Intrinsic F0 and Sound Change: Evidence from Australian Languages

Sarah Babinski
Yale University

1 Introduction

Vowels are known to vary systematically in their mean f0 as a function of vowel height as well as voicing of the previous consonant, a property known as *intrinsic f0*. Vowel height has been known to have a positive correlation with f0 since at least Black (1949), who found that high vowels in English had a higher intrinsic f0 than low vowels on average. Much work has since supported this claim, finding similar correlations in many languages of the world (Whalen & Levitt (1995), among others). Differences in intrinsic f0 also vary systematically based on the voicing of a preceding stop, with voiceless stops correlating with higher intrinsic f0 and voiced stops with lower (Ohde, 1984).

When f0 differences are related to the voicing of the previous consonant, these systematic differences have been observed as a common phonetic precursor to tonogenesis and tone splits (Erickson, 1975; Maddieson, 1984; Kingston, 2011). Tonogenesis from this source occurs as a reanalysis of the perturbations in intrinsic f0 seen on a vowel immediately following a voiced or voiceless consonant. Following a voiceless consonant, the intrinsic f0 of a vowel will tend to be higher than following a voiced consonant; this then results in high tone following voiceless stops, and low tone following voiced ones. Kingston (2011:2317) notes that voicing of the preceding consonant is not commonly the trigger for tonogenesis *de novo,* although there are some examples of this in Yabem (Austronesian) and Korean. Instead, stop voicing is often involved in tone splits after a language has already developed a tone system. Such splits have occurred in Hmong, Tai, and Sino-Tibetan language families, among others. Kingston suggests that heightened awareness of f0 perturbations can lead to a phonologization of this intrinsic f0 variability.

Intrinsic f0 related to vowel height, on the other hand, is not known as a common phonetic precursor to any type of sound change. High vowels tend to have a higher intrinsic f0 than low vowels, but no languages have been known to make use of this difference for the purposes of language change.[1] This mismatch of intrinsic f0 as a strong phonetic precursor to change in the case of preceding stop voicing, and a weak to nonexistant precursor in the case of vowel height, has been the subject of research in recent years (Sonderegger et al., 2017).

To investigate the relative robustness of these two types of intrinsic f0, Sonderegger et al. (2017) conducted a cross-linguistic study of intrinsic f0 in 14 languages using automated data processing methods. They found that languages vary widely with respect to their intrinsic f0 effects, but that there is some systematicity in the variation due to the voicing of a preceding stop (henceforth *consonant f0* or *CF0*). Tone languages like Mandarin and Thai showed less variation in consonant f0 than non-tone languages, supporting the claim that tone and consonant f0 are related. Intrinsic f0 variation with respect to vowel height (henceforth *vowel f0* or *VF0*) was more variable and did not show a clear difference between tone and non-tone languages, supporting the observation that this type of intrinsic f0 does not seem to participate in tonogenesis or other sound changes.

In this paper, I conduct a study similar to Sonderegger et al. (2017) looking at both CF0 and VF0 in sixteen Australian languages. The data, natural language field recordings, were sourced from archival deposits, as

* Many thanks to research assistants Shayley Martin, Jeremiah Jewell, and Ronnie Rodriguez, and HistLing@Yale members Dr. Claire Bowern, Dr. Rikker Dockum, and J. Hunter Craft, for their contributions in processing the forced alignment data for this project. Many thanks to PhonoloGroup@Yale for feedback. Data is from Paradisec, ELAR, and AIATSIS archives. This is an ongoing project related to the author's PhD dissertation.

[1] Though see Kingston (2011:2318) for brief mention of a tone split in Hu (Palaungkic) in which high vowels retained high tone when non-high vowels' tone became low.

summarized in §2. The statistical methods used are intended to be comparable to those in Sonderegger et al. (2017) and are discussed in §3. Results are presented in §4 and show cross-linguistic variation in intrinsic f0 effects even among languages that are closely related. The implications of these results for research on language variation and change are discussed in §5.

