1367

Original Paper

Phonetica 2004;61:147–171
DOI: 10.1159/000082561

Received: August 2, 2004
Accepted: November 7, 2004

Voice Register in Suai (Kuai): An Analysis of Perceptual and Acoustic Data

Arthur S. Abramson[a, b] Theraphan L-Thongkum[c] Patrick W. Nye[a]

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

Abstract

Analyses of the perceptual and acoustic characteristics of the Register 1 ('clear') versus Register 2 ('breathy') distinction have been carried out on the Kuai dialect of Suai, a Mon-Khmer language. The perception results were obtained from five-parameter synthesized stimuli. They showed that the primary parameter underlying the distinction is the frequency of onset of laryngeal excitation (F0). One other parameter making a significant contribution was the open quotient. The F0 result was confirmed by an acoustic analysis of eight pairs of natural utterances produced by native speakers. We conclude that the Suai language is in a state of flux with respect to the voice registers, although the distinction has not disappeared. The perceptual data reveal mixed levels of sensitivity, and the production data indicate that some speakers maintain a fairly good distinction, while others do not. The language seems to be replacing the register distinction with a prosodic one of pitch accent, possibly as a stage leading to tonogenesis.

Introduction

A register language can be defined as a language with lexically distinctive voice registers, mostly two-way distinctions: clear or modal voice vs. breathy voice as their dominant features. Some Mon-Khmer languages, such as Chong [Thongkum, 1988, 1991], have as many as four voice registers. Henderson [1952] was the first linguist to use the term 'register' as a phonological concept. It is a word that describes a cluster of laryngeal and supralaryngeal activities. Thus, the term register is generally understood to mean a 'register complex' one property of which may be dominant and the rest secondary. The complex of phonetic characteristics typically includes such features as phonation type, pitch, vowel quality, vowel length, loudness, and perhaps others. Defined in this way, the construct register seems to offer a useful conceptual framework for the analysis of such phonological distinctions. However, despite the fact that several

KARGER

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

Accessible online at:
www.karger.com/journals/pho

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 abramson@uconnvm.uconn.edu

Table 1. Examples of Suai words conventionally distinguished by registers

Minimal pairs			
R1 (modal)	gloss	R2 (breathy)	gloss
luː	to howl	lṳː	thigh; the lap
luːm	a mouthful	lṳːm	to gobble chunks of food
Quasi-minimal pairs with initial voiceless stops			
R1 (modal)	gloss	R2 (breathy)	gloss
tʰɔŋ	to remember by heart	tɔ̤ŋ	to hold
tʰeː	jar	te̤ː	no

acoustic studies of register phenomena [e.g., Ladefoged et al., 1988; Thongkum, 1988, 1989] have appeared since the concept was first introduced, to the best of our knowledge, no perceptual experimentation on phonologically distinctive voice registers has been reported in the literature.[1] Hence, wishing to embark on such research, we chose the relatively simple case of a language, Suai, that has just two registers traditionally labeled Register 1 ('modal' or 'clear') and Register 2 ('breathy'). Suai of the Mon-Khmer family is spoken in several northeastern provinces of Thailand, chiefly Surin, Buriram, Sisaket Ubonratchathani, Mahasarakham, and Nakhonratchasima [Sriwises, 1978]. Some Suai people have migrated from the northeastern region of Thailand to Suphanburi province in the central region. The language comprises two major dialects, Kui and Kuai.

This paper focuses on the Kuai dialect of Suai. Recordings of words that engage the register distinction were made by native speakers of Kuai from the village of Samrong in Surin province of northeastern Thailand. The phonation types are salient on the vowel of a syllable, but coarticulatory effects are also discernable in the surrounding consonants and vowels. Examples of minimal pairs are given in the upper section of table 1. The lower section of table 1 illustrates quasi-minimal pairs resulting from diachronic effects on initial stop consonants in Kuai. In the other major dialect, Kui, both members of the pairs have initial aspirated stops.

Yantreesingh [1980] has studied the phonology of Kuai, but we have relied on the more readily available recent analysis of Sukkasame [2003, pp. 166–169; although this work is dated 2003, it did not become publicly available until 2004, several years after the inception of our study] and present his phonemic inventory, translated from the Thai in 'Appendix 1'; we have replaced a few symbols to agree with our transcription system. Through his extensive sociolinguistic study of Suai, using auditory phonetic analysis of speakers in 12 age groups, Sukkasame [2003] finds that the Kuai dialect still has the register distinction, although it may not be very stable. He does predict [Sukkasame, 2003, p. 246] that the variety of Kuai spoken in our village of Samrong will develop four tones in the future: high rising falling, high level, low rising falling, and low rising. The Kui dialect, he asserts, is also clearly moving toward a tonal system, with younger speakers leading the way.

---

[1] A possible exception is Bickley's [1982] synthesis experiments with speakers of Gujarati. The contrast between clear and breathy vowels of that language might indeed be equivalent to a voice-register contrast.

The sociolinguistic setting of the village of Samrong is described by Sukkasame [2003, pp. 164–166]; it fits well with our own much briefer observations. The village, in its two subdivisions, includes 276 households with a population of 1,380 people, who are mostly farmers and laborers. The level of education attained varies with age: 30–60, grade 4; 20–30, grade 6; below 20, grades 9 and 12. The people of Samrong are trilingual in Suai (Kuai), Northeastern Thai (Lao Isaan; Thai and Lao are closely related members of the Tai language family), and Thai, the national language. They speak Kuai among themselves, Lao Isaan to local people of Thai ethnicity, and Thai in classrooms, government offices, and in regions outside the northeast (Isaan). Young people feel that it is more prestigious to speak Northeastern Thai. In most of the neighboring villages the language is Northeastern Thai. In Kham, the village to the south, the language is the Kui dialect of Suai.

We ourselves had a warm welcome in Samrong. The people, who had had little or no experience with curious linguists and anthropologists and certainly none with perception tests, were delighted by our scientific interest in their language and culture. Suai itself is not known to them in any written form, but all the younger people are literate in Thai. Some above the age of 60 are not.

Our original plan was just to do a perceptual study by manipulation of the apparently relevant parameters of a speech synthesis computer program. However, some surprising features in the resulting data led us also to undertake an acoustic analysis of the utterances of several speakers of Kuai.

Members of word-pairs conventionally described as being differentiated by the two voice registers were randomized and recorded in citation form by native speakers of Kuai. After down-sampling from 22,050 samples per second to a rate of 11,025, the pairs of words were prepared for analysis. At this juncture the analysis procedures applied to the words followed both perceptual and instrumental paths. A brief preliminary account of some of this work is to be found in Abramson [2004].

Methods

Speech Synthesis

As models for our perceptual experiments, words spoken by 3 native speakers of Kuai were used. Acoustic parameters including formant frequencies, fundamental frequency (F0), amplitude of turbulence, and spectral tilt were used to resynthesize the vowels contained in each word-pair. The parameter extraction and resynthesis was achieved by SynthWorks® software produced by Scicon Research and Development (http://www.sciconrd.com). Our basic stimulus 'luː 1' was a resynthesis of an utterance of our principal informant in Register 1 (clear voice). Using the cascade/parallel configuration [Klatt, 1980], a sequence of 50 other stimuli ('luː 2'–'luː 51') was derived from 'luː 1' in the manner described below. The frequency courses and bandwidths of the five formants of 'luː 1'were readily heard as the syllable [luː] and identified as the Kuai word in Register 1 meaning 'to howl'. For additional stimuli that might yield not only that word but also its counterpart in Register 2, [lṳː] 'thigh', we made use of the following five parameters of the synthesis program [Klatt and Klatt, 1990]:

ATU. Amplitude of turbulence (breathiness); choices: 40, 65, and 70 dB; default: 40 dB.

OQ. Open quotient of voicing waveform; lengths: 20, 40, and 60 samples; default: 20.

F0. Fundamental frequency contour; start at 70, 101, and 135 Hz; default: 101 Hz.

TL. Spectral tilt of 0, 18, and 24 dB; applied to output above 3 kHz; default: 0 dB.

TG. Time of 4–60 dB gain increase; periods range 0, 130, and 200 ms; default: 0 ms.

In 'luː', with a duration of 580 ms, the parameters OQ, TL, and TG were set at their default values. The F0 contour of luː 1 followed the pattern of our model speaker: it began at 101 Hz and reached 120 Hz at 310 ms where it remained for a further 40 ms and then fell slowly to 99 Hz at the 580 ms point.

In all other stimuli whose F0 trajectories were altered, the F0 parameter began at 70 or 135 Hz, moved to 120 Hz at 350 ms and then followed the F0 path taken by 'luː l'. Meanwhile, for the purpose of synthesizing the stimulus set, the first five formant frequencies extracted from the basic stimulus 'luː l' were used. For each parameter the two departures from the default setting were empirically chosen by 2 of the investigators who endeavored to select intervals that were perceptually equal.

