Volume 8, Issue 05, May 2023 ISSN (online): 2456-0006

International Journal of Science Technology Management and Research

Available online at: www.ijstmr.com

Corrected By Code: A Quasi Experimental Study Of Ai Pronunciation Feedback On L2 Speakers' Confidence And Accent Perception (Pre Genai Baseline)

Pallavi Sharad Joshi
Department of Science and Humanities
Government Polytechnic, Nashik

Dr. Sharayu Omprakash Sonawane
Department of English and Languages
School of Engineering and Technology, Sandip University, Nashik

Abstract: **The integration of artificial intelligence (AI) into language learning has expanded rapidly, yet the affective consequences of real-time algorithmic pronunciation feedback remain largely unquantified. This quasi-experimental study investigates the impact of AI-powered real-time pronunciation feedback on oral communication confidence and accent perception among 128 L2 English speakers. Participants were assigned to either an experimental group (4-week intervention using a non-generative AI pronunciation app with immediate corrective feedback) or a control group (delayed human feedback on recorded speech). Pre- and post-intervention assessments measured oral communication confidence (OCCI), accent perception identity (APIS), and willingness to communicate (WTC). Results of repeated-measures ANOVA revealed a significant decrease in oral communication confidence for the experimental group (F*(1,126)=12.47, *p*<.001, η²=.09) and a significant increase in negative accent self-perception (*F*(1,126)=8.93, *p*=.003, η²=.07). In contrast, the control group showed no significant changes. Frequency of algorithmic corrections was negatively correlated with post-intervention confidence scores (*r*=−.42, *p*=.002). These findings suggest that real-time AI pronunciation feedback, while potentially improving accuracy, may inadvertently undermine learners' affective states and self-concept. The study provides a pre-generative AI baseline essential for future longitudinal research.**

Keywords: AI pronunciation feedback; oral communication confidence; accent perception; L2 English; quasi experimental study; computer assisted language learning

I. Introduction

Artificial intelligence has become deeply embedded in language education, giving rise to applications that deliver instantaneous, algorithm-driven feedback on pronunciation, intonation, and fluency. Tools such as Elsa Speak, Duolingo's advanced pronunciation features, and various cloud-based speech recognition systems promise to improve learners' accuracy through real-time corrections. Research has extensively documented the effectiveness of such tools for enhancing pronunciation accuracy [6], [9]. However, the *affective* dimension—how learners perceive themselves and their communicative abilities when constantly evaluated by a machine—remains largely unexplored.

Prior to the widespread availability of generative AI (late 2022), these systems relied on rule-based or statistical pattern matching, offering a type of feedback distinct from conversational AI. Understanding learners' responses to this technology is crucial not only for ethical pedagogical practices but also for establishing a baseline against which the effects of newer generative AI tools can be assessed. This study captures a pivotal moment: the period immediately preceding the public release of ChatGPT (November 2022), which transformed the landscape of AI-mediated communication.

The present research addresses a gap in the literature by quantitatively investigating the impact of real-time AI pronunciation feedback on two affective constructs: oral communication confidence (the belief in one's capacity to communicate effectively in spoken English) and accent perception (how individuals evaluate their own accent and its social acceptability). The following hypotheses were formulated:

- H_1 : L2 English speakers exposed to real-time AI pronunciation feedback will demonstrate a statistically significant reduction in oral communication confidence compared to a control group receiving delayed human feedback.

- H_2 : The experimental group will show a statistically significant increase in negative accent self-perception relative to the control group.

- H_3 : The frequency of algorithmic corrections received during the intervention will be negatively associated with post-intervention oral communication confidence scores.

II. Literature Review

A. AI in Pronunciation Training

Computer-assisted pronunciation training (CAPT) has progressed from simple playback tools to sophisticated systems that utilize automatic speech recognition (ASR) [4]. Early meta-analyses confirmed that ASR-based feedback can improve segmental accuracy [15]. More recent work has examined the effectiveness of gamified AI applications [8] and the role of visual feedback [3]. Nonetheless, most studies have concentrated on objective performance gains rather than learner perceptions.

B. Affective Factors in L2 Oral Communication

Confidence, anxiety, and self-concept are central to L2 speaking performance [7]. Woodrow [16] demonstrated that self-efficacy significantly predicts oral communication outcomes. More recently, Dewaele and MacIntyre [1] emphasized the importance of "positive psychology" variables, such as self-esteem and willingness to communicate (WTC). Despite this, few studies have examined how technology-mediated feedback influences these affective variables.