2 Data

The sixteen languages surveyed in this paper are listed in (1-7) grouped by genetic affiliation. There are seven language families represented, including five languages in the largest Australian language family, Pama Nyungan.

(1) **Daly**
- a. Malak Malak
- b. Murrinh Patha
- c. Ngan'gi

(2) **Gunwinyguan**
- a. Dalabon
- b. Gunwinggu
- c. Kunbarlang

(3) **Maningrida**
- a. Burarra
- b. Gun-nartpa

(4) **Pama Nyungan**
- a. Wanyjirra
- b. Warlpiri
- c. Warnman
- d. Yannhangu
- e. Yidiny

(5) **Jarrakan:** Gija

(6) **Nyulnyulan:** Bardi

(7) **Tangkic:** Kayardild

The audio used for this study consist of natural speech field recordings from three online language archives: AIATSIS (2020), ELAR (SOAS, 2020), and PARADISEC (2016). These data were transcribed at the utterance level by the researchers who made the recordings, and were force aligned at the segment level by the author using the Montreal Forced Aligner (McAuliffe et al., 2017). Resulting segmentations were manually checked to ensure accuracy, which can ensure that segmentations are just as accurate as manual alignments, especially when working with endangered language data (Babinski et al., 2019; DiCanio et al., 2013; Johnson et al., 2018).

Number of speakers for each language range from 1 to 10 (average no. of speakers: 3). The number of word tokens for each language range from about 1,000 to about 25,000 (average no. of word tokens: 6,885). All of these languages lack a voicing distinction in their obstruents, but most have been found to show robust allophonic variation, with word-initial stops being voiceless and intervocalic stops being voiced (Fletcher & Butcher, 2014; Kakadelis, 2018). This fact has implications for consonant coding in the consonant-related f0 portion of this project, as discussed further in the following section.

3 Methods

Acoustic measurements were extracted from TextGrids using Praat scripts from Christian DiCanio and Daniel McCloy, with some slight modifications by the author. Post-extraction scripts were written in R (Team, 2015) using the packages 'lme4' (Bates et al., 2015) for regression models, 'ggplot2' (Wickham, 2016) for graphs, 'dotwhisker' (Solt & Hu, 2015) for dot-and-whisker plots, and 'hqmisc' (Quene, 2014) for conversion from values in Hertz to semitones. Any vowel shorter than 50ms in duration was excluded from this analysis.

Measurement of f0 was done in two different ways in order to capture the differing effects of vowel-related f0 and consonant-related f0. For vowel f0, maximum and minimum f0 measurements were taken in Hertz over the whole vowel. In Praat, pitch minimum was set at 80 Hz and maximum at 300 Hz. Within these parameters, the script found the maximum and minimum f0 readings across each vowel. These maxima and minima were averaged to calculate the mean f0 over the vowel, and this value was converted to semitones with base 50.

Consonant f0 measurements were taken only over the first 50 ms of the vowel in order to determine the previous consonant's effect on the vowel. The Praat script identified the maximum and minimum f0 readings across this 50 ms window using the same Praat settings as above. The maxima and minima were averaged to get the mean f0 in the first 50 ms of the vowel after a stop consonant. These values were converted to semitones with base 50, which allows for greater comparability across speakers and languages.

Because consonant-related f0 has been studied primarily with stop consonants, these were also targeted here. Differences in intrinsic f0 in this context are related to the voicing of the previous stop, but most Australian languages do not have a voicing distinction here (Fletcher & Butcher, 2014). As a proxy for phonological voicing distinctions, stops were categorized into voicing categories based on their position within the word. Australian languages have often been found to have consistent patterns of allophonic voicing in which initial stops are voiceless and medial (intervocalic) stops are voiced (Kakadelis, 2018). Therefore, intial stops were coded as 'voiceless' and medial stops as 'voiced' for the purposes of this analysis.

The effects of vowel height and consonant voicing were determined using regression models. These models were structured with a random intercepts for word and speaker, and a fixed effect for the test variable, either vowel height ('high' or 'low') or allophonic stop voicing ('initial/voiceless' or 'medial/voiced'). In some of the languages in this study, such as Murrinh Patha, only one speaker's data was available, so the random variable for speaker was not included in the model.

