

HIỆU QUẢ CỦA GOOGLE VOICE TYPING ĐỐI VỚI VIỆC CẢI THIỆN PHÁT ÂM TIẾNG ANH CỦA SINH VIÊN ĐẠI HỌC

Đặng Thị Hương Thảo¹

Email: dththao@hou.edu.vn, ORCID: 0009-0001-5630-6389

Ngày tòa soạn nhận được bài báo: 15/03/2026. Ngày phản biện đánh giá: 15/05/2026.

Ngày bài báo được duyệt đăng: 01/06/2026

DOI: 10.59266/houjs.2026.1277

Tóm tắt: Nghiên cứu này nhằm đánh giá hiệu quả Google Voice Typing (GVT), một công cụ chuyển lời nói thành văn bản dựa trên công nghệ nhận diện giọng nói tự động, đối với việc cải thiện phát âm tiếng Anh của sinh viên. Nghiên cứu được thực hiện với 50 sinh viên năm thứ hai chuyên ngành Ngôn ngữ Anh tại Trường Đại học Mở Hà Nội. Phương pháp nghiên cứu kết hợp định lượng và định tính được sử dụng để thu thập và phân tích dữ liệu. Dữ liệu định lượng được thu thập thông qua bảng hỏi gồm 14 mục sử dụng thang đo Likert 5 mức độ và số lượng lỗi phát âm của sinh viên trong các lần luyện tập. Dữ liệu định tính được thu thập thông qua câu hỏi mở và các lỗi phát âm cụ thể từ các bản ghi âm của 28 sinh viên. Kết quả cho thấy Google Voice Typing giúp sinh viên cải thiện phát âm và sinh viên có đánh giá tích cực về hiệu quả của việc sử dụng công cụ này trong luyện phát âm. Nghiên cứu cho thấy việc tích hợp công nghệ nhận diện giọng nói vào hoạt động luyện phát âm có thể góp phần nâng cao hiệu quả học tập tiếng Anh.

Từ khóa: Google Voice Typing, công nghệ nhận diện giọng nói tự động, phát âm tiếng Anh, hoạt động luyện phát âm, học tiếng Anh

I. Đặt vấn đề

Trong bối cảnh toàn cầu hóa và sự phát triển của công nghệ 4.0, việc ứng dụng công nghệ vào dạy và học tiếng Anh ngày càng trở nên phổ biến. Tuy nhiên, phát âm vẫn là một trong những thách thức lớn đối với người học do yêu cầu sự phối hợp chính xác giữa các yếu tố ngữ âm. Nhiều nghiên cứu cho thấy sinh viên thường gặp khó khăn trong việc phát âm chính xác do ảnh hưởng của tiếng mẹ đẻ

(Rosyid, 2016), thiếu môi trường thực hành và hạn chế về phản hồi kịp thời từ giáo viên (Arifah & Budiastuti, 2025). Bên cạnh đó, Gilakjani (2012) xác định bốn yếu tố ảnh hưởng đến việc học phát âm, bao gồm thái độ, động lực và mức độ tiếp xúc với tiếng Anh, hoạt động giảng dạy, và sự tiếp xúc với ngôn ngữ đích.

Sự phát triển của công nghệ và trí tuệ nhân tạo đã mở ra nhiều cơ hội hỗ trợ cải thiện phát âm. Leis (2025) cho thấy

¹ Trường Đại học Mở Hà Nội, Hà Nội, Việt Nam

công nghệ chuyên giọng nói thành văn bản có thể cải thiện đáng kể độ chính xác phát âm và nâng cao nhận thức của người học về lỗi sai. Tương tự, Dennis (2024) khẳng định công nghệ nhận diện giọng nói tích hợp AI có tác động tích cực đến phát âm và kỹ năng nói của người học tiếng Anh như một ngoại ngữ. Tuy nhiên, các nghiên cứu về hiệu quả của Google Voice Typing (GVT) trong bối cảnh giáo dục đại học vẫn còn hạn chế.

Phát âm đóng vai trò quan trọng trong giao tiếp vì ảnh hưởng trực tiếp đến mức độ dễ hiểu và hiệu quả truyền đạt thông tin (Dang, 2024). Mặc dù vậy, thời lượng luyện tập phát âm trên lớp còn hạn chế, đòi hỏi các giải pháp hỗ trợ ngoài giờ học. Xuất phát từ thực tiễn này, nghiên cứu nhằm đánh giá hiệu quả của GVT trong việc cải thiện phát âm tiếng Anh của sinh viên, đồng thời phân tích nhận thức của người học về việc sử dụng công cụ này trong quá trình luyện tập.

