Appendices

# Appendix A: Modeling procedure

Our modeling strategy was tailored to our research questions on the nature of change over time in Southland /r/. In particular, we started with baseline models that included a single fixed effect, *Generation*, and random intercepts of *Speaker* and *Word*. We then tested additional fixed-effect predictors in two “bands”: first, predictors that had been shown to coincide with increased rhoticity in Bartlett (2002), then additional exploratory predictors. After all possible fixed effects were exhausted, random slopes were added where possible.

We first attempted to add *Preceding vowel*, *Following segment*, and *Region* to the model as fixed effects. We treated *Preceding vowel* as a four-way predictor: nurse, start, north, and an Other category that collapsed three smaller lexical sets (near, square, cure) that share in common a centering offglide (when the /r/ is absent). *Following segment* was initially fit in terms of individual segments, then collapsed into five levels: labial obstruents, coronal obstruents, dorsal obstruents, sonorants, and pauses. As mentioned above, *Region* was a two-way factor: urban versus rural. Once those predictors were exhausted, we attempted to add several exploratory predictors: *Word frequency*, *Lemma frequency*, *Word Southlandness*, *Lemma Southlandness*, and *Word-final*. The inclusion of *Word*/*Lemma frequency* was motivated by past findings indicating the role of frequency in mediating sound change (e.g., Hay, Pierrehumbert, Walker, & LaShell, 2015; Pierrehumbert, 2001); these frequencies were drawn from CELEX (Baayen, Piepenbrock, & van Rijn, 1995) and were logged and standardized (z-scored). The inclusion of *Word*/*Lemma Southlandness* was motivated by past findings that the social distribution of words impacts their participation in sound change (e.g., Hay & Foulkes, 2016; Walker & Hay, 2011). Southlandness (also logged and standardized) was calculated as a ratio of word/lemma frequency in Southland to word/lemma frequency in General New Zealand English; the latter was drawn from the Canterbury Corpus, a subset of the Origins of New Zealand English corpus (Gordon, Maclagan, & Hay, 2007). The inclusion of the binary *Word-final* predictor was motivated by past findings of the influence of prosodic position on variation in other sociolinguistic variables (e.g., in coronal stop deletion, Wolfram & Schilling-Estes [2006]).

Predictors were only added if the resultant model converged (with singular fits counting as nonconvergence), did not suffer from excessive multicollinearity, and significantly improved upon the fit of the previous model. If multiple candidate predictors satisfied these criteria, the predictor that yielded the greatest improvement in model fit was added first, as measured by the Akaike information criterion (Akaike, 1974). Multicollinearity was assessed on the basis of the generalized variance inflation factor via the car package for R (Fox & Monette, 1992; Fox & Weisberg, 2011; R Core Team, 2019), using the usual “less than 10” rule of thumb for variance inflation factors (O’Brien, 2007).1 Significant improvement upon model fit was assessed via a likelihood ratio test at = .05. Predictors were added as the maximal interaction that satisfied these criteria, with interactions reduced to lower orders where necessary based on whether the higher-order terms significantly improved model fit.2

We fit the nurse models following the same procedure described above, except without the *Preceding vowel* or *Word-final* predictors, and without the pause *Following segment* level, due to the scarcity of word-final and/or pre-pausal /r/ after nurse in our data. nurse-f also lacked a Word random intercept due to singular fits.

Notes

**1.** Readers should note that the generalized variance inflation factor (GVIF) returned by the vif function in the car package needs to be squared before comparing to 10 or any other rule of thumb. Fox and Weisberg (2011:325) noted, “if there are *p* regressors in a term, then GVIF1/2p is a one-dimensional expression of the decrease in the precision of estimation due to collinearity–analogous to taking the square root of the usual VIF. When *p* = 1, the GVIF reduces to the usual VIF.”

**2.** The exclusion of a given effect from a model does not necessarily indicate nonsignificance; predictors were not assessed for significance if adding them to the model caused nonconvergence or excessive multicollinearity. In other words, it is possible that some effects that actually influence rhoticity are nevertheless excluded from the model because their inclusion causes the model to be untenable. We have made the data available to allow readers to test the data under different assumptions (e.g., attempting to fit *Word frequency* before *Generation*) at <https://github.com/nzilbb/Sld-R-Data>.