The three parameters ATU, OQ, and TL were chosen because of their obvious relevance to laryngeal control for the difference between modal and breathy voice, properties conventionally viewed as paramount differentiators of Register 1 and Register 2. F0 was chosen because of its frequently asserted association with the differentiation of the two registers. As for TG, not only the mention of loudness in the literature but also the attested role of F0 made us suspect that an important aspect of the register distinction could be relative auditory salience, a kind of accentual prominence, which could well entail the dimension of overall amplitude.

Stimulus Generation

Given five parameters, each of which provided three levels of adjustment, in principle it was possible to generate 243 different stimuli, a total number that was deemed too large to be delivered to inexperienced listeners. Hence, a subset of all possible combinations was selected which permitted level changes by parameters taken only one and two at a time and left all other parameters set to their default values. By this strategy the total number of tested stimuli was reduced to a more manageable 51. Full details of the parameter settings used to produce the synthetic stimuli are given in 'Appendix 2'.

Stimulus Presentation

The stimuli were recorded for presentation to 16 native speakers of Kuai in the village of Samrong who were paid to participate but voluntarily donated half their earnings to the village elementary school. Their ages ranged from 15 to 45 years. Nine were men and 7 were women. Three tests, each containing a randomization of the 51 stimuli, were reproduced through headphones in the language laboratory of the local school. The language has no standard written form, but, since the listeners were all literate in Thai, for each stimulus they decided whether it was the word for 'to howl' or 'thigh' and checked the box next to the appropriate Thai gloss on the answer sheet. To help assure their understanding we also posted cartoons of a dog with an uplifted snout and Miss Thailand with an arrow pointing to one of her thighs.

Control Test

Although we had no reason to doubt the validity of the conventional description of the phonological status of the voice registers, we took the precaution of running a control test with a randomization of 5 tokens of the natural productions of each of 4 speakers of the test words. These utterances were assessed to be good exemplars of the register distinction.

Preparation for Acoustic Analysis

Because the recordings that formed the raw material for perceptual tests contained in-field noise they were deemed to be unsuitable for instrumental analysis. Thus, a new set of utterances was recorded containing additional words featuring the same register contrast. The words, culled from a Suai dictionary based on the Kui dialect [Sriwises, 1978] and our own field notes, were presented one at a time as Thai glosses from a randomized list to each speaker for recording in Kuai (a Kui dictionary was used because none has been published specifically for the Kuai dialect). Thus, two members of any pair distinguished by register never followed one another. The list of 11 word-pairs shown in table 2 was recorded 3 times by each speaker. Note that, as seen in, e.g., /samiː/, some Kuai words have an initial unstressed syllable in one of the registers. To facilitate alignment and measurement in all such words, only the register-bearing stressed syllable was subjected to analysis. Each of the word-pairs was examined and both members of a pair were discarded if: (i) a word showed evidence of amplitude limiting, or (ii) the variability of both the first and second formant frequencies in a selected steady-state region of a word (defined below) was in excess of the difference limens 70 and 170 Hz, respectively [Flanagan, 1955; Mermelstein, 1978].

Any violation of these criteria led to the elimination of that word from the data sets of all speakers together with its paired register contrast. The output emerging from this preliminary examination was a

Table 2. List of Suai (Kuai) word-pairs recorded for acoustic analysis

Register 1		Register 2		Comments
transcription	gloss	transcription	gloss	
ciː	to go	ci̤ː	nun	
huːjʔ	to sip	hṳːjʔ	to walk	
kʰɔː	trousers	kɔkɔ̤ː	millipede	
luː	to howl	lṳː	thigh	
saʔmiː	to steam	mi̤ː	rich	
samuːjʔ	ant	kamṳːjʔ	ghost	*
kanaː	path	na̤ː	what	*
pʰuʔ	rotten	pṳʔ	moustache	
tʰaːh	tray	ta̤ːh	to clap	
tʰeː	close in space	te̤ː	empty	*
tʰɔŋ	to recite	tɔ̤ŋ	to collect	

* Mismatched formants revealed by the difference-limen criterion led to the exclusion of these word-pairs.

collection of eight word-pairs, each recorded 3 times by each one of our 6 speakers. Three speakers were female (speakers B, C, E) and 3 were male (speakers A, D, F). Thus, the total number of acoustically analyzed utterances was 288.

The principal analysis software employed in the study was Praat v 4.0.26 downloaded from the Praat web page (http://www.fon.hum.uva.nl/praat/) and executed on a Macintosh computer. Praat scripts were written that automatically extracted the first three formants, F0, and the amplitude contour as a function of time. The results were written into text files that were eventually imported by StatView© 5.0 (SAS Institute, Inc., 1998), a statistical software package. Matlab® (http://themathworks.com) was used to perform additional calculations that will be described later.

Spectrograms of each of the 288 utterances were also generated by Praat and subsequently examined by one of us who, using a mouse and cursor, inserted head and tail markers to identify a segment from each utterance that featured a minimum amount of formant change. These segments of the speech file were of variable length and are referred to here as 'steady-state regions'. The head and tail locations of these regions were recorded in text files together with the names of the sound files to which they belonged.

Amplitude and F0

Glottal frequency values (F0) covering a 75- to 600-Hz range together with overall amplitude trajectories were also obtained by Praat at 10-ms intervals and stored in text files. A program written for execution by Matlab retrieved those trajectories and normalized each one to a length of 100 arbitrary time samples. For each of the two voice registers, grand means of the F0 trajectories (expressed in semitones) were calculated and plotted. Conversion to a semitone scale was performed by the formula $F0_{st} = 3.32 \times 12 \times \log_{10}(F0_{Hz})/\text{base}$, where the base was the mean of the minimum average $F0_{Hz}$ value of each speaker's utterances across the two voice registers. This conversion, done as part of the normalization across speakers, was intended to more satisfactorily approximate listeners' sensation of pitch. Frequency doubling errors and dropouts were identified by eye and corrected whenever they occurred. The corresponding amplitude values were obtained at 7-ms intervals, also by Praat, and mapped onto a 100-point time scale and plotted. The graphs of these data are shown in figures 4 through 9.

Formant Frequencies

For the formant frequency measurements within steady-state segments, Praat employed the Burg LPC algorithm specifying 11 pole-pairs and a window length of 25 ms. Following the application of a

6-dB per octave pre-emphasis filter whose transfer function began its rise from unity at 50 Hz, the first three formant frequencies of the sound files were obtained at 10-ms intervals. The upper formant frequency limit was set at 5 kHz. Occasionally, Praat overlooked or misclassified formants. Hence, as a precaution, an additional set of formant frequency estimates was obtained visually from the pre-emphasized spectrograms. The visual estimates frequently diverged by small amounts from those obtained by Praat, but no systematic differences were found between the analyses conducted on the two sets of formant frequencies. Consequently, both sets of data yielded very similar statistical results.

Ratios of Harmonic Intensities

For each of the steady-state vowel regions, Matlab was employed to compute the intensities of the fundamental (H_1), the second harmonic (H_2), and the harmonic located closest to the peak of the first formant (H_{F1}). The ratios H_2/H_1 and H_{F1}/H_1 were then calculated and the results recorded in a disk file for statistical analysis. Harmonic intensity ratios are a useful metric for assessing the slope of the voice spectrum [Ladefoged et al., 1988]; some investigators [e.g., Shrivastav and Sapienza, 2003] include a comparison of H_1 with the third formant. Large ratios indicate a less steep spectral falloff at high frequencies.

The harmonic intensity extraction procedure began by computing a 512-point spectrum vector for each steady-state data segment. A Hamming window was then applied to each segment that was padded with zeros to a length of 1,024 points. A fast Fourier transform (FFT) was applied to each 1,024-point vector, thus yielding a spectrum vector with a frequency resolution of 21.5 Hz. A cepstrum of the 1,024-point vector was also employed to find the frequency of the harmonic structure in the spectrum vector. Then a sine wave with that same frequency was aligned with the harmonic structure of the spectrum vector by autocorrelation.

Indexes to the locations of the first six peaks of the sine wave were found and used to retrieve the intensities of the fundamental frequency H_1 and its harmonics from the spectrum vector. To find the location of the first formant (F1), an 11-pole-pair linear predictive coding (LPC) analysis was applied to the steady-state speech segment, and the index n of the first formant peak was identified by a simple peak-finding algorithm (i.e., n is the point such that $F1_{n-2} < F1_{n-1} < F1_n > F1_{n+1} > F1_{n+2}$). Then the harmonic lying closest to the F1 peak was found, and the ratios H_2/H_1 and H_{F1}/H_1 were computed and stored. The cepstrum and sine wave were plotted for each of the words, together with the results of the LPC analysis and peak-finding algorithm. These plots were all inspected by eye and, in 4 or 5 cases where obvious errors had occurred, recalculations were made by hand and the results substituted for those obtained by the computer program.

