Level Up Your Critical Thinking around AI for Speech Sound Disorders: Part 2

Nina Benway
PhD CCC-SLP
University of Maryland, College Park

Follow.the.Speech.Production.Lab.on.Instagram.to.find.updates.on.this.work.® speechproductionlabsyr

1. What does systematic review say about AI speech therapy?

● The systematic review by Deka et al (2024) examined AI-based speech therapy tools using articles published between January 2007 and February 2022 that were written in English and centered on AI-based therapy for Speech Sound Disorders (SSD). Studies that focused only on assessment or did not use AI were excluded. Key findings include:

○ Types of SSD Addressed: Articulation disorders were most frequently addressed (9 out of 24 studies), followed by dysarthria (4 studies) and childhood apraxia of speech (4 studies).

○ AI Techniques Used: Automatic speech recognition, classification for tasks like lexical stress evaluation and pronunciation assessment, and the "Wizard of Oz" technique were used

○ Level of Autonomy: Only 4 out of 24 studies had limited autonomy, with SLPs involved in tasks like selecting words, reviewing recordings, and monitoring progress. The majority (13 out of 24) were fully automated.

○ Modes of Intervention: Computer-based interventions were most common, with some using mobile platforms and robots. Gamification was frequently used to increase engagement.

○ Effectiveness: Only 5 out of 24 studies directly compared AI speech therapy to SLP therapy. Of these, two found no significant difference, one noted articulation improvement after using the AI tool, another showed improvement in phonological and morphosyntactic skills with robot-assisted therapy compared to SLP treatment, and the last demonstrated that AI feedback was as effective as at-home practice with a parent.

2. What is so hard about AI speech therapy?

● Data Scarcity: AI speech analysis systems need large amounts of speech data for training, but most publicly available data comes from adults. Children's speech is more variable, reflects developmental and clinical processes, and is harder to collect.

● Limited Technical Description of Child Speech Systems: The research literature often lacks detailed descriptions of the technology, making it hard to evaluate and compare systems.

● Unknown Therapeutic Benefit: The methodological limitations of existing research make it difficult to determine the true benefit of many AI speech therapy tools. Previous reviews found that the overall quality of research is moderate to low, and many commercially available apps lack evidence of therapeutic potential.

3. The Say Bananas! AI Clinician (say66.com, Kirrie Ballard and Beena Ahmed)

● Say Bananas is a "brightly themed platformer game with interwoven speech exercises" designed to engage children. The evidence for Say Bananas comes from a study with 10 Australian English speakers aged 5-12, most with idiopathic SSD. None were receiving outside treatment during the study.

● The study design involved SLPs selecting 1-2 stimulable sounds, caregivers conducting pre-practice to remind children of correct production and game feedback, and children playing until they completed the exercises (100 production attempts).

● The game provided "knowledge of results" feedback ("Good job!" vs. "Not quite"). Say Bananas uses template matching for automated speech analysis. Before gameplay, users record "correct" and "incorrect" templates for each word.

● During gameplay, users collect stars and record words. Recordings are compared to the templates, and the system determines if the recording is more similar to the correct or incorrect template, providing corresponding feedback

Key Findings:

○ Engagement: Most users maintained their average play time and reported trying "very hard," with most playing after treatment ended.

○ Pronunciation Improvement: Participants improved similarly in both conditions (template matching and caregiver feedback) on practiced words recorded immediately before and after a treatment phase.

○ Feedback Accuracy: The automated speech analysis algorithm agreed more with expert clinicians than the parent feedback did, but fewer than half of the recordings could be analyzed for accuracy due to audio quality issues.

4. The ChainingAI Clinician (Syracuse Speech Production Lab, Nina Benway and Jon Preston)

● ChainingAI is designed to simulate clinician-led Speech Motor Chaining therapy but in a home setting. It uses the PERCEPT AI system as its "ears". The evidence for ChainingAI comes from a single-case experimental study with 5 participants between the ages of 10 and 19. These participants, 4 male and 1 female, were all speakers of American English and had residual /r/ errors.

● The study design involved one "Orientation to /ɹ/" session led by a human clinician, and then nine sessions which started with 10 minutes of prepractice followed by up to 30 minutes of practice with the ChainingAI clinician.

● The ChainingAI platform provided two types of feedback: knowledge of results feedback ("Correct"/"Not Quite") and approximated knowledge of performance feedback, which was delivered through text, audio, and animation. ChainingAI uses a deep neural network binary classifier for automated speech analysis. The classifier was personalized for each speaker based on word lists the participants read during the assessment sessions.

● During the session, Speech Motor Chaining adapted the difficulty of practice according to the principles of motor learning based on the recordings made during the session. These recordings were predicted to be

either "correct" or "incorrect" based on the formant patterns for the /ɹ/ in each recording. The feedback provided was based on the "correct' or "incorrect" predictions. Note that ChainingAI now includes a treasure hunting game, but that was not part of the platform that was tested in this study.

Key findings:

○ Pronunciation Improvement in Practiced Words: ChainingAI resulted in immediate, statistically significant improvement for the practiced chain.

○ Pronunciation Improvement in Unpracticed Words: The entire treatment package resulted in improvement in unpracticed words for an average of 30% improvement in all participants.

○ Feedback Accuracy: The level of agreement between PERCEPT and human clinicians was largely in the same range as any two clinicians, but audio quality issues prevented the analysis of ~8% of audio recordings.

5. Notes from the session's group discussion on AI Speech Therapy: