

Supplemental Material S2. Supplemental methods and exploring parent and participant end-user experience.

Supplemental Table S-2. Single Case Analysis Metrics

ID	Trend			Mean (SD) Level			Nonoverlap
	BL	TX	Post	BL	TX	Post	TX vs BL
1107	.017	.024	.022	.20 (.08)	.36 (.12)	.68 (.03)	72/80, $NAP_r = 80$, $W = 8$, $p = .002$
1111	-.007	.000	-.013	.17 (.05)	.35 (.03)	.35 (.04)	98/100, $NAP_r = 96$, $W = 2$, $p < .0001$
1112	.058	-.017	.096	.46 (.13)	.57 (.10)	.58 (.11)	38/50, $NAP_r = 52$, $W = 12$; $p = .063$
1121	.069	.020	.028	.47 (.17)	.74 (.12)	.87 (.03)	56/60, $NAP_r = 86.6$, $W = 4$; $p = .002$
1130	-.013	-.006	.007	.33 (.08)	.34 (.05)	.47 (.05)	35/70, $NAP_r = 34.5$, $W = 34.5$; $p = .5$

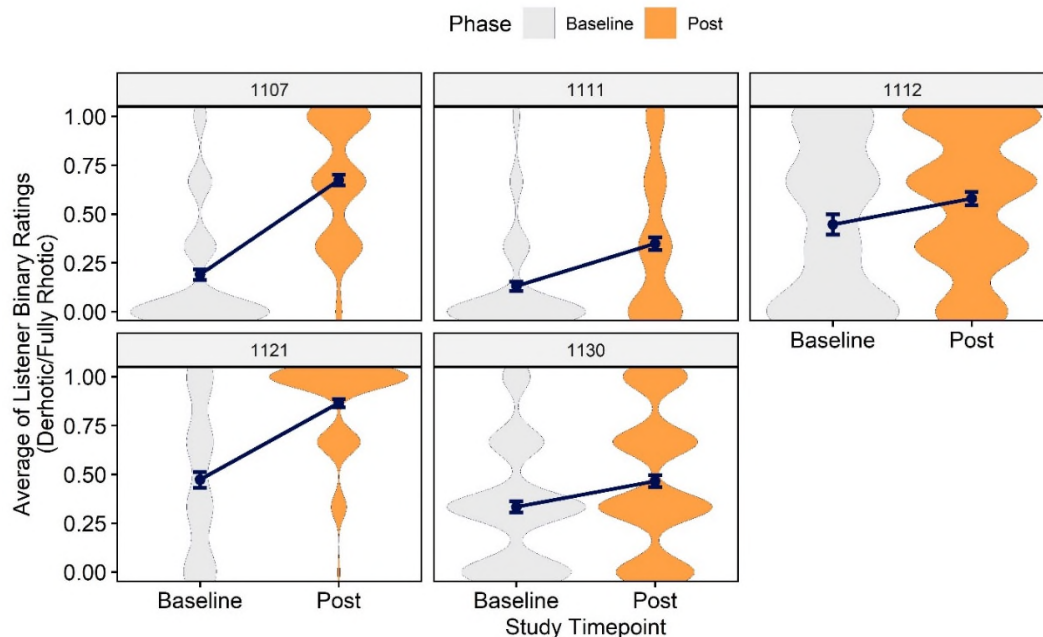
Note. BL = baseline, TX = treatment, Post = post-treatment. Non-overlap reported as number of BL-TX pairs where treatment point is higher versus all possible BL-TX pairs, rescaled nonoverlap of all pairs value ($NAP_{rescaled}$, ranging 0–100), Wilcoxon signed rank test statistic (W), p value.

Feedback delay due to PERCEPT processing time

The average time for PERCEPT to complete a prediction was 3.6 seconds. 99.3% of files returned a PERCEPT prediction in less than 10 seconds. These estimates, however, only include the time the PERCEPT Engine was processing. Extrapolating from the average file size (462,788 Bytes), the average round trip time of the data from the user's computer to the PERCEPT Engine and back would fall between 4.72 seconds per trial on a dial-up connection and 3.61 seconds per trial on an ethernet connection.

The average duration of KR text to speech prompts was 1.8 seconds, while the average duration of KP text to speech prompts was 4.39 seconds.

Figure S-3: Amount and Distribution of Pre–Post Change for Mean Listener Rating of Untreated Words



Note. This figure illustrates the distribution of data within the no-treatment baseline and post phases shown in Figure 7. All slopes were significantly different than zero; mean level change exceeded the clinically significant threshold for two participants, 1107 ($d_2 = 1.6$) and 1121 ($d_2 = 1.3$). Error bars represent the 95% confidence interval around the mean. Table 4 in the main paper illustrates the linguistic complexity of the words represented by these data points.

