour using discriminant analysis. The success rate is almost as good as with their F0 (Coster & Kratochvil, 1984, p. 124). The Ar contour of Modern Standard Chinese tones also appears to be a valid perceptual cue. Standard Chinese contour tones, i.e., tones 2, 3 and 4, can be perceptually identified from their Ar contour (Whalen & Xu, 1992). As well as providing the perceptual cues for speech, acoustics reflects production. Chinese and PPhN tones must therefore differ in some productional aspect or aspects. The section below investigates this further.

PHYSIOLOGICAL INFERENCES

Although the perceptual relevance of the PPhN Ar shapes may be minimal, they remain important for the inferences they allow from their relationship with the F0 shapes as to the way the tones are produced. The acoustic dimensions of F0 and Ar are related, as reflected in the name of the received "myoelastic-aerodynamic" theory of phonation, to parameters of vocal cord tension (VCT) and subglottal pressure (Ps). It is normally assumed that, other (usually supralaryngeal) things being equal, F0 in speech is primarily controlled by VCT, and that Ar reflects changes in Ps. (Examples of F0 and Ar differences reflecting supralaryngeal factors have been mentioned above, in the correlations with vowel differences.) However, these correspondences (F0 to VCT; Ar to Ps) are complicated by the fact that Ps can also cause changes in F0, and VCT can occasion changes in Ar. The former case often occurs, for example, in the realisation of primary, emphatic, or contrastive stress; the latter is said to occur as a result of positive correlation between VCT and glottal duty cycle (Ohala, 1978, pp. 18–19). As VCT increases to implement a rise in F0, the amount of time the cords are closed within each glottal period can also increase, thus increasing the mean glottal resistance and, thereby, the Ps. This correlation introduces an additional complication, since it implies that Ps can be controlled by two sources—pulmonically, or laryngeally via VCT. Since, however, glottal area is an independently controllable parameter, it must be assumed that changes in Ar can also occur as a result of changes in Ps brought about by deliberate changes in glottal area, such as when the cords are abducted towards the end of phonation, or execute contrastive phonatory gestures. Deliberate laryngeal control of Ps from mean glottal area is said to be the main mechanism underlying the control of loudness in speech (Isshiki, 1964).

The complexity of relationships just outlined makes interpretation of Ar and F0 in productional terms problematic. In particular, when Ar and F0 are congruent it is difficult to tell whether the Ar reflects extrinsic, pulmonic Ps, or intrinsic control via VCT. (It might be possible under these circumstances to tell what mechanism underlies the congruence from the magnitude of the effect. That is, there might be less Ar per unit increase of F0 under conditions of VCT control than under conditions of pulmonic control. This has yet to be tested. Changes in glottal resistance will also be revealed by controlled airflow studies.) However, when F0 and Ar change in different directions—specifically, when F0 rises and Ar falls, or remains constant—it is reasonable to infer VCT control of F0. In this paper, I have explored a method, developed by Monsen, Engebretson, and Vemula (1978), of investigating the separate contributions of VCT and Ps to the production of observed F0 and Ar. Monsen et al. (1978) used the Ishizaka-Flanagan model of the vocal cords to generate maps that relate physiological parameters of VCT and Ps to acoustical parameters of F0 and RMS glottal wave amplitude. Details of the application, together with the results of an analysis on another, Chinese, tone language, are in Rose (1984). The method involves plotting the observed PPhN acoustical values on the map and reading off corresponding "VCT" and "Ps" values. (I have enclosed these in quotes to indicate their inferred nature.) Amplitude was aligned so that 25dB Ar in the PPhN data corresponded to 14 dB RMS glottal amplitude on the map. This allowed the discontinuity in the Ar derivative reflecting abduction of the cords at the end of phonation to correspond realistically to a Ps value a little more (1 to 3 cms H_2O) than the minimum transglottal pressure drop of 2 cms H_2O required for phonation. One of the limitations of the application of the model is that the parameter of mean glottal area is held constant. This means that any changes in Ar in the data that reflect changing mean glottal area would be attributed, incorrectly, to pulmonic Ps. However, since there are no examples of ambiguous, congruent Ar and F0, this is of no consequence. A further limitation is that physiologically relevant dimensions, e.g., vocal cord length and mass, are assumed to be those of the typical female values in Monsen et al. (1978, p.69). As will be seen, this assumption yields rather natural results.

Figure 4, parts a-c, shows the inferred "VCT" and "Ps" values as functions of absolute duration for the seven PPhN tones as read off the female map in Monsen et al. (1978, Figure 4, p. 69). Values have been plotted for the tonally relevant portion of the contours (this was taken to be from about csec. 5 to the 80% point, see above).

