THE INTERACTION OF TIMING AND SCALING IN A LEXICAL TONE SYSTEM: AN EXAMPLE FROM SHILLUK

Jonathan Barnes[1], Nanette Veilleux[2], Alejna Brugos[1], Stefanie Shattuck-Hufnagel[3]
[1] Boston University, [2] Simmons College, [3] Massachusetts Institute of Technology
jabarnes@bu.edu, abrugos@bu.edu, veilleux@simmons.edu, sshuf@mit.edu

ABSTRACT

This paper presents data demonstrating that in Shilluk (Western Nilotic, South Sudan), a tonal contrast previously described as a typologically unusual distinction between two falling contours of identical shape and magnitude, differing only in the timing of the fall within the syllable, in fact involves distinctions in both the F0 timing and scaling domains. This interaction, furthermore, resembles a pattern that is common cross-linguistically, but difficult to account for given the strict separation between tonal timing and scaling patterns embodied in autosegmental representations. We offer a unified account both of this pattern ("later = higher"), and of its opposite ("earlier = higher"), also apparently attested. Our account turns on perceptual integration of disparate acoustic properties of the F0 contour simultaneously influencing perceived timing and scaling of tonal targets. We suggest that traditional AM approaches to the phonetics-phonology interface require revision in light of such evidence.

Keywords: tone perception, F0 timing/scaling Shilluk

1. INTRODUCTION

Within the Autosegmental-Metrical tradition, scaling and timing of tonal targets are orthogonal dimensions of phonological representation. As illustrated equally in (1) and (2) below for the lexical tone melodies of Mende nominals (Leben 1973), and realizations of the English "incredulity contour" (Ladd 1996/2008), there is an important sense in which we want to be able abstract tone strings away from their segmental associations, to capture the fundamental phonological sameness of F0 contours that are nonetheless shaped very differently on the surface. An H is an H, in other words, whether associated to one Tone-Bearing Unit, or to more.

(1) Lexical LH contours in Mende

```
 L  H        L  H          L  H
  \/         |  |          | |\
mba 'rice'   fande 'cotton'  ndavula 'sling'
```

(2) The English Incredulity Contour

```
L+H* L-H%            L+H*       L-H%
Sue?!               A  driving  instructor?
```

This basic insight has been carried over in AM theory to the study of phonetic realization as well. In the standard target-and-interpolation model (e.g., Pierrehumbert & Beckman 1988), each phonological tone in a string projects a target, to be implemented at a particular time, and at a particular pitch level. Targets are frequently identified with F0 turning points, between which the contour is simply interpolated along something like the shortest or articulatorily cheapest path. As Bruce (1977:132) memorably puts it "… reaching a certain pitch level at a particular point in time is the important thing, not the movement (rise or fall) itself."

1.1 F0 Scaling and Timing Interactions

Timing and scaling may be separable at some relatively abstract level of phonological analysis. In the signal, however, they are inextricably linked. Phonetically, timing and scaling are known to interact in a host of complex ways. The psychoacoustic literature is rife with perceptual interactions between pitch scaling and timing.[1] The implications of these phenomena for phonological systems, however, are still not broadly appreciated. Interactions between timing and scaling can notably muddy the representational distinction, however, between contrasts in the two domains, creating indeterminacies that AM theory has trouble accommodating. Here we present a case study of one such interaction, the key to understanding which, we argue, involves perceptual integration of F0 information over a relatively broad time window.

1.2. Case study: Tonal timing contrasts in Shilluk

The tone system of Shilluk (Western Nilotic, South Sudan) has received careful and insightful phonetic description by Remijsen and colleagues (2011, 2014). Its system of contrasts is complex, with 8 tone patterns contrasting on monosyllabic transitive verb stems: High, mid and low level tones, a rise, a low fall, a high fall-to-mid, and of particular interest here, two High-to-Low falls that Remijsen &

Ayoker (2014, hereafter R&A) describe as differing solely in the timing of movement onsets and offsets. These two falling contours, the Early High Fall and Late High Fall (EHF and LHF), given in Fig. 1, contrast equally on syllables of all vowel lengths, which, as R&A observe, is typologically unusual, if otherwise attested at all (cf. Hyman 1988).

