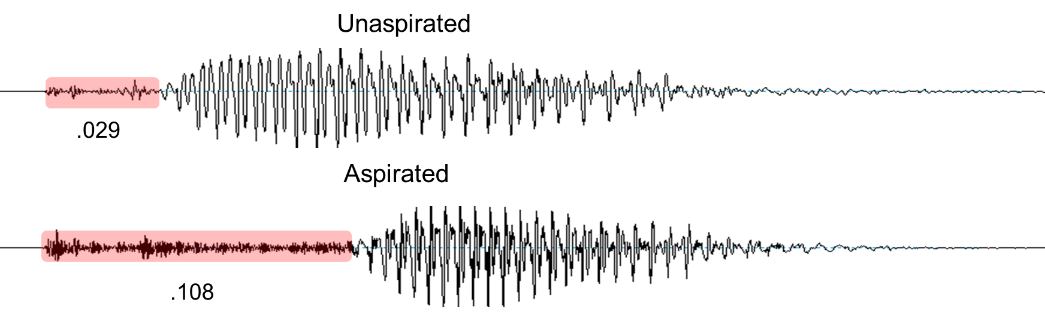
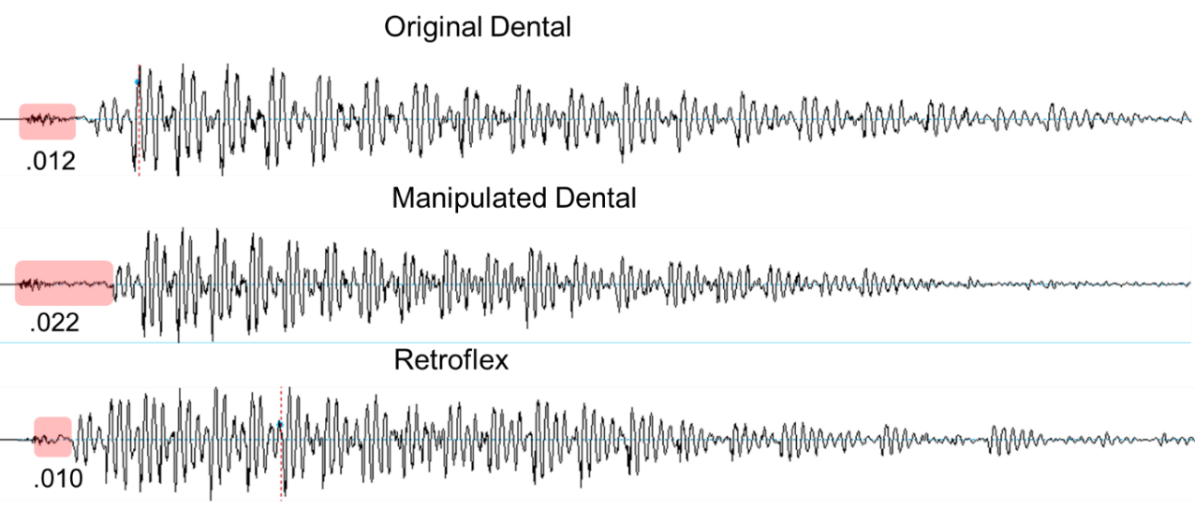
**Supplementary Material 1**



Example sound waveforms of ‘k’ phoneme recordings from one speaker.

The red box shows the VOT (s). The aspirated phoneme has a longer VOT than does the unaspirated phoneme. Listen to Supplementary Audio Files 3 and 4 to hear examples of the unaspirated and aspirated phonemes, respectively.

