PHONETIC IMPLEMENTATION OF PHONOLOGICALLY DIFFERENT HIGH TONE SPANS IN LUGANDA

Cerys Hughes, Seung Suk Lee, Alessa Farinella & Kristine M. Yu

University of Massachusetts, Amherst, USA
ceryshughes@umass.edu; seungsuklee@umass.edu; afarinella@umass.edu; krisyu@linguist.umass.edu

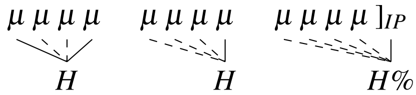
ABSTRACT

Substantial phonetic work exists both on long-distance phonological processes and on incomplete neutralization for segments, but not for tone. To help fill these gaps, we re-analyze Myers et al. 2018's Luganda production experiment on high tone spans derived by distinct phonological processes. They found differences in peak delay between two kinds of surface-identical high tone spans.We build on their analysis by analyzing the shape of the whole high span f0 trajectory using functional data analysis (FDA) and generalized additive mixed models (GAMMs). A previous reanalysis suggested the tone spans also differ in degree of sag in their plateau. We found this wasn't supported with different f0 normalization and incorporating richer random effects in trajectory shapes with GAMMs, although GAMMs supported different rise shapes to the plateau. In addition to addressing an empirical gap, we thus provide a cautionary tale on studying incomplete neutralization in dynamic trajectories.

Keywords: Tone, phonetics-phonology interface, incomplete neutralization, FDA, GAMMs

1. INTRODUCTION

A string of like-tones in the surface representation, e.g., a high tone span, can arise from multiple phonological tonal spreading processes. For example, [1] describes high tone spans in Luganda derived from three different spreading processes, as schematized in Fig. 1. Tones in Luganda are either high (H) or low (L), with both lexical and boundary tones. HH spans are high tone plateaus, which occur between two lexical high tones that spread to all of the syllables between them (subject to syntactic conditioning). LH spans occur when there is a low tone followed by a lexical high tone within a phrase and the high tone spreads leftward. LL high tone spans occur when there are no lexical high tones but instead an intonational H% that spreads leftward. Our current investigation does not focus on an additional "short" span discussed in [1], HL, but we include it in our replication of [2].



(a) HH (b) LH (c) LL

Figure 1: Relevant high tone span types in Luganda from [1], each derived by different tone spreading processes.

There is some work on the phonetic implementation of sequences of like-tones in languages with lexical and/or grammatical tone where each tone bearing unit (TBU) in the tone span is already specified in the underlying representation (e.g., in Yoruba and Mandarin [3, 4]). But [1] is, to our knowledge, the only instrumental study of like-tone spans arising from different spreading processes. It offers a case study for exploring whether underlying phonological contrasts are incompletely neutralized in spans with the same surface tones, in line with findings of incomplete neutralization of segmental contrasts.

There is a proliferation of studies of segmental incomplete neutralization, as evidenced by the feasibility of [5]'s meta-analysis of fourteen studies of German final devoicing alone. In contrast, there are just a handful of studies of incomplete neutralization of tonal processes, and these seem to be limited to a few Chinese varieties. Unlike the point-based acoustic measures of interest in segmental studies of incomplete neutralization such as in devoicing or flapping, the acoustic object of interest in tonal studies is the f0 trajectory shape. Statistical methods for analyzing shape components of trajectories, e.g., smoothing spline ANOVAs, functional data analysis (FDA) [6], and generalized additive mixed effects models (GAMMs) [7], have enabled prosodic analyses to address f0 trajectory shape in addition to f0-based measures at inflection points [8, 9, 10, 11], a.o., including in studies of incomplete neutralization in tonal sandhi [12, 13].

[1] focused on f0 turning points and pointwise measures in their analysis of high tone spans in

Luganda, which characterized the timing and f0 excursion size of the rise into and fall out of the high f0 plateau. They found that the lexical spans (e.g. HH and LH) have similar normalized f0 levels, HH and LH spans have a long plateau between f0 rise and fall, and that intonational span (LL) f0 contours are below those of the lexical spans (see Fig. 4). [1] also found that LH spans have more peak delay (with respect to the onset of the syllable initiating the H span) relative to all other classes, and took this as evidence that underlying phonological contrasts are not fully neutralized between LH and HH spans, which differ only in whether the initial syllable of the H span is underlyingly H-toned or not.

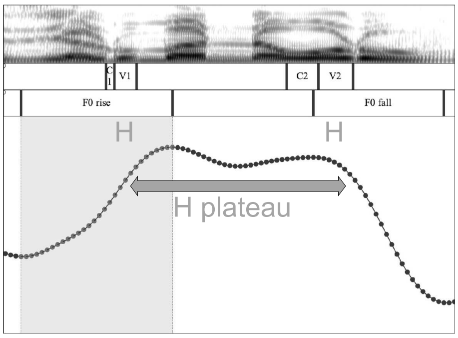


Figure 2: An HH token with noticeable sag in its f0 plateau, figure adapted from [1], Fig. 4

While [1] did not analyze the f0 trajectory over the high f0 plateau, a re-analysis of the same data from [1] in [2, 14] noticed that exemplar f0 contours of HH and LH spans in [1] both showed "sag" (a dip) in the middle of the plateau, with a greater degree of sag for the HH tokens (Fig. 2). This "sag" is reminiscent of the description of a "sagging" transition between two high tone targets in American English intonation in [15] that is a function of the distance between them, with greater distance resulting in a greater degree of sag, see also [16, 17]. [2, 14] used FDA to characterize shape components of the high tone plateau f0 trajectories and provided additional evidence for incomplete neutralization between HH and LH spans, due to more sag in the plateau of HH spans than LH spans.

This paper shows that: (i) [2, 14]'s finding of a difference in the degree of sag of the high plateaus of HH and LH span is not supported with the addition of richer random effects structures, (ii) but that GAMMs support a difference in the shape of the rise trajectory into the plateau which could have a perceptual effect [18] consistent with [1]'s reported acoustic LH-HH peak delay difference. We also make a methodological contribution in providing a case study of the effect of researcher degrees of freedom in assessing incomplete neutralization, an issue raised in [5]'s meta-analysis of German final devoicing. Specifically, the findings about sag and incomplete neutralization are greatly impacted by the choices of including speaker or item as random effects and of normalizing f0 by speaker or token.

2. METHODS

We re-analyzed the recordings of 10 native speakers of Luganda from [1], also analyzed in [2, 14]. Data consisted of approximately 20 sentences per span type for each of the four span types (the 3 shown in Fig. 1 in addition to the short span HL) for a total of 80 items. All items were elicited from all 10 speakers, and [1] included 798 tokens for analysis in total. We additionally discarded 4 HL tokens with span durations greater than 3SD higher than the mean HL span duration, and 1 LL token with heavy laryngealization. We followed the data pre-processing procedure in [2, 14]. F0 measurements were extracted using speaker-specific f0 floors and ceilings with STRAIGHT implemented in VoiceSauce [19]. A handful of clear gross f0 tracking errors were manually corrected. For both FDA and GAMMs, the analysis window was the span duration—the interval from the rise onset through the fall offset marked in Praat TextGrids [20] from [1] (Fig. 2).

FDA studies of f0 contours have varied in how they process f0, e.g., raw f0 values in Hz [9], conversion to semitones and centering about the mean within each token [12, 13], and by-speaker z-score normalization [10, 11]. While [2, 14] followed [12, 13]'s within-token mean-centering, [1] used by-speaker z-score normalization. For comparison to previous analyses of [1]'s data, we therefore ran parallel analyses with both types of f0 transformation for FDA and GAMMs. Span type contrasts were defined as treatment contrasts with HH-span as the reference level, since we were interested in comparing the LH-span (and other span types) to the HH-span. Statistical significance was assessed at an α-level of 0.05.

Data analysis was done in R [21]. Like in [2, 14], FDA was implemented with the *fda* package [6] on the full data set, including HL spans, B-spline basis functions were used for smoothing F0 measurements and parameterizing the principal components, and mixed effects regressions were conducted using *lme4* [22] and *lmerTest* [23]. For the long span subset (we obtained the same pattern of results with the full data set) we built GAMMs with by-speaker z-scored f0 values or by-token centered f0 values (30 evenly spaced samples, with normalized time defined as proportion of span duration) as the dependent variable, span and duration as parametric variables, an interaction between proportion of span duration and span type, and random smooths of proportion of span duration for participants and items with *mgcv* [7]. We corrected for autocorrelation in the residuals (due to the non-independence of time series data) and heavy tails in the distribution of measurements observed in GAMMs diagnostics by including an autoregressive (AR1) model and specifying a scaled-*t* distribution [24]. We coded the interaction between proportion of span duration and span type as nominal factor difference smooths for *itsadug* [25] visualization of f0 trajectory shape differences, and as ordered factor difference smooths for inferential tests on the difference between spans relative to the HH reference level [26]. Further details for data processing and modeling (including for other span type comparisons besides HH vs. LH), can be found in the OSF repository at https://osf.io/p4gj5/?view_only=95e88cf6724447fc93eb8684723762b6.

