Tone and Vowel Interaction in Northern Lisu

Rael Stanley, Marija Tabain, Defen Yu, David Bradley

Department of Languages and Linguistics, La Trobe University, Melbourne, Australia

r.stanley@latrobe.edu.au, m.tabain@latrobe.edu.au, d.yu@latrobe.edu.au, d.bradley@latrobe.edu.au

Abstract

This study presents data for f0 and H1*-H2* in Northern Lisu Tones, and compares the results between back and front vowels. F0 and H1*-H2* tracks are extracted for four of the six tones in Northern Lisu - 33, 33̲, 21, 21̲ʔ (where underlining indicates creaky voice quality). These are compared between tones that appear on raised vowels /ɯ, u, ɤ, o/, front vowels /i, y, e, ø, ɛ/, and the retracted vowel /ɑ/. Results show that there is a difference in tone contour when comparing across the three different vowel contexts.

Index Terms: acoustic phonetics, tone, voice quality, Northern Lisu, Tibeto-Burman languages, vowels

1. Introduction

Lisu is a language in the Loloish branch of the Tibeto-Burman language family [1] and has speakers in Northeast India, Northern Myanmar, Southwest China, and Thailand [2]. There are three main dialects of Lisu - Northern, Central, and Southern [3].

Most varieties of Lisu have six tones, though some have fewer [2]. Northern Lisu is one of the varieties with six tones, and they are: high-level (55); mid-rising (35); mid-level (33); mid-level, creaky voiced (33̲); low-falling (21); and low-falling, creaky voiced with final glottal stop(21̲ʔ) [2]. However, speakers of Northern Lisu have a tendency to merge the 33 and 33̲ tones [3].

There are 10 monophthongal vowels in Northern Lisu /i, y, e, ø, ɛ, ɑ, ɤ, o, ɯ, u/, as well as a fricative vowel /ʐ̩/ [2]. However, the contrast between /ɯ/ and /ɤ/ is a marginal one. There are also diphthongs /iɑ/ and /uɑ/. However, these do not appear in native Lisu words and occur only in loanwords [4, 2, 3].

In Esling et. al.'s Laryngeal Articulator Model [5], the authors classify vowels into categories which account for both lingual and laryngeal properties. Using this model, the vowels of Northern Lisu fall into three categories: Front vowels /i, y, e, ø, ɛ/, Raised vowels /ɯ, u, ɤ, o/, and the Retracted vowel /ɑ/. According to this model, retracted vowels may be more likely to be affected by laryngeal constriction, resulting in a higher likelihood of creaky voice, we explore this possible interaction with a language that has a relatively large set of tonal contrasts, as well as front, raised, and retracted vowel contrasts.

2. Method

2.1. Speakers

All participants were native speakers of Northern Lisu, aged between 20 and 22 at the time of recording in 2017 (N=8, four female speakers and four male speakers). They were all born in the Nujiang Lisu Autonomous Prefecture, in Western/Northwestern Yunnan Province, China, and all were students at Yunnan Minzu University, except for one who was a student at Yunnan Normal University, both located in the Chenggong district of Kunming.

2.2. Word List

Participants were recorded reading from a list of target words in Northern Lisu that was generated using [4]. For this analysis, each word in the list was selected in such a way that each of Northern Lisu's ten monophthongal vowels (excluding the fricative vowel) was accompanied by each of the six tones. In order to minimise co-articulatory effects, every effort was made to choose words where a bilabial stop preceded the target vowels. However, this was not possible in every case. In total, there were 54 words in the list.

The words were then translated into Chinese orthography, so that any influence from Lisu orthography could be minimised.

Participants were asked to say each word three times in isolation, in order to avoid potential effects of tone sandhi. However, not all participants knew every word, and some words had multiple potential Northern Lisu words to choose from. Because of this, not every speaker produced tokens of every word listed, and some words produced were different to what may have been expected based on the entry in the dictionary. In these cases, these unexpected words were noted and compared with the Sino-Tibetan Etymological Dictionary and Thesaurus (STEDT) [6]. Where the production clearly matched one of the various definitions for the same word in the STEDT, it was included for analysis. In cases where the produced word did not match an entry in the STEDT, it was left out of the analysis, so that judgements on tone or vowel quality by a non-native speaker would not influence the results. In total there were 969 tokens analysed.

