

Voice Register in Mon: Acoustics and Electroglottography

Arthur S. Abramson^{a, b} Mark K. Tiede^a
Theraphan Luangthongkum^c

^aHaskins Laboratories, New Haven, Conn., and ^bDepartment of Linguistics, University of Connecticut, Storrs, Conn., USA; ^cDepartment of Linguistics, Faculty of Arts, Chulalongkorn University, Bangkok, Thailand

Abstract

Mon is spoken in villages in Thailand and Myanmar. The dialect of Ban Nakhonchum, Thailand, has 2 voice registers, modal and breathy; these phonation types, along with other phonetic properties, distinguish minimal pairs. Four native speakers of this dialect recorded repetitions of 14 randomized words (7 minimal pairs) for acoustic analysis. We used a subset of these pairs in a listening test to verify the perceptual robustness of the register distinction. Acoustic analysis found significant differences in noise component, spectral slope and fundamental frequency. In a subsequent session 4 speakers were also recorded using electroglottography, which showed systematic differences in the contact quotient. The salience of these properties in maintaining the register distinction is discussed in the context of possible tonogenesis for this language.

© 2015 S. Karger AG, Basel

Introduction

Mon (ISO 639-3 code mnw) is spoken in many scattered villages in Thailand and Myanmar. The total number of speakers is 851,000, with 108,000 in central and western Thailand and 743,000 in southern Myanmar (www.ethnologue.com). Along with its single sister language, Nyah Kur, Mon belongs to the Monic branch of the Mon-Khmer family, which itself is a member of the Austroasiatic family. Other names for the Mon people are Talaing, Taleng, Peguan and Raman.

The dialect of Mon investigated is that of Ban Nakhonchum, Ban Pong District, Ratchaburi Province, Thailand. An inventory of its phonemes is given in the Appendix.

The Concept of Voice Register

Although earlier scholars (Blagden, 1910) were aware of distinctive phonation types, the concept of voice register was introduced by Eugénie Henderson (1952, 1965), who needed a term to cover a complex of laryngeal and supralaryngeal properties

functioning together to minimally distinguish utterances. Such a complex might include some or all of these phonetic properties: phonation type (voice quality), vowel quality, vowel length, pitch and perhaps others. How well a set of such properties might be perceptually integrated is an active research topic (Brunelle, 2012). Often it appears that phonation type is the dominant property; thus, it is that such terms as ‘clear’, ‘breathy’, ‘tense’ and ‘creaky’ have come into widespread use as labels for the registers. Indeed, in Southeast Asia, an area dominated by tone languages, it is sometimes difficult to establish a sharp boundary between the latter and voice register languages (Abramson and Luangthongkum, 2009). In some tone languages (e.g. Vietnamese) one or more of the tones may have a pitch contour coupled with a particular phonation type, while in voice register languages pitch may be an important auditory cue even in the presence of a phonation type or other phonetic properties. The term ‘voice quality’ is thus very hard to define rigorously. It is clearly a psychoacoustic concept that involves the speaker’s phonation types and the listener’s perception of them (Kreiman and Sidtis, 2011, pp. 1–24). Such properties as voice quality and pitch may also vary during the course of a syllable. If one feature is primary and under the active control of the speaker, other features that emerge during analysis can in some cases still be related to the primary one in the sense that they are automatic consequences of the primary one, although in other cases, such co-occurrences are likely to have correlational rather than causal relationships.

Linguistic and Auditory-Phonetic Investigations of Voice Register

Most register languages have 2 registers, although as many as 4 are attested in 2 languages, one being Chong, in the Mon-Khmer group (Thongkum, 1991; DiCanio, 2009), the other !Xóǝ (Traill, 1985). The historical development of voice registers in the Mon-Khmer family of languages is clearly described by Ferlus (1980; see also Diffloth, 1982). The major factor in their emergence is generally taken to be the gradual devoicing of initial obstruents. Because this explanation is so similar to descriptions of tonogenesis (Haudricourt, 1954; Matisoff, 1973), controversy over the importance of register in the origin of tones is not surprising (Thurgood, 2007). The possibility of an incipient tonal system emerging from the registers of the variety of Mon studied here is considered below. Given the question of whether the auditory property of pitch contributes to the register distinction, as well as its role as a possible harbinger of tonogenesis, in addition to examining F0 derived indirectly from the acoustic signal, we have also made use of electroglottography (EGG), a more accurate method of measuring glottal cycle patterns.

Early treatments of linguistically distinctive voice registers (Blagden, 1910, p. 479; Halliday, 1922; Henderson, 1952; Huffman, 1976; Ferlus, 1980; Gregerson and Smith, 1973) were impressionistic. Experienced field linguists worked closely and intensively with native speakers, listening repeatedly and mimicking the utterances. They produced useful descriptions, but they had no way of assessing quantitatively for production and perception the relative power of each putative property of a register complex. Such phonetic properties as voice quality, loudness, pitch, vowel quality and vowel length were found to participate in distinguishing registers from each other. Thus, it was that linguists conceived of registers as complexes, even though one property, voice quality, tended to be dominant in their descriptions. The labels ‘clear’ (modal), ‘breathy’, ‘creaky’, ‘tense’ and others were used for voice quality. A survey by John Laver (1980) provides auditory, physiological and acoustic rationales for them.

Of course, aside from their linguistic role in phonological voice registers, voice qualities (phonation types) have long been a matter of concern in speech pathology (Murry, 1982).

Instrumental-Phonetic Investigations of Voice Register

Speech scientists, availing themselves of modern instrumentation for studying the acoustics, physiology and aerodynamics of speech, eventually turned their attention to variation in phonation types (Hollien, 1974). They were not typically concerned with phonologically distinctive functions. Experimental phoneticians based in the discipline of linguistics, however, did begin to do instrumental research on phonologically relevant phonation types and, more generally, voice registers (Fischer-Jørgensen, 1967; Huffman, 1987; Ladefoged et al., 1988; Watkins, 1997; Brunelle, 2005; Keating et al., 2010).

In recent years a number of studies (Gordon and Ladefoged, 2001; Wayland and Jongman, 2003) have provided good descriptions of techniques of acoustic analysis relevant to the differentiation of phonation types. Others have done research on factors of physiological control by means of EGG (DiCanio, 2009) and the functioning of various valves of the throat, such as retraction of the tongue and epiglottis, in voice registers by means of fiberoptic laryngoscopy (Edmondson and Esling, 2006).

Of special interest to this paper are existing acoustic studies of Mon-Khmer languages (Gårding and Lindell, 1977; Thongkum, 1987, 1988, 1989, 1990, 1991; Abramson et al., 2004, 2007; Edmondson, 2009). There have been two such studies specifically on Mon (Lee, 1983; Thongkum, 1987). Our earlier studies on two other Mon-Khmer languages, Suai (Abramson et al., 2004) and Khmu' (Abramson et al., 2007) used acoustic analysis to seek relevant dimensions for the register distinction together with speech synthesis for testing the perceptual power of those dimensions. For Suai the general finding was that this dialect is in a state of flux with regard to the voice registers. There is much variation among speakers in how well it is maintained. The major differentiating properties are the fundamental frequency (F0) early in the syllable, which is higher for the modal register, and the open quotient of the glottal cycle, i.e. phonation type. All this may indicate either a weakening of the distinction or the onset of tonogenesis. As for Khmu', only one property, F0, was found to differentiate the two registers both acoustically and perceptually, with F0 higher for the modal register. Thus, this dialect of Khmu', Rawk, may have actually moved into a state of tonogenesis.

Motivation for the Current Study

The main aim of the present study is to use EGG to explore laryngeal behavior underlying the voice register distinction in the Mon-Khmer language group. For this, we have turned to the Mon language because its phonologically distinctive voice registers seem less affected by tonogenesis than those of the closely related Khmer group, and should thus provide a clearer testbed for exploring the voice register contrast itself. Mon's register 1 is called 'clear' or 'modal,' while register 2 is called 'breathy' (Thongkum, 1990). The present work represents, to the best of our knowledge, the first application of EGG to the investigation of register differences in the Mon language.

