

An initial exploration of the interaction of tone and intonation in Kera'a

Constantijn Kaland¹, Naomi Peck², T. Mark Ellison¹, Uta Reinöhl²

¹Institute of Linguistics, University of Cologne, Germany ²Department of General Linguistics, University of Freiburg, Germany

ckaland, naomi.peck, tellison, uta.reinoehl
[@uni-koeln.de | @linguistik.uni-freiburg.de]

Abstract

This paper provides a first acoustic analysis of the lexical tones found in Kera'a, an endangered language spoken in the North-East of India (Arunachal Pradesh). Minimal existing work provides impressionistic descriptions of the tone system and leads to diverging claims. Apart from lexical contrasts, tones are reported to be affected by sociolinguistic variables such as clan and gender, as well as by elicitation setting. The present study explores these factors using f0 contours found in monosyllabic words that were produced by speakers that differed in gender, clanlectal background and in different elicitation contexts. Cluster analyses were used to explore the f0 variation. Preliminary results suggest interactions between speaker and elicitation context and shed a new light on the realisation of tones in Kera'a.

Index Terms: lexical tone, prosody, gender, elicitation, cluster analysis, f0 contour.



1. Introduction

Figure 1: Map of the Dibang Valley and the towns (red) and villages (blue) where Kera'a is spoken (from [1]).

Kera'a (iso: clk) is an endangered Trans-Himalayan language spoken in the Dibang Valley (state of Arunachal Pradesh, India), e.g. [1], see black bordered area in Figure 1. The language is known elsewhere as "Idu (Mishmi)", based on an exonym for its speakers [2]. The number of Kera'a speakers is estimated at 11 000 [3], covering multiple dialects [2]. The main reasons for its endangered status are social, economic, and political in nature, with (Arunachali) Hindi and English as prestige varieties that people shift to. One particularly important reason is the Hindi-based education system. Despite proposals for an orthography [4], the language still lacks a standardized script. Middle-aged and elderly predominantly speak Kera'a, as younger generations have adopted Hindi or English, in some cases fully, but normally in combination with extensive codeswitching with Kera'a [5].

The available linguistic work on Kera'a is limited and the documentation of its phonology is rather impressionistic. The most extensive works available to date are a phrasebook with basic language description, word lists and phrases [6], and a community handbook covering Kera'a phonology, morphology and syntax [7]. These two sources agree that Kera'a makes use of three lexical tones, either termed falling, level and rising [6] or low, mid and high [7] respectively (à, ā and á are used as diacritics), but see [8] for an analysis with five tones. However, there is no consistent agreement between these sources on how the tones are used and the literature lack thorough linguistic analysis. An illustrative number of minimal triplets are given in Example 1, where (a) and (b) show the disagreement. Polysyllabic minimal tonal triplets are reported in [6], although there is no reference to the tonal realisation, i.e. whether a tone is found on a single syllable or spreads over multiple syllables or whether both syllables are assigned separate tones. Given the minimal work on lexical tone in Kera'a, the current study further explores this topic by taking into account several sources of variation.

(1) Examples of lexical tones in Kera'a.

a.	Taken from	[6]
	prà	well
	prā	salt
	prá	bird
b.	Taken from	[7]
	prà	bird
	prā	leaves, paper, flat things
	nrá	calt

1.1. Sociolinguistic differences

The literature on other Sino-Tibetan languages reports gender effects on the realisation of lexical tone. For example, these effects were reported for several Bodish languages, spoken in the northern part of Nepal (see [9] for an overview). Furthermore, a study on 5 female and 5 male speakers of the Karenic Sino-Tibetan language Pa-O, as spoken in Bangkok (Thailand) reported similar effects [10]. These effects did not result from physiological differences in f0 in that men tend to have overall lower f0 levels than women, i.e. the f0 values were gender normalized. Despite this, differences were found in the way some of the six tones in Pa-O were realised. The clearest gender effects were found in the overall shape as well as in the convexity of the f0 contour. That is, tone 4 was consistenly produced as a rise-fall by the female speakers and as a fall by the male speakers. Tone 3 was consistently produced as a linear fall by the female speakers and as a concave or convex fall by the male speakers [10]. These differences were ascribed to different speaking styles as a result of the segregated lives that men and women lead: "Pa-O males and females are kept separate with different activities most of the day" ([10], p.90).

