Tonogenesis: the perception of tone and the role of place of articulation in Kurtöp

William Peralta[1]

[1]University of Sydney, Australia
william.peralta@sydney.edu.au

Abstract

Previous work on tonogenesis in Kurtöp production [1, 2] has addressed questions of propagation. They describe patterns whereby environments following more sonorous onsets appear to develop tone prior to less sonorous onsets. Additionally, the production study by Plane [2] revealed that place of articulation may also play a role in the order of tonogenetic sound change for production. Perceptual evidence for these processes in Kurtöp has not yet been researched. As such, this paper assesses how the sequence of tone propagation manifests in perception for Kurtöp.

A two-way forced choice perceptual identification task was designed to investigate the degree to which fundamental frequency (f_0) has phonologized in Kurtöp perception. It tested the hypothesis that contrastive tone develops following dorsal onsets prior to bilabial-dental onsets.

The results of the study show a division between the dorsal and bilabial-dental onset environments, whereby the dorsal onset environments exhibit more advanced phonologization of f_0 than the bilabial-dentals, which corroborate the findings from previous studies.

Index Terms: tonogenesis, place of articulation, Kurtöp, perception, phonologization, tone

1. Introduction

Kurtöp (Lhuntse dzongkhag, the Kingdom of Bhutan) is argued to be currently undergoing tonogenesis. Previous research has shown that this process had been completed in environments following the sonorants first, and subsequently following the palatal fricative [1]. These processes, however, are still being observed following the remaining obstruents. Research into such processes in speaker production have been undertaken, however, these trends have not been confirmed through perceptual evidence.

Research on tonogenesis has primarily focused on the retrospective historical development of tonal languages such as in Jabem [3], Athabaskan [4], or Chinese [5], on predictions of incipient development in non-tonal languages [6, 7], or on the phonetic precursors that initiate such change [8, 9]. However, the sequence in which such phonologization occurs as part of the tonogenetic process is understudied. For instance, we do not know whether tone spreads equally and simultaneously across the entire phonology of the language, or whether this is a gradual process affecting certain environments prior to others. Furthermore, if it is a gradual process, which environments are likely to undergo development before others, and is this tied to specific phonetic biases? As the literature argues, the tonal contrast in Kurtöp first developed following sonorants, subsequently expanded to the palatal fricative, and is currently spreading to the dental fricatives and all stops. Plane [2] confirms that the dental fricatives are phonologizing f_0 before the dental stops, but how is it developing in the category of stops?

The primary aim of this study is to assess how the sequence of tonal propagation manifests in perception for a language that is argued to be currently undergoing tonogenesis. For this paper, a perception experiment was undertaken to address the research question: do certain environments undergo phonologization of f_0 prior to others? The study thus tests the following hypothesis: the phonologization of f_0 following dorsal onsets occurs before bilabial-dental onsets.

2. Methods

To investigate the degree to which speakers of Kurtöp utilize pitch and voicing in perception, a two-way forced choice perception identification task was designed. Participants listened to 240 randomized stimuli such that no repetition of a token would occur in succession. Subsequently, they were presented with a choice between two alternatives – a visual representation of the word meanings of both the voiceless and voiced words within the minimal pair.

2.1. Stimuli

2.1.1. Recordings

The stimuli were recorded from one female and one male talker, both in their early thirties. This was to control for gender and ensure a balance sample in the stimuli. At the time of recording T1 resided in Canberra, Australia, but has lived in Bhutan for the majority of her life, while T2 has spent most of his life in Thimphu, Bhutan. T1 was recorded in Canberra, while T2 was recorded in Thimphu. Both recordings were made using a Zoom H4N recorder at a 24-bit 96kHz sampling rate into an uncompressed WAV format.

