Supplementary Materials

Appendix A: Annotator 2 Results

The main text presents model correlations and results compared against Annotator 1. Here we present analogous results comparing the models against Annotator 2.

1 DDK Speech Rate Correlations

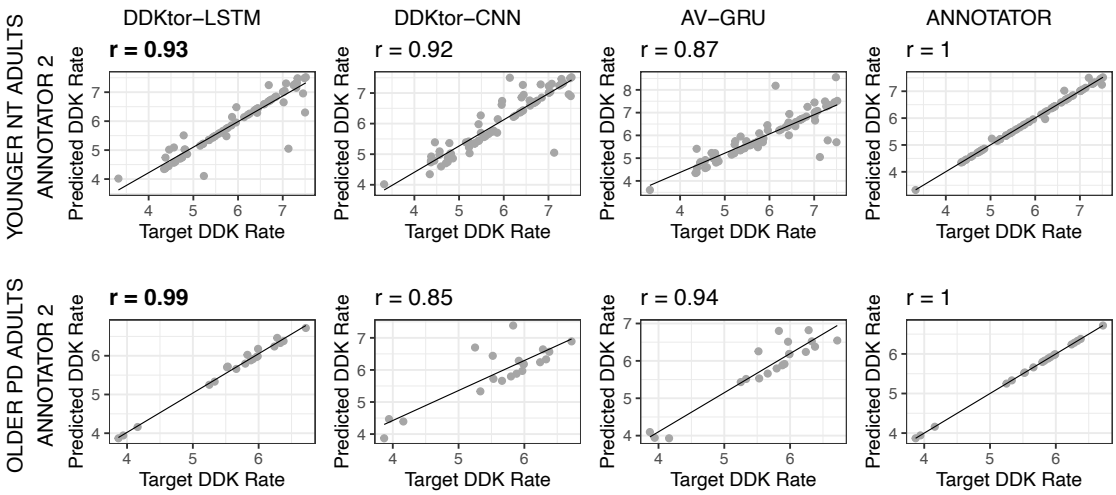

Fig. S1. Correlations between model and Annotator 2 DDK rates for the Younger NT Adults (top) and Older PD Adults (bottom) test sets. Each point represents one participant's set of productions of a particular type (i.e., pa, ta, ka, or pataka). Bold indicates the best performing model within each row, or test set.

2 F1-scores for VOTs and Vowels

Table S1. F1-scores for VOT and Vowel by dataset with Annotator 2 annotations treated as the gold standard.

	Younger NT Adults Test		Older PD Adults Test	
	VOT F1-scores	Vowel F1-scores	VOT F1-scores	Vowel F1-scores
DDKtor-LSTM	**97.7**	**98.4**	**99.3**	**99.8**
DDKtor-CNN	95.8	95.6	97.4	96.4
AV-GRU	94.4	-	95.0	-