## References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–23. <https://doi.org/10.1109/TAC.1974.1100705>

Baayen, H., Piepenbrock, R., & van Rijn, H. (1995). *The CELEX Lexical Database (Release 2, CD-ROM), LDC Catalogue No.: LDC96L14*.

Bartlett, C. (2002). *The Southland Variety of New Zealand English: Postvocalic /r/ and the BATH vowel* (Thesis).

Fox, J., & Monette, G. (1992). Generalized collinearity diagnostics. *Journal of the American Statistical Association*, *87*(417), 178–83. <https://doi.org/10.2307/2290467>

Fox, J., & Weisberg, S. (2011). *An R companion to applied regression* (2nd ed.). Retrieved from <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion>

Gordon, E., Maclagan, M., & Hay, J. (2007). The ONZE corpus. In J. C. Beal, K. P. Corrigan, & H. L. Moisl (Eds.), *Creating and digitizing language corpora: Volume 2: Diachronic Databases* (pp. 82–104). <https://doi.org/10.1057/9780230223202_4>

Hay, J., & Foulkes, P. (2016). The evolution of medial /t/ over real and remembered time. *Language*, *92*(2), 298–330. <https://doi.org/10.1353/lan.2016.0036>

Hay, J., Pierrehumbert, J. B., Walker, A. J., & LaShell, P. (2015). Tracking word frequency effects through 130 years of sound change. *Cognition*, *139*, 83–91. <https://doi.org/10.1016/j.cognition.2015.02.012>

O’Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, *41*(5), 673–90. <https://doi.org/10.1007/s11135-006-9018-6>

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. L. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 137–57). Amsterdam: John Benjamins.

R Core Team. (2019). *R: A language and environment for statistical computing*. Retrieved from <https://www.R-project.org/>

Walker, A., & Hay, J. (2011). Congruence between ’word age’ and ’voice age’ facilitates lexical access. *Laboratory Phonology*, *2*(1), 219. <https://doi.org/10.1515/labphon.2011.007>

Wolfram, W., & Schilling-Estes, N. (2006). *American speech* (2nd ed.). In *Language in Society* (2nd ed.). Malden, MA: Blackwell.

# Appendix B: Model summaries

The following tables display coefficients for the models reported in this paper. Colons denote interaction terms. Blanks in the right three columns denote reference levels (terms for which the model did not calculate coefficients due to treatment coding).

## Table B1. *Model all-f: Results of mixed-effects logistic regression model: factors predicting Present (overall*n*= 4402)*

