Original Paper

1473

Phonetica 2007;64:80–104
DOI: 10.1159/000107911

Received: February 17, 2007
Accepted: June 9, 2007

Voice Register in Khmu': Experiments in Production and Perception

Arthur S. Abramson[a, b] Patrick W. Nye[a] Theraphan Luangthongkum[c]

[a]Haskins Laboratories, New Haven, Conn., and [b]Department of Linguistics, University of Connecticut, Storrs, Conn., USA; [c]Department of Linguistics, Faculty of Arts, Chulalongkorn University, Bangkok, Thailand

Abstract

Some Khmu' dialects have phonologically distinctive voice registers. Auditory observations have claimed a stable distinction between clear voice and high pitch for Register 1 and breathy voice and low pitch for Register 2 in the Khmu' Rawk dialect of northern Thailand. Word pairs distinguished only by register were recorded by 25 native speakers. Acoustic analysis yielded F0 and overall amplitude contours, frequencies of F1 and F2 in quasi-steady states of the vowels, relative intensities of higher harmonics to that of the first harmonic, and vowel durations. When circumstances caused early attention to perception testing, the words of only 8 speakers had been analyzed for properties other than amplitude and F0. Since the only significant factor that had emerged by then was F0 contour, the synthetic stimuli were made with just a series of seven contours. The labeling by 32 native speakers yielded two categories, demonstrating the sufficiency of F0 as an acoustic cue. The completed acoustic analysis showed a significant effect of one of the harmonic ratios for the women only, suggesting a conservative bias. The language has been shifting toward tonality and may have reached it.

Introduction

Voice register as a phonological unit in a number of languages is interesting for its phonetic complexity and variability, as well as for its possible role in tonogenesis. The concept of register itself was apparently first formulated by Eugénie Henderson [1952], who described it as a complex of laryngeal and supralaryngeal properties, one of which may be dominant. Traditional, largely impressionistic descriptions of register complexes have included such characteristics as voice quality (phonation type), vowel quality, vowel length, pitch, and perhaps others. Although for the most part, such languages have two registers, there are some that have more, even as many as four [Thongkum, 1991].

Our interest in the topic has been stimulated not only by published instrumental studies [e.g., Thongkum, 1989; Gordon and Ladefoged, 2001; Wayland and Jongman,

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

0031–8388/07/0643–0080
$23.50/0

Accessible online at:
www.karger.com/pho

Prof. Arthur S. Abramson
Haskins Laboratories
270 Crown Street
New Haven, CT 06511-6610 (USA)
Tel. +1 203 865 6163, Fax +1 203 865 8963
E-Mail arthur.abramson@uconn.edu

2003; Brunelle, 2005] and its possible relevance to tonogenesis [Thongkum, 1990; Svantesson, 2001; Premsrirat, 2003; Thurgood, 2002, 2007; Abramson, 2004] but also by the apparent lack or scarcity of perceptual studies of register complexes in the public domain. Indeed, there have been perceptual studies of dialects of Khmu', but these dialects have been taken to be either tonal or nontonal. Thus, on the assumption of the existence of phonemically distinct tones in a northern dialect of Khmu', perceptual experiments with three sets of fundamental frequency (F0) patterns were run with 3 native speakers to explore the possible role of differences in shape of the contours [Gandour et al., 1978]; this was a sequel to a study of the F0 contours of its tones [Gårding and Lindell, 1977]. This dialect, however, is a member of a subgroup considered to have undergone tonogenesis perhaps as early as 700 years ago [Svantesson and House, 2006, p. 312]. In a recent study just called to our attention [Svantesson and House, 2006], the investigators used synthetic F0 contours for perception testing with speakers of nontonal Eastern Khmu' and speakers of tonal Western and Northern Khmu'. As far as we can tell, in the latter study too, only the absence or probable presence of distinctive tones was recognized; presumably no Western dialect conventionally described as having register complexes [Premsrirat, 2003] was involved. Rather, two tones differentiated essentially by pitch were the units to be studied.

For our first foray into this area [Abramson et al., 2004], we synthesized stimuli that varied incrementally along dimensions traditionally said to be relevant for the register distinction of the Kuai dialect of Suai, a Mon-Khmer language, and we did acoustic analyses of those properties in a large sample of speakers. We found the distinction to be in a state of flux both in production and perception. To the extent that it was still viable, the registers were mainly distinguished by F0 contours with some contribution from phonation type.

Responding to reports of greater stability in the register distinctions of some Western dialects of another Mon-Khmer language, Khmu' [Premsrirat, 2002, 2003], we have undertaken a study of the Khmu' Rawk dialect, the home language of 3,197 people in ten villages of the northern Nan Province of Thailand. Specifically, we have worked with speakers in the village of Huay Sataeng, which has 482 speakers, in the Thung Chang District. The Khmu' language belongs to the Khmuic branch of the Mon-Khmer language family. Khmu' in its many dialects is spoken in southern China and in northern areas of Vietnam, Laos, and Thailand. In Thailand about 10,000 Khmu' speakers live in the provinces of Chiangrai and Nan near the Lao border. The inventory of Khmu' Rawk phonemes is given in the 'Appendix'. Other dialects have been described [e.g., Smalley, 1961; Gårding and Lindell, 1977; Premsrirat, 1987, 2002].

With regard to our research topic, it is important to note that the languages of the Mon-Khmer family are of three types [Ferlus, 1980]: (1) nonregister and nontonal, (2) register (usually two register complexes), and (3) tonal (two tones). Indeed, the dialects of Khmu' itself are distributed across the three types. Type 1 is historically the most conservative with maintenance of the old voicing contrasts in consonants, while type 2 can be seen as a transitional stage of a movement toward tonality. Bringing linguistic findings and laryngeal mechanisms to bear on the matter, Graham Thurgood [2002, 2007] has argued plausibly for the probable role of voice register as a stage of tonogenesis in general. Evidence for its role in the Mon-Khmer family has been published [Thongkum, 1990] and specifically for its role in Khmu' [Premsrirat, 2003, table 4]. Premsrirat's [2003] listing of dialects places Khmu' Rawk in the subgroup of Western dialects with register complexes [see also L-Thongkum et al., 2007, p. 232]. Preliminary observations of the Khmu' Rawk spoken in the village of Huay Sataeng, accessible through the third author's project Linguistic Diversity in Nan Province: A Foundation for Tourism Development, did indeed strike the ear as belonging to type 2 with variation among its speakers as to how much Register 1 is characterized by modal (clear) voice or by high pitch and how much Register 2 is characterized by breathy voice or by low pitch. For some speakers interviewed, impressions included utterances with both phonetic properties present for each register, while for others it was just pitch.

Our research had two goals. First, we wished to determine what acoustic properties serve to differentiate the two voice registers in speech production. Are these registers indeed complexes of more or less equally important properties, such as phonation type and F0, or, rather, does one property dominate in the differentiation even if one or more other concomitant properties participate? If, in the latter case, the dominant property is F0 (the principal physical correlate of pitch), is tonogenesis in progress? Our second goal was to design experiments to determine the perceptual efficacy of any significant properties found in our acoustic analysis. Although we could not conduct physiological or aerodynamic studies in the village, we nevertheless expected to be able to rationalize our findings in the light of what is known about the acoustic consequences of articulatory gestures.

Methods

Production Studies

Recorded Materials

With a few refinements noted herein, the acoustic recording and analysis methods closely followed those employed in an earlier study [Abramson et al., 2004]. The recordings we analyzed were of nine word pairs; one member of each was uttered in Register 1 (clear voice, high pitch) and the other member in Register 2 (breathy voice, low pitch) (table 1). The words were recorded in randomized sequences by 25 native speakers of Khmu' (identified by the letters A through Z excluding D due to the excessively poor quality of his recording). Five of the 25 speakers were employed in two recording sessions that generated six tokens of each word; meanwhile the remaining speakers were engaged in one recording session that yielded three tokens per word. The speakers were of both sexes (9 men and 16 women), and they ranged in age from 19 to 72 years. To minimize the influence of surrounding languages of the Tai family (Standard Thai, Northern Thai, or Lao) on the voice registers of the Khmu' dialect it was necessary to obtain data from speakers who have spent their lives in the dialect region of interest. Such people are plentiful in rural villages, remote from the linguistic mélange of an urban environment and also, it must be noted with some regret, frequently unavoidably remote from such basic technical amenities as a sound-damped room or walled enclosure that, in an urban area, can be relied upon to attenuate ambient noise. Indeed, our recordings had to be made in an open-walled Buddhist worship hall. Although it was up the mountain slope at the edge of the village, people in vehicles, domestic animals, wild birds, and insects contributed to the ambient noise. Consequently, despite the use of a directional microphone, our recordings from time to time were contaminated by some low level domestic noises typical of those that arise in a farming community. All the recordings, originally sampled at a rate of 44,100/s, were downsampled to 11,025/s prior to analysis. The total number of analyzed tokens exceeded 1,670 whereas in the earlier study [Abramson et al., 2004] only 288 were ultimately examined.

Noise contamination, exacerbated by the habit of our Southeast Asian subjects to speak softly in face-to-face exchanges and when confronted with a microphone, undermined the robustness of several computer-implemented analysis procedures, particularly those involving formant frequency extraction. Formant frequencies extracted by computer analysis were among the least reliable data. Consequently, all formant frequencies were obtained by visual inspection. We might note here that while we lack any systematic evidence of a general cultural tendency among people of the region to favor passive manners of interpersonal communication, we believe that we have much anecdotal evidence to support such a claim.

Table 1. List of Khmu' word pairs spoken from 3 to 6 times each by 25 native speakers

Register 1		Register 2	
transcription	gloss	transcription	gloss
khɛ́ɛm	Johnson grass	khɛ̀ɛm	waterway edge
lɔ́ɔk	haunted	lɔ̀ɔk	to peel off
már	sterile	màr	snake
míəŋ	fermented tea leaves	mìəŋ	to chew
phák	vegetable	phàk	to ride
phɛ́ɛ	wound	phɛ̀ɛ	raft
phlúʔ	firework	phlùʔ	thigh
ráaŋ	tooth	ràaŋ	flower
wák	to hang	wàk	abyss

Each word was inspected in a combined waveform and spectrographic format, and its visible beginning and ending points were identified and stored. A similar procedure was used to visually identify and store the onset and offset points of 'steady-state' regions of the spectrum, regions in which the variation in first and second formants was less than the difference limens of 70 and 170 Hz [Flanagan, 1955a; Mermelstein, 1978].

Praat, Matlab®, and StatView© software was used throughout to perform acoustic and statistical analyses in the manner described in greater detail in an earlier paper [Abramson et al., 2004]. Some empirical modifications were made to portions of the Matlab software in an effort to improve the precision of formant extraction. However, we ultimately chose to rely upon formant-frequency data that we obtained by the combined visual inspection of broad-band and narrow-band spectrograms and narrow-band cross sections of steady-state regions of the speech data.