Vowel Duration

Vowel duration was the final acoustic property that was measured because of its potential to differentiate the two voice registers [Thongkum, 1988; Gordon and Ladefoged, 2001]. We examined the eight selected pairs of words with virtually the same formant patterns in each pair. Thus, we hoped to avoid distortions of the data caused by possible correlations between vowel duration and the configuration of the supraglottal vocal tract. Our criteria for defining the span of a vowel were gestural. That is, we measured the duration of the formant pattern of the vowel, including formant transitions, from the release of any initial consonant[2] to the completion of the closing gesture of any final consonant, by inspecting both a wideband spectrogram and the wave form of each utterance. Consequently, the type of excitation of the vowel, or any portion thereof, whether the quasi-periodicity of voicing or the turbulence of aspiration, had no bearing on our measurement of the duration [this is in agreement with an early stand taken by Fischer-Jørgensen, 1954]. In Kuai, aspiration can be manifested in the initial and final portions of the vowel as [h] or in the release phases of the initial aspirated voiceless stop consonants [pʰ tʰ kʰ]. The usual somewhat blurry formant pattern was seen during aspiration. As will be seen in the 'Results' section, we later found cause to present a set of measurements with the initial aspiration omitted. All our words began with an oral consonant or /h/ and ended with a vowel, a vocalic glide, /h/, glottal stop, or a

---

[2] Lacking a release, the sole exception is initial /h/, which is normally treated phonologically as a member of the class of consonants, although phonetically, of course, it may simply be the noise-excited early portion of a vowel. For such words then, we included the aspiration in the duration of the vowel.

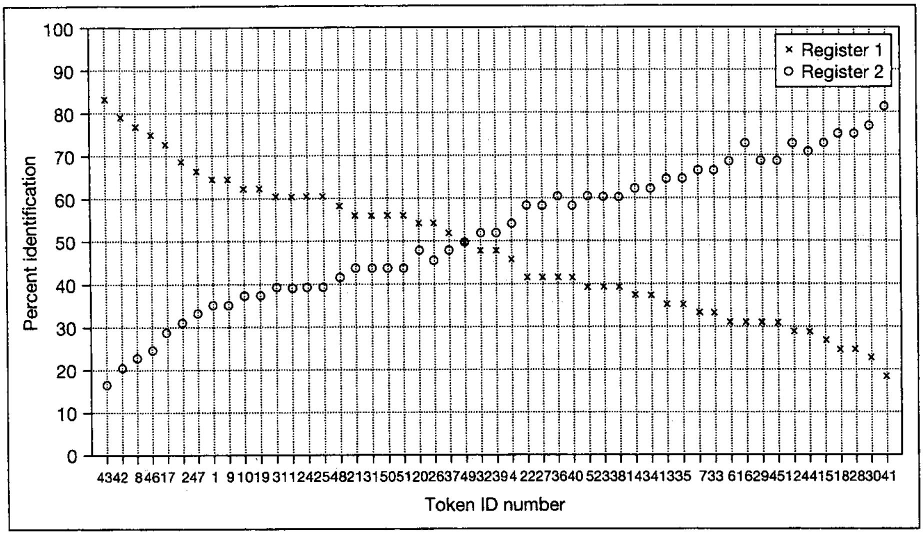

Fig. 1. Response percentages of 16 Suai-speaking listeners to all 51 synthesized stimulus tokens. Percentages plotted along the y axis have been rank ordered. Stimulus tokens are identified by number along the x axis and their parameter values are shown in 'Appendix 1'.

nasal consonant. Our criterion for the termination of a vowel in an open syllable was the detectability of at least two of the first three formants.

Statistical Analysis

A statistical examination of listeners' responses to the synthesized stimuli employed correlation and factor analysis to identify the contribution made to the response variance by each of the five parameters. For the acoustic data, an analysis of variance (ANOVA) procedure was applied in which Word and Register were identified as independent variables. The dependent variables were the following: (i) ratios of the amplitudes of the second harmonic H_2 and H_{F1} (the harmonic lying closest to F1) to the amplitude of the first harmonic H_1; (ii) mean formant frequencies in the steady-state regions of words; (iii) gradients of the normalized F0 and amplitude functions in four regions; (iv) mean levels of the normalized F0 and amplitude trajectories in four regions, and (v) mean vowel durations.

Auditory Phonetics

While processing our perceptual data, we became aware of the possible instability of the Kuai register distinction. We therefore engaged the first two authors, both experienced field phoneticians, to carry out a traditional impressionistic assessment of the utterances of the 6 speakers. This procedure was performed in parallel with the acoustic analyses before the word list had been pruned down to 8 pairs of words. Thus, the phonetic observations were made on each of the 3 repetitions of each member of the 11 original pairs of words shown in table 2. Comments and classifications were written on charts prepared for the purpose. Our focus was on voice quality (modal vs. breathy) and pitch (mid/high vs. low/falling).

Results

Perception Tests

The averaged responses of all 16 listeners are shown in figure 1. The stimulus labels are rank-ordered along the x axis according to an increasing and decreasing number of

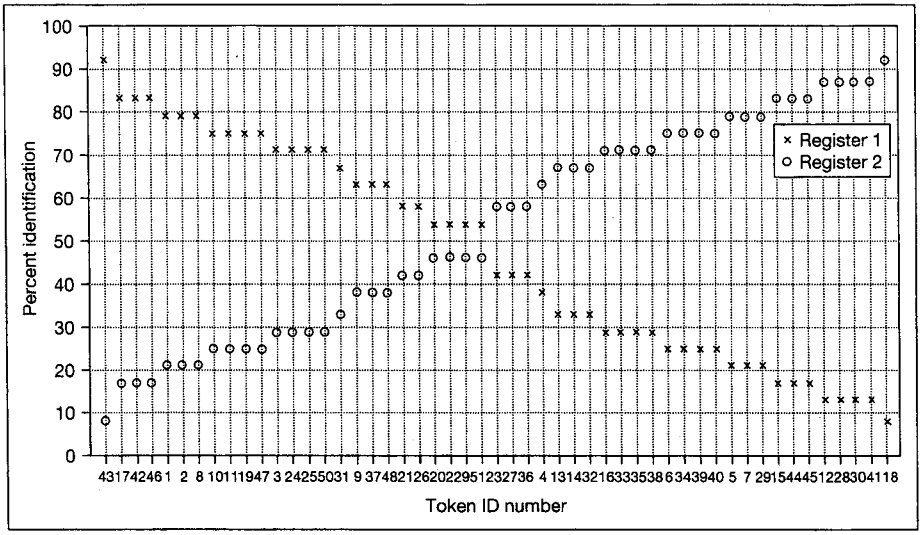

Fig. 2. Response percentages of 8 good Suai-speaking listeners for all 51 synthesized stimulus tokens. Percentages plotted along the y axis have been rank-ordered. Stimulus tokens are identified by number along the x axis and their parameter values are shown in 'Appendix 1'.

'Register 1' (clear) and 'Register 2' (breathy) responses, respectively. Thus, on the y axis we see the percentage of responses in each category. Notwithstanding the resemblance of this graph to identifications along a conventional stimulus continuum, the x axis of the present graph does not represent ordinal stimulus values. Hence, to emphasize this property, the plot points in figures 1–3 are not connected by a continuous line. Underlying the very gradual slopes of the functions in figure 1 are the nearly random performances of a number of listeners. An examination of the data from the control test revealed that only 8 of our listeners had identified the naturally spoken words with a level of accuracy somewhat better than chance (defined as 65% correct). This result appears to add further evidence of a weakening or destabilizing of the distinction within the community.

A display of the overall response data of the 8 'good' listeners who performed well on the control test is shown in figure 2. The response percentages at the extreme ends of the graph are seen to be better than those of figure 1, and the register category boundary is seen to lie in the vicinity of the stimuli labeled 20, 22, 49, and 51. The first two of these ambiguous items contain settings of ATU and TL, and the last two TL and TG. The underlying parameter specifications of the stimulus labels are to be found in the chart of parameter settings ('Appendix 2'). To obtain an estimate of the efficiency of each of the five synthesis parameters as cues to the register distinction, we computed the correlation coefficients of each of the parameters with respect to the response scores plotted in figure 2. The results, shown in table 3, reveal that the F0 parameter provided the most salient cue by contributing 37.7% of the response variance, a figure that substantially exceeded the 16.5% contribution made by the second ranking parameter OQ. The remaining parameters in descending order of effectiveness were TG, TL, and ATU, each contributing less than 5% of the variance.

