

ỨNG DỤNG HỆ THỐNG PHẢN HỒI DỰA TRÊN TRÍ TUỆ NHÂN TẠO NHẪM NÂNG CAO ĐỘ CHÍNH XÁC PHÁT ÂM TRONG LỚP HỌC TIẾNG ANH NHƯ MỘT NGOẠI NGỮ

TS. Nguyễn Thanh Tú - Tạ Thị Hoà

Bộ môn ngoại ngữ, Trường Đại học Công nghệ Giao thông vận tải

Email: tunt@utt.edu.vn, hoatt@utt.edu.vn.

Tóm tắt: Năng lực sản xuất lời nói chính xác là yêu cầu nền tảng cho hiệu quả giao tiếp trong tiếng Anh như một Ngoại ngữ (EFL), đặc biệt đối với sinh viên chuyên ngành. Môi trường giảng dạy truyền thống thường không thể cung cấp phản hồi chẩn đoán chi tiết và nhất quán ở cả cấp độ âm vị và siêu âm đoạn, dẫn đến những hạn chế trong việc sửa lỗi phát âm. Nghiên cứu này trình bày việc thiết kế, triển khai và đánh giá thực nghiệm một hệ thống Phản hồi phát âm tự động (A-CPT) tiên tiến, ứng dụng mô hình Học sâu (Deep Learning) và Nhận dạng giọng nói tự động (ASR) để khắc phục thách thức này. Dựa trên thiết kế thực nghiệm tiên và hậu kiểm ngẫu nhiên trên 60 sinh viên chuyên ngành EFL, chúng tôi đo lường sự khác biệt về chỉ số Chính xác phát âm trung bình (MPIA). Kết quả cho thấy nhóm can thiệp sử dụng A-CPT đạt mức tăng trưởng MPIA lớn hơn đáng kể (Cohen's $d = 2.05$) so với nhóm đối chứng (Cohen's $d = 1.25$). Phân tích khẳng định rằng A-CPT, thông qua việc cung cấp chỉ số Goodness of Pronunciation (GOP) chi tiết, đã kích hoạt hiệu quả quá trình Nhận biết lỗi (Noticing) của người học, từ đó thúc đẩy việc tái cấu trúc ngữ âm một cách khoa học.

Từ khóa: Trí tuệ nhân tạo, phản hồi chẩn đoán, ngữ âm ứng dụng, GOP, ASR, EFL chuyên ngành.

Nhận bài: 15/12/2025; Biên tập: 16/12/2025; Phản biện: 22/12/2025; Duyệt đăng: 29/12/2025.

1. Đặt vấn đề

Năng lực phát âm chuẩn xác là một trong những thành tố cốt lõi quyết định mức độ thông hiểu và hiệu quả giao tiếp của người học tiếng Anh như một ngoại ngữ (EFL). Phát âm không chỉ là quá trình tạo âm thanh mà còn là bộ phận quan trọng của năng lực giao tiếp, ảnh hưởng trực tiếp đến khả năng truyền tải ý nghĩa và tính dễ hiểu của thông điệp. Đối với sinh viên chuyên ngành Ngôn ngữ Anh tại các trường đại học, yêu cầu về phát âm không dừng lại ở mức độ thông hiểu tối thiểu mà cần tiệm cận chuẩn mực bản ngữ nhằm đáp ứng các yêu cầu học thuật và nghề nghiệp như giảng dạy, biên phiên dịch hay truyền thông quốc tế. Những sai lệch trong phát âm, đặc biệt là các lỗi ảnh hưởng đến ý nghĩa, không chỉ gây khó khăn cho người nghe mà còn làm suy giảm sự tự tin của người nói, hình thành tâm lý lo lắng về phát âm, từ đó khiến người học né tránh các hoạt động nói và làm chậm quá trình phát triển năng lực giao tiếp.

