Acoustic Changes in Speech Prosody Produced by Children with Autism after Robot-assisted Speech Training

Sarah Si Chen[1,7], Bruce Xiao Wang[2,7], Yitian Hong[1], Fang Zhou[1], Angel Chan[1], Tempo Tang[1], Bin Li[3], Chen Chunyi[4], James Cheung[4], Yan Liu[5], Zhuoming Chen[6]

1 Department of Chinese and Bilingual Studies, Hong Kong Polytechnic University 2 Department of English and Communication, Hong Kong Polytechnic University 3 Department of Linguistics and Translation, City University of Hong Kong 4 Department of Biomedical Engineering, Hong Kong Polytechnic University 5 Department of Computing, Hong Kong Polytechnic University 6 Brain Hospital, The First Hospital of Jinan University 7 Research Institute for Smart Ageing, The Hong Kong Polytechnic University

INTERSPEECH 2024

Introduction

Focus Marking

- A strategy to signal the prominence of information

- Prosodic marking of focus involving acoustic correlates such as fundamental frequency (F0), intensity, duration etc.

Autism Spectrum Disorder (ASD)

- A neurodevelopmental disorder affecting various aspects of life, such as language acquisition and social interaction

- Unable to understand emotion norms and exhibit atypical prosodic patterns compared to typically-developing (TD) counterparts

Robot-assisted Training with Social Interactions

- Essential to language acquisition and may leads to positive behavioural outcome and brain changes

- Using a social robot to control the consistency of such social behavioural information

Method

Participant

- 20 ASD and 23 TD native Hong Kong Cantonese children

- Assessed using the Autism Spectrum Quotient (AQ) and non-verbal Intelligence Quotient (IQ)

Table 1 *Mean and standard deviation of age, IQ and AQ scores.*

	TD	ASD
Age	7.82	9.53
	±1.16	±1.61
IQ	113.61	109.89
	±13.82	±18.91
AQ	58.95	80.83
	±14.64	±28.44

Table 2 *Examples of prompt questions and target sentences in relation to different focus conditions.*

Focus Condition	Prompt Questions	Target Sentence
a. Broad	What do you see in the picture?	**Mr. Cheung is flying the plane**.
b. 1) Initial narrow	Who is flying the plane?	**Mr. Cheung** is flying the plane.
b. 2) Medial narrow	What is Mr. Cheung doing to the plane?	Mr. Cheung is **flying** the plane.
b. 3) Final narrow	What is Mr. Cheung flying?	Mr. Cheung is flying the **plane**.
c. 1) Initial contrastive	Miss Chan is flying the plane?	**Mr. Cheung** is flying the plane.
c. 2) Medial contrastive	Mr. Cheung is buying the plane?	Mr. Cheung is **flying** the plane.
c. 3) Final contrastive	Mr. Cheung is driving the bus?	Mr. Cheung is **flying** the plane.

Production

- Fifteen target sentences, each with seven prompt questions as in Table 2

- Participants were instructed to answer questions from the trained speech therapist using the target sentences.

- Recorded once for TD and twice (before and after the training) for ASD

Data Extraction & Statistical Analyses

- Word duration, fundamental frequency (F0) and intensity were extracted

- Linear Mixed-Effects Models (LMM) were fitted to acoustic data and a Likelihood Ratio (LR) test was conducted. Word duration, mean F0 and mean intensity were assigned as the response variables in each model, while groups (ASD Robot Speech Training and TD) and focus conditions (i.e., narrow/contrastive pre-, on- and post-focus) were assigned as explanatory variables.

Training

- Six out of fifteen target sentences were used for training

Figure 1: *Example of setup used for Phase 1.*

Phase 1 – congruent pairs

1) Ask the robot a question based on the picture

2) Identify the focus condition based on the robot's response

3) Click on the corresponding button

Figure 2: *Example of setup used for Phase 2.*

Phase 2 – congruent & incongruent pairs

1) Ask the robot a question based on the picture

2) Determine whether the focus type of the robot's response is correct

3) Input the judgment and view the feedback

Results

Duration

- For TD, the main effect of focus condition had a significant effect on mean duration ($\chi 2$= 335.85; df = 6; p < 2.2e-16).

- For ASD, focus condition had a significant effect on mean duration in both pre- ($\chi 2$ = 120.96; df = 6; p < 2.2e-16) and post-training ($\chi 2$ = 208.78; df = 6; p < 2.2e-16) production.
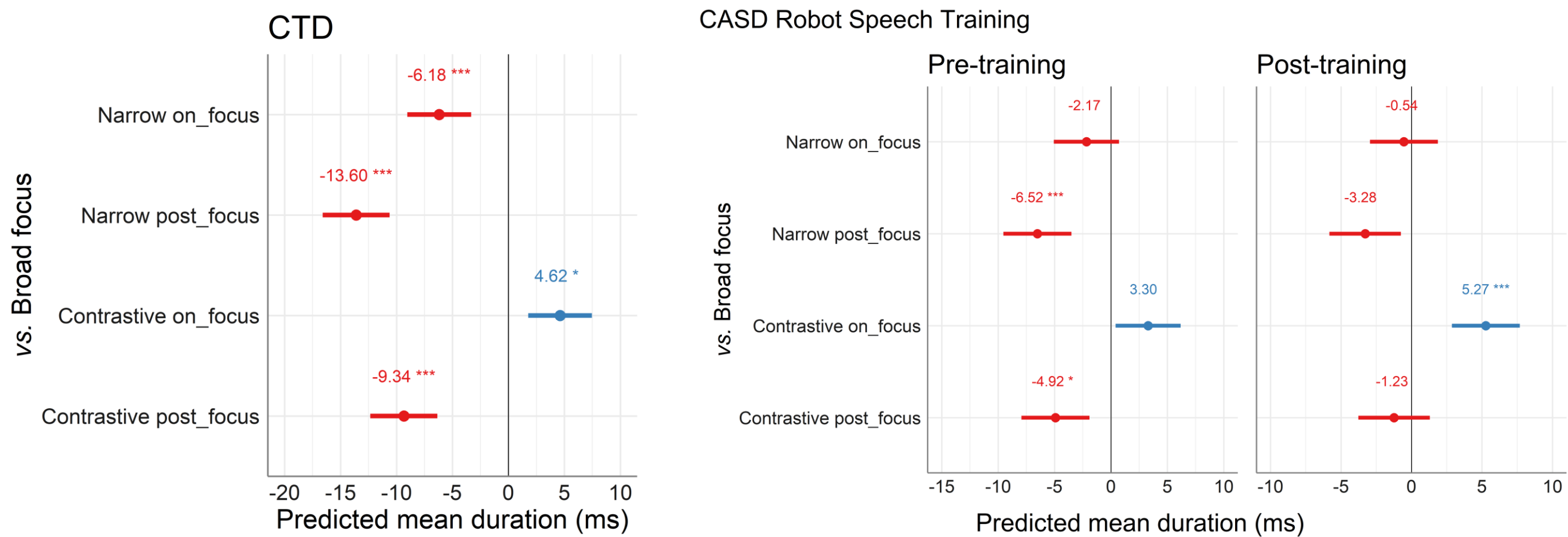

Figure 3: *Difference in predicted mean duration between broad focus and narrow and contrastive on- and post-focus across groups and training sessions.*

Intensity

- For TD, the main effect of focus condition had a significant effect on mean intensity ($\chi 2$ = 1377.8; df = 6; p < 2.2e-16).

- For ASD, focus condition had a significant effect on mean intensity in both pre- ($\chi 2$ = 535.86; df = 6; p < 2.2e-16) and post-training ($\chi 2$ = 387.94; df = 6; p < 2.2e-16) sessions.
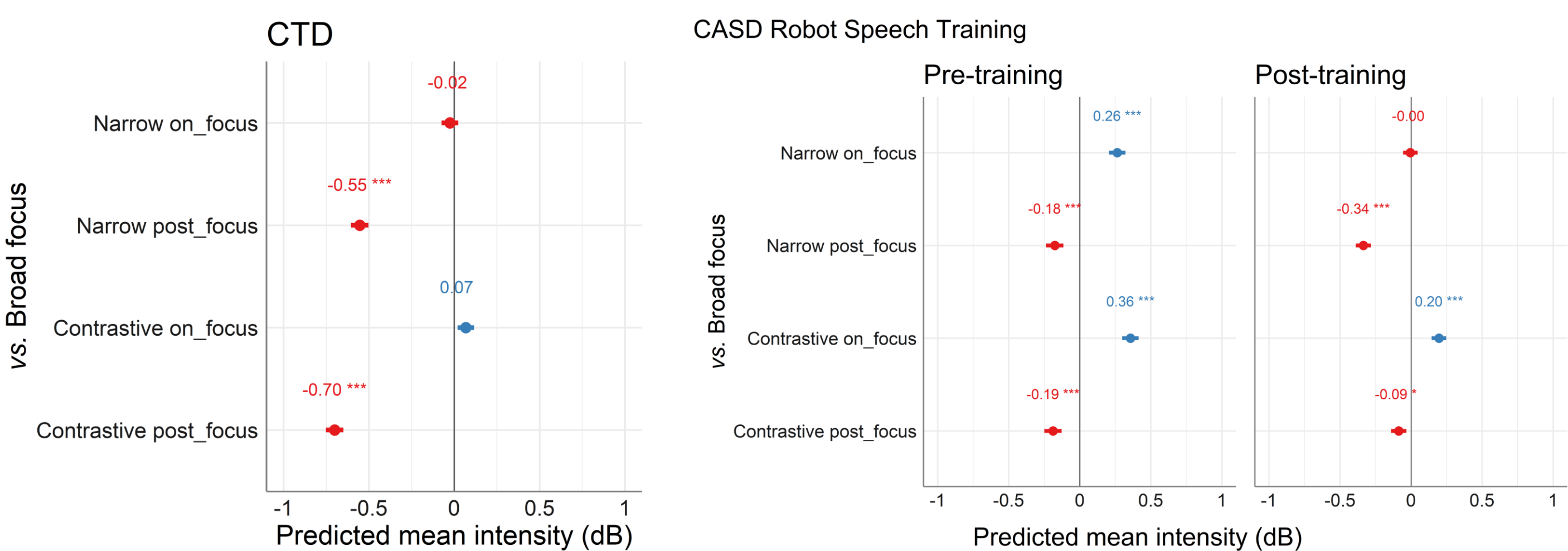