2.3. Recordings

Recordings were made using an H4n Zoom in a quiet hotel room in Kunming. All recordings were done in mono with a sampling rate of 48 kHz (with one exception, due to the Zoom running low on battery and automatically lowering the sampling rate to 24 kHz, which was later upsampled to 48 kHz using Audacity, so that the affected speaker could be included with other speakers in the analysis), and a bit depth of 1536 kbps. These files were saved as uncompressed .wav files.

2.4. Analyses

Prior to analysis, each participant's recording was split into individual utterances using Praat [7], by utilising scripts to detect

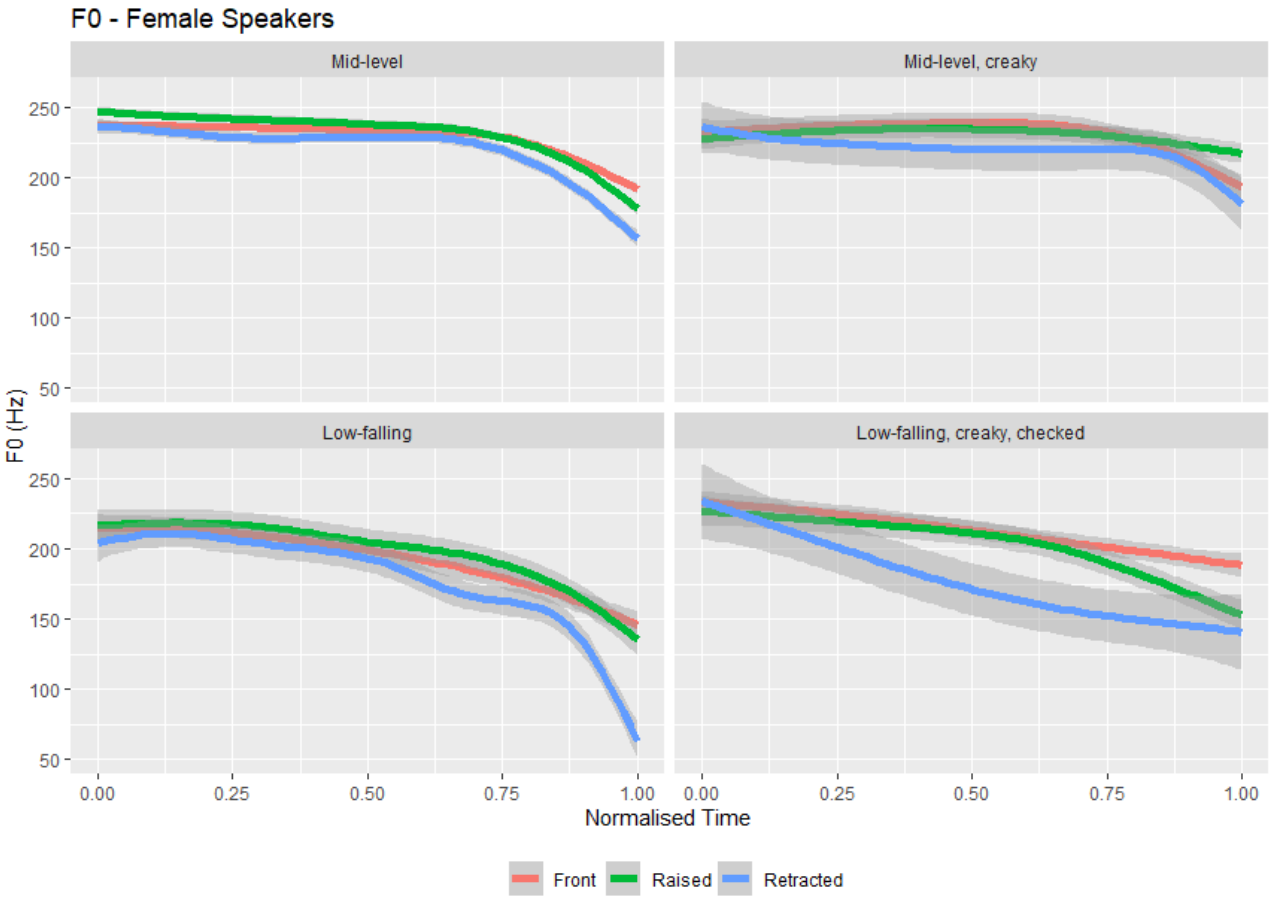

Figure 1: F0 across normalised time for female speakers of Northern Lisu

pauses, mark boundaries, and split the file into individual words [8]. TextGrids were then created using Praat, with tiers: Word, Sex, Tone, Repetition, Segment, Speaker, and Syllable.

Pitch traces across normalised time were produced using the emuR package [9] in R [10]. H1*-H2* values were produced using VoiceSauce [11] in MATLAB [12]. Plots were made using the ggplot2 [13] package. F0 and H1*-H2* data will be presented as GAM (Generalised Additive Model)-smoothed trajectories, with grey intervals indicating 95% confidence levels.

3. Results

3.1. Female Speakers - F0

As can be seen in the top panels of Figure 1, there is little difference in f0 contour for mid (top-left panel) versus mid-level creaky (top-right panel) tones, of female speakers. In both these tones, f0 values in the retracted vowel context (blue line) are overall lower than the front (red line) or raised (green line) vowel contexts. For the mid-level tone, f0 values in the retracted vowel context are significantly lower than in either of the raised or front vowel contexts from approximately 70% of total duration, until the end of the vowel.

In the low-falling (bottom-left panel) and low-falling creaky checked (bottom-right panel) tones, f0 values are also overall lower in the retracted vowel context than in the other vowel contexts. In the low-falling tone, there is a signifcant drifference from approximately 70% of total duration until the end of the vowel, as with the mid-level tone.

The low-falling creaky checked tone has a more complicated difference in these vowel contexts than the other tones. Between approximately 35% and 95% of total duration, f0 in the retracted vowel context is significantly lower than in the raised or front vowel contexts. From approximately 80% of total duration until the end of the vowel, the low-falling creaky checked tone has a significantly higher f0 in the front vowel context than either of the other vowel contexts.

3.2. Male Speakers - F0

As can be seen in Figure 2, male speakers produce all four of the tones examined with an overall lower f0 in the retracted vowel context as opposed to the front or raised vowel contexts, like the female speakers.

For the mid-level tone (top-left panel), productions in the retracted vowel context are significantly lower than in the front or raised vowel contexts, from the beginning of the vowel until approximately 75% of total duration, after which point there is no difference between the retracted vowel context and the raised vowel context. From approximately 70% of total duration until the end of the vowel, the mid-level tone has a significantly higher f0 in the front vowel context than in the other vowel contexts.