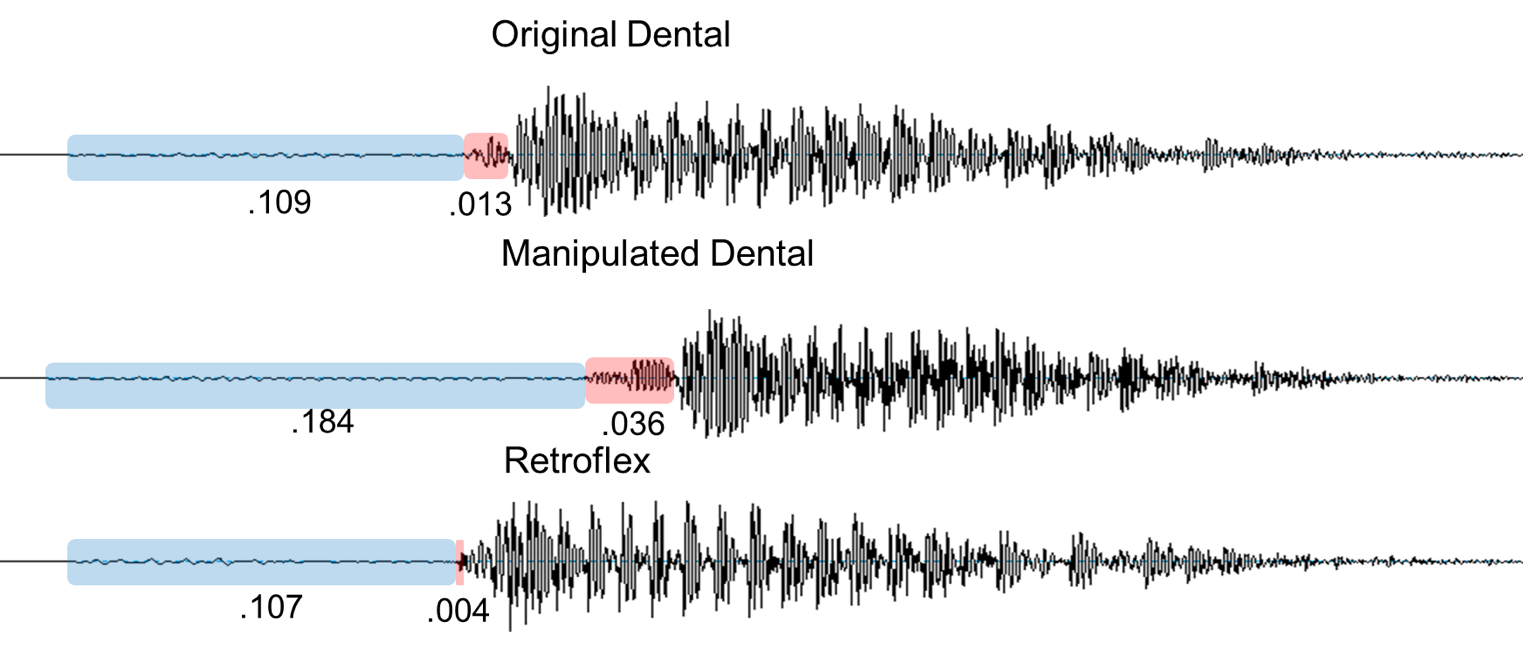
**Supplementary Material 2**

~~~~

Example sound waveforms of ‘t’ phoneme recordings from one speaker.

The red box shows the VOT. The retroflex has a slightly shorter VOT than the original dental. The length of the VOT is increased in the manipulated dental. Listen to Supplementary Audio Files 5, 6, and 7 to hear examples of the original dental, manipulated dental, and retroflex, respectively.

**Supplementary Material 3**



Example sound waveforms of ‘d’ phoneme recordings from one speaker.

The red box shows the VOT (s). The retroflex has a slightly shorter VOT than the original dental. The blue box shows the pre-voicing (s). The retroflex and original dental have similar pre-voicing durations. The length of the pre-voicing duration and the VOT is increased in the manipulated dental. Listen to Supplementary Audio File 8, 9, and 10 to hear examples of the original dental, manipulated dental, and retroflex, respectively.

