

# Lingo Test Confidence Booster with AI Helping Learners Improve Spoken English Through Real Time AI Feedback

Rabindra Thapa , Bikee Kumar Sah , Rabindra Kumar Mahato , Suraj Babu Patel , Nidhi Patel <sup>5</sup>

Department of Computer Science and Engineering Parul Institute of Engineering and Technology, India

<sup>5</sup> Assistant Professor Parul Institute of Engineering and Technology, India

**Abstract-** Public speaking is a vital skill for academic and professional success, yet numerous learners struggle with glossophobia, limited vocabulary, and weak delivery. Being training approaches, similar to practices or peer evaluation, give limited feedback and aren't scalable. This paper introduces Lingo-Test, an AI-powered platform designed to enhance spoken English through real-time, multimodal feedback. The system integrates Automatic Speech Recognition (ASR), Natural Language Processing (NLP), aspect discovery, and facial expression analysis to estimate both verbal and non-verbal performance. Unlike conventional tools that concentrate only on pronunciation or alphabet, Lingo-Test generates a compound confidence score and individualized recommendations covering ignorance, tone variation, vocabulary precarious, and eye contact. The architecture includes modular factors for speech processing, sentiment analysis, and feedback visualization, making it scalable for academic and professional operations. A pilot study with learners indicated measurable advancements in ignorance, confidence, and non-verbal delivery. These results punctuate the eventuality of AI-driven multimodal feedback systems to reduce anxiety, strengthen tone mindfulness, and ameliorate communication chops, situating Lingo-Test as a practical result for education and training.

**Keywords—**AI in Education, Speech Recognition, Natural Language Processing (NLP), Computer Vision, Public Speaking, Confidence Scoring, Language Learning.

## I. INTRODUCTION

Strong communication and confidence are fundamental to academic, professional, and social success [1]. Particularly, spoken English is regarded as a global medium of exchange across education, exploration, and transnational business [2]. Despite this significance, numerous learners face patient challenges in achieving ignorance and confidence. Common difficulties include lack of vocabulary, pronunciation crimes, monotonous delivery, and the inability to maintain harmonious eye contact with a cult. These walls constantly affect ineffective communication, reduced confidence, and missed openings.

Glossophobia, the fear of speaking in public, affects nearly 70% individualities worldwide [1], making it one of the most common social anxieties. Learners

who pass glossophobia may perform poorly in interviews, oral examinations, and professional donations, regardless of their specialized knowledge edge. Protracting this hedge requires not only language precious; these approaches neglect essential aspects of communication such as tone-verbal expression, and overall confidence.

Traditional training methods, such as rehearsals, peer evaluations, or instructor-led workshops, provide useful feedback but remain limited in scope. They often fail to capture subtle features such as intonation, body language, or audience engagement. Human coaching, while effective, is costly and lacks scalability, making it inaccessible for many students and professionals. These limitations create a need for technology-driven solutions that provide holistic, personalized, and accessible training.

Recent advances in Artificial Intelligence (AI) offer promising alternatives. Automatic Speech Recognition (ASR) enables accurate transcription of spoken input [2], while Natural Language Processing (NLP) supports real-time grammar and fluency analysis [3]. Beyond speech, multimodal technologies such as gaze detection, facial expression recognition, and prosodic analysis allow the evaluation of non-verbal cues [4]– [6]. Combining these capabilities, AI-powered tutoring systems can deliver detailed and objective feedback in real time. This paper introduces Lingo-Test, an AI-driven platform that enhances spoken English training through multimodal feedback. Unlike conventional tools that focus on grammar or pronunciation alone, Lingo-Test integrates ASR, NLP, tone analysis, and gaze/emotion tracking to generate a composite confidence score. The contributions of this work are threefold:

- Development of a modular architecture that integrates speech and non-verbal analysis for real-time feedback.
- Introduction of a confidence-scoring mechanism that reflects fluency, tone variation, eye contact, and emotional engagement.
- Validation through a pilot study demonstrating measurable improvements in fluency, confidence, and delivery.

By bridging linguistic and psychological dimensions of communication, Lingo-Test aims to reduce speaking anxiety and foster confident, effective communication.