The study comprised two recording sessions of Mon utterances, one with audio alone, and the other obtained 4 months later together with simultaneous EGG. A

Table 1. Nakhonchum speakers

Speaker	Age, years	Sex	Repetitions	Sessions
B	59	M	9	both
N	71	M	9	both
D	64	M	4	1st (acoustic only)
S	63	M	9	both
P	64	F	5	2nd (with EGG)

There were two recording sessions, the first acoustic only and the second with concurrent EGG recording.

separate perception experiment was also conducted. There are thus three separate but related analyses of these data:

- 1 Acoustic analysis of both sets of production data
- 2 EGG analysis of data from the second recording session
- 3 Perceptual validation of the registers' efficacy for distinguishing these utterances assessed through listening tests by native speakers.

Collection of Acoustic and EGG Recordings

Four men and 1 woman of the village of Nakhonchum served as our speakers of Mon. They are listed in table 1 with their ages and the number of tokens of each word recorded. They were recommended to us as good and authentic speakers of the local dialect. Given the much more extensive contact of younger members of the community with the Thai language, we preferred to work with somewhat older speakers. Even they were bilingual in Mon and Thai. Indeed, we communicated with them in Thai.

Recordings were made in a quiet room in the local Buddhist monastery on 2 occasions. In the first session, no EGG data were collected. On a separate occasion 3 of the original speakers and 1 more were recorded with simultaneous EGG acquisition. In both sessions 14 minimal pairs contrasting breathy and modal phonation were recorded at a sampling rate of 44,100 Hz with a directional microphone placed about 10 cm from the mouth. The microphone was attached to headphones and thus stayed at the same place relative to the mouth even with head movements. The utterances were recorded by means of the Cool Edit Pro program onto a laptop computer, and included words with monophthongs and diphthongs. They were recorded not as minimal pairs but in a randomized sequence. The words, all highly frequent in vernacular Mon, were elicited by prompting the speakers with previously rehearsed Thai glosses one word at a time, which they then produced the Mon word for, in isolation.

For the EGG session we used a portable electroglottograph made by Laryngograph[®] Ltd. of London. Such a device works by passing a weak electric current between 2 electrodes placed on the neck. Our equipment included electrodes of 3 sizes to make it easy to choose a suitable pair for each person. After the front of the neck had been

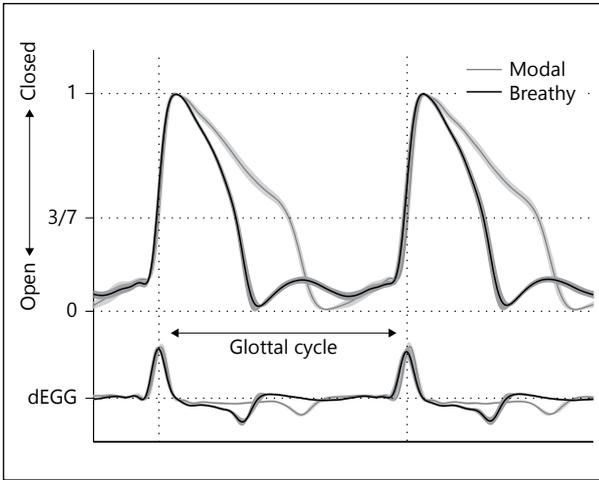


Fig. 1. Five averaged cycles of EGG following linear time and amplitude normalization, contrasting modal /bi/ and breathy /b̥i/ for speaker S, with corresponding derivatives (dEGG) underneath. Note the shorter closed portion of the cycle associated with breathy production.

Table 2. Mon word minimal pairs contrasting modal and breathy registers

Register 1 (modal)		Register 2 (breathy)	
transcription	gloss	transcription	gloss
bi	river	b̥i	you (impolite)
klan	lick	kl̥an	naughty
sak	rich	s̥ak	rancid
b̥oŋ	give birth	b̥oŋ	hookah
əto	ear	ət̥o	jujube
cut	put on	c̥ut	bone
p̥əŋ	rice	p̥əŋ	stomach

swabbed with alcohol and coated with an electrically conductive gel, the 2 electrodes were placed on the neck on either side of the thyroid cartilage. 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 the glottal pulse is seen rising to a peak. As the contact between the folds begins to decrease, less current flows and the wave shape of the glottal pulse changes in a downward movement until the glottis is open. The signal then remains low during the open portion of the cycle until the rise begins for the next closing movement of the folds. Figure 1 illustrates this cycle.

For analysis we chose the 7 minimal pairs with monophthongal vowels shown in table 2. The reason for this restriction was our concern that differences in vowel quality might be one of the factors serving to differentiate the registers, and such effects would be easier to distinguish in steady-state monophthongs than over the large frequency shifts found in diphthongs. For the single disyllabic pair we measured only the register-bearing second syllable.

Perceptual Validation Tests

Because our speakers are bilingual in Mon and Thai, a tone language, it is important to verify how well the register distinction is still maintained in vernacular usage. To assess the robustness of the register distinction in this dialect of Mon, we conducted forced-choice perceptual tests to determine how well listeners could distinguish between the registers. Utterances of the first 2 word pairs in table 2 by our 4 male native speakers were chosen for this experiment. The choice was arbitrary, and any 2 pairs would have sufficed, but the limitation was necessary to minimize the time needed for the test.

The productions of each speaker and each word pair were tested in separate listening blocks. Within a given block, 4 different modal productions and 4 different breathy productions from the same speaker and word pair were presented in randomized order. The listening test participants heard 2 presentations of a given stimulus before being prompted to respond to an image graphic displayed on a computer screen, which showed the 2 possible meanings written in Thai numbered 1 and 2. Thus, altogether, each participant gave a total of 64 responses (4 speakers \times 4 productions \times 2 word pairs \times 2 registers). The participants were instructed to give oral responses in Thai of '1' or '2' depending on how they interpreted what they heard. Thai was preferred over Mon for the responses as our research assistants recording them had little or no experience with the Mon language. The tester entered the responses into that person's test file. Entering the response caused the next stimulus to be presented. Each test was preceded by a short practice session to help the listener accommodate to the voice of the speaker. The tests were administered under control of the DMDX program of the University of Arizona (Forster and Forster, 2003).

We chose 36 listeners, bilingual in Mon and Thai, to participate in our listening tests. They were all native speakers of the Nakhonchum dialect of Mon whose parents were also native speakers. To control for possible effects of sex and age we recruited 18 men and 18 women. In each of these sex-based groups there were 9 participants in the relatively young group within the age range of 30–45 years and 9 in the relatively old group within the range of 60–75 years.

The perceptual validation tests were administered by 4 research assistants sitting with their laptop computers at widely separated stations in the dining hall of the local Buddhist monastery. At each station there were 2 pairs of full-sized ear covering headphones, one for the tester and the other for the current participant. The first and third authors had tried them in a practice run and found the attenuation of extraneous sound to be satisfactory. Two additional assistants registered the participants as they arrived, provided them with an orientation to the project, and paid them as they left. Every test was preceded by a brief practice session. Instructions were given in Thai. They are rendered in English here:

We are doing research on how certain sounds of your language are made and heard. There will be 4 listening tests. In each of them you are going to hear 2 Mon words meaning (1) 'river' and (2) 'you' spoken by a man in one part of the test and 2 words meaning (1) 'lick' and (2) 'naughty' spoken by the same man in the other part of the test. The utterances are randomized¹. Each time you hear a word, it

¹ This somewhat technical term in English was rendered in Thai by a commonplace colloquial expression.

will be said twice by the same person. Listen carefully. The 2 possible meanings numbered 1 and 2 will appear on the screen of the computer. If you understand the word to have the first meaning, please say 'one'. If you understand the word to have the second meaning, please say 'two'.