Supplementary Material 4

*One heritage Marathi speaker’s performance on the tasks*

|  |  |  |  |
| --- | --- | --- | --- |
| Measure | Score K | Score T | Score D |
| Phoneme PN Test (out of 24)  Word PN Test (out of 20)  Word PN Learning (out of 20)  Mispronunciation Detection (out of 20) | 24  20  20  20 | 21  20  18  18 | 22  20  18  19 |

*Note.* She performed well on all tasks, demonstrating that the materials were good examples of the contrasts. She did not do the Phoneme PN Learning because the same items were used at learning and test.

### Supplementary Material 5: Repetition

*Fixed effect estimates (top) and variance estimates (bottom) for logit model of the repetition analyses with individual differences*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fixed Effect | Coefficient | SE | *z* | *p* |
| Intercept | .547 | .521 | 1.050 | .294 |
| Phoneme (K > D, T) | 2.517 | 1.110 | 2.268 | .023\* |
| Phoneme (T > D) | .295 | 1.240 | .238 | .812 |
| Orthography (N, M > E) | .071 | .178 | .397 | .692 |
| Orthography (M > N) | .214 | .207 | 1.031 | .303 |
| Test (post > pre) | .610 | .182 | 3.348 | <.001\* |
| Test (immediate > delayed) | .192 | .201 | .954 | .340 |
| Rise Time Discrimination | -.412 | .121 | -3.405 | <.001\* |
| PAD | .448 | .139 | 3.231 | .001\* |
| Orthography (N, M > E) \* Test (post > pre) | -.380 | .382 | -.993 | .321 |
| Orthography (M > N) \* Test (post > pre) | .375 | .439 | .856 | .392 |
| Orthography (N, M > E) > Test (immediate > delayed) | -.160 | .422 | -.379 | .705 |
| Orthography (M > N) \* Test (immediate > delayed) | -.136 | .497 | -.274 | .784 |
| Test (post > pre) \* Rise Time Discrimination | .245 | .190 | 1.290 | .197 |
| Test (immediate > delayed) \* Rise Time Discrimination | -.195 | .202 | -.965 | .335 |
| Test (post > pre) \* PAD | -.291 | .220 | -1.324 | .186 |
| Test (immediate > delayed) \* PAD | -.201 | .232 | -.866 | .387 |
| Random Effect | s2 |  |  |  |
| Subject | .321 |  |  |  |
| Item | 1.501 |  |  |  |

*Note.* n = 46, log-likelihood: -447.6, \* *p* < .05.

*Fixed effect estimates (top) and variance estimates (bottom) for logit model of the repetition analyses with all participants*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fixed Effect | Coefficient | SE | *z* | *p* |
| Intercept | .636 | .519 | 1.226 | .220 |
| Phoneme (K > D, T) | 2.039 | 1.086 | 1.877 | .061† |
| Phoneme (T > D) | .126 | 1.235 | .102 | .919 |
| Orthography (N, M > E) | .141 | .146 | .965 | .335 |
| Orthography (M > N) | .274 | .169 | 1.627 | .104 |
| Test (post > pre) | .679 | .146 | 4.646 | < .001\* |
| Test (immediate > delayed) | .206 | .167 | 1.233 | .218 |
| Orthography (N, M > E) \* Test (post > pre) | -.272 | .309 | -.878 | .380 |
| Orthography (M > N) \* Test (post > pre) | .164 | .353 | -.809 | .419 |
| Orthography (N, M > E) > Test (immediate > delayed) | -.285 | .353 | -.809 | .419 |
| Orthography (M > N) \* Test (immediate > delayed) | .293 | .409 | .716 | .474 |
| Random Effect | s2 |  |  |  |
| Subject | .778 |  |  |  |
| Item | 1.500 |  |  |  |

*Note.* n = 69, log-likelihood: -683.3, \* *p* < .05, † *p* < 0.10.

