CI mispronunciation sensitivity: results

Meg Cychosz

17 May 2022

Table 1: Participant demographic information (N=19 matches), Mean (SD), range.

Hearing Status	Girls, Boys	Chronological Age: months	Hearing Age: months	Maternal Education Level	EVT-2
CochlearImplant	13 , 6	56.32 (6.71) 44 - 66	44.47 (8.69) 29 - 56	6.26 (0.56)	134.84
NormalHearing	13 , 6	45.47 (7.46) 36 - 57	45.47 (7.46) 36 - 57	6.05 (0.62)	132.68

1 Results

1.1 Analysis plan

We evaluate the mispronunciation sensitivity of children with CIs in comparison to their hearing age- and vocabulary size-matched peers. The outcome variable is the proportion of looks to the familiar object versus the unfamiliar object, over time (300–1,800 ms after target word onset), which we modeled using Generalized Additive Mixed Models (GAMMs). GAMMs have become an important tool to model time series data, such as eyetracking trajectories, because they can incorporate nonlinear relationships between variables such as time, response, and relevant covariates (i.e., effects of group and/or condition) (rijChildrenEyeGaze2016;zahnerAlignmentF0Peak). GAMMs are composed of (fixed) *parametric* terms that model static relationships between two variables, as is common in generalized linear modeling, and *smooth* terms that model nonlinear effects by using penalized basis functions (smoothing splines). Parametric terms can be interpreted from model summaries, as in traditional regression, but smooth terms must be interpreted visually. Wieling (2017) and Sóskuthy (2021) provide tutorials for GAMMs in language research and Wood (2017) provides a comprehensive textbook treatment of the approach.

GAMMs also allow for autocorrelation between observations to be factored into the modeling. Incorporating autocorrelation is of particular importance to eyetracking data where we anticipate large amounts of within-trial correlation between measurements over time: the area where the child is looking at 500ms is highly correlated with where they are looking at 550ms. As such, GAMMs are a significant improvement upon other polynomial regression models common in time series analysis, such as Growth Curve Models (GCMs). The standard approach for estimating GCMs cannot factor in this inherent correlational structure within the data and, as a result, recent work has shown that they result in inflated Type I error rates (Huang & Snedeker, 2020).

The current data were analyzed in the RStudio computing environment (R version 4.0.2; Rcite). All computing and statistical analyses are included in the GitHub repository affiliated with this project <github.com/megseekosh/ci-mispron>. Visualizations were made using the ggplot2 and cowplot packages. Modeling was conducted and presented using the mgcv, itsadug, and tidymv packages. For all modeling, the proportion of children's looks to the familiar object versus the unfamiliar object was calculated for each frame (every 50 ms) and transformed to empirical logit (elog), or the log-odds of looking to the familiar object at each sample (Barr, 2008). Random effects ("factor smooths" in GAMMs) included by-participant, by-observation (first [at age 3], second [at age 4], or third [at age 5] visit to the lab), and by-item trajectories. These factor smooths modeled variability stemming from individual children and lexical items, and took into account the repeated observations from some children at two different ages. To ensure assumptions were met

and to avoid overfitting, model criticism was conducted using the gam.check() function; when necessary, the basis function number (k or knots) was increased.

As is common in eye-tracking data, the response data were distributed with heavy tails. Consequently, all models were fit using a scaled-t model using the "scat" link function, which substantially improved data distribution (woodSmoothingParameterModel2016). Finally, for each model, autocorrelation between model residuals was calculated; all models showed high amounts of autocorrelation. These dependencies were factored into the modeling by fitting an autoregressive error model (AR(1)) that modeled the degree of autocorrelation (*rho*) between timepoints in each trial. Subsequent model inspection demonstrated that specifying this autocorrelation value in the model sufficiently factored out autocorrelation between residuals (Wieling, 2017).

1.2 Evaluating the effect of phonetic detail on mispronunciation sensitivity

To evaluate how access to fine, phonetic detail may affect mispronunciation sensitivity, a series of GAMMs were fit comparing children with CIs and their hearing age- and vocabulary size-matched TH peers. **Condition** (Correct Pronunciation vs. Mispronunciation) was contrast-coded to facilitate model interpretation and the 2x2 relationship of **Group** (Children with CIs vs. TH) and **Condition** was modeled using ordered factors. A model with parametric and smooth terms for **Group** and **Condition** improved upon a **Condition**-only model, suggesting that children with CIs and TH responded differently to correct- versus mis-pronunciations.

