Variation of [v] in Cook Islands Māori

Quartz Colvin[Rutgers University]

Variation of [v] in Cook Islands Māori

1.0 Introduction

In this paper, we will do a statistical analysis of [v] across a sample of islands in Cook Islands Māori (CIM). It's known that in many dialects and other varieties of Māori, this phoneme can be realized as [w] or [v]. This paper aims to take a statistical approach to this generalization.

This paper has four sub-questions to investigate. First, how does w~v duration vary by island and second, how does intensity for these phonemes vary by island. The other two questions are about identifying information about the surface forms of w~v by island. Specifically, we will model f0 and f2 to determine whether certain islands have a voiced phoneme realized. Finally, the f2 model will help us determine which islands have higher rates of [w]s surfacing and which have more [v]s surfacing.

1.1 Background (CIM)

Cook Islands Māori is an Eastern Polynesian language classified as *Endangered*. It is very closely related to Aotearoa Māori, but is definitely different from it.

It's known across both Aotearoa (NZ) Māori and CIM that some speakers regularly pronounce [v] as [w], but it isn't clear to anyone who does this more and what conditions it.

???fill in

2.0 Methods

???fill in

3.0 Data

???fill in

4.0 Analysis

This section readdresses the main questions about the w~v alternation. First, how does duration differ by island and second, how does intensity vary by island? Finally, do any islands have regular occurrence of [w] instead of [v] surfacing. For this final question, we will compare f0 and f2 separately.

Note that in all of my models, I did not control for where w~v occurs in the word, since I am looking at general frequency information and not doing a phonological analysis of specific environments in which [w] or [v] occurs more.

???discuss autocorrelation issue

4.1 Duration by island

???fill in

speaker	island	word	duration
TA	Atiu	ava	0.08
TA	Atiu	ava	0.07
TA	Atiu	ava	0.14
TA	Atiu	ava	0.11
TA	Atiu	ava	0.05
TA	Atiu	ava	0.13

term	estimate	std.error	statistic	p.value
(Intercept)	0.0660169	0.0033898	19.4750727	0.0000000
islandMauke	-0.0003532	0.0038067	-0.0927931	0.9260887
islandPenrhyn	0.0179242	0.0044121	4.0624899	0.0000528
islandRarotonga	-0.0208246	0.0044925	-4.6354104	0.0000041

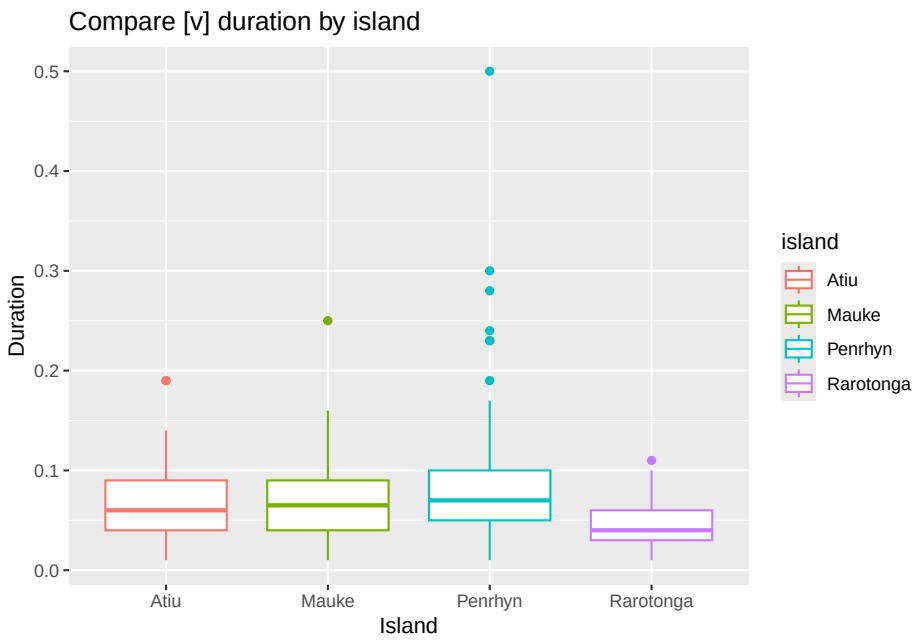

Figure 1

4.2 Intensity by island

???fill in

speaker	island	word	intensity
TA	Atiu	ava	54.08202
TA	Atiu	ava	58.06061
TA	Atiu	ava	48.40481
TA	Atiu	ava	54.18315
TA	Atiu	ava	49.59224
TA	Atiu	ava	53.07335