#### Supplementary Material 6: Phoneme perception

*Fixed effect estimates (top) and variance estimates (bottom) for logit model of the phoneme perception analyses with individual differences*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fixed Effect | Coefficient | SE | *z* | *p* |
| Intercept | 1.735 | 0.217 | 7.989 | <.001\* |
| Phoneme (K > D, T) | 1.741 | 0.320 | 5.434 | <.001\* |
| Phoneme (T > D) | 0.791 | 0.306 | 2.583 | .010\* |
| Orthography (N, M > E) | 0.319 | 0.179 | 1.779 | 0.075† |
| Orthography (M > N) | -0.168 | 0.218 | -0.770 | 0.441 |
| Test (post > pre) | 0.783 | 0.098 | 7.974 | <.001\* |
| Test (immediate > delayed) | -0.019 | 0.089 | -0.215 | 0.830 |
| Order (2 > 1) | 0.372 | 0.211 | 1.757 | 0.079† |
| Order (3 > 1) | 0.570 | 0.218 | 2.616 | 0.009\* |
| Rise Time Discrimination | -0.161 | 0.149 | -1.080 | 0.280 |
| PAD | 0.232 | 0.170 | 1.368 | 0.171 |
| Orthography (N, M > E) \* Test (post > pre) | 0.237 | 0.126 | 1.888 | 0.059† |
| Orthography (M > N) \* Test (post > pre) | -0.149 | 0.153 | -0.974 | 0.330 |
| Orthography (N, M > E) > Test (immediate > delayed) | -0.232 | 0.152 | -1.527 | 0.127 |
| Orthography (M > N) \* Test (immediate > delayed) | 0.099 | 0.191 | 0.521 | 0.603 |
| Test (post > pre) \* PAD | 0.273 | 0.110 | 2.494 | 0.013\* |
| Test (immediate > delayed) \* PAD | -0.053 | 0.097 | -0.546 | 0.585 |
| Test (post > pre) \* Rise Time Discrimination | -0.176 | 0.097 | -1.822 | 0.068† |
| Test (immediate > delayed) \* Rise Time Discrimination | 0.059 | 0.086 | 0.678 | 0.498 |
| Random Effect | s2 |  |  |  |
| Phoneme (K > D, T) | Subject | 2.089 |  |  |  |
| Phoneme (T > D) | Subject | 1.198 |  |  |  |
| Test (post > pre) | Subject | 0.263 |  |  |  |
| Test (immediate > delayed) | Subject | 0.097 |  |  |  |
| Subject | 0.950 |  |  |  |
| Phoneme Pair | 0.019 |  |  |  |
| Phoneme Pair in Order | 0.171 |  |  |  |

*Note.* n = 46, log-likelihood: -4011.4, \* *p* < .05, † *p* < 0.10.

*Fixed effect estimates (top) and variance estimates (bottom) for logit model of the phoneme perception analyses with all participants*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fixed Effect | Coefficient | SE | *z* | *p* |
| Intercept | 1.645 | .185 | 8.902 | <.001\* |
| Phoneme (K > D, T) | 1.819 | .278 | 6.547 | <.001\* |
| Phoneme (T > D) | .666 | .287 | 2.323 | .020\* |
| Orthography (N, M > E) | .103 | .147 | .697 | .486 |
| Orthography (M > N) | .044 | .172 | .258 | .797 |
| Test (post > pre) | .823 | .092 | 8.970 | <.001\* |
| Test (immediate > delayed) | -.097 | .068 | -1.155 | .248 |
| Order (2 > 1) | .354 | .171 | 2.076 | .038\* |
| Order (3 > 1) | .662 | .172 | 3.840 | <.001\* |
| Orthography (N, M > E) \* Test (post > pre) | .023 | .104 | .222 | .824 |
| Orthography (M > N) \* Test (post > pre) | -.175 | .123 | -1.425 | .154 |
| Orthography (N, M > E) > Test (immediate > delayed) | -.162 | .126 | -1.292 | .196 |
| Orthography (M > N) \* Test (immediate > delayed) | .060 | .149 | .400 | .689 |
| Random Effect | s2 |  |  |  |
| Phoneme (K > D, T) | Subject | 1.291 |  |  |  |
| Phoneme (T > D) | Subject | 1.741 |  |  |  |
| Test (post > pre) | Subject | .070 |  |  |  |
| Test (immediate > delayed) | Subject | .402 |  |  |  |
| Subject | .945 |  |  |  |
| Phoneme Pair | .031 |  |  |  |
| Phoneme Pair in Order | .164 |  |  |  |

*Note.* n = 69, log-likelihood: -6109.2, \* *p* < .05.

