INTRINSIC FUNDAMENTAL FREQUENCY DIFFERENCES IN TWO TONAL AUSTRONESIAN LANGUAGES

Laura Arnold[1, a)], Jiayin Gao[2, 3, b)], James Kirby[3, 1, c)]

[1]University of Edinburgh, [2]LPP (UMR7018, CNRS/Sorbonne Nouvelle), [3]Institute for Phonetics and Speech Processing (IPS), LMU Munich
a) laura.arnold@ed.ac.uk, b) jiayin.gao@cnrs.fr, c) jkirby@phonetik.uni-muenchen.de

ABSTRACT

We report on intrinsic fundamental frequency (IF0) differences in Butlih Salawati and Biga, two closely related Austronesian languages. Both languages have small inventories of six underlying vowels; they both also have simple tone systems, with two specified tones. We show that IF0 differences between high and low vowels in the languages range from 1.5 to 2.8 ST, depending on tonal specification and utterance context. Overall, these results are larger than the mean crosslinguistic difference of 1.65 ST. Based on the previous literature, strong claims that IF0 is under the control of the speaker predict that IF0 differences are attenuated in tone languages, and not enhanced in languages with small vowel inventories. The results presented here do not support these claims. However, we suggest that these results may explain the comparatively frequent tonal developments conditioned by vowel height attested in Austronesian languages.

Keywords: Austronesian, intrinsic fundamental frequency, vowels, tone, production.

1. INTRODUCTION

1.1. Intrinsic fundamental frequency

Intrinsic fundamental frequency (IF0) is a near-universal phenomenon in which high vowels are produced with a higher F0 than low vowels. The mean crosslinguistic difference in F0 between the high vowels /i, u/ and the low vowel /a/ is 1.65 semitones (ST) [1]; however, IF0 differences vary between languages and between speakers [1, 2].

While IF0 has been documented in over 40 genealogically and typologically diverse languages, the mechanisms are the subject of much debate. Traditionally, there are two camps. In the first are advocates of the *biomechanical hypothesis*, who argue that IF0 is an automatic phonetic process caused by an interaction of the articulatory and phonatory systems [1, 3]. The main evidence supporting this hypothesis is the sheer number of languages in which IF0 is found—it is attested in nearly every language in which it has been investigated (e.g. [1, 4, 5])—as well as the presence of IF0 in the babbling of infants [3].

In the second camp are those who argue that IF0 is under speaker control, and that it is more-or-less consciously exploited—for example, to maximize distinctions elsewhere in the phonological system [6, 7]. We refer to this position as the *speaker control hypothesis*. Support for this hypothesis includes the presence of IF0 in esophageal speech, which can have no biomechanical explanation [8]; higher levels of cricothyroid activity during the articulation of high vowels [9], which is associated with active control of F0 [10]; and the existence of at least one language without IF0 differences (Mambila, [4]).

Previous literature suggests that IF0 may be enhanced or attenuated, depending on the phonological profile of a language. One proposed link relates to vowel inventory size. Because vowel height is better perceived when IF0 is enhanced [11], the following trend is predicted: a language with a larger vowel inventory enhances IF0 differences, in order to maximize phonological differentiation within a more clustered vowel space. While Dutch (with 12 vowels, mean IF0 = 21 Hz) and Arabic (with three vowels, mean IF0 = 8 Hz) [5] have been taken as evidence to illustrate such an effect of vowel inventory size, cross-linguistically the correlation between IF0 and number of vowels is not significant [1].

Whether a language has tone, and if so the nature of the tone system, may also have an effect on language-specific IF0 differences. As F0 has a contrastive function in tonal languages, it is reasonable to assume that IF0 differences might be attenuated so as not to jeopardize tonal distinctions. While IF0 differences are attested in most tone languages (e.g. [1, 4]), they tend to be smaller when compared with non-tone languages [2]; and the unusually crowded tone system of

Mambila is offered as the explanation for the lack of IF0 differences in this language [4]. This position predicts that IF0 will be attenuated in tone languages, with increased inhibition in languages with more complex tone systems. However, crosslinguistically, such a tendency is only weakly observed [1, 2]. Furthermore, these observations are largely based on data from a limited number of tone languages spoken in Africa and East and Southeast Asia. More typologically diverse data is required to develop stronger hypotheses on the relationship between tone and IF0.