Twelve words were chosen for a balanced representation of place of articulation, consisting of initial stops at all five contrastive locations - labial, dental, retroflex, palatal, and velar. Only monosyllabic words were chosen for the experiment, to control for word stress and tone variation inherent in multisyllabic Kurtöp words. Possible durational influences were also controlled for by only using words with phonemically short vowels. All but one of the minimal pairs chosen for the study consists of the syllable structure CV the exception being the minimal pair /cɐt/ - /ɟɐt/ consisting of a CVC structure as there are no appropriate minimal pairs for the palatal stop in the CV structure. Lastly, the stimuli could not be fully controlled for vowel quality due to limitations in the lexicon, and for the perceptual clarity of lexical items.

10.21437/TAL.2018-17

2.1.2. Manipulation

Each word token was resynthesised in Praat [10] to generate five individual stimuli with different f_0 values. The f_0 values for the stimuli were determined by calculating the mean minimum and maximum f_0 of each talker. It was then rounded up and divided into five equidistant values. This was to ensure f_0 levels reflected accurate levels, and to maintain a natural sound for resynthesised tokens. For T1, the rounded mean maximum f_0 was 250Hz, and the rounded mean minimum f_0 equated to 210Hz. For T2, the rounded mean maximum f_0 was 170Hz, and the rounded mean minimum f_0 was 130Hz. The f_0 of the stimuli was flat throughout to maintain consistency for each set of minimal pairs. To test the naturalness of the tokens, a pilot experiment was also undertaken by a separate native speaker, who did not indicate any significant abnormality in the stimuli.

2.2. Participants

There were 24 participants in the perception study, all of whom are native speakers of Kurtöp, and reside in the capital of Bhutan, Thimphu. The mean age of the sample is 38.29 years (range = 13-73). This spans a 60-year difference that can be grouped into three categories representing 20 year intervals: 13-33 years, 34-54 years, and 55+ years. In this sample of 24 participants, sixteen were female, and eight were male. As with much of the population in Bhutan, all participants are multilingual. Participants are at least bilingual in Dzongkha (the national language of Bhutan) and Kurtp, with the majority also fluent in a varying combination of Nepali, English, Hindi, Tshangla, Chocangaca, Bumthap, Khengkha, and Tibetan.

2.3. Procedure

The two-way forced choice perception identification task required each participant to listen to a given stimulus (the presentation of which was facilitated by Praat) and subsequently respond by clicking on one of two alternative visual representations of the audio stimulus they had heard. For instance, if the stimulus presented was /gɐ/ at 250Hz, the participant would be presented with a visual representation of both /gɐ/ *'saddle'* and /kɐ/ *'snow'*. They were instructed to click on the picture that represented the word they heard from the audio stimulus. As such, if they responded by clicking on the picture representing /kɐ/ *'snow'*, the interpretation is that they relied on f_0 as their cue to differentiate between the minimal pair since the stimulus was in fact /gɐ/ but consisting of a relatively high f_0. However, if they chose /gɐ/ *'saddle'*, the interpretation is that they relied on voicing as their cue since it was still regarded as the voiced alternative in the minimal pair regardless of the f_0 value of the stimulus.

3. Results

A total of 5760 responses to the stimuli presented to the participants (480 per word) were collected and analysed. These responses reveal that the role of pitch in the perception of stimuli is statistically significant across all places of articulation - that is, pitch is used as a cue for words with obstruent initials. It further shows that place of articulation affects the amount of weight participants give to pitch as a perceptual cue in distinguishing between minimal pairs.