Figure 1. Top: EHF. 'Somebody has beaten it in this place.' gìn-ání á-**lɛ́ŋ** kì-kêŋ, Bottom: LHF 'Somebody went to the village to beat it.' gìn-ání á-**lɛ́ŋ** pâac
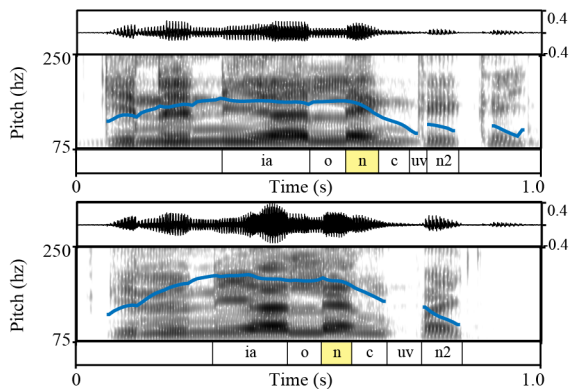

Fig. 2 illustrates the robustness of the timing contrast between the two falls, as manifest in data collected from nine Shilluk speakers by R&A. Fig. 2 further demonstrates that the scaling of the onset of the fall (the F0 turning point taken there as a phonetic proxy for the High tonal target) does not differ between the two lexical tone categories. The contrast between EHF and LHF thus appears to be purely based on tonal timing.

Figure 2. Left: Timing of fall onset for EHF and LHF relative to host syllable onset. Right: Scaling of same, z-transformed F0.
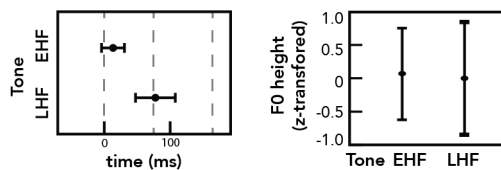
Reliance on a single turning point as a proxy for F0 target scaling, however, can be misleading. It is well known, for example, that both peak height and shape influence perceived scaling of F0 events: plateau-shaped highs (Fig. 3, right) are systematically perceived as higher in pitch than analogous sharp peaks with identical maximum F0 (Fig. 3, left, 't Hart et al. 1990, Knight 2008).

Consider now the average overall shapes of Shilluk EHF and LHF (Fig. 4), from the R&A's 9 speakers. Two things should be immediately clear: First, whether or not EHF and LHF differ in scaling of whatever point we take to instantiate the onsets of their falls, LHF nonetheless appears systematically

higher throughout its duration than EHF. Secondly, even if this were not the case, if instead the high plateaux preceding the fall in both contours turned out to be identical in scaling, the plateau for LHF nonetheless persists longer into the duration of the host syllable than does the one for EHF.

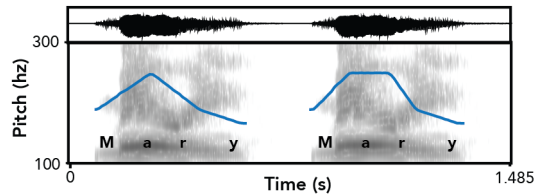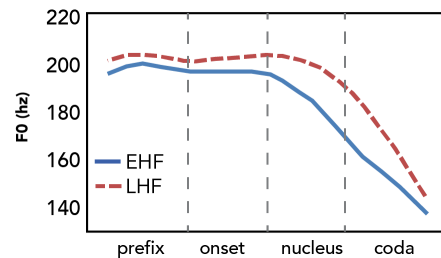**Figure 3**: Example: Sharp Peak vs. Plateau

Figure 4: Time-normalized F0 contours for EHF and LHF averaged across nine speakers

Because Shilluk LHF spends relatively more of its host syllable's duration near its maximum F0, we predict it should sound higher to listeners than does EHF.[2] There could, in other words, be a scaling component to this tonal contrast in Shilluk that is only evident from a view of tonal implementation over which "targets" emerge in perception from the integrated characteristics of the entire shape of the F0 contour within some region of interest (e.g., the host syllable.) It also seems, however, that this inherent difference may be enhanced by increased maximum F0 for LHF. To verify this possibility, we conducted a reanalysis of the R&A's data.[3]