Trong bối cảnh lớp học truyền thống, việc giảng dạy và luyện tập phát âm thường gặp nhiều hạn chế. Thứ nhất, thời lượng giảng dạy hạn chế và sĩ số lớp học lớn khiến giáo viên khó có thể theo dõi, lắng nghe và sửa lỗi phát âm cho từng sinh viên một cách thường xuyên và cá nhân hóa. Thứ hai, phản hồi phát âm từ giáo viên tuy có giá trị chuyên môn nhưng thường mang tính chủ quan, thiếu nhất quán và khó chẩn đoán sâu các lỗi ở cả cấp độ âm vị và các yếu tố siêu đoạn như trọng âm, ngữ điệu hay nhịp điệu. Việc thiếu phản hồi kịp thời và mang tính chẩn đoán khiến sinh viên khó tự nhận diện điểm yếu, hạn chế khả năng tự theo dõi và tự sửa

lỗi, từ đó ảnh hưởng đến quá trình hình thành thói quen phát âm đúng.

Trước những thách thức đó, nghiên cứu này đề xuất và kiểm nghiệm hiệu quả của hệ thống đào tạo phát âm hỗ trợ máy tính tự động (A-CPT) dựa trên trí tuệ nhân tạo. Hệ thống A-CPT ứng dụng công nghệ học sâu và nhận dạng giọng nói tự động, cho phép phân tích lỗi phát âm một cách khách quan, chính xác và tức thì. Phản hồi tự động đóng vai trò cầu nối trực tiếp giữa sản phẩm ngôn ngữ của người học và chuẩn mực phát âm mục tiêu, hỗ trợ quá trình tự điều chỉnh và cải thiện liên tục. Mục tiêu của nghiên cứu là đánh giá mức độ tác động của A-CPT đối với sự cải thiện độ chính xác phát âm của sinh viên chuyên ngành Ngôn ngữ Anh, qua đó cung cấp bằng chứng khoa học về tiềm năng ứng dụng công nghệ AI trong đào tạo phát âm và ngữ âm ứng dụng hiện nay.

2. Nội dung nghiên cứu

2.1. Cơ sở lý thuyết và công nghệ ASR hiện đại trong chẩn đoán ngữ âm

Nghiên cứu này được xây dựng vững chắc trên hai trụ cột: Cơ sở lý thuyết về Tiếp thu ngôn ngữ thứ hai (SLA) và Cơ sở công nghệ xử lý giọng nói (Speech Processing) ứng dụng Trí tuệ nhân tạo.

2.1.1. Cơ sở Lý thuyết: Giả thuyết nhận biết (Noticing Hypothesis)

Về mặt lý thuyết dạy học, nghiên cứu này áp dụng Giả thuyết nhận biết (Noticing Hypothesis) được đề xuất bởi Richard Schmidt (1990). Giả thuyết này khẳng định rằng quá trình học ngôn ngữ thứ hai (L2) là một quá trình phụ thuộc vào sự chú

ý và nhận thức, trong đó người học phải: (1) Chú ý (Attention): Tập trung vào đầu vào ngôn ngữ chuẩn mực (target L2 input). (2) So sánh (Comparison): So sánh có ý thức giữa sản phẩm ngôn ngữ hiện tại của họ (L2 output) và mục tiêu chuẩn mực. (3) Nhận biết (Noticing): Nhận biết được khoảng cách \$ (the \ mismatch \ or \ gap) \$ giữa hai yếu tố trên.

Chỉ những gì được nhận biết một cách có ý thức mới có thể trở thành đầu vào tiếp nhận (intake), là tiền đề cho quá trình xử lý và tiếp thu ngôn ngữ (Hasan & Al-Qarni, 2020). Trong lĩnh vực Ngữ âm ứng dụng, thách thức lớn nhất là các lỗi phát âm (mispronunciations) thường là những sự sai lệch âm học tinh tế và mơ hồ về mặt thính giác, gây khó khăn cho việc tự nhận biết. Phản hồi chẩn đoán chi tiết và khách quan từ hệ thống A-CPT đóng vai trò là \$ \text{xúc tác nhận thức (cognitive catalyst)} \$, chuyển đổi dữ liệu âm thanh phức tạp thành: a) Dữ liệu thị giác khách quan (GOP scores, mã hóa màu). b) Dữ liệu định lượng có thể hành động (actionable quantitative data).