Analysis

We subjected recordings of the words with monophthongal vowels to acoustic and electroglottographic analyses to quantify their phonetic properties. Acoustic measures were obtained using the freely available UCLA-developed VoiceSauce toolbox (Shue et al., 2009). EGG measurements were obtained using Matlab (Mathworks) loosely following the methods of Mooshammer (2010), and through the use of the UCLA EGGWorks program (Tehrani, 2009). Statistical analysis of these measurements was performed in R (R Core Team, 2014) in order to evaluate the relevance of their properties to the register distinction.

Acoustic Analysis

A number of dimensions of the speech signal that may serve to differentiate the Mon voice registers were analyzed using VoiceSauce, including F0, formant frequencies, bandwidths and amplitudes (e.g. F1, B1, A1); harmonic amplitudes (e.g. H1) and amplitude differences characteristic of spectral tilt (e.g. H2–H4). Because the nonmodal register of Mon has been said to be 'breathy', it is also important to examine utterances in that register for the presence of turbulence. For this purpose VoiceSauce implements de Krom's algorithm (1993) to obtain the amplitude ratio of harmonics to noise (HNR). A related measure of cepstral peak prominence (Hillenbrand et al., 1994) is obtained from cepstral peak amplitude normalized by overall cepstral energy. For both measures, higher values are expected for the harmonically more coherent modal register.

As for spectral tilt, given the shorter closed quotients for the breathy register, a steeper fall in the amplitude of its spectral slope from low to high frequencies is expected. To measure for this effect, it is necessary to be sure that the vowels of the contrasting word pairs do not have significantly different formant frequencies. That is, although we chose word pairs with allegedly the same vowel and differentiated solely by voice register, it could be the case that there is a systematic difference between the registers in formant pattern for all the vowels or at least some of them. To our ears this was not suggested by audible differences in vowel quality; nevertheless, such a qualitative difference could be one of the bundle of phonetic properties distinguishing the registers.

It is therefore important to use a method that can correct for any differences in vocalic formant frequencies between the two registers. VoiceSauce accomplishes this through its use of the STRAIGHT algorithm (Kawahari et al., 1998) to estimate F0 and the SNACK toolkit (Sjölander, 2004) to find the frequencies and bandwidths of the formants. The harmonics nearest these formant frequencies are found, and their amplitudes are taken as the amplitudes of the formants. Finally, the formant frequencies and bandwidths are used in an algorithm that corrects harmonic amplitudes for the effects of the formants by applying the extension of Iseli et al. (2007) to Hanson's (1997) method. This correction, indicated as an asterisk appended to a particular measure as in, e.g., H1*, avoids inverse filtering. Instead, a formula is derived to 'undo' the effects of the formants on the magnitudes of the source spectrum, with corrections applied

to both harmonic and formant amplitudes (Iseli et al., 2007, p. 2285). The frequency values of the formants themselves, however, are not adjusted by this procedure. Since we have included the identity of each word as a random intercept in the model (nested within each speaker), any systematic differences in formant values across the registers are likely to be valid and may reflect part of the complex of phonetic properties distinguishing the 2 registers, as discussed in the Introduction.

The acoustic measures were evaluated by VoiceSauce over the voiced interval of each word at 1-ms intervals. With the exceptions of vowel duration, peak and average root mean square (RMS), which were measured separately from the thresholded RMS envelope, a single average value was computed for each VoiceSauce measurement type over the central third of the voiced interval to consolidate them for statistical comparison.

EGG Analysis

Glottal cycles within the EGG signal were delimited by identifying maxima within its derivative exceeding 40% of the peak value, associated in our data with the rapid rise in current flow due to the glottal closing gesture. Because the more gradual opening gesture did not provide a reliable cue within the derivative, we followed Heinrich et al. (2004) and Mooshammer (2010) in defining the end of the closed phase as a threshold (3/7 of the normalized range) applied to the EGG signal, normalized in amplitude over the cycle. Similarly, the closing and opening glottal gestures were delimited by thresholding the cycle at 10 and 90% of normalized amplitude (see fig. 2 for an example).

We follow previous research applying EGG to voice quality (Mooshammer, 2010) in obtaining the following measurements from each delimited cycle:

- $F0_{\text{EGG}}$ – fundamental frequency (1/glottal cycle)
- CQ – closure quotient (closed phase/glottal cycle)
- CQH – closure quotient (hybrid method) obtained using UCLA EGGWorks
- SQ – speed quotient (opening gesture/closing gesture)
- SLP_{CLOSE} – closing gesture slope ($\text{amplitude}_{\text{CLOSE}}/\text{duration}_{\text{CLOSE}}$)
- SLP_{OPEN} – opening gesture slope ($\text{amplitude}_{\text{OPEN}}/\text{duration}_{\text{OPEN}}$)

As with the acoustic measurements, these values were then averaged over the central third of the voiced interval to obtain a single value for each word instance.

Results

Acoustical Results

We performed statistical tests of how well the acoustic measures distinguished the register contrast using linear mixed models in R (Pinheiro and Bates, 2000), in which register (modal vs. breathy) provided the fixed contrast effect. Words nested within speakers were modeled as random effects, and byspeaker slopes and intercepts were included². Outliers with residuals exceeding 3 standard deviations (SD) were excluded from the analysis. A measure of effect size was obtained by subtracting corresponding register values, grouped by word and speaker and averaged across 5 repetitions (modal – breathy), with the differences then pooled across words and speakers. Bonferroni-corrected results are provided in table 3 (part a shows results from the acoustics-only session, part b from the EGG session).

² For example, $\text{lme}(F0 \sim \text{REG}, \text{random} = \sim \text{REG}|\text{SPKR}/\text{WORD}, \text{data})$.

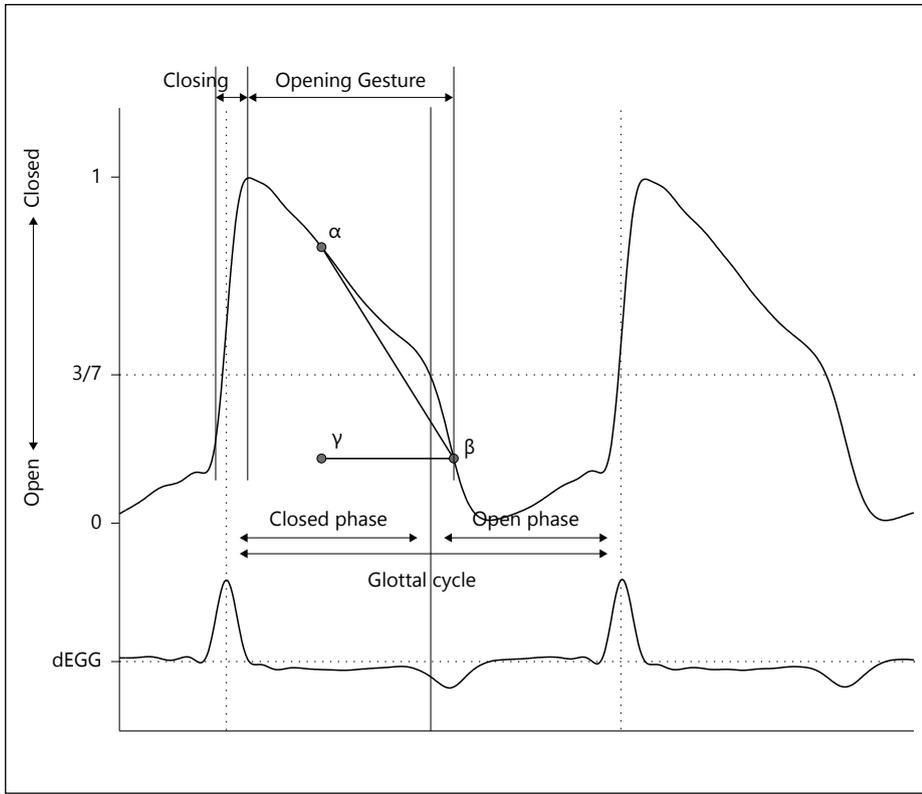


Fig. 2. EGG measurements exemplified by modal /bi/, speaker S. The onset of each glottal cycle was determined by the steep rise in current flow associated with the derivative peak. Glottal opening was linked to 3/7 of the normalized cycle, separating the glottal cycle into closed and open phases. The durations of the closing and opening gestures were determined by thresholding the normalized cycle at 10% (β) and 90% (α) of peak closing amplitude, with corresponding slopes determined by rise ($\alpha-\beta$) over duration ($\beta-\gamma$).