While there is evidence on both sides of the debate, the emerging consensus is that IF0 has a mixed basis: there is a biomechanical foundation and IF0 is thus (near-)universal; but it is also under speaker control, and thus varies by language, dialect, or speaker [2, 5, 12, 13]. With this context in mind, we provide some background on the languages investigated in this study.

1.2. Target languages and phonological profiles

Biga and the Butlih variety of Salawati are spoken in the Raja Ampat archipelago, northwest New Guinea. Within Austronesian, they belong to the South Halmahera-West New Guinea (SHWNG) branch, which is the sister of Oceanic; and within SHWNG, the languages belong to the Raja Ampat-South Halmahera (RASH) subbranch [14, 15]. Both languages have 14 consonants (Butlih /p b t d k f s m n h r l w j/; Biga /p b t d k g f s m n r l w j/). Butlih contrasts five vowels /i e a o u/ plus a schwa in sesquisyllables; Biga contrasts six vowels /i ɪ e a o u/. Unusually for the Austronesian family—although not so unusually for the RASH languages of Raja Ampat [14, 16]—both languages have lexical tone. The tone systems are modest, and specifications are generally culminative and limited to word-final syllables. In Butlih, there is a contrast between High, Rise, and toneless syllables; in Biga, the contrast is between High, Extra-High, and toneless syllables.

The phonological profiles of the two languages lead us to predict that IF0 differences will be comparatively small in both Butlih and Biga. On the one hand, the vowel inventories of both languages are relatively small and uncrowded, so there is no need for speakers to disambiguate a crowded vowel space by enhancing IF0 differences. On the other, both languages have tone systems. While the systems are not complex in either language—so there is no need to neutralize IF0 differences to compensate for a crowded tonal space, as argued for Mambila [4]—we still expect to see some attenuation of the differences.

Based on the preceding discussion, we use novel data from these two underdocumented languages to address the following research questions: (1) Are there IF0 differences in Butlih Salawati and Biga? If so, what is the magnitude of the differences? (2) Are IF0 differences in Butlih Salawati and Biga in line with predictions based on the phonological profiles of the languages?

2. METHOD

2.1. Participants

Six speakers (3 male, 3 female) were recorded for each language during a fieldtrip to Raja Ampat in 2019. The mean age of the Butlih speakers was 47 (range: 36-65), and of the Biga speakers was 51 (range: 44-66). Informed consent was gained from participants, and all were remunerated for their time.

2.2. Speech material

In both languages, the target words were monosyllables of the shape CV, VC, and CVC. As far as the lexical inventories of the languages allowed, target words were selected that combined each vowel with each tonal specification. Each target word was recorded multiple times in two contexts: utterance medially and utterance finally. The carrier sentences used were as follows: 'I say X' (utterance-final; Butlih /ine jəwe (bo) X/, Biga /ɪn bɪtɪn X/) and 'I don't say X' (utterance-medial; Butlih /ine jəwe (bo) X po/, Biga /ɪn bɪtɪn X apo/). Prompts were provided orally by the first author, who asked for translations of equivalent sentences from Papuan Malay, the local lingua franca.

2.3. Data processing and analysis

This study reports a subset of the data, in which each syllable contains a monophthong high or low vowel /i (ɪ) u a/ and carries a specified tone (High or Rise in Butlih; Extra-High or High in Biga). Toneless words were excluded because the number of tokens was too small. In total, 525 tokens (Butlih) and 703 tokens (Biga) were retained for analysis. Recordings were stored as an EMU-SDMS database [17]. Annotations were performed either manually or using WebMAUS [18] with hand corrections. F0 values were extracted at 5 msec intervals using the `ksvF0` function of the `wrassp` package [19], and then converted to semitones with reference to the mean F0 of each speaker.

