

CẢI THIỆN LIÊN LẠC THOẠI THÔNG QUA HỆ THỐNG NHẬN DẠNG GIỌNG NÓI TỰ ĐỘNG

Phạm Minh Hằng*, Nguyễn Thị Lan Anh, Nguyễn Hoàng Phát, Phan Thanh Minh

Khoa Khai thác Hàng không, Học viện Hàng không Việt Nam

**Email: 2056060003@vaa.edu.vn, Số điện thoại: 0763083745*

TÓM TẮT

Số lượng các chuyến bay với mật độ dày đặc trên toàn thế giới đặt ra những thách thức tương ứng về năng lực đối với các dịch vụ bảo đảm hoạt động bay, đặc biệt là dịch vụ không lưu. Trong đó, liên lạc giữa kiểm soát viên không lưu (KSVKL) và phi công luôn được xem như yếu tố cốt lõi, tác động trực tiếp trong công tác điều hành bay. Nhằm tăng hiệu quả công việc cho KSVKL, bài báo này đề ra phương án áp dụng hệ thống nhận dạng giọng nói tự động (ASR) vào việc xác định và khắc phục các lỗi do liên lạc thoại giữa KSVKL và phi công. Đây là một công nghệ tiên tiến được nhiều quốc gia, hãng hàng không,... thử nghiệm và triển khai trên toàn cầu. Việt Nam cũng đã từng bước tham gia nghiên cứu hệ thống này, trong đó nổi bật là báo cáo thí nghiệm hệ thống dựa trên dữ liệu được trích xuất từ các cơ sở điều hành bay thuộc Công ty Quản lý bay miền Trung. Kết quả của nghiên cứu đã cho thấy tỷ lệ phát hiện giọng nói ở mức độ chấp nhận được và tỷ lệ nhận dạng lỗi rất đáng ghi nhận.

Từ khóa: Nhận dạng giọng nói tự động, KSKL, Liên lạc thoại, Lỗi đọc lại/nghe lại, Phi công.

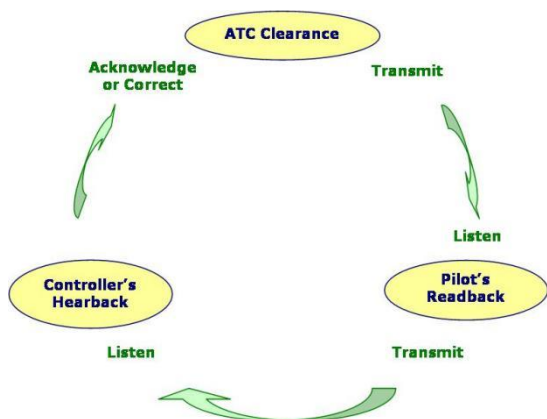
ABSTRACT

The increasing volume of air traffic worldwide poses corresponding capacity challenges for air traffic management service, especially air traffic control service. In particular, communication between air traffic controllers and pilots is always the core factor that directly affects the flight operation. In order to support air traffic controllers, the article proposes to apply automatic speech recognition techniques to detect errors caused by voice communication errors between air traffic controllers and pilots. This is an advanced technology that many countries, airlines,... embarked on testing and researching around the world. In Vietnam, a study based on data collected from the flight control stations of the Middle Region Air Traffic Services Company. The initial experimental results of the study gave an acceptable recognition rate and a very positive error detection rate.

Keywords: Automatic speech recognition, Air traffic control, Voice communication, Readback/Hearback errors, Pilot.