II. Cơ sở lý thuyết

2.1. Phát âm trong việc học ngoại ngữ

2.1.1. Khái niệm phát âm

Theo Kelly (2000), phát âm không chỉ bao gồm việc tạo ra các âm riêng lẻ mà còn bao hàm các yếu tố siêu đoạn tính như trọng âm, ngữ điệu và nhịp điệu, vốn ảnh hưởng trực tiếp đến mức độ dễ hiểu trong giao tiếp. Nhiều nhà nghiên cứu cũng cho rằng phát âm là quá trình tạo ra âm thanh nhằm truyền tải ý nghĩa. Từ góc độ ngôn ngữ học, phát âm là một thành phần cốt lõi của năng lực giao tiếp. Việc kiểm soát tốt cả yếu tố âm đoạn và siêu đoạn giúp người học truyền đạt thông điệp rõ ràng, chính xác và hiệu quả hơn.

2.1.2. Các yếu tố ảnh hưởng đến việc cải thiện phát âm

Nhiều nghiên cứu trong lĩnh vực giảng dạy tiếng Anh như một ngoại ngữ cho thấy việc cải thiện phát âm chịu ảnh hưởng bởi nhiều yếu tố. Trước hết, luyện tập thường xuyên được xem là yếu tố then chốt, giúp người học hình thành thói quen phát âm chính xác và nâng cao khả năng kiểm soát các đặc điểm âm vị (Derwing & Munro, 2005). Bên cạnh đó, phản hồi kịp thời từ giáo viên hoặc các công cụ hỗ trợ học tập đóng vai trò quan trọng trong việc điều chỉnh phát âm. Nhờ phản hồi, người học có thể nhận ra sự khác biệt giữa cách phát âm của mình và chuẩn ngôn ngữ đích, từ đó cải thiện một cách hiệu quả.

Ngoài ra, việc tiếp xúc với các nguồn tài liệu xác thực như phim, podcast và bài nói của người bản ngữ cũng góp phần nâng cao khả năng phát âm và sự tự tin trong giao tiếp (Arifah & Budiastuti, 2025). Các nghiên cứu trong lĩnh vực ngữ âm học ứng dụng đã khẳng định rằng ảnh hưởng của tiếng mẹ đẻ là một trong những yếu tố quan trọng tạo nên sự khác biệt trong phát âm tiếng Anh của người học ngoại ngữ. Những khác biệt về đặc điểm phát âm và giọng điệu chịu ảnh hưởng của L1 có thể làm suy giảm tính dễ hiểu của lời nói và gây khó khăn cho quá trình xử lý ngôn ngữ của người nghe, đặc biệt trong các bối cảnh giao tiếp không quen thuộc (Munro & Derwing, 2011; Bent & cộng sự, 2024). Như vậy, việc cải thiện phát âm không chỉ phụ thuộc vào năng lực cá nhân mà còn chịu ảnh hưởng bởi môi trường học tập, cơ hội luyện tập và các yếu tố ngôn ngữ học.

2.2. Công nghệ nhận diện giọng tự động nói trong dạy học ngôn ngữ

2.2.1. Khái niệm công nghệ nhận diện giọng nói tự động

Trong những năm gần đây, công nghệ nhận diện giọng nói tự động (Automatic Speech Recognition - ASR) đã thu hút sự quan tâm đáng kể trong lĩnh vực giảng dạy ngoại ngữ. Nhiều nghiên cứu đã xem xét tác động của công nghệ này đối với năng lực nói của người học (Evers & Chen, 2020; Jiang & cộng sự, 2022). ASR là công nghệ dựa trên máy tính nhằm chuyển đổi ngôn ngữ nói thành văn bản, thuộc lĩnh vực trí tuệ nhân tạo và xử lý ngôn ngữ tự nhiên, và được tích hợp rộng rãi trong các ứng dụng như trợ lý ảo Siri hay hệ thống nhận diện giọng nói (Evers & Chen, 2020).