Statistical analyses were conducted using the lmerTest [20] and emmeans [21] packages in R [22]. For both languages, F0 values were averaged over one portion of the vowel (see the VOWELREGION predictor below), and linear mixed-effect models were fitted to the averaged F0 data. The full models included Helmert-coded predictors VOWEL HEIGHT (VH) (highV /i (ɪ) u/ vs. lowV /a/), SEX, ONSETVOICING (zero, voiceless, or voiced), POSITION in the utterance (medial vs. final), VOWELREGION in the vowel (Vonset vs. Vmid), TONE (H vs. R or EH vs. H), and the interaction terms VH:ONSETVOICING, VH:SEX, and VH:POSITION:VOWELREGION:TONE. (Vonset and Vmid indicate the first third and the middle third of the vowel, respectively.) All these predictors have been found to affect F0 in previous studies (e.g., [1, 23, 24, 25]). Random intercepts were included for SPEAKER, ONSET, and CODA. By-speaker random slopes were included for VH, TONE (for both), and POSITION (for Butlih only). The final models for each language were selected based on backward elimination of non-significant effects and will be provided in the relevant result subsections.

3. RESULTS

3.1. Butlih Salawati

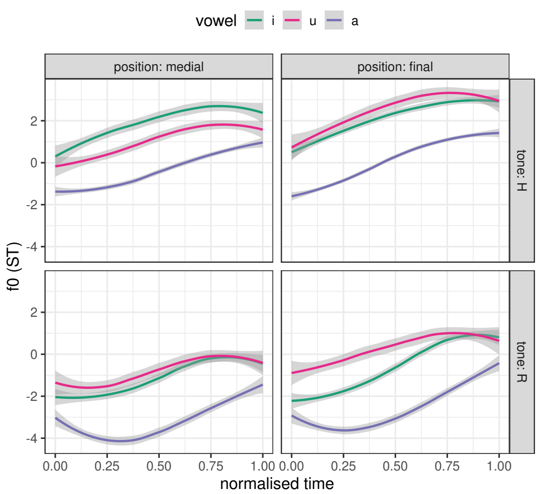

Figure 1: Aggregated F0 curves (loess-smoothed) by vowel, tone, and position from 6 speakers of Butlih Salawati. Shading: 95% CI.

As shown in Fig. 1, the two high vowels carry higher F0 than the low vowel almost throughout the entire trajectory. To quantify the magnitudes of IF0 in different contexts, pairwise comparisons were conducted on the finally selected model with the following predictors: VH, TONE, VOWELREGION, POSITION, ONSETVOICING, and the interaction terms VH:TONE:VOWELREGION and VH:ONSETVOICING.

In both utterance-medial and -final positions, IF0 is overall larger in H than R tones. When the syllable carries R tone, IF0 is larger at Vmid ($est. = 2.71$, SE $= 0.40$, $t = 6.81$, $p < .0001$) than at Vonset ($est. = 1.96$, SE $= 0.40$, $t = 4.93$, $p = .0002$). When the syllables carries H tone, IF0 is realized with a similar extent at Vonset ($est. = 2.70$, SE $= 0.36$, $t = 7.52$, $p < .0001$) and Vmid ($est. = 2.60$, SE $= 0.36$, $t = 7.24$, $p < .0001$).

3.2. Biga

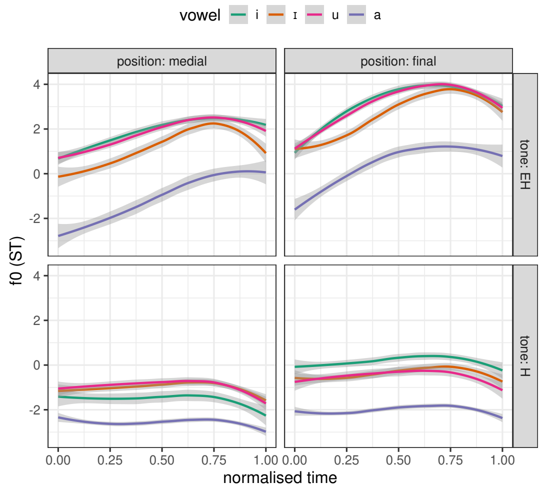

Figure 2: Aggregated F0 curves (loess-smoothed) by vowel, tone, and position from 6 speakers of Biga. Shading: 95% CI.