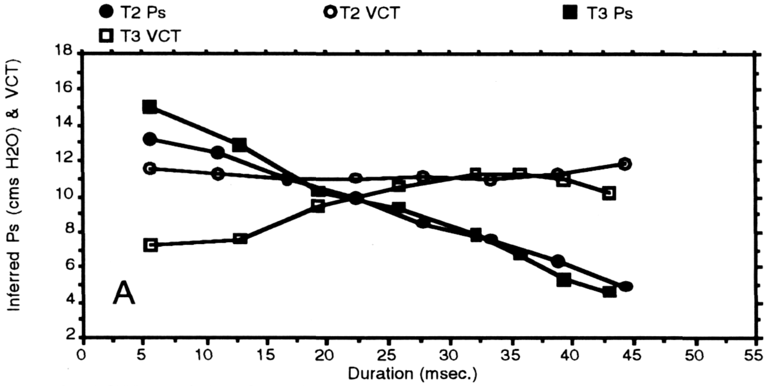
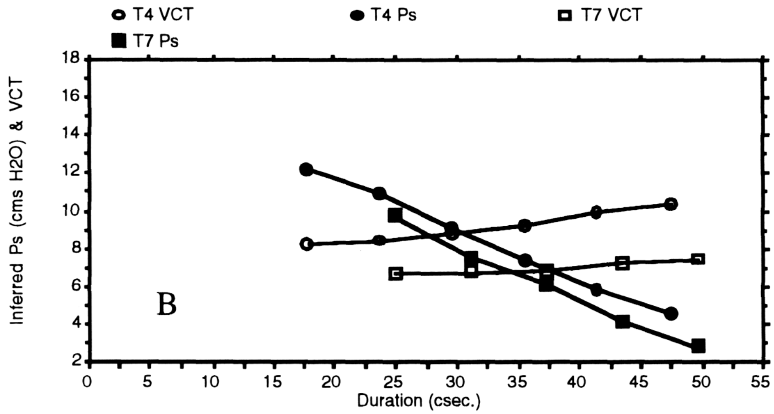


Figure 4a. Inferred VCT and Ps for Tones 2 and 3.

An interpretation of the Ar and F0 shapes in terms of pulmonic Ps and VCT is plausible for all except the initial parts of tones 4 and 7, which have been omitted in the figures. The observed F0 and Ar of the upper level tone 2, for example, are resolved as if they had been produced by a combination of constant VCT and gradually falling Ps; and the F0 and Ar of the low convex tone 3 are resolved as, again, gradually falling Ps and a convex VCT (figure 4a). (In tones 4 and 7, the amount of Ar relative to F0 from 5% to 30% of duration is so high that it would require the implausible interpretation of a compensatory drop in VCT. This suggests that Ar values over this initial portion may still also reflect non-source features, like the effect of the initial consonant.)

Figure 4, parts a–c, shows that in general the VCT shapes mirror the F0, while the Ps, like the Ar, is largely undifferentiated, and shows gradually falling profiles. These correspondences in overall pattern suggest that the observed PPhN Ar shapes are explainable in terms of the gradual intrinsic drop in Ps that is usually observed in a declarative utterance. The resolution by the model has resulted in a slight redistribution of the original variance in the input F0 and Ar shapes. For example, as pointed out above, the Ar shapes of tones 2, 3, 4, and 7 differ variously in peak and contour, but their corresponding "Ps" shapes are similar in slope and intercept. There are also some changes in the "VCT" shapes compared to the F0. Tone 5, with a level-falling F0 contour is resolved with a simple falling "VCT" shape, and the low dipping tone 6, with a level F0 component, is resolved with concave "VCT" (Figure 4c).

Tones 2, 3, 4, and 7 are all resolved with similar falling "Ps" shapes. This allows the inference that they are all produced with the same minimal extrinsic Ps involvement. Linear regression of "Ps" on duration for these tones shows "Ps" to decline at very similar rates: -0.22 (T2), -0.27 (T3,T4), -0.28 (T7) cms H_2O/csec., with similar intercepts: 14.6 (T2), 16.3 (T3), 17.0 (T4), 16.4 (T7) cms H_2O. The mid falling tone 5 and low dipping tone 6, however, show slightly different "Ps" shapes, with initial level portions before decay. Tone 5 has a rate of decay of its falling part of -0.56 cms H_2O/csec., much greater than the range of values in tones 2, 3, 4, and 7. The rate of decay of the falling part of the tone 6 "Ps" is much more like the others: -0.3 cms H_2O/csec..