Exploring parent and participant end-user experience

Parents and participants: Is there anything you would like us to know about how the study may be impacting [the participant/you], positively or negatively?

Three parents volunteered that they perceived functional improvement in their child's /ɹ/ sound during study, citing increased confidence, self-monitoring, clarity of speech, and noticeable carry over; one parent did not answer this question. Two participants volunteered that they perceived improvement in their /ɹ/ production, citing increased intelligibility when speaking with their parents, and that the study helped "a lot". Three participants indicated there was nothing they wished to share in response to this question.

Parents: What do you think would be the right balance of clinician-led sessions and computer-led sessions for children with speech sound disorders?

Three of the four parents (plus the adult participant himself) indicated that computer-led sessions had some place in treatment for children with speech sound disorders (multiple choice selection: *Sometimes a Person/Sometimes a Computer*). These individuals identified that computerized components would be especially useful for practice between sessions with a clinician. The adult participant elaborated: *if the technology improved[,] I would much prefer speech lessons from a computer because it would allow me to practice anytime without having to schedule in advance and would allow me to spend as much time per week as I wanted*

practicing. One parent, of participant 1111, indicated they would rather have speech lessons from a person, expressing preference for human connection in each speech therapy interaction.

Parents: How do you think the use of artificial intelligence in speech therapy, generally, would impact daily life for children and young adults with speech sound disorders?

Three of the four parents, plus the adult participant, indicated they foresaw a neutral-to-positive impact on daily life for children and young adults with speech sound disorders. Responses ranged from 50 (*neutral*) to 99 (*make daily life better*) on the visual analogue scale, with an average response of 83.5. When asked to elaborate, themes that arose included accessing speech therapy without stigma and/or in a lower-pressure environment, increasing access to services, the benefit of repeated/home practice, and only using computerized treatment with those it is a good fit for. One parent omitted an answer to this question.

Parents: Is there anything else we should know about your thoughts on computerized speech therapy?

One parent provided a response, indicating that it was "hugely helpful" to have the clinician review PERCEPT's predictions from the previous session at the start of the following session. The parent felt that this alleviated their child's frustration with the computer letting them know their production was not quite right.

Participants: Would you rather have speech lessons from a person, a computer or sometimes a person/sometimes a computer?

Three of the child participants indicated they would rather have a balance of person-led and computer-led sessions (multiple choice selection: *Sometimes a Person/Sometimes a Computer*). As one of our participants put it: *the person makes you good, and the computer tells you how good you are*. One child participant indicated a full preference for person-led sessions (multiple choice selection: *Person*), explaining *being in-person is a nice experience for everyone, even if it's a bit of a drive*. As a note, this participant was homeschooled.

Participants: How often was the speech app awesome/terrible?

These questions were rated with a Likert scale: never (1), sometimes (2), often (3), always (4). The average participant response to "the app was awesome" fell between "sometimes" and "often" (2.6). The average participant response to "the app was terrible" fell between "never" and "sometimes" (1.4).

Participants: What were the three best/three worst things about the website?

A variety of themes were cited as one of the three best things about the website, including: *it was easy to use, nothing was wrong with it, helped me get better at /ɹ/, the computer said the sound and also the prompt, I liked my emoji and the drawings of the clinician, it was accurate, it adjusts to the learner, and it offers multiple difficulty levels*. Some of the same themes were repeated for the three worst things, including: *it was sometimes slow; I didn't always think it was accurate; it would always tell me I was wrong; the [prosody prompts] were confusing; I wasn't able to use it on my own outside of the study, view my progress, or choose which sounds I wanted to work on; it told me I was correct too frequently, and it only said "correct" or "not quite", never "in-between"*.

Exploring parent and participant end-user experience: Discussion

Exploration of survey data indicates that parent and participants largely feel that computerized intervention can positively impact service delivery for children with speech sound disorders, most frequently mentioning hybrid clinician-AI models in which computerized systems facilitate

at-home practice. Future survey research on this topic, however, might expand stakeholder polling beyond a self-selected group of people who would seek to enroll in a research study with computerized speech lessons. Even so, comparison of responses herein indicates that participants have differing views on ChainingAI, which supports clinical intuition that automated treatment may not meet everyone's personal preferences or clinical profile. Future studies can elucidate the social, emotional, and motivational preferences that make a learner a candidate for computerized treatment, and our ongoing work will adapt the ChainingAI interface into an interactive game for participants.

Exploring parent and participant end-user experience: Conclusion

Exploration of survey data indicated that parents and participants largely felt that computerized intervention could positively impact service delivery for children with speech sound disorders, most frequently endorsing hybrid models in which computerized systems facilitate at-home practice.