C. Accent Perception and Identity

Accent serves as a significant marker of identity for L2 speakers [14]. Negative evaluations of one's own accent can lead to communication avoidance [2]. Although some CAPT research mentions learner satisfaction, no studies have systematically measured the effect of algorithmic feedback on accent self-perception using validated scales.

D. Research Gap

The intersection of AI pronunciation feedback and learner affect remains an underexplored area. Existing investigations are often qualitative or limited in scale [10], [12]. A quasi-experimental study that quantifies the effects of real-time, non-generative AI feedback on both confidence and accent perception using validated instruments and appropriate statistical controls has been lacking. This study fills that gap by providing a rigorous baseline from the pre-generative AI era.

III. Methodology

A. Research Design

A quasi-experimental design with pre-test and post-test measures and a control group was employed. Participants were allocated based on course enrolment to minimize disruption, but baseline equivalence was confirmed through statistical tests.

B. Participants

One hundred twenty-eight L2 English speakers, all university students, were recruited from two institutions. Inclusion criteria were: (a) self-reported intermediate or higher English proficiency; (b) no significant prior experience with AI pronunciation applications; (c) not currently enrolled in a pronunciation-focused course. Sample size was determined *a priori* using G*Power [5] for a medium effect size (*f*=.25) with 80% power and α=.05, yielding a required *N* of 128. Table I presents participant demographics.

TABLE I
PARTICIPANT DEMOGRAPHICS

Variable	Experimental (*n*=64)	Control (*n*=64)
Age (M, SD)	22.4 (3.1)	22.8 (3.3)
Female (%)	56.3	53.1
L1 background		
– Mandarin	31	29
– Spanish	18	20
– Arabic	9	8
– Other	6	7
Years of English study	8.2 (2.5)	8.0 (2.7)

No significant differences were observed between groups at baseline for any demographic or pre-test measure (all *p* > .05).

C. Instruments

1. **Oral Communication Confidence Index (OCCI)** – A 12-item Likert scale (1=strongly disagree, 5=strongly agree) developed for this study, adapted from the Self-Perceived Communication Competence Scale [13]. Items assess confidence in one-on-one, group, and public speaking contexts. Cronbach's α in pilot testing was .89.

2. **Accent Perception Identity Scale (APIS)** – A 10-item scale measuring self-perception of accent acceptability, pride, and anxiety. A sample item is: "My accent makes me feel self-conscious when speaking English." Cronbach's α = .87.

3. **Willingness to Communicate (WTC) Scale** – Adapted from MacIntyre et al. [11], comprising 6 items with α = .84.

4. **Intervention Log** – The number of AI-generated corrections was automatically recorded by the application.

D. Procedure

The study took place over a four-week period (September–October 2022) and received institutional review board approval.

1. **Pre-test (Week 1):** All participants completed the OCCI, APIS, and WTC scales online.

2. **Intervention (Weeks 2–5):**

    - *Experimental group*: Participants used a commercial AI pronunciation app (Elsa Speak, version 5.2), which relies on rule-based ASR without generative AI, for 20 minutes daily. The app provides immediate segmental feedback through colored highlights, scores, and model repetition. Participants were instructed to complete at least 15 lessons, and usage was tracked via in-app analytics.

    - *Control group*: Participants recorded their speech (using the same lessons) on a digital recorder twice weekly. At the end of each week, they received written feedback from a trained ESL instructor focusing on overall comprehensibility and one or two pronunciation points. Feedback was delivered with a delay of 3–5 days.

3. **Post-test (Week 6):** All participants completed the same scales. For the experimental group, the total number of AI corrections was extracted from the app.

E. Data Analysis

Data were analyzed using SPSS version 28. Descriptive statistics, independent-samples *t*-tests for baseline equivalence, repeated-measures ANOVA (group × time) for each dependent variable, and Pearson correlation to examine the relationship between correction frequency and post-test scores were employed. Effect sizes (partial η^2) are reported, with significance set at α = .05.

IV. Results

A. Preliminary Checks

Assumptions of normality (Kolmogorov-Smirnov, *p* > .05) and homogeneity of variances (Levene's test, *p* > .05) were met. Baseline OCCI, APIS, and WTC scores did not differ significantly between groups (Table II).