Quá trình này giúp người học xác định chính xác điểm yếu của mình, từ đó kích hoạt quá trình tái cấu trúc ngữ âm (phonological restructuring) một cách có chủ đích, trực tiếp hỗ trợ cho Giả thuyết nhận biết.

2.1.2. Cơ sở Công nghệ: ASR, căn chỉnh bắt buộc và Chỉ số Goodness of Pronunciation (GOP)

Nền tảng công nghệ của A-CPT là khả năng phân tích tiên tiến của hệ thống Nhận dạng giọng nói tự động (Automatic Speech Recognition - ASR), đặc biệt là chức năng Đánh giá phát âm chẩn đoán (Diagnostic Pronunciation Assessment).

Căn chỉnh bắt buộc (Forced Alignment): Để đánh giá phát âm chi tiết, A-CPT phải sử dụng quy trình Căn chỉnh bắt buộc (Forced Alignment). Quy trình này sử dụng Thuật toán Viterbi (Viterbi Algorithm) để tìm ra chuỗi trạng thái (sequence of states) có khả năng xảy ra cao nhất, đồng bộ hóa đoạn âm thanh được thu âm $\mathbf{O} = (o_1, o_2, \dots, o_T)$ của người học với chuỗi âm vị tham chiếu đã biết $\mathbf{P} = (p_1, p_2, \dots, p_N)$ theo trục thời gian. Kết quả là việc xác định chính xác vị trí thời gian bắt đầu và kết thúc của từng âm vị ($\text{start}_i, \text{end}_i$) trong luồng lời nói.

Chỉ số Goodness of Pronunciation (GOP): Sau khi căn chỉnh bắt buộc, chất lượng của từng âm vị p_i được định lượng hóa bằng Chỉ số Goodness of Pronunciation (GOP). Về mặt kỹ thuật, GOP được định nghĩa là log-likelihood của quan sát âm thanh tại thời điểm t cho âm vị p_i (theo mô hình âm học Target Acoustic Model - TAM), chuẩn hóa bằng một tham số tối ưu hóa hoặc so sánh với một mô hình thay thế.

2.1.3. Vai trò của Kiến trúc Học sâu Transformer trong Phân tích lỗi phát âm (PEA)

Trong những năm gần đây, Học sâu (Deep

Learning), đặc biệt là kiến trúc Transformer Sequence-to-Sequence (Seq2Seq), đã thay thế các mô hình dựa trên HMM truyền thống, vốn chỉ dựa vào Mô hình hỗn hợp Gauss (Gaussian Mixture Models - GMMs) và bị giới hạn trong việc mô hình hóa các đặc trưng âm học phức tạp và phụ thuộc ngữ cảnh.

Xử lý phụ thuộc dài hạn (Long-Range Dependencies): Kiến trúc Transformer, với cơ chế Self-Attention (Tự Chú ý), cho phép mô hình xử lý toàn bộ chuỗi đầu vào (features âm học) cùng một lúc. Điều này khác biệt so với RNNs (Mạng Nơ-ron Hồi quy) cần xử lý tuần tự, giúp Transformer vượt trội trong việc nắm bắt các mối quan hệ ngữ cảnh và phụ thuộc dài hạn trong lời nói.

Mô hình hóa Siêu Âm đoạn: Khả năng này là cực kỳ quan trọng đối với phân tích Lỗi Phát âm (PEA), đặc biệt là các yếu tố Siêu âm đoạn (Suprasegmental Features) như trọng âm từ, trọng âm câu, ngữ điệu (Intonation), và nhịp điệu (Rhythm). Các lỗi siêu âm đoạn thường là lỗi toàn cục (global errors) ảnh hưởng đến một đoạn lời nói dài, và chỉ có các kiến trúc Deep Learning tiên tiến như Transformer mới có thể mô hình hóa chúng với độ chính xác cao.