1. ĐẶT VẤN ĐỀ

Số lượng các chuyến bay ngày càng tăng trên toàn thế giới đặt ra những thách thức tương ứng về năng lực đối với các dịch vụ bảo đảm hoạt động bay, đặc biệt là dịch vụ không lưu. Theo báo cáo “Triển vọng Vận tải Hàng không đến Năm 2025” (“*Outlook for Air Transport to the Year 2025*”) của Tổ chức Hàng không Dân dụng Quốc tế (ICAO), lưu lượng hành khách trên các đường bay quốc tế chính dự kiến sẽ tăng khoảng 3% đến 6% mỗi năm cho đến năm 2025 [1]. Vì vậy, việc thử nghiệm, rà soát và đổi mới luôn được đẩy mạnh nhằm đáp ứng yêu cầu đặt ra ngày càng khắt khe hơn đối với dịch vụ không lưu. Đặc biệt, liên lạc giữa KSVKL và phi công luôn là yếu tố cốt lõi, tác động trực tiếp trong công tác điều hành bay. Chất lượng liên lạc ảnh hưởng đáng kể đến hiệu suất cũng như sự an toàn của hoạt động kiểm soát không lưu (KSKL).



Hình 1: Vai trò của liên lạc thoại trong KSKL

Việc tích hợp các công nghệ nhận dạng giọng nói tự động (*Automatic Speech Recognition - ASR*) trong KSKL đã được nghiên cứu để cải thiện hiệu suất liên lạc giữa KSVKL và phi công cũng như tăng khả năng tự động hóa của các hệ thống KSKL. Sự cải thiện liên tục về độ chính xác và hiệu suất của các công nghệ nhận dạng giọng nói tự động đã mở ra nhiều cơ hội tiềm năng để nghiên cứu, xem xét và cải thiện mức độ an toàn trong công tác điều hành bay.

Trong thời gian gần đây, việc áp dụng hệ thống ASR (*Automatic Speech Recognition*) vào KSKL là một lối đi mới mẻ, thu hút được

các hãng hàng không và nhiều công ty, doanh nghiệp hàng không trên toàn thế giới. Điển hình như vào năm 2018, Airbus hợp tác cùng nhiều doanh nghiệp đã tổ chức cuộc thi với quy mô toàn cầu với mục tiêu nâng cao an toàn, điều hòa và hiệu quả trong công tác điều hành bay. Cuộc thi này nhận về nhiều ý tưởng, dự án đến từ những doanh nghiệp trong ngành, cũng như các nhóm và cá nhân nghiên cứu đến từ nhiều trường đại học trên toàn thế giới [2].

Ngoài ra, ở Trung Quốc, hệ thống ASR được nghiên cứu và thử nghiệm tại nhiều cơ sở cung cấp dịch vụ không lưu ở Trung Quốc, chẳng hạn như Thành Đô, Trùng Khánh, Lệ Giang,... trong các giai đoạn cất hạ cánh, tiếp cận và đường dài. Tại đây, ASR được xem như là một hệ thống hỗ trợ, phối hợp làm việc hoạt động với hệ thống KSKL thu được kết quả độ chính xác cao nhất của giọng nói tiếng Trung và tiếng Anh lần lượt đạt trên 97% và 96%, có thể coi là hiệu suất vượt trội so với khả năng của con người [3].

Thêm vào đó, Việt Nam có thể được xem là một trong những quốc gia có vùng trời rộng lớn trong khu vực Đông Nam Á cũng như tần suất bay ngày một tăng. Song, lãnh thổ của Việt Nam còn liền kề với Trung Quốc với dân số hơn 1 tỷ dân, kết nối với quốc gia này để tiếp tục thực hiện chuyến bay đến các quốc gia khác trên thế giới. Do đó, số lượng và mật độ không lưu trên vùng trời lãnh thổ do Việt Nam cung cấp dịch vụ không lưu là vô cùng lớn.

Dựa vào văn bản báo cáo Bộ Giao thông vận tải về hoạt động vận chuyển hàng không hai tháng đầu năm 2023 của Cục Hàng không Việt Nam, lượng hành khách thông qua các cảng hàng không ghi nhận tăng trưởng 91,5% so với cùng kỳ năm 2022 [4]. Trong tình hình lưu lượng hàng không phát triển mỗi năm, điều này vô hình trung tạo nên sức ép không hề nhỏ cho các KSVKL. Họ phải đối mặt với yêu cầu cực kỳ khắt khe về việc liên lạc và kiểm soát nhiều tàu bay cùng một lúc trên sóng vô tuyến. Điều này được xem như một nguyên nhân làm tăng khả năng xuất hiện sai sót trong liên lạc thoại để cấp huấn lệnh và kiểm soát tàu bay. Vì vậy, hệ thống ASR là một ứng dụng cần thiết để loại trừ sai sót, đảm bảo an toàn hàng không.