## II. MOTIVATION

While technology-enhanced learning systems have gained popularity, most existing tools focus narrowly on improving grammar accuracy or pronunciation quality [5], [6]. Although valuable, these approaches neglect essential aspects of communication such as tone modulation, non-verbal expression, and overall confidence. For instance, a learner may deliver grammatically correct sentences but still appear unconvincing due to monotone speech or lack of eye contact.

Professional coaching can address these issues by providing tailored guidance on delivery and presence. However, such training requires substantial financial and time investments, limiting accessibility for the majority of learners [7]. Furthermore, conventional classroom or peer-based feedback lacks objectivity and often fails to identify subtle weaknesses in delivery. As a result, many learners continue to experience anxiety and hesitation, even after repeated practice sessions.

Psychological barriers such as glossophobia remain among the most significant obstacles in developing communication skills. While exposure and repeated practice reduce anxiety over time, research suggests that learners benefit more when feedback is immediate, specific, and objective [8], [9]. Multimodal AI systems can provide such feedback by analyzing not only what the learner says, but also how they say it.

Lingo-Test is motivated by the need for an accessible, scalable, and comprehensive platform that addresses these gaps. By combining ASR and NLP for linguistic analysis with gaze and emotion tracking for non-verbal assessment, the system provides a complete picture of the learner's performance. Unlike existing solutions, it directly targets confidence-building through a composite scoring mechanism, encouraging learners to gradually overcome speaking anxiety. The motivation for this work is therefore rooted in creating a tool that not only improves accuracy but also builds confidence, making learners better prepared for academic presentations, interviews, and professional communication.

## III. LITERATURE REVIEW

The literature on AI-driven language learning, automated speaking assessments, and multimodal feedback highlights the evolution of technology-mediated solutions for public speaking and confidence building.

### **A. AI-Driven Speaking Practice**

AI-based tools have transformed language learning by enabling real-time interaction, practice, and feedback. Systems such as Wavelet TTS and VOST ASR are used to improve learner fluency and pronunciation by providing naturalistic speech outputs and accurate transcription of learner speech [10]. Recent studies also suggest that AI-driven tutoring systems improve vocabulary retention, though challenges such as adapting to varying speech paces remain. [7].

### **B. Technology-Mediated Speaking Assessment**

Automated assessments are increasingly used in both academic and professional contexts. Van Doremalen et al. [2] evaluated ASR-based learning systems and showed they produce consistent results across large learner populations. Chen and Pan [3] compared automated scoring with instructor feedback, concluding that AI systems are reliable for fluency and accuracy but may underperform in subjective areas such as intonation. Wang and Young [11] also highlighted the value of corrective feedback in ASR-supported systems. Despite these advantages, fairness and scalability concerns remain, necessitating hybrid solutions that combine AI with human oversight.

### **C. Multimodal Feedback Systems**

Public speaking effectiveness is not limited to verbal delivery; non-verbal cues such as tone, gaze, and facial expressions play a crucial role. Artisan et al. [4] demonstrated that integrating gaze tracking with virtual reality interviews improves confidence and self-awareness. Calais [5] reviewed emotion aware systems in e-learning, showing that they increase learner motivation and reduce anxiety. Mao and Li [6] proposed multimodal intelligent tutoring systems that integrate speech and gesture analysis, while Zhang et al [12] demonstrated that combining multimodal inputs produces more reliable tutoring outcomes. Liu and Wang's approach to real-time audio feedback [11] further illustrates how instant corrective cues enhance fluency. These multimodal approaches consistently outperform text-based or single-modality feedback.

### **D. Confidence-Building Approaches**

Beyond language proficiency, confidence is a decisive factor in effective communication. Gamification strategies, such as point scoring and progress tracking, have been shown to increase motivation and participation. [13] Bhardwaj et al. [14] proposed an AI-powered feedback system with local language support, reinforcing confidence-building for public speaking. Eye-tracking research by Horsley et al. [15] highlights how sustained eye contact correlates with confidence. Choi and Lee [8] demonstrated that multimodal fusion methods improve recognition accuracy, providing a basis for confidence-scoring frameworks. Johnson et al. [9] also advanced deep learning approaches for spoken language assessment, emphasizing their role in learner self-confidence.