Amplitude and F0 Contours

Overall amplitude values in decibels and glottal frequency (F0) values over a 75- to 600-Hz range were obtained at 10-ms intervals for each utterance. Visual inspection was employed to eliminate some instances of frequency doubling, and the resulting frequency values were then converted from Hertz to semitones using the formula $P_{semitones} = 3.32 \times 12 \times \log_{10}((F0_{Hz})/\text{base})$. This conversion, done as part of the normalization across speakers, was meant to more satisfactorily approximate the listeners' sensation of pitch. In the case of each speaker the 'base' was the minimum F0 frequency measured for that speaker across all utterances in both voice registers. Thus, for each speaker, a single base value was computed and applied in the conversion of all F0 contours in both registers. Next, both the frequency and amplitude data were time-normalized to fit 100-point scales, and, using these scales, the amplitude and F0 data for different subgroups of subjects (males vs. females; old speakers vs. youthful speakers) and the entire coterie of speakers were averaged. Graphs of these data are shown in figures 1–10. Finally, to help us examine the differences between the time-normalized contours in each register, we borrowed a procedure from the previous study. Basically, this involved the examining of plots of the averaged data and selecting sectors along the normalized time scale where the largest differences in rate of change of amplitude or F0 between the two Registers were apparent. Thus, in this instance, we divided the time scale into four sectors (1–25, 26–50, 51–85, and 86–100) and computed the gradients of the Register 1 and Register 2 contours in each sector. Thus, for each utterance by each speaker, four gradients were calculated by linear regression. Then for each of the gradients 1–4, each speaker's repetitions of a given word in a given register were averaged. This procedure led to the formation of a data set of 1,800 measurements (25 speakers $\times$ 9 words $\times$ 2 registers $\times$ 4 gradients) that was subsequently submitted to a repeated-measures analysis of variance (ANOVA).

Formant Frequencies

Shortcomings in the quality of the speech, primarily due, we believe, to the generally soft speech of the speakers but also partially due to background noise, led to numerous small errors in the results produced by a usually effective linear predictive coding method of formant extraction. To correct these errors, a Matlab program was created to display broad- or narrow-band spectrograms and narrow-band cross sections of each member of the entire steady-state data set. Using this display, both first (F1) and second (F2) formant frequencies were identified by eye, and their values were stored in a file linked to each utterance of origin. These data were subsequently used in statistical tests of the stability of the F1 and F2 values as a function of voice register and also in calculating harmonic ratio values. That is, mismatched formant frequencies between the members of a word pair would invalidate any measure of spectral tilt for that pair.

Ratios of Harmonic Intensities

Following the procedure adopted earlier [Abramson et al., 2004], for each steady-state region of the vowel formants the intensity of the fundamental (H_1) and its second harmonic (H_2) were identified and combined with the intensity of the F1 peak harmonic (H_{F1}) to form two ratios, H_2/H_1 and H_{F1}/H_1. For this study, however, a third ratio involving F2, H_{F2}/H_1, was added. The harmonic extraction procedure applied a Hamming window to each steady-state data segment and extended it to a length of 1,024 points by the addition of zeros. A fast Fourier transform of the 1,024-point array led to the production of a spectrum vector with a 21.5-Hz frequency resolution. The harmonic frequency was extracted from the cepstrum and a sine wave with that frequency was aligned or synchronized by autocorrelating the sine wave with the harmonic structure of the spectrum vector. Finally, peaks in the sine wave representing the fundamental and its second harmonic were used to locate those specific frequencies in the harmonic spectrum and extract their intensities.

Vowel Duration

To determine whether differences in vowel length might be being used to differentiate the two voice registers [Thongkum, 1988; Gordon and Ladefoged, 2001] we repeated our vowel-duration measurement procedure based on a gestural criterion for defining the span [Abramson et al., 2004, pp. 152–153]. Thus, the vowel was taken to run from the release of the initial consonant to the closing gesture of any final consonant. For the one pair with initial /w/, the onset of the vocalic glide was marked as the beginning, and for the one pair with no final consonant, the last detectable voice excitation of at least one formant was marked as the end. All vowel onset and offset points were identified by visually inspecting a wide-band spectrogram of each utterance. The positions of a hairline cursor placed at the onset and offset boundaries of each vowel were recorded.

Perception Studies

Natural Speech Perception

Whatever the phonetic properties of the two traditional voice registers might be today, we had no reason to doubt the validity of the phonological distinction itself in Khmu' Rawk; nevertheless, as a basis for our perceptual experiment, we wished to have an assessment of its auditory robustness. In addition, dealing with a similar populace with little schooling in our earlier study of Suai, we had encountered a small number of listeners who lacked either the ability to hear the distinction or an understanding of the task, and we wished to verify that each Khmu' listener understood and could carry out the instructions correctly. The natural stimuli were two utterances of the words /ráaŋ/ 'tooth' in Register 1 and /ràaŋ/ 'flower' in Register 2; one pair of utterances was spoken by a male (speaker J) and the other pair by a female (speaker K). Chosen for the salience of the register distinction, each of the selected utterances was presented seven times in one of two random orders to 32 native speakers (10 men and 22 women) for identification. Pictures of teeth and flowers lay in front of the listeners, and, to indicate what they had heard, they made their selections known by pointing at the appropriate picture. The pointing responses were observed by the experimenter and recorded via the computer's keyboard.

Speech Synthesis

As a starting point for synthesis, a production of the word /ráaŋ/ in (clear voice) spoken by speaker A was selected as the model utterance. The SynthWorks® software package was used to extract its synthesis parameters, among which was included an F0 contour. Although in our previous study of Suai the behavior of F0 was found to be the most significant factor in a multifactor analysis of listeners' Register 1 vs. Register 2 responses to synthetic stimuli, we had no reason at the outset of our work

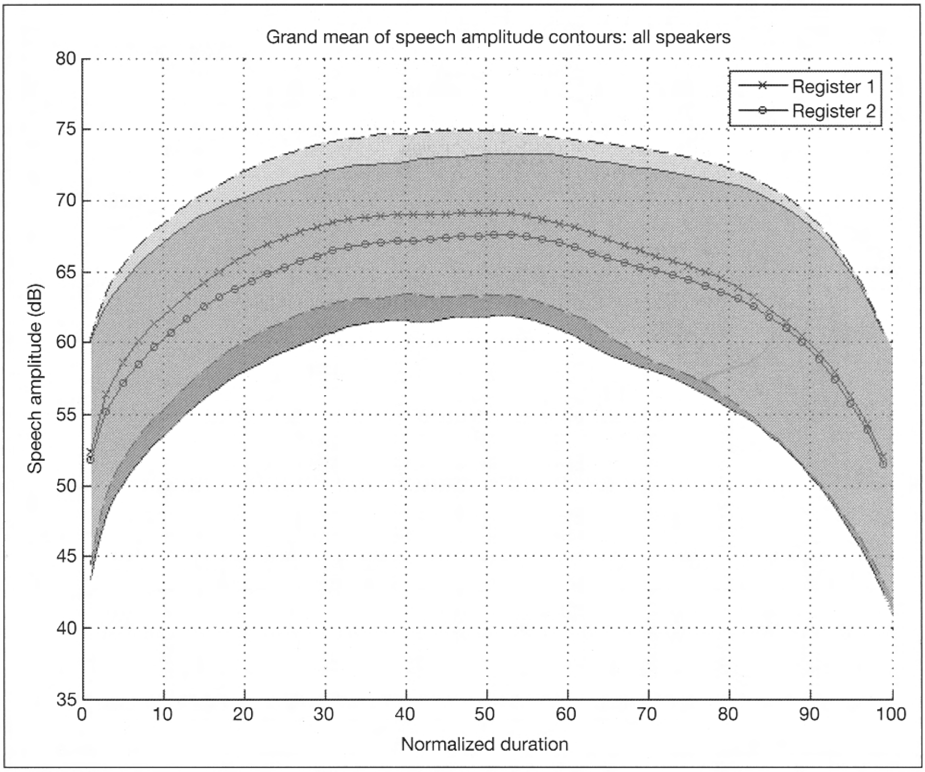

Fig. 1. Average amplitude contours for all Register 1 and Register 2 utterances produced by 25 Khmu' speakers of both sexes. The shaded regions within dashed lines and continuous lines indicate the standard deviations of the Register 1 and Register 2 contours, respectively.

to believe that the same would be true of our dialect of Khmu'. We were led to focus our attention solely on the F0 parameter in this study because of the exigencies of scheduling. When it was nearly time for the first author to take up residence in Thailand to prepare the test orders and logistics for carrying out the perceptual experiments, we had not done much of the analysis of harmonic ratios. Thus, at this stage of the work, we had no analytic basis for assuming that phonation type is relevant to the phonological distinction. Therefore, it seemed prudent at the time to proceed only with F0, the significance of which we had found through acoustic analysis. The averaged F0 contours obtained from measurements of natural productions of the two registers (fig. 6) provided the basis for synthesizing a set of seven different F0 contours representing steps along a hypothetical F0 continuum. The interpolated contours were obtained by computing increments representing one sixth of the differences between the frequency values of Registers 1 and 2 at each time sample. These increments were used to create seven contours that, in addition to those of the original averaged F0 Register 1 and Register 2 contours, included four intermediate contours and one super stimulus that followed a trajectory lying one increment above that of Register 2 (fig. 13).

Perception of Synthetic Speech

The same 32 native speakers of Khmu were paid to listen to sequences of synthesized words presented in one of two random orders and to identify them. Each random order contained seven tokens of each stimulus. Again, pictures of teeth and flowers lay in front of the listeners and, to indicate what they had heard, they made their selections known by pointing at the appropriate picture. The synthesized

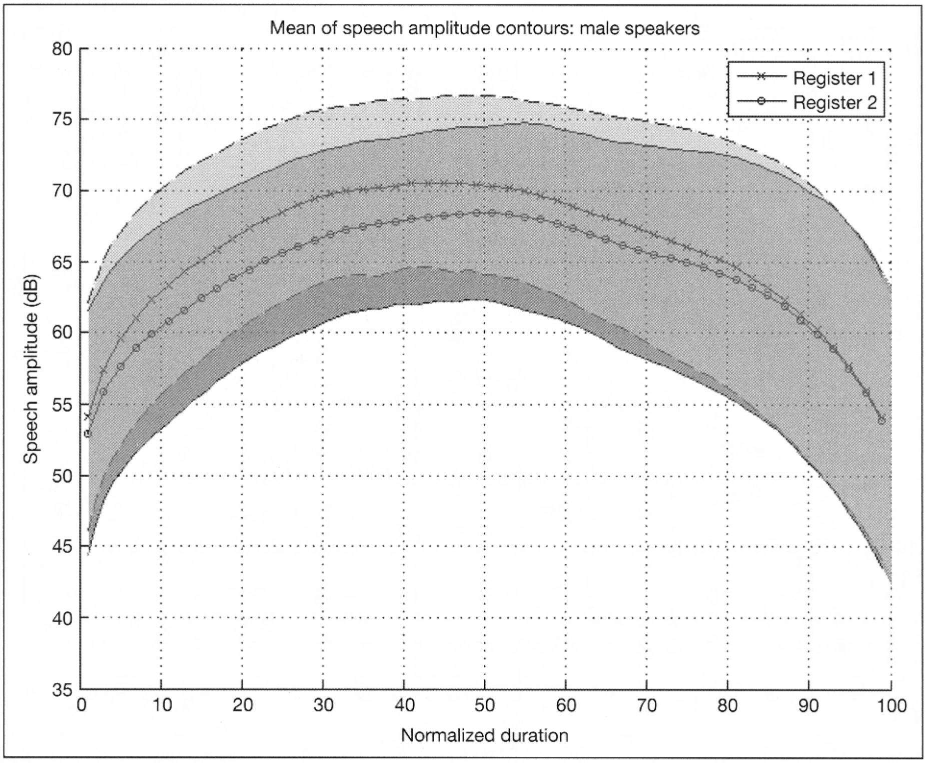

Fig. 2. Average amplitude contours for all Register 1 and Register 2 utterances produced by 9 male Khmu' speakers. The shaded regions within dashed lines and continuous lines indicate the standard deviations of the Register 1 and Register 2 contours, respectively.

words were reproduced by a Macintosh laptop computer in sequences controlled by a program written for PsyScope [Cohen et al., 1993].