Results are presented in the following section as dot-and-whisker plots showing the estimate and standard error values of the test variable in the output of each language's regression model. This presentation shows the estimated strength and variability of the effect of each variable, taking into account the random intercepts included in these models.

4 Results

Sonderegger et al. (2017) identifies two crucial characteristics that are needed for a phonetic property to be a strong phonetic precursor to sound change. First, the property must be *robust;* that is, the effects of the phonetic phenomenon must be consistent across speakers of the same language, and the effect must be seen to apply across languages as well. Second, the property must be *variable,* meaning that there are individual speaker differences within a language, as well as cross-linguistic differences in the strength of the property's effect or realization. Robustness and variability are in tension with one another, as one requires consistency and the other requires difference. However, Sonderegger et al. (2017) find both to be true of the intrinsic f0 effects in the languages they analyze. For example, their CF0 results show robustness in the direction of the effects across languages, and across speakers of tone languages, but variability in the size of the effect across languages, and high variability across speakers of non-tone languages.

This section presents the results for CF0 and VF0 effects in the sixteen Australian languages considered for this project. They are presented as overall results across languages in this study, and one language (Ngan'gi) is investigated for its interspeaker variability. Robustness and variability of intrinsic f0 effects are also observed in these languages, with some caveats discussed as they arise.

4.1 *CF0 results across languages* Figure 1 presents the distribution of average f0 measurements across the first 50 ms window of vowels following stops categorized as 'voiced' and 'voiceless' based on the positional criteria already mentioned. In some languages, higher f0 averages after voiceless consonants can be observed, although it should be noted that in most of these cases there is wide variation in the actual f0 values.

The results of regression modeling for CF0 across each language are shown in Figure 2. The figure shows the difference (in semitones) between the f0 measurements at the beginnings of vowels following initial versus medial stops, as estimated by the model structured as discussed in the previous section. The reference state is voiced (stops in the medial position), which means that a positive value in Figure 2 indicates higher intrinsic f0 on vowels following a voiceless (initial) stop.

The results in Figure 2 show variation in the effect size, from a non significant (zero) effect, to rather large effects around 1.5 semitones. In Sonderegger et al. (2017), smaller CF0 effects were observed generally in tonal languages, while larger effects were seen in non-tonal languages. None of the languages in this data set have tone, but five of them have effects that cross the zero line, indicating a non significant effect. While more detailed research into the typology of these effects is certainly needed, these results suggest preliminarily that CF0 effects may be more complex than previous work has revealed.

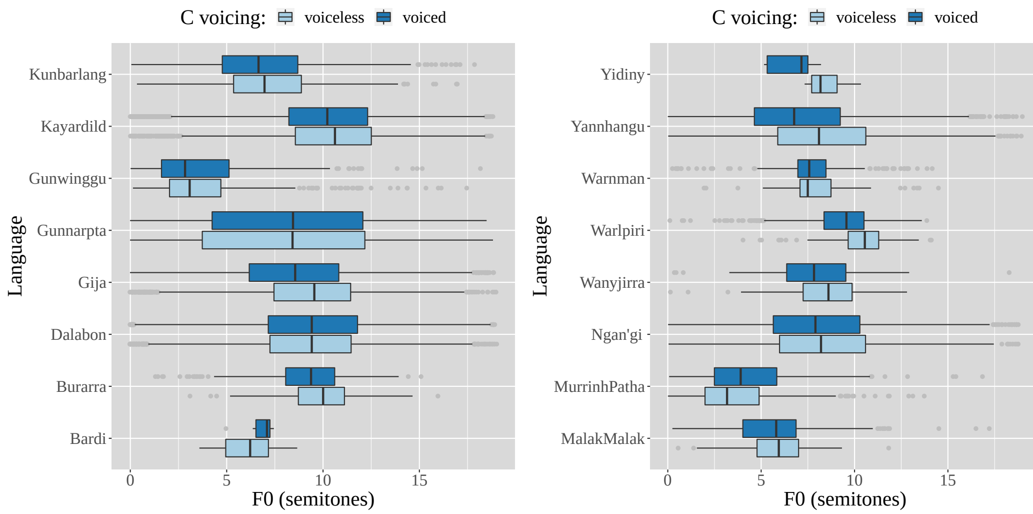

Figure 1: Average f0 measurements at beginnings of vowels following voiceless vs. voiced stops.