Hiệu suất và Độ bền (Robustness): Mô hình Transformer được huấn luyện trên kho ngữ liệu lớn (bao gồm lời nói chuẩn và lời nói bị lỗi) có thể trích xuất các đặc trưng âm học phong phú hơn (ví dụ: các biến thể của MFCCs hoặc Filter Bank features), dẫn đến chỉ số GOP ổn định và chính xác hơn, ngay cả khi đối mặt với các lỗi phát âm chuyển giao (L1-transfer errors) phức tạp.

Sự tích hợp của ASR dựa trên Transformer và chỉ số GOP đã tạo nên một hệ thống A-CPT có khả năng cung cấp phản hồi chẩn đoán chi tiết và đáng tin cậy ở cả cấp độ âm vị (segmental) và siêu âm đoạn (suprasegmental), là chìa khóa cho hiệu quả thực nghiệm của nghiên cứu này.

2.2. Phương pháp nghiên cứu thực nghiệm và thiết kế hệ thống A-CPT

Nghiên cứu áp dụng thiết kế thực nghiệm tiền kiểm - hậu kiểm ngẫu nhiên (Randomized Pretest-Posttest Control Group Design) kéo dài 8 tuần, một thiết kế mạnh mẽ nhằm đảm bảo tính nội giá trị (internal validity) cao.

Đối tượng nghiên cứu: 60 SV năm thứ ba chuyên ngành NNA tại một trường ĐH lớn. Đối tượng này được chọn vì họ có trình độ ngôn ngữ trung cấp đến cao cấp và có nhu cầu cải thiện phát âm để đạt được mục tiêu học thuật chuyên sâu. Các SV được phân bổ ngẫu nhiên (random assignment) thành hai nhóm: Nhóm can thiệp (N=30) và Nhóm đối chứng (N=30).

Quy trình can thiệp:

- **Nhóm Can thiệp (A-CPT):** Thực hành phát âm tự học 3 giờ/tuần bằng hệ thống A-CPT do nhóm nghiên cứu phát triển. Hệ thống này cung cấp các

bài tập được thiết kế riêng biệt để xử lý các lỗi chuyển giao (L1-transfer errors) phổ biến của người Việt (ví dụ: thiếu âm cuối, phân biệt /l/-/r/, /s/-/ʃ/).

- **Nhóm Đối chứng:** Thực hành tự học 3 giờ/tuần bằng các tài liệu truyền thống (sách giáo trình, audio luyện nghe-nhắc lại) mà không có sự can thiệp của bất kỳ công cụ phản hồi tự động nào.

- **Đo lường:** Chỉ số chính xác phát âm trung bình (MPIA) được sử dụng làm biến phụ thuộc chính. MPIA là điểm đánh giá khách quan, độc lập từ ba giáo viên bản ngữ được chứng nhận TESOL/CELTA và có kinh nghiệm giảng dạy EFL tối thiểu 5 năm. Điểm số được tính trên thang 100, tập trung đánh giá cả âm vị (segmental features) và các yếu tố siêu âm đoạn (suprasegmental features) trong các đoạn hội thoại và bài đọc.

- **Thiết kế Công nghệ A-CPT:** Hệ thống được xây dựng trên Kiến trúc Transformer Seq2Seq, được huấn luyện trên một kho ngữ liệu lớn bao gồm cả lời nói chuẩn và lời nói bị lỗi (mispronunciation corpus).

Cơ chế Phản hồi Âm vị: Sử dụng GOP để đánh giá từng âm vị. Phản hồi được trực quan hóa bằng Mã hóa Màu trên IPA (International Phonetic Alphabet): Xanh (chính xác), Vàng (cần cải thiện), Đỏ (lỗi nghiêm trọng).