Results

Amplitude and F0

Figure 1 shows the overall averages of the time-normalized amplitude contours of 25 speakers' productions of nine utterances in both Register 1 and Register 2. With the issue of register stability in mind, subsets of the data were plotted. They are shown in figures 2–5 and represent the collective performances of male, female, young, and old speakers, respectively. In every instance, the peak of the Register 1 contour is consistently higher than that of Register 2. Nevertheless, to address the question of whether the rate of change of amplitude might differ, we took our examination of the data a step further and divided each amplitude contour into four sectors along the time dimension and conducted an ANOVA. The results of this analysis are summarized in table 2, which shows in the 25-speaker group an inconsistent pattern of register-effect probabilities that reaches significance only for sector 3. (In this paper, an effect is termed

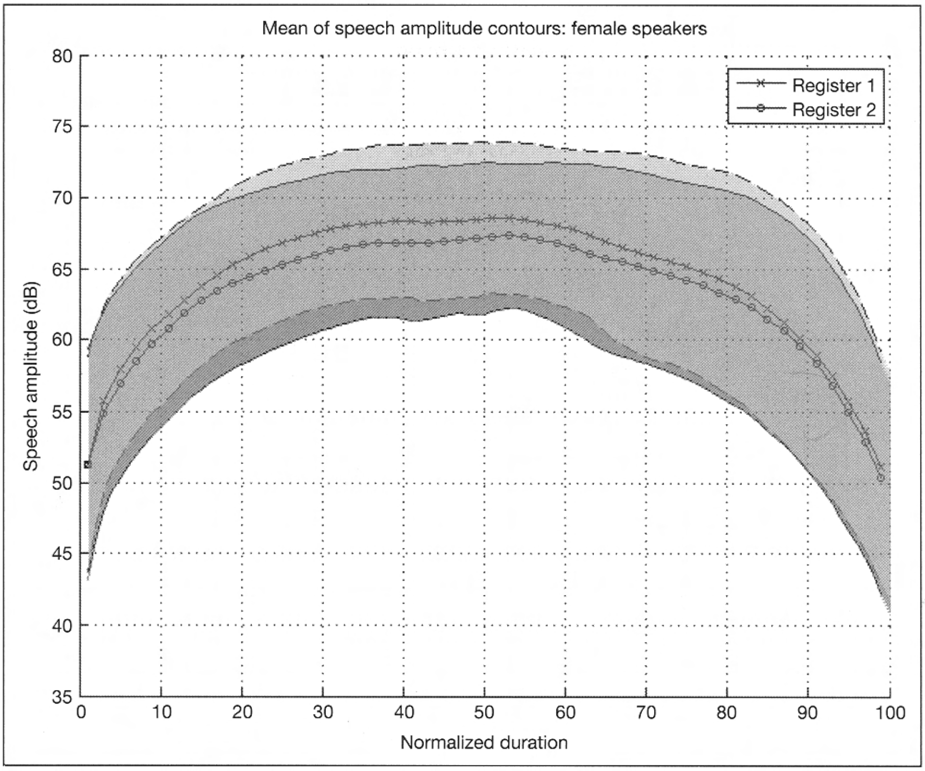

Fig. 3. Average amplitude contours for all Register 1 and Register 2 utterances produced by 16 female Khmu' speakers. The shaded regions within dashed lines and continuous lines indicate the standard deviations of the Register 1 and Register 2 contours, respectively.

significant when the probability p that a null hypothesis is true is less than 0.005, i.e., $p < 0.005$.) This inconsistency is also apparent for the smaller subgroups.

In figure 6 the time-normalized average F0 contours of the Register 1 and Register 2 productions of 25 speakers are shown plotted on a semitone scale. A mean difference of 3.27 semitones is apparent over the entire duration scale. The maximum difference of 4.2 semitones occurs at time datum 18. Plots of data in the age and sex categories are shown in figures 7–10. The female contours are a few semitones higher in pitch than those of the males and exhibit larger standard deviations. The average difference across all 100 points of the normalized time scale ranges from a maximum of 3.54 semitones (young speakers) to 2.56 semitones (old speakers). Meanwhile, the maximum Register 1–Register 2 difference (4.37 semitones) occurs at datum 26 of the young dataset, while the minimum difference (0.97 semitones) is found in datum 9 of the old speakers' dataset. To examine differences in F0 contour shape, we computed the gradients of the two contours in each of the four normalized time periods or sectors. The results, which appear in table 3, reveal that for the group consisting of all 25 speakers there are significant differences between Register 1 and Register 2 in each of the four sectors. For a group composed exclusively of male speakers, only the two mid sectors of the normalized time scale yield significant differences. In contrast, the corresponding results for

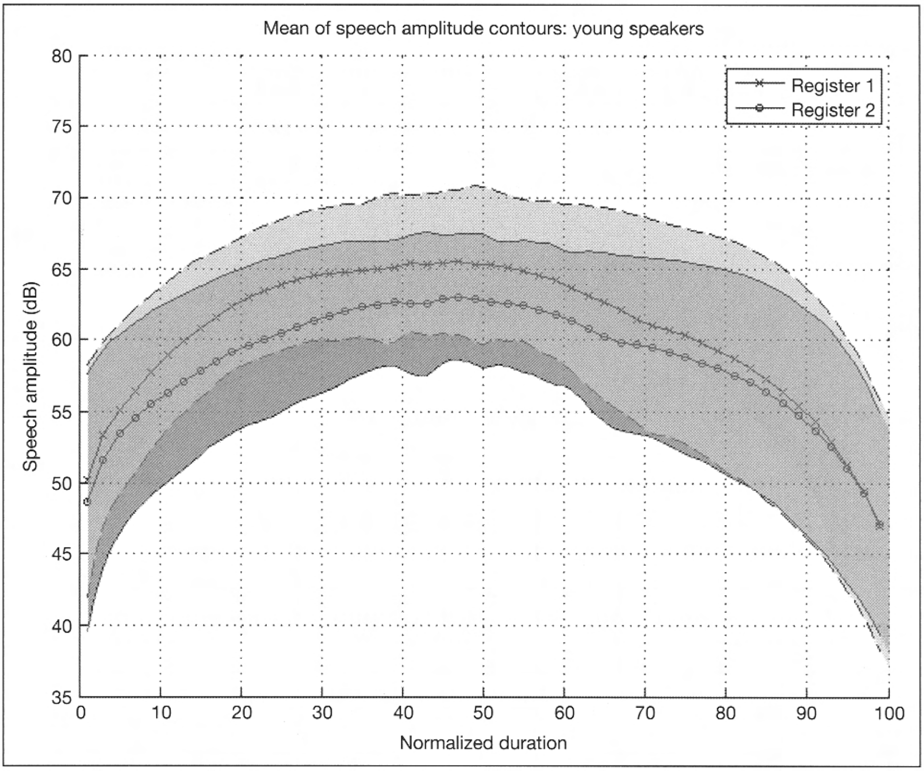

Fig. 4. Average amplitude contours for all Register 1 and Register 2 utterances produced by 5 young Khmu' speakers. The shaded regions within dashed lines and continuous lines indicate the standard deviations of the Register 1 and Register 2 contours, respectively.

the female group show significant effects across all four sectors. Elderly speakers as a group produce significant mid-section results like those of male speakers, while younger speakers have somewhat less clear results, given the probability $p = 0.009$ in sector 2. However, the fact that $p << 0.005$ in both sectors 1 and 3 suggests that it may be reasonable to conclude that the shape of the F0 contour over all of the first three sectors is important.

Analysis of Vowel Duration

Vowel duration data derived from onset and offset times established by the visual inspection of spectrograms were analyzed by an ANOVA. The test was designed to determine whether the two registers were indistinguishable with regard to vowel duration. The results, shown in table 4, reveal that Register-2 vowels are longer with a mean difference in duration of 22 ms for the full 25-speaker dataset. Moreover, within both age and sex subcategories, the difference, ranging from 28 to 21 ms, persists. An analysis to determine whether this duration difference might have arisen by chance shows that, with the exception of the old and young datasets that are marred by small sample sizes ($n = 6$ and 5, respectively), the probability is exceedingly small ($p < 0.0003$).

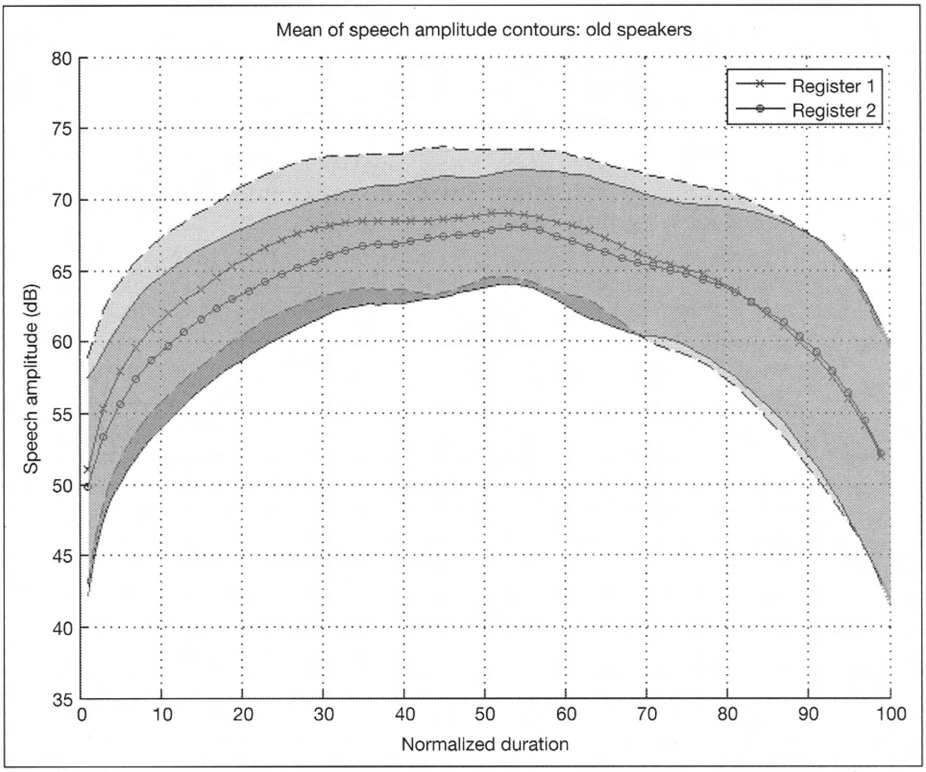

Fig. 5. Average amplitude contours for all Register 1 and Register 2 utterances produced by 6 older Khmu' speakers. The shaded regions within dashed lines and continuous lines indicate the standard deviations of the Register 1 and Register 2 contours, respectively.