| Predictor | Levels | *n* | % Present | *β* | SE | *p*-value |
| --- | --- | --- | --- | --- | --- | --- |
| (Intercept) |  |  |  | −3.388 | 0.746 | < .001 |
| Generation | 1900-1935 | 1924 | 17.7 |  |  |  |
|  | 1936-1955 | 1256 | 15.4 | −1.074 | 0.755 | .155 |
|  | 1956-1985 | 1222 | 21.3 | −0.534 | 0.747 | .475 |
| Preceding vowel | START | 911 | 7.1 |  |  |  |
|  | NORTH | 1437 | 4.0 | −0.019 | 0.616 | .976 |
|  | NURSE | 1237 | 47.0 | 1.607 | 0.492 | < .01 |
|  | Other | 817 | 11.0 | −0.428 | 0.608 | .482 |
| Region | Invercargill | 1835 | 14.2 |  |  |  |
|  | RuralSld | 2567 | 20.7 | 0.029 | 0.627 | .963 |
| Following segment | LabialObs | 204 | 16.2 |  |  |  |
|  | CoronalObs | 2433 | 12.7 | −0.430 | 0.326 | .187 |
|  | DorsalObs | 537 | 38.5 | 0.413 | 0.386 | .284 |
|  | Sonorant | 966 | 20.8 | −0.548 | 0.348 | .115 |
|  | Pause | 262 | 16.8 | 0.997 | 0.398 | < .05 |
| Generation : Preceding vowel | 1900-1935 : START | 414 | 10.9 |  |  |  |
|  | 1900-1935 : NORTH | 589 | 5.1 |  |  |  |
|  | 1900-1935 : NURSE | 555 | 37.7 |  |  |  |
|  | 1900-1935 : Other | 366 | 15.3 |  |  |  |
|  | 1936-1955 : START | 251 | 3.2 |  |  |  |
|  | 1936-1955 : NORTH | 426 | 2.8 | 0.070 | 0.687 | .919 |
|  | 1936-1955 : NURSE | 341 | 46.3 | 1.758 | 0.521 | < .001 |
|  | 1936-1955 : Other | 238 | 6.3 | 0.614 | 0.613 | .317 |
|  | 1956-1985 : START | 246 | 4.9 |  |  |  |
|  | 1956-1985 : NORTH | 422 | 3.6 | −0.098 | 0.665 | .883 |
|  | 1956-1985 : NURSE | 341 | 62.8 | 2.245 | 0.509 | < .001 |
|  | 1956-1985 : Other | 213 | 8.9 | 0.633 | 0.603 | .294 |
| Preceding vowel : Region | START : Invercargill | 358 | 5.0 |  |  |  |
|  | START : RuralSld | 553 | 8.5 |  |  |  |
|  | NORTH : Invercargill | 571 | 4.7 |  |  |  |
|  | NORTH : RuralSld | 866 | 3.5 | −1.447 | 0.592 | < .05 |
|  | NURSE : Invercargill | 557 | 36.1 |  |  |  |
|  | NURSE : RuralSld | 680 | 55.9 | 1.383 | 0.441 | < .01 |
|  | Other : Invercargill | 349 | 4.3 |  |  |  |
|  | Other : RuralSld | 468 | 16.0 | 1.164 | 0.544 | < .05 |

## Table B2. *Model all-m: Results of mixed-effects logistic regression model: factors predicting Present (overall*n*= 5908)*