Hòa vào sự phát triển của ngành hàng không trên thế giới, ngành hàng không tại Việt Nam đã có kế hoạch áp dụng từng bước các công nghệ tự động hóa trong công tác điều hành bay. Để chuẩn bị cho yêu cầu đó, ta cần giải quyết thách thức thử nghiệm hệ thống ASR tại môi trường KSKL.

Phần kế tiếp của bài báo này được viết gồm các phần: Mục 2 khái quát về lý thuyết nhận dạng giọng nói tự động trong KSKL. Mục 3 nêu ra nguyên lý hoạt động cơ bản của hệ thống. Mục 4 là trình bày kết quả của các thí nghiệm áp dụng hệ thống. Sau cùng là kết luận rút ra được sau qua trình nghiên cứu được thể hiện qua Mục 5.

2. NHẬN DẠNG GIỌNG NÓI TỰ ĐỘNG

Nhận dạng giọng nói tự động là quy trình phân tách đoạn giọng nói đầu vào thu được thành chuỗi các mẫu theo thứ tự. Đây có thể được coi là một công việc thách thức, đòi hỏi năng lực và trình độ chuyên môn cao do giọng nói con người phát ra luôn thay đổi theo thời gian, do sự khác biệt lớn giữa những chủ thể (sự đa dạng về độ tuổi, giới tính của người nói,...), tốc độ nói (nhanh, trung bình, chậm), ngữ cảnh của cuộc hội thoại (các bối cảnh, chuyên ngành khác nhau), môi trường (môi trường bình thường, trò chuyện qua điện thoại, liên lạc qua sóng vô tuyến,...) và sự tác động của tiếng địa phương trong bối cảnh môi trường làm việc tập hợp nhiều nhân viên đến từ nhiều tỉnh thành khác nhau. Công nghệ nhận dạng giọng nói tự động có khả năng dịch lời nói của con người thành các chuỗi từ, đã phát triển vượt bậc trong những thập kỷ qua. Trong thời gian gần đây, công nghệ và hiệu suất của các hệ thống nhận dạng giọng nói đã được cải thiện đáng kể. Điều đó giúp ASR được áp dụng thành công và rộng rãi trong nhiều lĩnh vực của đời sống như như đọc chính tả, ra lệnh và điều khiển, xử lý lời nói thành văn bản (ví dụ: trình xử lý văn bản hoặc email) và trợ lý cá nhân trên điện thoại di động (ví dụ: Siri của Apple trên iOS, Cortana của Microsoft trên Window Phone, Google Hiện hành trên Android) [5].

2.1 Ứng dụng của nhận dạng giọng nói trong KSKL

Hệ thống ASR ứng dụng trong lĩnh vực KSKL có những khía cạnh đặc trưng, khác biệt so với sự ứng dụng hệ thống này trong các lĩnh vực, ngành nghề khác. Nguyên nhân đầu tiên có thể kể đến là chất lượng giọng nói thu được không ổn định do môi trường truyền âm thanh là sóng vô tuyến VHF độ nhiễu cao. Tiếp theo là do ngôn ngữ sử dụng chính trong KSKL là tiếng Anh được phát âm từ người nói đến từ đa dạng các quốc gia, dân tộc trên toàn cầu, nên đôi khi người sử dụng có thể sử dụng trộn lẫn với tiếng bản địa do thói quen, ngoài ra, nhịp độ giao tiếp của KSVKL và phi công thường nhanh hơn nhịp độ trong đời sống hằng ngày hay trong các lĩnh vực khác, có thể dẫn đến xảy ra hiểu nhầm. Bên cạnh các bất lợi đã được liệt kê thì việc ứng dụng hệ thống nhận dạng giọng nói trong KSKL cũng có một số điểm mạnh, ví dụ như số lượng từ vựng chuyên ngành sử dụng trong lĩnh vực này ít hơn, ngữ nghĩa rõ ràng hơn và cấu trúc của các cuộc hội thoại phải tuân thủ theo một tập hợp các quy chuẩn của quốc tế [6].