Formant Frequencies

To determine whether productions of the two registers might be distinguished by systematic variations in formant frequencies, we performed independent repeated-measures ANOVAs on the visually extracted F1 and F2 data. The results are shown in table 5. The probabilities that the populations of F1 and F2 frequencies in Register 1 are indistinguishable from their corresponding populations in Register 2 are very high for all speaker categories. Thus, no matter the age or sex of the speaker, Registers 1 and 2 do not differ significantly with respect to either their F1 or F2 frequencies.

Harmonic Intensity Ratios

As a result of the presence of recorded noise coupled with the speakers' tendency toward soft speech, the periodicity of the harmonic spectrum was sometimes buried in the noise, and the autocorrelation procedure failed to find a plausible alignment. Therefore, graphs of all the spectra and their autocorrelation results were visually examined for instances of synchronization failure which, when found, led to the omission of that ratio value. The frequency distributions of the surviving Register 1 and Register 2 ratio values of all types (H_2/H_1, H_{F1}/H_1, and H_{F2}/H_1) were positively skewed and, to achieve closer approximations to the Gaussian or normal distributions underlying

Table 2. Amplitude gradients

Groups	Source	Sector 1				Sector 2			
		F(d.f.)	F	MSe	p	F(d.f.)	F	MSe	p
All speakers	Register	(1, 24)	8.529	0.206	0.0075	(1, 24)	5.647	0.055	0.0258
	Word	(8, 192)	19.503	0.791	0.0001	(8, 192)	93.487	1.507	0.0001
	Reg × Word	(8, 192)	2.001	0.032	0.0482	(8, 192)	2.349	0.016	0.0295
Male	Register	(1, 8)	7.431	0.125	0.0260	(1, 8)	18.067	0.058	0.0028
	Word	(8, 64)	18.993	0.594	0.0001	(8, 64)	31.556	0.533	0.0001
	Reg × Word	(8, 64)	0.898	0.014	0.5236	(8, 64)	1.668	0.009	0.1570
Female	Register	(1, 15)	3.153	0.092	0.0961	(1, 15)	0.972	0.012	0.3398
	Word	(8, 120)	8.526	0.333	0.0001	(8, 120)	63.357	0.991	0.0001
	Reg × Word	(8, 120)	1.872	0.032	0.0883	(8, 120)	2.040	0.015	0.0473
Old	Register	(1, 5)	1.168	0.013	0.3292	(1, 5)	41.177	0.088	0.0014
	Word	(8, 40)	4.141	0.166	0.0011	(8, 40)	22.509	0.313	0.0001
	Reg × Word	(8, 40)	0.832	0.016	0.5796	(8, 40)	3.220	0.015	0.0064
Young	Register	(1, 4)	15.358	0.182	0.0173	(1, 4)	14.316	0.025	0.0194
	Word	(8, 32)	10.976	0.458	0.0001	(8, 32)	40.282	0.450	0.0001
	Reg × Word	(8, 32)	0.512	0.009	0.8381	(8, 32)	0.890	0.005	0.5361

Groups	Source	Sector 3				Sector 4			
		F(d.f.)	F	MSe	p	F(d.f.)	F	MSe	p
All speakers	Register	(1, 24)	15.090	0.091	0.0007	(1, 24)	0.102	0.005	0.7525
	Word	(8, 192)	103.308	0.178	0.0001	(8, 192)	10.629	1.182	0.0001
	Reg × Word	(8, 192)	2.346	0.008	0.0244	(8, 192)	1.954	0.079	0.0886
Male	Register	(1, 8)	33.005	0.075	0.0004	(1, 8)	1.089	0.030	0.3272
	Word	(8, 64)	38.519	0.672	0.0001	(8, 64)	4.088	0.224	0.0065
	Reg × Word	(8, 64)	1.018	0.003	0.4317	(8, 64)	2.142	0.067	0.1136
Female	Register	(1, 15)	3.910	0.030	0.0667	(1, 15)	0.027	0.002	0.8724
	Word	(8, 120)	67.303	1.252	0.0001	(8, 120)	18.944	1.779	0.0001
	Reg × Word	(8, 120)	2.578	0.009	0.0124	(8, 120)	1.212	0.054	0.2978
Old	Register	(1, 5)	4.641	0.026	0.0838	(1, 5)	0.100	0.001	0.7652
	Word	(8, 40)	26.624	0.417	0.0001	(8, 40)	0.760	0.073	0.6395
	Reg × Word	(8, 40)	1.496	0.006	0.1893	(8, 40)	2.280	0.079	0.0410
Young	Register	(1, 4)	34.575	0.065	0.0042	(1, 4)	4.076	0.126	0.1136
	Word	(8, 32)	35.724	0.510	0.0001	(8, 32)	4.120	0.414	0.0018
	Reg × Word	(8, 32)	1.354	0.004	0.2540	(8, 32)	0.433	0.024	0.8923

Probability p that the hypothesis of a zero difference between the overall amplitude gradients of Registers 1 and 2 is true. The analysis performed in four sectors of time-normalized utterance duration. Results are shown for a group of 25 speakers and four subgroups of those speakers.

repeated-measures ANOVAs, a log transformation of the ratio data derived from each repetition of each of the words was applied prior to analysis. Thus, for each of the three ratio types, the log ratio for each repetition of a given word uttered by a given speaker was calculated and averaged over all repetitions of that word to yield one datum for each ratio type, word, and register. As a result of rejecting all the ratios affected by synchronization failure, a total of 3.05% of the Register 1 data derived from individual

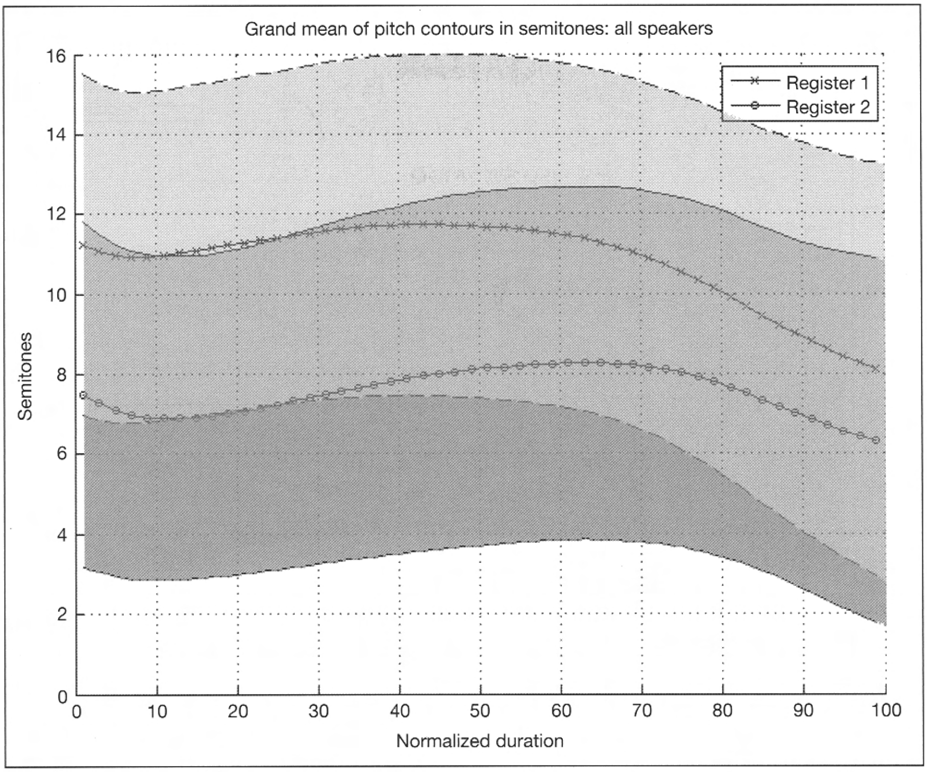

Fig. 6. Normalized pitch contours averaged over the Register 1 and Register 2 utterances of 25 Khmu' speakers of both sexes. The shaded regions within dashed lines and continuous lines indicate the standard deviations of the Register 1 and Register 2 contours, respectively.

word repetitions and 5.00% of the corresponding Register 2 data was eliminated before averaging. Histograms of Register 1 and Register 2 log-ratio data obtained from all 25 speakers are shown in figure 11.

Table 6 summarizes the results of an analysis of the logarithmically transformed harmonic intensity ratios. The pooled data of the entire group of 25 speakers reveal significant differences ($p < 0.005$) for *Register* for the first two ratios but not for the third. Significant differences are found for *Word* in all three cases as expected, since the nine words were all chosen to be distinctively different from one another. The *Register* × *Word* interaction, meanwhile, lacks significance for all three ratios. Harmonic ratios are a measure of the slope of the voice spectrum, and significant differences between the two voice registers suggest that the durations of the open and closed quotients of the glottal cycle undergo change [Stevens, 1998, pp. 85–92]. As a consequence of their different social roles in the community, male and female speakers are exposed to different influences; females may spend more time within their village, while males may travel further afield and become exposed to greater linguistic variation. Thus, it is of interest to analyze the data of males and females independently. In this case the analysis shows that, in male speakers, the glottal duty cycle makes no contribution toward the

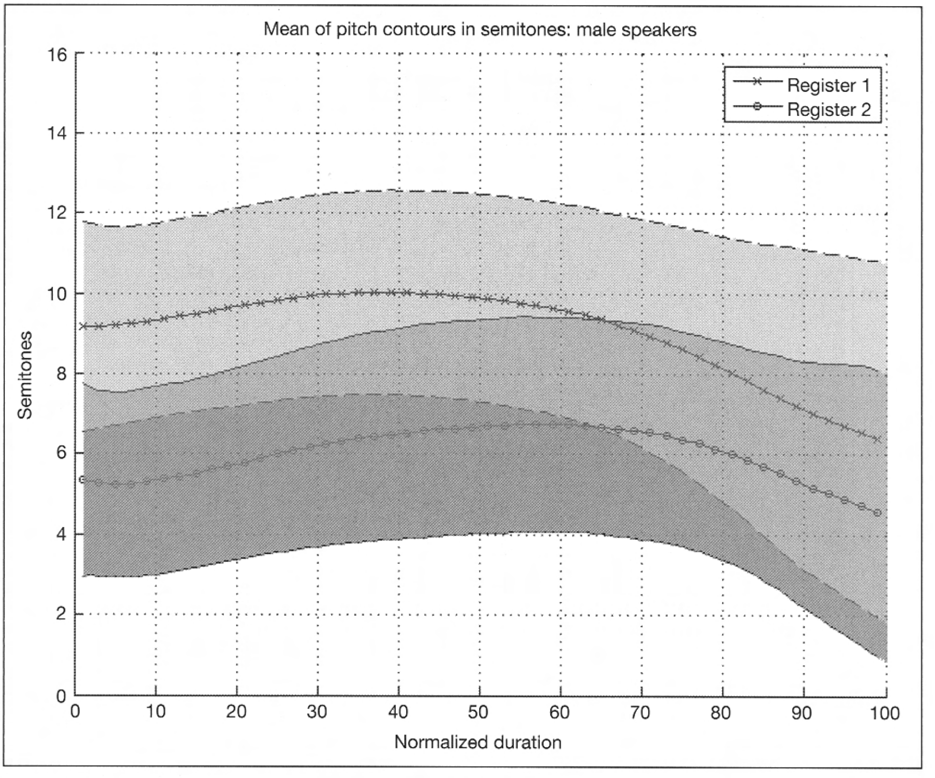

Fig. 7. Normalized pitch contours averaged over the Register 1 and Register 2 utterances of 9 male speakers of Khmu'. The shaded regions within dashed lines and continuous lines indicate the standard deviations of the Register 1 and Register 2 contours, respectively.

production of voice register distinctions. The situation is shown to be different for females whose data achieve the $p < 0.005$ level of significance for the first of the three log ratios. Less reliable, due to the small quantity of data available for analysis, are the results of old vs. young speakers. All probability levels for register fall substantially short of significance and do not permit any conclusions to be drawn as to the effects of changes in linguistic norms over time.