| Predictor | Levels | *n* | % Present | *β* | SE | *p*-value |
| --- | --- | --- | --- | --- | --- | --- |
| (Intercept) |  |  |  | −2.622 | 0.763 | < .001 |
| Generation | 1900-1935 | 3786 | 38.8 |  |  |  |
|  | 1936-1955 | 1517 | 39.0 | −1.713 | 1.104 | .121 |
|  | 1956-1985 | 605 | 27.4 | −2.096 | 1.520 | .168 |
| Preceding vowel | START | 1393 | 31.9 |  |  |  |
|  | NORTH | 1767 | 12.6 | −0.984 | 0.396 | < .05 |
|  | NURSE | 1686 | 69.5 | 1.959 | 0.316 | < .001 |
|  | Other | 1062 | 36.5 | 0.773 | 0.367 | < .05 |
| Following segment | LabialObs | 371 | 38.0 |  |  |  |
|  | CoronalObs | 3318 | 30.3 | −0.436 | 0.424 | .304 |
|  | DorsalObs | 711 | 66.2 | 0.770 | 0.508 | .129 |
|  | Sonorant | 1257 | 40.2 | −0.564 | 0.459 | .219 |
|  | Pause | 251 | 41.4 | 0.724 | 0.649 | .264 |
| Region | Invercargill | 2367 | 28.4 |  |  |  |
|  | RuralSld | 3541 | 43.9 | 2.614 | 0.922 | < .01 |
| Word-final | FALSE | 5164 | 37.9 |  |  |  |
|  | TRUE | 744 | 36.4 | 0.115 | 0.257 | .656 |
| Generation : Preceding vowel | 1900-1935 : START | 968 | 39.5 |  |  |  |
|  | 1900-1935 : NORTH | 1142 | 17.3 |  |  |  |
|  | 1900-1935 : NURSE | 1050 | 62.3 |  |  |  |
|  | 1900-1935 : Other | 626 | 37.5 |  |  |  |
|  | 1936-1955 : START | 300 | 18.0 |  |  |  |
|  | 1936-1955 : NORTH | 437 | 4.6 | −0.623 | 0.413 | .131 |
|  | 1936-1955 : NURSE | 454 | 81.7 | 3.534 | 0.380 | < .001 |
|  | 1936-1955 : Other | 326 | 44.8 | 0.593 | 0.340 | .081 |
|  | 1956-1985 : START | 125 | 7.2 |  |  |  |
|  | 1956-1985 : NORTH | 188 | 2.1 | −0.335 | 0.816 | .681 |
|  | 1956-1985 : NURSE | 182 | 80.2 | 3.947 | 0.682 | < .001 |
|  | 1956-1985 : Other | 110 | 6.4 | −1.617 | 0.859 | .060 |
| Generation : Following segment | 1900-1935 : LabialObs | 201 | 35.8 |  |  |  |
|  | 1900-1935 : CoronalObs | 2071 | 31.8 |  |  |  |
|  | 1900-1935 : DorsalObs | 499 | 66.9 |  |  |  |
|  | 1900-1935 : Sonorant | 896 | 39.4 |  |  |  |
|  | 1900-1935 : Pause | 119 | 43.7 |  |  |  |
|  | 1936-1955 : LabialObs | 130 | 45.4 |  |  |  |
|  | 1936-1955 : CoronalObs | 880 | 29.3 | 0.930 | 0.633 | .142 |
|  | 1936-1955 : DorsalObs | 161 | 67.1 | 0.411 | 0.786 | .601 |
|  | 1936-1955 : Sonorant | 254 | 46.5 | 1.733 | 0.722 | < .05 |
|  | 1936-1955 : Pause | 92 | 52.2 | 0.537 | 0.910 | .555 |
|  | 1956-1985 : LabialObs | 40 | 25.0 |  |  |  |
|  | 1956-1985 : CoronalObs | 367 | 24.3 | 0.370 | 1.149 | .748 |
|  | 1956-1985 : DorsalObs | 51 | 56.9 | 1.283 | 1.371 | .349 |
|  | 1956-1985 : Sonorant | 107 | 31.8 | −0.724 | 1.217 | .552 |
|  | 1956-1985 : Pause | 40 | 10.0 | 1.734 | 1.696 | .307 |
| Generation : Region | 1900-1935 : Invercargill | 1146 | 21.8 |  |  |  |
|  | 1900-1935 : RuralSld | 2640 | 46.2 |  |  |  |
|  | 1936-1955 : Invercargill | 764 | 38.4 |  |  |  |
|  | 1936-1955 : RuralSld | 753 | 39.6 | −0.422 | 1.415 | .766 |
|  | 1956-1985 : Invercargill | 457 | 28.4 |  |  |  |
|  | 1956-1985 : RuralSld | 148 | 24.3 | 1.772 | 1.957 | .365 |
| Preceding vowel : Region | START : Invercargill | 471 | 11.5 |  |  |  |
|  | START : RuralSld | 922 | 42.4 |  |  |  |
|  | NORTH : Invercargill | 669 | 4.6 |  |  |  |
|  | NORTH : RuralSld | 1098 | 17.4 | −0.958 | 0.421 | < .05 |