Có thể phân loại ứng dụng của hệ thống ASR trong KSKL thành 04 nhóm như dưới đây [7]:

- *Ứng dụng trong công tác đào tạo KSKL:* Ngày nay việc đào tạo kỹ năng thực hành cho KSVLL nhờ vào phần mềm mô phỏng (*simulation*) và với sự hỗ trợ của các phi công giả định - một người trong vai trò là phi công, cách hành xử và giao tiếp tương tự như phi công thật. Tuy nhiên, các phi công này có thể mang lại linh hoạt kém hơn, ngoài ra có thể tạo ra tình huống không sát với thực tế do những người đóng vai phi công này chỉ dựa vào kiến thức lý thuyết mà chưa được đào tạo và thực hành như phi công thật. Khó khăn này được khắc phục bằng cách sử dụng hệ thống ASR để lập trình phần mềm “phi công tự động” - là phần mềm có thể hiểu, xử lý giọng nói của các KSVKL sử dụng phần mềm và phản hồi lại họ bằng dữ liệu tổng hợp giọng nói.

- *Ứng dụng tính toán và cân bằng khối lượng công việc của KSVKL:* Trong dịch vụ không lưu, khối lượng công việc của KSVKL là điều kiện tiên quyết tác động đến hiệu quả

của cả quy trình, đặc biệt là khi lưu lượng hàng không gia tăng cao thì bài toán phân tích và điều hòa khối lượng công việc của KSVKL càng trở nên cấp thiết hơn. Để thực hiện điều này một cách thủ công thì sẽ mang lại khó khăn và tốn kém. Tuy nhiên, với sự trợ giúp của hệ thống ASR, việc kiểm soát các tàu bay thông qua các huấn lệnh do KSVKL cấp sẽ được nhận dạng tự động, từ đó khối lượng công việc sẽ được xác định và tính toán.

- *Ứng dụng phân tích và diễn giải cuộc hội thoại giữa KSVKL và phi công:* Với hệ thống ASR, hệ thống có thể tự động phân tích số lượng lớn dữ liệu liên lạc, phục vụ cho mục đích nghiên cứu và cải thiện các quy định đối với KSVKL. Việc phân tích hội thoại này cũng là một bước trong quy trình chuyển đổi dữ liệu âm thanh sang dạng chữ viết, dựa vào đoạn văn bản mà hệ thống hiển thị, KSVKL có thể so sánh huấn lệnh đã cấp với câu đọc lại của phi công từ tàu bay để xác định lỗi.

- *Ứng dụng KSVKL dự phòng:* Kết quả đầu ra của hệ thống ASR kết hợp với các dữ liệu khác trong môi trường KSKL có thể được sử dụng làm nguồn dữ liệu cho hệ thống KSVKL dự phòng. Hệ thống dự phòng này có thể nhận dạng các nguy cơ tiềm ẩn đe dọa an toàn hàng không mà KSVKL thực đã bỏ sót do sơ suất hay mất tập trung, sau đó cung cấp các gợi ý giải pháp và thông tin để KSVKL xem xét.

2.2 Tiêu chí lựa chọn hệ thống ASR cho KSKL

Các hệ thống ASR có thể được phân loại theo loại lời nói, loại mô hình người nói và loại từ vựng mà hệ thống có thể nhận ra [8].