Perception of Control Stimuli

Figure 12 shows the overall percentage of responses given by 32 listeners to utterances spoken in Register 1 and Register 2 by speakers J and K. At a level of 96.9% versus 93.1%, the Register 1/Register 2 distinction was apparently perceived more readily in the productions of the female speaker (K) than in those of the male speaker. Eleven of the 32 listeners made errors. Eight of those listeners made three errors or fewer, while the remaining 3 listeners made six errors or more. When the responses of the 3 lowest performing listeners are removed, the overall response levels to the productions of male and female speakers rise to 97.2 and 100%, respectively. Accordingly, the responses of the 3 least acute listeners were excluded from our analysis of the perception test using synthesized stimuli.

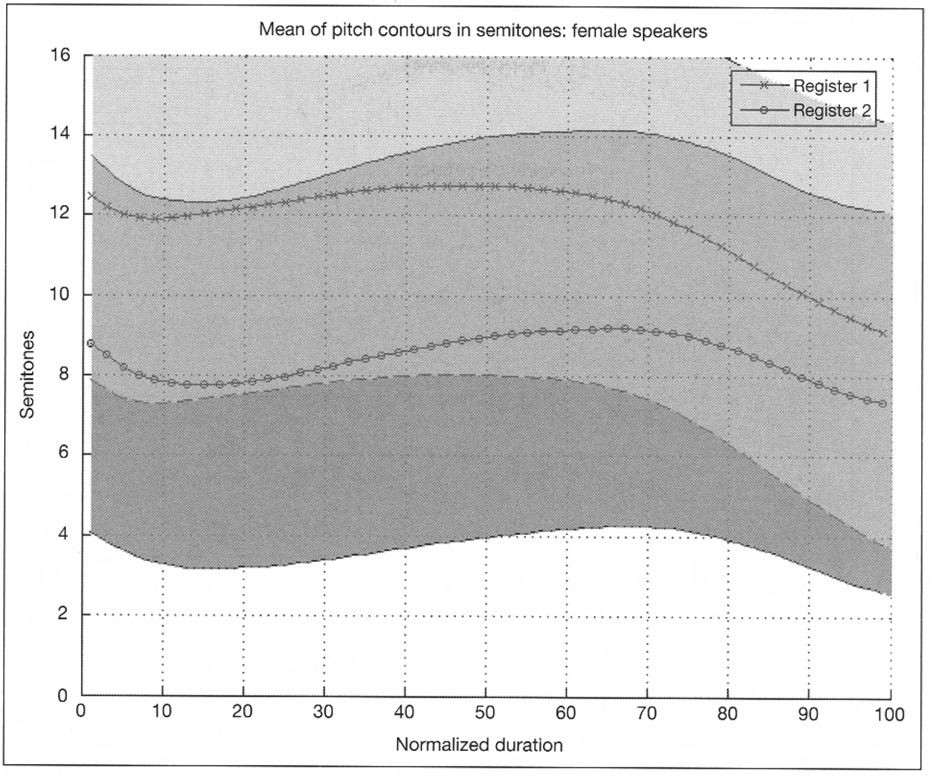

Fig. 8. Normalized pitch contours averaged over the Register 1 and Register 2 utterances of 16 female speakers of Khmu'. The shaded regions within dashed lines and continuous lines indicate the standard deviations of the Register 1 and Register 2 contours, respectively.

Perception of Synthetic Stimuli

The same group of 32 listeners was engaged to judge the synthetic utterances. These utterances differed only in their F0 contours and were made to match the overall F0 contours of male speakers in the two registers (fig. 13), with the addition of five interpolated F0 trajectories to bring the number of intervals along the F0 continuum to a total of seven. The response data, omitting the responses of 3 of the 32 listeners for the reasons noted above, are shown plotted in figure 14. The plot lines are quasi-sigmoidal in shape and indicate that, at the extreme, Register 1 performance lay at the 90% level. At the Register 2 end of the scale, the percentage of correct responses is slightly lower than that for Register 1, and it is evident that no enhancement of recognition performance was achieved by the F0-elevated super stimulus.

Discussion

Amplitude

Although the amplitude peak in at least one sector for all speakers, as well as for all the subgroups, is significantly greater in Register 1 (table 2), the small difference

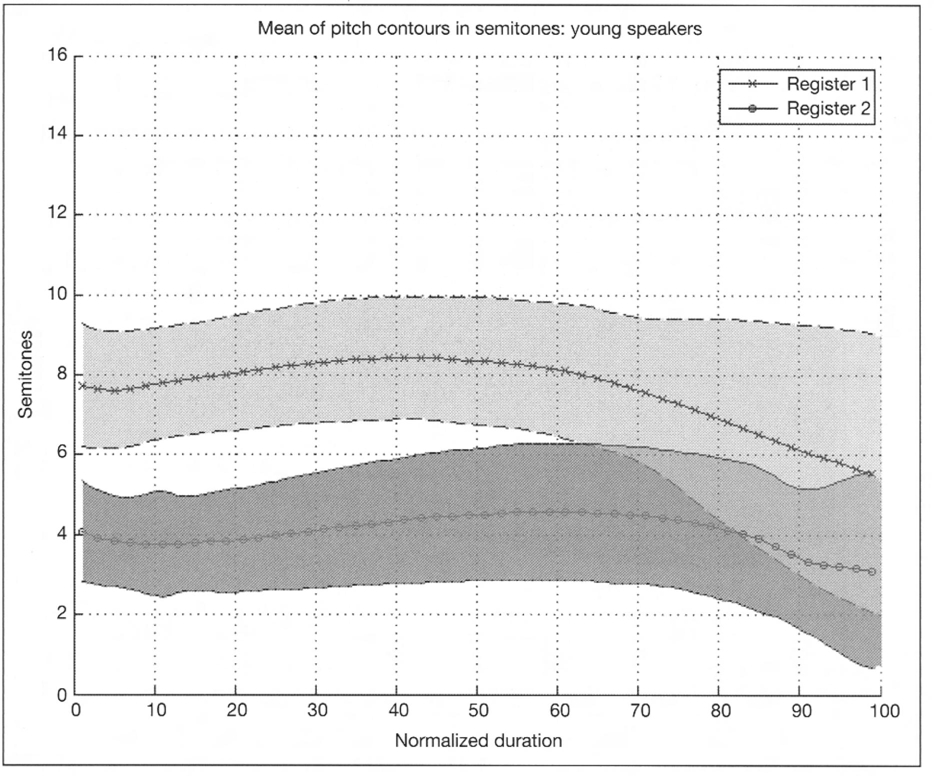

Fig. 9. Normalized pitch contours averaged over the Register 1 and Register 2 utterances of 5 young speakers of Khmu'. The shaded regions within dashed lines and continuous lines indicate the standard deviations of the Register 1 and Register 2 contours, respectively.

makes us believe that it can have little or no value in differentiating the two registers. These differences are seen to be smaller than the standard deviations of the individual utterances. In each graph, dark and light shaded areas indicate the span of standard deviations for Registers 1 and 2, respectively. The graphs of data in all categories reveal a broadly identical pattern, a predominantly positive Register 1–Register 2 difference, whose mean ranges from a maximum of 2.05 dB (young speakers) to a minimum of 1.10 dB (female speakers). The largest single difference (3.4 dB) occurs in the young speakers' plot at datum 23 on the normalized time scale. Flanagan [1955b] measured the perceptible difference limen for vowel amplitude and concluded that, for a synthesized neutral vowel of fixed duration, the threshold lay in the neighborhood of 1.5 dB. Natural vowels extracted from spoken words have variable durations and a variety of vocal damping and excitation characteristics that would be expected to give rise to a higher limen that exceeds the differences found here. In brief, the difference limens of steady-state synthetic vowels are smaller than those of the typically dynamic vowels of natural speech [O'Shaughnessy, 2000, pp. 134–136]. Thus, it appears plausible that, notwithstanding the persistence of an absolute Register 1–Register 2 difference, it is unlikely to be perceptible to native speakers of Khmu'. That is, as a parameter of the

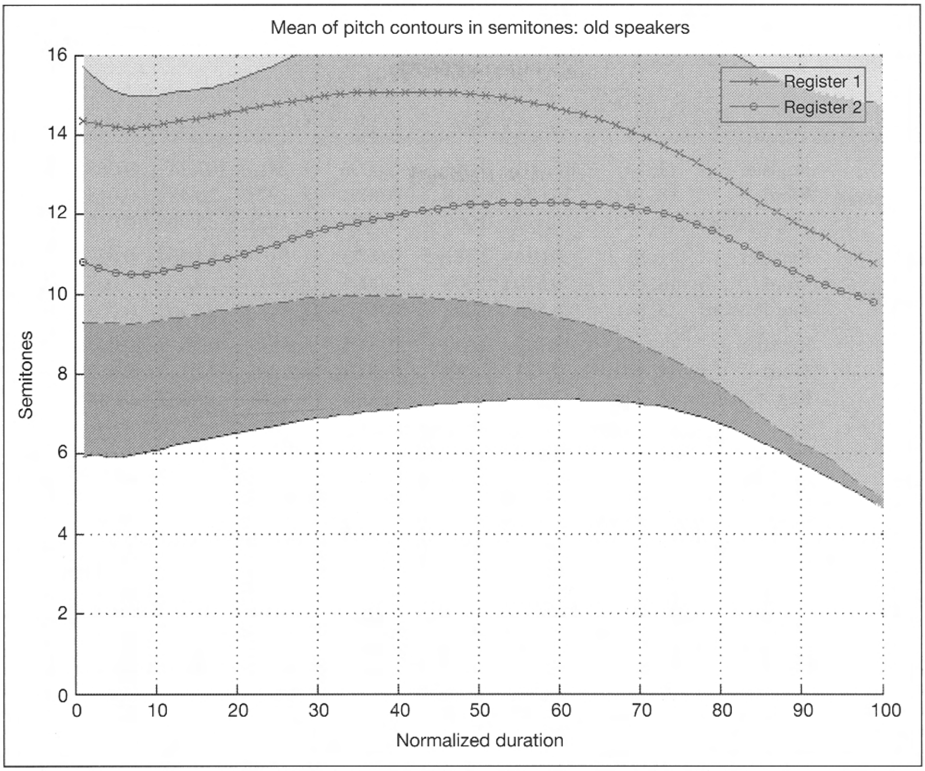

Fig. 10. Normalized pitch contours averaged over the Register 1 and Register 2 utterances of 6 elderly speakers of Khmu'. The shaded regions within dashed lines and continuous lines indicate the standard deviations of the Register 1 and Register 2 contours, respectively.

speech mechanism, it is not here under voluntary control. Rather, it seems to be tightly correlated with the F0 contours.