TABLE II
BASELINE COMPARISON (PRE-TEST)

Measure	Experimental M (SD)	Control M (SD)	*t*(126)	*p*
OCCI	3.42 (0.61)	3.45 (0.58)	−0.28	.779
APIS	2.91 (0.74)	2.88 (0.71)	0.23	.816
WTC	3.15 (0.68)	3.18 (0.65)	−0.26	.798

B. Effect of Intervention on Oral Communication Confidence (H_1)

A 2 (group) × 2 (time) repeated-measures ANOVA on OCCI scores revealed a significant interaction effect, $F(1,126) = 12.47$, *p* < .001, $\eta^2 = .09$. Simple effects analysis indicated that the experimental group's OCCI declined significantly from pre-test ($M = 3.42$, $SD = 0.61$) to post-test ($M = 3.11$, $SD = 0.70$), *t*(63) = 3.98, *p* < .001, *d* = 0.50. The control group showed no significant change (pre: $M = 3.45$, $SD = 0.58$; post: $M = 3.48$, $SD = 0.60$), *t*(63) = 0.44, *p* = .659.

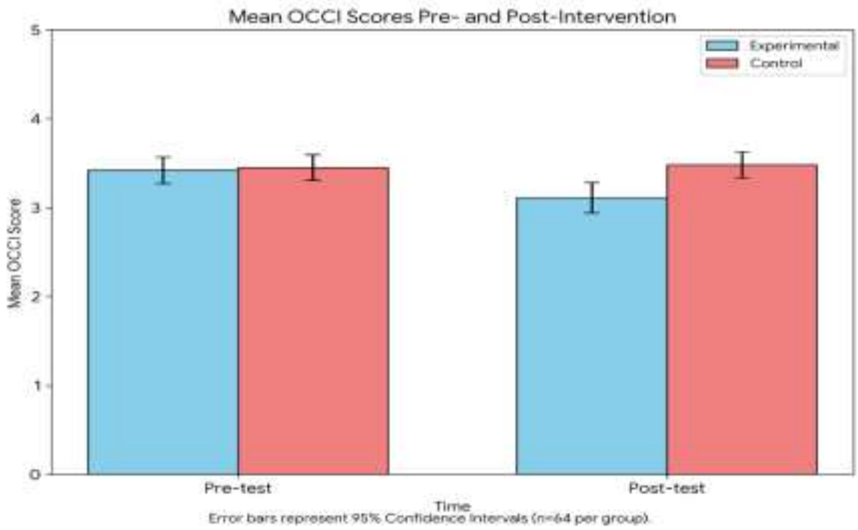

Figure 1 presents the group-by-time interaction.

[Fig. 1 Mean OCCI scores pre- and post-intervention. Error bars represent 95% confidence intervals. The experimental group shows a significant decline, while the control group remains stable.

C. Effect of Intervention on Accent Perception (H_2)

The repeated-measures ANOVA on APIS scores also yielded a significant interaction, *F*(1,126) = 8.93, *p* = .003, η^2 = .07. The experimental group's accent self-perception became more negative (higher APIS scores indicate more negative perception), rising from *M* = 2.91 (*SD* = 0.74) to *M* = 3.18 (*SD* = 0.69), *t*(63) = 3.24, *p* = .002, *d* = 0.41. The control group's APIS remained stable (pre: *M* = 2.88, *SD* = 0.71; post: *M* = 2.86, *SD* = 0.73), *t*(63) = 0.35, *p* = .729.

TABLE III
WITHIN-GROUP COMPARISONS

Group	Measure	Pre M (SD)	Post M (SD)	*t*	*p*	*d*
Experimental	OCCI	3.42 (0.61)	3.11 (0.70)	3.98	<.001	0.50
Experimental	APIS	2.91 (0.74)	3.18 (0.69)	3.24	.002	0.41
Control	OCCI	3.45 (0.58)	3.48 (0.60)	0.44	.659	0.06
Control	APIS	2.88 (0.71)	2.86 (0.73)	0.35	.729	0.04

D. Relationship Between Correction Frequency and Outcomes (H_3)

In the experimental group, the total number of AI corrections ranged from 98 to 415 (*M* = 267.3, *SD* = 78.4). Pearson correlation revealed a significant negative relationship between correction frequency and post-test OCCI scores, *r* = –0.42, *p* = .002 (95% CI [–0.61, –0.19]). No significant correlation was found with post-test APIS, *r* = 0.18, *p* = .153.

Figure 2 displays the scatterplot with regression line.