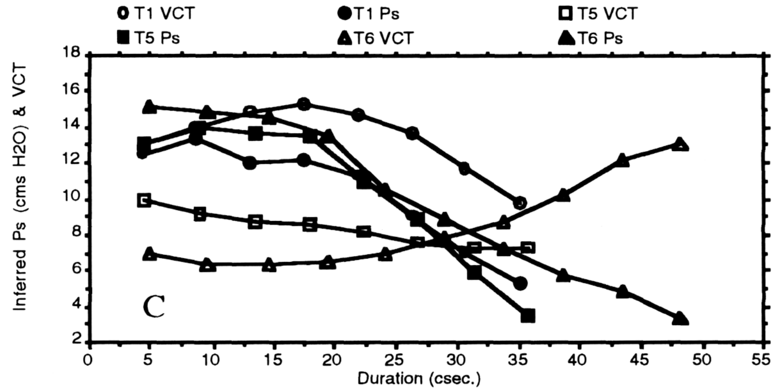


Figure 4b-c. Inferred VCT and Ps for Tones 4 & 7 (B), and 1, 5, and 6 (C).

In the mid falling tone 5, unlike all the other tones, the F0 contour is reflected more in the "Ps" than the "VCT" (Figure 4c). This is because of its large amount of Ar drop (9.7 dB) relative to F0 drop (54 Hz). Under conditions of constant VCT, this magnitude of Ar drop would account in the model for 60% of the observed F0 decrease. As explained above, it is not clear whether this reflects extrinsic pulmonic control of Ps, or whether it results from changes in Ps attendant upon changes in mean glottal area, together with intrinsically declining pulmonic Ps. That is, the speaker could be lowering F0 in tone 5 by a combination of slight relaxation of VCT and deliberate decrease in pulmonic Ps, or—more likely, given the fact that F0 is controlled primarily by VCT—by decreasing the VCT in a way that results in decreasing glottal resistance and consequently gradually decreasing Ps.

The F0 shape of the low dipping tone 6 (Figure 4c) is reflected in its "VCT" profile, just like tones 2, 3, 4, and 7. In addition, its rate of decay is close to that of these tones. It therefore differs in its initial Ps shoulder, and this may, as in tones 4 and 7, be the result of too high Ar values relative to F0 from the influence of the initial consonant.

The productional interpretation of the high convex tone 1 (Figure 4c) is not totally clear. Its "Ps" profile is ambiguous, being interpretable as either level-falling, like tones 5 and 6; or overall falling, like tones 2, 3, 4, and 7. In the latter interpretation, it has a rate of decay (-0.25 cms H_2O/csec) and intercept (17.4 cms H_2O) very similar to the other overall falling tones. In the former, the rate is considerably steeper: -0.42 cms H_2O/csec, thus showing a possible productional similarity to the mid falling tone 5. However, tone 1 is clearly unlike tone 5 in having its F0 contour reflected in the inferred "VCT," and it is best to assume that it too is produced with minimal extrinsic Ps involvement.

In all except the mid falling tone 5, then, the observed F0 and Ar can be explained in terms of an intrinsically declining pulmonic Ps and extrinsic VCT. Only in tone 5 is it necessary to posit the additional factor of glottal resistance covarying with VCT. This confirms what can already be suspected from the several cases where F0 rises accompany Ar falls, namely that in the citation tones of this variety of Thai, F0 is primarily controlled by VCT with minimal Ps involvement. This has also been demonstrated by Erickson (1976) for Standard Thai with an electromyographic investigation of intrinsic and extrinsic laryngeal musculature. There is a contradiction in this result, however, because the PPhN dipping tone 6 and Standard Thai Rise tone are very similar in F0 shape, but, as pointed out above, have differing Ar shapes. This must indicate that they differ in production, unless it is to be attributed to a sex-related difference.

These results contrast markedly with inferred "VCT" and "Ps" shapes derived from the Ar and F0 in the Zhenhai dialect of Chinese, which, as mentioned above, has highly contrastive Ar contours. Rose (1984, pp.151–157) compared the F0 and Ar shapes of the PPhN low dipping and low convex tones with tones of comparable pitch from Zhenhai, and showed that they had effectively the same F0, but very different Ar shapes. The inescapable conclusion is that tones are being produced with different gestures in the two languages. In terms of the conceptual framework outlined above, the production of Zhenhai tones may involve more pulmonic involvement, or some kind of covariation of mean glottal area with VCT. The Southern Thai data appear to support very little extrinsic Ps involvement, all tones being produced with decreasing Ps.