Việc áp dụng hệ thống ASR trong KSKL đi kèm với nhiều thách thức và cơ hội do các đặc điểm giao tiếp đặc trưng giữa KSVKL và phi công, chẳng hạn như vốn từ vựng chuyên ngành giới hạn, yêu cầu độ chính xác cao, gắn với nhu cầu thời gian thực và các định dạng tiêu chuẩn hóa để giao tiếp [6].

Dựa trên những khía cạnh này, các nghiên cứu đã gợi ý rằng một hệ thống ASR phù hợp

với KSKL phải là một hệ thống độc lập với người nói, có thể nhận dạng các từ vựng có số lượng trung bình và lời nói liên tục [7].

- *Phụ thuộc vào người nói:* Mặc dù các ứng dụng nhận dạng huấn lệnh trong KSKL chỉ yêu cầu một KSVKL cùng lúc, nhưng có những trường hợp cần nhiều KSVKL trong môi trường vận hành.

- *Nhận dạng giọng nói liên tục:* Mặc dù các hệ thống nhận dạng từ riêng lẻ và từ liên kết thường có độ chính xác cao hơn hệ thống nhận dạng giọng nói liên tục, nhưng chúng không phù hợp trong lĩnh vực KSKL. Điều này là do yêu cầu công việc khiến KSVKL tạm dừng giữa mỗi từ khi đưa ra huấn lệnh. Do đó, các từ độc lập và các hệ thống nhận dạng từ được kết nối sẽ gây ra sự chậm trễ trong giao tiếp giữa phi công và KSVKL. Hệ thống nhận dạng giọng nói liên tục, cho phép KSVKL nói một cách tự nhiên mà không bị tạm dừng, là hệ thống được lựa chọn khi áp dụng hệ thống ASR trong lĩnh vực KSKL.

- *Số lượng từ vựng:* Trong KSKL, các từ vựng được sử dụng trong liên lạc thoại giữa KSVKL và phi công tuân theo Cụm từ tiêu chuẩn (*phraseology*) của ICAO. Toàn bộ vốn từ vựng chỉ khoảng vài trăm từ. Do đó, một hệ thống nhận dạng giọng nói, từ vựng có kích thước trung bình là phù hợp trong bối cảnh KSKL.

- *Tính năng:* Trong KSKL, điều quan trọng không phải là hệ thống ASR có thể nhận ra từng từ đơn lẻ mà điều quan trọng là các thông tin quan trọng được truyền tải đúng và lỗi sai được phát hiện chính xác. Ví dụ: hệ thống ASR không bắt buộc phải nhận dạng tất cả các từ trong câu sau: “Good morning Lufthansa one zero one descend level one two three”, tuy nhiên, hệ thống phải có khả năng trích xuất thông tin “DLH101 DESCEND FL 123”.

3. NGUYÊN LÝ HOẠT ĐỘNG

Trong số 4 nhóm ứng dụng đã được liệt kê trên, do đặc thù yêu cầu nghe, nói giữa KSVKL và phi công nên bài báo này sẽ đi sâu vào vào chức năng hỗ trợ phát hiện lỗi trong liên lạc thoại [8].

Ứng dụng này thuộc nhóm thứ ba, được vận hành trên cơ sở hai giai đoạn như sau:

- *Bước 1*: Chuyển đổi dữ liệu giọng nói của KSVKL và sự đọc lại (*readback*) của phi công thành dạng chữ viết.

- *Bước 2*: Đối chiếu hai dữ liệu này để phát hiện lỗi liên lạc thoại.

Do đó nhiệm vụ quan trọng nhất của bài báo là phải giải quyết vấn đề chuyển đổi dữ liệu âm thanh thành dữ liệu chữ viết, mà trong đó mỗi lời nói sẽ được chuyển thành một chuỗi theo trật tự các từ. Quy trình này có thể được thể hiện qua hai giai đoạn như sau:

- *Bước 1*: Xác định một chuỗi vector $X = X_1, X_2, \dots, X_n$ từ tín hiệu tiếng nói đã nhận.