Despite the foregoing, the question arises as to why the amplitude should be consistently higher for Register 1 along with a higher F0. If we take the position that the language has given up its register system and has shifted to a tonal system with F0 contours as the paramount differentiators, as implied at least by the male data, one possibility is that higher subglottal air pressure, the major contributor to sound pressure level, makes enough of a contribution to the higher F0 to account for the higher output amplitude. That is, even though most of the control of the rate of glottal pulsing resides in the laryngeal muscles, subglottal pressure may play a sufficient role to account for the small but statistically significant differences in amplitude. Indeed, a lengthy exposition [Fant et al., 2000, pp. 122–125] of covariation of subglottal pressure, F0, and sound pressure level explains that up to approximately the mid point of a speaker's voice range the voice-excitation amplitude increases with F0. Of course, aside from these trends toward covariation, subglottal pressure and F0 can vary independently. In a follow-up study [Fant and Kruckenberg, 2004, pp. 79–80] we find, '. . . we may now predict that in connected speech there should exist conformity between intonation

Table 3. F0 gradients

Groups	Source	Sector 1				Sector 2			
		F(d.f.)	F	MSe	p	F(d.f.)	F	MSe	p
All speakers	Register	(1, 24)	15.601	0.046	0.0006	(1, 24)	107.74	0.068	0.0001
	Word	(8, 192)	16.973	0.020	0.0001	(8, 192)	20.482	0.015	0.0001
	Reg × Word	(8, 192)	10.553	0.009	0.0001	(8, 192)	35.905	0.010	0.0001
Male	Register	(1, 8)	0.014	2.7E−5	0.9073	(1, 8)	54.564	0.022	0.0001
	Word	(8, 64)	9.510	0.008	0.0001	(8, 64)	9.665	0.005	0.0001
	Reg × Word	(8, 64)	4.823	0.003	0.0001	(8, 64)	12.362	0.003	0.0001
Female	Register	(1, 15)	32.787	0.070	0.0001	(1, 15)	58.819	0.047	0.0001
	Word	(8, 120)	9.421	0.013	0.0001	(8, 120)	17.146	0.012	0.0001
	Reg × Word	(8, 120)	8.366	0.008	0.0001	(8, 120)	27.657	0.008	0.0001
Old	Register	(1, 5)	0.013	3.1E−5	0.9141	(1, 5)	73.418	0.025	0.0004
	Word	(8, 40)	7.075	0.006	0.0001	(8, 40)	3.007	0.003	0.0097
	Reg × Word	(8, 40)	5.460	0.004	0.0001	(8, 40)	5.507	0.002	0.0001
Young	Register	(1, 4)	237.24	0.013	0.0001	(1, 4)	21.960	0.005	0.0094
	Word	(8, 32)	3.035	0.003	0.0116	(8, 32)	8.880	0.003	0.0001
	Reg × Word	(8, 32)	5.889	0.003	0.0001	(8, 32)	14.587	0.002	0.0001

Groups	Source	Sector 3				Sector 4			
		F(d.f.)	F	MSe	P	F(d.f.)	F	MSe	p
All speakers	Register	(1, 24)	257.02	0.217	0.0001	(1, 24)	10.794	0.055	0.0031
	Word	(8, 192)	118.80	0.178	0.0001	(8, 192)	20.645	0.193	0.0001
	Reg × Word	(8, 192)	13.367	0.006	0.0001	(8, 1 92)	1.669	0.006	0.1083
Male	Register	(1, 8)	86.231	0.052	0.0001	(1, 8)	0.505	0.002	0.4975
	Word	(8, 64)	73.970	0.062	0.0001	(8, 64)	12.643	0.103	0.0001
	Reg × Word	(8, 64)	7.719	0.003	0.0001	(8, 64)	0.247	0.001	0.9799
Female	Register	(1, 15)	227.6	0.170	0.0001	(1, 15)	12.766	0.069	0.0028
	Word	(8, 120)	64.722	0.009	0.0001	(8, 120)	12.993	0.116	0.0001
	Reg × Word	(8, 120)	7.419	0.003	0.0001	(8, 120)	3.032	0.011	0.0039
Old	Register	(1, 5)	42.921	0.052	0.0012	(1, 5)	2.177	0.030	0.2001
	Word	(8, 40)	25.711	0.048	0.0001	(8, 40)	7.171	0.078	0.0001
	Reg × Word	(8, 40)	3.391	0.001	0.0046	(8, 40)	0.449	0.003	0.5930
Young	Register	(1, 4)	102.27	0.034	0.0005	(1, 4)	1.743	0.006	0.2573
	Word	(8, 32)	21.119	0.026	0.0001	(8, 32)	3.470	0.033	0.0055
	Reg × Word	(8, 32)	4.818	0.003	0.0006	(8, 32)	0.151	0.001	0.9956

Probability p that the hypothesis that of a zero difference between the F0 gradients of Registers 1 and 2 is true. The analysis was performed in four sectors of each time-normalized utterance duration. Results are shown for a group of 25 speakers and four subgroups of those speakers.

contours and intensity contours, with 1 semitone in F0 corresponding to about 1 dB in SPL. This rule has a support in our data on prose reading.'

To this may be added another possibility on the assumption of a state of linguistic flux in which a transition from registers to tones is not complete. To the extent that breathy voice *is* a property of Register 2, especially for the women, one would expect

Table 4. Vowel duration

Groups	Avg. Reg. 1 s	Avg. Reg. 2 s	Mean diff. s	Variables	F(d.f.)	F	MSe	p
All speakers	0.2491	0.2714	0.022	Register	(1, 25)	79.871	0.058	0.0001
				Word	(8, 200)	469.60	0.595	0.0001
				Reg × Word	(8, 200)	12.242	0.003	0.0001
Male	0.2126	0.2374	0.025	Register	(1, 8)	38.084	0.025	0.0003
				Word	(8, 64)	268.688	0.167	0.0001
				Reg × Word	(8, 64)	3.971	0.001	0.0007
Female	0.2696	0.2905	0.021	Register	(1, 15)	37.522	0.031	0.0001
				Word	(8, 120)	303.169	0.418	0.0001
				Reg × Word	(8, 120)	8.30	0.002	0.0001
Old	0.2408	0.2685	0.028	Register	(1, 5)	44.035	0.021	0.0012
				Word	(8, 40)	85.208	0.155	0.0001
				Reg × Word	(8, 40)	3.028	0.001	0.0093
Young	0.2388	0.2597	0.021	Register	(1, 4)	22.001	0.010	0.0094
				Word	(8, 32)	68.421	0.107	0.0001
				Reg × Word	(8, 32)	7.028	0.001	0.0001

Given a hypothesis that the durations of vowels in Register 1 are drawn from the same population as corresponding values in Register 2, the column headed by the symbol p shows the probabilities that the hypothesis is true for different groups of speakers.

Table 5. Formant frequencies

Groups	Analysis of Variance							
	Variables	F(d.f.)	formant 1			formant 2		
			F	MSe	p	F	MSe	p
All speakers	Register	(1, 24)	2.60	5,287.21	0.1197	1.185	121,407	0.2871
	Word	(8, 192)	163.86	21E+5	0.0001	160.496	19E+6	0.0001
	Reg × Word	(8, 192)	8.75	16,629.4	0.0001	0.779	82,740.1	0.6213
Male	Register	(1, 8)	1.56	598.965	0.2469	0.245	523.297	0.6337
	Word	(8, 64)	157.27	70E+5	0.0001	293.95	46E+5	0.0001
	Reg × Word	(8, 64)	1.55	603.80	0.1568	6.531	10,462.1	0.0001
Female	Register	(1, 15)	4.66	11,934.9	0.0475	1.100	175,047	0.3109
	Word	(8, 120)	80.56	14E+5	0.0001	96.098	15E+6	0.0001
	Reg × Word	(8, 120)	9.18	21,440.9	0.0001	0.866	142,372	0.5469
Old	Register	(1, 5)	0.34	260.183	0.5868	0.753	32E+4	0.4253
	Word	(8, 40)	25.80	39E+4	0.0001	14.226	51E+5	0.0001
	Reg × Word	(8.40)	1.59	2,905.683	0.1574	0.868	37E+4	0.5508
Young	Register	(1, 4)	1.11	1,123.03	0.3513	1.362	2,316.18	0.3081
	Word	(8, 32)	34.60	49E+4	0.0001	65.92	37E+5	0.0001
	Reg × Word	(8, 32)	3.23	1,395.06	0.0084	2.947	4,686.05	0.0138

Given a hypothesis that formant 1 and formant 2 frequencies in Register 1 are indistinguishable from their corresponding values in Register 2, the columns headed by the symbol p show the probabilities that the hypothesis is true for different groups of speakers.

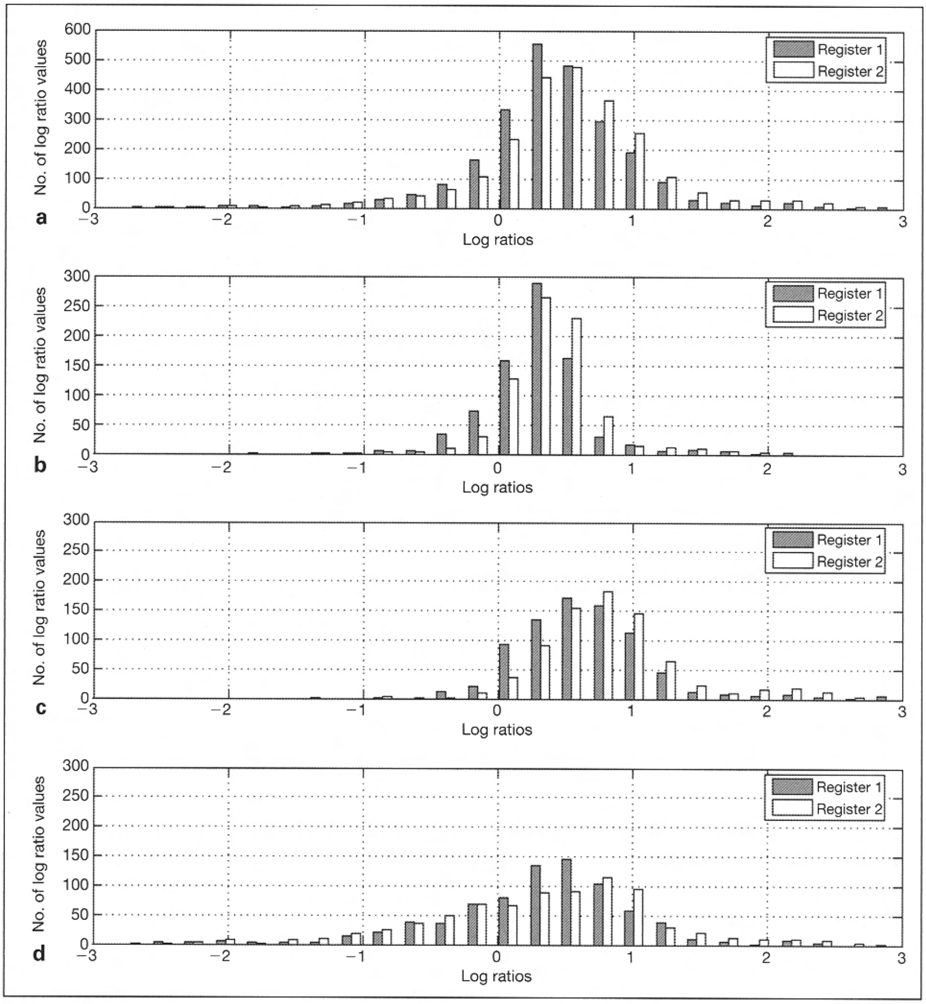

Fig. 11. Distributions of log ratio values after the elimination of all instances of synchronization failure prior to the application of an ANOVA. **a** Combined distribution of all ratio values. **b** Distribution of H_2/H_1 ratio values. **c** Distribution of H_{F1}/H_1 ratio values. **d** Distribution of H_{F2}/H_1 ratio values.

weak medial compression of the vocal folds with a resultant small opening along much of the glottis, to cause, through wasted air in less efficient phonation, a drop in the amplitude of the radiated signal [Laver, 1980, pp. 132–135]. No support, however, appears for such a difference between the men and women in table 2. A study of the clear and breathy vowels of Gujarati [Fischer-Jørgensen, 1970, pp. 101–103] yields inconsistent differences in amplitude between the two categories.