3 Segment duration correlations

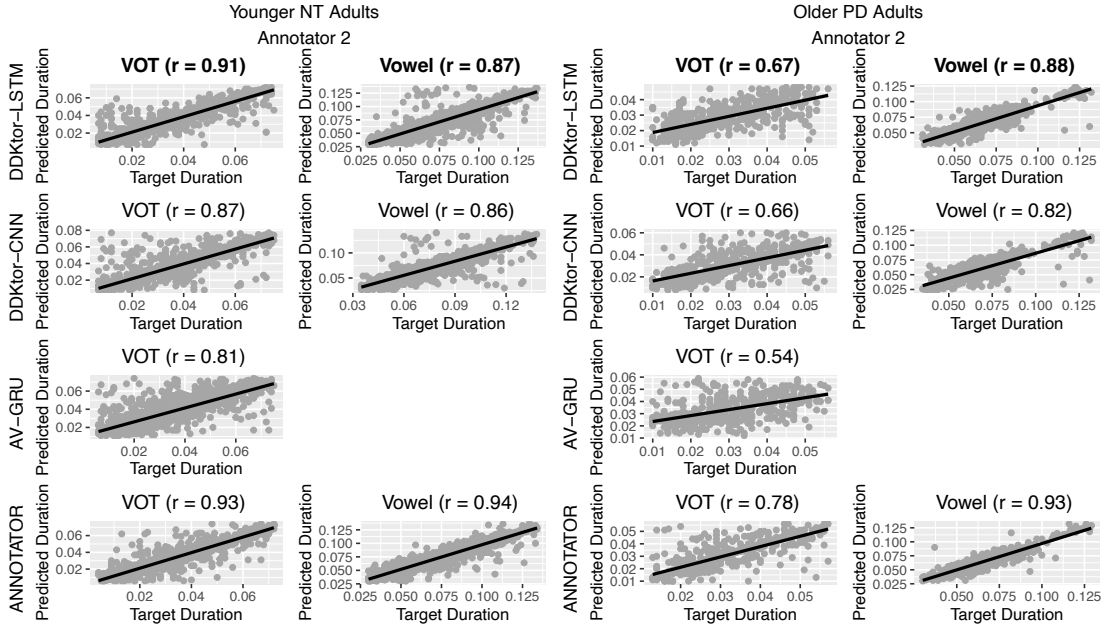

Fig. S2. Correlations between model and Annotator 2 speech sound durations for the (left) Younger NT Adults test set and the (right) Older Adults with PD test set. Each point represents the VOT or vowel of a syllable. The DDKtor-LSTM, DDKtor-CNN, and annotator rows include both VOT and vowel durations, but the AV-GRU row only includes VOTs, as the AV-GRU model does not predict vowels. Bolded values represent the best-performing models within each column (VOT or Vowel in each dataset).

4 Mean absolute deviations in boundaries

Table S2. Mean absolute deviation in boundary offsets (milliseconds) from Annotator 2 by test set.

	Younger NT Adults Test			Older PD Adults Test		
	VOT Onset	VOT Offset/Vowel Onset	Vowel Offset	VOT Onset	VOT Offset/Vowel Onset	Vowel Offset
DDKtor-LSTM	**1.81**	**2.53**	**5.79**	2.93	5.69	**3.17**
DDKtor-CNN	2.05	2.59	6.92	**2.92**	**5.63**	4.94
AV-GRU	3.82	3.98	-	5.99	5.74	-
Annotators	1.10	2.30	2.62	1.95	2.99	2.04

Appendix B: Results on Younger NT Adults Validation Set

The main text reports results from the test sets (the Younger NT Adults and Older PD Adults test sets). Here, we report results from the validation set, which came from a different subset of the Younger NT Adults dataset. Results are presented against both Annotator 1 and Annotator 2.

1 DDK speech rate correlations

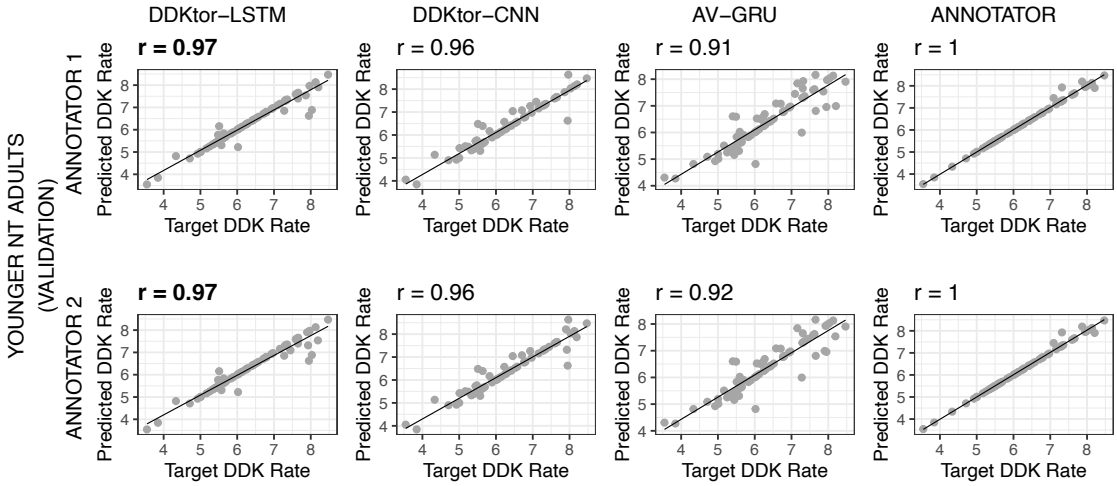

Fig. S3. Correlations between model and annotator DDK rates for the Younger NT Adults validation set (top: Annotator 1; bottom: Annotator 2). Each point represents one participant's set of productions of a particular type (i.e., pa, ta, ka, or pataka). Bold indicates the best performing model within each row, or annotator.