Although the (sociolinguistic) effects reported above are striking as they interfere with tonal realisation at the lexical level, the literature has shown gender effects on other prosodic levels as well. American English (AE) speaking females were found to have different preferences for boundary tones at the intermediate phrase (ip) and intonational phrase (IP) levels compared to male speakers [11]. The gender effects at the ip-level were, however, also dependent on the dialect of the speaker in that Southern AE females used a H phrase accent more often than Midland AE females. These results were comparable to the gender effects reported for high rising terminals (IP-final) in New Zealand English (e.g. [12]).

In addition to gender, an important aspect regarding the background of the Kera'a speakers concerns their clanlect. Increasing work in communities outside the normal purview of sociolinguistics has shown that the 'clan' that a speaker belongs to may play an important role in social variation [13]. For example, the clanlects associated with Bininj Kun-wok (Gunwinyguan, Australia) exhibit different syntactic rules [14]. In Kera'a, a number of clanlects are bundled into larger dialectal groups based on a number of salient lexical, phonological and phonetic characteristics; the two primary dialects are Midu and Mithu (with Mindri a potential third) (see [2]). As for phonetic differences, members of differents clans are reported to realise tonal categories in different ways ([2]). These clanlects appear to be more primary to speakers than the dialect group that the clanlect is associated with.

1.2. Phonetic realisation of tones

The sociolinguistic factors can challenge phonological accounts in which prosodic structure is the primary focus to describe the tone and intonation patterns of a certain language ([15], [16]). That is, because sociolinguistic variables such as gender interfere with the choice of those f0 movements that are traditionally analysed as having a mainly linguistic meaning (e.g. demarcating), our thinking in terms of phonological inventories is then lacking an important factor and only applies to an abstract representation of natural speech. A similar observation holds for the use of speech segments, which are also traditionally analysed in terms of inventories (e.g. [17]). It has been shown repeatedly that sociolinguistic variables strongly correlate with the segmental choices of speaker, where even a single speech sound can be a reliable indicator of the speaker's sociolinguistic background (e.g. [18]). As illustrated above, sociolinguistic variables play a large role in the phonetic realisation of tone as well. It seems therefore crucial to take them into account in early stages of language documentation.

Recently, a method to categorise f0 contours based solely on their shape was proposed [19]. This approach was shown to be able to capture multiple sources of variation in f0 contours as long as they led to different shapes in the surface productions. It provides therefore a reliable and reproducible result for further hypothesizing and testing of how f0 variation can be related to underlying *prototypical* contours. This method is applied in the current study.

1.3. Research aims

Given the little work on tone in Kera'a, this study further explores two sources of f0 variation at the lexical level. One source is linguistic in nature and pertains to the context in which the lexical tone is realised; either in isolation or in a phrase. The other sources is sociolinguistic and concerns the gender of the speaker and their clanlect. Using cluster analysis, this study investigates how the surface realisations of contours correlate with these potential sources of variation. Secondarily, the results are discussed in the light of the proposed three-way tonal inventory ([6], [7]).

2. Methodology

2.1. Data collection

The speech data collected for this study concerns 189 monosyllabic words. They were extracted from an elicitation recording in which they were either produced in isolation or in a phrase, following the methods described in [20]. The phrase context was as follows: "I said [word]", which corresponds to a phrasemedial position of the target word in Kera'a. In both elicitation contexts the word was repeated three times. In order to minimize effects of list intonation, the second occurrence was selected for analysis. Some words were repeated in multiple triplets and therefore occurred multiple times in the data. The total number concerned 665 syllables selected for analysis (328 in isolation, 337 in phrase). They were collected from one female speaker (27, from Anini, Mipi/Tayu clan, 347 syllables) and one male speaker (30, from Etalin, Milli clan, 318 syllables) of the Mindri dialect. Recordings were saved as 44.1 kHz mono wave-files.