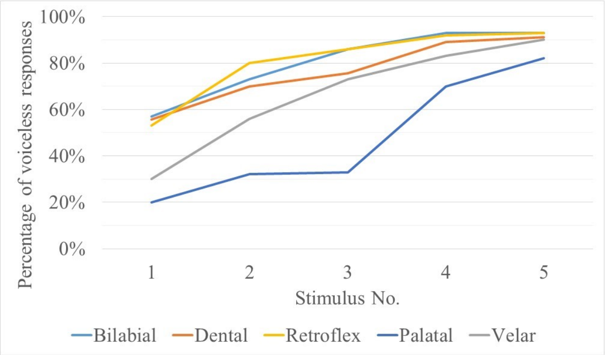

Figure 1: *Responses to voiceless stimuli*

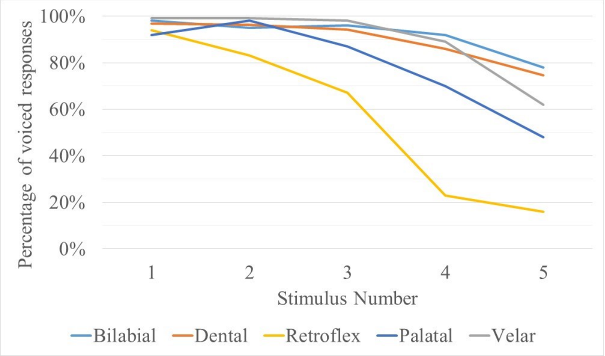

Figure 2: *Responses to voiced stimuli*

Figures 1-2 detail the responses to the stimuli (1-5) with voiceless and voiced onsets – stimulus 1 possesses the lowest f_0 value, and stimulus 5 the highest. The first observation here is that, within its own series of stimuli, there is a strong tendency for lower f_0 valued stimuli to be interpreted by participants as voiced, and for higher f_0 stimuli to be interpreted as voiceless. Furthermore, this trend occurs regardless of place of articulation. For example, in Figure 1, there is only 20% identification of the voiceless stimulus 1 as voiceless for the palatal tokens, whilst this percentage increases as the f_0 of the stimuli also increases. Participants therefore seem to be placing weight on f_0 as a cue in perception. Similarly, the inverse is seen with the voiced stimuli: tokens retain 90% and above voiced response rates with stimulus 1 (the lowest f_0 stimulus), which decreases as the f_0 of the stimulus increases. Again, this reveals that participants are placing weight on pitch as a cue for differentiation, which also seems greater in the voiceless series than the voiced.

3.1. Bilabial and dental articulations

There are, however, substantial differences in the perception of different onsets. Firstly, (as seen in Figure 1) the stimuli with labial and dental onsets seem to show relatively similar behaviours – clustering for both stimulus 1 and stimulus 5, with roughly 50-60% of participant responses being voiceless for stimulus 1 and 90% for stimulus 5. A similar trend is also observed for the voiced stimuli in Figure 2. Again, the labial and dental articulations tend to cluster together, and have been interpreted by participants mostly as voiced.

To further examine the strength of the relationship between pitch and the identification of voicing by participants, a χ^2 test of independence was performed, with α = 0.05 as the criterion for significance. The test was conducted to test the hypothesis (H_1) that the identification of the word was dependent on pitch. As such, the null hypothesis (H_0) is that the identification of the word within a minimal pair is independent of pitch.

For the voiceless bilabial stimuli, a significant relation was found, χ^2(4, N = 480) = 58.30, p = $<$ 0.001, and for the voiceless dental χ^2(4, N = 480) = 45.96, p = $<$ 0.001. This confirms that there is a strong correlation between the pitch of the stimuli and the identification of the voicing of the token. In other words, pitch is a significant factor in influencing the perception of stimuli. For their voiced counterparts, a significant relation was also found – bilabial: χ^2(4, N = 480) = 31.36, p = $<$ 0.001; dental: χ^2(4, N = 480) = 37.87, p = $<$ 0.001. Therefore, this is evidence to suggest the rejection of the null hypothesis.