As shown in Fig. 2, here again, the three high vowels carry higher F0 than the low vowel almost throughout the entire trajectory. The effect of tonal context is more visible here than in Butlih. Pairwise comparisons were conducted on the finally selected model with the following predictors: VH, TONE, VOWELREGION, POSITION, ONSETVOICING, and the interaction terms VH:TONE:POSITION:VOWELREGION and VH:ONSETVOICING.

IF0 is larger in EH than H tones. When the syllable carries EH tone, IF0 is larger in utterance-medial (Vonset: $est. = 2.83$, SE $= 0.49$, $t = 5.79$, $p < .0001$; Vmid: $est. = 2.44$, SE $= 0.49$, $t = 4.99$, $p < .0001$) than -final position (Vonset: $est. = 2.40$, SE $= 0.42$, $t = 5.71$, $p < .0001$; Vmid: $est. = 2.39$, SE $= 0.42$, $t = 5.67$, $p < .0001$). When the syllable carries H tone, IF0 is smaller in utterance-medial (Vonset: $est. = 1.56$, SE $= 0.27$, $t = 5.75$, $p < .0001$; Vmid: $est. = 1.70$, SE $= 0.27$, $t = 6.30$, $p < .0001$) than -final position (Vonset: $est. =$

1.86, SE = 0.32, $t = 5.87$, $p < .0001$; Vmid: *est.* = 1.94, SE = 0.32, $t = 6.14$, $p < .0001$).

4. DISCUSSION

1. Are there IF0 differences in Butlih Salawati and Biga? If so, what is the magnitude of the differences?

Yes, both languages exhibit IF0 differences to a similar extent, ranging from 1.5 to 2.8 ST depending on tonal specification and utterance context. Overall, this is larger than the mean crosslinguistic difference of 1.65 ST reported in careful speech [1], and the differences obtained for most languages from a study on large corpora of read speech [2]. While the nature of the speech material and the methods for measurements differ across studies, which would always remain an uncontrolled source of variability, we can conclude that IF0 in Butlih and Biga is probably larger than the crosslinguistic average.

IF0 also varies according to tone or the portion of the tonal contour. It is overall larger in syllables with a higher than a lower tone. For the Rise tone in Butlih, it is higher in the mid portion than the onset portion of the vowel, that is, it is higher in the portion of the vowel where overall F0 is higher. Previous studies report the disappearance of IF0 in the low F0 range [26], in low tones or the low portion of the tonal contour [1]. Although we do not observe an absence of IF0 even in the lowest portion of the tone, the increase of IF0 in higher-pitched tones or the higher portion is in line with previous findings. This relation between IF0 and the F0 range is widely attested, and thus has been taken as evidence for the biomechanical hypothesis, arguably because the activity of the strap muscles involved in larynx lowering counteracts the tongue-pull effect [1, 27] which is primarily responsible for the IF0 differences [28].

2. Are IF0 differences in Butlih Salawati and Biga in line with predictions based on the phonological profiles of the languages?

The results do not support a strong position linking speaker enhancement of IF0 differences to the phonological profiles of the languages. Following such a position, we would expect to see smaller IF0 differences in both Butlih and Biga, resulting from the combination of (1) a lack of enhancement due to relatively small vowel inventories and thus a lower need to maximize contrasts; and (2) attenuation due to the contrastive use of F0 in the tone systems. However, this is not what the results show: both Butlih and Biga have IF0 differences that are, overall, larger than average. These results indicate that, to the extent that IF0 is under speaker control, the effect of phonological profile may be limited, or perhaps overridden by some other factor—for example, the socio-cultural explanations offered by [29, 30]. The findings presented here emphasize the need to study IF0 in a cross-linguistically diverse sample of languages, in order to develop our theoretical understanding of this proposed universal.

