My focus here is on the languages of Southeast Asia, which comprise tone languages and voice-register languages, as well as languages that fall into neither category. These languages include not only official national languages with their regional and social dialects but also the languages of numerous ethnic minorities (see, e.g., Smalley, 1994).

Tones

In a tone language distinctive pitch levels and contours serve, along with vowels and consonants, to make up a word. Such languages vary as to how many phonologically relevant tones they have. Even within one language, all its regional dialects need not have the same number. For example, Standard Thai, the official language of Thailand, which is based on the Central dialect of Bangkok and its environs, has five tones, but dialects of most other areas have six tones. For many speakers of Western languages it is hard to grasp the concept of pitch as an integral component of every word or morpheme. Thus, the Thai or Chinese child, for example, through mimicry of his or her caregivers must acquire each new word as a combination of a syllable type formed from the stock of vowels and consonants with a particular tone from the set provided by the language. That is, the tone is an integral part of the word or morpheme, not something added to a syllable. Published studies of children’s acquisition of their tonal systems together with sequences of consonants and vowels do not readily come to mind. An early one (Li & Thompson, 1978) is still important in providing guide posts for such work. Another study of the same period (Tuanycharoen, 1977) is a broad treatment of a child’s acquisition of Thai phonology in general, including tones.

It is often asked whether one has to be musically gifted to learn a tone language. The populations of such countries as Thailand, Myanmar (Burma), Laos, and Vietnam include people with various levels of musical training and skill, but all normal individuals have full control of the phonological distinctions, including tones, of their languages. As for a foreigner wishing to learn his or her first tone language, the task is not like learning to sing
songs, because there is no fixed musical score involved. Indeed, a person's “tonal space” can rise and fall with sentence intonation. In addition, co-articulation (Abramson, 1979) and speech rate (Gandour, Tuntavavitkul & Satthamnuwong, 1999) can affect the tonal shapes. The differences between the pitch levels and contours of such a language are relative, not absolute. A big man with a large larynx and a little child with a small larynx will seldom if ever overlap with each other in absolute pitch heights, yet the two of them can talk to each other, although their tonal spaces are well separated on the frequency axis. It is to be understood, of course, that for intonation and perhaps stress pitch has a role in non-tonal languages too (Lehiste, 1970).

It is important to distinguish between pitch, an auditory property, and its principal acoustic correlate, frequency. In the acoustic analysis of speech the repetition rate of the glottal pulses is the fundamental frequency (F₀) of the utterance moment by moment. This F₀ is stated as the number of cycles per second, which are conventionally labeled as Hertz (Hz). In order to treat the F₀s of a number of speakers as if they were in the same frequency space, investigators often convert the acoustic data in Hertz to a scale closer to the auditory sensation of pitch, such as semitones. This non-linear relation between F₀ and pitch is illustrated in Figure 21.1. Let me stress that the way was paved for experimental work by linguists, experimental phoneticians, psychologists, and engineers by fieldwork done by practical phoneticians with a keen ear for pitch patterns.

Jackson T. Gandour (1978) gives a very good overview of early experimental work on the perception of tones. He includes (pp. 63–71) his own multidimensional scaling analysis of tone perception. He finds five dimensions that account for subjects’ judgments of heights in the voice range and contour shapes: average pitch, direction, length, extreme endpoint, and slope.

Finally, it must be said that while a pitch pattern is normally the major, or even the only, perceptually important phonetic attribute of a tone, there may also be audible concomitant properties. Examples are voice quality, syllable length, and loudness. The question arises as to whether in the absence, of course, of pitch in whispered speech such concomitant properties are retained and enable speakers to produce and perceive at least a somewhat diminished set of tones (Giet, 1956). For Thai the answer is that tonally contrasting words in the same sentence carrier are to a very limited extent identifiable in whisper in the absence of pitch (Abramson, 1972).

Registers

The concept of phonologically relevant voice register in certain languages was first mentioned in print, it seems, by Eugénie Henderson (1952). She defined it as a complex of laryngeal and supralaryngeal properties, one of which may be dominant. Fieldworkers’ impressionistic descriptions include such characteristics as voice quality, vowel quality, vowel length, and pitch. Voice quality is mainly an auditory correlate of phonation type, the selling of the larynx to yield a particular kind of voice. Deviation from normal (“modal” or “clear”) voice includes such types as “breathy” or “creaky” voice (Laver, 1980). Perceived vowel quality or “color” has as its major acoustic correlate the frequency locations of the resonances (“formants”) of the supraglottal vocal tract. An impression of length has as its major acoustic correlate duration, usually measured in milliseconds (ms). Pitch has already been discussed.