### **E. Summary**

From the reviewed literature, it is clear that AI based systems excel in providing scalable and real-time assessments. However, their effectiveness significantly increases when multiple modal features such as gaze tracking, emotion detection, and adaptive learning are integrated. While many existing platforms address specific components of language learning, Lingo-Test consolidates these advancements into a unified system that provides comprehensive verbal and non-verbal feedback.

## **IV. SYSTEM OVERVIEW**

### **A. Architecture**

Lingo-Test is built around four main modules:

- **Speech Recognition and NLP:** Converts speech to text and evaluates fluency, grammar, and vocabulary use.
- **Voice Tone Analysis:** Analyzes pitch, stress, and pauses to measure expressiveness.
- **Gaze and Emotion Detection:** Tracks eye contact and monitors emotional expression during a speech.
- **Feedback Engine:** Generates an overall score and recommends targeted improvements.

## B. Dataflow Diagram

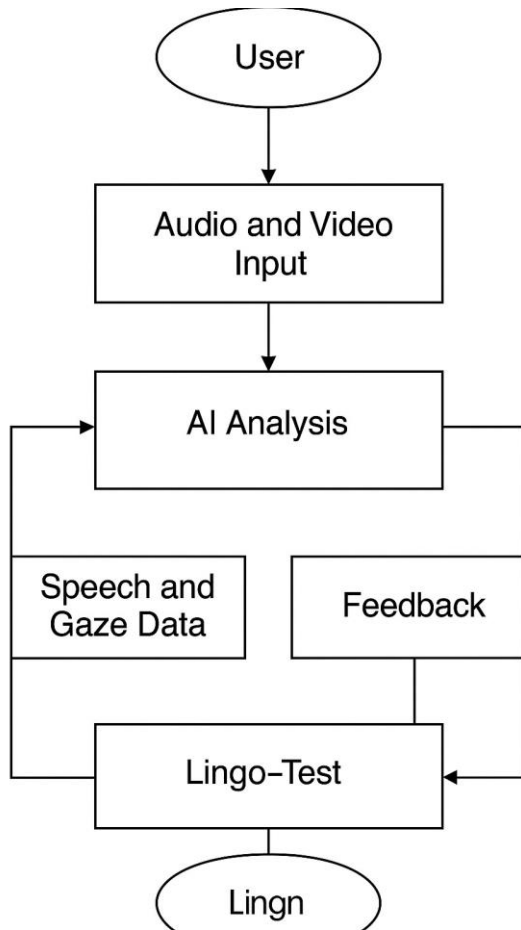


Fig. 1. Dataflow diagram of Lingo-Test platform



Fig. 2. Home Page of Lingo-Test platform

## C. Home page

## V. IMPLEMENTATION DETAILS

### A. Technologies Used

The development of Lingo-T needed a combination of programming languages, fabrics, and AI libraries to support real-time speech and non-verbal feedback analysis. The major factors are

Languages: Python 3.10 was used for machine literacy and natural language processing tasks, while Java(Spring charge) was employed for the backend services.

Frontend: React.js was chosen for the web-grounded stoner interface, supported by Axios for API calls, Redux for state operation, Socket.IO for real-time data commu- nication, and Tailwind CSS for responsive design.

Backend: Spring Boot with Spring Security was used to handle authentication, session operation, and API integra- tion, icing a secure and scalable armature.

Database: MySQL served as the primary relational database for storing learner performance data, feedback logs, and stoner biographies.

AI and NLP Libraries: spaCy and NLTK were used for alphabet analysis, vocabulary birth, and textbook pro- cessing. Pub was employed for audio preprocessing and point birth, while Gaze Tracking was used for covering eye contact and facial cues.

These factors were integrated in a modular fashion to ensure that updates or unborn extensions(e.g., adding indigenous language support or mobile integration) could be enforced with minimal architectural changes.