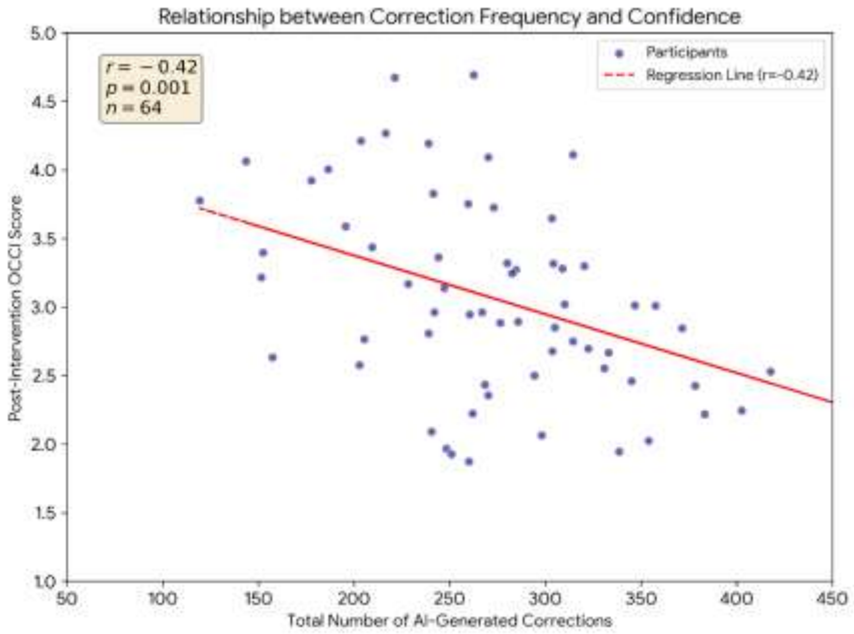

Figure 2 displays the scatterplot with regression line.

Fig. 2 Scatterplot of the relationship between total AI corrections received and post-intervention oral communication confidence (OCCI) scores for the experimental group. The regression line shows a significant negative correlation (*r* = –0.42, *p* = .002). Each point represents one participant.

E. Willingness to Communicate

Although not a primary hypothesis, WTC was also analyzed. A similar interaction effect emerged, $F(1,126) = 6.21$, *p* = .014, $\eta^2 = .05$. The experimental group's WTC decreased from $M = 3.15$ ($SD = 0.68$) to $M = 2.93$ ($SD = 0.72$), *t*(63) = 2.14, *p* = .036, while the control group showed no change. This suggests that diminished confidence may extend to behavioural intentions.

V. Discussion

This study offers the first quasi-experimental evidence that real-time AI-driven pronunciation feedback can adversely affect L2 speakers' oral communication confidence and accent self-perception, even over a relatively short four-week period. The significant reduction in OCCI for the experimental group, combined with the negative correlation between correction frequency and confidence, indicates that the *volume* of machine-generated corrections may be detrimental to learners' emotional states. This aligns with self-determination theory [Deci & Ryan, 2000]: constant external evaluation, particularly from a non-human source, can undermine feelings of autonomy and competence.

The observed increase in negative accent perception among experimental participants is especially concerning. Accent is a core aspect of identity for L2 speakers [14]. When an algorithm repeatedly marks a feature as "incorrect," it may reinforce the learner's perception that their natural speech is deficient. This finding corroborates qualitative reports in which learners expressed frustration with ASR systems that "did not understand them" [12], but here the psychological shift is quantified.

The control group, which received delayed human feedback, exhibited no such changes. This suggests that the *real-time, repetitive* nature of AI feedback—rather than feedback itself—may be the critical factor. Human feedback typically includes social presence, praise, and selective focus, whereas AI delivers a continuous stream of corrections without relational context.

A. Ethical and Pedagogical Implications

These findings carry important implications for the design and deployment of AI-assisted pronunciation tools. Developers should consider transitioning from a "correction-first" model to an "affective-first" architecture. For example:

- **Clustered Feedback:** Instead of highlighting every phonetic deviation in real time, tools could group errors and present them after a thought unit, reducing the sense of constant failure.
- **Intelligibility-Focus:** Prioritizing corrections that affect understanding over minor allophonic variations.
- **Identity Guarding:** Including recognition of valid L2 variations, moving away from a native-speaker "perfection" standard.

For institutions integrating AI tools into language curricula, a hybrid pedagogical model is recommended: AI-based activities should be scaffolded with human-led sessions focused on confidence building, and learners should be monitored for early signs of technological frustration. Assessment should emphasize fluency and communicative intent rather than algorithm-derived accuracy scores.