Các nghiên cứu thực nghiệm đã chỉ ra rằng ASR có thể hỗ trợ hiệu quả cho việc học ngôn ngữ (Jiang & cộng sự, 2021; Ranchal & cộng sự, 2013). Công nghệ này mang lại nhiều lợi ích, đặc biệt trong việc nâng cao kỹ năng nói và hỗ trợ người học cải thiện độ chính xác ngôn ngữ.

2.2.2. Ứng dụng Google Voice Typing trong việc cải thiện phát âm tiếng Anh

Nghiên cứu của Nguyễn và cộng sự (2025) cho thấy việc sử dụng GVT có thể hỗ trợ cải thiện phát âm cho sinh viên chuyên ngành tiếng Anh, với sự tiến bộ rõ rệt ở nguyên âm đơn, nguyên âm đôi, nguyên âm ba và các phụ âm phức tạp. Tương tự, các nghiên cứu khác cũng khẳng định hiệu quả của công nghệ nhận diện giọng nói dựa trên AI trong việc nâng cao phát âm (Abimanto & Sumarsono, 2024; Hoang & cộng sự, 2025).

Một lợi ích quan trọng của GVT là cung cấp phản hồi tức thì. Khi phát âm chính xác, hệ thống chuyển đổi lời nói thành văn bản đúng; ngược lại, lỗi phát âm dẫn đến kết quả nhận diện sai, giúp người học nhận ra và điều chỉnh. Bên cạnh đó, công cụ này tạo điều kiện cho việc luyện tập lặp lại, giúp hình thành thói quen phát âm đúng và cải thiện độ chính xác theo thời gian.

III. Phương pháp nghiên cứu

Nghiên cứu áp dụng phương pháp hỗn hợp (định lượng và định tính) với thiết kế đo lặp nhằm đánh giá hiệu quả của GVT đối với việc cải thiện phát âm tiếng Anh của sinh viên. Dữ liệu định lượng được thu thập thông qua bảng hỏi gồm 14 mục theo thang Likert 5 mức độ và thống kê lỗi phát âm từ các bản ghi âm. Bảng hỏi được cấu trúc thành ba tiểu thang đo: (i) nhận thức về sự cải thiện phát âm (6 mục; Cronbach's Alpha = 0,94), (ii) nhận thức về hiệu quả của GVT (5 mục; Cronbach's Alpha = 0,97), và (iii) mức độ tham gia vào quá trình luyện tập (3 mục; Cronbach's Alpha = 0,90), cho thấy độ tin cậy nội tại cao. Dữ liệu định tính được thu thập từ 4 câu hỏi mở nhằm làm rõ trải nghiệm học tập và các dạng lỗi phát âm.

Đối tượng nghiên cứu gồm 50 sinh viên năm thứ hai chuyên ngành Ngôn ngữ Anh tại Trường Đại học Mở Hà Nội, đang theo học học phần Nói 4 và chưa từng sử dụng GVT cho mục đích luyện phát âm. Nghiên cứu được triển khai trong 8 tuần, trong đó sinh viên được hướng dẫn sử dụng công cụ và thực hành với ngữ liệu phù hợp trình độ.

Dữ liệu phát âm được thu thập thông qua nhiệm vụ đọc một đoạn văn có độ dài cố định 362 từ ở cả hai thời điểm trước

và sau can thiệp. Số lỗi phát âm được xác định là tổng số lỗi trong mỗi bài ghi âm. Lỗi phát âm được xác định dựa trên các tiêu chí thống nhất, bao gồm lỗi nguyên âm đơn, nguyên âm đôi, phụ âm cuối và cụm phụ âm. Mỗi lỗi được tính là một đơn vị lỗi, trong đó mỗi từ chỉ được tính tối đa một lỗi nhằm đảm bảo tính nhất quán trong quá trình đánh giá. Các lỗi không liên quan đến phát âm không được đưa vào phân tích. Trong số 50 sinh viên, 28 sinh viên đáp ứng đầy đủ tiêu chí về dữ liệu và chất lượng ghi âm được đưa vào

phân tích. Các bản ghi âm được hai giảng viên đánh giá dựa trên cùng tiêu chí. Kết quả được đối chiếu và thống nhất sau khi chấm. Do dữ liệu ở dạng tổng số lỗi trên mỗi bài ghi âm, nghiên cứu không sử dụng hệ số Cohen's Kappa.