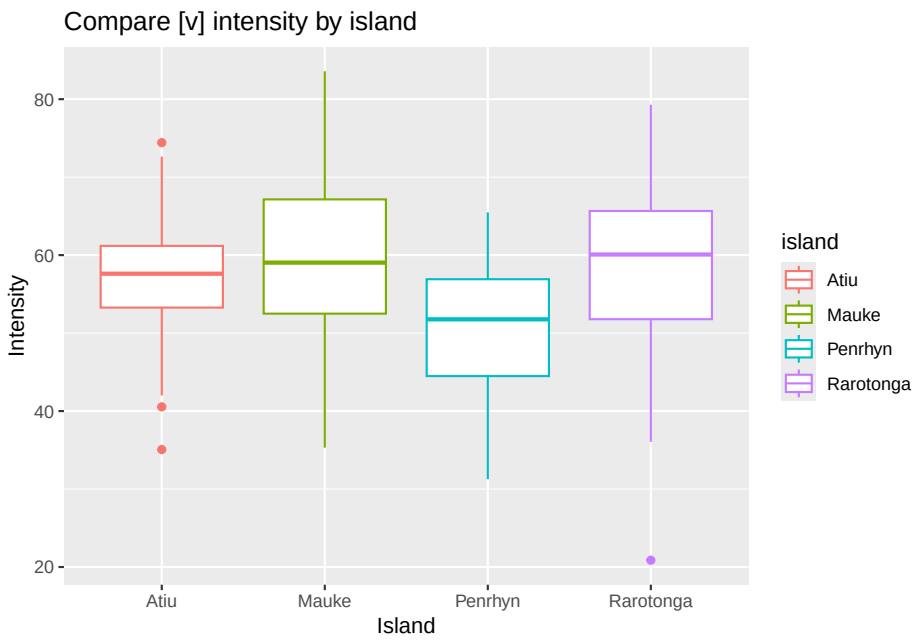

Figure 2

term	estimate	std.error	statistic	p.value
(Intercept)	57.4682172	0.8418656	68.2629370	0.0000000
islandMauke	2.2841738	0.9458247	2.4150075	0.0159356
islandPenrhyn	-6.9840858	1.0957577	-6.3737500	0.0000000
islandRarotonga	0.0787305	1.1157213	0.0705647	0.9437601

4.3 Voicing by island

???fill in

speaker	island	word	f0	duration
TA	Atiu	ava	128.5916	0.08
TA	Atiu	ava	127.0908	0.08
TA	Atiu	ava	127.4540	0.08
TA	Atiu	ava	160.6464	0.07
TA	Atiu	ava	157.4397	0.07
TA	Atiu	ava	158.2820	0.07