Overall Amplitude

The literature in general, including our own work on Suai and Khmu' (Abramson et al., 2004, 2007), does not show overall amplitude of the syllable to be relevant to the distinction between modal and breathy registers. Here too, for both the average RMS (AVGRMS) and the peak RMS (PKRMS) table 3 shows no significant differences.

Fundamental Frequency

F0 was observed to be higher for the modal register in both recording sessions, but the contrast under Bonferroni correction does not reach significance. To estimate the magnitude of this effect, the mean difference between breathy and modal productions was computed, averaged across repetitions paired by register for each word and speaker – modal productions were higher than breathy ones in both sessions, but the differences are small: acoustic session mean 9.0 Hz (SD 5.9); EGG session mean 5.4 Hz (SD 3.4).

Table 3. The acoustic measurement sessions**a** Acoustic-only session

	M-B	Coeff.	t	p	Sig.	R ²	Dir.
Duration	-12.03	-12.98	-1.38	0.168	n.s.	0.83	
AVGRMS	0.69	1.73	1.15	0.252	n.s.	0.59	
PKRMS	0.02	0.02	1.77	0.078	n.s.	0.56	
F0	9.05	8.99	2.17	0.032	n.s.	0.90	
F1	60.38	59.22	6.66	0.000	*	0.97	M > B
B1	-34.75	-35.29	-1.98	0.049	n.s.	0.79	
A1	0.18	0.19	0.29	0.769	n.s.	0.80	
F2	37.28	43.50	1.75	0.082	n.s.	0.91	
B2	-22.61	-13.49	-1.32	0.189	n.s.	0.73	
A2	-0.36	-0.15	-0.23	0.822	n.s.	0.78	
F3	157.23	157.23	2.07	0.040	n.s.	0.45	
B3	17.16	4.28	0.07	0.947	n.s.	0.51	
A3	0.68	0.90	0.55	0.586	n.s.	0.61	
H1	-4.78	-4.95	-4.33	0.000	*	0.63	M < B
H2	-0.48	-0.71	-1.48	0.139	n.s.	0.45	
H4	-0.42	-0.31	-0.45	0.651	n.s.	0.65	
H1*-A1*	-4.94	-4.94	-5.17	0.000	*	0.76	M < B
H1*-A2*	-4.42	-4.56	-3.16	0.002	*	0.73	M < B
H1*-A3*	-5.46	-5.27	-2.54	0.012	n.s.	0.56	
H1*-H2*	-4.30	-4.23	-3.68	0.000	*	0.88	M < B
H2*-H4*	-0.06	-0.07	-0.12	0.908	n.s.	0.68	
HNR35	6.15	6.05	8.80	0.000	*	0.87	M > B
CPP	2.88	2.93	5.32	0.000	*	0.84	M > B

b EGG session

	M-B	Coeff.	t	p	Sig.	R ²	Dir.
Duration	-24.30	-20.79	-1.08	0.281	n.s.	0.96	
AVGRMS	-0.44	-0.45	-0.56	0.577	n.s.	0.97	
PKRMS	0.00	0.00	0.03	0.977	n.s.	0.95	
F0	5.36	5.21	2.99	0.003	n.s.	0.96	
F1	41.92	34.80	1.77	0.079	n.s.	0.96	
B1	-20.39	-14.80	-0.88	0.382	n.s.	0.87	
A1	0.96	1.03	1.22	0.223	n.s.	0.93	
F2	-97.18	-111.27	-1.20	0.230	n.s.	0.97	
B2	-47.44	-49.95	-2.07	0.040	n.s.	0.74	
A2	2.46	2.49	3.24	0.001	*	0.91	M > B
F3	98.62	88.90	1.41	0.159	n.s.	0.74	
B3	22.19	9.01	0.22	0.827	n.s.	0.59	
A3	3.97	3.99	1.68	0.094	n.s.	0.83	
H1	-5.02	-5.03	-10.37	0.000	*	0.97	M < B
H2	0.27	0.40	0.26	0.792	n.s.	0.93	
H4	0.70	0.90	1.01	0.313	n.s.	0.89	
H1*-A1*	-5.94	-6.00	-7.34	0.000	*	0.92	M < B

Table 3. (continued)

	M-B	Coeff.	t	p	Sig.	R ²	Dir.
H1*-A2*	-7.47	-7.54	-10.37	0.000	*	0.90	M < B
H1*-A3*	-8.99	-9.02	-4.03	0.000	*	0.84	M < B
H1*-H2*	-5.29	-5.31	-3.68	0.000	*	0.91	M < B
H2*-H4*	-0.43	-0.49	-0.34	0.732	n.s.	0.83	
HNR35	0.84	1.40	1.64	0.103	n.s.	0.93	
CPP	1.87	2.03	2.90	0.004	n.s.	0.80	

AVGRMS and PKRMS = Average and peak RMS; HNR35 = harmonics-to-noise ratio band-limited to 35 kHz; CPP = cepstral peak prominence. Statistical tests of how well acoustic measurements distinguished the register contrast using a linear mixed model in which register (M = modal vs. B = breathy) provided the fixed contrast effect. Words nested within speakers were modeled as random effects, and byspeaker slopes and intercepts were included. The direction (dir.) of the contrast is the modal result relative to the breathy baseline. Shaded rows show consistent results across the two data collection sessions. Significance was evaluated using a Bonferroni-corrected level of $p = 0.05/23 = 0.0022$, with significant results shown in bold. R² was estimated as the correlation between the model fitted and the observed values. An estimate of effect size (M-B) is given by the average difference between corresponding register types pooled by repetitions across words and speakers.

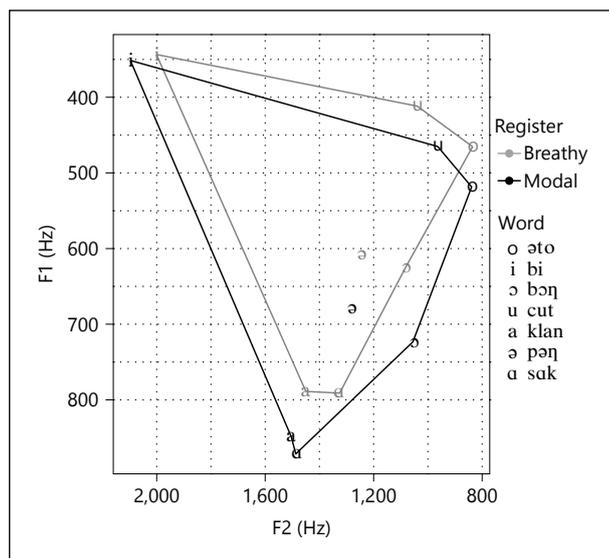


Fig. 3. Mon vowels plotting the first × second formant by register averaged across speakers and productions: breathy, modal.