3. RESULTS

3.1. FDA

We first replicated [2, 14]'s FDA procedure using within-token f0 centering and data from all four tone spans. FDA with by-speaker z-scored f0 values instead resulted in a fourth functional principal component (PC4) (5.8% of variability in data, see Fig. 3) that looked nearly identical to the PC3 indexing sag in [2, 14]. A higher coefficient for these PCs indexes a decrease in degree of sag.

A mixed effects logistic regression model in [14, 2] with span type as the dependent variable, the PCs as non-interacted fixed effects, and a by-speaker random intercept showed that more sag (as indexed by a PC like in Fig. 3) significantly increased the probability of HH relative to an LH span. However, once we also included by-item random intercepts, none of the PCs reached significance for classifying span type. Similarly, for the by-speaker f0 normalized data, the PC4 sag component was significant when only by-speaker random intercepts were included, but no PCs were significant when by-item intercepts were included (the model didn't converge with both types of random intercepts).

Since incomplete neutralization studies typically estimate potential differences in acoustic measures between different categories, we also ran linear regressions with each of the sag components (PC3 for by-token centering; PC4 for by-speaker

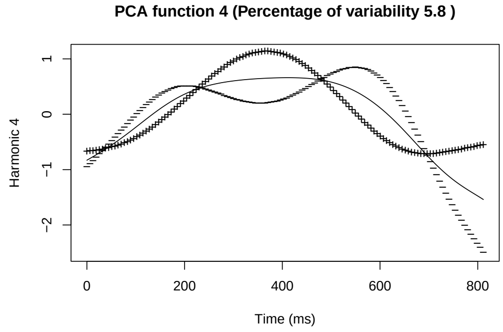


Figure 3: PC4 from by-speaker f0 normalization indexes a sag shape component nearly identical to one found in [2, 14] with within-token f0 centering. The minus sign curve indicates the effect on the resulting curve when the PC's coefficient is decreased, and the plus sign curve shows the effect when it is increased.

normalization) as the dependent variable, span type interacted with span duration as the fixed effects, and random intercepts by-speaker and by-item (random slopes models did not converge). For the by-token centering model, all span types, duration, and span type-duration interactions were significant. In particular, for the LH vs. reference level HH contrast, the sag component coefficient was higher for LH (less sag) than HH spans, particularly for shorter span durations. However, for the by-speaker normalization model, the LH vs. HH span contrast was marginal (β=11.2, (SE 3.5), t = 1.8, p = 0.07) and its interaction with duration was insignificant. Thus, while both f0 normalization choices showed the same direction of effects for the sag component, they differed in whether LH vs. HH span significantly affected the degree of sag. Regression diagnostics indicated non-normality—a heavy-tailed distribution for both linear models, which we addressed with GAMMs.

3.2. GAMMs

With richer random effects structures in GAMMs (by-speaker and by-item adjustments for trajectory shape), as well as corrections for heavy tails and the non-independence of f0 values within the same time course, differences in the degree of sag between the LH and HH spans are not supported with either f0 normalization choice. Estimated GAMMs partial effects for the two kinds of normalization are in Fig. 4.

The LL span trajectory is clearly lower than other span trajectories, as found in [1]. The overlapping confidence intervals between the LH

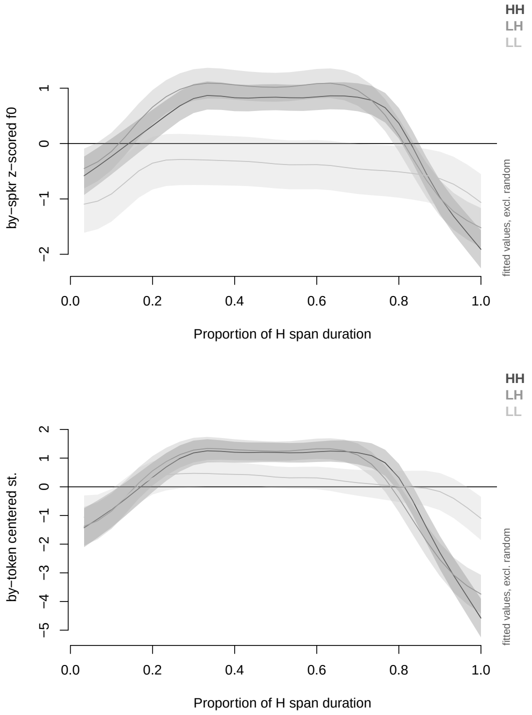


Figure 4: Estimated GAMMs partial effects of span type among long spans. (top) by-speaker z-scores (top); (bottom) by-token centered st.