To statistically evaluate the source of the **Group** effect (i.e., stemming from overall vs. time-varying response to the stimuli), another model was fit that included parametric terms for **Group**, and the ordered factors of *Correct Pronunciation* for children with CIs and *Correct Pronunciation* for children with TH (Wieling, 2017). These parametric effects modeled the constant effect (the intercept) of the covariates upon the response variable. Smooth model terms included non-linear effects of **Time** and **Time** by **Group**. The latter allowed us to model the non-linear difference between the two different groups' responses to mispronunciations. Finally, the model included difference smooths, which allowed us to separately model how each hearing group responded to correct- versus mis-pronunciations, over time. See Table **??** for model summary.

```
## % latex table generated in R 4.0.2 by xtable 1.8-4 package
## % Tue May 17 12:23:51 2022
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrr}
##    \hline
## A. parametric coefficients & Estimate & Std. Error & t-value & p-value \\
##   Intercept (Cochlear Implant: Mispronunciation) & 0.2662 & 0.1920 & 1.3868 & 0.1655 \\
##   Typical Hearing & -0.3259 & 0.1924 & -1.6932 & 0.0904 \\
##   Typical Hearing: Correct & 1.1826 & 0.2259 & 5.2363 & $<$ 0.0001 \\
##   Cochlear Implant: Correct & 0.6789 & 0.2252 & 3.0138 & 0.0026 \\
##    \hline
## B. smooth terms & edf & Ref.df & F-value & p-value \\
##   s(Time) & 0.0002 & 0.0004 & 755.4057 & 0.5803 \\
##   s(Time,Cochlear Implant) & 1.0001 & 1.0002 & 2.6265 & 0.1055 \\
##   s(Time,Typical Hearing) & 3.2556 & 4.1432 & 1.3554 & 0.2429 \\
##   s(Time,Typical Hearing; Correct) & 5.0601 & 6.3159 & 8.6538 & $<$ 0.0001 \\
##   s(Time,Cochlear Implant; Correct) & 3.7598 & 4.8315 & 7.2809 & $<$ 0.0001 \\
##   s(Time,Child) & 52.1753 & 341.0000 & 0.4680 & $<$ 0.0001 \\
##   s(Time,Item) & 25.5375 & 108.0000 & 0.6854 & $<$ 0.0001 \\
##   s(Time,Observation) & 0.0010 & 27.0000 & 0.1394 & $<$ 0.0001 \\
##    \hline
## \end{tabular}
## \caption{Model summary predicting the difference between proportion of looks to the familiar object b
## \label{2-by-2-model-summary}
## \end{table}
```