In Figure 21.2 we see electroglottographic (EGG) displays of two glottal periods in each of two words distinguished by register in one variety of the Mon language. An electroglottograph works by passing a weak electrical current from one electrode on the neck through the larynx to another electrode. The current is conducted more easily through human tissue than through air. As the vocal folds come together and close the impedance is lessened and more current flows, so we see the glottal pulse rising to a peak. As the contact between the folds begins to loosen, less current flows and the shape of the glottal pulse changes in a downward movement until the glottis is...
open, when the trace more or less levels off for the open quotient until the rise begins for the next closing movement of the folds. In the modal voice in the lower part of the figure we see that the open quotient of the vocal folds between pulses is very brief, whereas in the breathy voice of the upper part we have a much longer virtually level open quotient that permits a somewhat continuous flow of air during voicing to yield breathy voice or murmur. A new book (Kreiman & Sidtis, 2011) is an excellent compendium of research on phonation. Certain articles (e.g., Ladefoged, Maddieson, & Jackson, 1988; Wayland & Jongman, 2003) provide a helpful introduction to the acoustic analysis of phonation types, while others (e.g., Gilbert, Potter, & Hoodin, 1984) describe electroglottography.

Conventionally, as suggested by the name voice register, one expects voice quality to be the dominant property; however, in certain register languages pitch seems to be taking over the dominant role, suggesting the possible ultimate emergence of tones. Indeed, it has been plausibly argued (Thurgood, 2007) that the process of tonogenesis commonly includes an earlier stage of voice register. Also, the register languages of concern here, generally the languages of ethnic minorities, are very likely to be in contact with regional tone languages. That is, the speakers are largely bilingual in their own language and the language of the surrounding major culture, including sometimes both its standard form and its local regional dialect. Thus, such contact with a prestigious tone language might help promote an increase of the salience of pitch among the properties of the registers. Important other contributions to the topic are Gandour’s work with colleagues on speaking rate and tones (Gandour, Tumtavavitikul, & Satthamnuwong, 1999) and studies of tone and the brain (e.g., Gandour, Petty, & Dardarananda, 1988; Gandour et al., 2002). As for the brain, one study (Van Lancker & Fromkin, 1973) showed lateralization to the left hemisphere for Thai tones.

The voice-register languages of Southeast Asia are essentially to be found in the Mon–Khmer subfamily of the Austro-Asiatic family (Thongkum, 1988). Its languages fall into three types (Ferlus, 1980): (1) non-tonal and non-register, (2) register (usually two register complexes), and (3) tonal (two tones). Indeed, it turns out that one language, Khmu’, has a distribution of its dialects over all three types. Type I is historically the most conservative in that it maintains the old consonantal voicing contrast, while the other two types are departures from that stage.

### Fuzzy boundary

The boundary between the two typological categories, tone and voice register, is rarely clear and normally rather fuzzy (Abramson & Luangthongkum, 2009). Indeed, even the well-trained and highly experienced field linguist may at times be bewildered as to whether he or she is working with informants who speak a tone language or a register language. Ilse Lehiste (1970, p. 79) comments, “It is not impossible that in two words presumably differing in tone, the fundamental frequency differences may happen to be minimal, while concomitant features of intensity, quantity, or segmental quality may carry the chief distinctive burden.” Consequently, since language is always in a state of flux with changes taking place in spite of any conservative force that may exist in the culture, it can happen that the phonetic properties distinguishing particular phonemes will shift in their relative salience; also, certain properties may in fact disappear altogether or may be replaced by others. Thus, if what has always been known as a tone language shows one or more voice qualities working, as it were, with pitch or even in place of pitch, to distinguish one or two tones from the rest of the set, are we observing the
Some studies on tone languages

As far as I can tell, experimental studies of the production and perception of phonological distinctions in the languages of the area did not begin to appear in print until after 1960. There is, however, an early forerunner of this development, Cornelius Beach Bradley, who must be mentioned. He was born in Bangkok on November 18, 1843 and spent his childhood there as an English–Thai bilingual before going with his parents to their home in the United States for his further education. Returning to Thailand in 1871, he served as a missionary until 1874 and then returned to various posts in the United States.