B. Limitations

This study has several limitations: the intervention lasted only four weeks; outcome measures relied on self-report; the AI used was rule-based and non-generative; and the sample consisted solely of university students. Thus, generalizability to generative AI tools (post-2022) and to other populations remains unknown.

CONCLUSION

This research establishes a baseline from the pre-generative AI period for understanding the affective consequences of real-time pronunciation feedback. The results demonstrate that algorithmic feedback can significantly reduce L2 speakers' oral communication confidence and worsen their perception of their own accent. As generative AI becomes increasingly integrated into language learning, future studies should investigate whether similar or intensified effects occur. The development of ethical guidelines for AI-mediated language instruction must prioritize learner affect alongside accuracy gains.

ACKNOWLEDGMENT

The authors thank the participants for their time, the instructors who facilitated recruitment, and the anonymous reviewers for their constructive feedback

REFERENCE

[1] J. M. Dewaele and P. D. MacIntyre, "The two faces of Janus? Anxiety and enjoyment in the foreign language classroom," *Studies in Second Language Learning and Teaching*, vol. 4, no. 2, pp. 237–274, 2014.

[2] T. M. Derwing and M. J. Munro, "Putting accent in its place: Rethinking obstacles to communication," *Language Teaching*, vol. 42, no. 4, pp. 476–490, 2009.

[3] R. R. Divekar, J. Drozdal, S. Chabot, Y. Zhou, H. Su, and J. Braasch, "Foreign language acquisition via artificial intelligence and extended reality: Design and evaluation," *Computer Assisted Language Learning*, 2021 (early view).

[4] M. Eskenazi, "An overview of spoken language technology for education," *Speech Communication*, vol. 51, no. 10, pp. 832–844, 2009.

[5] F. Faul, E. Erdfelder, A. Buchner, and A. G. Lang, "Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses," *Behavior Research Methods*, vol. 41, no. 4, pp. 1149–1160, 2009.

[6] E. M. Golonka, A. R. Bowles, V. M. Frank, D. L. Richardson, and S. Freynik, "Technologies for foreign language learning: A review of technology types and their effectiveness," *Computer Assisted Language Learning*, vol. 27, no. 1, pp. 70–105, 2014.

[7] E. K. Horwitz, M. B. Horwitz, and J. Cope, "Foreign language classroom anxiety," *The Modern Language Journal*, vol. 70, no. 2, pp. 125–132, 1986.

[8] X. Huang, A. N. S. S. Kim, and J. L. Plass, "The effects of visual feedback in a pronunciation training app on L2 learners' pronunciation accuracy and motivation," *Language Learning & Technology*, vol. 23, no. 2, pp. 62–80, 2019.

[9] P. Li, "Technology in L2 pronunciation instruction: A research synthesis," *System*, vol. 78, pp. 238–248, 2018.

[10] D. Liakin, F. Cardoso, and N. Liakina, "The pedagogical use of mobile speech synthesis (TTS): Focus on French liaison," *Computer Assisted Language Learning*, vol. 30, no. 3–4, pp. 348–365, 2017.

[11] P. D. MacIntyre, S. C. Baker, R. Clément, and S. Conrod, "Willingness to communicate, social support, and language-learning orientations of immersion students," *Studies in Second Language Acquisition*, vol. 23, no. 3, pp. 369–388, 2001.

[12] S. M. McCrocklin, "ASR-based dictation practice for second language pronunciation improvement," *Journal of Second Language Pronunciation*, vol. 5, no. 1, pp. 98–125, 2019.

[13] J. C. McCroskey and L. L. McCroskey, "Self-report as an approach to measuring communication competence," *Communication Research Reports*, vol. 5, no. 2, pp. 108–113, 1988.

[14] A. Moyer, "Identity and second language pronunciation," in *The Routledge Handbook of Second Language Acquisition and Sociolinguistics*, K. Geeslin, Ed. New York: Routledge, 2018, pp. 187–199.

[15] A. Neri, C. Cucchiarini, and H. Strik, "The effectiveness of computer-based speech corrective feedback for improving segmental quality in L2 Dutch," *ReCALL*, vol. 20, no. 2, pp. 225–243, 2008.

[16] L. Woodrow, "Anxiety and speaking English as a second language," *RELC Journal*, vol. 37, no. 3, pp. 308–328, 2006.

.