Dữ liệu được xử lý bằng SPSS 25, sử dụng thống kê mô tả và kiểm định t cho mẫu ghép. Cỡ ảnh hưởng được ước lượng bằng hệ số Cohen's dz. Tuy nhiên, việc chỉ sử dụng dữ liệu của 28 sinh viên có thể dẫn đến sai lệch mẫu, do đó kết quả cần được diễn giải thận trọng.

IV. Kết quả và thảo luận

4.1. Kết quả phân tích lỗi phát âm tiếng Anh từ các bản ghi âm

Bảng 1. Bảng thống kê lỗi phát âm tiếng Anh của sinh viên

| Thời điểm luyện tập | Số lượng | Giá trị nhỏ nhất | Giá trị lớn nhất | Trung bình | Độ lệch chuẩn |
|--------------------------------|----------|------------------|------------------|------------|---------------|
| Lỗi phát âm luyện tập lần đầu | 28 | 9 | 39 | 24,04 | 8,82 |
| Lỗi phát âm luyện tập lần cuối | 28 | 4 | 28 | 15,82 | 6,87 |

Kết quả cho thấy số lỗi phát âm trung bình của sinh viên giảm từ 24,04 lỗi ở lần ghi âm ban đầu xuống còn 15,82 lỗi ở lần ghi âm cuối. Sự giảm này phản ánh xu hướng cải thiện rõ rệt về độ chính xác phát âm sau quá trình luyện tập.

Do tất cả sinh viên đều thực hiện cùng một đoạn văn có độ dài cố định 362 từ, số lỗi tuyệt đối có thể được so sánh trực tiếp giữa hai thời điểm. Khi quy đổi tương đối, số lỗi trung bình giảm từ khoảng 6,64 lỗi/100 từ xuống còn 4,37 lỗi/100 từ, cho thấy mức cải thiện nhất quán.

Bảng 2. Kiểm định sự khác biệt giữa lỗi phát âm trước và sau luyện tập

| Cặp so sánh | Sự khác biệt về lỗi giữa hai bản ghi âm | | | | | t | df | p |
|---|---|---------------|--------------|--------------------|----------|------|----|--------|
| | Trung bình chênh lệch | Độ lệch chuẩn | Sai số chuẩn | Khoảng tin cậy 95% | | | | |
| | | | | Cận dưới | Cận trên | | | |
| Lỗi phát âm luyện tập lần đầu Lỗi phát âm luyện tập lần cuối | 8,21 | 4,40 | 0,83 | 6,51 | 9,92 | 9,88 | 27 | <0,001 |

Kết quả kiểm định t cho mẫu ghép cho thấy số lỗi phát âm của sinh viên giảm đáng kể sau quá trình luyện tập, với sự khác biệt có ý nghĩa thống kê giữa hai lần ghi âm ($t(27) = 9,88$, $p < 0,001$).

Bên cạnh đó, cỡ ảnh hưởng được ước lượng bằng hệ số Cohen's dz cho thiết kế đo lặp. Kết quả cho thấy Cohen's dz = 1,87, phản ánh mức ảnh hưởng rất lớn theo phân loại của Cohen (1988), cho

thấy sự cải thiện phát âm không chỉ có ý nghĩa thống kê mà còn có ý nghĩa thực tiễn rõ rệt.

Kết quả phân tích bản ghi âm cho thấy sinh viên gặp nhiều khó khăn trong phát âm, đặc biệt ở nguyên âm đôi, nguyên âm đơn và phụ âm cuối. Thứ nhất, người học thường gặp khó khăn với nguyên âm đôi như /ei/, /ai/ và /əʊ/, có xu hướng đơn giản hóa thành nguyên âm đơn, dẫn đến sai nghĩa từ. Nguyên nhân chủ yếu là do sự khác biệt giữa hệ thống nguyên âm tiếng Anh và tiếng Việt. Thứ hai, các lỗi liên quan đến nguyên âm đơn cũng

phổ biến, đặc biệt là sự nhầm lẫn giữa các cặp âm gần nhau như /i/ - /i:/ và /ʊ/ - /u:/, làm giảm độ rõ ràng trong giao tiếp (ví dụ: *live* → *leave*, *full* → *fool*). Bên cạnh đó, việc lược bỏ hoặc phát âm không đầy đủ phụ âm cuối như /t/, /s/, /z/ diễn ra khá thường xuyên (ví dụ: *packed* → *pack*, *starts* → *start*). Ngoài ra, sinh viên cũng gặp khó khăn với cụm phụ âm như /kr/, /kt/ (ví dụ: *crowd* → *cowd*, *looked* → *look*). Tuy nhiên, việc sử dụng GVT cung cấp phản hồi tức thì, giúp người học nhận diện và điều chỉnh lỗi phát âm, từ đó cải thiện độ chính xác theo thời gian.