Using himself as the speaker of Thai for his study, Bradley, using a kymograph, made $F_0$ tracings of his own utterances of the five tones of Bangkok Thai (Bradley, 1911) with “Abbe Rousselot’s apparatus” (p. 284). Apparently he means the kymograph, a battery-operated device with a sheet of smoked paper wrapped around a drum that rotated at a constant speed. A stylus, wiggling in response to movements of attached body part or the sound waves of a human voice, traced patterns on the smoked paper (Rousselot, 1897–1901). Bradley obviously deserves credit for his pioneering work, yet it is hard to evaluate the results as typical of Thai speech of the end of the nineteenth century and a bit thereafter. To say that someone is a bilingual does not necessarily mean native-like proficiency in both languages. Unfortunately, there seems to be no independent evidence on this matter available for Bradley. Also, how adept was he in handling the apparatus? Unfortunately, certain scholars (e.g., Teeranon & Rungrojsuwan, 2009; Thepboriruk, 2010) in their treatments of ongoing changes in the Thai tonal system, have accepted without hesitation the authenticity of Bradley’s $F_0$ curves.

In an acoustic and perceptual study of Thai vowels and tones, Abramson (1962) published average $F_0$ contours for the five tones on short and long vowels. Since at that time, really reliable programs for automatically extracting $F_0$ from speech were not yet available, measurements were made at closely spaced intervals across harmonics on narrow band spectrograms. Figure 21.3 shows the results for long vowels. Their satisfactoriness was validated by native speakers in listening tests with synthetic speech. Actually, the 1962 monograph was a slight revision of the author’s 1960 doctoral dissertation at Columbia University, so the data for the $F_0$ contours were obtained no later than 1960. Fifteen years later (Abramson, 1975) the same contours in new synthetic stimuli achieved close to 100% identification as intended, although included was an experiment in which suitable contours of amplitude, a concomitant property, slightly enhanced the intelligibility of the $F_0$ contours. Somewhat later a study (Tingsabadh & Deeprasert, 1997) showed very similar shapes for the tones in citation forms; not surprisingly, this work showed deviations in running speech, as did the small samples of such speech in Abramson (1962, 1975). I hasten to add, however, that current observers (Zsiga & Nitisaroj, 2007; Teeranon & Rungrojsuwan, 2009; Thepboriruk, 2010) have indeed observed ongoing changes without hesitation the authenticity of Bradley’s $F_0$ curves.
of tonogenesis by applying experimental methods to other modern Asian languages. Acoustic studies of the tones of some regional dialects of Thai have appeared. One (Rose, 1997) examines a variety of Southern Thai with seven tones. It has also been shown that in Thai and other languages of the area there is some systematic interaction between distinctive vowel length and the pitch of tones (Gandour, 1977; L-Thongkum, Teeranon, & Intajamornrak, 2007).

Another tone language of Southeast Asia is Vietnamese, which, depending on one’s analytic stance, has either six or eight tones. Anyway, for the six tones occurring contrastively in the same environment (Pham, 2003), pitch patterns are clearly relevant; nevertheless, for at least part of the system two voice qualities, breathiness and creakiness, are equally salient and relevant to the distinctions. This takes us back to the discussion of fuzziness above. A similar situation is found in Green Mong, a dialect of Hmong (Andruski & Ratliff, 2000; Andruski & Costello, 2004). Later work on Vietnamese (Brunelle, Nguyen, & Nguyen, 2010) focuses on their examination of laryngeal behavior.

Some studies on voice-register languages

Theraphan Luangthonkum, who used to write under a shortened form of her surname, L-Thongkum or just Thongkum, is a pioneer in the acoustic study of minority languages of Thailand and surrounding countries, with attention to the voice registers of Mon–Khmer languages (Thongkum, 1988, 1989, 1991). Although these languages commonly have just two voice registers, one exception is Chong, which has four: modal, creaky (or tense), breathy, and breathy-creaky (Thongkum, 1991; Edmondson, 1996; DiCanio, 2009). The fourth one is dynamic in that it shifts from one laryngeal setting to another in the same syllable.