In the first part of the results, we ask: are both children with CIs and their TH matches sensitive to mispronunciations? Parametric effects in the model summary show that there are, overall, significantly more looks to the familiar photo for *Correct Pronunciation* trials than *Mispronunciation* trials, for both children with CIs and TH (CI logit Est.=0.68, p<.001, or proportion Est: 0.2106274; TH logit Est.=1.18, p<.001, or proportion Est: 0.30632). we interpret the smooth terms by first considering effective degrees of freedom (EDF) and the significance test for each smooth. The EDF indicates how much wiggliness there is in a smooth where EDF=1 indicates a linear relationship and a larger value indicates more wiggliness in the smooth. Interpretation of the non-linear smooths shows that there are significant, non-linear differences in looks to the familiar object between correct- and mis-pronunciations for children with CIs (smooth of **Time** by the ordered **Cochlear Implant; Correct**) and children with TH (smooth of **Time** by the ordered **Typical Hearing; Correct**) (Figure 1). Thus, both children with CIs and TH are sensitive to mispronunciations.
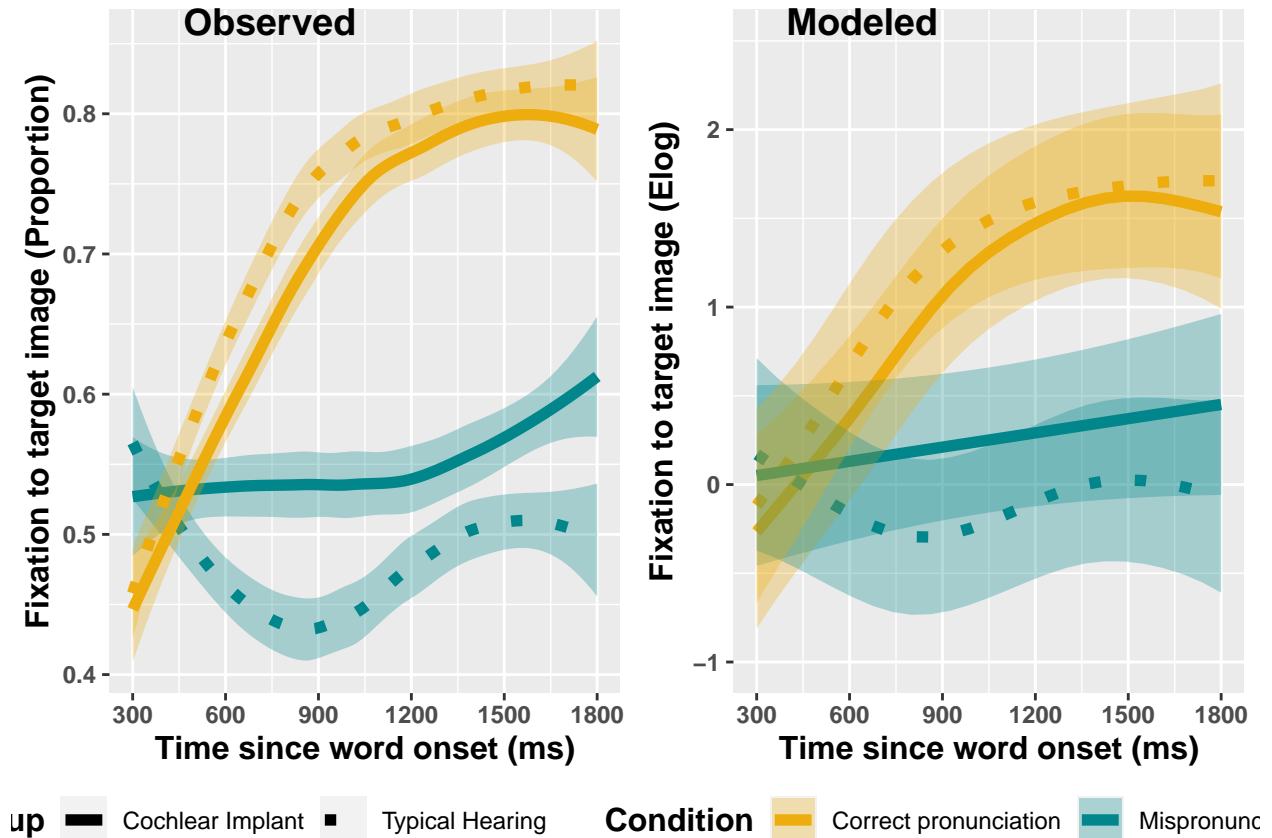

Figure 1: Observed data and GAMM predictions for proportion of looks to familiar object, by word condition and hearing status. Fixations on the y-axis are plotted as the empirical logit values (elog). Shaded ribbons represent 95% confidence intervals.

Nevertheless, the above modeling cannot tell us if these children with CIs are *less* sensitive to mispronunciations than their TH peers; the modeling demonstrates only that both groups show sensitivity. To evaluate differences in mispronunciation sensitivity by group, another GAMM was fit, with a binary difference smooth, which allowed us to evaluate the *difference* between smooths (Correct- vs. Mis-pronunciations) for children with CIs and TH, over time. Model fit included parametric effects of **Group**, as well as smooths of **Time**, **Time** by **Group**, **Time** by **Condition**, and **Time** by the ordered variable of **Group** by **Condition** (to model the difference between real- and mis-pronunciations for each group). Model results are plotted in Figure 2; the model summary is included in supplementary materials. Overall, the model-estimated difference smooths show smaller differences between correct- and mis-pronunciations for the children with CIs—and that these differences take longer to manifest during online processing (left panel of Figure 2).