- *Bước 2*: Từ chuỗi vector đặc trưng này, tìm chuỗi từ tương ứng $W = W_1, W_2, \dots, W_n$ sao cho chuỗi từ này khớp nhất với chuỗi vector X , hay nói cách khác là tối đa hóa xác suất hậu nghiệm $P(W|X)$:

$$\hat{W} = \underset{W}{\operatorname{argmax}} P(W|X) \quad (1)$$



Hình 2: Sơ đồ nguyên lý hệ thống ASR

Hoạt động của hệ thống diễn ra theo hai giai đoạn, đào tạo và nhận dạng, như trong Hình 2 và được diễn tả bên dưới:

- *Xử lý*: Giọng nói trước khi phân tích sẽ được xử lý trước nhằm khử nhiễu và tạp âm, nhận dạng điểm cuối của lời nói (*end point*), rồi sau đó được phân ra thành các khung chồng lên nhau. Độ dài một khung thường gặp nhất là $25ms$ và khoảng cách giữa nhiều khung liền kề là $10ms$.

- *Trích đặc trưng*: Sau khi xử lý, đoạn âm thanh này sau đó được cho qua khối trích đặc trưng để trích xuất các đặc trưng của giọng nói. Và việc chọn bộ đặc trưng ảnh hưởng rất nhiều đến hiệu quả hoạt động của cả hệ thống.

- *Giai đoạn huấn luyện*: Trong giai đoạn này, thông số của các mẫu giọng nói sẽ được ước lượng. Cụ thể, một câu là sự pha trộn các đơn vị ngôn ngữ nhỏ hơn là như từ ngữ, âm tiết, âm vị, ... Để hình thành nên một câu ta cần phải kết hợp các mẫu âm thanh của chúng với các quy luật ngôn ngữ. Do đó, để đảm bảo tính chính xác, thì giọng nói cần được thể hiện qua hai mô hình. Đầu tiên là mô hình âm thanh, đóng vai trò quyết định trong việc cải thiện hiệu quả nhận dạng của hệ thống bằng cách liên kết các đặc điểm giọng nói với các âm vị trong câu. Các mẫu âm thanh biểu thị bởi xác suất quan sát được âm thanh X (thể hiện trong đặc trưng X) khi từ W được nói, kí hiệu là $P(X|W)$. Mô hình tiếp theo là mô hình ngôn ngữ, nhằm tìm ra từ vựng thỏa mãn các yêu cầu về ngôn ngữ.

- *Giai đoạn nhận dạng*: Nếu như trong giai đoạn trên ta có hai mô hình được thực hiện thì trong giai đoạn này ta cũng có hai bước so sánh tương ứng. Cụ thể là ở hai mức độ, ngôn ngữ và âm thanh.

Ở cấp độ âm thanh, mỗi vector đặc trưng được trích từ một khung giọng nói cần nhận dạng được đối chiếu với mô hình âm thanh của tất cả các từ ngữ và sau đó vector đặc trưng này sẽ được gán cho từ phù hợp nhất sau khi phân tích. Chuỗi từ giả định sau đó được kết hợp với mô hình ngôn ngữ để đối chiếu và cho ra câu khớp nhất. Áp dụng định luật Bayes:

$$P(W|X) = \frac{P(X|W)P(W)}{P(X)} \quad (2)$$

Ở đây $P(X)$ là xác suất tiên nghiệm của vector đặc trưng. Lưu ý rằng xác suất này độc lập với các mô hình giọng nói. Kết hợp (2) và (1), ta sẽ có xác suất chuỗi từ W xấp xỉ tích của hai mô hình.

4. KẾT QUẢ THỬ NGHIỆM

Theo kết quả thử nghiệm thực tế tại Công ty Quản lý bay miền Trung [9], nghiên cứu đi sâu vào nhiệm vụ chính của KSVKL là duy trì phân cách giữa những tàu bay, bằng việc

nhận dạng lỗi đọc lại (*readback*) của KSVKL đối với các huấn lệnh thay đổi mực bay.