Formant Frequencies and Bandwidths

Although table 3 shows no systematic effects of formant frequency and bandwidth between registers, the contrasting vowel space plots of F1 × F2 shown in figure 3 suggest that a vowel quality difference might be associated with the register distinction. The mean Euclidean distance between corresponding modal and breathy productions, computed in mels on values averaged across repetitions by word and speaker,

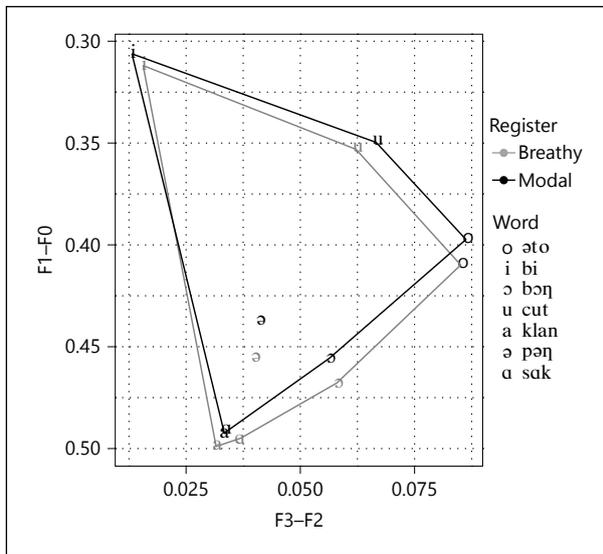


Fig. 4. Vowel spaces from the contrasting registers averaged across speakers and productions represented with the Syrdal and Gopal (1986) bark difference transformation: breathy, modal.

was 94 Hz (SD 32) for the acoustics-only session and 135 Hz (SD 75) for the EGG session.

However, a different interpretation is possible. In work intended to support a perceptual basis for vowel quality, Syrdal and Gopal (1986) developed a transformation employing bark differences on F0 and the first 3 formants. Figure 4 shows this transform applied to the same data averaged across words and speakers presented in figure 3, from which it is apparent that under the transform the vowel spaces overlap between the registers to a greater degree.

As a means of quantifying overlap, we followed Lane et al. (2001) in computing the average vowel space as the mean distance between all vowel pairs, contrasting the $F1 \times F2$ measures alone with the transformed bark difference measures. In addition, the area of each vowel space polygon (as illustrated by the peripheral vowels plotted in fig. 3 and 4) overlapping between the two registers was scaled by the total area of both polygons. Results are shown in table 4. The highest degree of overlap is found using the transformed bark difference measures. This suggests that within this transformed perceptual space the shift in F1 and F2 is offset to some degree by counterbalancing trends in F0 and F3, and thus that vowel quality is not an important difference distinguishing the registers.

Spectral Tilt

A register-related spectral slope difference is supported in table 3 by the amplitude ratio of the first two harmonics ($H1^*-H2^*$), and the amplitude ratios of the first harmonic with the first and second formants ($H1^*-A1$, $H1^*-A2$). In each of these measures, as expected, the breathy values are higher for both recording sessions; that is the spectral slope falls off faster in the breathy register. The relative amplitude of the first harmonic H1 was also systematically higher for the breathy register in both recording sessions.

Table 4. Average vowel space and percentage of the common area of each register's vowel space polygon

	Average vowel space			Vowel space overlap
	breathy	modal	breathy/modal	% overlap/total
F1 × F2 (bark)	2.91	3.07	0.94	0.65
F3–F2 × F1–F0 (bark)	2.95	3.11	0.95	0.72

(1) The average vowel space was computed between all vowel pairs for each register, showing the scaled (breathy/modal) quotient for F1 versus F2 in barks and the normalization of Syrdal and Gopal (1986). (2) The percentage of the common area of each register's vowel space polygon was scaled by the total area of both polygons. Note that the degree of overlap is highest using the transformed bark difference measures incorporating F0 and F3.

Noise Component

The ratio of the energy in the harmonics to the noise floor is shown in table 3 as HNR35 (band-limited to 3,500 Hz³). Given the turbulent airflow associated with breathy production, HNR is expected to be smaller for that condition. While results confirm smaller HNR for breathy production in both sessions, the difference reaches significance only for the acoustics-only session. Similarly cepstral peak prominence in table 3, another measure of relative noise, shows greater amplitude for the (less noisy) modal register in both sessions; however, once again the effect reaches significance only in the acoustics-only session. It is possible that the presence of EGG electrodes on both sides of the thyroid cartilage may have interfered with this aspect of the register distinction, either directly, or through some type of speaker compensation for their presence.

EGG Results

Paralleling the acoustic analysis we performed statistical tests of how well the EGG measures distinguished the register contrast using the same linear mixed model, in which register (modal vs. breathy) provided the fixed contrast effect. Words nested within speakers were modeled as random effects, and byspeaker slopes and intercepts were included. Outliers with residuals exceeding 3 SDs were excluded from the analysis. A measure of effect size was obtained by subtracting corresponding register values, grouped by word and speaker and averaged across 5 repetitions (modal – breathy), with the differences then pooled across words and speakers. Results are provided in table 5.

Fundamental Frequency of EGG

The F0 values obtained from EGG represent a direct measurement of glottal cycle counts, as opposed to deriving them from the composite waveform, and as such can be taken as being more accurate. The direction of the results (modal > breathy) is significant and consistent with those obtained acoustically. However, the effect size, estimated as before as the average difference between corresponding register types (modal – breathy) pooled across words and speakers, is just 5.2 Hz (SD 3.6).

³ As observed by de Krom (1993), the influence of jitter on harmonic structure increases with harmonic number. The VoiceSauce implementation of his HNR algorithm is band-limited to 3,500 Hz, as at higher frequencies jitter cannot be effectively distinguished from noise.

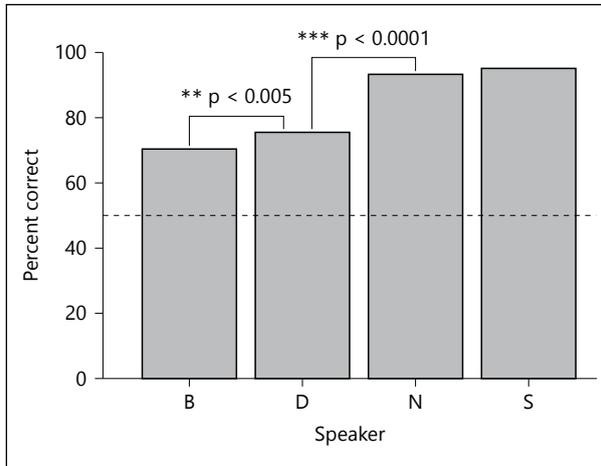


Fig. 5. Identification rates for the utterances of each speaker ($n = 1,152$). Interspeaker differences were evaluated using a generalized linear mixed model in R with correct response as the dependent variable, speaker as an independent fixed effect and listener as a random effect.

Table 5. Statistical tests of how well EGG measurements distinguished the register contrast using the same linear mixed model applied to the acoustic measures

	M-B	Coeff.	t	p	Sig.	R ²	Dir.
F0 _{EGG}	5.18	5.03	2.97	0.003	*	0.96	M > B
CQ	0.12	0.12	4.19	0.000	*	0.95	M > B
CQH	0.12	0.12	4.21	0.000	*	0.96	M > B
SQ	0.68	0.71	1.57	0.119	n.s.	0.90	
SLP _{CLOSE}	-0.06	-0.04	1.55	0.123	n.s.	0.95	
SLP _{OPEN}	0.32	0.31	-0.30	0.767	n.s.	0.93	

The direction of the contrast is the modal result relative to the breathy baseline. Significance was evaluated using a Bonferroni-corrected level of $p = 0.05/6 = 0.0083$, with significant results shown in bold. R² was estimated as the correlation between the model fitted and the observed values. An estimate of effect size (M-B) is given by the average difference between corresponding register types pooled by repetitions across words and speakers.