#### Supplementary Material 7: Word perception

*Fixed effect estimates (top) and variance estimates (bottom) for logit model of the word perception analyses with individual differences*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fixed Effect | Coefficient | SE | *z* | *p* |
| Intercept | 0.919 | 0.163 | 5.650 | <.001\* |
| Phoneme (K > D, T) | 1.526 | 0.269 | 5.668 | <.001\* |
| Phoneme (T > D) | 0.293 | 0.250 | 1.171 | 0.241 |
| Orthography (N, M > E) | 0.168 | 0.119 | 1.416 | 0.157 |
| Orthography (M > N) | 0.042 | 0.145 | 0.288 | 0.774 |
| Test (post > pre) | 0.602 | 0.090 | 6.720 | <.001\* |
| Test (immediate > delayed) | -0.073 | 0.069 | -1.069 | 0.285 |
| Order (2 > 1) | 0.334 | 0.143 | 2.344 | 0.019\* |
| Order (3 > 1) | 0.531 | 0.146 | 3.629 | <.001\* |
| Rise Time Discrimination | -0.197 | 0.106 | -1.855 | 0.064† |
| PAD | 0.162 | 0.121 | 1.333 | 0.182 |
| Orthography (N, M > E) \* Test (post > pre) | 0.226 | 0.120 | 1.891 | 0.059† |
| Orthography (M > N) \* Test (post > pre) | 0.142 | 0.143 | 0.990 | 0.322 |
| Orthography (N, M > E) > Test (immediate > delayed) | -0.111 | 0.142 | -0.779 | 0.436 |
| Orthography (M > N) \* Test (immediate > delayed) | 0.309 | 0.171 | 1.808 | 0.071† |
| Test (post > pre) \* Rise Time Discrimination | -0.060 | 0.089 | -0.679 | 0.497 |
| Test (immediate > delayed) \* Rise Time Discrimination | 0.027 | 0.067 | 0.407 | 0.684 |
| Test (post > pre) \* PAD | -0.119 | 0.102 | -1.165 | 0.244 |
| Test (immediate > delayed) \* PAD | -0.076 | 0.078 | -0.977 | 0.329 |
| Random Effect | s2 |  |  |  |
| Phoneme (K > D, T) | Subject | 1.200 |  |  |  |
| Phoneme (T > D) | Subject | 0.403 |  |  |  |
| Test (post > pre) | Subject | 0.217 |  |  |  |
| Subject | 0.474 |  |  |  |
| Word Pair | 0.149 |  |  |  |
| Speaker | 0.080 |  |  |  |

*Note.* n = 46, log-likelihood: -4244.9, \* *p* < .05, † *p* < 0.10.

