ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

JOURNAL OF APPLIED LINGUISTICS AND TESOL (JALT)
Vol.9.No.1 2026

EVALUATING USER TRUST AND PSYCHOLOGICAL IMPACT OF AUTOMATED SPEECH RECOGNITION (ASR) FEEDBACK ON NON-NATIVE ENGLISH LEARNERS IN PAKISTAN

Sadia Ghaffar[1], Amna Ali[2], Syeda Noor Fatima[3]
University of Education, Lahore, Pakistan
[1]Email: sadiaghafar@gmail.com
[2]Email: amnaali2504@gmail.comp
[3]Email: syedanoor7521@gmail.com

Abstract
This research examines the link between speech technologies and human psychology by investigating the relationship between user trust and the psychological effects of feedback from an Automated Speech Recognition (ASR) tool for non-native English-speaking students in Pakistan. Using a quantitative correlational approach, the experiment involved a sample size of $N = 40$ college students enrolled in English linguistics at the undergraduate level, who were engaged with an accessible ASR tool based on the Whisper system from OpenAI. Psychological variation was determined through modified versions of the TAM model and FLCAS anxiety scale. Statistical analysis conducted in SPSS showed a significant positive relationship between the reliability of the ASR system and user trust ($r = .68$, $p < .01$). At the same time, frequent mistakes made by the ASR system caused speaking anxiety among low-level language learners.

Introduction
In modern SLA, CAPT has undergone a paradigm shift owing to developments in both artificial intelligence and automated speech recognition. The model by OpenAI known as Whisper can give instant phonetic feedback to non-native speakers learning English. In countries such as Pakistan where there is a lack of native speakers and language laboratories, ASR systems present a scalable alternative to micro-level pronunciation correction.

However, existing literature continues to be very one-sided towards technical error-rate minimization while ignoring the HCI element entirely. Following the theory of Schmidt’s Noticing Hypothesis, language acquisition happens through learners’ noticing the difference between their target output and the feedback provided. However, if there is any lack of trust from the learner in the digital tool, or if automatic correction leads to major emotional problems in the learner, the effectiveness of such technology for pedagogy suffers. This research aims to fill this significant void by measuring trust levels and FLSA of ESL learners in Pakistan.

Literature Review
The integration of ASR technology into language teaching from the point of view of theories is fundamentally based on Long's Interaction Hypothesis and Schmidt's Noticing Hypothesis. As noted by Schmidt (1990), unless learning input is consciously 'noticed', it will not lead to intake. Through ASR feedback in computer programs, the non-native speaker is compelled to cognitively acknowledge the phonetic differences between the learner's interlanguage and native speech.

According to empirical studies in CAPT, such immediate feedback increases acoustic accuracy dramatically (Derwing & Munro, 2015). The use of technology, however, is more than a cognitive act in education, as technology adoption involves the socio-psychology of learning. As per the model presented by Davis (1989), the continuous use of any technology depends on the perceived usefulness (PU) and perceived ease of use (PEOU) of that technology, which together make up the user trust level. When there is no trust built by the users for any AI model, cognitive resistance arises.

Within the framework of Pakistani society, language learning is highly influenced by socio-cultural constraints where the prospect of a negative evaluation leads to high levels of anxiety (Horwitz, Horwitz, & Cope, 1986). Traditional theories about SLA have considered teacher-

ISSN E: 2709-8273
ISSN P:2709-8265

induced and peer-induced anxiety as critical sources of such anxiety, which arise within classroom settings. However, algorithm-induced anxiety, where there is continuous criticism by an automated algorithm against a non-native accent of the students, remains significantly unexplored psychologically. This research aims to fill this gap in literature.

Research Methodology & Data Analysis Framework

In order to ensure that this research is empirically rigorous in line with experiment-based applied linguistic conventions, a Quantitative Correlational and Survey-Based Experimental Design is utilized here. Rather than focusing on the programming part of technological interaction through algorithms, the human-computer interaction (HCI) aspect is isolated for measurement.