4.1 Xây dựng cơ sở dữ liệu

- *Giới hạn nghiên cứu:* các đoạn hội thoại của phi công và kiểm soát viên ở 10 mực bay thông dụng trong vùng trời kiểm soát ở Đà Nẵng từ 15.000ft đến 25.000ft (FL150-FL250). Trong đó cách phát âm sẽ theo tiêu chuẩn phát âm của ICAO, ví dụ mực bay 21.000ft sẽ phát âm là “flight level two one zero” và được viết là FL210.

- *Dữ liệu:* bao gồm dữ liệu thực tế và giả lập.

+ *Dữ liệu thực tế:* là các đoạn giao tiếp thoại của phi công và KSVKL thực tế trong quá trình điều hành bay tại đài KSKL Đà Nẵng. Các chủ thể dữ liệu là các phi công, kiểm soát viên trong độ tuổi từ 25 đến 40, đối tượng người nói bao gồm ở cả trong và ngoài nước. Số lượng thu thập thực tế là 519 đoạn giọng nói (2018).

+ *Dữ liệu giả lập:* là các đoạn giao tiếp thoại được ghi âm của 5 TNV bao gồm 2 nam giới và 3 nữ giới đều là công dân Việt Nam trong độ tuổi tương tự như ở dữ liệu thực tế với số lượng các đoạn mô phỏng là 601 mẫu.

Tổng số dữ liệu đã tổng hợp được là 1120 đoạn. Tỷ lệ lượng dữ liệu thực tế trên tổng số lượng mẫu biến thiên từ 0% với FL220 đến 100% với FL150 (trình bày chi tiết trong bảng 1).

Mực bay (FL)	220	250	160	...	140	180	240
Số lượng mẫu thực tế	0	15	18	...	53	70	101
Số lượng mẫu giả lập	100	85	82	...	47	70	39
Tỷ lệ mẫu thực tế	0	0,15	0,18	...	0,53	0,50	0,72

Bảng 1. Thành phần dữ liệu

Về chất lượng, dữ liệu mô phỏng có nhiều và lượng tạp âm xung quanh thấp hơn so với dữ liệu thực và không có sự khác biệt đáng kể về cách phát âm, vì cả hai loại mẫu giọng nói đều được phát âm từ những chủ thể không dùng tiếng Anh như ngôn ngữ thứ nhất của họ.

Về sự ảnh hưởng của nhiễu tín hiệu lên dữ liệu thực tế thì cả giọng nói của phi công và kiểm soát viên đều bị ảnh hưởng bởi những tác động nhất định. Cụ thể, dữ liệu giọng nói thu được từ các phi công thường bị rè đi khi truyền qua sóng vô tuyến VHF, còn giọng nói của kiểm soát viên bị ảnh hưởng bởi tiếng ồn xung quanh trong môi trường làm việc như tiếng trao đổi qua điện thoại của kiểm soát viên hiệp đồng, tiếng điện thoại reng,...

Triển khai thử nghiệm

Việc đánh giá hiệu quả của hệ thống dựa trên hai thí nghiệm, tương ứng với hai bước xử lý chính trong đề xuất hệ thống. Đầu tiên, giọng nói của phi công và KSVKL sẽ được chuyển đổi thành dạng chữ viết dựa trên nền tảng Azure, sau đó ta đối chiếu chúng với nhau, từ đó nhận dạng ra các lỗi đọc lại để phát tín hiệu báo động cho kiểm soát viên mục đích để ngăn chặn lỗi nghe lại (*hearback*).

Kết quả thử nghiệm

Trong thử nghiệm 1, ta đã đưa 100 đoạn giọng nói giả lập chỉ mực bay FL220 vào mô hình sẵn có của Azure để chuyển đổi sang văn bản. Tính đến tháng 9/2022 thì đây là phiên bản tiên tiến nhất của dịch vụ đám mây Azure. Kết quả đạt được tỷ lệ lỗi WER (*Word Error Rate*) là 36,47%.