3.2. Dorsal articulations

There is a clear division in the perceptual weight of pitch between the bilabial-dental cluster, and the dorsal articulations in both the voiceless and voiced series. Figures 1-2 reveal that pitch is being used as a perceptual cue for the dorsal places of articulation far more than the bilabial-dental cluster. In other words, the variation in pitch appears to be less significant as a perceptual cue in the bilabial and dental places of articulation when compared to the palatal and velar articulations. The χ^2 scores also confirm this: voiceless palatal, χ^2(4, N = 480) = 111.91, p = $<$ 0.001; voiced palatal, χ^2(4, N = 480) = 94.85, p = $<$ 0.001; voiceless velar, χ^2(4, N = 480) = 98.17, p = $<$ 0.001; voiced velar, χ^2(4, N = 480) = 99.03, p = $<$ 0.001.

3.3. Retroflex articulations

The results of the retroflex series do not seem to align easily with either the bilabial-dental or the dorsal articulations. Within the voiceless series, it appears to align with the bilabial-dental articulations: χ^2(4, N = 480) = 65.57, p = $<$ 0.001, whilst the voiced series displays the largest reliance on pitch and does not align with either the bilabial-dental or dorsal articulations: χ^2(4, N = 480) = 195.63, p = $<$ 0.001. It is possible to conflate the voiceless retroflex into the bilabial-dental grouping as a labio-coronal group, however, the idiosyncratic behaviour of the voiced retroflex must also be explained. Tables 1-2 summarize the χ^2 results for all places of articulation and voicing series:

Table 1: χ^2 *values for voiceless articulations*

Place of Articulation	χ^2 **value**
Bilabial	58.30
Dental	45.96
Retroflex	65.57
Palatal	111.91
Velar	98.17

$df = 4; \alpha = 0.05$
$\therefore critical value = 9.488$

Table 2: χ^2 *values for voiced articulations*

Place of Articulation	χ^2 **value**
Bilabial	31.36
Dental	37.87
Retroflex	195.63
Palatal	94.85
Velar	99.03

$df = 4; \alpha = 0.05$
$\therefore critical value = 9.488$

All results show statistical significance with varying degrees of strength depending on place of articulation. As such, the hypothesis Tables 1-2 demonstrate the close clustering of the labio-coronal group, and their separation from the dorsal places of articulation.

3.4. Age data

A division of the complete data set by age shows that the statistical significance of pitch as a cue for perception is not present for all generations of participants. The two youngest groups of participants (Group A and B) maintain consistent statistical significance across all words, while the oldest generation (Group C) does not show such consistency. Tables 3–4 below summarizes the results based on age:

Table 3: χ^2 *values for voiceless articulations (Group A: youngest; Group B: middle; Group C: oldest group)*

Place of Articulation	**A**	**B**	**C**
Bilabial	32.4821	25.1389	9.7222
Dental	38.8571	39.0831	26.6667
Retroflex	29.9582	45.2335	7.8947
Palatal	71.8621	37.3332	12.8
Velar	60.4609	50.8311	14.9816

$df = 4; \alpha = 0.05$
$\therefore critical value = 9.488$

Table 4: χ^2 *values for voiced articulations (Group A: youngest; Group B: middle; Group C: oldest group)*

Place of Articulation	**A**	**B**	**C**
Bilabial	18.3333	10.1335	0.9217
Dental	36.2642	33.5285	8.5106
Retroflex	91.226	65.8796	38.2418
Palatal	49.5743	45.9259	2.1739
Velar	49.3304	24.4244	7.8144

$df = 4; \alpha = 0.05$
$\therefore critical value = 9.488$

The apparent–time data seem to provide further evidence for the division of the bilabial-dental and dorsal articulations. There is a clear increase in the utilisation of f_0 as a perceptual cue as the age of the group decreases, and furthermore, there appears to be variation in the level of increase depending on the place of articulation. In Table 3 for example, the χ^2 value for the velar place of articulation of Group C is 14.9816, which increases to 50.8311 for Group B and to 60.4609 for Group A. On the other hand, the dental place of articulation appears to show a larger χ^2 value for Group C with 26.6667. However, this only increases to 39.0831 for Group B, and appears to slightly lower for Group A with a χ^2 value of 38.8571. With the exception of the voiced retroflex, these behaviours are also replicated in the voiced series (Table 4).