Further inspection of the first model, as plotted in Figure 3, demonstrates why this is the case. The children with CIs and TH do not respond significantly differently to correct pronunciations: once vocabulary size and hearing age are controlled, both groups of children respond similarly to correctly-pronounced words. Instead, children with CIs—who are listening with a degraded speech signal via electric hearing—are less sensitive to *mis*-pronunciations (Figure 3), resulting in smaller difference smooths between correct- and mis-pronunciations.
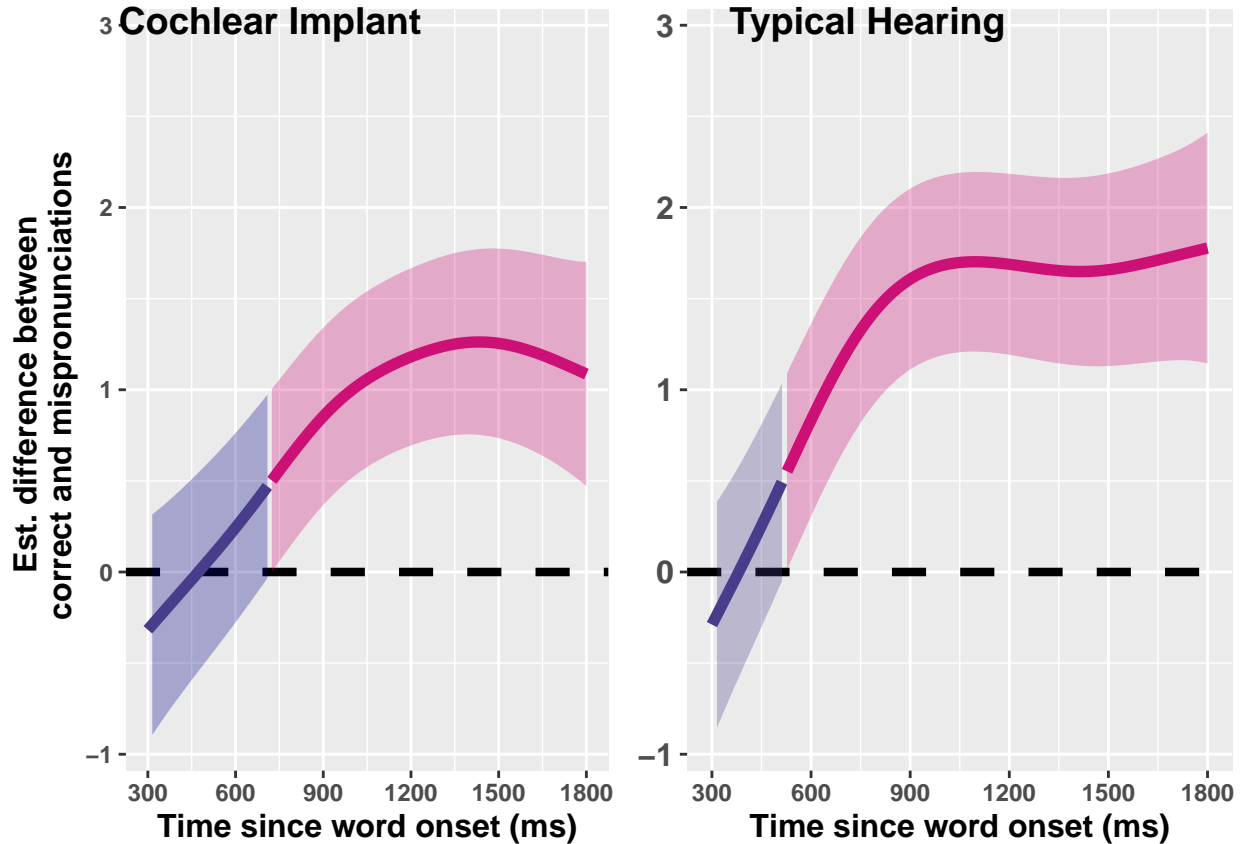

Figure 2: Difference smooths (GAMM predictions) by condition (correct- vs. mis-pronunciations) for children with CIs (L) and TH (R). Pink smooths represent the point when correct and mis-pronunciation smooths differ (i.e., reliable effect of condition) for each group: there is a larger difference between correct and mispronunciation responses for children with TH than CIs. Shaded ribbons represent 95% confidence intervals.

1.3 Explaining individual differences in mispronunciation sensitivity

Table 2: Summary statistics of standardized speech-language measures, by hearing status (N=33 children with CIs and N=24 with TH). Mean (SD), range.

Hearing Status	EVT-2 standard score	EVT-2 GSVs	GFTA-2 standard score
CochlearImplant	95.7 (18.71) 46 - 127	120.76 (25.99) 42 - 159	73.61 (19.33) 39 - 107
NormalHearing	116.17 (12) 88 - 134	140.46 (15.46) 117 - 164	90.04 (12.04) 67 - 113