Table 3. Coefficients of correlation between parameters and response percentages for the 8 good listeners

Synthesis parameter	Correlation coefficient	Percentage of variance
ATU	−0.026	0.1
OQ	0.421	16.5
F0	−0.613	37.7
TL	−0.171	3.4
TG	−0.202	4.8

Table 4. Unrotated factor analysis of the good listeners' response data

Parameters	Factor 1	Factor 2	Factor 3	SMC	3FFE
Response %	0.910	0.024	−0.182	0.561	0.861
ATU	−0.047	0.597	−0.094	0.158	0.367
OQ	0.641	−0.352	0.564	0.358	0.853
F0	−0.667	−0.286	0.583	0.455	0.866
TL	−0.323	−0.610	−0.640	0.211	0.886
TG	−0.287	0.544	0.065	0.202	0.382

The Register 2 response percentages[3] plotted in figure 2 together with the values of each of the five parameters were submitted to a principal components factor analysis. This analysis delivered three factors determined by a rule that stopped computing any additional factors when 75% of the original variance was accounted for. Table 4 shows the values of the unrotated (i.e., orthogonal) factors, the squared multiple correlation (SMC), and the three-factor final estimate of variance (3FFE). The SMC column contains the total proportion of the variance estimated by obtaining the squared multiple correlation of each parameter value with all the other parameter values. Thus, 56% of the response data is predicted by a linear regression on the five synthesizer parameters.

The 3FFE column shows the proportion of the variance captured by the five parameters when all three factors are used to predict the response percentage. As expected, the factor 1 in table 4 reflects the results of the correlation analysis in revealing that dominant contributions to the response variance were made by the F0 and OQ parameters.

In table 5 the values of the three factors have been made more distinctive by performing an oblique rotational transformation. Following such a transformation, the factors are no longer orthogonal but now lend themselves to some plausible categorizations. First, it becomes clear that by virtue of its dominant F0 loading, the first and most important factor can be identified as the contribution made by the F0 onset frequency. Factor 2, on the other hand, contains three loadings of approximately equal weight, namely those derived from the OQ, ATU, and TG parameters. Finally, factor 3 receives its principal loadings from the TL and OQ parameters. In the 'Discussion' section of this paper it will be suggested that, if category names are sought to broadly characterize

---

[3] An arc-sine transformation of the response percentages applied prior to factor analysis did not significantly alter the distribution of parameter loading among the three factors.

Table 5. Rotated factor analysis of the good listeners' response data

Parameters	Factor 1	Factor 2	Factor 3
Response %	0.909	0.404	0.380
ATU	0.084	−0.580	0.019
OQ	0.254	0.741	0.765
F0	−0.921	0.025	0.039
TL	−0.060	0.275	−0.844
TG	−0.216	−0.625	<0.001

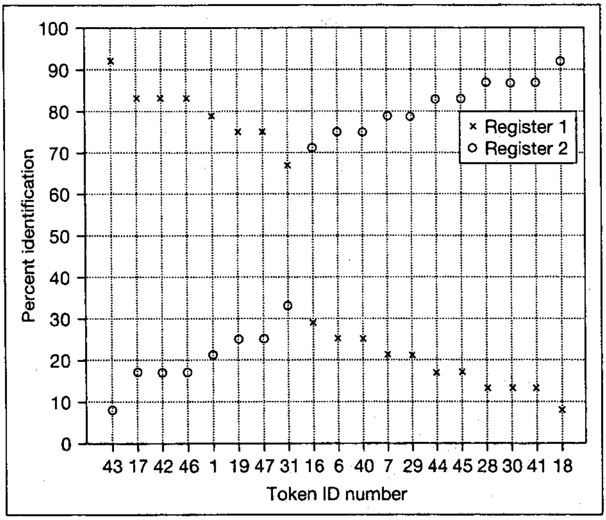

Fig. 3. Rank-ordered response percentages of 8 Suai-speaking listeners to the 19 synthesized stimulus tokens (identified on the x axis) whose F0 contours commenced at either 70, 101 or 135 Hz and reached a common target frequency of 120 Hz after 310 ms.

the roles of the three factors, candidates might be 'F0 Onset Frequency' for factor 1, 'Excitation Amplitude' for factor 2 and 'Aeroacoustic Interaction' for factor 3.

Figure 3 contains a plot of all the responses associated with adjustments made to the F0 parameter. Once again the stimuli are ordered along the x axis in accordance with their response magnitudes. The graph shows that the register responses closely approach 100% at both ends. The 6 rightmost stimuli on the abscissa have in common the fact that their F0 contour rise begins at a frequency of 70 Hz. Meanwhile, the 4 stimuli identified at the leftmost end of the abscissa have in common a starting frequency of 135 Hz. Hence, the higher the F0 frequency at the start of its contour the greater is the listener tendency to label the stimulus as belonging to Register 1.

Ratios of Harmonic Intensities

The ratios obtained from the three repetitions of each utterance were first averaged, thus producing a ratio data set comprised of 8 words × 2 registers × 6 speakers. An ANOVA with Word and Register treated as repeated measures was then performed and the results presented in table 6. The table shows that the probability that the Register effect could have arisen by chance is less than 5% for both ratios. However, in the H_2/H_1 condition, there is evidence of a significant interaction between Register and Word. The source of this interaction lies in the individual speech behaviors of the 6 speakers as revealed by a subsequent examination.

Table 6. Ratios of harmonics for all 6 speakers

Variable	F(d.f.)	H_2/H_1			H_{F1}/H_1		
		F	MSe	p	F	MSe	p
Register	(1, 5)	0.7675	0.438	0.0393	7.772	1.259	0.0385
Word	(7, 35)	1.206	0.347	0.3255	8.123	1.204	0.0001
Register × Word	(7, 35)	2.389	0.120	0.0416	1.439	0.330	0.2214

Table 7. Ratios of harmonics for speakers A, C, and E

Variable	F(d.f.)	H_2/H_1			H_{F1}/H_1		
		F	MSe	p	F	MSe	p
Register	(1, 2)	25.72	0.017	0.0368	24.58	1.445	0.0384
Word	(7, 14)	0.787	0.248	0.6092	2.424	0.078	0.0732
Register × Word	(7, 14)	1.729	0.035	0.1815	2.807	0.189	0.0476

Table 8. Ratios of harmonics for speakers B, D, and F

Variable	F(d.f.)	H_2/H_1			H_{F1}/H_1		
		F	MSe	p	F	MSe	p
Register	(1, 2)	5.689	0.662	0.1398	3.876	2.385	0.1878
Word	(7, 14)	1.261	0.448	0.3359	4.982	1.348	0.0052
Register × Word	(7, 14)	1.583	0.211	0.2198	0.796	0.434	0.6031

Table 9. Register 1 minus Register 2 differences of means for H_2/H_1 and H_{F1}/H_1 across speakers

Speakers	H_2/H_1	H_{F1}/H_1
A, B, C, D, E, and F	0.374	0.638
A, C, and E	0.188	0.399
B, D, and F	0.561	0.878

The results of that examination, shown in tables 7 and 8, indicate that our speakers exhibit two types of behavior. In table 7 speakers A, C, and E show a significant Register effect (at just under the 5% level) in both the H_{F1}/H_1 and H_2/H_1 conditions. In contrast, the Register effect in the data of speakers B, D, and F seen in table 8 falls substantially short of the same level of significance in both conditions. Meanwhile, the interaction between Register and Word observed earlier in table 6 for H_2/H_1 has disappeared from table 8, and a significant interaction effect has now emerged for H_{F1}/H_1 in table 7, the results for speakers A, C, and E.

Finally, the differences between means of the ratios computed for each of the two registers are shown in table 9 for all three groupings of the 6 speakers. The consistently positive sign of the differences indicates that the slope of the speech spectrum tends to be less steep in the higher frequencies in Register 1 than in Register 2.

Table 10. ANOVA of mean formant frequencies for speakers A, B, C, D, E, and F

Variable	F(d.f.)	Formant 1			Formant 2			Formant 3		
		F	MSe	p	F	MSe	p	F	MSe	p
Register	(1, 5)	8.534	1,903.3	0.0330	6.290	813.7	0.0540	12.04	5,481.1	0.0179
Word	(7, 35)	70.34	5,488.7	0.0001	243.6	21,157	0.0001	17.37	32,318	0.0001
Register × Word	(7, 35)	1.174	923.3	0.3427	1.514	2,898.6	0.1949	0.977	8,517.5	0.4629

Table 11. ANOVA of mean formant frequencies for speakers A, C, and E

Variable	F(d.f.)	Formant 1			Formant 2			Formant 3		
		F	MSe	p	F	MSe	p	F	MSe	p
Register	(1, 2)	3.507	3,898.3	0.2020	0.403	1,138.1	0.5906	8.673	2,318.0	0.0986
Word	(7, 14)	30.74	5,614.7	0.0001	157.5	16,217	0.0001	8.765	34,663	0.0003
Register × Word	(7, 14)	0.757	1,315.8	0.6309	0.407	4,282.8	0.8821	0.985	4,272.4	0.4797

Table 12. ANOVA of mean formant frequencies for speakers B, D, and F

Variable	F(d.f.)	Formant 1			Formant 2			Formant 3		
		F	MSe	p	F	MSe	p	F	MSe	p
Register	(1, 2)	28.26	141.87	0.0336	141.9	44.83	0.0070	5.006	9,797.8	0.1547
Word	(7, 14)	50.75	4,354.2	0.0001	90.96	28,746	0.0001	7.989	35,023	0.0005
Register × Word	(7, 14)	1.062	676.83	0.4350	1.519	2,435.7	0.2392	1.152	12,107	0.3875

Formant Frequencies

The first, second, and third formant frequency measurements were each averaged across the three repetitions of a given word in a given register. The results of a series of ANOVAs performed on each one of the first three formant frequencies with Word and Register treated as repeated measures are shown in table 10. With the exception of formant 2, where variation due to Register just fails to meet the 5% level of significance, the significance of the voice register effect is quite robust. Meanwhile, the variation due to word identity is highly significant throughout, and no interaction effects exceed the 5% criterion.