2.2. F0 measurements

The boundaries of the selected syllables were labeled using EMU-R [21] and Praat [22] textgrids. Thereafter, f0 measures were taken using the script provided in [19]. That is, 20 measures were taken per syllable, which were equally divided over the duration of the syllable (stylization resolution: 1 ST, f0 range: 75-500, time-step: 0.01 s.). This resulted in one vector of f0 measurements per syllable (time-series), which could then be analysed in the cluster analysis. This method ensures that the f0 contour is stylized and interpolated such that microprosodic variation is kept to a minimum and that there is no missing f0 data due to voiceless consonants (as required by the cluster analysis).

2.3. Cluster analysis

The time-series f0 measures were analysed in a cluster analysis using the tool provided in [19]. Speaker normalisation was applied using standardization [23], which was proposed as a particularly suitable speaker correction in lexical tone productions. In this way, physiological differences between the male and female speaker were taken into account (see also [10]). Several rounds of clustering were performed on the same dataset, in order to find the most suitable number of clusters for further analysis. Note that this process is inherent in cluster analyses in general and has limitations (see discussion in [19]). One approach takes into account the number of contours in each cluster and performs rounds with increasing numbers of clusters (starting with two clusters). No more clustering rounds are performed when the distribution of contours over the cluster starts to be skewed. In this coarse grained approach, the current analysis stopped with six clusters. In this analysis, the smallest cluster (cluster 5, Figure 2) consisted of 11 contours and the largest cluster consisted of 344 contours.

An additional assessment of the number of clusters was performed using a novel method based on information cost [24]. This evaluation metric compares the information cost (entropy) of several clustering rounds in order to find the round with the lowest cost. The metric takes into account several aspects of the contour clustering to compute this measure and is meant as an evaluation of both the description length [25] and the best fit of the individual contours in a given cluster. This metric was run on clustering rounds ranging from two to ten clusters and indicated that the round with six clusters had the lowest information cost (Table 1).

 Table 1: Information cost per clustering round. Shaded row indicates lowest information cost.

N clusters	Information cost
2	12084.54
3	11767.49
4	11587.94
5	11432.51
6	11372.18
7	11494.41
8	11529.95
9	11503.22
10	11571.06

3. Results



Figure 2: Mean f0 contour (and standard deviation) per cluster with 6 clusters assumed.

The results show that with six clusters, the largest f0 movements are found in cluster 2 (fall) and cluster 6 (rise). Inspection of Tables 2 and 3 reveals that these contours were mostly produced in isolation and that the fall was almost always produced by the female speaker whereas the rise was always produced by the male speaker. See examples in Figure 3. Other clusters did not show such a skewed distribution of the contours. These results indicate for the words in isolation the male and female speaker produced different tonal realisations.

Table 2: N contours per cluster and elicitation context.

Context	Cluster number					
	1	2	3	4	5	6
isolation	61	41	144	42	4	36
phrase	73	8	200	49	7	0

Table 3: N contours per cluster and gender.

Gender	Cluster number					
	1	2	3	4	5	6
female	63	47	188	43	6	0
male	71	2	156	48	5	36



Figure 3: *Example contours of the word 'so' ('to boil vegetables') produced in isolation by the female (top) and male (bot-tom) speaker.*

It remains unclear why the speaker difference only appears in isolation contexts. In order to further explore this difference, another cluster analysis was carried out on the syllables produced in phrases only. This analysis was run with three clusters assumed, in order to investigate whether the outcome provides support for the three-way tonal contrast proposed in [6] and [7], and/or the alleged gender differences.