Having established that children with CIs are less sensitive to mispronunciations than their TH peers, we next correlated the children's responses with two different standardized speech-language assessments:
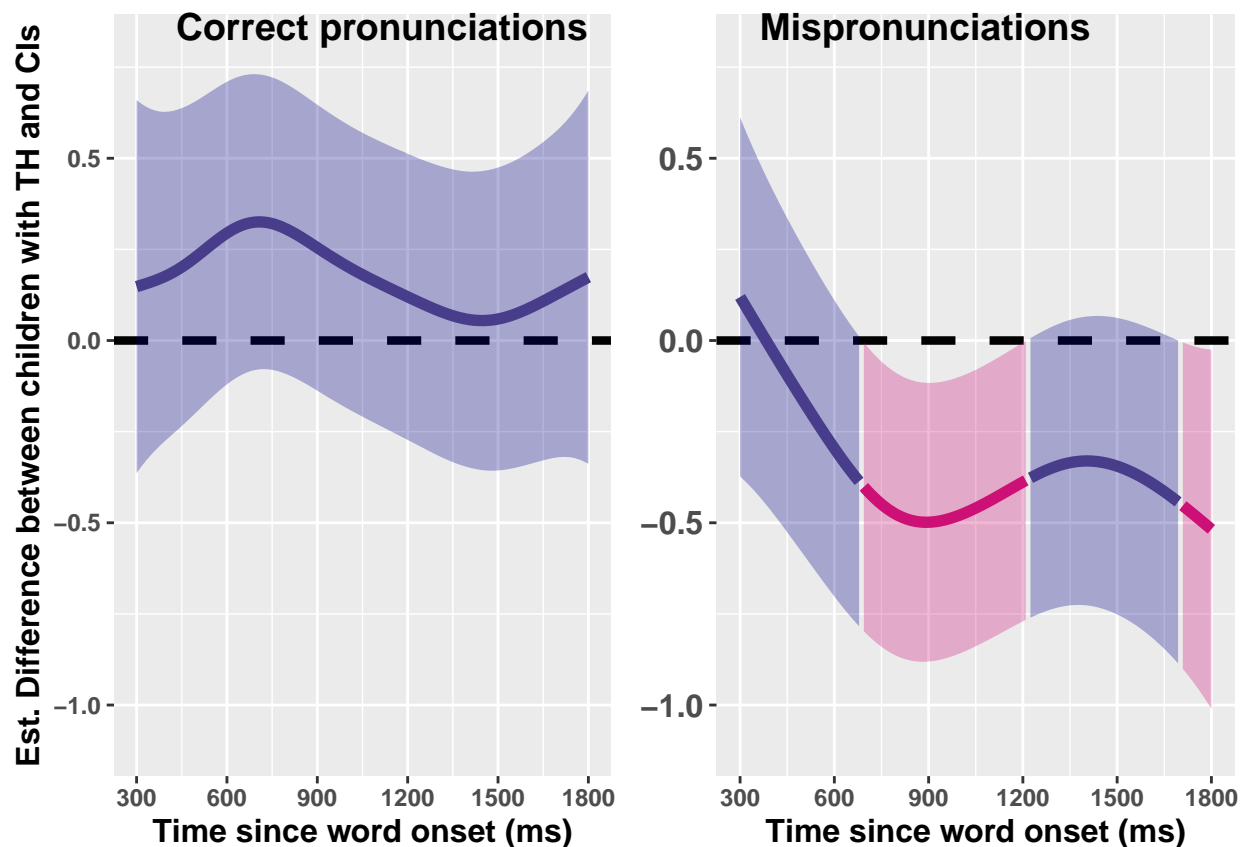
Figure 3: Difference smooths (GAMM predictions) by hearing status for correct pronunciations and mispronunciations. The pink smooth represents the point when the smooth for children with TH differs from children with CIs (i.e., reliable effect of group): there is an effect of group upon mispronunciations (L plot: see also difference between yellow lines in Figure 1), but not correct pronunciations (R plot: see also difference between turquoise lines in Figure 1). Shaded ribbons represent 95% confidence intervals.

Table 3: Demographic information of children who completed standardized speech-language measures, by hearing status (N=33 children with CIs and N=24 with TH). Mean (SD), range.

Group	Girls, Boys	Chronological Age: months	Hearing Age: months	Maternal Education Level
CochlearImplant	21 , 12	50.91 (10.01) 34 - 66	33.33 (14.91) 5 - 56	5.76 (1.09)
NormalHearing	15 , 9	52.71 (10.81) 36 - 66	52.71 (10.81) 36 - 66	5.67 (1.34)

expressive vocabulary size (EVT-2) and spoken phonetic/articulatory accuracy (GFTA-2). Because we took an individual differences approach, we examined the children with CIs and TH separately.

We modeled the effects of vocabulary size and articulation on the children who completed both assessments (N=33 with CIs and N=24 with TH) by using stepwise GAMM fitting. Specifically, we assessed the non-linear interaction between **Time**, **Condition**, and **Vocabulary Score**/**Phonetic Accuracy** to evaluate if children's vocabulary sizes and/or phonetic accuracy predicted their looks to the target over time for the correct- and mis-pronunciation conditions. As before, all models included factor (random) smooths by participant, observation (visit to the lab), and item. Each additionally included a difference smooth of **Time** and **Participant** by **Condition** (*Correct-* versus *Mis-pronunciation*). A baseline model was fit with a parametric term for **Condition** (estimating the average looking probability in each condition), smooth terms for **Time** and **Time** by **Condition**, as well as a non-linear interaction (tensor product) of **Time** and **Child Age** (modeled continuously) by **Condition**. In all models we included the Age by Condition tensor product smooth term because our child-level variables (vocabulary score and phonetic accuracy) are confounded with age, and we wanted to evaluate the potential influence of these speech-language abilities independent of child age.