The foregoing material gives us little reason to believe that our significantly different amplitude contours give evidence of voluntary control over an independent mechanism in the production of the registers.

Table 6. Harmonic intensity ratios

Groups	Source	F(d.f.)	H_2/H_1			H_{F1}/H_1			H_{F2}/H_1		
			F	MSe	p	F	MSe	p	F	MSe	p
All speakers	Register	(1, 23)	14.681	1.851	0.0009	12.049	4.452	0.0017	1.327	0.385	0.2611
	Word	(8, 184)	5.268	0.349	0.0001	12.769	1.542	0.0001	78.207	13.820	0.0001
	Reg × Word	(8, 184)	1.601	0.085	0.1269	0.840	0.043	0.8247	3.790	0.324	0.0068
Male	Register	(1, 7)	0.621	0.073	0.4566	4.806	0.754	0.0645	0.802	0.138	0.4002
	Word	(8, 56)	6.092	0.184	0.0001	14.406	0.599	0.0001	56.952	4.983	0.0001
	Reg × Word	(8, 56)	3.306	0.044	0.0037	2.412	0.045	0.0259	3.962	0.116	0.0009
Female	Register	(1, 15)	19.442	2.176	0.0005	8.508	3.882	0.0106	0.678	0.247	0.4232
	Word	(8, 120)	5.508	0.396	0.0001	7.740	1.000	0.0001	41.816	9.044	0.0001
	Reg × Word	(8, 120)	1.394	0.099	0.2061	0.646	0.069	0.7380	2.771	0.307	0.0404
Old	Register	(1, 5)	1.522	0.181	0.2721	2.636	1.195	0.1654	0.123	0.034	0.7397
	Word	(8, 40)	1.994	0.097	0.0723	5.137	0.414	0.0002	17.682	2.599	0.0001
	Reg × Word	(8, 40)	1.185	0.019	0.3319	0.792	0.048	0.6127	0.959	0.060	0.4809
Young	Register	(1, 3)	16.768	0.820	0.0263	6.658	1.296	0.0818	1.537	0.510	0.3033
	Word	(8, 24)	1.125	0.065	0.3821	8.627	0.587	0.0001	25.362	3.097	0.0004
	Reg × Word	(8, 24)	2.037	0.080	0.0850	1.741	0.074	0.1400	3.225	0.198	0.0124

Given a hypothesis that the log-transformed ratios H_2/H_1, H_{F1}/H_1, and H_{F2}/H_1 are independent of voice register, the columns headed by the symbol p show the probabilities that the hypothesis is true for different groups of speakers.

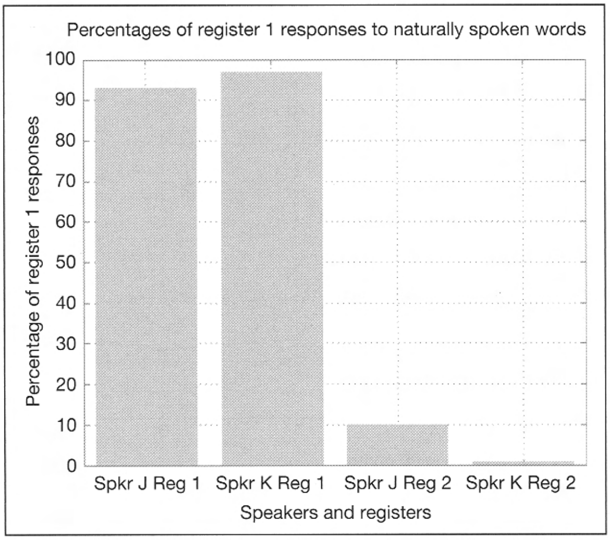

Fig. 12. The register identification scores of 32 listeners who heard a randomized presentation of natural exemplars of the two registers produced by male and female speakers.

Fundamental Frequency

Our acoustic analysis of the speech of a good sampling of the Khmu' populace of Huay Sataeng revealed F0 to be the most salient and consistent property differentiating the two voice registers. Register 1 is uttered on a higher pitch than Register 2. As for the shapes of the contours, the gradual drop toward the end of each of them is most likely

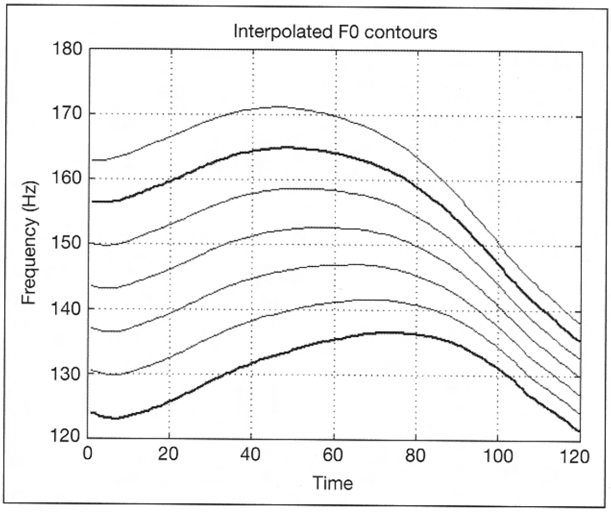

Fig. 13. The series of F0 contours imposed on the synthesized syllable [raːŋ]. The thick lines show the limits of the set before the addition of the super stimulus (upper line).

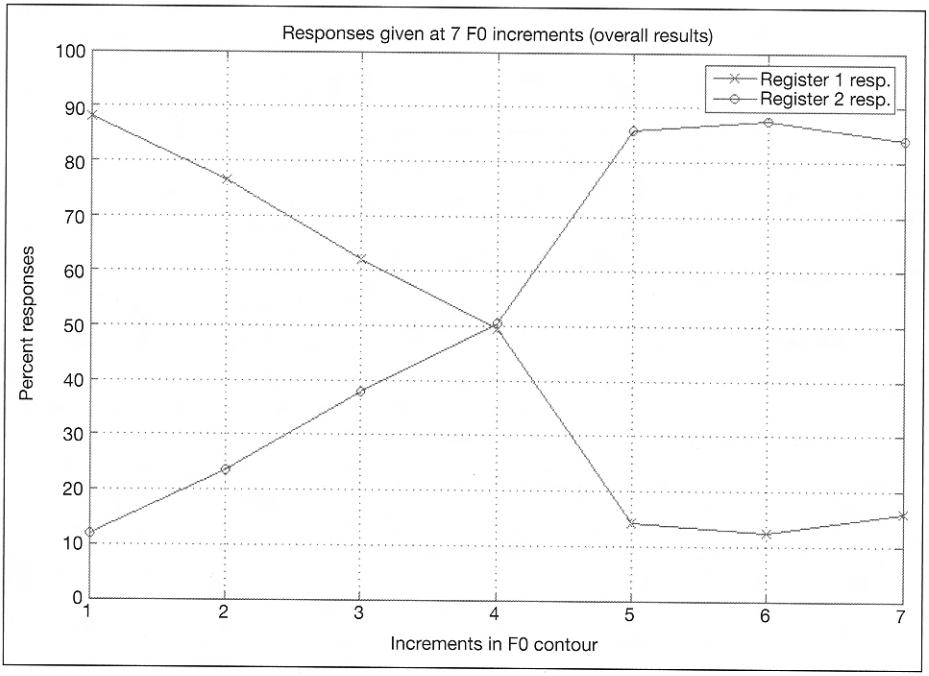

Fig. 14. The percentages of Register 1 vs. Register 2 responses of 29 listeners to incremental changes in the F0 contours of synthesized versions of the Khmu' syllable [raːŋ].

merely a sentence-intonational fall, since the words were recorded one by one as citation forms. Perceptual validation of this pitch difference was obtained from the results of our experiment with synthetic speech. The fact that the peaks of the two identification functions did not quite reach 90% is, we believe, either an artifact of the substandard testing conditions or, perhaps, our failure to make the stimuli natural enough. In particular, to the extent that breathy voice still occurs as a property of Register 2 in the community, the complete absence of it in the stimuli at that end of the pitch series may have been dissatisfying. The same subjects achieved somewhat higher peaks in our control test with natural speech, especially with the utterances of speaker K, a woman.

Vowel Duration

The vowels of Register 2 are longer than those of Register 1. Although the difference is significant, it is so small that it is psychoacoustically unlikely to have value as an acoustic cue. At first blush, one might assume that this is simply a matter of covariation of F0 height and duration like the apparent covariation of F0 and amplitude. Indeed, as reported by Gandour [1977], several studies have demonstrated the effects on duration of the height and shape of F0 contours. One tone language in which low F0 is found with longer vowels is Thai [Abramson, 1962, p. 107; Gandour, 1977]. This is an especially interesting example, because Thai has a phonological distinction between short and long vowels. This distinction apparently suffers no interference in the current state of the language from the slightly greater duration of vowels with low F0, although, as shown by Gandour [1977], historically there has been interaction between tones and vowel length in some dialects of Thai.

We are concerned here with registers that are traditionally described as differing not only in pitch but also in phonation type; therefore, it is interesting to note that Fischer-Jørgensen [1970, pp. 92–95] finds the murmured vowels of Gujarati to be longer than the clear ones. This is, she says [p. 138] 'probably due to the historical origin of the murmured vowels as a fusion of vowel with [h]...'. On the other hand, Thongkum [1989, 1991] thinks that length distinctions between voice registers are much less likely in languages with phonemic distinctions between short and long vowels. Thus, in the Kui dialect of Suai she finds the difference limited to short vowels, while in Chong there is no length difference between the clear and breathy registers. In addition, in our recent study [Abramson et al., 2004] of the Kuai dialect of Suai, which has distinctive vowel length, once our duration data had been corrected for the presence of consonantal aspiration in Register 1 but not Register 2, there was no length difference between the registers.

In an interesting paper Alan C.L. Yu [in press] presents evidence for an explanation of the duration phenomenon. Drawing on work in the psychoacoustic literature with pure tones as well as his own experiments with synthetic speech, he reports that stimuli with a low F0 are heard as shorter than those of the same duration produced with a higher F0. Consequently, he explains the contrary finding of longer durations at lower F0s to be the effect of hypercorrection on the part of speakers to make syllables of rather different pitches sound equally long. There are some complications when it comes to pitting 'flat tones' against 'dynamic tones'. The latter would seem irrelevant to our Khmu' contours, which show little movement before the final intonational dip. If Yu is right, we can only suppose that speakers start making this adjustment as children during the acquisition of language and that it eventually becomes automatic.