The speaker differences do not show as clearly in the phrase context as they did for the syllables in isolation (Figure 4 and Table 4). Cluster 2 (low fall) has the most skewed distribution with the majority of the contours produced by the female speaker. However, the division is not as clear as in the isolation context. The analysis on the words produced in phrase context with three clusters shows both shape and register differences between the contours. In cluster 1 a fall around the speaker mean (zero) can be observed. In cluster 2 a falling movement well below the speaker's mean is found. And in cluster 3 a high register rise-fall is found. These three contours do correspond to some extent with the mid, low and high tones proposed in [6] and [7]. However, their overall distribution is highly skewed, with most of the contours in cluster 1, indicating that the cluster output is still asymmetric and does not reflect all f0 variation in the data.



Figure 4: Mean f0 contour (and standard deviation) per cluster with 6 clusters assumed.

Table 4: N contours per cluster and gender (phrase context).

Gender	Cluster number				
	1	2	3		
female	125	40	12		
male	123	28	9		

4. Discussion

This study showed differences between the male and female speaker of Kera'a in the way lexical tones were produced. In particular for monosyllabic words produced in isolation the female speaker produced falling contours whereas the male speaker produced rising contours. This division was almost exclusive and hinted at gender or clanlectal differences in tonal realisation. However, as this study is a first exploration using acoustic measurements for this language, some issues remain unanswered.

First, it remains to be seen why the differences did not show up in phrase contexts, or at least not in a clear manner. This observation may indicate that the skewed distribution of rises and falls originates from a speaking style differences in repeating words in a list. In this way, the results might indicate individual differences that happen to coincide with sociolingistic factors. From the two speakers reported in this study it is not possible to disentangle the two.

As for the methodology in this study, it was shown that cluster analysis is able to distinguish multiple sources of variation in f0 contours and produced useful results for further research

on Kera'a. For example, it is now clear that speaker variation can lead to opposite tonal realisations and co-occur with realisations that are found across speakers (Figure 2). Although isolation was reported as a useful elicitation context for Tai tones in [20], it remains to be seen whether this holds for Kera'a tones. If the same monosyllabic word can be produced with either a steep rise or a steep fall (Figure 3), depending on the speaker, these realisations are likely masking or at least interacting with additional underlying linguistically meaningful lexical tones. Although this interaction could provide a challenge for future research, the current study does not exclude a threeway tonal contrast as proposed in the literature ([6] and [7]). It should be noted, however, that a linguistic analysis based on minimal tonal pairs is still lacking. Preliminary evidence from recorded minimal sets hints at the possibility of a four- or five-way contrast (see footnote 13 in [2]).

The next logical steps for Kera'a tone would be to analyse different elicitation contexts, including more spontaneous speech from a larger sample of speakers. This would shed more light on the alleged gender differences and other sources of variation in the realisation of lexical tones.

5. Acknowledgements

Research for this paper was funded by the German Research Foundation (DFG) Project-ID 281511265 (SFB-1252 "Prominence in language") and Project-ID 406074683 (Emmy Noether-Gruppe "Non-hierarchicality in grammar: Construction formation without word class distinction across categories and languages").