4. Discussion

Firstly, the results have shown that voicing is no longer the sole perceptual cue participants use to distinguish between the minimal pairs for all onset environments. This is indicative of a redundant overlap between voicing and f_0 cues which suggests the incipient phonologization of f_0 as the primary cue: a shift from voicing to f_0. This is further substantiated by a division of results by age, with the older generation showing the retention of a voicing primary perceptual cue, whilst the younger generation maintained greater levels of pitch use; a clear indicator of tonogenesis currently occurring in the language.

Next, the bilabial-dental articulations depicted less variation regardless of the f_0 value of the stimuli, and thus seem to retain use of voicing as their primary cue for discrimination. Inversely, for the dorsal articulations, participants appear to identify stimuli by their f_0 value and therefore seem to rely less on voicing to maintain the distinction between the minimal pairs. Consequently, although the previous data show that the trajectory of both groups seems to be towards an f_0 primary distinction, f_0 currently appears to be further phonologized in the dorsal group than in the bilabial-dental group.

The regularity in the division between the two groups may have its roots in phonetic origins, as place of articulation has previously been shown to correlate with the duration in which voicing can be maintained for stop obstruents. The build-up of oral pressure during an obstruent closure reduces the transglottal pressure within the vocal apparatus, therefore terminating voicing. The maximum duration for which voicing can be maintained depends on the place of articulation of the obstruent, as each possess a different oral capacity. The smaller capacity results in a more rapid termination of voicing, whereas a large capacity allows for a longer effective duration. The maximum volume of the vocal tract stretches from the lips to the larynx, and therefore, the labial obstruents operatively have the potential for the most voicing [11].

The idiosyncratic behavior of the voiced retroflex in the tonogenetic process in Kurtöp has also been noted in Plane [2]. In her study, the retroflex appeared to be the most phonologized environment in the production of tone, as participants in her study neutralized their voicing distinctions in favor of a tonal contrast at the retroflex place of articulation more than any other place of articulation. In light of the results from both this study and Plane [2] , why does f_0 appear to be phonologizing following the voiced retroflex prior to other places of articulation in Kurtöp?

This phenomenon may be the result of historical development. The simplification of complex onsets into the retroflex series /kr, khr, gr/ > /ʈ, ʈ^{h}, ɖ/ is a relatively recent sound change in Kurtöp [12]. Plane [2] suggests that there could be a correspondence between this simplification of complex onsets and the collapse of the voicing contrast in the retroflex place of articulation. If this is the case, the simultaneous simplification of complex onsets and occurrence of tonogenesis could have thus affected the voiced retroflex obstruent differently. Since the complex onsets were originally velar-plus-liquid, arguably tonogenesis could have been initiated while the sonorant was still present. Recall, however, that unlike its voiced counterpart, the voiceless retroflex is behaving as predicted by the hypothesis; that is, its relationship to f_0 is stronger than the bilabial and dental, but weaker than the palatal and velar. The reason for this remains unclear. However, one may propose that during the simultaneous development of the retroflex segments and a motion towards f_0 usage, the voiceless qualities of the voiceless segment were necessarily enhanced to maintain its contrast with the voiced segment since they were presumably both merging with the voiced liquid. Consequently, as the voiceless segment developed more voicelessness, the voiced retroflex possibly underwent similar enhancement (see [13]). This enhancement could have drawn characteristics from the voiced liquid, prompting its progression into tonogenesis. However, this hypothesis is mostly speculation and beyond the scope of this study. Future study into the specificities of the historical development of the retroflex series in Kurtöp may provide and explanation of this idiosyncrasy.