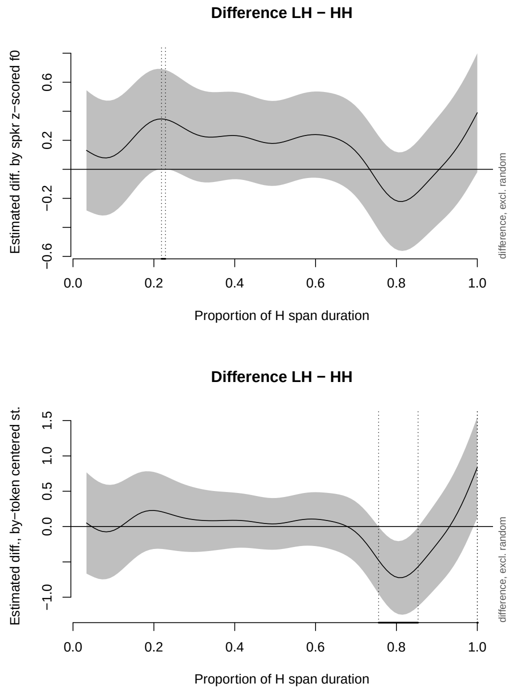


Figure 5: GAMMs estimated differences for LH span trajectories relative to HH ones; pointwise 95% CIs shown as shaded bands; by-speaker z-scored f0 (top); by-token centered st. (bottom)

and HH span plateaus indicate a lack of support for differences in trajectory shape in the plateau (formally assessed by non-significance of the LH-HH ordered factor difference smooths), but suggest potential differences in the rise/fall. We can better visualize the regions of differences with the difference smooths between LH and HH shown in Fig. 5. The difference smooth for by-speaker z-scores shows that significant differences between the LH and HH spans occur only in the rise to the plateau—the dashed region at 22-24% of the H span duration—but the difference smooth for by-token centered st. shows that significant differences occur only in the fall out of the plateau, at about 80% of the H span duration. While the robustness of these small differences at the contour edges is unclear, they demonstrate the impact of normalization choice.

By combining Figs. 4, 5, we can see that the tiny shape difference between the by-speaker z-scored LH and HH spans is that the LH span rise into the plateau bulges out more (a "domey" shape) than the straighter HH span rise (perceptually consistent with [1]'s peak delay difference finding [18]). But for by-token centered st., the difference is that the LH span fall bulges out downward more than the straighter HH span fall. (Differences at the end are likely an artifact of edge effects in f0 estimation.)

4. DISCUSSION AND CONCLUSION

GAMMs do support incomplete neutralization between LH and HH spans through small but consistent f0 differences. However, the effect of f0 standardization choice on whether the difference was in the rise or fall demonstrates the impact of analysis choices in incomplete neutralization studies. It's not obvious that one choice is more theoretically-motivated than the other. Together, with our failure to replicate the difference in degree of sag between the two spans reported in [2, 14] with richer random effects structures, these results show the need for caution in interpreting implications for phonological representations based on small phonetic implementation differences. As we have shown, even where phonetic differences appear to be statistically significant, different methodological choices can result in non-significant effects. A broader concern is that even statistically significant differences in phonetic implementation do not necessarily support the existence of underlying phonological contrasts [27]. Irrespective of incomplete neutralization effects, our trajectory-shape analyses enrich our understanding of the very limited data available on like-tone spans arising from different spreading processes.

5. ACKNOWLEDGEMENTS

We thank Scott Myers for helpful correspondence as well as sharing the Luganda tone span data he collected and analyzed.