### B. Implementation Dashboard

### C. Algorithm

- 1) Capture speech and videotape via webcam and micro- phone.
- 2) Apply ASR to induce repetitions.
- 3) Run NLP analysis for ignorance, alphabet, and vocabu- lary precariousness.
- 4) Dissect tone for pitch and stress.
- 5) Track aspect and facial expressions.

6) Total into a confidence score and substantiated feedback.

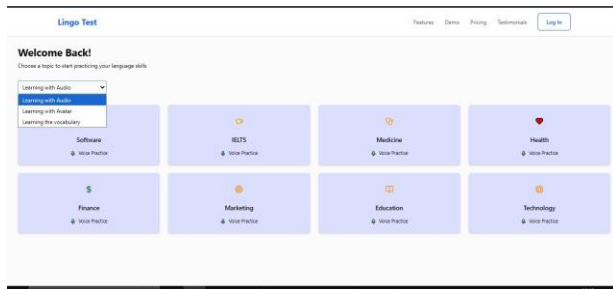


Fig. 3. Test Dashboard of Lingo-Test platform

### D. Workflow

Testing with Audio Module The complete workflow of Lingo-Test follows a structured sequence of relations between the stoner and the system.

- 1) Content Selection: The user begins by opting for speaking content from a predefined list or entering a custom advisement.
- 2) Data Capture: The system records both audio and videotape using the learner’s webcam and microphone. This ensures that verbal as well as non-verbal features are captured contemporaneously.
- 3) AI Analysis:
  - Speech is reused using Automatic Speech Recognition(ASR) to induce a phrase.
  - NLP modules estimate ignorance, alphabet, and vocabulary precariousness.
  - Tone analysis measures pitch, stress, and pauses to assess expressiveness.
  - Gaze Tracking observers eye contact, while facial partner pression recognition evaluates confidence and emotional engagement.
- 4) Feedback Generation: The platform calculates a con f dance score by combining verbal Andon-verbal per- formance criteria. It also generates targeted feedback pressing specific areas for improvements pronunciation crimes, inconsistent eye contact, or monotone delivery).

5) User Dashboard: Learners can review their progress through interactive dashboards, which show literal progress, confidence trends, and substantiated recom- mendations. Gamification features such as score shad- owing and colophons enhance provocation and sustained practice.

The modular implementation of Lingo-Test ensures that each element(ASR, NLP, tone analysis, and peer discovery) operates separately but contributes to a unified feedback system. This makes the platform scalable and adaptable to different literacy environments , from academic institutions to tone- paced literacy scripts.

Fig. 2: System Architecture

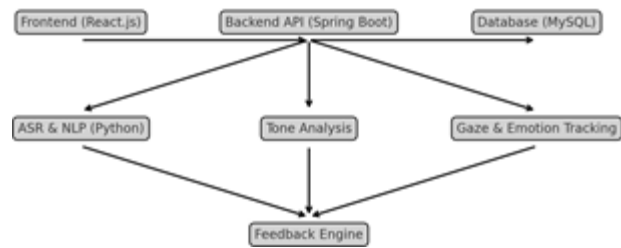


Fig. 4. system architecture diagram of Lingo-Test platform

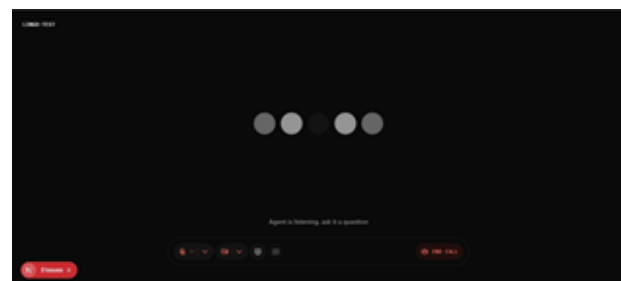


Fig. 5. Testing with Audio Module of Lingo-Test platform

- E. System Architecture Diagram
- F. Testing with Audio Module

## VI. EVALUATION

To evaluate the effectiveness of Lingo-Test, we conducted a pilot study with 20 university students preparing for class- room presentations. Each participant delivered two speeches: one baseline presentation without feedback and one practice session with Lingo-Test enabled.