Cơ chế Phản hồi Siêu Âm đoạn: Trực quan hóa Pitch Contour (F0) và Energy Contour của người học, được vẽ chồng lên đường tham chiếu của người bản xứ. Sự so sánh trực quan này giúp người học điều chỉnh trọng âm từ, trọng âm câu và ngữ điệu một cách chính xác.

2.3. Phân tích kết quả định lượng và thảo luận chuyên sâu

Sau giai đoạn can thiệp 8 tuần, các dữ liệu hậu kiểm được thu thập và phân tích bằng các phương pháp thống kê định lượng. Độ tin cậy Liên giám khảo (Inter-rater Reliability) đã được kiểm tra bằng thống kê Kappa của Cohen và đạt mức Kappa = 0.85, khẳng định tính thống nhất cao và khách quan của dữ liệu đánh giá từ con người.

Phân tích Thống kê và Kết quả: Cả hai nhóm đều có mức tăng trưởng MPIA sau can thiệp, nhưng sự khác biệt là rất lớn.

Nhóm Đối chứng: Điểm tiền kiểm là 65.5 ± 3.9 , tăng lên 70.8 ± 4.5 ở hậu kiểm, với mức cải thiện là +5.3 điểm. Kiểm định *t* cho thấy sự cải thiện này có ý nghĩa thống kê ($t(58) = 8.12, p < 0.001$) và kích thước ảnh hưởng là Cohen's $d = 1.25$ (Mức ảnh hưởng Lớn).

Nhóm Can thiệp (A-CPT): Điểm tiền kiểm là 65.2 ± 4.1 , tăng lên 79.5 ± 3.8 ở hậu kiểm, với mức cải thiện đột phá là +14.3 điểm. Kiểm định *t* độc lập giữa hai nhóm xác nhận sự khác biệt này là đáng kể ($t(58) = 4.88, p < 0.001$). Quan trọng nhất, kích thước ảnh hưởng của can thiệp A-CPT là Cohen's $d = 2.05$, được phân loại là mức ảnh hưởng "Rất Lớn"

(Very Large Effect Size) theo tiêu chuẩn của Cohen.

Phân tích sâu hơn về mức cải thiện theo thành phần Ngữ âm: Để hiểu rõ hơn về nguồn gốc của sự cải thiện vượt trội này, chúng tôi tiến hành phân tích phương sai đơn biến (ANOVA) trên hai thành phần ngữ âm chính: Độ chính xác Âm vị (Segmental Accuracy) và Độ chính xác Siêu Âm đoạn (Suprasegmental Accuracy). Kết quả chỉ ra rằng, trong khi cả hai nhóm đều cải thiện độ chính xác âm vị (Âm vị Nhóm Can thiệp cải thiện +8.5 điểm so với +4.0 điểm của nhóm đối chứng), sự khác biệt lớn nhất nằm ở thành phần siêu âm đoạn. Nhóm Can thiệp có mức tăng trưởng độ chính xác siêu âm đoạn là +5.8 điểm, trong khi nhóm Đối chứng chỉ là +1.3 điểm. Sự khác biệt này là rất quan trọng vì các yếu tố siêu âm đoạn (như trọng âm và ngữ điệu) thường khó sửa nhất trong quá trình học ngôn ngữ thứ hai và là nơi mà phản hồi của GV truyền thống thường thiếu chi tiết nhất. Việc trực quan hóa Pitch Contour do A-CPT cung cấp đã chứng minh vai trò quyết định trong việc giúp người học làm chủ các khía cạnh ngữ điệu phức tạp này.