Closure Quotient (Closed Phase/Glottal Cycle)

CQ is greater for the modal register indicating that the closed phase of the glottal cycle is longer. The results obtained separately using the hybrid method (CQH) from the UCLA EGGworks program are nearly identical to those measured directly. In both cases the results are in the expected direction (modal > breathy), and the effect is significant.

Perceptual Validation Tests

Listeners identified the test words well above chance level (3,852 out of 4,608 correct, 83.6%). Figure 5 shows that this was consistent across all 4 speakers, though at different levels of accuracy: the utterances of speakers N and S were significantly

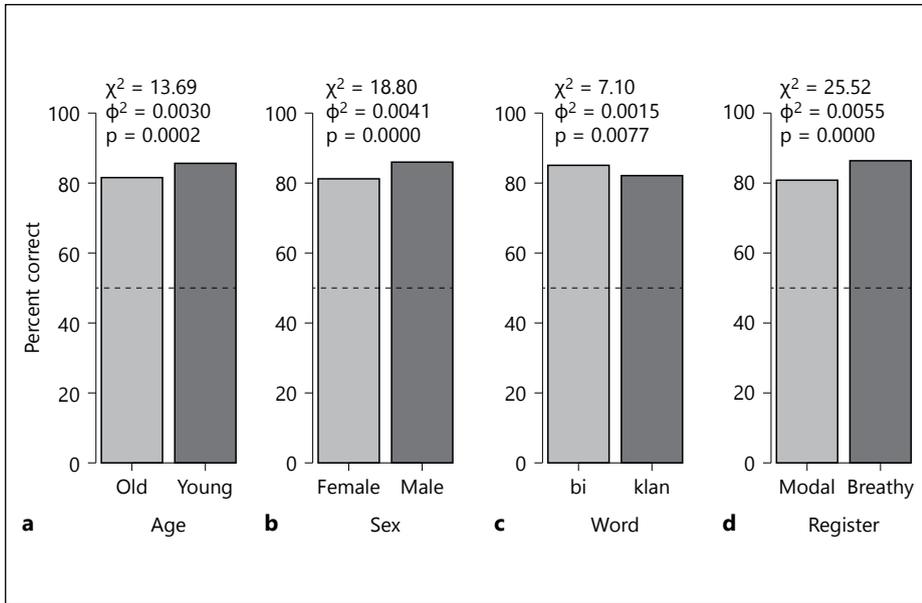


Fig. 6. Identification rates by listener participant groups (**a**, old vs. young; **b**, female vs. male), word pairs (**c**, /bi:/bi/ vs. /klan:/kɿn/) and registers (**d**, modal vs. breathy). Group differences assessed by a two-sided proportion test ($n = 2,304$).

better labeled by listeners than those of speaker D, and his in turn were better than for speaker B. This leads us to speculate that we have here a condition common to many cultures, namely varying levels of clarity in speech (Bradlow et al., 1996).

Figure 6 shows response rates for the participants grouped by age and sex, as well as for the two word pairs used and the two registers. For each such category the results assessed by two-sided proportion tests are significant, although the associated effect sizes (Φ^2) are small. The young group responded more accurately than the old group. We had expected that the older speakers with their presumably somewhat smaller exposure to Thai would be more conservative, while the younger speakers might show a weakening of the distinction. It is also possible that some of our older speakers may have suffered hearing losses. Unfortunately we were unable to obtain audiograms.

Also somewhat unexpected was the greater accuracy demonstrated by male listeners, possibly due to all 4 speakers in this test themselves being male.

Across the minimal word pairs, the /bi:/bi/ pair was identified slightly more consistently than the /klan:/kɿn/ pair, and breathy productions were identified with consistently greater accuracy than modal productions. As the initial /k/ of the /klan/ pair is devoiced, the possibility exists that the poorer discrimination observed for this pair might be due to devoicing or residual aspiration interpreted by listeners as breathy voice even within modal productions. However, in our acoustic analysis we observed no aspiration whatsoever in any /klan/ token for either register; voice onset time is effectively zero for these words.

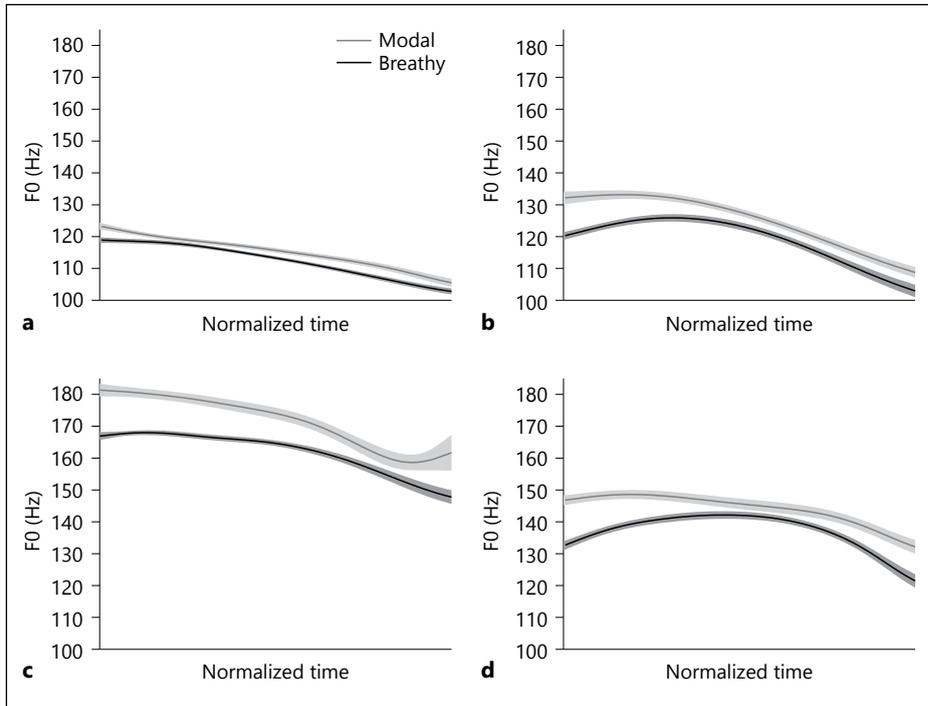


Fig. 7. Linearly normalized F0 vowel time courses determined by measured EGG cycles, averaged across all words for the 4 participants in the EGG session; shaded bars show standard error at each normalized sample offset. **a** Speaker B. **b** Speaker N. **c** Speaker P. **d** Speaker S.

Discussion

Our findings show that the two voice registers of this variety of Mon are well distinguished by the acoustic properties associated with spectral tilt, which follow from different glottal cycle patterns associated with each register. Consistent with previous investigations of the characteristics of breathy register [e.g. Fischer-Jørgensen (1967) for Gujarati, Huffman (1987) for Hmong], we found enhanced first harmonic amplitude for the breathy register. We also found a clear pattern of reduced amplitude at higher frequencies for breathy register associated with increased spectral tilt in comparisons normalized by H1, which is consistent with the modeling of breathy airflow as reported by Stevens (1998).

All 4 speakers show a significant contrast between registers in glottal cycle closure patterns (CQ). The CQ is shorter for the breathy register, reflecting greater duration of the separation of the vocal folds. The associated steeper fall in the spectral slope of the breathy register is explained by Stevens (1998, pp. 85–92) as the result of glottal leakage during closure of the vocal folds. That is, even though the membranous folds are fully closed, there can be a space between the posterior surfaces of the arytenoid cartilages through which air can flow. If the cartilages are also separated at their anterior ends, the leak is larger. In either case, the folds are adducted more gradually, less

abruptly than for the modal register. As a consequence, energy in the higher frequencies is attenuated, resulting in increased spectral tilt.