4.2. Kết quả khảo sát về nhận thức của sinh viên đối với việc sử dụng Google Voice Typing

Bảng 3. Nhận thức của sinh viên về việc cải thiện phát âm

| Mục | Giá trị nhỏ nhất | Giá trị lớn nhất | Trung bình | Độ lệch chuẩn |
|---|------------------|------------------|------------|---------------|
| Khả năng phát âm tổng thể của tôi được cải thiện sau khi sử dụng GVT. | 1 | 5 | 3,60 | 0,81 |
| Phát âm của tôi trở nên rõ ràng hơn đối với người nghe. | 1 | 5 | 3,68 | 0,82 |
| Tôi phát âm các âm tiếng Anh chính xác hơn. | 1 | 5 | 3,76 | 0,85 |
| Trọng âm từ (word stress) của tôi được cải thiện. | 1 | 5 | 3,74 | 0,88 |
| Trọng âm câu và nhịp điệu khi nói của tôi được cải thiện. | 1 | 5 | 3,68 | 0,87 |
| Tôi cảm thấy tự tin hơn khi nói tiếng Anh. | 1 | 5 | 3,72 | 0,83 |

Kết quả ở Bảng 3 cho thấy sinh viên có nhận thức tích cực về việc cải thiện phát âm khi sử dụng GVT, với điểm trung bình dao động từ 3,60 đến 3,76 trên thang Likert 5 mức. Tiêu chí “Tôi phát âm các âm tiếng Anh chính xác hơn”

đạt giá trị cao nhất ($M = 3,76$, $SD = 0,85$), phản ánh sự cải thiện rõ rệt về độ chính xác. Độ lệch chuẩn từ 0,81 đến 0,88 cho thấy mức độ phân tán thấp và sự đồng thuận tương đối cao giữa các người học.

Bảng 4. Nhận thức của sinh viên về hiệu quả của Google Voice Typing

| Mục | Giá trị nhỏ nhất | Giá trị lớn nhất | Trung bình | Độ lệch chuẩn |
|--|------------------|------------------|------------|---------------|
| GVT cung cấp phản hồi ngay lập tức về phát âm của tôi. | 1 | 5 | 3,74 | 0,94 |
| Khi hệ thống gõ sai từ, tôi nhận ra rằng mình đã phát âm chưa chính xác. | 1 | 5 | 3,86 | 0,99 |
| Phản hồi từ hệ thống giúp tôi xác định các lỗi phát âm. | 1 | 5 | 3,90 | 0,93 |
| Tôi điều chỉnh cách phát âm của mình dựa trên phản hồi của hệ thống. | 1 | 5 | 3,80 | 0,99 |
| GVT giúp việc luyện tập phát âm trở nên hiệu quả hơn. | 1 | 5 | 3,84 | 0,98 |

Kết quả bảng 4 chỉ ra sinh viên đánh giá cao hiệu quả phản hồi của GVT, với điểm trung bình dao động từ 3,74 đến 3,90, phản ánh mức độ đồng ý khá cao. Đặc biệt, tiêu chí *xác định lỗi phát âm* đạt giá trị cao nhất ($M = 3,90$), cho thấy công cụ hỗ trợ tốt trong việc nhận diện

sai sót. Ngoài ra, khả năng nhận ra lỗi khi hệ thống gõ sai ($M = 3,86$) và điều chỉnh phát âm ($M = 3,80$) cũng được ghi nhận tích cực. Tuy nhiên, độ lệch chuẩn tương đối cao (0,93-0,99) cho thấy vẫn tồn tại sự khác biệt nhất định trong trải nghiệm của người học

Bảng 5. Nhận thức của sinh viên về sự tham gia vào quá trình luyện tập

| Mục | Giá trị nhỏ nhất | Giá trị lớn nhất | Trung bình | Độ lệch chuẩn |
|--|------------------|------------------|------------|---------------|
| Tôi luyện tập phát âm thường xuyên hơn khi sử dụng GVT. | 1 | 5 | 3,52 | 0,93 |
| Tôi lặp lại các từ cho đến khi hệ thống nhận diện đúng. | 1 | 5 | 3,80 | 0,88 |
| Tôi sẽ tiếp tục sử dụng GVT để cải thiện phát âm của mình. | 1 | 5 | 3,76 | 0,85 |