2. SHILLUK: AN ALTERNATIVE ANALYSIS

One way of understanding the difference between sharp peaks and plateaux involves treating the perception of F0 events in speech as involving not the extraction of values at particular points in time, but rather some form of averaging of F0 information over time (Rossi 1971/78, d'Alessandro, Rosset, and Rossi 1998, d'Alessandro & Mertens 1995).[4] A simple way of operationalizing such integration would be to use mean F0 within the target syllable to represent perceived scaling (Barnes, et al. 2014).[5]

Mean F0 during the host syllable for the two contours in Fig. 4 is indeed different, LHF being higher than EHF, as is visually apparent. Beyond this "built-in" scaling distinction created by the longer plateau of LHF, however, we wished to determine whether the apparently higher maximum

F0 of LHF was indeed systematically higher than that of EHF, in a manner consistent with enhancement of the scaling distinction inherent in the two shapes. We thus reanalyzed R&A's data, using maximum F0 during the accented syllable as a measure of scaling, rather than F0 at the moment of fall onset. R&A's Shilluk data comes from 9 talkers—8 male, 1 female, containing 133 instances of EHF, and 96 of LHF, all realized on closed short vowel syllables, for 229 total utterances analyzed.

Mixed-level logistic regression was used to predict tone category (EHF vs. LHF) from two fixed effects, timing of the onset of the fall (as per R&A), and max. F0 during the target syllable. Random intercepts for speakers and a random slope by speaker for scaling was also included. This model finds significant effects for timing ($\beta = 34.85$, $SE = 5.04$, $z = 6.92$, $p < .001$) and for scaling ($\beta = 2.08$, $SE = 0.7$, $z = 2.96$, $p = 0.003$). The distribution for the scaling measure is depicted in Fig. 5.

Figure 5. Mean and 95% confidence interval for EHF and LHF, expressed as z-transformations done for each speaker over measured F0 in semitones.
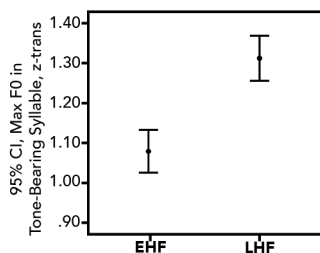
Interestingly, however, the scaling distinction between the two contours is not equally robust for all speakers. Fig. 6 represents time-normalized mean F0 for EHF and LHF for one speaker, representative of a group we call "scalers", in that maximum F0 of the two contours is clearly different.

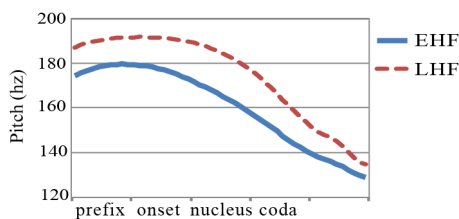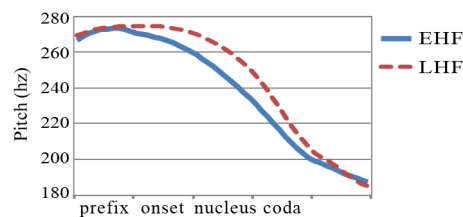**Figure 6**. Time-normalized average F0 for EHF and LHF for one representative "scaler".

Fig. 7 shows the same for another speaker, representing a group we call "timers", in that their contrast appears to be conveyed by timing alone.[6]

Figure 7. Time-normalized average F0 for EHF and LHF for one representative "timer".