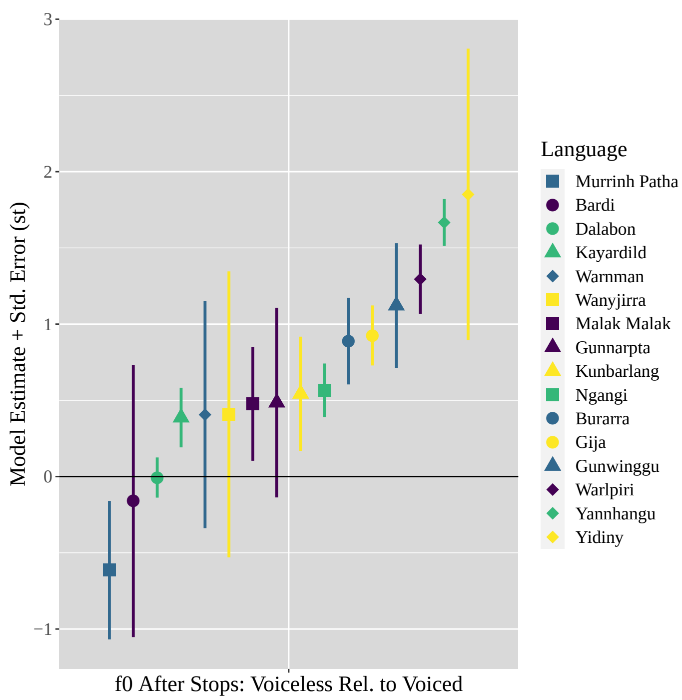

Figure 2: Difference (in semitones) between f0 measures in initial versus medial obstruents - positive numbers have higher f0 in initial position.

Two of these languages, Bardi and Murrinh Patha, have average estimates in Fig. 2 below zero. This suggests that the voiced category of medial stops in fact has a higher f0 than the voiceless initial stops, running counter to the typological generalization. In the case of Murrinh Patha, this effect is significant, although the data require greater scrutiny. It is likely that the heuristic for categorizing allophonic voicing used for this project does not hold for Murrinh Patha. Investigation is currently underway to determine if the voicing of obstruents in these word positions follows a different pattern than the other languages. MurrinhPatha may show voiced obstruent allomorphs in initial position and voiceless allomorphs in medial position, while the other languages show the opposite pattern. Furthermore, the MurrinhPatha data only include audio from one speaker, so strong conclusions cannot be drawn from these results.

The case of Bardi is likely a result of insufficient data and perhaps other intersecting phenomena confounding the results here. I draw this conclusion based on the large standard error value combined with the non-signficant effect. Bardi has been studied extensively for its allophonic voicing patterns in Kakadelis (2018), in which the coded initial-medial pattern of variation was found to be quite robust. Therefore, it is unlikely that the negative estimate value here is a result of improper voicing categorization, as I suspect is the case for Murrinh Patha.

It is important to note that these regression models are at least partly picking up on differences in stress. In all of these languages except Murrinh Patha, stress is consistently word-initial, and is often cued primarily by raised f0. Therefore, it is difficult to tease apart what portion of f0 differences come from stress, and what portion from the allophonic voicing distinction in obstruents, as these two parameters pick out the exact same grouping of vowels. A cursory investigation of both initial and medial nasals and glides in these languages was performed, and these voiced consonants also patterned with the 'voiced' category of stops, but further research is needed.

4.2 *CF0 results across Ngan'gi speakers* This section presents interspeaker variation across speakers of one language in this project, Ngan'gi. The Ngan'gi data had audio from ten different speakers, all of whom are female. Separate regression models were run for each speaker with the same setup as described in §3. The results are shown in the dot-and-whisker plot in Figure 3.