In a voice register distinction that is not described as including among its phonetic properties differences in vowel quality, we should expect to find that the formant frequency variation of our speakers' utterances would be independent of voice register. Thus, we closely examined the formant patterns of our speakers in a search for those who met this criterion and those who did not. Once again, we found that by placing the speakers A, C and E into one group (table 11) and the speakers B, D and F into another (table 12), the variation due to Register for the first group fell below the 5% level of

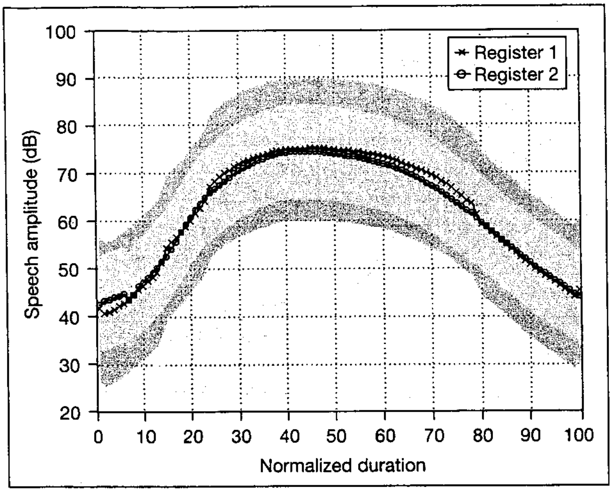

Fig. 4. A plot of the grand mean of the normalized amplitude trajectories of 8 pairs of words spoken in Register 1 (clear voice) and Register 2 (breathy voice) by 6 speakers. The shaded areas indicate the range of the standard error of the mean. The dark area represents the error for modal voice and the light area the corresponding error for breathy voice.

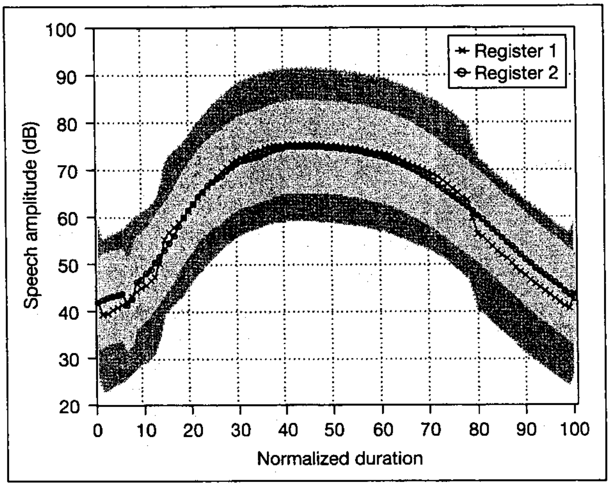

Fig. 5. A plot of the grand mean of the normalized amplitude trajectories of 8 pairs of words spoken in Register 1 (clear voice) and Register 2 (breathy voice) by speakers A, C, and E.

significance, whereas the F1 and F2 levels for the second group did not. The two groups contained mixed sexes; however, sex differences with respect to the formant data cannot be the source of the difference between the two groups. This is because what we examined is the set of within-speaker differences with respect to formant frequencies as a function of register. That is to say, each speaker served as his or her own control.

Amplitude and F0

Figures 4 and 7 show the normalized amplitude and F0 trajectories of the words plotted along the 100-point arbitrary time scale for all 6 speakers. The data for amplitude are broken down into our two subgroups in figures 5 and 6 and for F0 in figures 8 and 9. Figures 4, 5, and 6 indicate that there is essentially no difference in overall speech

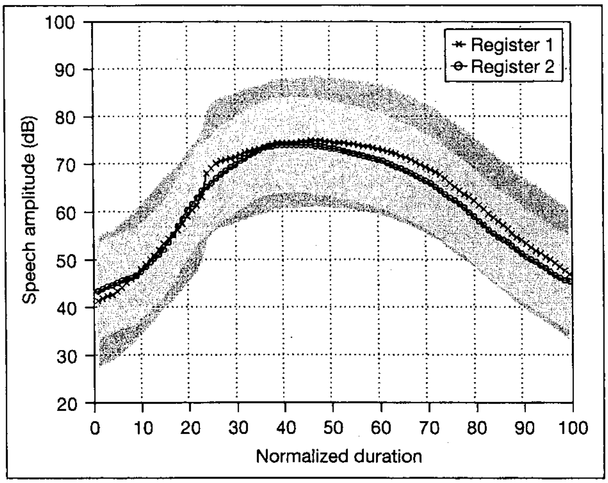

Fig. 6. A plot of the grand mean of the normalized amplitude trajectories of 8 pairs of words spoken in Register 1 (clear voice) and Register 2 (breathy voice) by speakers B, D, and F.

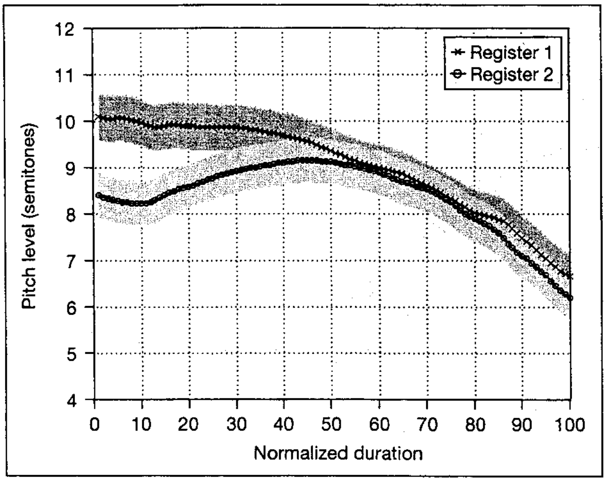

Fig. 7. A plot of the grand mean in semitones of the normalized F0 trajectories extracted from 8 pairs of words spoken in Register 1 (clear voice) and Register 2 (breathy voice) by 6 speakers. The shaded areas indicate the range of the standard error of the mean. The dark area represents the error for modal voice and the light area the corresponding error for breathy voice.

amplitude with respect to the two registers as a function of time. However, figures 7, 8, and 9 strongly suggest that there is a register-based difference in the F0 trajectory over the first half of the normalized time scale but a somewhat greater difference for speakers A, C, and E than for speakers B, D, and F.

An ANOVA was used to test the validity of these observations. On the strength of our visual inspection of figure 5, the length of the time scale was divided into four distinct segments (samples 1–15, 16–50, 51–75, and 76–100). In each of the four segments or normalized time intervals, the means and the gradients of the signal amplitude and F0 were calculated. With respect to the four means and gradients of the amplitude trajectories, we found that there was no significant difference between the two registers under any speaker grouping. With respect to the means of the F0 trajectories, significant differences were found for the mean frequency of the first interval (segment 1 in

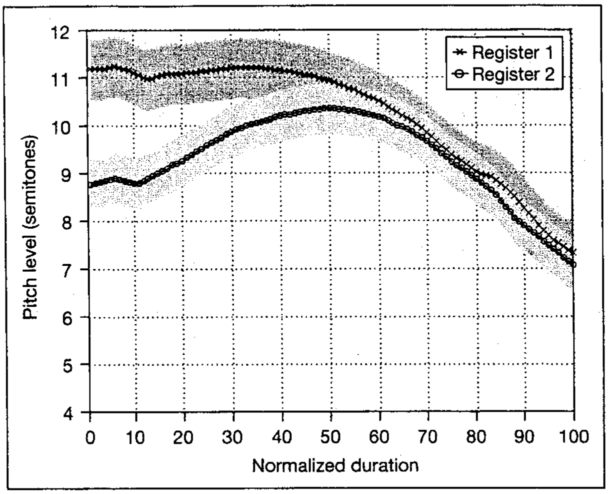

Fig. 8. A plot of the grand mean in semitones of the normalized F0 trajectories extracted from 8 pairs of words spoken in Register 1 (clear voice) and Register 2 (breathy voice) by speakers A, C, and E.