Investigating this difference further, we reapplied the same regression model described, this time separating the nine speakers impressionistically but exhaustively into two subgroups, scalers and timers. These tests yielded statistical significance for both timing ($\beta = 30.32$, $SE = 6.27$, $z = 4.84$, $p < .001$) and scaling ($\beta = 2.82$, $SE = 1$, $z = 2.8$, $p = 0.005$) of for the group identified as scalers, but significance only of timing properties for the timers ($\beta = 41.15$, $SE = 8.37$, $z = 4.91$, $p < .001$). We take this to mean that while the contrast between EHF and LHF in Shilluk likely involves a scaling dimension, by virtue of the contour shapes involved, for all speakers, a subset of speakers appears to have begun enhancing the inherent perceived scaling distinction between the two tonemes by altering max. F0 as well. It is difficult to say for a sample this small whether speakers are best divided into two camps in terms of realization strategies, or whether timers and scalers instead exist along a continuum of possible implementations of this contrast.

Figure 8. EHF and LHF, host syllable mean F0 maximum for "scalers" (left), and "timers" (right).
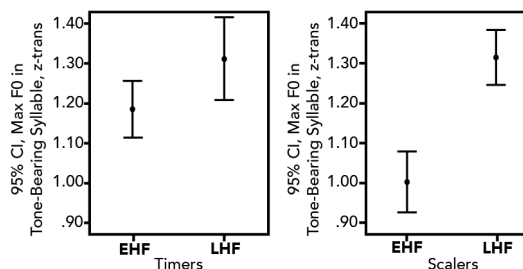

The Shilluk contrast, interestingly, turns out to be another instance of a cross-linguistically common pattern whereby later-timed contours are systematically scaled higher than earlier-timed analogues, or where later timing otherwise trades with higher scaling in some way. In fact, it is strikingly similar in this respect to the lexical pitch accent contrast in Gothenburg Swedish (Segerup & Nolan 2006), which Barnes et al. (2015) account for using a similar integration, mean-F0-based strategy.

The present account of Shilluk is based on the notion of Tonal Center of Gravity, advanced by Barnes et al. (2010, 2012), whereby a broad, seemingly disparate array of properties of global F0 contour shape (including relative timing and scaling of pitch movement onsets and offsets, and shape or curvature of the movements in between) are seen to

integrate perceptually to yield holistically earlier or later tonal timing patterns. Just as, in that model, apparently unconnected acoustic properties of the contour work synergistically to enhance the timing profile of an F0 event, so here the same set of acoustic properties is simultaneously integrating to create later perceived timing on the one hand, and higher perceived scaling on the other. Integration of F0 properties of the signal to create a holistic percept of pitch event scaling may be called TCoG-F (for frequency), and is modeled, as suggested above, by a weighted average of F0 values during a window of interest. Figure 9 shows how a single acoustic property of a contour, here F0 fall curvature, simultaneously shifts the Tonal Center of Gravity of an otherwise symmetrical rise-fall pattern either later and higher (in the case of the convex or domed fall), or earlier and lower (in the case of the concave fall).

Figure 9. Simultaneous shifts in TCog-T(ime) and TCoG-F(requency) as a function of F0 fall curvature.
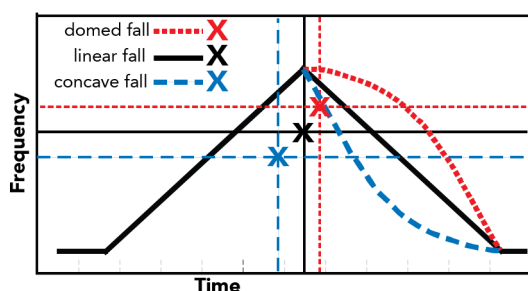
One advantage of this account of the "later = higher" pattern over others (e.g., Gussenhoven 2006) is that it should be able to handle both "later = higher", and the similarly attested "earlier = higher" pattern, seen in cases such as Egyptian Arabic (Cangemi, et al. 2016) or Spanish (Face 2006), where contours associated with raised peaks (e.g., in narrow or contrastive focus contexts) typically involve peak retraction as well. In those cases the retraction of an otherwise slightly delayed peak might have the effect of raising mean F0 for the host syllable, just as raising the peak would.