We fit the three-way interaction of **Time**, **Condition**, and **Vocabulary Score** and **Time**, **Condition**, and **Phonetic Accuracy** using tensor product terms. For the children with TH, neither the vocabulary nor phonetic accuracy term improved upon a baseline model controlling for child age. This result indicates that, for the children with TH, mispronunciation sensitivity—the difference in looks to the target image in correct- versus mis-pronunciation conditions—is not moderated by vocabulary size or phonetic accuracy over and above age effects. For the children with CIs, best model fit included **Phonetic Accuracy**; **Vocabulary Score** did not improve upon model fit. The final model summary for the children with CIs is included in the supplementary materials.

Given the multiple non-linear effects at play, it is necessary to plot the model predictions in order to interpret GAMM outputs, in particular how phonetic accuracy mediates mispronunciation sensitivity for the children with CIs. To facilitate interpretation of the non-linear three-way interaction, the children with CIs were divided into tertiles by vocabulary score and phonetic accuracy. Predictions from the model, by articulatory tertile, are plotted in Figure 4 and raw response curves are plotted in Figure 5. The model predictions demonstrate that children with better articulation scores show larger differences between looks to the target for correct- versus mis-pronunciations (higher overall *y*-intercept value) and that these children show significant differences between correct- and mis-pronunciations slightly earlier in the analysis window (cross-over from purple to pink smooth occurs sooner in the analysis window). Thus, for the children with CIs, phonetic/articulatory accuracy predicts mispronunciation sensitivity, independent of age and language ability.