Given the current status of Mon embedded in a Thai-speaking culture in which phonemic tones are contrastive, it is also useful to examine characteristic shapes of pitch contours derived from complete productions. Accordingly, in a departure from our focus on averaged values from the middle third of the vowel of each syllable, in figure 7 we have plotted the linearly normalized time courses of the F₀ contours from contrasting registers averaged over all words for the 4 participants in the EGG session. It will be observed that the 2 on the left differ from the 2 on the right in not having a rising onset, though all 4 have a higher F₀ for the modal (clear) register. The lack of consistency in tonal contours suggests that this parameter is idiosyncratic by speaker.

Independently of contour shape however, F₀ is slightly higher in the modal register for both sessions, as measured acoustically and from glottal cycles. The question arises as to why this should be so. This effect has been noted in a number of languages, e.g. Chong (Thongkum, 1991; DiCanio, 2009), Suai (Abramson et al., 2004), Khmu' (Abramson et al., 2007) and Cham (Brunelle, 2012). A physiological explanation is offered by Wayland and Jongman (2003, p. 187): 'During the production of breathy phonation, to allow the vocal folds to vibrate while they stay relatively far apart, the vocal folds have to be relatively less taut. Thus, the fundamental frequency of a breathy vowel is expected to be lower than that of a clear vowel.' F₀, then, may be one of the constituent properties distinguishing the Mon voice registers, although perhaps not to the same degree for all speakers. However, the small differences observed (7.2 Hz averaged across sessions) contrast markedly with the range of pitch excursion encountered in true tone languages, such as the Thai dialect spoken by our participants, and thus suggest a facilitating or enhancing role (Stevens and Keyser, 2010) in maintaining the register contrast, rather than acting as a primary cue.

Given the role of F₀ in differentiating the two registers, one might wonder what the prospects for tonogenesis in Mon could be. A factor favoring such a phonological shift is the close contact between Mon as spoken in scattered villages in Thailand with the Thai language and in Myanmar with Burmese. Both of the latter are tonal languages, although Burmese, with intertwined pitch and other properties in its tonal system (Watkins, 2001) could arguably be better described as a language with both tones and registers. Linguistic interference between languages in contact takes place with the growth of bilingualism (Weinreich, 1953). For our Nakhonchum dialect of Mon there seem to be no available data for the ratio of monolingual Mon speakers to bilingual speakers over the many years of contact. Certainly in recent years, with greater attendance of Mon children in schools with Thai as the medium of instruction, essentially all the local ethnic Mons are quite fluent in Thai, which is used in school, government offices and work outside the village, although Mon continues to be the language of the village. Thus, if distinctive voice register is in fact a probable early stage in tonogenesis (Thurgood, 2007), conditions may well be ripe for its onset. On the other hand, a conservative factor working against tonogenesis could be Mon speakers' pride in their culture, and their history of an ancient Mon kingdom in Southeast Asia.

However, at this time, even with our observed F₀ differences between the registers, there is no obvious evidence of increasing dominance of F₀ in making the distinction that would indicate incipient tonogenesis in the Mon language of Nakhonchum. Consistency in the shapes of the F₀ contours would suggest learned behavior possibly associated with the onset of tonogenesis, but we instead observed varying patterns across our speakers,

which would seem to rule out the possibility of learned behavior. If the aforementioned physiological explanation of the F0 difference between the registers is correct, we can assume that it is an automatic consequence of the difference in laryngeal control for the phonation types involved rather than behavior learned by children and subsequently controlled to make the phonemic distinction. Thus, our data do not suggest the dominance of F0 over other phonetic properties serving to make the distinction.

The vowel quality differences across the registers shown in the F1 × F2 plot of figure 3 suggest that vowel quality may also be one of the distinguishing properties, though the differences are relatively small. The vowel space corresponding to breathy register is somewhat offset in the formant space from that of the modal one; however, this degree of overlap is reduced under the Syrdal and Gopal (1986) bark difference plot shown in figure 4. If in fact vowel quality is a distinguishing characteristic of register in Mon, we should expect it to maximize the difference between registers regardless of the transformation used. Because it appears that projected into a perceptual space the contrast is instead reduced, we consider this support for a general conclusion that differences in formant patterns do not play an important role in maintaining the register distinction.

Conclusion

Even though certain dialects of some Mon-Khmer languages, e.g. Khmu' (Svantesson and House, 2006), are undergoing tonogenesis, our results do not support such a finding for our dialect of Mon. Indeed, as demonstrated by the validation tests as well as the acoustic and electroglottographic data, the register distinction, as manifested in a bundle of acoustic and laryngeal properties, seems to be quite stable for this language at this time.

Acknowledgments

This work was supported by NIH grants HD-01994 and DC-02717 to Haskins Laboratories. The third author's participation was supported by a grant from the Thailand Research Fund, which also furnished the funds for the first author's travel to Thailand as well as the living expenses for his 2-month stay there to work on the gathering of data. For their excellent help in Nakhonchum with the recording of Mon speech and the running of perception tests, we are grateful to Yanin Sawanakunanon, Narinthorn Sombutnon Behr, Chommanad Intajamornrak, Nattaphon Phuengnoi, Charttreeya Churat and Theekayu Jiamjuankhao. In addition, we are indebted to Miguel Moreno for his patient and expert guidance in the construction of listening tests with the DMDX computer program. Discussions with Douglas H. Whalen, Anders Löfqvist, Gordon Ramsay, Christine Shadle, Terrance M. Nearey, Richard Lissemore, Laura Koenig and Christian DiCano have been enlightening. The helpful comments of Sarah Hawkins and 3 anonymous reviewers are gratefully acknowledged.

Appendix

The Phonemes of Nakhonchum Mon

Consonants

Initials

p t c k ? ph th ch kh b d m n ɲ ŋ ɲ
s e h w l ʃ r j

pr pl kr kl kw (~f) phr phl
The occurrence of the clusters /phr/ and /phl/ is rare.

Finals

p t c k ʔ m n ɲ ŋ h

Vowels

i e ε i ə ɜ a u o ɔ

Voice Registers

Register 1: modal (clear), e.g. /sak/ 'rich'

Register 2: breathy, e.g. /sək/ 'rancid'