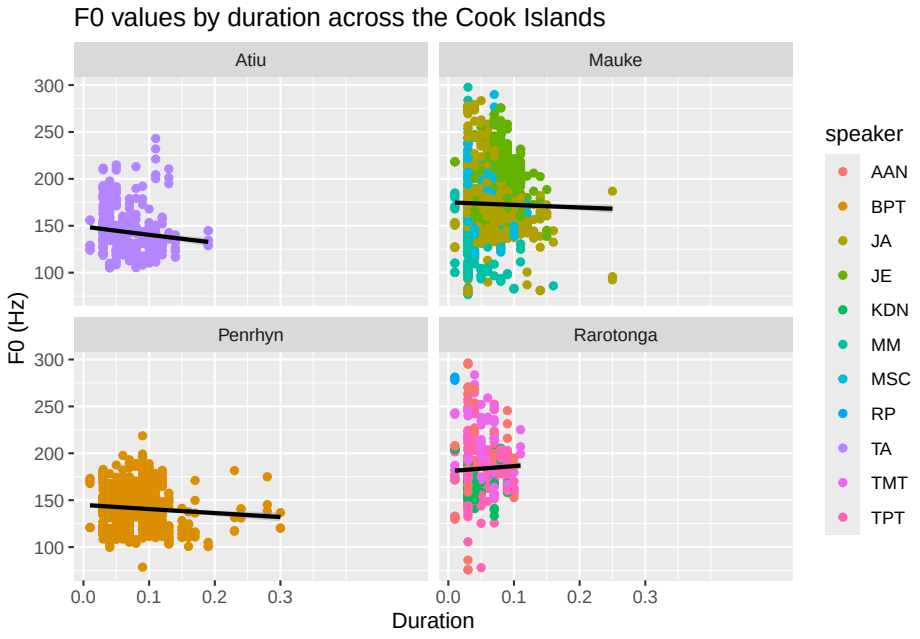

Figure 3

term	estimate	std.error	statistic	p.value
(Intercept)	143.226548	1.779445	80.4894349	0.0000000
islandMauke	29.922893	2.005496	14.9204478	0.0000000
islandPenrhyn	-1.840639	2.304081	-0.7988602	0.4244468
islandRarotonga	40.179828	2.400413	16.7387166	0.0000000

4.4 w~v distribution by island

???fill in

speaker	island	word	f2
TA	Atiu	ava	964.3925
TA	Atiu	ava	950.2333
TA	Atiu	ava	2575.9621
TA	Atiu	ava	961.5736
TA	Atiu	ava	984.4541
TA	Atiu	ava	896.0441

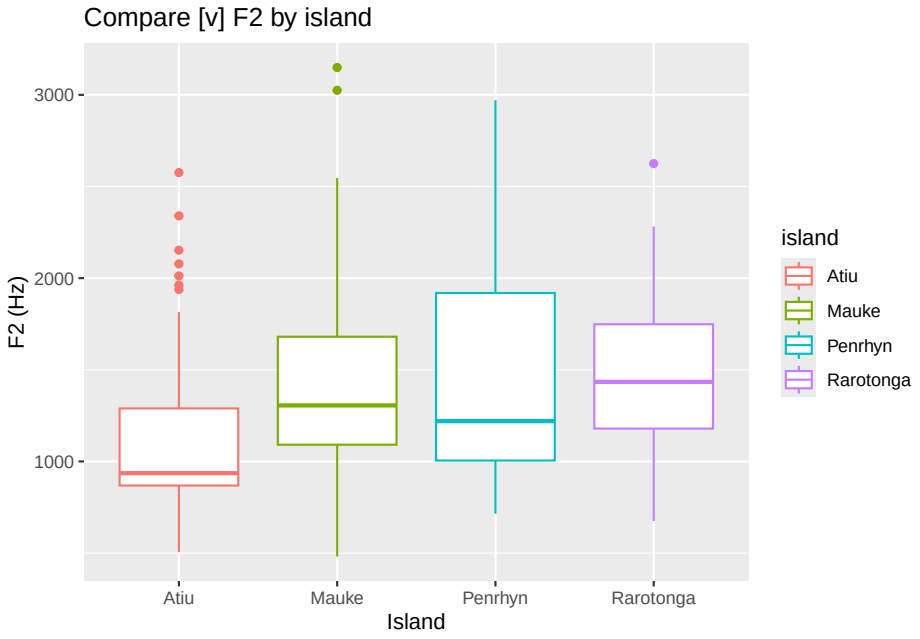

Figure 4

term	estimate	std.error	statistic	p.value
(Intercept)	1090.9432	39.79288	27.415534	0
islandMauke	308.5190	44.70677	6.900944	0
islandPenrhyn	373.9431	51.79373	7.219853	0
islandRarotonga	360.9223	52.73736	6.843768	0