2 F1-scores for VOTs and Vowels

Table S3. F1-scores for VOT and Vowel by Annotator on the Younger NT Adults validation set

	Younger NT Adults Validation (Ann. 1)		Younger NT Adults Validation (Ann. 2)	
	VOT F1-scores	Vowel F1-scores	VOT F1-scores	Vowel F1-scores
DDKtor-LSTM	**97.6**	**98.1**	**97.5**	**98.0**
DDKtor-CNN	97.0	97.4	97.1	97.4
AV-GRU	95.2	-	95.2	-

3 Segment duration correlations

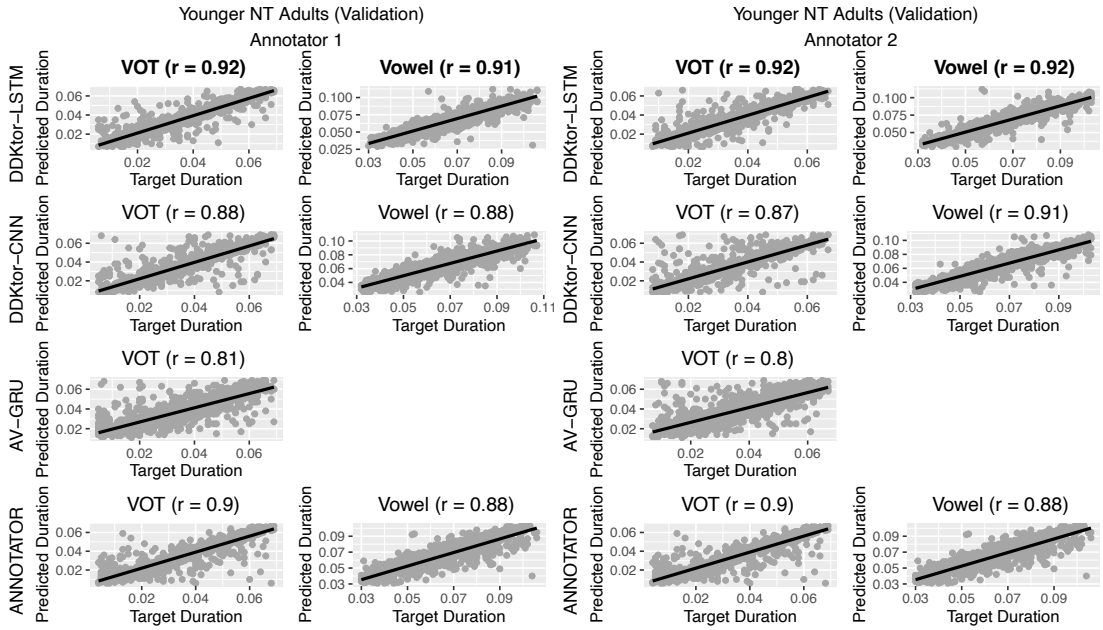

Fig. S4. Correlations between (left) model and Annotator 1 durations and (right) model and Annotator 2 durations for the Younger NT Adults validation set. Each point represents the VOT or vowel of a syllable. The DDKtor-LSTM, DDKtor-CNN, and annotator rows include both VOT and vowel durations, but the AV-GRU row only includes VOTs, as the AV-GRU model does not predict vowels. Bolded values represent the best-performing models within each column (VOT or Vowel against each annotator).

4 Mean absolute deviations in boundaries

Table S4. Mean absolute deviation in boundary offsets (milliseconds) against Annotator 1 (left) and Annotator 2 (right) on the Younger NT Adults validation set.

	Younger NT Adults Validation (Ann. 1)			Younger NT Adults Validation (Ann. 2)		
	VOT Onset	VOT Offset/Vowel Onset	Vowel Offset	VOT Onset	VOT Offset/Vowel Onset	Vowel Offset
DDKtor-LSTM	**1.81**	**2.53**	**5.79**	2.93	5.69	**3.17**
DDKtor-CNN	2.05	2.59	6.92	**2.92**	**5.63**	4.94
AV-GRU	3.82	3.98	-	5.99	5.74	-
Annotators	2.45	1.56	4.70	2.45	2.99	2.04