Hypotheses Development

To ensure scientific rigor and directionality within the empirical analysis, the following research hypotheses have been proposed:

H1: There is a positive correlation between the perceived reliability of ASR phonetic feedback and UTI.

H2: Increased phonetic errors flagged by the system strongly predict higher FLSA.

Operationalization Variables Matrix

Structural configuration of the experiment requires a distinction between variables, avoiding any overlapping in their empirical representation per the requirements of international reviewers:

- **Independent Variable (IV):** Frequency and types of ASR Phonetic Correction (operationalized via controlled oral interaction with OpenAI's Whisper Base model core).
- **Dependent Variable 1 (DV1):** User Trust Index (UTI) – representing system reliability, objective accuracy, and dependency.
- **Dependent Variable 2 (DV2):** Foreign Language Speaking Anxiety Scale (FLSAS) – representing situational stress, self-consciousness, and emotional barriers.

Target Participants & Purposeful Sampling Methodology

The number of participants $N = 40$, enrolled in the BS English Linguistics undergraduate program at the University of Education, Lahore, was chosen using a purposeful sampling methodology. The use of this non-random sampling methodology resulted in homogeneity among all participants regarding their socio-linguistic background in the following aspects:

- **L1 Origin:** Native Urdu/Punjabi speakers.
- **Academic Background:** All participants have standardized intermediate English language competence.
- **Technical Background:** None had any prior experience using automated pronunciation feedback software.

Experimental Procedure

Prior to conducting the actual experiment, a pilot test involving ten participants was done. In this regard, the tool achieved Cronbach's Alpha internal consistency of $\alpha = .78$, which met international validation standards ($\alpha \geq .70$). Following this, the actual empirical experiment was conducted in a well-controlled lab setting through two separate stages of operations.

- **Stage 1 (Stimulation):** This stage involved conducting a controlled oral reading activity among participants for 15 minutes using standardized English phonetics test batteries that had minimal pairs and phonetic clusters. Speech input was done using an open-source ASR technology called the Whisper Base Architecture.
- **Stage 2 (Psychometric Evaluation):** Following completion of the first stage, a 5-scale Likert Scale psychometric evaluation tool was administered immediately to collect

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

unbiased psychological responses before any environmental decay could affect participants' perceptions.

Data Collection Tools

The questionnaire items were based on the standard instrument of Horwitz's Foreign Language Classroom Anxiety Scale (FLCAS) and Davis's Technology Acceptance Model (TAM) to ensure construct validity.

Sub-Scale A: User Trust Index (UTI)

- **Item 1:** I think that the phonetic corrections made by the ASR model are reliable and accurate.
- **Item 2:** I have faith in the accuracy of the model's ability to detect my mistakes with respect to pronunciation relative to the teacher.
- **Item 3:** I am confident that the corrections made by the automated system will help me improve my pronunciation skills in English language.

Sub-Scale B: Foreign Language Speaking Anxiety (FLSAS)

- **Item 4:** It makes me anxious or uncomfortable when the ASR system detects that my pronunciation style is incorrect.
- **Item 5:** Getting instant automated corrections motivates me to try again to fix my pronunciation problems.
- **Item 6:** It creates less anxiety for me to be corrected via AI technology compared to making pronunciations in the presence of a human teacher.

Statistical Data Analysis & Tables

The collected quantitative data was then encoded and analyzed through the use of SPSS (Statistical Package for the Social Sciences). Descriptive and inferential statistical tests were used to analyze the psychological constructs.

Table 1: Descriptive Statistics of Key Psychological Variables

Variable Measure	Mean (M)	Standard Deviation (SD)	Verbal Interpretation
User Trust Index (UTI)	3.82	0.64	High Trust Level
Speaking Anxiety Score (FLSAS)	3.45	0.78	Moderate-to-High Anxiety
L2 Learning Motivation	4.12	0.55	Strong Motivation

Statistical Note for Table 1: As seen from the figures in Table 1, although Pakistani ESL students have a very strong psychological sense of ASR accuracy ($M = 3.82$), the consistent use of automated algorithmic feedback also provokes a considerable sense of anxiety ($M = 3.45$). Nonetheless, their motivation levels stay remarkably high ($M = 4.12$).