6. References

- N. Peck, "Keraa (Arunachal Pradesh, India) Language Snapshot," *Language Documentation and Description*, vol. 19, pp. 26– 34, 2020.
- [2] U. Reinöhl, "Locating Keraa (Idu Mishmi) in its linguistic neighbourhood: Evidence from dialectology," in *Ethnolinguistic prehistory of the Eastern Himalaya*, M. W. Post, S. Morey, and T. Huber, Eds. Leiden: Brill, 2022.
- [3] M. P. Lewis, G. F. Simons, and C. D. Fennig, *Ethnologue: lan-guages of asia*. Dallas: Summer Inst Linguistics, 2015.
- [4] R. Blench, "Attempts to write the Idu Mishmi language and a proposal for a modern orthography," in *Change and continuity among tribes: the Idu Mishmis of the Eastern Himalayas*, T. Mene and S. K. Chaudhuri, Eds. Bhopal: Rawat Publications, 2018, pp. 115–130.
- [5] R. Sarma, "Disappearing Dialect: the Idu- Mishmi Language of Arunachal Pradesh (India)," *International Journal of Intangible Heritage*, vol. null, no. 10, pp. 61–72, May 2015.
- [6] J. Pulu, A hand book on Idu Mishmi language. Itanagar: Dept. of Cultural Affairs, Directorate of Research, Govt. of Arunachal Pradesh, 2002.
- [7] R. Blench and M. Lingi, A grammar of Idu, a language of Arunachal Pradesh. Roing: Mishmi Publishing House, 2018.
- [8] D. Jiang, Yidu yu yanjiu. Beijing: The Nationalities Press, 2005.
- K. A. Hildebrandt, "Tone in Bodish languages: Typological and sociolinguistic contributions," in *New Challenges in Typology: Broadening the Horizons and Redefining the Foundations*, M. Miestamo and B. Wlchli, Eds. De Gruyter Mouton, 2008, pp. 77–102.
- [10] S. Jitwiriyanont, "Gender Variation in Acoustic Characteristics of Southern Pa-O Lexical Tones," *MANUSYA*, vol. 18, no. 2, pp. 71–91, 2015.

- [11] C. G. Clopper and R. Smiljanic, "Effects of gender and regional dialect on prosodic patterns in American English," *Journal of Phonetics*, vol. 39, no. 2, pp. 237–245, Apr. 2011.
- [12] D. Britain, "Linguistic change in intonation: The use of high rising terminals in New Zealand English," *Language Variation and Change*, vol. 4, no. 1, pp. 77–104, 1992.
- [13] J. Mansfield and J. Stanford, "Documenting sociolinguistic variation in lesser-studied indigenous communities: Challenges and practical solutions," *Language Documentation & Conservation*, pp. 116–136, 2017.
- [14] M. Garde, "Kun-dangwok: clan lects and Ausbau in western Arnhem Land," *International Journal of the Sociology of Language*, vol. 2008, no. 191, Jan. 2008.
- [15] S.-A. Jun, Ed., Prosodic typology: the phonology of intonation and phrasing, ser. Oxford linguistics. Oxford; New York: Oxford University Press, 2005.
- [16] —, Prosodic typology II: the phonology of intonation and phrasing, ser. Oxford linguistics. Oxford: Oxford University Press, 2014.
- [17] I. Maddieson, *Patterns of sounds*, 1st ed., ser. Cambridge studies in speech science and communication. Cambridge: Cambridge University Press, 2009.
- [18] C. Kaland, V. Galatà, L. Spreafico, and A. Vietti, "Which Language R You Speaking? /r/ as a Language Marker in Tyrolean and Italian Bilinguals," *Language and Speech*, vol. 62, no. 1, pp. 137–163, 2019.
- [19] C. Kaland, "Contour clustering: A field-data-driven approach for documenting and analysing prototypical f0 contours," *Journal of the International Phonetic Association*, pp. 1–30, 2021.
- [20] S. Morey, "Studying tones in North East India: Tai, Singpho and Tangsa," *Language Documentation & Conservation*, vol. 8, pp. 637–671, 2014, publisher: University of Hawai'i Press.
- [21] R. Winkelmann, J. Harrington, and K. Jnsch, "EMU-SDMS: Advanced speech database management and analysis in R," *Computer Speech & Language*, vol. 45, no. Supplement C, pp. 392 – 410, 2017.
- [22] P. Boersma and D. Weenink, "Praat: doing Phonetics by Computer," 2019, http://www.praat.org/.
- [23] P. Rose, "Considerations in the normalisation of the fundamental frequency of linguistic tone," *Speech Communication*, vol. 6, no. 4, pp. 343–352, Dec. 1987.
- [24] C. Kaland and T. M. Ellison, "Evaluating cluster analysis on f0 contours: an information theoretic approach on three languages."
- [25] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, no. 5, pp. 465–471, Sep. 1978.