Kết quả bảng 5 thể hiện GVT có tác động tích cực đến mức độ tham gia học tập của sinh viên, với điểm trung bình từ 3,52 đến 3,80. Sinh viên có xu hướng lặp lại từ cho đến khi được nhận diện đúng ($M = 3,80$), cho thấy sự kiên trì trong luyện tập. Ý định tiếp tục sử dụng công cụ cũng ở mức cao ($M = 3,76$). Tuy nhiên, tần suất luyện tập chỉ ở mức trung bình ($M = 3,52$), cho thấy mức độ duy trì chưa thực sự mạnh. Độ lệch chuẩn tương đối thấp phản ánh sự đồng thuận khá cao giữa người học.

4.3. Kết quả từ câu hỏi mở

Khi trao đổi về tác động của GVT, sinh viên cho rằng công cụ này hỗ trợ tích cực việc cải thiện phát âm thông qua phản hồi tức thì và khả năng nhận diện lỗi. Khi phát âm chưa chính xác, hệ thống ghi nhận sai giúp người học nhanh chóng nhận ra và điều chỉnh. Nhờ đó, họ nâng cao ý thức về độ chính xác, phát âm chậm rãi hơn và chú ý đến trọng âm, ngữ điệu và từng âm riêng lẻ. Việc luyện tập lặp lại kết hợp phản hồi ngay lập tức giúp cải thiện độ chính xác và tăng sự tự tin trong giao tiếp.

Về các khía cạnh cải thiện, sinh viên ghi nhận tiến bộ rõ rệt ở phụ âm cuối (/s/, /t/, /z/), trọng âm từ và độ rõ ràng của lời nói. Ngoài ra, khả năng phân biệt nguyên âm dài - ngắn, cũng như các yếu tố siêu đoạn tính như ngữ điệu, trọng âm câu và nhịp điệu, cũng được cải thiện, giúp lời nói tự nhiên và dễ hiểu hơn.

Tuy nhiên, sinh viên cũng gặp một số khó khăn. Công cụ chủ yếu cung cấp phản hồi ở mức nhận diện từ, thiếu thông tin chi tiết về lỗi âm vị, trọng âm và ngữ điệu. Độ chính xác của hệ thống có thể bị ảnh hưởng bởi tiếng ồn, giọng nói và kết nối mạng. Bên cạnh đó, công cụ ưu tiên độ chính xác từ hơn tính tự nhiên của lời nói, hạn chế hỗ trợ các đặc điểm như nối âm.

Từ đó, sinh viên đề xuất cần cải thiện công cụ theo hướng cung cấp phản hồi chi tiết hơn, tích hợp mô hình phát âm chuẩn, ký hiệu ngữ âm và chức năng nghe chậm. Đồng thời, vai trò hướng dẫn của giáo viên, môi trường luyện tập phù hợp và các tính năng như theo dõi tiến độ hay chấm điểm phát âm cũng được xem là cần thiết để nâng cao hiệu quả học tập.

4.4. Thảo luận

Kết quả phân tích bản ghi âm cho thấy sự cải thiện trong phát âm của sinh viên sau quá trình sử dụng GVT. Số lượng lỗi có xu hướng giảm, đặc biệt ở nguyên âm đôi, nguyên âm đơn, phụ âm cuối và một số cụm phụ âm phức tạp. Tuy nhiên, do nghiên cứu không sử dụng nhóm đối chứng, sự cải thiện này không thể quy hoàn toàn cho việc sử dụng công cụ, mà có thể chịu ảnh hưởng của các yếu tố khác như quá trình luyện tập tự nhiên theo thời gian hoặc hoạt động học tập trên lớp. Trong bối cảnh đó, cơ chế phản hồi tức thì của công nghệ nhận diện giọng nói có thể là một yếu tố hỗ trợ giúp người học nhận diện và điều chỉnh lỗi thông qua việc luyện tập lặp lại. Tuy nhiên, việc phân tích bản ghi âm chỉ được thực hiện trên 28/50 sinh viên có thể ảnh hưởng đến tính đại diện của dữ liệu. Do đó, các kết quả liên quan đến phát âm cần được diễn giải một cách thận trọng và chưa nên khái quát hóa cho toàn bộ người học. Kết quả bảng hỏi cũng cho thấy sinh viên có nhận thức tích cực về hiệu quả của công cụ, với các giá trị trung bình ở mức khá cao. Người học nhìn nhận GVT như một phương tiện hỗ trợ nhận diện lỗi nhanh chóng, cải thiện độ chính xác và thúc đẩy việc tự học. Dữ liệu định tính phần nào củng cố nhận định này khi sinh viên cho rằng công cụ giúp họ dễ dàng phát hiện lỗi, đặc biệt ở phụ âm cuối, trọng âm và nguyên âm, đồng thời tăng sự tự tin nhờ có thể luyện tập nhiều lần.