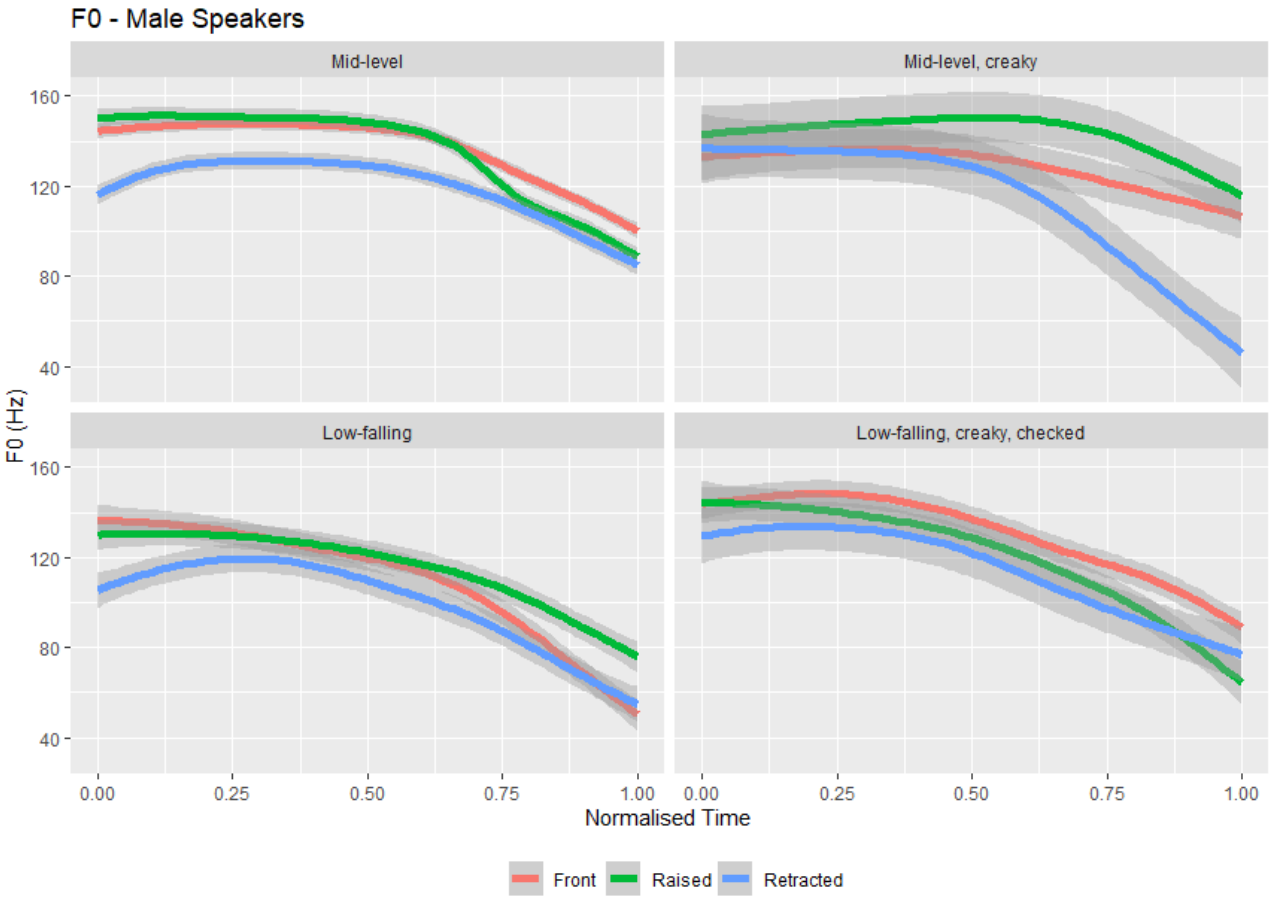

Figure 2: F0 across normalised time for male speakers of Northern Lisu

For the mid-level creaky tone (top-right panel) f0 of productions in the retracted vowel context are significantly lower than in the other vowel contexts from approximately 70% of total duration, until the end of the vowel.

For the low-falling tone (bottom-left panel), f0 of in the retracted vowel context is significantly lower for the first 20% of total duration, and for the last 20% of duration f0 is significantly higher in the front vowel context than in the other vowel contexts.

While f0 of the low-falling creaky checked tone is overall lower in the retracted vowel context, it is only significantly different when compared to the front vowel context and not the raised vowel context.

3.3. Male Speakers - H1*-H2*

As can be seen in Figure 3, H1*-H2* values for all tones except for the low-falling creaky checked tone are overall lower in the retracted vowel context than in the front or raised vowel contexts.

For the mid-level tone (top-left panel), H1*-H2* values are significantly lower in the retracted vowel context for almost the entire duration of the vowel. There is also a short portion of the vowel for which H1*-H2* values are significantly higher in the raised vowel context than in the front vowel context, from approximately 5% to 15% of total duration.

For the mid-level creaky tone (top-right panel), there is no significant difference in H1*-H2* values between productions in the retracted vowel context and the front vowel context. However, H1*-H2* values for both these vowel contexts are significantly lower than the raised vowel context from approximately 25% of total duration until the end of the vowel (aside from a small degree of overlap in the grey confidence intervals between the front and raised vowel contexts between 55% and 80% of total duration).