ACKNOWLEDGEMENTS

I record here my gratitude to Dr. Tony Diller for his tremendous help and encouragement. He first drew my attention to the complex Southern Thai tone systems; wrote the Thai prompts for the recording and co-supervised it; made two more sets of recordings in Thailand for me; and made his field notes available. I also want to thank Ms. Sutira Wacharaboworn for acting as informant for this study during her stay as visiting fellow at the Australian National University, and subsequently in Southern Thailand; and Wipa Treerat for recording a second Pakphanang speaker for me in Bangkok.

REFERENCES

Abramson, A.S. (1975). The tones of Central Thai: Some perceptual experiments. In J.G. Harris & J. Chamberlain (Eds.), *Studies in Thai Linguistics* (pp.1–16). Bangkok: Central Institute of English Language.

Coster, D.C., & Kratochvil, P. (1984) Tone and stress discrimination in normal Beijing dialect speech. In B. Hong (Ed.), *New Papers on Chinese Language Use* (pp. 119–132). Canberra: Contemporary China Papers no. 18, Australian National University.

Diller, A.V.N. (1979a). Tones, segments and Thai regional society. In T.L. Thongkum, P. Kullavanijaya, V. Panupong, & K. Tingsabadh (Eds.), *Studies in Tai and Mon-Khmer Phonetics and Phonology in Honour of Eugénie J.A. Henderson*. (pp. 60–93). Bangkok: Chulalongkorn University Press.

Diller, A.V.N. (1979b). How many tones for Southern Thai? In D.L. Nguyen (Ed.), *South East Asian Linguistic Studies*, vol. 4 (pp. 117–29). Canberra: *Pacific Linguistics*, Series C. Australian National University

Diller, A.V.N., (1987). Reflections on Tai diglossic mixing. *Orbis, 32*, 147–166.

Erickson, D. (1976). *A physiological analysis of the tones of Thai.* Unpublished doctoral dissertation, University of Connecticut, Storrs.

Fant, G. (1960). *Acoustic theory of speech production.* The Hague: Mouton.

Haas, M. R. (1958). The tones of four Thai dialects. *Bulletin of the Institute of History and Philology Academia Sinica, 29,* 817–826.

Hombert, J-M. (1977). A model of tone systems. In *Studies on Tone, UCLA Working papers in Phonetics, 36,* 20–32.

Isshiki, N. (1964). Regulatory mechanism of voice intensity variation. *Journal of Speech and Hearing Research, 7,* 17–29.

Monsen, R.B., Engebretson, A.M., & Vemula, N.R. (1978). Indirect assessment of the contribution of subglottal air pressure and vocal cord tension to changes of fundamental frequency in English. *Journal of the Acoustical Society of America, 64,* 65–80.

Nolan, F. (1983). *The phonetic bases of speaker recognition.* Cambridge Studies in Speech Science and Communication. Cambridge: Cambridge University Press.

Ohala, J. (1978). The production of tone. In V.A Fromkin (Ed.), *Tone: A linguistic survey* (pp. 5–39). New York: Academic Press.

Rose, P.J. (1982). Acoustic characteristics of the Shanghai-Zhenhai syllable-types. In D. Bradley (Ed.), *Tonation* (pp. 1–53). Papers in South-East Asian Linguistics No. 8. *Pacific Linguistics,* Series A. Canberra: Australian National University.

Rose, P.J. (1984). The role of subglottal pressure and vocal chord tension in the production of tones in a Chinese dialect. In B. Hong (Ed.), *New papers on Chinese language use* (pp. 133–168). Canberra: Contemporary China papers, no. 19, Australian National University.

Rose, P.J. (1985). Comparing the tones of Central and Southern Thai: Data from a bidialectal speaker. Paper presented at the 18th Conference on Sino-Tibetan Languages and Linguistics, Bangkok.

Rose, P.J. (1990). Thai-Phake tones: Acoustic, aerodynamic and perceptual data on a Tai dialect with contrastive creak. In R. Seidl (Ed.), *Proceedings of the Third Australian International Conference on Speech Science and Technology* (pp. 394–398). Canberra: Australian Speech Science and Technology Association.

Sundberg, J., & Nordstroem, P.E. (1976). Raised and lowered larynx—the effect on vowel formant frequencies. *Quarterly Progress Status Report, Speech*

Transmission Laboratory, 2-3, 3539. Stockholm: Research Institute of Technology.
Whalen, D.H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica, 49*, 25-47.