|  | NURSE : Invercargill | 808 | 61.1 |  |  |  |
|  | NURSE : RuralSld | 878 | 77.1 | −0.300 | 0.357 | .400 |
|  | Other : Invercargill | 419 | 22.4 |  |  |  |
|  | Other : RuralSld | 643 | 45.7 | −0.430 | 0.385 | .264 |
| Following segment : Region | LabialObs : Invercargill | 154 | 30.5 |  |  |  |
|  | LabialObs : RuralSld | 217 | 43.3 |  |  |  |
|  | CoronalObs : Invercargill | 1367 | 21.1 |  |  |  |
|  | CoronalObs : RuralSld | 1951 | 36.8 | 0.821 | 0.504 | .103 |
|  | DorsalObs : Invercargill | 295 | 57.3 |  |  |  |
|  | DorsalObs : RuralSld | 416 | 72.6 | 0.276 | 0.611 | .651 |
|  | Sonorant : Invercargill | 439 | 31.4 |  |  |  |
|  | Sonorant : RuralSld | 818 | 44.9 | 0.651 | 0.543 | .230 |
|  | Pause : Invercargill | 112 | 27.7 |  |  |  |
|  | Pause : RuralSld | 139 | 52.5 | 0.693 | 0.769 | .368 |
| Generation : Word-final | 1900-1935 : FALSE | 3382 | 39.3 |  |  |  |
|  | 1900-1935 : TRUE | 404 | 34.7 |  |  |  |
|  | 1936-1955 : FALSE | 1269 | 36.9 |  |  |  |
|  | 1936-1955 : TRUE | 248 | 49.6 | 1.127 | 0.378 | < .01 |
|  | 1956-1985 : FALSE | 513 | 30.8 |  |  |  |
|  | 1956-1985 : TRUE | 92 | 8.7 | −0.249 | 0.949 | .793 |
| Generation : Following segment : Region | 1900-1935 : LabialObs : Invercargill | 66 | 22.7 |  |  |  |
|  | 1900-1935 : LabialObs : RuralSld | 135 | 42.2 |  |  |  |
|  | 1900-1935 : CoronalObs : Invercargill | 617 | 11.8 |  |  |  |
|  | 1900-1935 : CoronalObs : RuralSld | 1454 | 40.2 |  |  |  |
|  | 1900-1935 : DorsalObs : Invercargill | 180 | 55.0 |  |  |  |
|  | 1900-1935 : DorsalObs : RuralSld | 319 | 73.7 |  |  |  |
|  | 1900-1935 : Sonorant : Invercargill | 245 | 22.4 |  |  |  |
|  | 1900-1935 : Sonorant : RuralSld | 651 | 45.8 |  |  |  |
|  | 1900-1935 : Pause : Invercargill | 38 | 21.1 |  |  |  |
|  | 1900-1935 : Pause : RuralSld | 81 | 54.3 |  |  |  |
|  | 1936-1955 : LabialObs : Invercargill | 64 | 43.8 |  |  |  |
|  | 1936-1955 : LabialObs : RuralSld | 66 | 47.0 |  |  |  |
|  | 1936-1955 : CoronalObs : Invercargill | 468 | 30.1 |  |  |  |
|  | 1936-1955 : CoronalObs : RuralSld | 412 | 28.4 | −1.210 | 0.812 | .136 |
|  | 1936-1955 : DorsalObs : Invercargill | 75 | 60.0 |  |  |  |
|  | 1936-1955 : DorsalObs : RuralSld | 86 | 73.3 | −0.339 | 1.021 | .740 |
|  | 1936-1955 : Sonorant : Invercargill | 109 | 54.1 |  |  |  |
|  | 1936-1955 : Sonorant : RuralSld | 145 | 40.7 | −1.525 | 0.921 | .098 |
|  | 1936-1955 : Pause : Invercargill | 48 | 41.7 |  |  |  |
|  | 1936-1955 : Pause : RuralSld | 44 | 63.6 | −0.420 | 1.134 | .711 |
|  | 1956-1985 : LabialObs : Invercargill | 24 | 16.7 |  |  |  |
|  | 1956-1985 : LabialObs : RuralSld | 16 | 37.5 |  |  |  |
|  | 1956-1985 : CoronalObs : Invercargill | 282 | 26.2 |  |  |  |
|  | 1956-1985 : CoronalObs : RuralSld | 85 | 17.6 | −2.505 | 1.536 | .103 |
|  | 1956-1985 : DorsalObs : Invercargill | 40 | 62.5 |  |  |  |
|  | 1956-1985 : DorsalObs : RuralSld | 11 | 36.4 | −7.190 | 2.104 | < .001 |
|  | 1956-1985 : Sonorant : Invercargill | 85 | 28.2 |  |  |  |
|  | 1956-1985 : Sonorant : RuralSld | 22 | 45.5 | −1.153 | 1.812 | .524 |
|  | 1956-1985 : Pause : Invercargill | 26 | 11.5 |  |  |  |
|  | 1956-1985 : Pause : RuralSld | 14 | 7.1 | −4.444 | 2.144 | < .05 |