*Fixed effect estimates (top) and variance estimates (bottom) for logit model of the word perception analyses with all participants*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fixed Effect | Coefficient | SE | *z* | *p* |
| Intercept | 0.828 | 0.145 | 5.704 | <.001\* |
| Phoneme (K > D, T) | 1.613 | 0.246 | 6.550 | <.001\* |
| Phoneme (T > D) | 0.331 | 0.244 | 1.355 | 0.175 |
| Orthography (N, M > E) | 0.084 | 0.106 | 0.790 | 0.429 |
| Orthography (M > N) | 0.129 | 0.124 | 1.035 | 0.301 |
| Test (post > pre) | 0.648 | 0.066 | 9.787 | <.001\* |
| Test (immediate > delayed) | -0.042 | 0.055 | -0.756 | .450 |
| Order (2 > 1) | 0.358 | 0.125 | 2.858 | 0.004\* |
| Order (3 > 1) | 0.607 | 0.125 | 4.854 | <.001\* |
| Orthography (N, M > E) \* Test (post > pre) | 0.182 | 0.098 | 1.845 | 0.065† |
| Orthography (M > N) \* Test (post > pre) | -0.011 | 0.115 | -0.092 | 0.927 |
| Orthography (N, M > E) > Test (immediate > delayed) | -0.083 | 0.117 | -0.713 | 0.476 |
| Orthography (M > N) \* Test (immediate > delayed) | 0.149 | 0.137 | 1.091 | 0.275 |
| Random Effect | s2 |  |  |  |
| Phoneme (K > D, T) | Subject | 1.118 |  |  |  |
| Phoneme (T > D) | Subject | 0.556 |  |  |  |
| Test (post > pre) | Subject | 0.153 |  |  |  |
| Subject | 0.478 |  |  |  |
| Word | 0.151 |  |  |  |
| Speaker | 0.080 |  |  |  |

*Note.* n = 69, log-likelihood: -6396.6, \* *p* < .05, † *p* < 0.10.

Supplementary Material 8: Mispronunciation detection

*Fixed effect estimates (top) and variance estimates (bottom) for logit model of the mispronunciation detection task*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fixed Effect | Coefficient | SE | *z* | *p* |
| Intercept | 0.845 | 0.106 | 7.994 | <.001\* |
| Phoneme Perception Immediate Post-test Performance | 3.547 | 0.506 | 7.014 | <.001\* |
| Orthography (E, M > N) | 0.300 | 0.079 | 3.812 | <.001\* |
| Orthography (E > M) | -0.060 | 0.091 | -0.665 | 0.506 |
| Test (immediate > delayed) | 0.281 | 0.067 | 4.208 | <.001\* |
| Order (2 > 1) | -0.122 | 0.099 | -1.223 | 0.221 |
| Order (3 > 1) | -0.061 | 0.101 | -0.606 | 0.544 |
| Phoneme Perception Immediate Post-test Performance \* Orthography (E, M > N) | 0.520 | 0.952 | 0.546 | 0.585 |
| Phoneme Perception Immediate Post-test Performance \* Orthography (E > M) | 0.921 | 1.316 | 0.700 | 0.484 |
| Phoneme Perception Immediate Post-test Performance \* Test (immediate > delayed) | 1.345 | 0.755 | 1.782 | 0.075† |
| Orthography (E, M > N) \* Test (immediate > delayed) | 0.277 | 0.139 | 2.000 | 0.045\* |
| Orthography (E > M) \* Test (immediate > delayed) | 0.043 | 0.163 | 0.265 | 0.791 |
| Phoneme Perception Immediate Post-test Performance \* Orthography (E, M > N) \* Test (immediate > delayed) | -0.140 | 1.448 | -0.097 | 0.923 |
| Phoneme Perception Immediate Post-test Performance \* Orthography (E > M) \* Test (immediate > delayed) | 3.002 | 2.009 | 1.494 | 0.135 |
| Random Effect | s2 |  |  |  |
| Immediate Post-test | Subject | 0.144 |  |  |  |
| Delayed Post-test | Subject | 0.198 |  |  |  |
| Subject | <.001 |  |  |  |
| Item | 0.127 |  |  |  |

*Note.* n = 66, log-likelihood: -2907.6, \* *p* < .05, † *p* < 0.10.