This last point is highlighted by the diachronic relationships between tone and vowel height in several languages related to Butlih and Biga. Tonogenesis or tone changes conditioned by vowel height are typologically rare compared to the phonologization of co-intrinsic effects of consonants [31, 32]. Despite this rarity, a diachronic relationship between tone and vowel height is attested in several other eastern Austronesian languages: Ambel [33], Maˈya [34] (both also RASH languages of Raja Ampat), Yerisiam [15] (a SHWNG language of northwest New Guinea), and Cèmuhî [35] (an Oceanic language of New Caledonia). One obvious phonetic source for the development of tone from vowel height is IF0 differences [31, 32]. The large IF0 differences of Butlih Salawati and Biga reported in this study may represent more general trends in the family. If so, this would indicate a longstanding and complex relationship between vowel height and F0 in Austronesian languages—the nature of which requires further data and research.

5. ACKNOWLEDGMENTS

This study was funded by a British Academy Postdoctoral Fellowship (PF19\100004) and a Moray Endowment Grant (University of Edinburgh) to L. Arnold, and an ERC Starting Grant (758605) to J. Kirby. Thanks to Dorlince Drimlol, Ahad Filis, Hasim Filis, Siti Hariani Klasin, Yunus Klasin, Penina Logof, Hodija Malakabu, Argenus Malbanat, Etresia Malbanat, Semuel Mandurun, Melianuis Samigita, and Aifa Warfandu for speech data, and Brisa Speier-Brito for help with data annotation.

6. REFERENCES

[1] D. H. Whalen and A. G. Levitt, "The universality of intrinsic F_0 of vowels," *J. Phon*, vol. 23, pp. 349–366, 1995.

[2] C. Ting, M. Clayards, M. Sonderegger, and M. McAuliffe, "The cross-linguistic distribution of vowel and consonant intrinsic F0 effects," PsyArXiv. March 24, 2023.

[3] D. H. Whalen, A. G. Levitt, P.-L. Hsiao, and I. Smorodinsky, "Intrinsic F0 of vowels in the babbling of 6-, 9-, and 12-month-old French- and English-learning infants," *JASA*, vol. 97, no. 4, pp. 2533–2359, 1995.

[4] B. Connell, "Tone languages and the universality of intrinsic F0: evidence from Africa," *J. Phon*, vol. 30, pp. 101–29, 2002.

[5] S. V. Hoof and J. Verhoeven, "Intrinsic vowel F0, the size of vowel inventories and second language acquisition," *J. Phon*, vol. 39, no. 2, pp. 168–177, 2011.

[6] R. L. Diehl and K. R. Kluender, "On the objects of speech perception," *Ecological Psychology*, vol. 1, no. 2, pp. 121–144, 1989.

[7] J. Kingston, "The phonetics and phonology of perceptually motivated articulatory covariation," *Lang Speech*, vol. 35, no. 1-2, pp. 99–113, 1992.

[8] J. Gandour and B. Weinberg, "On the relationship between vowel height and fundamental frequency: Evidence from esophageal speech," *Phonetica*, vol. 37, no. 5-6, pp. 344–354, 1981.

[9] N. Dyhr, "The activity of the cricothyroid muscle and the intrinsic fundamental frequency in Danish vowels," *Phonetica*, vol. 47, no. 3-4, pp. 141–154, 1990.

[10] A. Löfqvist, T. Baer, N. S. McGarr, and R. S. Story, "The cricothyroid muscle in voicing control," *JASA*, vol. 85, no. 3, 1989.

[11] H. Traunmüller, "Perceptual dimension of openness in vowels," *JASA*, vol. 69, no. 5, 1981.

[12] E. Fischer-Jørgensen, "Intrinsic F0 in tense and lax vowels wth special reference to German," *Phonetica*, vol. 47, no. 3-4, pp. 99–140, 1990.

[13] P. Hoole and K. Honda, "Automaticity vs. feature-enhancement in the control of segmental F0," in *Where do phonological features come from?*, N. G. Clements and R. Ridouane, Eds. John Benjamins, 2011, pp. 131–172.