Tuy nhiên, một số hạn chế cũng được ghi nhận, bao gồm độ chính xác chưa ổn định, ảnh hưởng của tiếng ồn và hạn chế trong phản hồi về ngữ điệu và trọng âm. Do đó, việc sử dụng công cụ cần được xem xét trong mối quan hệ với

vai trò hướng dẫn của giảng viên nhằm nâng cao hiệu quả điều chỉnh lỗi.

Nhìn chung, kết quả của nghiên cứu có xu hướng tương đồng với một số nghiên cứu trước (Abimanto & Sumarsono, 2024; Hoang & cộng sự, 2025; Nguyễn & cộng sự, 2025), cho thấy tiềm năng hỗ trợ của công nghệ nhận diện giọng nói trong học phát âm. Tuy nhiên, các kết luận cần được giới hạn trong phạm vi mẫu nghiên cứu và bối cảnh cụ thể, và công cụ này nên được xem như một phương tiện bổ trợ hơn là giải pháp thay thế cho phương pháp giảng dạy truyền thống.

V. Kết luận

Nghiên cứu cho thấy việc sử dụng GVT có liên quan đến sự cải thiện độ chính xác phát âm tiếng Anh của sinh viên, có thể thông qua cơ chế phản hồi tức thì của công nghệ nhận diện giọng nói. Công cụ này hỗ trợ người học nhận diện và điều chỉnh lỗi phát âm trong quá trình luyện tập, đồng thời tạo thêm cơ hội thực hành và thúc đẩy sự tham gia chủ động. Tuy nhiên, do nghiên cứu không sử dụng nhóm đối chứng, kết quả cần được diễn giải một cách thận trọng và chưa cho phép khẳng định mối quan hệ nhân quả. Sự cải thiện quan sát được có thể chịu ảnh hưởng của các yếu tố khác như quá trình học tập trên lớp hoặc sự tiến bộ tự nhiên theo thời gian.

Từ góc độ sư phạm, giảng viên có thể cân nhắc tích hợp GVT như một công cụ hỗ trợ trong hoạt động luyện phát âm nhằm tăng cơ hội thực hành và tự học cho sinh viên. Tuy nhiên, công cụ này nên được sử dụng kết hợp với hướng dẫn của giáo viên để đạt hiệu quả cao hơn. Các nghiên cứu trong tương lai cần sử dụng thiết kế có nhóm đối chứng, mở rộng

quy mô mẫu và kéo dài thời gian can thiệp để đánh giá rõ hơn vai trò của công cụ này trong việc hỗ trợ cải thiện phát âm.