References

- Abramson AS, L-Thongkum T, Nye PW (2004): Voice register in Suai (Kuai): an analysis of perceptual and acoustic data. *Phonetica* 61:147–171.
- Abramson AS, Nye PW, Luangthongkum T (2007): Voice register in Khmu': experiments in production and perception. *Phonetica* 64:80–104.
- Abramson AS, Luangthongkum T (2009): A fuzzy boundary between tone languages and voice-register languages; in Fant G, Fujisaki H, Shen J (eds): *Frontiers in Phonetics and Speech Science*. Beijing, Commercial Press, pp 149–155.
- Blagden CO (1910): Quelques notions sur la phonétique du Talian et son évolution historique. *J Asiat Dix Sér* 15:477–505.
- Bradlow AR, Torretta GM, Pisoni DB (1996): Intelligibility of normal speech. I. Global and fine-grained acoustic-phonetic talker characteristics. *Speech Commun* 20:255–272.
- Brunelle B (2005): Register and tone in Eastern Cham: evidence from a word game. *Mon-Khmer Stud* 35:121–132.
- Brunelle B (2012): Dialect experience and perceptual integrality in phonological registers: fundamental frequency, voice quality and the first formant in Cham. *J Acoust Soc Am* 131:3088–3102.
- De Krom G (1993): A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *J Speech Hear Res* 36:254–265.
- DiCanio CT (2009): The phonetics of register in Takhian Thong Chong. *J Int Phonet Assoc* 39:162–188.
- Diffloth G (1982): Registres, dévoisement, timbres vocaliques: leur histoire en Katouique. *Mon-Khmer Stud* 11:1–33.
- Edmondson JA (2009): Voice qualities and inverse filtering in Chong. *Mon-Khmer Stud* 26:107–116.
- Edmondson JA, Esling JH (2006): The valves of the throat and their functioning in tone, vocal register, and stress: laryngoscopic case studies. *Phonology* 23:157–191.
- Ferlus M (1980): Formation des registres et mutations consonantiques dans les langues Mon-Khmer. *Mon-Khmer Stud* 8:1–76.
- Fischer-Jørgensen E (1967): Phonetic analysis of breathy (murmured) vowels in Gujarati. *Indian Linguist* 28:71–139.
- Forster KL, Forster JC (2003): DMDX: a windows display program with millisecond accuracy. *Behav Res Methods Instr Comput* 35:116–124 (www.web.arizona.edu/~cul/dmdx.htm).
- Gårding E, Lindell K (1977): Tones in Northern Kammu: a phonetic investigation. *Acta Orientalia* 38:321–332.
- Gordon M, Ladefoged P (2001): Phonation types: a cross-linguistic overview. *J Phonet* 29:383–406.
- Gregerson K, Smith K (1973): The development of To'drah register. *Mon-Khmer Stud* 4:143–184.
- Halliday R (1922): *A Mon-English Dictionary*. Bangkok, Siam Society.
- Hanson H (1997): Glottal characteristics of female speakers: acoustic correlates. *J Acoust Soc Am* 101:466–481.
- Haudricourt A-G (1954): De l'origine des tons en vietnamien. *J Asiat* 242:69–82.
- Heinreich N, d'Alessandro C, Doval B, Castellengo M (2004): On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation. *J Acoust Soc Am* 115:1321–1332.
- Henderson EJA (1952): The main features of Cambodian pronunciation. *Bull School Orient Stud* 17:149–174.
- Henderson EJA (1965): The topography of certain phonetic and morphological characteristics of South East Asian languages. *Lingua* 15:400–434.
- Hillenbrand J, Cleveland RA, Erickson RL (1994): Acoustic correlates of breathy voice quality. *J Speech Hear Res* 31:769–778.
- Hollien H (1974): On vocal registers. *J Phonet* 2:125–143.
- Huffman FE (1976): The register problem in fifteen Mon-Khmer languages; in Jenner PN, Thompson LC, Starosta S (eds): *Austroasiatic Studies*. Special Publication No 13. Honolulu, University of Hawaii, Oceanic Linguistics, vol 1, pp 575–589.
- Huffman MK (1987): Measures of phonation in Hmong. *J Acoust Soc Am* 81:495–504.
- Iseli M, Shue Y-L, Alwan A (2007): Age, sex, and vowel dependencies of acoustic measures related to the voice source. *J Acoust Soc Am* 121:2283–2295.

- Kawahari H, de Cheveigné A, Patterson RD (1998): An instantaneous-frequency-based pitch extraction method for high quality speech transformation: revised TEMPO in the STRAIGHT-suite. Proc ICSLP'98, Sydney, December 1998.
- Keating P, Esposito C, Garelick M, Khan S, Kuang J (2010): Phonation contrasts across languages. *UCLA Work Papers Phonet* 108:188–202.
- Kreiman J, Sidtis D (2011): *Foundations of Voice Studies*. Chichester, Wiley-Blackwell.
- Ladefoged P, Maddieson I, Jackson M (1988): Investigating phonation types in different languages; in Fujimura H (ed): *Vocal Physiology: Voice Production, Mechanisms and Functions*. New York, Raven Press, vol 2, pp 297–317.
- Lane H, Matthies M, Perkell J, Vick J, Zandipour M (2001): The effects of changes in hearing status in cochlear implant users on the acoustic vowel space and CV coarticulation. *J Speech Lang Hear Res* 44:552–563.
- Laver J (1980): *The Phonetic Description of Voice Quality*. Cambridge, Cambridge University Press.
- Lee T (1983): An acoustic study of the register distinction in Mon. *UCLA Work Papers Phonet* 57:79–96.
- Lofqvist A, Baer T, McGarr NS, Seider Story R (1989): The cricothyroid muscle in voicing control. *J Acoust Soc Am* 85:1314–1321.
- Matisoff JA (1973): Tonogenesis in Southeast Asia; in Hyman L (ed): *Consonant Types and Tone*. Southern Calif Occas Papers in Linguistics, No 1. Los Angeles, Linguistics Program, USC, pp 71–95.
- Mooshammer C (2010): Acoustic and laryngographic measures of the laryngeal reflexes of linguistic prominence and vocal effort in German. *J Acoust Soc Am* 127:1047–1058.
- Murry T (1982): Phonation assessment; in Lass NJ, McReynolds LV, Northern JL, Yoder DE (eds): *Speech, Language, and Hearing*. Philadelphia, Saunders, vol 2: *Pathologies of speech and language*, pp 477–488.
- Pinheiro JC, Bates DM (2000): *Mixed-Effects Models in S and S-PLUS*. New York, Springer.
- R Core Team (2014): R: a language and environment for statistical computing. Vienna, R Foundation for Statistical Computing. <http://www.R-project.org/>.
- Sjölander K (2004): Snack sound toolkit. KTH Stockholm, Sweden. <http://www.speech.kth.se/snack/>.
- Shue Y-L, Keating P, Vicenik C (2009): VoiceSauce: a program for speech analysis (abstract). *J Acoust Soc Am* 126:2221.
- Stevens KN (1998): *Acoustic Phonetics*. Cambridge, MIT Press.
- Stevens KN, Keyser SJ (2010): Quantal theory, enhancement and overlap. *J Phonet* 38:10–19.
- Svantesson J-O, House D (2006): Tone production, tone perception and Kammu tonogenesis. *Phonology* 23:309–333.
- Syrdal AK, Gopal HS (1986): A perceptual model of vowel recognition based on auditory representation of American English vowels. *J Acoust Soc Am* 79:1086–1100.
- Tehrani H (2009): EGGWorks: a program for automated analysis of EGG signals. <http://www.linguistics.ucla.edu/faciliti/facilities/physiology/EggWorksSetup.exe>.
- Thongkum TL⁴ (1987): Another look at the register distinction in Mon. *UCLA Work Papers Phonet* 67:132–165.
- Thongkum TL (1988): Phonation types in Mon-Khmer languages; in Fujimura H (ed): *Vocal Fold Physiology: Voice Productions, Mechanisms and Functions*. New York, Raven Press, vol 2, pp 319–333.
- Thongkum TL (1989): An acoustical study of the register complex in Kui (Suai). *Mon-Khmer Stud* 15:1–19.
- Thongkum TL (1990): The interaction between pitch and phonation type in Mon: implications for a theory of tonogenesis. *Mon-Khmer Stud* 16–17:11–24.
- Thongkum TL (1991): An instrumental study of Chong registers; in Davidson JHCS (ed): *Austroasiatic Languages: Essays in Honor of H.L. Shorto*. London, School of Oriental and African Studies, University of London, pp 141–160.
- Thurgood G (2007): Tonogenesis revisited: reviving the model and the analysis; in Harris JG, Somsong B, Harris JE (eds): *Studies in Tai and Southeast Asian Linguistics*. Bangkok, Ekphim Thai, pp 241–262.
- Trails A (1985): *Phonetic and phonological studies of !Xóǀ Bushman*. Vol. 1. Philadelphia, John Benjamins Publishing Company.
- Watkins J (1997): Can phonation types be reliably measured from sound spectra? Some data from Wa and Burmese. *SOAS Work Papers Linguist Phonet* 7:321–339.
- Watkins J (2001): Burmese. *J IPA* 31:291–295.
- Wayland R, Jongman A (2003): Acoustic correlates of breathy and clear vowels: the case of Khmer. *J Phonet* 31:181–201.
- Weinreich U (1953): *Languages in Contact: Findings and Problems*. Publications of the Linguistic Circle of New York, No 1. New York, Linguistic Circle of New York.

⁴T.L. Thongkum and Theraphan Luangthongkum are the same person.