2 Appendices

3 Supplementary Materials

```
## % latex table generated in R 4.0.2 by xtable 1.8-4 package
## % Tue May 17 12:24:02 2022
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrr}
```
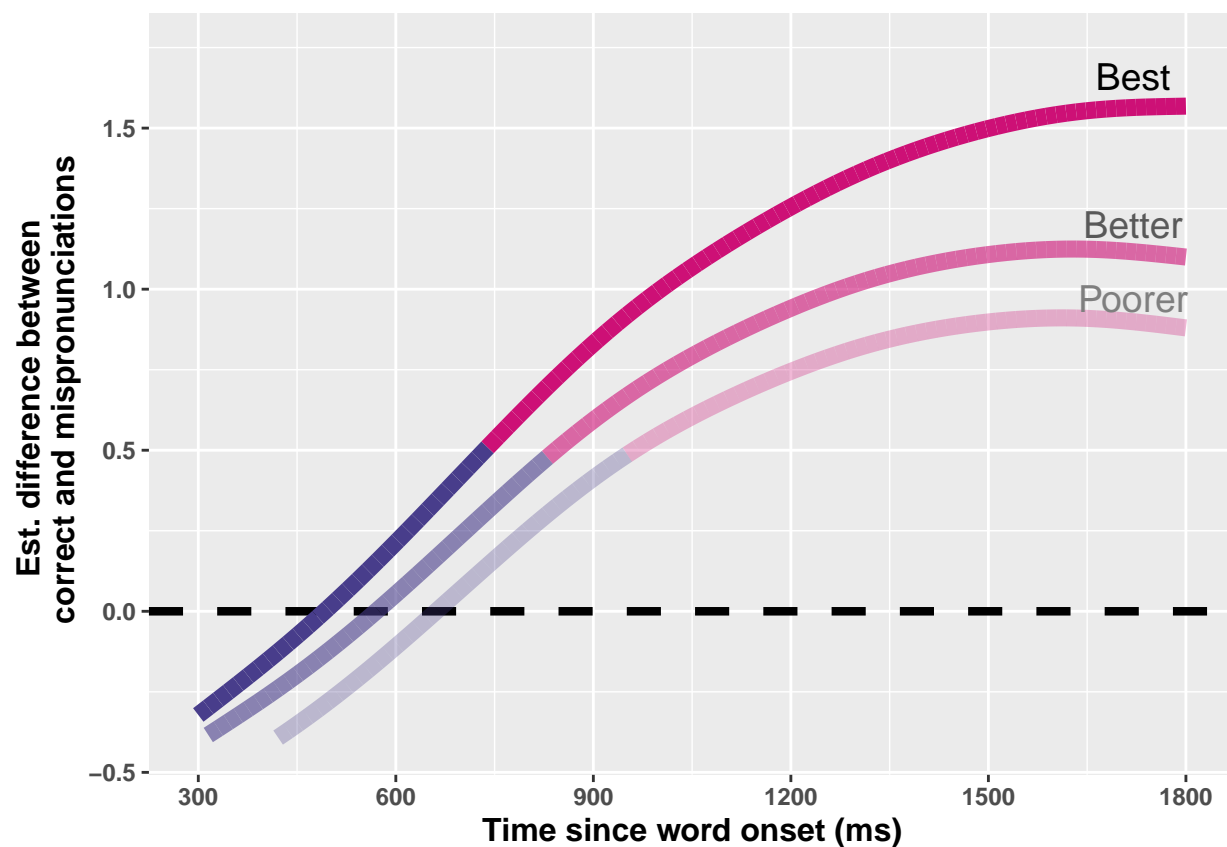
Figure 4: Difference smooths (GAMM predictions) between correct- and mis-pronunciations for children with CIs, by standardized articulation score. Pink smooths represent the point when correct- and mis-pronunciations smooths significantly differ (i.e., reliable effect of condition). Children were divided into tertiles by score with smooths representing the median score for children with poorer (median score=57), better (72), and best (96) articulation scores.
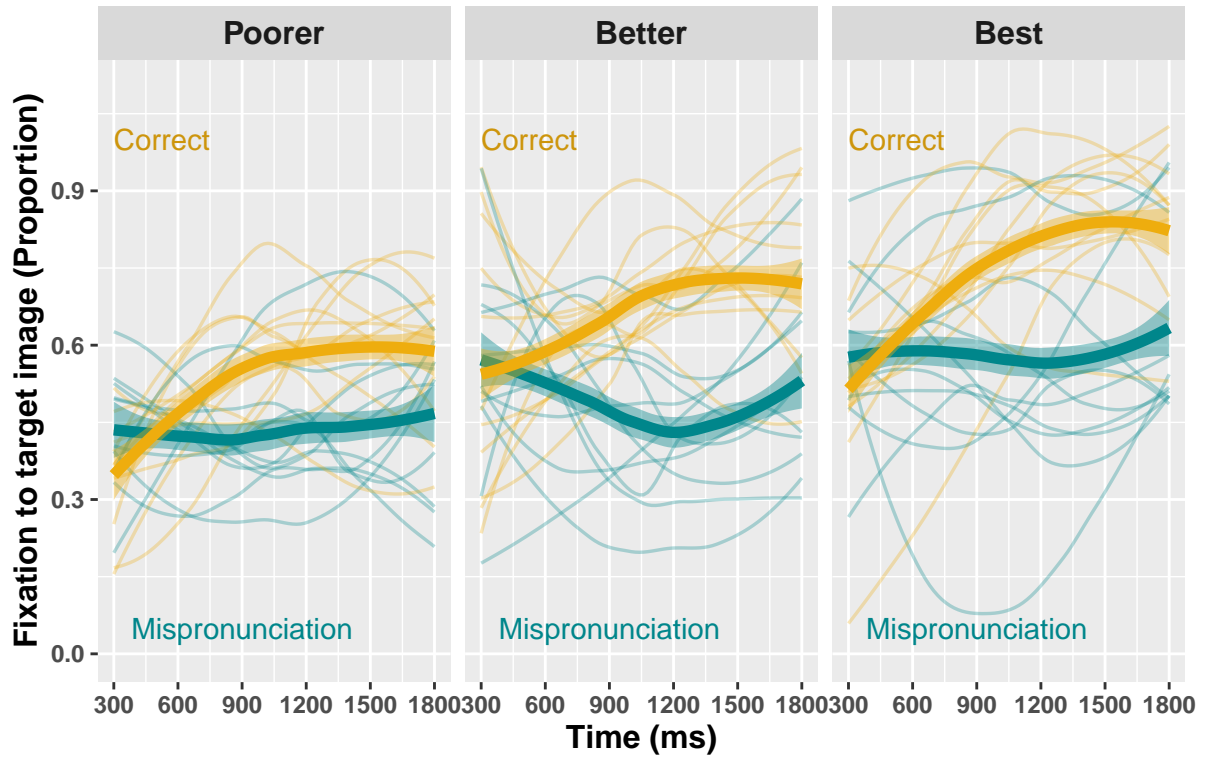
Figure 5: Raw response trajectories for proportion of looks to familiar object for children with CIs, by word condition and standardized articulation score. Children were divided into tertiles by score: poorer (median score=57), better (72), and best (96) articulation scores.