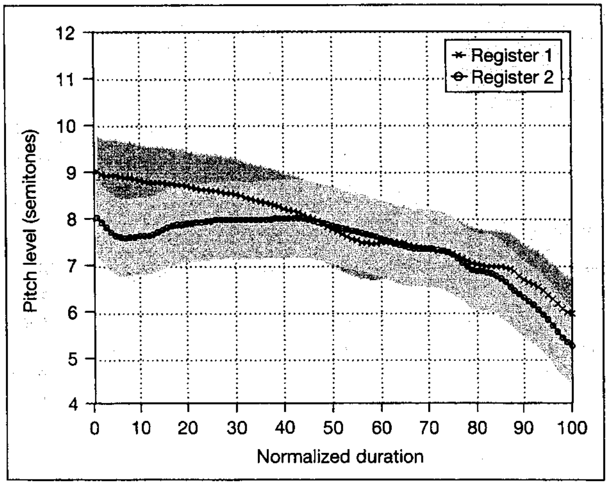

Fig. 9. A plot of the grand mean in semitones of the normalized F0 trajectories extracted from 8 pairs of words spoken in Register 1 (clear voice) and Register 2 (breathy voice) by speakers B, D, and F.

table 13). Furthermore, the F0 gradient of the second segment (segment 2 in table 14) also shows evidence of a significant Register effect. Meanwhile, the data reveal no significant interaction between voice Register and Word.

Table 14 shows the results of an ANOVA of Register and Word with respect to the gradients of the normalized F0 function in each of the four segments. The relationship of gradient with respect to Word is significant across all segments of the F0 contour and is consistent with the expectation that inherently different production features must distinguish the test words one from another. Meanwhile, the Register variable is significant only with respect to the gradient of segment 2. This suggests that the acoustic feature chiefly responsible for the perception of voice register is most in evidence early in the utterance, shortly after the speaker has made the laryngeal adjustments necessary to successfully sustain voicing. The interaction between Word and Register fails to rise

Table 13. Mean F0 of four segments: speakers A, B, C, D, E, and F

Variable	F(d.f.)	Segment 1			Segment 2		
		F	MSe	p	F	MSe	p
Register	(1, 5)	7.956	884.7	0.0371	3.503	473.6	0.1202
Word	(7, 35)	3.721	74.81	0.0041	8.010	125.7	0.0001
Register × Word	(7, 35)	1.847	89.08	0.1089	1.868	71.48	0.1050

Variable	F(d.f.)	Segment 3			Segment 4		
		F	MSe	p	F	MSe	p
Register	(1, 5)	0.397	286.8	0.3562	1.379	285.0	0.2931
Word	(7, 35)	11.61	352.4	0.0001	24.96	680.8	0.0001
Register × Word	(7, 35)	1.191	160.0	0.3335	0.753	284.3	0.6297

Table 14. Gradients of F0 trajectories: speakers A, B, C, D, E, and F

Variable	F(d.f.)	Segment 1			Segment 2		
		F	MSe	p	F	MSe	p
Register	(1, 5)	0.203	0.026	0.671	51.40	0.001	0.0008
Word	(7, 35)	7.392	0.218	0.0001	7.745	0.001	0.0001
Register × Word	(7, 35)	0.377	0.010	0.9098	1.228	0.001	0.0727

Variable	F(d.f.)	Segment 3			Segment 4		
		F	MSe	p	F	MSe	p
Register	(1, 5)	0.079	0.004	0.7900	2.104	0.002	0.2066
Word	(7, 35)	18.79	0.004	0.0001	4.356	0.015	0.0015
Register × Word	(7, 35)	0.059	0.005	0.9996	0.536	0.008	0.8013

to a significant level in all four segments. However, in the case of segment 2, the interaction is almost significant at the 5% level.

Tables 15 and 16 show the results of an ANOVA applied to the two previously isolated groups of speakers. The group comprised of speakers A, C, and E retains evidence of a Register effect significant at the 5% level for segment 2 while in the case of speakers B, D, and F, the probability for the same segment also exceeds the 5% criterion level.

Vowel Duration

The ANOVA of the vowel duration data in table 17 reveals that there is a significant difference between the two registers. That is, in the underlying data, the vowels of Register 1 are significantly longer than those of Register 2. Close inspection of the data, however, suggested that the four word-pairs containing an initial aspirated voiceless stop for Register 1 and an unaspirated voiceless stop for Register 2 (cf. the lower part of table 1) were responsible for the effect on the total group of eight word-pairs. Indeed,

Table 15. Gradients of F0 trajectories: speakers A, C, and E

Variable	F(d.f.)	Segment 1			Segment 2		
		F	MSe	p	F	MSe	p
Register	(1, 2)	0.155	0.053	0.732	27.34	0.001	0.0347
Word	(7, 14)	14.61	0.004	0.0001	11.18	0.0004	0.0001
Register × Word	(7, 14)	0.111	0.022	0.996	1.036	0.0003	0.4384

Variable	F(d.f.)	Segment 3			Segment 4		
		F	MSe	p	F	MSe	p
Register	(1, 2)	1.129	0.003	0.3993	0.352	0.0002	0.6130
Word	(7, 14)	17.74	0.002	0.0001	0.996	0.0210	0.4730
Register × Word	(7, 14)	0.474	0.001	0.838	0.989	0.0050	0.9742

Table 16. Gradients of F0 trajectories: speakers B, D, and F

Variable	F(d.f.)	Segment 1			Segment 2		
		F	MSe	p	F	MSe	p
Register	(1, 2)	0.014	0.009	0.918	41.82	0.0002	0.0231
Word	(7, 14)	3.139	0.003	0.0326	2.208	0.0020	0.0982
Register × Word	(7, 14)	1.372	0.002	0.290	1.657	0.0010	0.1994

Variable	F(d.f.)	Segment 3			Segment 4		
		F	MSe	p	F	MSe	p
Register	(1, 2)	0.233	0.004	0.6771	2.071	0.004	0.2867
Word	(7, 14)	25.72	0.002	0.0001	11.41	0.006	0.0001
Register × Word	(7, 14)	0.178	0.010	0.9857	0.0351	0.014	0.9161

Table 17. Vowel durations across the 8-word corpus

Variable	F(d.f.)	Vowel durations		
		F	MSe	p
Register	(1, 5)	7.718	878.28	0.0390
Word	(7, 35)	41.924	2,037.0	<0.0001
Register × Word	(7, 35)	8.459	863.64	<0.0001

an ANOVA of the difference between the two subgroups of word-pairs (table 18) showed them to be significantly different with regard to vowel duration. This led us to measure the voicing lags of the onsets of the four pairs with initial voiceless stops and subtract them from the corresponding vowels represented in table 17. That is, we were led to deviate in this instance from the procedure described in the 'Methods' section that stipulated the inclusion of the aspiration phase in all measurements of vowel duration. Even the 'unaspirated' stops had short voicing lags, i.e., brief periods of aspiration. The

Table 18. Vowel durations for words divided into two groups

Variable	F(d.f.)	Vowels with aspiration			Vowels without aspiration		
		F	MSe	p	F	MSe	p
Register	(1, 5)	37.662	866.16	0.0017	6.361	647.56	0.0530
Word	(3, 15)	66.257	2,308.1	<0.0001	2.818	2042.7	0.0747
Register × Word	(3, 15)	4.063	1,233.4	0.0268	3.592	569.94	0.0389

Four words contained vowels with initial aspiration and the remaining 4 did not.

Table 19. Vowel durations of the 8-word corpus with aspiration subtracted

Variable	F(d.f.)	Vowel durations		
		F	MSe	p
Register	(1, 5)	0.224	609.19	0.6562
Word	(7, 35)	53.196	1,960.8	<0.0001
Register × Word	(7, 35)	4.448	902.79	0.0013

Table 20. Auditory phonetic labels in percentages given to 198 utterances in each voice register by two phoneticians

Phonetician	Register 1				Register 2			
	phonation type		pitch		phonation type		pitch	
	modal	breathy	high	low	modal	breathy	high	low
ASA	98	2	98	2	51.5	48.5	8.1	91.9
TLT	98.5	1.5	74.2	25.8	33.3	66.7	43.3	56.6

ANOVA of the 8 pairs with aspiration removed from the subgroup (table 19) showed that there was no longer a significant difference between the two registers with regard to vowel duration.

Auditory Phonetics

As can be seen in table 20, the two phoneticians agreed in finding modal (clear) voice to be dominant for Register 1. All judgments in the 'mid/high' range were collapsed in the 'high' columns; all judgments in the 'low/falling' range were likewise collapsed in the 'low' columns. This was done because of rather random placements within each of the ranges. Although both observers also agreed in finding higher pitch to be another characteristic of Register 1, ASA heard it virtually 100% of the time, while TLT did not. As for phonation type in Register 2, ASA's response data were random, while TLT gave breathy voice responses two thirds of the time. She assigned very little weight to low pitch for Register 2, while ASA labeled 92% of the items low. The scattered additional comments are not systematic enough for tabulation.

Discussion

In the apparent absence of published research on the perception of phonologically relevant phonation types or voice registers, our wish was to fill this gap by beginning with the simplest possible case, a language with just two registers, 'clear' and 'breathy', and, preferably, one within fairly easy reach of Bangkok. To the best of our knowledge, the Mon-Khmer language Suai in its Kuai dialect as spoken in the village of Samrong satisfied both our requirements. In addition, access to native speakers was facilitated by the fact that the principal of the village school is the brother of a student of the second author.