Figure 4: *Difference in predicted mean intensity between broad focus and narrow and contrastive on- and post-focus across groups and training sessions.*

F0

- For TD , the main effect of focus condition had a significant effect on mean f0 ($\chi 2$ = 1558.2; df = 6; p < 2.2e-16).

- For ASD, the main effect of focus condition also had a significant effect on mean F0 in pre- ($\chi 2$ = 126502; df = 6; p < 2.2e-16) and post-training ($\chi 2$ = 703.44; df = 6; p < 2.2e-16) production.
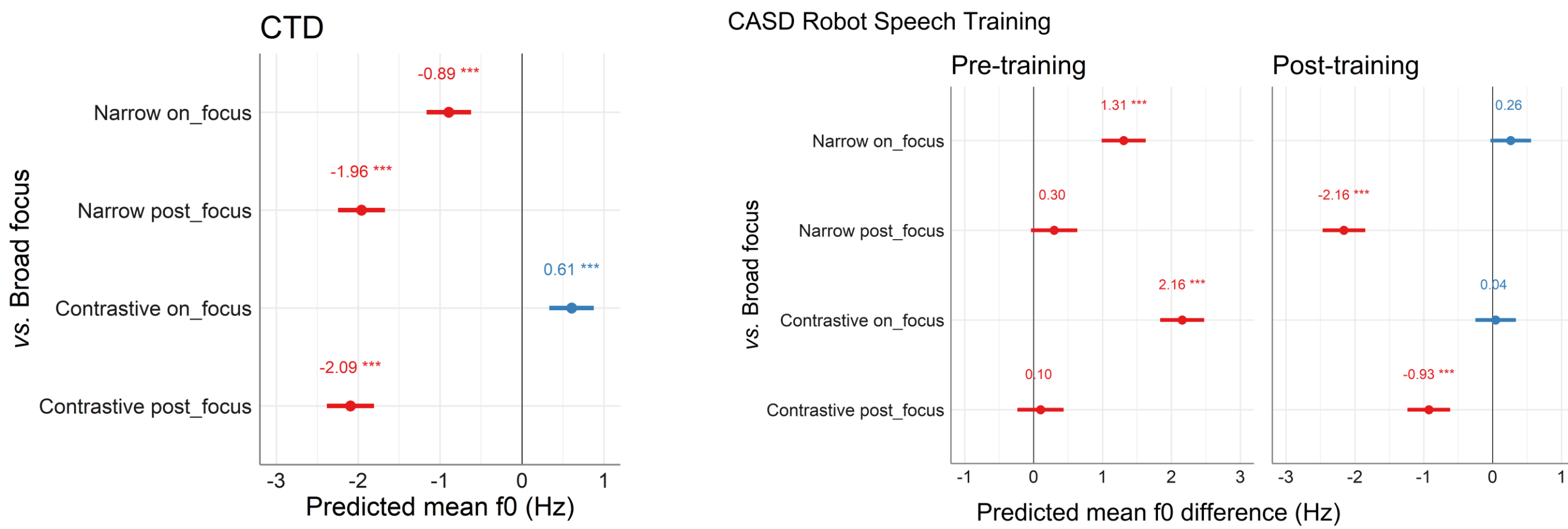

Figure 5: *Difference in predicted mean duration between broad focus and narrow and contrastive on- and post-focus across groups and training sessions.*

Discussion & Conclusion

- As for duration, ASD produced on-focus words with significantly longer duration than the broad focus words under contrastive focus condition which is more similar to that of the TD. However, the the pattern of TD's shorter duration on post-focus words is absent among ASD in the post-training.

- In the use of intensity, ASD produced post-focus words with significantly lower intensity than the broad focus words in both pre- and post-training, which resembles the TD. However, ASD produced on-focus words with significantly higher intensity which is not observed among TD.

- In terms of F0, the ASD produced on-focus words with significantly higher f0 than the broad focus words as TD in pre-training, but not in post-training. Even though, ASD picked up significant F0 lowering with the post-focus words in the post-training like TD.

- ❑ **Overall, it seems that using Robot-assisted Training with Social Interactions to improve the speech prosody of Cantonese-speaking ASD children is effective on certain acoustic cues. Future studies should consider training designs from a global perspective to enhance their ability to manage sentence production globally.**