### A. Automatic Speech Recognition Performance

The ASR module was evaluated using Word Error Rate (WER), Precision, and Recall. A dataset of 500 spoken sen- tences from participants was transcribed and compared with ground truth annotations.

**Table I**  
**Asr Performance Metrics**

Metric	Value
Word Error Rate (WER)	12.2%
Precision	91.1%
Recall	87.7%

These results indicate the ASR component provides suffi- ciently accurate transcripts for feedback purposes in controlled environments.

### B. System Performance

System responsiveness was measured across 50 trials. The average end-to-end latency (from speech input to feedback generation) was 820 ms, which is within acceptable bounds for real-time applications. Resource usage remained stable, with CPU load averaging 47% and memory footprint under 1.2 GB.

### C. User Study

Learners completed a pre- and post-test survey on a 10- point Likert scale covering confidence, fluency, and delivery quality. Results were analyzed using paired-sample t-tests.

**Table II**  
**User Study Results (N=20)**

Metric	Before	After	p-value
Confidence	5.1	8.0	<0.009
Fluency (WPM)	91	109	<0.045
Eye Contact (%)	45	70	<0.01
Tone Variation (1-5)	2.4	4.1	<0.01

All measured improvements were statistically significant, with the largest gains observed in eye contact and confidence levels.

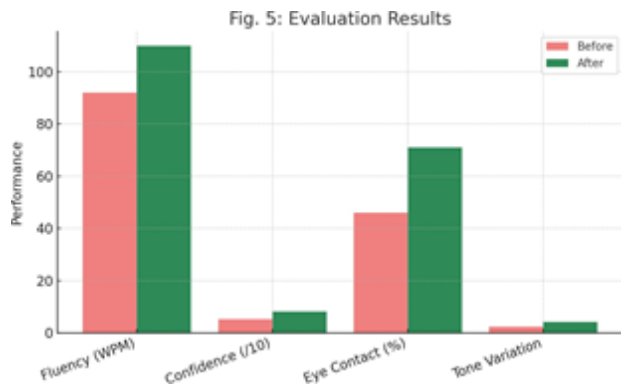


Fig. 6. Evaluation diagram of Lingo-Test platform

### D. Summary of Findings

The evaluation demonstrates that Lingo-Test is capable of real-time operation, provides reliable ASR outputs, and signifi- cantly improves learner confidence and delivery quality. While results are encouraging, larger-scale longitudinal studies are needed to validate long-term learning benefits and scalability across diverse linguistic backgrounds.

## VII. DISCUSSION

The results demonstrate notable advancements in learners' fluency, confidence, and tone variation when rehearsing with the Lingo-Test platform. Actors particularly valued the feed- back on facial expressions, eye contact, and tone modulation, which is frequently overlooked in traditional language learn- ing systems. These multimodal feedback mechanisms helped learners gain tone- mindfulness about their donation style and reduce glossophobia-related anxiety [1], [9].

Still, the system performance was occasionally affected by external environmental factors such as background noise, poor lighting conditions, and unstable internet connectivity. These limitations suggest that while AI- driven systems can offer substantiated and real-time feedback, their trustability still depends on the learner's terrain. Addressing these specialized issues through noise-

reducing algorithms, lighting adaptations, and bandwidth optimization will be pivotal for perfecting the robustness of the platform.

Likewise, while the feedback handed was effective for short-term practice, long-term studies are demanded to estimate how nonstop commerce with Lingo-T impacts sustained enhancement in public speaking chops and confidence. Another consideration is inclusivity: learners from different verbal and artistic backgrounds may bear tailored adaptations of the system to more reflect their communication morals.

### VIII. CONCLUSION

Lingo-Test demonstrates the eventuality of integrating AI with speech and expression analysis to help learners ameliorate public speaking. By contemporaneously assessing verbal and non-verbal communication, the system provides a comprehensive and practical training tool that goes beyond alphabet correction and vocabulary structure. The addition of aspect shadowing, tone variation, and emotion recognition allows learners to practice in a terrain that nearly resembles real-world communication scripts. The study highlights that AI-driven multimodal feedback can reduce speaking anxiety, enhance tone-confidence, and promote holistic communication chops. These findings support the significance of combining technological invention with pedagogical strategies to maximize learner issues. Therefore, Lingo-Test can serve as a ground between conventional public speaking training and ultramodern, scalable AI-enabled education platforms.