In the early stages of the research our understanding was that the language had a stable distinction between the two voice registers but, as we made progress with our data analysis, evidence began to emerge that suggested ongoing change. Given this awareness, we felt the need to look for the possible relevance of age to the three classes of our informants: the 8 listeners who met our criterion in the control test for the perceptual experiment, the other 8 listeners who failed to identify the control stimuli at a rate better than chance, and the two groups of 3 speakers who differed in their performances according to the acoustic analysis. Unfortunately, because the surviving records for the 16 listeners did not include that information, the opportunity to explore this possibility eluded us. As for the 6 speakers recorded for acoustic analysis and identified by the letters A through F, records show their ages to have been 49, 50, 52, 60, 31, and 63. Given such a small sample, however, it is not possible to extract convincing evidence of an age effect. Had we been better informed of the linguistic situation at the outset of our work, we certainly would have taken greater care to retain data on the ages of our listeners and in future work intend to incorporate a sufficiently broad sampling of ages into our experimental design for both listening and speaking.

Perception Tests

Although the first acoustic recordings obtained for us of the speech of a few people from Samrong were not very good, we had the impression that the distinction was alive and well. We therefore went ahead to design five-parameter synthetic stimuli to explore the perceptual efficacy of three characteristics, phonation type, F0, and overall amplitude, that had previously been found to be prominent acoustic differentiators of such voice registers. As a matter of routine, our perceptual experiments included a control test with the utterances in the natural speech of 4 speakers. The overall responses of our 16 listeners to the control test were so ambiguous that we had to sift through them to find 8 listeners whose identification of the registers was somewhat better than chance; the display in figure 2 is limited to their data.

Our factor analysis of the identification scores clearly showed that the parameter that controlled the F0 onset frequency provided the most important acoustic cue. As F0 onset frequency increased, the more likely it became that listeners would identify a stimulus as belonging to Register 1, the 'clear' voice register. This finding was subsequently confirmed by a statistical analysis of the acoustic data extracted from natural speech tokens recorded by native speakers (fig. 8). In addition to the F0 onset frequency, which appeared in table 5 as factor 1, combinations of other parameters (at substantially lower levels of performance) emerged in groupings that suggested the possible nature of their supplementary contributions to the register distinction.

Factor 2 was suggested as a candidate for the label Excitation Amplitude. The justification for this label was the involvement of the parameters ATU and TG (both of which had a direct influence on signal amplitude) and the parameter OQ that also made a contribution to the overall amplitude and to the distribution of energy across the voice spectrum. Bearing in mind that the Register 2 response data were employed in the analysis, the positive sign of the largest term of factor 2, OQ, suggests that an increase in the open quotient of the voicing waveform led (as expected) to an increase in the proportion of Register 2 ('breathy') responses. Similarly, the TG term indicates that as the initial stimulus gain increased over longer periods of time, fewer Register 2 responses occurred. Meanwhile, the smallest of the three principal terms of factor 2, ATU, suggests (counterintuitively) that an increase in Register 2 responses associated with an increase in the amplitude of turbulence led to a *lower* proportion of Register 2 ('Breathy') responses. However, as the correlation coefficients in table 3 indicate, the settings explored by the ATU parameter contributed such a small proportion of the response variance that for all practical purposes the parameter's anomalous behavior may be ignored.

Dubbed the Aeroacoustic Interaction factor, factor 3 contains principal loadings from the parameters OQ and TL. In this instance, increases in the open quotient gave rise to increases in 'breathy' responses, and increases in the spectral tilt parameter, TL, led to increases in Register 1 or 'clear' responses. From a production viewpoint, the two parameters are interrelated, inasmuch as, in the human being, as the open quotient of the glottal cycle increases, its effect on the spectral distribution of energy is such as to increase the attenuation of high frequencies. Thus, the two synthesis parameters ought to act in harmony to enhance 'breathy' responses, yet we have here a paradoxical result in that increases in TL led, instead, to Register 1 responses! Although we cannot with great assurance explain the latter result, we are much taken with the following reasoning by an anonymous reviewer. The attenuation of higher frequencies (spectral tilt) in a stimulus presented acoustically through a playback system is likely to give the listener the impression of somebody 'turning down the treble knob', which for the listener need not reflect any change in phonation type. The open-quotient parameter, however, is more specifically related to the voice source and thus leads to more direct inferences about phonation type. It might even be true that manipulation of the relative amplitudes of the first harmonic and specific upper harmonics would be a better synthesis strategy than adjustments of overall spectral tilt to yield the perception of breathiness, even though in natural speech spectral tilt as such is a normal consequence of longer spans of open quotient of the glottis. Indeed, as things stand, we do wish to emphasize, as shown in table 3, that the puzzling effect of the TL parameter is relatively minor, accounting for only 3.4% of the response variance, while OQ accounts for 16.5%.

Ratios of Harmonic Intensities

The positive Register 1 minus Register 2 differences for each of the two ratios (H_2/H_1 and H_{F1}/H_1) show that the words spoken in Register 1 (clear voice) tend to have a less sloping spectrum than those in Register 2 (breathy voice). The interaction between Word and Register in table 7 does not contradict the fact that Register 2 always has a greater spectral slope than Register 1. Breathy voice is usually the result of increasing the open quotient of the glottal cycle which, in turn, leads to an increase in the slope of the speech spectrum. This phenomenon arises because the open quotient of the glottal cycle is usually accompanied by a less rapid closure, and a slower closure produces less high frequency energy to excite the cavities of the vocal tract. Consequently, the energy of the speech spectrum declines more rapidly as a function of frequency [Ladefoged et al., 1988; Wayland and Jongman, 2003]. Thus, the results of the ratio analysis are also consistent with the results of the perception study that highlighted the importance of the OQ parameter in contributing to the register distinction.

Formant Frequencies

Our primary concern in measuring the formant frequencies of the vowels in our corpus of utterances was to be sure that we were not obtaining distorted ratios of harmonic intensities by examining spectra within word-pairs with dissimilar formant patterns. Another concern was the possibility of a correlation of vowel quality with register. (Vowel quality, of course, is the principal auditory correlate of formant pattern.) We had to be sure that there were no significant differences in formant frequency, including any that might not be very discernable by ear. This was found to be the case for speakers A, C, and E (table 11) but not in the case of speakers B, D, and F (table 12) for whom significant differences were evident in formants 1 and 2. This register-sensitive difference in the formants of speakers B, D, and F might cast doubt upon the validity of the harmonic-ratio results for those 3 speakers.

Amplitude and F0

No difference in overall amplitude appears for either group of speakers, thus seemingly undermining our reason for choosing TG as one of our synthesis parameters. This is true even though the latter had an effect in the perception tests but only a very minor one, accounting for less than 5% of the variance. On the other hand, F0 is dominant as a cue in perception and as a phonetic property in speech production.

Vowel Duration

Once spans of voicing lag (aspiration) are removed from vowels following voiceless stops, vowel duration is not a significant differentiator of the voice registers. This raises the question of whether noise excitation of the first portion of a vowel universally tends to lengthen the duration of the vocalic gesture. It is not easy to find published data on the topic. One study [Peterson and Lehiste, 1960, pp. 700–701] does indeed find that American English syllabic nuclei after voiceless stops are longer than those after voiced stops if aspiration is included in the measurements.

Auditory Phonetics

The impressionistic observations of the two phoneticians require further comment. Unfortunately, we could not obtain the services of additional trained and experienced field workers; nevertheless, given the awareness emerging from our own perceptual and acoustic data of the state of flux of the Kuai voice registers, we thought that these findings fitted well with that flux. What can we say about table 20? TLT is a native speaker of a tone language, Thai, and has done much research on Mon-Khmer languages, including both fieldwork and instrumental studies, as well as diachronic phonology. Meanwhile, ASA, a native speaker of American English, has developed much sensitivity to the distinctive use of pitch in tonal systems through his practical experience with Thai and his work on the tones of that language; however, his experience with Mon-Khmer languages is meager. We conjecture that, in the face of the aforementioned flux, the two observers attended more diligently to different aspects of the speech. TLT, perhaps influenced by her knowledge of the history of Suai in its matrix of closely related voice-register languages, focused sharply on voice quality. Also, perhaps subconsciously, she may have tried to avoid a tone-language speaker's bias toward accepting too readily the emergence of an embryonic tonal system characterized by pitch differences. ASA, on the other hand, already very doubtful about the stability of phonation type as a differentiator, may have paid too little attention to that property and concentrated on pitch. In summary, all this may be best expressed by saying that the data of table 20 are not at all surprising for a language obviously undergoing phonological change.

Conclusion

We originally set out to study the acoustic and perceptual features of the register distinction in the Kuai dialect of Suai. As our analysis of the data evolved, we found ourselves obliged to shift our focus due to the discovery that our chosen language was in a state of flux. The shift appeared to be from a register language to one with some kind of accentual salience, which itself may be a transitional stage to one of phonologically distinctive tones. Indeed, one investigator [Sukkasame, 2003] asserts that the process is well on its way in Kui, the other major dialect of Suai.