With earlier work (Thongkum, 1989) as a starting point, a somewhat more elaborate acoustic analysis of the Kuai dialect of Suai, a Mon–Khmer language spoken in Thailand, was carried out together with perceptual experiments (Abramson, L-Thongkum, & Nye, 2004). The language has two registers conventionally labeled modal and breathy. The testing of the perceptual value of various potential cues to phonetic properties of registers was apparently an innovation in the literature. For example, incremental changes in spectral tilt, an acoustic correlate of the duration of the open quotient of the glottal period, provided one set of stimuli, as did variants in contours of $F_2$ and amplitude. Spectral tilt means the intensity ratio of upper harmonics to that of the fundamental harmonic. As the open quotient is lengthened for breathiness, the upper slope of the spectrum falls more in intensity. Of course, included in the project was a detailed acoustic analysis of a large corpus of

spoken Kuai. The general conclusion was that the language, at least as spoken in the village of Samrong, is in a state of flux. The mixture of results for both production and perception yields the impression that it may be undergoing a shift from a register language to one with some kind of accentual salience; the latter, then, could be a transitional stage to phonologically distinctive tones. Insofar as the distinction is still viable, the registers are mainly distinguished by $F_2$ contours with some help from phonation types.

Not long thereafter, the same investigators turned their attention to another Mon–Khmer language, Khu’ (Abramson, Nye, & Luangthongkum, 2007). The dialect chosen was Khu’ Rawk, the home language of about 3,200 people in the province of Nan in northern Thailand. The specific variety was that of the village Huay Sataeng. Khu’ was of special interest, because other investigators have worked on a northern dialect that has clearly undergone tonogenesis, while still others have neither tones nor registers (Gandour, Gårding, & Lindell, 1978; Svantesson & House, 2006). Here too the work included acoustic analysis and perceptual experiments. For the statistical treatment of the latter, the responses of the 32 native-speaker participants were handled not only as a whole but also in subgroups by sex and two age groups, young and old. Only the women showed a difference in harmonic ratios, implying involvement of differences in glottal open quotients for breathy phonation. The men, presumably less linguistically conservative, showed none. As for age, there were not enough old speakers available to get possibly significant results over two generations. The single robust differentiating property in this study is pitch. Although Khu’ Rawk is in the subgroup of the regional dialects of Khu’ known to have given rise to distinctive voice registers, at least the variety of it spoken in Huay Sataeng seems to be in a late transitional stage. With a breakdown in the old register complexes and the increased dependence on pitch, it is surely on the verge of tonogenesis.

A study now in progress (Abramson, Tiede, & Luangthongkum, in preparation) looks at the voice registers of Mon, another Mon–Khmer language. The dialect is that of the village Nakhonchum, Ratchaburi Province, Thailand. Special processing of the recordings of the four speakers has been done in order to have acoustic signals that are close approximations to the equivalent glottal source waveform without the effects of vocal-tract resonances (formants). To accomplish this for another Mon–Khmer language, Chong, inverse filtering was used (Edmondson, 1996). Another new feature is the electroglottographic analysis of three of the four speakers and one other speaker; a small sample is to be seen in Figure 21.2. For this language at least, see Table 21.1 to see a set of minimal word pairs under analysis. It is hoped to have a paper ready for submission fairly soon. It is to be followed by a study of Mon perception.
Table 21.1 List of Mon word pairs spoken four or five times each by four male native speakers

<table>
<thead>
<tr>
<th>Register 1: Modal</th>
<th></th>
<th>Register 2: Breathy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Transcription</strong></td>
<td><strong>Gloss</strong></td>
<td><strong>Transcription</strong></td>
</tr>
<tr>
<td>hi</td>
<td>river</td>
<td>hi</td>
</tr>
<tr>
<td>klan</td>
<td>lick</td>
<td>klan</td>
</tr>
<tr>
<td>sak</td>
<td>rich</td>
<td>sak</td>
</tr>
<tr>
<td>boy</td>
<td>give birth</td>
<td>boy</td>
</tr>
<tr>
<td>ato</td>
<td>ear</td>
<td>ato</td>
</tr>
<tr>
<td>cut</td>
<td>put on</td>
<td>cut</td>
</tr>
</tbody>
</table>

Conclusion

For some years now, growing attention has been given to the experimental phonetic study of the production and perception of speech in Southeast Asian languages. Here the focus is on the categories of phonologically relevant tones and voice registers found in many of the languages of the region. The role of registers in diachronic shifts toward tones is discussed in connection with the fuzzy boundary between the two typological categories. Brief accounts of a number of studies are given. There remains much more to be done.

Acknowledgment

This work was supported by NIH Grant DC-01995 to Haskins Laboratories. The author is also grateful to his Thai colleagues, especially Professors Theraphan Luangthongkum and M. R. Kalaya Tingsabadh, for long years of enriching colleagueship and collaboration as well as warm friendship.
References