For the low-falling tone (bottom-left panel), H1*-H2* values in the retracted vowel context are significantly lower than the other two vowel contexts from the start of the vowel until approximately 20% of total duration. After this point, there is no significant difference between the retracted and front vowel contexts. However, both these vowel contexts are significantly lower in H1*-H2* values than the raised vowel context from approximately 15% of total duration until the end of the vowel.

For the low-falling creaky checked tone (bottom-right panel), there is no difference in H1*-H2* values between the retracted and raised vowel contexts. However, between approximately 15% and 50% of total duration H1*-H2* values are higher in the front vowel context than in either the raised ro retracted vowel contexts.

Due to reasons of space, only male data is presented here. The female plots show differences based on vowel context. However, it appears that tones in the retracted vowel context tend to be breathier (that is, displaying an overall higher H1*-

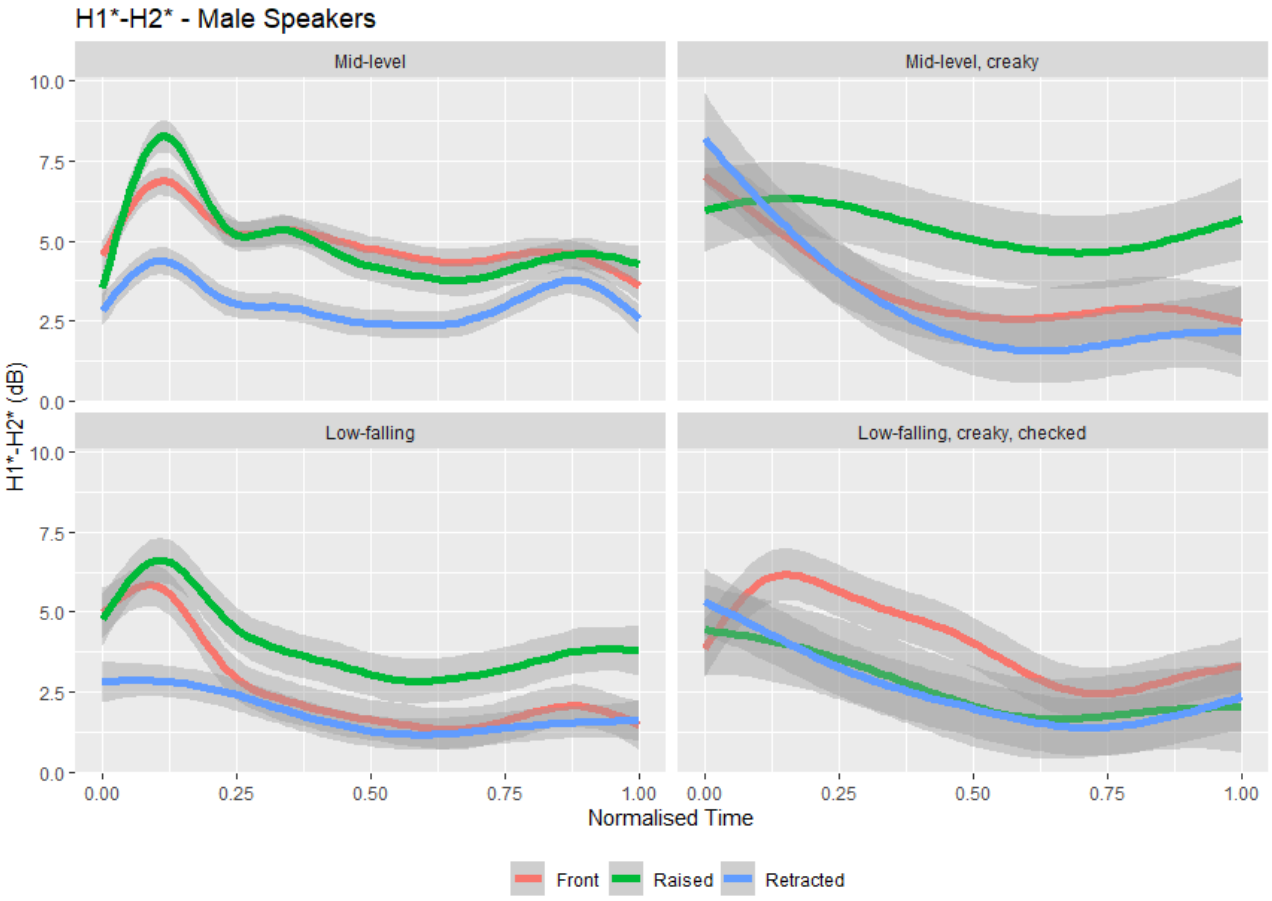

Figure 3: H1*-H2* across normalised time for male speakers of Northern Lisu

H2*) than in the other vowel contexts.