3. CONCLUSIONS

We have argued that the contrast between EHF and LHF in Shilluk, particularly for some speakers, is realized both in the timing, and in the scaling dimensions. The timing/scaling interaction in Shilluk is furthermore reminiscent of a broader pattern, which we call "later = higher", found in other languages as well. We offer an account of this and other timing/scaling interactions in terms of how the collective raw properties of F0 contour shape integrate perceptually to yield distinction in the location of the Tonal Center of Gravity of a pitch

event at once both in time, and in frequency space. Critically, depending on the shapes in question, a single change to the F0 curve may alter the perception of TCoG in both dimensions. In Shilluk, delaying the onset of a fall for LHF, thereby pushing TCoG later, has the simultaneous effect of pushing it higher as well. This distinction may then be further enhanced by raising the maximum F0 itself.

In an important sense, then, tonal timing and tone scaling are inextricably linked in implementation, a fact that is obscured by their rigid separation in AM representations. Recognition of this link, we argue, can lead to an understanding of tone patterns that might otherwise have seemed mysterious or contradictory. We also suspect that further descriptive work on tone systems will yield far more examples of such interactions than are currently appreciated. If true, this should also make us wonder about the nature of the claims embodied in AM representations of these patterns. To the extent that the Shilluk contrast is both about timing and scaling, it is not immediately clear which dimension should encode it in the phonology. Is this a phonological timing contrast that happens to be enhanced with a phonetic scaling distinction? Or is it a scaling contrast (e.g., High Fall vs. Super-High Fall), that happens to be enhanced by timing. Perhaps both? Or perhaps, in some deeper sense, neither.[7] To the extent that other aspects of the phonology remain silent on the matter (which may or may not ultimately be true in Shilluk), we might even wonder how necessary it is that the phonology encode explicit implementational information about the contrast to begin with. It is by now abundantly clear, contra the assumptions of Chomsky and Halle (1968), that a great many aspects of phonetic implementation are under speaker control, may differ across languages, and yet don't necessarily warrant abstract, symbolic representation in the phonology. But once we take away from the phonology its erstwhile job of providing exhaustive implementational instructions to the phonetics, then absent natural-class behaviors, or clear patterned recycling of articulatory/acoustic properties across all members of some set of segments, we are left with a scenario in which, from a phonological point of view, Toneme A and Toneme B may be just as effective a representation of Shilluk EHF and LHF as any combination of tonal autosegments might. A smart phonetics, in other words, may provide a more insightful model of contrasts, and their attendant enhancement and variation patterns across languages, than can an arbitrarily constrained phonology.

4. REFERENCES

[1] d'Alessandro, C. & Mertens, P. 1995. Pitch Contour Stylization Using A Tonal Perception Model. *Proceedings of the 13th International Congress of Phonetic Sciences*, Stockholm, Volume 4.

[2] d'Alessandro, C., Rosset, S. & Rossi, J.P. 1998. The pitch of short-duration fundamental frequency glissandos. *The Journal of the Acoustical Society of America* 104(4): 2339-48.

[3] Barnes, J., Veilleux, N., Brugos, A. & Shattuck-Hufnagel, S. 2010. The effect of global F0 contour shape on the perception of tonal timing contrasts in American English intonation. *Speech Prosody 2010, Chicago.*

[4] Barnes, J., Veilleux, N., Brugos, A. & Shattuck-Hufnagel, S. 2012. Tonal Center of Gravity: A global approach to tonal implementation in a level-based intonational phonology. *Laboratory Phonology* 3(2): 337-383.

[5] Barnes, J., Veilleux, N., Brugos, A. & Shattuck-Hufnagel, S. 2014. Segmental Influences on the Perception of Pitch Accent Scaling in English. *Speech Prosody 2014, Dublin.*

[6] Barnes, J., Brugos, A., Veilleux, N. & Shattuck-Hufnagel, S. 2015. Perception of Pseudoswedish tonal contrasts by native speakers of American English: Implications for models of intonation perception. *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow.

[7] Bruce, G. 1977. Swedish word accents in a sentence perspective. (*Travaux de l'Institut de Linguistique de Lund 12.)* Lund, Sweden: CWK Gleerup.

[8] Cangemi, F., El Zarka, D., Wehrle, S., Baumann, S., & Grice, M. 2016. Speaker-specific intonational marking of narrow focus in Egyptian Arabic. *Speech Prosody 2016*, Boston.