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Table 2: Pearson Correlation Matrix of Feedback Reliability and User Trust

Variables	**Perceived Algorithmic Reliability**	**User Trust Index (UTI)**
Perceived Algorithmic Reliability	1.00	.68
User Trust Index (UTI)	.68	1.00

Note. Correlation is significant at the 0.01 level (2-tailed).
Interpretation of Statistical Data in Table 2: For testing hypothesis H1, Pearson Correlation was run. It clearly indicates a very strong and statistically significant positive correlation between the accuracy of the system and trust in the system ($r = .68$, $p < .01$), and therefore, hypothesis H1 is validated.

Table 3: Linear Regression Model (Tech Anxiety Variance Prediction)

Model	Unstandardized Coefficients (B)	Standard Error	Standardized Coefficients (β)	t-value	Significance (p)
(Constant)	1.20	0.34	—	3.53	< .01
ASR Error Frequency	0.45	0.12	0.42	3.75	< .05

Note. $R^2 = .176$; Adjusted $R^2 = .155$; $F(1, 38) = 14.06$, $p < .05$.
Statistical Analysis of Table 3: To prove H2, a linear regression model was used. Findings show that the frequency of automatic error flags predicts increased anxiety among students ($\beta = .42$, $t = 3.75$, $p < .05$). This proves H2 and shows that high-frequency corrective loops enhance the learner's affective filter.

Discussion

The empirical evidence generated through this study provides invaluable understanding of the dynamics associated with CAPT in the Pakistani ESL environment. The confirmation of H1 ($r = .68$, $p < .01$) bears significant relevance with Davis's (1989) TAM theory, in which he establishes that user trust is intrinsically contingent upon system reliability. In this case, undergraduate students taking courses in linguistics can establish trust and become system-dependent when they perceive that the Whisper technology used by OpenAI is capable of identifying their phonetic discrepancies. As such, the objective feedback received from the technology is considered as more unbiased than subjective human judgment, corroborating the arguments by Derwing and Munro (2015).

On the other hand, the confirmation of H2 ($\beta = .42$, $p < .05$) creates a pedagogical dilemma in the teaching process. On one hand, instantaneous computer-based corrections help satisfy Schmidt's noticing hypothesis since students will be compelled to focus on the phonetic errors; however, these instant notifications have been identified as increasing situational anxiety (FLSA) levels among users. According to Horwitz et al. (1986), elevated anxiety levels activate the affective filter, which hinders language production.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

When a system consistently highlights issues in a learner's speech patterns, they face a type of induced performance anxiety brought about by an algorithm. However, the descriptive baseline for L2 motivation was remarkably resilient despite such anxieties ($M = 4.12$). This suggests that although there is considerable psychological stress in using such software, non-native speakers can easily appreciate the importance of these applications.

Conclusion & Pedagogical Implications

The present study is able to provide an accurate measure of the subtle relationship between technology and psychology in practical linguistics. The study proves that though feedback loops powered by ASR have gained significant trust among Pakistani English as a Second Language undergraduate students, they pose the threat of causing cognitive fatigue and speech anxiety if not moderated by emotion and pedagogy.

Pedagogical Recommendations:

- **Adaptive Feedback Gateways:** CAPT programmers must incorporate adaptive feedback gateways that will reduce the instances of displaying errors in lower-proficiency baselines to avoid anxiety-induced shutdowns.
- **Hybrid Approach:** Instead of substituting humans with technology, ASR programs can be employed as supplementary software for reducing classroom anxiety.

Limitations and Future Research:

Although this experiment involves a well-controlled purposive sampling process ($N = 40$), it is necessary for future research to adopt longitudinal studies involving more than one Pakistani university to observe changes in user trust over time.

References

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-340.

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. John Benjamins Publishing Company.

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, 70(2), 125-132.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129-158.