4. Discussion

As shown above, vowel context affects f0 in both the mid-level tone pairs and the low-falling tone pairs, in Northern Lisu, resulting in lower overall f0 for both in the retracted vowel context. It has a greater effect on the f0 of low-falling, creaky, checked tones, making them lower overall than the low-falling tones, for female speakers. For male speakers, this effect is stronger in the mid-level and mid-level creaky tones.

Female speakers displayed almost no difference between the mid-level and the mid-level, creaky tones, overall. This indicates that for female speakers, there is further evidence of merger between these two tone categories.

Male speakers, however display larger differences between mid-level and mid-level, creaky tones. This may be another example where female speakers appear to be leading a sound change in progress (i.e. the female speakers appear to be merging the mid-level and mid-level creaky tones, as had previously been noted by Bradley [3]).

These results also fit with Esling et. al.'s [5] conclusion that the laryngeal constriction involved in producing retracted vowels shortens vocal folds and lowers pitch. There is also some evidence for this constriction influencing creakiness, with H1*-H2* values in this vowel context being overall lower in even the modal tones. Although, the nature of this interaction is, as Esling states, a complex one.

5. Acknowledgements

We would like to thank He Jian Xin and Hu Shi Yu for their patience and assistance in the data recording, as well as the participants for their hard work.

6. References

[1] Bradley, D., "Lisu", in G. Thurgood and R. LaPolla [Eds], The Sino-Tibetan Languages, 222-235, Routledge, 2003.

[2] Yu, D., Aspects of Lisu Phonology and Grammar, La Trobe, 2007.

[3] Bradley, D., "Lisu Language", in R. Sybesma, W. Behr, Y. Gu, Z. Handel, and C. -T. Huang [Eds], Encyclopedia of Chinese Languages and Linguistics, Brill, 2018.

[4] Bradley, D., A Dictionary of the Northern Dialect of Lisu (China and Southeast Asia), Australian National University, 1994.

[5] Esling, J.H., Moisik, S.R., Benner, A., and Crevier-Buchman, L., The Laryngeal Articulator Model, Cambridge University Press, 2019.

[6] STEDT Project, "Sino-Tibetan Etymological Dictionary and Thesaurus", Online: https://stedt.berkeley.edu, accessed on 4 May 2022.

[7] Boersma, P. and Weenink, D., Praat: doing phonetics by computer [Computer program]., Verstion 6.2.12, Online: http://www.praat.org/, retrieved 17 April 2022.

[8] Lennes, M., SpeCT - Speech Corpus Toolkit for Praat (v1.0.0), Online: https://doi.org/10.5281/zenodo.375923, accessed 4 May 2022.

[9] Winkelmann, R., Jaensch, J., Cassidy, S., and Harrington, J., emuR: Main Package of the EMU Speech Database Management System., R package version 2.3.0, 2021.

[10] R Core Team, R: A Language and Environment for Statistical Computing., Online: https://www.R-project.org., accessed 4 May 2022.

[11] Shue, Y.-L., Keating, P., Vicenik, C., and Yu, K., VoiceSauce: A Program for Voice Analysis, Proceedings of the ICPhS XVII, 1846-1849, 2011.

[12] MATLAB., version 7.10.0 (R2010a), Natick, Massachusetts: The MathWorks Inc., 2010.

[13] Wickham, H., ggplot2: Elegant Graphics for Data Analysis., Springer-Verlag., Online: https://ggplot2.tidyverse.org., accessed 4 May 2022.

[14] Pinheiro, J.C. and Bates, D.M., Mixed-Effects Models in S and S-Plus., Springer, doi: 10.1007/b98882., accessed 4 May 2022.