Thảo luận về cơ chế cải thiện và giả thuyết nhận biết: Kết quả thống kê cung cấp bằng chứng thực nghiệm mạnh mẽ chứng minh tính hiệu quả vượt trội của A-CPT. Sự khác biệt +9.0 điểm so với nhóm đối chứng được giải thích thông qua cơ chế phản hồi công nghệ cao. Phản hồi chẩn đoán dựa trên GOP và Trực quan hóa Pitch Contour đã trực tiếp kích hoạt quá trình Nhận biết (Noticing) lỗi của người học. SV chuyên ngành NNA, với khả năng nhận thức ngôn ngữ cao, có thể sử dụng dữ liệu GOP mã hóa màu để: (1) Định vị lỗi chính xác đến từng âm vị, chuyển đổi lỗi thính giác thành một điểm dữ liệu thị giác khách quan; (2) Thực hiện tái cấu trúc ngữ âm tức thì (micro-adjustments) thông qua luyện tập lặp đi lặp lại có mục tiêu. Sự can thiệp lặp đi lặp lại này, được gọi là Luyện tập phân tán có mục tiêu (Targeted Distributed Practice), đã củng cố các kết nối thần kinh (neural pathways) cần thiết cho việc sản xuất âm thanh chính xác. Tính liên tục và tức thì của phản hồi AI tạo điều kiện cho việc sửa lỗi diễn ra trong khung thời gian ngắn nhất sau khi lỗi được tạo ra...

Phản hồi từ A-CPT có ba lợi thế quyết định: Tính khách quan (dựa trên thuật toán GOP thay vì cảm tính giáo viên), Tính tức thì (ngay lập tức sau khi nói) và Tính không đánh giá cá nhân (non-judgmental), cho phép người học thực hành với tần suất cao hơn mà không gặp rào cản tâm lý lo lắng về phát âm. Sự kết hợp của các yếu tố này đã tối ưu hóa chu trình luyện tập, từ đó dẫn đến mức tăng trưởng MPIA ấn tượng. Sự thành công của A-CPT cũng là một sự khẳng định thực nghiệm về tính ứng dụng cao của các mô hình Deep Learning mới nhất trong lĩnh vực giáo dục ngôn ngữ.

3. Kết luận

Nghiên cứu này đã thành công trong việc thiết kế, triển khai và đánh giá thực nghiệm hệ thống Phản hồi phát âm tự động (A-CPT) ứng dụng Trí tuệ nhân tạo dựa trên kiến trúc Transformer Seq2Seq và chỉ số Goodness of Pronunciation (GOP). Kết quả định lượng đã xác nhận hiệu quả đột phá của A-CPT, thể hiện qua kích thước ảnh hưởng “Rất Lớn” (Very Large Effect Size) được đo lường bằng Cohen’s $d = 2.05$ đối với sự cải thiện Chỉ số chính xác phát âm trung bình (MPIA) của SV chuyên ngành NNA. Sự khác biệt này không chỉ có ý nghĩa thống kê mà còn mang tính sự phạm sâu sắc, đặc biệt khi sự cải thiện tập trung đáng kể vào các yếu tố siêu âm đoạn phức tạp (trọng âm, ngữ điệu), vốn là điểm yếu cố hữu trong quá trình học tập truyền thống.

Đóng góp chính của nghiên cứu là cung cấp bằng chứng thực nghiệm (empirical evidence) mạnh mẽ và một mô hình ứng dụng công nghệ thành công trong việc giải quyết thách thức lâu dài về phản hồi chẩn đoán ngữ âm trong môi trường EFL. Nghiên cứu không chỉ định lượng hóa mức độ hiệu quả của can thiệp đối với nhóm người học có yêu cầu cao về chuyên môn ngôn ngữ mà còn củng cố mô hình lý thuyết về Giả thuyết Nhận biết (Noticing Hypothesis) trong bối cảnh học tập có hỗ trợ AI, minh chứng rằng phản hồi chẩn đoán khách quan, tức thì là xúc tác mạnh mẽ nhất cho quá trình tái cấu trúc ngữ âm.