6. REFERENCES

[1] S. Myers, E. Selkirk, and Y. Fainleib, "Phonetic implementation of high-tone spans in Luganda," *Laboratory Phonology*, vol. 9, no. 1, p. 19, Dec. 2018.
[2] C. Hughes, A. Farinella, S. S. Lee, and K. M. Yu, "Phonetic implementation of phonologically different high tone plateaus in Luganda," LabPhon 18, Virtual, Jun. 2022.
[3] Y. O. Laniran, "Intonation in tone languages: The phonetic implementation of tones in Yoruba," Ph.D. dissertation, Cornell University, 1992.
[4] C. Shih, "A declination model of Mandarin Chinese," in *Intonation: Analysis, Modeling and Technology*, A. Botinis, Ed. Kluwer Academic Publishers, 2000, pp. 243–268.
[5] B. Nicenboim, T. B. Roettger, and S. Vasishth, "Using meta-analysis for evidence synthesis: The case of incomplete neutralization in German," *J. Phon.*, vol. 70, pp. 39–55, Sep. 2018.
[6] J. Ramsay and B. Silverman, *Functional Data Analysis*, 2nd ed. New York, N.Y.: Springer Verlag, 2005.
[7] S. Wood, *Generalized Additive Models: An Introduction with R*, 2nd ed. Chapman and Hall/CRC, 2017.
[8] J. E. Andruski and J. Costello, "Using polynomial equations to model pitch contour shape in lexical tones: An example from Green Mong," *JIPA*, vol. 34, no. 02, pp. 125–140, 2004.
[9] J. A. D. Aston, J.-M. Chiou, and J. P. Evans, "Linguistic pitch analysis using functional principal component mixed effect models," *Journal of the Royal Statistical Society: Series C*, vol. 59, no. 2, pp. 297–317, 2010.
[10] G. Lohfink, A. Katsika, and A. Arvaniti, "Variability and category overlap in the realization of intonation," in *Proc. ICPhS 19 Melbourne, Australia 2019*, S. Calhoun, P. Escudero, M. Tabain, and P. Warren, Eds., 2019, pp. 701–705.
[11] S. Gryllia, K. Marcoux, K. Jepson, and A. Arvaniti, "The many shapes of H*," in *Speech Prosody 2022*, May 2022, pp. 754–758.
[12] C. Cheng, J.-Y. Chen, and M. Gubian, "Are Mandarin Sandhi Tone 3 and Tone 2 the Same or Different? The Results of Functional Data Analysis," in *Proceedings of PACLIC27*, 2013, pp. 296–301.
[13] M. Gubian, F. Torreira, and L. Boves, "Using Functional Data Analysis for investigating multidimensional dynamic phonetic contrasts," *J. Phon.*, vol. 49, pp. 16–40, Mar. 2015.
[14] S. S. Lee, A. Farinella, C. Hughes, and K. M. Yu, "Phonetic implementation of phonologically different high tone plateaus in Luganda," 1st Intl. Conf. on Tone and Intonation, University of Southern Denmark, Dec. 2021.
[15] J. Pierrehumbert, "The phonology and phonetics of English intonation," Ph.D. dissertation, MIT, Cambridge, MA, 1980.
[16] ——, "Synthesizing intonation," *JASA*, vol. 70, no. 4, pp. 985–995, 1981.
[17] D. R. Ladd and A. Schepman, ""Sagging transitions" between high pitch accents in English: Experimental evidence," *J. Phon.*, vol. 31, pp. 81–112, 2003.
[18] J. Barnes, N. Veilleux, A. Brugos, and S. Shattuck-Hufnagel, "Tonal Center of Gravity: A global approach to tonal implementation in a level-based intonational phonology," *Laboratory Phonology*, vol. 3, no. 2, Jan. 2012.
[19] Y.-L. Shue, P. Keating, C. Vicenik, and K. Yu, "Voicesauce: A program for voice analysis," *Proceedings of ICPhS XVI*, 2011.
[20] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer (Version 6.3.03) [Computer program]." 2022.
[21] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2022.
[22] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.
[23] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, "lmerTest package: Tests in linear mixed effects models," *Journal of Statistical Software*, vol. 82, no. 13, pp. 1–26, 2017.
[24] J. van Rij, P. Hendriks, H. van Rijn, R. H. Baayen, and S. N. Wood, "Analyzing the Time Course of Pupillometric Data," *Trends in Hearing*, vol. 23, p. 233121651983248, Jan. 2019.
[25] J. van Rij, M. Wieling, R. H. Baayen, and H. van Rijn, "itsadug: Interpreting time series and autocorrelated data using GAMMs," 2022.
[26] G. Simpson, "Comparing smooths in factor-smooth interactions II," https://www.fromthebottomoftheheap.net/2017/12/14/difference-splines-ii/, Dec. 2017.
[27] J. A. Shaw and S. Kawahara, "Assessing surface phonological specification through simulation and classification of phonetic trajectories," *Phonology*, vol. 35, no. 3, pp. 481–522, 2018.