The results in Figure 3 show wide variation across Ngan'gi speakers. Some individuals show little to no CF0 effect, with estimates and standard errors at or below the zero mark. Other speakers show a trend in the same direction as the language overall, with some individuals showing even stronger effects. Each speaker shows internal variation of roughly 1-2 semitones, with the exception of Speaker 4, for which there was much less data than for the other speakers (around 450 tokens compared to 1,000-5,500 tokens).

4.3 *VF0 results across languages* Figure 4 presents the distribution of average VF0 measurements, taken across the entire vowel, split into 'high' and 'low' vowel height categories. Sonderegger et al. (2017) found that the effects of vowel height on intrinsic f0 were more variable than the effects for CF0, a generalization that appears to be supported in these data as well. They also found that some languages had little difference between high and low vowels, indicated by a difference of about zero. The regression results in Figure 5 show similar trends.

The VF0 results for these languages only reflect the expected VF0 effect - higher f0 in high vowels - significantly in three languages: Yidiny, Kayardild, and Warnman. Other languages have positive estimate values, but their standard errors extended below the zero mark, indicating that these effects are not statistically significant. Finally, a handful of languages have estimate values below zero, but only in Kunbarlang is this negative effect significant, predicting an effect that runs counter to the typological generalization.

The results in Figure 5 also show much wider variation than in the CF0 results generally. Malak Malak and Murrinh Patha both have especially large standard errors of around 3.0 semitones. This may indicate that other factors are determining the f0 of high and low vowels that are more robust than intrinsic f0 effects, such as stress or higher level prosody. More work is needed to determine what these confounding factors may be.

4.4 *VF0 results across Ngan'gi speakers* As in the CF0 results, Ngan'gi speakers show considerable individual differences in their VF0 effects as shown in Figure 6. Again, individuals range from having little to no VF0 effect to having effects of about +1.0 semitones. Some individuals show little variation in their VF0 effects, while others have more variation. Again, Speaker 4 has much more variation than the others, likely due to the small number of tokens for this individual. The language overall had a range of about -0.5 to +0.5

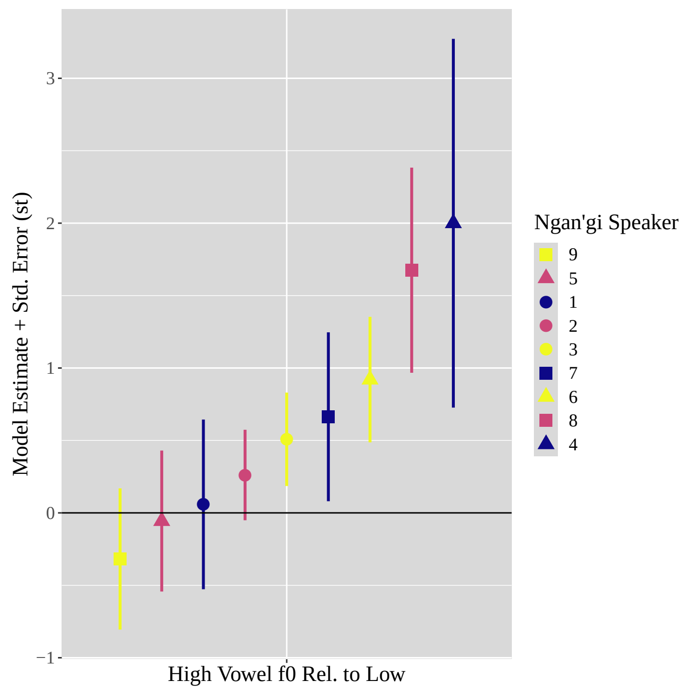

Figure 3: Difference (in semitones) between f0 measures in initial versus medial obstruents for each Ngan'gi speaker - positive numbers have higher f0 in initial position.