Table 4: Audiological information from the N=25 unique children with cochlear implants studied.

Participant	Matched to child with TH?	Chronological age	Age at hearing loss	Age at activation	Hearing age	E
300E	Y	57	0	13	44	G
302E	Y	37	0	13	24	U
303E	Y	65	6	13	52	U
304E	Y	48	0	12	36	G
305E	Y	44	0	22	22	U
306E	Y	49	0	8	41	U
307E	Y	44	0	15	29	G
309E	Y	59	0.5	7	52	G
311E	Y	62	9	13	49	U
314E	Y	38	10	17	21	U
608L	Y	55	0.5	9	46	C
665L	Y	40	0	12	28	G
801E	Y	39	1.5	15	24	U
804E	Y	56	0	7	49	G
809E	Y	64	6	8	56	M
301E	N	53	0	45	8	U
308E	N	37	0	13	24	G
310E	N	52	unknown	23	29	G
312E	N	57	0	24	33	G
679L	N	58	0	29	29	G
800E	N	65	30	37	28	G
803E	N	41	0	34	7	U
806E	N	42	14	34	8	G
807E	N	51	10	22	29	M
808E	N	37	0	6	31	G

```
##    \hline
## A. parametric coefficients & Estimate & Std. Error & t-value & p-value \\
##    Intercept (Cochlear Implant: Mispronunciation) & 0.2560 & 0.1922 & 1.3325 & 0.1827 \\
##    Typical Hearing & -0.3070 & 0.1918 & -1.6007 & 0.1095 \\
##    \hline
## B. smooth terms & edf & Ref.df & F-value & p-value \\
##    s(Time) & 1.0826 & 1.5820 & 6.3116 & 0.0212 \\
##    s(Time,Cochlear Implant) & 1.0001 & 1.0002 & 0.5928 & 0.4414 \\
##    s(Time,Typical Hearing) & 1.0039 & 1.0072 & 5.5500 & 0.0182 \\
##    s(Time,Cochlear Implant Correct - Cochlear Implant Mispronunciation) & 6.0112 & 7.2081 & 8.2213 & 
##    s(Time,Typical Hearing Difference - Cochlear Implant Difference) & 2.0002 & 2.0004 & 5.3745 & 0.004
##    s(Time,Child) & 52.1398 & 340.0000 & 0.4644 & $<$ 0.0001 \\
##    s(Time,Item) & 25.7681 & 106.0000 & 0.7128 & $<$ 0.0001 \\
##    s(Time,Observation) & 0.0014 & 26.0000 & 0.0003 & $<$ 0.0001 \\
##    \hline
## \end{tabular}
## \caption{Predicting the degree of mispronunciation sensitivity by hearing status.}
## \label{tab.gam}
## \end{table}

## % latex table generated in R 4.0.2 by xtable 1.8-4 package
## % Tue May 17 12:24:13 2022
```

```
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrr}
##    \hline
## A. parametric coefficients & Estimate & Std. Error & t-value & p-value \\
##   Intercept (Mispronunciation) & 0.0835 & 0.1704 & 0.4900 & 0.6241 \\
##   Correct & 0.7110 & 0.1915 & 3.7131 & 0.0002 \\
##    \hline
## B. smooth terms & edf & Ref.df & F-value & p-value \\
##   s(Time) & 0.0006 & 9.0000 & 0.0001 & 0.0413 \\
##   s(Time,Correct) & 3.1503 & 9.0000 & 2.3435 & $<$ 0.0001 \\
##   s(Age) & 0.6289 & 9.0000 & 0.1893 & 0.0081 \\
##   te(Time,Age; Correct) & 3.1548 & 24.0000 & 0.3734 & 0.0049 \\
##   te(Time,Phonetic Accuracy; Correct) & 3.0315 & 24.0000 & 0.4423 & 0.0013 \\
##   s(Time,Child) & 48.6931 & 296.0000 & 0.2923 & $<$ 0.0001 \\
##   s(Time,Item) & 12.9259 & 106.0000 & 0.2786 & 0.0002 \\
##   s(Time,Observation) & 0.5166 & 26.0000 & 0.0299 & 0.0223 \\
##   s(Time,Child; Correct) & 0.0027 & 296.0000 & 0.0255 & $<$ 0.0001 \\
##    \hline
## \end{tabular}
## \caption{Predicting the degree of mispronunciation sensitivity by phonetic accuracy score.}
## \label{tab.gam}
## \end{table}
```