Dựa trên những thành tựu này, hướng nghiên cứu tương lai nên tập trung vào ba lĩnh vực chính nhằm tối ưu hóa và mở rộng phạm vi ứng dụng của A-CPT: (1) Phân tích định tính và Tinh chỉnh mô hình PEA: Tiến hành phân tích định tính lỗi sâu hơn để tinh chỉnh mô hình PEA (Pronunciation Error Analysis) nhằm nhận diện các lỗi âm vị và siêu âm đoạn đặc trưng theo ngữ cảnh (ví dụ: các biến thể

của nguyên âm trong các môi trường ngữ âm khác nhau) và các lỗi chuyển giao (L1-transfer errors) tinh vi hơn. (2) Đánh giá Tính Hiệu quả Dài hạn (Retention and Transferability): Đánh giá Tính hiệu quả dài hạn (retention) của sự cải thiện phát âm sau khi dừng can thiệp A-CPT, đồng thời nghiên cứu khả năng người học chuyển giao (transfer) các kỹ năng ngữ âm được cải thiện sang các tình huống giao tiếp tự nhiên, không có sự hỗ trợ của công nghệ. (3) Tích hợp ngữ cảnh và giao tiếp thực tế: Phát triển các mô hình A-CPT có khả năng tích hợp ngữ cảnh (Context-Aware Models) để không chỉ đánh giá độ chính xác âm học mà còn đánh giá sự phù hợp ngữ âm trong bối cảnh giao tiếp thực tế, bao gồm việc sử dụng các hiện tượng liên kết âm (linking sounds) và giảm âm (reduction) một cách tự nhiên. Việc tích hợp sâu hơn A-CPT vào chương trình giảng dạy ĐH sẽ là bước đi chiến lược nhằm nâng cao chất lượng đào tạo ngôn ngữ trong kỷ nguyên công nghệ ■

Tài liệu tham khảo

- [1]. Hasan, M. S., & Al-Qarni, A. I. (2020). *Revisiting Schmidt’s Noticing Hypothesis in AI-mediated language learning environments*. *System*, 95, 102371.
- [2]. Jones, S. A., & Chen, L. (2023). *Developing contextual appropriateness assessment in ASR for non-native speech*. *Language Learning & Technology*, 27 (1), 101-120.
- [3]. Li, J., & Kim, D. (2019). *The impact of automated corrective feedback on L2 English speaking fluency and accuracy*. *Journal of Educational Technology & Society*, 22 (4), 1 - 13.
- [4]. Zhang, Y., & Wang, Q. (2022). *Deep learning-based ASR for L2 English pronunciation evaluation: A review and perspective*. *Computer Assisted Language Learning*, 35 (5 - 6), 1101 - 1125.

Applying an artificial intelligence-based feedback system to enhance pronunciation accuracy in English as a foreign language classrooms

Dr. Nguyen Thanh Tu - Ta Thi Hoa

Subject of Foreign Languages, University of Transport Technology

Email: tunt@utt.edu.vn, hoatt@utt.edu.vn.

Abstract: *Accurate speech production is a foundational requirement for effective communication in English as a Foreign Language (EFL), especially for students majoring in*

English. Traditional instructional settings frequently fail to deliver detailed and consistent diagnostic feedback at both the segmental and suprasegmental levels, leading to limitations in pronunciation error correction. This study presents the design, implementation, and experimental evaluation of an advanced Automatic Speech Response (A-CPT) system, applying Deep Learning and Automatic Speech Recognition (ASR) models to overcome this challenge. Based on a randomized pretest-posttest experimental design involving 60 EFL majors, we measured the differential gain in the Mean Pronunciation Accuracy Index (MPIA). The results demonstrate that the A-CPT intervention group achieved a significantly greater increase in MPIA (Cohen’s $d = 2.05$) compared to the control group (Cohen’s $d = 1.25$). This analysis confirms that A-CPT, by providing a detailed Goodness of Pronunciation (GOP) index, effectively activated the learner’s error Noticing process, thereby promoting systematic phonological restructuring.

Keywords: *Artificial intelligence, diagnostic feedback, applied phonology, GOP, ASR, EFL majors.*