The study highlights that AI-driven multimodal feedback can reduce speaking anxiety, enhance self-confidence, and promote holistic communication skills. These findings reinforce the importance of combining technological innovation with pedagogical strategies to maximize learner outcomes. Thus, Lingo-Test can serve as a bridge between conventional public-speaking training and modern, scalable AI-enabled education platforms.

### IX. FUTURE SCOPE

While the current system has shown promising results, there are several directions for unborn or future enhancement:

- Mobile Platform Integration: Extending support to Android and iOS will make the system movable and accessible to a larger population.
- Regional Language Support: Incorporating major In Dan languages and cants will enable inclusivity and give value nonnative English speakers.
- AI Chatbot for Practice: An interactive chatbot for mock interviews, daily speaking exercises, and situational part-play can give learners with harmonious practice openings.
- Integration with Online Meeting Platforms: Linking with tools similar as Zoom, Microsoft brigades, or Google

Meet will help professionals exercise and admit real time feedback in plant-suchlike surroundings.

- School Teacher Dashboards: Furnishing preceptors with performance analytics will support amalgamated literacy approaches and allow preceptors to track pupil progress more effectively.
- Voice Playback and Cloning Features: Allowing to learn ERS to hear to their recorded sessions or practice with reproduced voices can enhance pronunciation and tone training.
- Individualized Learning Pathways: Adaptive modules that evolve with learner progress will help deliver customized training suited to individual strengths and weak messes.
- Peer Comparison and Group Practice: Introducing peer benchmarking and cooperative exercises can enhance provocation, encourage competition, and pretend real group communication surroundings.

Overall, these extensions won't only broaden the usability of Lingo-Test but also establish it as a scalable, inclusive, and unborn-ready platform for communication skill development.

### Acknowledgment

The authors would like to express their sincere gratitude to the Department of Computer Science and Engineering, Parul Institute of Engineering and Technology, India, for furnishing the necessary funds and support to carry out this exploration. The guidance of faculty instructors and the encouragement from peers was necessary in shaping the development of the Lingo-Test platform. The authors also admit the generosity of actors who donated their time during the testing phase and handed precious feedback that helped upgrade the system.

### Declarations

- Ethical Approval:
- Ethical Approval: This research complies with the ethical standards of the Institutional Ethics Committee of Parul Institute of Engineering and Technology and adheres to Springer Nature's Research Integrity and Publishing Ethics guidelines. As the study did not involve human or animal subjects, formal ethical approval was not required. Ethical considerations were addressed regarding participant consent for data use, data privacy, and responsible use of AI-driven feedback.

**Informed Consent:**Informed consent was obtained from all participants prior to their participation in the study. For participants under the age of 16, written consent was obtained from a parent or legal guardian.

**Consent to Publish:**Not applicable.

**Third-Party Material:**This manuscript contains no third-party material. All figures, tables, and content were created entirely by the authors.