Tài liệu tham khảo

- Abimanto, D., & Sumarsono, W. (2024). Improving English pronunciation with AI speech-recognition technology. *Acitya: Journal of Teaching and Education*, 6(1).
- Arifah, T., & Budiastuti, R. E. (2025). The EFL students' difficulties of pronunciation practice. *English Language and Literature International Conference (ELLiC) Proceedings*, 8(1). <https://jurnal.unimus.ac.id/index.php/ELLIC/article/view/18510>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bent, T., Holt, R. F., Henry, M., & Lind-Combs, H. (2024). Relating pronunciation distance metrics to intelligibility across English accents. *Journal of Phonetics*, 107, 101357. <https://doi.org/10.1016/j.wocn.2024.101357>.
- Dang, D. T. V. (2024). Teaching English pronunciation for international intelligibility. *Science & Technology Development Journal*, 27(SI), 1-7. <https://doi.org/10.32508/stdj.v27iSI.4410>.
- Dennis, N. K. (2024). Using AI-powered speech recognition technology to improve English pronunciation and speaking skills. *IAFOR Journal of Education*, 12(2), 85-102. <https://doi.org/10.22492/ije.12.2.05>.
- Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, 39(3), 379-397.
- Evers, K., & Chen, S. (2020). Effects of an automatic speech recognition system with peer feedback on pronunciation instruction for adults. *Computer Assisted Language Learning*, 1-21. <https://doi.org/10.1080/09588221.2020.1839504>.
- Gilakjani, A. P. (2012). A study of factors affecting EFL learners' English pronunciation learning and the strategies for instruction. *International Journal of Humanities and Social Science*, 2(3), 119-128.
- Hoang, T. T., Nguyen, T. H. H., & Gonzales, V. O. (2025). Integrating technology into English pronunciation learning: A combination of both synchronous and asynchronous activities. *International Journal of TESOL Studies*, 25(0912), 1-20. <https://doi.org/10.58304/ijts.250912>.
- Jiang, M. Y. C., Jong, M. S. Y., Lau, W. W. F., Chai, C. S., & Wu, N. (2021). Using automatic speech recognition technology to enhance EFL learners' oral language complexity in a flipped classroom. *Australasian Journal of Educational Technology*, 37(2), 110-131. <https://doi.org/10.14742/ajet.6798>
- Jiang, M. Y. C., Jong, M. S. Y., Wu, N., Bin, S., Chai, C. S., Lau, W. W. F., & Huang, B. Y. (2022). Integrating automatic speech recognition technology into vocabulary learning in a flipped English class for Chinese college students. *Frontiers in Psychology*, 13, 902429. <https://doi.org/10.14742/ajet.6798>.
- Kelly, G. (2000). *How to teach pronunciation*. Longman.
- Leis, A. (2025). How speech-to-text technology affects pronunciation gains and self-confidence in EFL learners. *Computer Assisted Language Learning*. Advance online publication. <https://doi.org/10.1080/09588221.2025.2534498>.
- Munro, M. J., & Derwing, T. M. (2011). The foundations of accent and intelligibility in pronunciation research. *Language Teaching*, 44(3), 316-327. <https://doi.org/10.1017/S0261444811000103>.

- Nguyễn, T. L., Nguyễn, T. T. H., & Đặng, T. Q. H. (2025). Leveraging Google voice typing technology to enhance pronunciation skills among English language majors at Hung Vuong University. *Middle East Research Journal of Humanities and Social Sciences*, 5(3). <https://doi.org/10.36348/merjhss.2025.v05i03.002>.
- Ranchal, R., Taber-Doughty, T., Guo, Y. R., Bain, K., Martin, H., Robinson, J. P., & Duerstock, B. S. (2013). Using speech recognition for real-time captioning and lecture transcription in the classroom. *IEEE Transactions on Learning Technologies*, 6(4), 299-311. <https://doi.org/10.1109/TLT.2013.21>.
- Rosyid, A. (2016). Factors affecting EFL learners in learning English pronunciation. *Pedagogia: Jurnal Ilmu Pendidikan*, 8(2), 436-443. <https://journal.unpak.ac.id/index.php/pedagogia/article/view/4924>.

THE EFFECTIVENESS OF GOOGLE VOICE TYPING IN IMPROVING ENGLISH PRONUNCIATION AMONG UNIVERSITY STUDENTS

Dang Thi Huong Thao¹

Abstract: *This study aims to evaluate the effectiveness of Google Voice Typing, a speech-to-text tool that uses Automatic Speech Recognition (ASR) technology, in improving students' English pronunciation. The study was conducted with 50 second-year English-major students at Hanoi Open University. A mixed-methods approach was employed to collect and analyze the data. Quantitative data were collected through a 14-item questionnaire on a five-point Likert scale and by recording the number of pronunciation errors students made during practice sessions. Qualitative data were obtained through 4 open-ended questions and by analyzing specific pronunciation errors identified in recordings of 28 students. The results indicate that Google Voice Typing helps students improve pronunciation accuracy, and participants reported positive perceptions of its effectiveness in developing their pronunciation skills. The study suggests that integrating voice typing into pronunciation practice activities may improve the effectiveness of English language learning.*

Keywords: *Google Voice Typing, ASR technology, English pronunciation, pronunciation practice activities, English language learning*

¹ Hanoi Open University, Hanoi, Vietnam