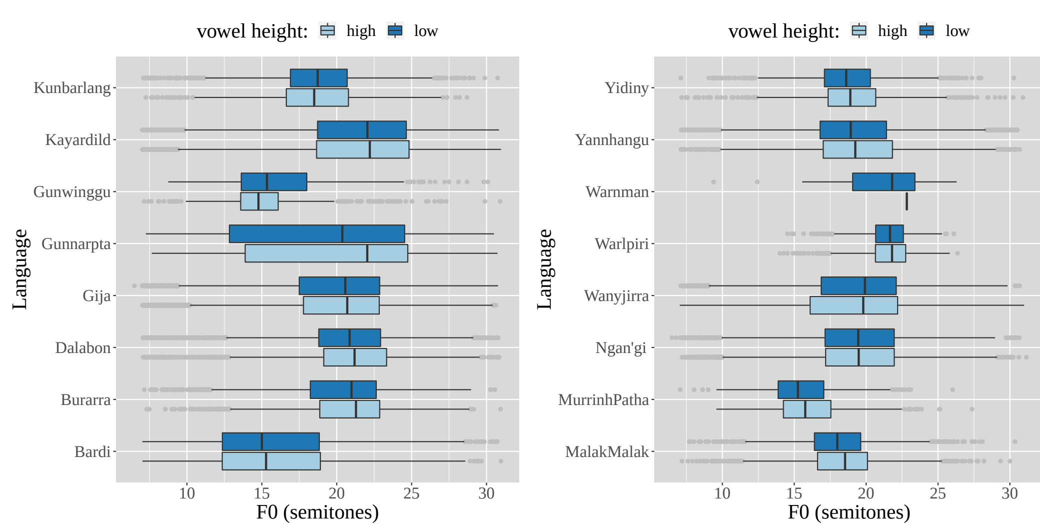

Figure 4: Average f0 measurements in high versus low vowels.

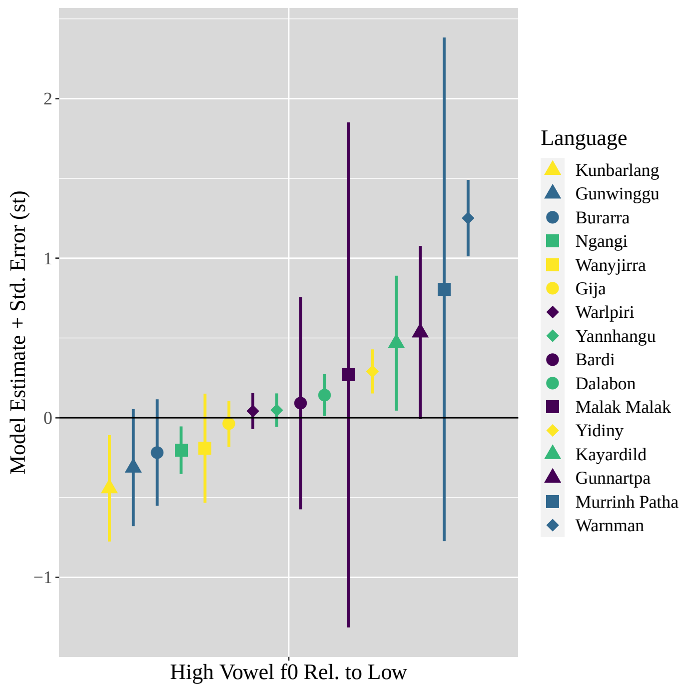

Figure 5: Difference (in semitones) between f0 measures in high versus low vowels - positive values have higher f0 in high vowels.

semitones, an overall non-significant effect, and this is generally reflected in Figure 6.

5 Discussion

The results of this investigation into intrinsic f0 effects in Australian languages show a trend that is comparable to the results in Sonderegger et al. (2017): effects of previous consonant voicing tend to be more robust and less variable than effects of vowel height on intrinsic f0. CF0 results show a more robust trend for positive estimate values, while VF0 results have a mix of positive and negative values, i.e. effects going in opposite directions, and have a higher proportion of languages with no significant effect. As obstruent voicing seems to be a more reliable indicator of a vowel's f0 than vowel height does, it follows that CF0 would be a stronger precursor to change than VF0.