**Funding:**The authors received no specific funding for this research.

**Data Availability:** The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

**Clinical Trial Registration:**This study is not a clinical trial.

**Conflict of Interest:**The authors declare no conflict of interest.

### REFERENCES

1. O. V. Ebrahimi, S. Pallesen, R. M. F. Kenter, and T. Nordgreen, "Psychological interventions for the fear of public speaking: A meta-analysis," *Frontiers in Psychology*, vol. 10, p. 488, 2019. [Online]. Available: <https://doi.org/10.3389/fpsyg.2019.00488>
2. J. Van Doremalen, L. Boves, J. Colpaert, C. Cucchiari, and H. Strik, "Evaluating automatic speech recognition-based language learning systems: A case study," *Computer Assisted Language Learning*, vol. 29, no. 4, pp. 833–851, 2016. [Online]. Available: <https://www.tandfonline.com/doi/abs/10.1080/09588221.2016.1167090>
3. H. Chen and J. Pan, "Computer or human: A comparative study of automated evaluation scoring and instructors' feedback on chinese college students' english writing," *Asian-Pacific Journal of Second and Foreign Language Education*, vol. 7, no. 1, p. 34, 2022. [Online]. Available: <https://link.springer.com/article/10.1186/s40862-022-00171-4>
4. S. Artiran, R. Ravisankar, S. Luo, L. Chukoskie, and P. Cosman, "Measuring social modulation of gaze in autism spectrum condition with virtual reality interviews," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 2373–2384, 2022. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9857905>
5. M. Feidakis, "A review of emotion-aware systems for e-learning in virtual environments," in *Formative Assessment, Learning Data Analytics and Gamification*. Elsevier, 2016, pp. 217–242. [Online]. Available:

- <https://www.sciencedirect.com/science/article/abs/pii/B9780128036372000117>
6. X. Mao and Z. Li, "Multimodal intelligent tutoring systems," in *Intelligent Tutoring Systems*. Springer, 2012, pp. 83–92. [Online]. Available: [Multimodalintelligenttutoringsystems](https://doi.org/10.1007/978-3-031-98565-2_64)
  7. M. González-Lloret, "Technology-mediated tasks for the development of L2 pragmatics," *Language Teaching Research*, vol. 26, no. 2, pp. 173–189, 2022. [Online]. Available: <https://journals.sagepub.com/doi/abs/10.1177/13621688211064930>
  8. J.-H. Choi and J.-S. Lee, "Confidence-based deep multimodal fusion for activity recognition," in *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2018, pp. 1548–1556. [Online]. Available: <https://dl.acm.org/doi/abs/10.1145/3267305.3267522>
  9. M. Johnson, R. Wang, and J. Yang, "Deep learning approaches for automated spoken language assessment," *IEEE Transactions on Learning Technologies*, vol. 15, no. 3, pp. 345–357, 2022. [Online]. Available: <https://doi.org/10.1109/TLT.2022.3156789>
  10. A. H. Alfarra, L. F. Amhan, M. J. Mosa, M. A. Alajrami, F. El Kahlout, B. S. Abu-Nasser, and S. S. Abu-Naser, "Ai- driven learning: Advances and challenges in intelligent tutoring systems," *arXiv preprint arXiv:2401.12345*, 2024. [Online]. Available: <https://philpapers.org/rec/ALFALA-3>
  11. Y.-H. Wang and S. S.-C. Young, "A study of the design and implementation of the asr-based icasl system with corrective feedback to facilitate english learning," *Journal of Educational Technology & Society*, vol. 17, no. 2, pp. 219–233, 2014. [Online]. Available: <https://www.jstor.org/stable/jeductechsoci.17.2.219>
  12. Y. Zhang, H. Chen, and L. Zhao, "Multimodal interaction in intelligent tutoring systems," *Computers & Education*, vol. 170, p. 104239, 2021. [Online]. Available: <https://doi.org/10.1016/j.compedu.2021.104239>
  13. C.-K. Chang and G.-J. Hwang, "Gamification in language learning: A systematic review," *Journal of Educational Computing Research*, vol. 57, no. 1, pp. 46–74, 2019. [Online]. Available: <https://elt.tabrizu.ac.ir/article/106761300.html?lang=en>
  14. A. Bhardwaj, M. S. Khan, K. Sanil, S. Y. Kotagiri, and L. Pranav, "Ai-powered speech feedback system for public speaking with local language support and confidence evaluation," in *International Conference on Intelligent and Fuzzy Systems*. Springer, 2025, pp. 593–600. [Online]. Available: [https://link.springer.com/chapter/10.1007/978-3-031-98565-2\\_64](https://link.springer.com/chapter/10.1007/978-3-031-98565-2_64)
  15. M. Horsley, M. Eliot, B. A. Knight, and R. Reilly, *Current Trends in Eye Tracking Research*. Springer, 2014. [Online]. Available: <https://link.springer.com/book/10.1007/978-3-319-02868-2>