It is true, of course, that a definitive way of testing the hypothesis that vowel duration is indeed relevant to the register distinction would be by means of perceptual experiments employing durational variants. In the light of the foregoing discussion together with the smallness of the mean difference of 22 ms in our data, and the presence of a phonemic vowel-length difference independent of register in Khmu', the probability of length as perceptually relevant to the register distinction is remote.

Formant Frequencies

In a register language one of the properties emerging over time to distinguish the lexical items in the registers could be diverging vowel quality. In the extreme case, a vowel of an earlier stage of the language could split into two vowel phonemes. In a less extreme case, the divergence could yield subphonemic differences that are concomitant with other properties of each register. For our Khmu' Rawk data the vowels in the minimal pairs studied are phonologically the same; however, even a significant subphonemic difference would have rendered invalid our method of determining acoustic indices of phonation-type differences through intensity ratios of harmonics. Our finding of no significant difference between the formant frequencies of the two registers failed to invalidate this metric.

Harmonic Intensity Ratios

Our analysis shows a significant difference in one of the three acoustic indices of spectral tilt for the women in our study but none for the men. In the sociolinguistic literature one comes across references to linguistic conservatism among women. For the most part, the women of Huay Sataeng limit their activities to the village, while the men largely work elsewhere in the vicinity where they are in contact with Northern Thai, Standard Thai, or Lao. Even so, the men do chat with the women, and it seems likely that they are sensitive to such phonation-type differences as remain. Thus, phonation type has weakened as a differentiator of the voice registers but seems not to have disappeared altogether.

Conclusion

Among the regional dialects of Khmu', Khmu' Rawk belongs to the subgroup known to have developed phonologically relevant voice registers. It, however, at least as represented by the variety spoken in Huay Sataeng, appears to be in a late transitional stage. It seems that the registers are dying out and being replaced by tones. Perhaps a better way of putting it is that one phonetic property of the traditional registers, pitch, now dominates for the community as a whole. Indeed, although we are not native speakers of the language, the two of us who are experienced field workers have listened many times to the recordings of our 25 speakers and are struck *mainly* by the pitch difference, even in the speech of the women, who seem not to have entirely lost phonation-type differences. The men and women of the village do certainly talk to each other and may be aware of qualitative differences between male and female speech. Pitch is what we normally take to be the essential property of a tonal system, although for one or more of the tones in a language taken to be tonal there may be other concomitant properties. The frequent difficulty in specifying exactly at what point along the evolutionary scale of tonogenesis a language stands, calls our attention to the unclear typological boundary between voice-register languages and tone languages. We were not able to find enough very old subjects to make our examination of subgroups by age very fruitful.

Unfortunately, at the time when we were preparing our perceptual experiments, we had not yet found the differences in one harmonic intensity ratio among the women, so we did not test for phonation type in our synthetic stimuli. We now hope to follow up this study with appropriate new listening tests. Pending further research, we must conclude then that Khmu' Rawk, at least the variety spoken in Huay Sataeng, has either shifted from a register-complex system to a tonal system or is on the brink of doing so.

Acknowledgments

This work was supported by NIH Grant DC-02717 to Haskins Laboratories and a grant from the Thailand Research Fund to the third author. We thank Ms. Chommanad Intajamornrak and Ms. Supakorn Panichkul for their making of the digital sound recordings in rather difficult circumstances. The excellent help of Ms. Chommanad Intajamornrak and Dr. Phinnarat Akharawatthanakun in handling the logistics of the perception testing in Huay Sataeng is much appreciated. We also wish to acknowledge the assistance of Dr. Steven Frost in the development of our Psyscope designs and Dr. Christine Shadle from whom we gained the benefit of a discussion about the interactions between air pressure, turbulence, amplitude, and fundamental frequency. Two anonymous reviewers made very helpful comments. As always, the first author's strong ties to the members of the Department of Linguistics, Faculty of Arts, Chulalongkorn University in Bangkok greatly facilitated matters during his stay for part of the research.

Appendix

The Phonemes of Khmu' Rawk

Consonants								
Initials					Examples of initial consonant clusters			
p	t	c	k	ʔ	C_1C_2			
ph	th	ch	kh		mp nr ns sʔ			
ʔb/ʔm	ʔd/ʔn				$C_1C_2C_3$			
m	n	ɲ	ŋ		ntr nsr ŋkw mpl			
f	s			h				
	r							
	l							
w		j						
		ʔj						
Finals								
p	t	c	k	ʔ				
m	n	ɲ	ŋ					
w		j						
		ç/jh		h				
	r							
	l							
Vowels								
i	e	ɛ	ɨ	ə	a	u	o	ɔ
ii	ee	ɛɛ	ɨɨ	əə	aa	uu	oo	ɔɔ
iə	ɨə	uə						

Registers: R1 = clear voice, high pitch; R2 = breathy voice, low pitch.
(Considerable variation as to presence of phonation types.)

References

Abramson, A. S.: The vowels and tones of Standard Thai: acoustical measurements and experiments. Indiana University Research Center in Anthropology, Folklore, and Linguistics, Pub. 20, Bloomington. Int. J. Am. Ling. *28:* No. 2 (1962).

Abramson, A. S.: The plausibility of phonetic explanations of tonogenesis; in Fant, Fujisaki, Cao, Xu, From traditional phonology to modern speech processing: Festschrift for Professor Wu Zongji's 95th birthday, pp. 17–29 (Foreign Language Teaching and Research Press, Beijing 2004).

Abramson, A.S.; L-Luangthongkum, T.; Nye, P.W.: Voice register in Suai (Kuai): an analysis of perceptual and acoustic data. Phonetica *61:* 147–171 (2004).

Brunelle, M.: Register in Eastern Cham: phonological, phonetic and sociolinguistic approaches; doct. diss. Cornell University, Ithaca (unpublished, 2005).

Cohen, J.D.; MacWhinney, B.; Flatt, M.; Provost, J.: PsyScope: a new graphic interactive environment for designing psychology experiments. Behav. Res. Methods Instrum. Comput. *25:* 257–271 (1993). See also http://psyscope.psy.cmu.edu/.

Fant, G.; Kruckenberg, A.: Analysis and synthesis of Swedish prosody with outlooks on production and perception; in Fant, Fujisaki, Cao, Xu, From traditional phonology to modern speech processing: Festschrift for Professor Wu Zongji's 95th birthday, pp. 73–95 (Foreign Language Teaching and Research Press, Beijing 2004).

Fant, G.; Kruckenberg, A.; Liljencrantz, J.: The source-filter frame of reference. Phonetica *57:* 113–127 (2000).

Ferlus, M.: Formation des registres et mutations consonantiques dans les langues Mon-Khmer. Mon-Khmer Stud. *8:* 1–76 (1980).

Fischer-Jørgensen, E.: Phonetic analysis of breathy (murmured) vowels in Gujarati. Indian Linguistics *28:* 71–139 (1970).

Flanagan, J.L.: A difference limen for vowel formant frequency (Letter). J. acoust. Soc. Am. *27:* 149–174 (1955a).

Flanagan, J.L.: Difference limen for the intensity of a vowel sound (Letter). J. acoust. Soc. Am. *27:* 1223–1225 (1955b).

Gandour, J.: On the interaction between tone and vowel length: evidence from Thai dialects. Phonetica *34:* 54–65 (1977).

Gandour, J.T.; Gårding, E.; Lindell, K.: Tones in Northern Kammu: a perceptual investigation. Acta orientalia *39:* 181–189 (1978).

Gårding, E.; Lindell, K.: Tones in Northern Kammu: a phonetic investigation. Acta orientalia *38:* 321–332 (1977).

Gordon, M.; Ladefoged, P.: Phonation types: a cross-linguistic overview. J. Phonet. *29:* 382–406 (2001).

Henderson, E.: The main features of Cambodian pronunciation. Bull. School Orient. Stud. *17:* 149–174 (1952).

Laver, J.: The phonetic description of voice quality (Cambridge University Press, Cambridge 1980).

L-Thongkum,[1] T.; Teeranon, P.; Intajamornrak, C.: The interaction between vowel length and pitch in Southeast Asian (SEA) languages: implications for tonogenesis; in Harris, Burusphat, Harris, Studies in Tai and Southeast Asian linguistics, pp. 335–340 (Ekphim Thai, Bangkok 2007).

Mermelstein, P.: Difference limens for formant frequencies of steady-state and consonant-bound vowels. J. acoust. Soc. Am. *63:* 572–580 (1978).

O'Shaughnessy, D.: Speech communications: human and machine (IEEE Press, Piscataway 2000).

Premsrirat, S.: Khmu, a minority language of Thailand (a Khmu grammar and a study of Thai and Khmu cutting words). Papers in Southeast Asian Linguistics, No. 10, Pacific Linguistic Series A, No. 75 (1987).

Premsrirat, S.: Dictionary of Khmu in Thailand (Institute of Language and Culture for Rural Development, Mahidol University, Nakhon Pathom 2002).

Premsrirat, S.: Khmu dialects: a case of register complex and tonogenesis; in Kaji, Proc. Symp. Cross Language Studies of Tonal Phenomena: Historical Development, Phonetics of Tone, and Descriptive Studies, 2002, pp. 13–36 (Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies, Tokyo 2003).

Smalley, W.A.: Outline of Khmuʔ structure (American Oriental Society, New Haven 1961).

Stevens, K.N.: Acoustic phonetics (MIT Press, Cambridge 1998).

Svantesson, J.-O.: Tonogenesis in Southeast Asia – Mon-Khmer and beyond; in Kaji, Proc. Symp. Cross Language Studies of Tonal Phenomena: Tonogenesis, Japanese Accentology, and Other Topics, 2000, p. 58 (Research Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies, Tokyo 2001).

Svantesson, J.-O.; House, D.: Tone, production, tone perception and Kammu tonogenesis. Phonology *23:* 309–333 (2006).

Thongkum, T.L.: Phonation types in Mon-Khmer languages; in Fujimura, Vocal fold physiology: voice productions, mechanisms and functions, vol. 2, pp. 319–333 (Raven Press, New York 1988).

Thongkum, T.L.: An acoustical study of the register complex in Kui (Suai). Mon-Khmer Stud. *15:* 1–19 (1989).

Thongkum, T.L.: The interaction between pitch and phonation type in Mon: implications for a theory of tonogenesis. Mon-Khmer Stud. *16–17:* 11–24 (1990).

Thongkum, T.L.: An instrumental study of Chong registers; in Davidson, Austroasiatic languages: essays in honor of H.L. Shorto, pp. 141–160 (School of Oriental and African Studies. University of London, London 1991).

Thurgood, G.: Vietnamese and tonogenesis: revising the model and the analysis. Diachronica *19:* 333–363 (2002).

Thurgood, G.: Tonogenesis revisited: reviving the model and the analysis; in Harris, Burusphat, Harris, Studies in Tai and Southeast Asian linguistics, pp. 241–262 (Ekphim Thai, Bangkok 2007).

Wayland, R.; Jongman, A.: Acoustic correlates of breathy and clear vowels: the case of Khmer. J. Phonet. *21:* 181–201 (2003).

Yu, A.C.L.: Tonal effects on perceived vowel duration; in Fougeron, Kühnert, D'Imperio, Vallée, Papers in Laboratory Phonology X: Variation, phonetic detail and phonological modeling (Mouton de Gruyter, Dordrecht, in press).

[1] In the References, L-Thongkum, T. and Thongkum, T.L. are the same person, Theraphan Luangthongkum.