## Table B3. *Model nurse-f: Results of mixed-effects logistic regression model: factors predicting Present (overall*n*= 1237)*

| Predictor | Levels | *n* | % Present | *β* | SE | *p*-value |
| --- | --- | --- | --- | --- | --- | --- |
| (Intercept) |  |  |  | −1.636 | 0.710 | < .05 |
| Generation | 1900-1935 | 555 | 37.7 |  |  |  |
|  | 1936-1955 | 341 | 46.3 | 0.551 | 0.650 | .396 |
|  | 1956-1985 | 341 | 62.8 | 1.639 | 0.662 | < .05 |
| Following segment | LabialObs | 44 | 52.3 |  |  |  |
|  | CoronalObs | 471 | 41.6 | −0.734 | 0.420 | .081 |
|  | DorsalObs | 309 | 58.3 | 0.469 | 0.427 | .271 |
|  | Sonorant | 413 | 44.1 | −0.552 | 0.425 | .194 |
| Region | Invercargill | 557 | 36.1 |  |  |  |
|  | RuralSld | 680 | 55.9 | 1.454 | 0.537 | < .01 |

## 

## Table B4. *Model nurse-m. Results of mixed-effects logistic regression model: factors predicting Present (overall*n*= 1686).*

| Predictor | Levels | *n* | % Present | *β* | SE | *p*-value |
| --- | --- | --- | --- | --- | --- | --- |
| (Intercept) |  |  |  | −0.919 | 0.885 | .299 |
| Following segment | LabialObs | 92 | 50.0 |  |  |  |
|  | CoronalObs | 650 | 64.8 | 0.082 | 0.634 | .898 |
|  | DorsalObs | 453 | 84.1 | 2.309 | 0.678 | < .001 |
|  | Sonorant | 491 | 65.8 | 0.382 | 0.630 | .544 |
| Generation | 1900-1935 | 1051 | 62.3 |  |  |  |
|  | 1936-1955 | 453 | 81.7 | 2.281 | 1.092 | < .05 |
|  | 1956-1985 | 182 | 80.2 | 1.651 | 1.627 | .310 |
| Region | Invercargill | 807 | 61.1 |  |  |  |
|  | RuralSld | 879 | 77.1 | 0.941 | 1.049 | .370 |
| Following segment : Generation | LabialObs : 1900-1935 | 40 | 32.5 |  |  |  |
|  | LabialObs : 1936-1955 | 46 | 60.9 |  |  |  |
|  | LabialObs : 1956-1985 | 6 | 83.3 |  |  |  |
|  | CoronalObs : 1900-1935 | 344 | 52.6 |  |  |  |
|  | CoronalObs : 1936-1955 | 209 | 76.1 | 0.047 | 0.760 | .951 |
|  | CoronalObs : 1956-1985 | 97 | 83.5 | −0.469 | 1.375 | .733 |
|  | DorsalObs : 1900-1935 | 326 | 80.1 |  |  |  |
|  | DorsalObs : 1936-1955 | 97 | 95.9 | 0.388 | 0.923 | .675 |
|  | DorsalObs : 1956-1985 | 30 | 90.0 | −0.682 | 1.619 | .674 |
|  | Sonorant : 1900-1935 | 341 | 58.7 |  |  |  |
|  | Sonorant : 1936-1955 | 101 | 89.1 | 0.544 | 0.845 | .520 |
|  | Sonorant : 1956-1985 | 49 | 67.3 | −2.022 | 1.390 | .146 |
| Following segment : Region | LabialObs : Invercargill | 63 | 54.0 |  |  |  |
|  | LabialObs : RuralSld | 29 | 41.4 |  |  |  |
|  | CoronalObs : Invercargill | 346 | 54.0 |  |  |  |
|  | CoronalObs : RuralSld | 304 | 77.0 | 1.813 | 0.805 | < .05 |
|  | DorsalObs : Invercargill | 183 | 82.5 |  |  |  |
|  | DorsalObs : RuralSld | 270 | 85.2 | −0.114 | 0.821 | .889 |
|  | Sonorant : Invercargill | 215 | 56.3 |  |  |  |
|  | Sonorant : RuralSld | 276 | 73.2 | 1.026 | 0.803 | .201 |