5. Conclusion

In summary, the results of the perception study confirm the hypothesis that the phonologization of f_0 following dorsal onsets occurs prior to bilabial-dental onsets in Kurtöp. It therefore suggests that there may be an organised order to which tonogenesis propagates within the sound system of a language. Results from the retroflex place of articulation reveal that there may also be a hierarchy of competing influences that determine the order to which the phonologization of f_0 is realized in tonogenesis in Kurtöp. This study raises the question of whether this sequence of propagation is unique to Kurtöp or perhaps can be observed in other languages currently undergoing tonogenesis. For example, it may be extended to other languages of the region Dakpa, Chali, Dzala, Bumthap, and Khengkha, which all appear to be developing tone in a similar manner [14, 15, 16].

6. Acknowledgements

I would like to thank Gwendolyn Hyslop and Sam Cook for their helpful discussion regarding this paper. Funding for fieldwork was received from the ARC Discovery Project Reconstructing Eastern Himalayan Histories: Languages, Plants and People (ARC DP 140103937). The study would not have been possible without their assistance.

7. References

[1] Hyslop, G. *Kurtöp Tone: A Tonogenetic Case Study*. Lingua, 119(6):827-845, 2009.

[2] Plane, S. M. *Role of Place and Manner in Tonogenesis: A Case Study with Kurtöp*. Master's Thesis. The University of Sydney. 2016.

[3] Bradshaw, J. *Obstruent Harmony and Tonogenesis in Jabem*. Lingua, 49(1):189-205, 1979.

[4] Leer, J. "Tonogenesis in Athabaskan", in *Proceedings of the Symposium Cross-Linguistic Studies of Tonal Phenomena, Tonogenesis, Typology, and Related Topics*, Tokyo, 1999.

[5] Sagart, L. *Austronesian Final Consonants and the Origin of Chinese Tones*. Oceanic Linguistics Special Publications, 24:47-59, 1993.

[6] Coetzee, A. W., P. S. Beddor, & D. P. Wissing. *Emergent Tonogenesis in Afrikaans*. The Journal of the Acoustical Society of America, 135(4):2421-2422, 2014.

[7] Kirby, J. *Incipient Tonogenesis in Phnom Penh Khmer: Acoustic and Perceptual Studies*. Journal of Phonetics, 43:69-85, 2014.

[8] Hombert, J., J. J. Ohala & W. G. Ewan. *Phonetic Explanations for the Development of Tones*. Language, 55(1):37-58, 1979.

[9] Kingston, J. ”The Phonetics of Athabaskan Tonogenesis”, in S. Hargus & K. Rice [Eds], *Athabaskan Prosody*, 137-184, J. Benjamins Publishing, 2005.

[10] Boersma, P. & D. Weenink. *Praat: Doing Phonetics by Computer*. Computer Program, Version 6.0.30, 2017.

[11] Ohala, J. J. & C. J. Riordan. ”Passive Vocal Tract Enlargement During Voiced Stops”, in J. J. Wolf & D. H. Klatt [Eds.], *Speech Communication Papers. Presented at the 97th Meeting of the Acoustical Society of America*, Massachusetts Institute of Technology, 1979.

[12] Hyslop, G. ”On the Internal Phylogeny of East Bodish” in G. Hyslop, S. Morey, & M. W. Post [Eds.], *North East Indian Linguistics Vol. 5*, 2013.

[13] Kirby, J. ”The Role of Probabilistic Enhancement in Phonologization” in A. C. L. Yu [Eds.], *Origins of Sound Change: Approaches to Phonologization*, p. 228-246, 2013.

[14] Hyslop, G. ”Sonorants, fricatives, and a tonogenetic typology”, in *LSA Annual Meeting Extended Abstracts*, 2010.

[15] Hyslop, G. & K. Tshering. ”Preliminary Notes on Dakpa (Tawang Monpa)”, in S. Morey & M. Post [Eds.], *North East Indian Linguistics Vol. 2*, 2010.

[16] van Driem, G. ”Dakpa and Dzala Form a Related Subgroup Within East Bodish, and Some Related Thoughts” in R. Bielmeier & F. Haller [Eds.], *Linguistics of the Himalayas and Beyond*, p. 71-84, 2007.