Despite the state of flux, we persisted in our efforts to perform perception tests and acoustic analyses. The data suggest that the distinction between clear and breathy registers has not entirely disappeared. In speech production some people maintain a fairly good distinction, while others do not. Likewise, the perceptual data reveal a mixture of levels of sensitivity. To the extent that inhabitants of the village who no longer produce the distinction respond differently to it as listeners, this may be because they are quite used to the speech of many of their elders who still have it in their normal speech.

It remains to be seen whether in future generations this state of flux will give way to some kind of stability, albeit probably temporary, as is commonly expected in diachronic phonology. If so, will that stability result from a complete loss of all aspects of the old register distinction with, presumably, a merger of the lexical classes formally minimally distinguished by Registers 1 and 2? Of course, another strong likelihood is that the phonological distinction will be maintained through a shift to a pitch accent or even the rise of phonemic tones. The latter could be helped by the external pressure of extensive contact through bilingualism with the tone languages Thai and Lao.[4] Notwithstanding the evidence of instability, we believe that the results may have some interest to phoneticians, phonologists, and specialists in Mon-Khmer linguistics. In addition, for historical linguists a glimpse of phonological change in progress is surely important.

Acknowledgments

This work was supported by NIH Grant DC-02717 to Haskins Laboratories and a grant from the Thailand Research Fund. The first author is also grateful to the Department of Linguistics, Faculty of Arts, Chulalongkorn University, Bangkok for its gracious hospitality during his stay for part of the research. Preecha Sukasame's help in working with the people of Samrong was invaluable. We appreciate the critical comments on various aspects of this paper from Claudia Carello, Carol Fowler, Leonard Katz, Bruno Repp, Khalil Iskarous, Klaus Kohler, and the writer of an excellent anonymous review.

---

[4] A somewhat similar process seems to be happening with another minority language of Thailand, Pattani Malay, through a shift from distinctive word-initial consonant length to a pitch accent [Abramson, 2003].

Appendix 1

The Phonemes of Kuai [adapted from Sukkasame, 2003, pp. 166–169]

Consonants

Initials

p	t	c	k	ʔ
pʰ	tʰ	cʰ	kʰ	
b	d			
m	n	ɲ	ŋ	
	l			
w		j		
	s			h

Initial clusters

pl	pʰ	bl	tr
kl	kʰl	kw	

Note: /r/ occurs only in the clustert /tr/.

Finals

p	t	c	k	ʔ
m	n		ŋ	
w	l	j		h

Vowels

Short

i	ɯ	u
e	ɤ	o
ɛ	ʌ	ɔ
æ	a	ɒ

Long

iː	ɯː	uː
eː	ɤː	oː
ɛː	ʌː	ɔː
æː	aː	ɒː

Diphthongs

ɪa ɯa ua

Registers

R1 = modal voice, R2 = breathy voice

Appendix 2

Parameter Settings

	ATU, dB		OQ, samples		F0, Hz at 0 ms		TL, dB		Time to reach 60 dB gain (TG), ms	
	65	70	40	60	70	135	18	24	130	200
lu:1	default	40	default	20	default	101	default	0	default	0
lu:2	65		default	20	default	101	default	0	default	0
lu:3		70	default	20	default	101	default	0	default	0
lu:4	default	40	40		default	101	default	0	default	0
lu:5	default	40		60	default	101	default	0	default	0
lu:6	default	40	default	20	70		default	0	default	0
lu:7	default	40	default	20		135	default	0	default	0
lu:8	default	40	default	20	default	101	18		default	0
lu:9	default	40	default	20	default	101		24	default	0
lu:10	default	40	default	20	default	101	default	0	130	

Appendix 2 (continued)

	ATU, dB		OQ, samples		F0, Hz at 0 ms		TL, dB		(TG), ms	
	65	70	40	60	70	135	18	24	130	200
lu:11	default	40	default	20	default	101	default	0		200
lu:12	65		40		default	101	default	0	default	0
lu:13	65			60	default	101	default	0	default	0
lu:14		70	40		default	101	default	0	default	0
lu:15		70		60	default	101	default	0	default	0
lu:16	65		default	20	70		default	0	default	0
lu:17	65		default	20		135	default	0	default	0
lu:18		70	default	20	70		default	0	default	0
lu:19		70	default	20		135	default	0	default	0
lu:20	65		default	20	default	101	18		default	0
lu:21	65		default	20	default	101		24	default	0
lu:22		70	default	20	default	101	18		default	0
lu:23		70	default	20	default	101		24	default	0
lu:24	65		default	20	default	101	default	0	130	
lu:25	65		default	20	default	101	default	0		200
lu:26		70	default	20	default	101	default	0	130	
lu:27		70	default	20	default	101	default	0		200
lu:28	default	40	40		70		default	0	default	0
lu:29	default	40	40			135	default	0	default	0
lu:30	default	40		60	70		default	0	default	0
lu:31	default	40		60		135	default	0	default	0
lu:32	default	40	40		default	101	18		default	0
lu:33	default	40	40		default	101		24	default	0
lu:34	default	40		60	default	101	18		default	0
lu:35	default	40		60	default	101		24	default	0
lu:36	default	40	40		default	101	default	0	130	
lu:37	default	40	40		default	101	default	0		200
lu:38	default	40		60	default	101	default	0	130	
lu:39	default	40		60	default	101	default	0		200
lu:40	default	40	default	20	70		18		default	0
lu:41	default	40	default	20	70			24	default	0
lu:42	default	40	default	20		135	18		default	0
lu:43	default	40	default	20		135		24	default	0
lu:44	default	40	default	20	70		default	0	130	
lu:45	default	40	default	20	70		default	0		200
lu:46	default	40	default	20		135	default	0	130	
lu:47	default	40	default	20		135	default	0		200
lu:48	default	40	default	20	default	101	18		130	
lu:49	default	40	default	20	default	101	18			200
lu:50	default	40	default	20	default	101		24	130	
lu:51	default	40	default	20	default	101		24		200

References

Abramson, A.S.: Acoustic cues to word-initial stop length in Pattani Malay. Proc. 15th Int. Congr. Phonet. Sci: (The 15th ICPhS Organizing Committee, Barcelona 2003), pp. 387–390.

Abramson, A.S.: Toward prosodic contrast: Suai and Pattani Malay; in Bel, Marlien, Proc. Int. Symp. on Tonal Aspects of Languages: Emphasis on Tone Languages, pp. 1–4 (The Institute of Linguistics, Chinese Academy of Social Sciences, Beijing, 2004).

Bickley, C.: Acoustic analysis and perception of breathy vowels. MIT Speech Commun. Working Papers *1:* 71–81 (1982).

Fischer-Jørgensen, E.: Acoustic analysis of stop consonants. Miscnea phonet. *2:* 42–59 (1954).

Flanagan, J.L.: A difference limen for vowel formant frequency. J. acoust. Soc. Am. *27:* 613–617 (1955).

Gordon, M.; Ladefoged, P.: Phonation types: a cross-linguistic overview. J. Phonet. *29:* 383–406 (2001).

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

Klatt, D.H.: Software for a cascade/parallel formant synthesizer. J. acoust. Soc. Am. *67:* 971–995 (1980).

Klatt, D.H.; Klatt, L.C.: Analysis, synthesis and perception of voice quality variations among female and male talkers. J. acoust. Soc. Am. *87:* 820–856 (1990).

Ladefoged, P.; Maddieson, I.; Jackson, M.: Investigating phonation types in different languages; in Fujimura, Vocal physiology: voice production, mechanisms and functions, vol. 2, pp. 297–317 (Raven Press, New York 1988).

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

Peterson, G.P.; Lehiste, I.: Duration of syllabic nuclei in English. J. acoust. Soc. Am. *32:* 693–703 (1960).

Shrivastav, R.; Sapienza, C.M.: Objective measures of breathy voice quality obtained using an auditory model. J. acoust. Soc. Am. *114:* 2217–2224 (2003).

Sriwises, P.: Kui (Suai)-Thai-English dictionary. Indigenous Languages of Thailand Research Project (Chulalongkorn University Language Institute, Bangkok 1978).

Sukkasame, P.: Phonological variation and change in Kuai–Kui (Suai) (in Thai); doct. diss. Chulalongkorn University, Bangkok (unpublished, 2003).

Thongkum, T.L.: Phonation types in Mon-Khmer languages; in Fujimura, Vocal fold physiology 2: voice production, 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.: 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).

Wayland, R.; Jongman, A.: Acoustic correlates of breathy and clear vowels: the case of Khmer. J. Phonet. *31:* 181–201 (2003).

Yantreesingh, P.: The phonology of the Kuai language of Suphanburi with comparison to the Kui language of Surin; MA thesis Mahidol University, Bangkok (unpublished, 1980).