It was found in Sonderegger et al. (2017) that vowel height related intrinsic f0 had greater variability overall than consonant related f0, presumably related to the minimal variation seen in tone languages. This was found for Australian languages as well. The mean standard error for VF0 results was about 0.34 semitones, while the mean error for CF0 results was about 0.16 semitones. As all of the languages investigated here are non-tonal, this suggests a broad typological trend in the differences between these two types of intrinsic f0 effects.

Interspeaker variation was investigated for one of the languages in this study, in order to explore the variation in intrinsic f0 effects seen within one language. Across speakers of Ngan'gi, substantial individual differences are clear for both CF0 and VF0 results. The mean standard error for CF0 was 0.30 semitones, while mean error for VF0 across Ngan'gi speakers was 0.50 st. In the language as a whole, CF0 results were decidedly positive (estimate about 0.5 st), while VF0 results were centered around zero (no effect). However, some individual speakers have internally consistent effects that differ from the language overall. Therefore, intrinsic f0 shows the variability that Sonderegger et al. (2017) cites as necessary for a phonetic precursor to change.

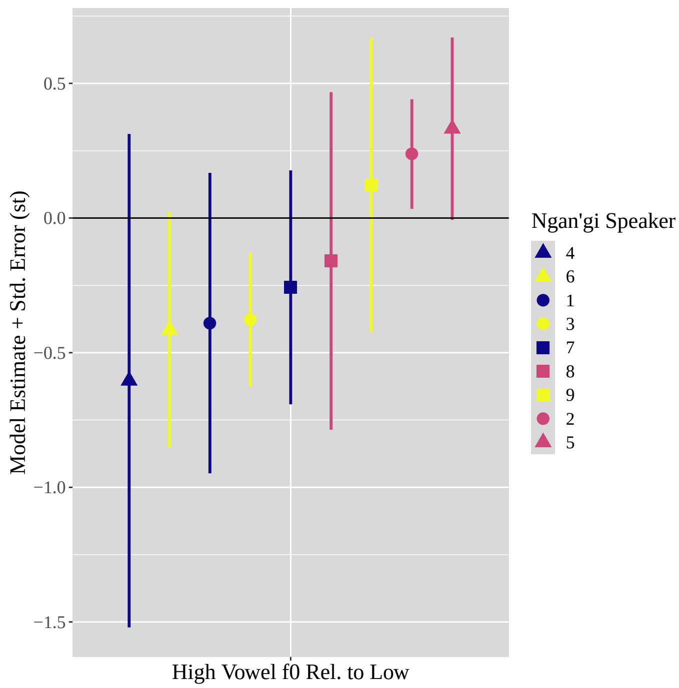

Figure 6: Difference (in semitones) between f0 measures in high versus low vowels in each Ngan'gi speaker - positive values have higher f0 in high vowels.

6 Conclusion

This paper uses automatic methods and statistical analyses to analyze the effects of intrinsic f0 in sixteen languages of Australia, based on the methods used in Sonderegger et al. (2017). According to the authors of the previous study, intrinsic f0 as a phonetic precursor to change should be both *robust* and *variable*, showing a general trend across languages, with variation in the strength of this result and variation across speakers of the same language.

Results as presented roughly match the results seen in Sonderegger et al.'s 14 languages. Intrinsic f0 related to the voicing of the preceding obstruent had a robust effect in which vowels following voiceless obstruents tend to have a higher f0 than those following voiced obstruents. However, individual speakers within one language, Ngan'gi, showed interspeaker variation in this CF0 effect. Some individuals had no effect or even a slight tendency for an opposite effect.

Likewise, intrinsic f0 effects relating to the height of the vowel were broadly robust across languages. However, the robustness of VF0 results was not as strong as that of the CF0 results, and VF0 was more variable than CF0. These cross-linguistic generalizations match those found in Sonderegger et al. (2017). Across speakers of Ngan'gi, there was again wide variation in each individual's effects of VF0. Individuals differed in the effect or lack thereof, as well as the amount of variation they showed.