[14] B. Remijsen, *Word-prosodic systems of Raja Ampat languages*. Utrecht: LOT, 2001.

[15] D. Kamholz, "Austronesians in Papua: Diversification and change in South Halmahera–West New Guinea," Ph.D. dissertation, University of California, Berkeley, 2014.

[16] L. Arnold, "Lexical tone in Metnyo Ambel," *Ocean. Linguistics*, vol. 57, no. 1, pp. 199–220, 2018.

[17] R. Winkelmann, J. Harrington, and K. Jänsch, "EMU-SDMS: Advanced speech database management and analysis in R," *Computer Speech and Language*, pp. 392–410, 2017.

[18] T. Kisler, U. Reichel, and F. Schiel, "Multilingual processing of speech via web services," *Computer Speech & Language*, vol. 45, pp. 326–347, 2017.

[19] L. Bombien, R. Winkelmann, and M. Scheffers, "wrassp: an R wrapper to the ASSP Library," *R package version 0.1*, vol. 8, 2018.

[20] 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, 2017.

[21] R. Lenth, *emmeans: Estimated marginal means, aka Least-squares means*, 2022.

[22] R Core Team, *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2022.

[23] H. M. Hanson, "Effects of obstruent consonants on fundamental frequency at vowel onset in English," *JASA*, vol. 125, no. 1, pp. 425–441, 2009.

[24] J. Kirby and D. R. Ladd, "Effects of obstruent voicing on vowel F0: evidence from "true voicing" languages," *JASA*, vol. 140, no. 4, pp. 2400–2411, 2016.

[25] Q. Luo, "Consonantal effects on F0 in tonal languages," Ph.D. dissertation, Michigan State University, 2018.

[26] D. R. Ladd and K. E. A. Silverman, "Vowel intrinsic pitch in connected speech," *Phonetica*, vol. 41, no. 1, pp. 31–40, 1984.

[27] M. Pettorino, "Intrinsic pitch of vowels: An Experimental study on Italian," *Procs. 11th ICPhS*, pp. 138–141, 1987.

[28] J. J. Ohala and B. W. Eukel, "Explaining the intrinsic pitch of vowels," in *In honor of Ilse Lehiste*, R. Channon and L. Shockey, Eds. Dordrecht: Foris, 1987, pp. 207–215.

[29] E. Jacewicz and R. A. Fox, "Intrinsic fundamental frequency of vowels is moderated by regional dialect," *JASA*, vol. 138, no. 4, pp. EL405–EL410, 2015.

[30] Y. F. Holt, E. Jacewicz, and R. A. Fox, "Sociocultural influences on fundamental frequency in vowels: The case of African American English," in *Procs. 19th ICPhS*, 2019, pp. 874–878.

[31] J. Kingston, "Tonogenesis," in *Blackwell Companion to Phonology*, M. van Oostendorp, C. J. Ewen, E. Hume, and K. Rice, Eds. Oxford: Blackwell publishing, 2011, pp. 2304–2333.

[32] A. Michaud and B. Sands, "Tonogenesis," in *Oxford research encyclopedia of linguistics*, M. Aronoff, Ed. Oxford University Press, 2020.

[33] L. Arnold, "Highs and lows: A previously unattested tone split from vowel height in Metnyo Ambel," *Trans Philol Soc*, vol. 118, pp. 141–158, 2020.

[34] ——, "A preliminary archaeology of tone in Raja Ampat," in *Contact and substrate in the languages of Wallacea, part 2*, ser. NUSA: Linguistic Studies of Indonesian and Other Languages in Indonesia 64, A. Schapper, Ed., 2018, pp. 7–39.

[35] J.-C. Rivierre, "Tonogenesis and evolution of tonal systems in New Caledonia, the example of Cèmuhî," in *Procs. of the symposium Cross-linguistic studies of tonal phenomena*, S. Kaji, Ed. Tokyo: Institute for the Study of Languages/Cultures of Asia/Africa, 2001, pp. 23–42.