5.1 Data analysis

I used R (Version 4.5.0; R Core Team, 2024) and the R-packages *broom* (Version 1.0.8; Robinson, Hayes, & Couch, 2025), *dplyr* (Version 1.1.4; Wickham, François, Henry, Müller, & Vaughan, 2023), *ds4ling* (Version 1.0; Casillas, 2025), *forcats* (Version 1.0.0; Wickham, 2023a), *ggplot2* (Version 3.5.2; Wickham, 2016), *ggpubr* (Version 0.6.0; Kassambara, 2023), *here* (Version 1.0.1; Müller, 2020), *lme4* (Version 1.1.37; Bates, Mächler, Bolker, & Walker, 2015), *lmerTest* (Version 3.1.3; Kuznetsova, Brockhoff, & Christensen, 2017), *lubridate* (Version 1.9.4; Grolemund & Wickham, 2011), *Matrix* (Version 1.7.3; Bates, Maechler, & Jagan, 2025), *papaja* (Version 0.1.3; Aust & Barth, 2024), *purrr* (Version 1.0.4; Wickham & Henry, 2025), *readr* (Version 2.1.5; Wickham, Hester, & Bryan, 2024), *stringr* (Version 1.5.1; Wickham, 2023b), *tibble* (Version 3.2.1; Müller & Wickham, 2023), *tidyr* (Version 1.3.1; Wickham, Vaughan, & Girlich, 2024), *tidyverse* (Version 2.0.0; Wickham et al., 2019) and *tinylabels* (Version 0.2.5; Barth, 2023) for all my analyses.

6.0 Conclusion

???fill in

References

Aust, F., & Barth, M. (2024). *papaja: Prepare reproducible APA journal articles with R Markdown*. https://doi.org/10.32614/CRAN.package.papaja

Barth, M. (2023). *tinylabels: Lightweight variable labels*. Retrieved from https://cran.r-project.org/package=tinylabels

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bates, D., Maechler, M., & Jagan, M. (2025). *Matrix: Sparse and dense matrix classes and methods*. https://doi.org/10.32614/CRAN.package.Matrix

Casillas, J. (2025). *ds4ling: Datasets and functions developed for SPAN658: Data science for linguists*. Retrieved from https://github.com/jvcasillas/ds4ling

Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, *40*(3), 1–25. Retrieved from https://www.jstatsoft.org/v40/i03/

Kassambara, A. (2023). *Ggpubr: 'ggplot2' based publication ready plots*. https://doi.org/10.32614/CRAN.package.ggpubr

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Müller, K. (2020). *Here: A simpler way to find your files*. https://doi.org/10.32614/CRAN.package.here

Müller, K., & Wickham, H. (2023). *Tibble: Simple data frames*. https://doi.org/10.32614/CRAN.package.tibble

R Core Team. (2024). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Robinson, D., Hayes, A., & Couch, S. (2025). *Broom: Convert statistical objects into tidy tibbles*. https://doi.org/10.32614/CRAN.package.broom

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from https://ggplot2.tidyverse.org

Wickham, H. (2023a). *Forcats: Tools for working with categorical variables (factors)*. https://doi.org/10.32614/CRAN.package.forcats

Wickham, H. (2023b). *Stringr: Simple, consistent wrappers for common string operations*. https://doi.org/10.32614/CRAN.package.stringr

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *Dplyr: A grammar of data manipulation*. https://doi.org/10.32614/CRAN.package.dplyr

Wickham, H., & Henry, L. (2025). *Purrr: Functional programming tools*. https://doi.org/10.32614/CRAN.package.purrr

Wickham, H., Hester, J., & Bryan, J. (2024). *Readr: Read rectangular text data*. https://doi.org/10.32614/CRAN.package.readr

Wickham, H., Vaughan, D., & Girlich, M. (2024). *Tidyr: Tidy messy data*. https://doi.org/10.32614/CRAN.package.tidyr