Based on these preliminary results, Australian languages generally seem to show intrinsic f0 effects in the same way that other languages of the world do. Consonant f0 is a robust and variable phonetic property that is a good candidate to trigger language change, while vowel height f0 is less robust and more variable, perhaps providing some explanation for why this phonetic property is much less likely to instigate change.

References

Babinski, Sarah, Rikker Dockum, J. Hunter Craft, Anelisa Fergus, Dolly Goldenberg & Claire Bowern (2019). A Robin Hood approach to forced alignment: English-trained algorithms and their use on Australian languages. *Proceedings of the Linguistic Society of America* 4:1, p. 3, URL `https://journals.linguisticsociety.org/proceedings/index.php/PLSA/article/view/4468`.

Bates, Douglas, Martin Machler, Ben Bolker & Steve Walker (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67:1, 1–48.

Black, John W. (1949). Natural frequency, duration, and intensity of vowels in reading. *Journal of Speech and Hearing Disorders* 14, 216–221.

DiCanio, Christian, Hosung Nam, Douglas H Whalen, H Timothy Bunnell, Jonathan D Amith & Rey Castillo García (2013). Using automatic alignment to analyze endangered language data: Testing the viability of untrained alignment. *The Journal of the Acoustical Society of America* 134:3, 2235–2246.

ELAR (2020). The Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS). URL `https://aiatsis.gov.au/`.

Erickson, Donna (1975). Phonetic implications for an historical account of tonogenesis in Thai. *Studies in Tai Linguistics in Honor of WJ Gedney*, 110–111.

Fletcher, Janet & Andrew R. Butcher (2014). Sound patterns of Australian languages. Nordlinger, Rachel & H. Koch (eds.), *The Languages and Linguistics of Australia: A Comprehensive Guide*, The World of Linguistics, De Gruyter.

Johnson, Lisa M, Marianna Di Paolo & Adrian Bell (2018). Forced alignment for understudied language varieties: Testing prosodylab-aligner with tongan data .

Kakadelis, Stephanie M (2018). *Phonetic Properties of Oral Stops in Three Languages with No Voicing Distinction*. PhD, CUNY.

Kingston, John (2011). Tonogenesis. van Oostendorp, Marc, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell Companion to Phonology*, Wiley-Blackwell, Oxford, 2304–2333, URL `https://blogs.umass.edu/jkingstn/files/2019/04/kingstonTonogenesis2011BlackwellCompanionPhonology.pdf`.

Maddieson, Ian (1984). The effects on F0 of a voicing distinction in sonorants and their implications for a theory of tonogenesis. *Journal of Phonetics* 12:1, 9–15, URL `http://www.sciencedirect.com/science/article/pii/S0095447019308459`.

McAuliffe, Michael, Michaela Socolof, Sarah Mihuc, Michael Wagner & Morgan Sonderegger (2017). Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. *Interspeech 2017*, ISCA, 498–502, URL `http://www.isca-speech.org/archive/Interspeech_2017/abstracts/1386.html`.

Ohde, Ralph N. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *Journal of the Acoustical Society of America* 75:1, 224–230.

PARADISEC (2016). The Pacific And Regional Archive for Digital Sources in Endangered Cultures (PARADISEC). URL `https://www.paradisec.org.au/`.

Quene, Hugo (2014). Miscellaneous convenience functions and dataset.

SOAS, University of London (2020). The Endangered Languages Archive (ELAR). URL `https://www.soas.ac.uk/elar/`.

Solt, Frederick & Yue Hu (2015). dotwhisker: Dot-and-whisker plots of regression results.

Sonderegger, Morgan, Michael McAuliffe & Hye-Young Bang (2017). Segmental influences on F0: p. 27.

Team, R Core Development (2015). R: A language and environment for statistical computing. URL `www.r-project.org`.

Whalen, Douglas H & Andrea G Levitt (1995). The universality of intrinsic f0 of vowels. *Journal of phonetics* 23:3, 349–366.

Wickham, Hadley (2016). *ggplot2: Elegant graphics for data analysis*. Springer Verlag, New York.