[9] Chomsky, N. & Halle, M. 1968. *The Sound Pattern of English*. Cambridge, MA: MIT Press.

[10] Cohen, J., Hansel, C., & Sylvester, J. 1953. A new phenomenon in time judgment. *Nature*, *172*: 901.

[11] Face, Timothy. 2006. Narrow Focus Intonation in Castilian Spanish Absolute Interrogatives. *Journal of Language and Linguistics* 5(2): 295-311.

[12] 't Hart, J., Collier, R. & Cohen, A. 1990. *A perceptual study of intonation*. Cambridge: Cambridge University Press.

[13] Helson, Harry. 1930. The tau effect—An example of psychological relativity. *Science* 71(1847): 536-537.

[14] Hyman, Larry M. 1988. Syllable structure constraints on tonal contours. *Linguistique Africaine* 1: 49–60.

[15] Knight, Rachael-Anne. 2008. The Shape of Nuclear Falls and their Effect on the Perception of Pitch and Prominence: Peaks vs. Plateaux. *Language and Speech* 51(3): 223-244.

[16] Ladd, D. Robert. 2008. *Intonational Phonology* (2nd ed.). Cambridge: Cambridge University Press. Original edition, Cambridge: Cambridge University Press, 1996.

[17] Leben, William. 1973. *Suprasegmental Phonology*. New York: Garland.

[18] Niebuhr, O., D'Imperio, M., Gili Fivela, B. & Cangemi, F. 2011. Are there 'shapers' and 'aligners'? Individual differences in signalling pitch accent category. *Proceedings of the 17th International Congress of Phonetic Sciences, August 2011, Hong Kong.* 120-123.

[19] Pierrehumbert, J. & Beckman, M. (1988),

[20] Remijsen, B., Ayoker, O.G. & Mills, T. 2011. Shilluk. *Journal of the International Phonetic Association* 41(1): 111-125.

[21] Remijsen, B. & Ayoker, O.G. 2014. Contrastive tonal alignment in falling contours in Shilluk. *Phonology* 31(3), 435-462.

[22] Rossi, M. 1971. Le seuil de glissando ou seuil de perception des variations tonales pour les sons de la parole. *Phonetica* 23: 1–33.

[23] Rossi, M. 1978. La perception des glissando descendants dans les contours prosodiques. *Phonetica* 35: 11–40.

[24] Segerup, M., & Nolan, F. 2006. Gothenburg Swedish word accents: a case of cue trading? In: G. Bruce & M. Horne (eds), *Nordic Prosody: Proceedings of the IXth Conference*. Frankfurt am Main.

---

[1] E.g., the Glissando Threshold ('t Hart, et al. 1990), the Kappa Effect (Cohen, et al. 1953), the Tau Effect (Helson 1930), etc.

[2] Assuming, perhaps incorrectly, that Shilluk speakers perceive the sharp peak vs. plateau contrast similarly to speakers of the intonation languages for which that effect has been verified.

[3] Remijsen and Ayoker generously made their data publicly available at the following web address: https://datashare.is.ed.ac.uk/handle/10283/3218. We wish to thank Bert Remijsen in particular for substantial assistance and discussion rendered to us throughout this project.

[4] For another type of explanation, see Knight 2008.

[5] In practice, mean F0 is almost certainly too coarse a measure. It is not clear that all F0 samples within so broad a window should receive equal weight, either owing to properties of the segmental string over which they are realized, dynamic properties of the F0 contour itself, or for other reasons. See, e.g., Barnes, et al. 2014 on an averaging procedure called TCoG-F, or Tonal Center of Gravity in the Frequency domain.

[6] The reference to the "shapers" and "aligners" of Niebuhr, et al 2011 should be apparent in this nomenclature.

[7] It might be argued that since the max. F0 distinction is less consistent among speakers, perhaps this contrast is still fundamentally about timing. Recall though that there is likely a perceptible scaling contrast inherent in the basic shapes of the two contours, arising from the timing distinction, which is not variable in this respect, making consistency a less compelling argument