Trong thử nghiệm thứ 2, thực hiện tương tự như thử nghiệm 1 nhưng với 100 đoạn giọng nói thực tế chỉ FL150. Kết quả thu được tỷ lệ WER là 63,41%.

Từ hai thử nghiệm trên, ta thấy rằng lỗi nhận dạng thu được từ dữ liệu thực tế là vượt trội hơn so với lỗi nhận dạng ở dữ liệu giả lập (hơn 27%).

Lý do là do tác động nhiễu khác nhau của môi trường thu thập hai loại dữ liệu. Điều này

đòi hỏi cần có sự cải tiến chất lượng thiết bị thu phát và khử nhiễu.

Thí nghiệm đánh giá mô hình chuyển đổi tín hiệu tiếng nói sang văn bản được huấn luyện mới.

Dữ liệu thử nghiệm: Như được trình bày trong Bảng 1.

Thực hiện: Dữ liệu sẽ được phân chia ngẫu nhiên vào 10 nhóm, sau đó thực hiện 10 lần huấn luyện và kiểm tra theo kiểu xoay vòng nghĩa là huấn luyện trên 9 tập dữ liệu và thử nghiệm trên tập dữ liệu còn lại. Tỷ lệ nhận dạng cuối cùng là trung bình cộng của 10 thử nghiệm (bảng 2).

Bảng 2. Đối chiếu tỷ lệ nhận dạng (%) của mô hình sẵn có và mô hình mới trên 10 mục bay

	FL 220	FL 200	FL 210	FL 250	...	FL 230	FL 150	Trung bình
Mô hình có sẵn	87,00	75,00	85,00	88,00	...	44,29	40,00	64,44
Mô hình mới	92,00	78,00	92,00	98,00	...	59,21	74,00	81,25
Hiệu suất cải thiện	5,00	3,00	7,00	10,00	...	14,92	34,00	16,81

Mặc dù có những hạn chế của môi trường như nhiễu sóng VHF, tiếng ồn nơi làm việc hoặc phát âm không chính xác, khi so sánh hai mô hình trên, ta thấy rằng tỷ lệ nhận dạng được nâng cao đáng kể ở cả 10 mục bay.

Ngoài ra các lỗi nhận dạng là vì đa dạng các lý do gây nên. Vì vậy cần có nhiều giải pháp cụ thể tương ứng đối với mỗi loại lỗi. Ví dụ như lỗi sai từ bắt nguồn từ việc phát âm không chính xác nên cần tập luyện thêm, lỗi thêm từ có thể đến từ dữ liệu thu được chứa những từ đệm không cần thiết (“ừm”, “à”) nên cần loại trừ chúng trong dữ liệu huấn luyện.

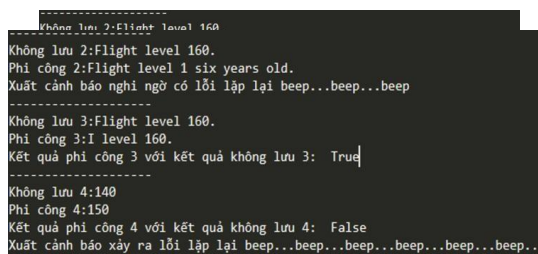
4.2 Thí nghiệm mô phỏng đối chiếu mục bay

Thí nghiệm này là giai đoạn thử nghiệm kế tiếp, sau khi đã chuyển đổi các đoạn giao tiếp thoại sang văn bản giữa phi công và KSVKL. Thí nghiệm diễn ra theo các bước:

- Lựa chọn những cặp dữ liệu dạng chữ đã được chuyển đổi từ hội thoại trong kết quả thí nghiệm sử dụng mô hình mới, hiển thị văn bản trên màn hình.

- Đối chiếu và hiển thị cảnh báo.

Hình 3 thể hiện ba kết quả mô phỏng cho ba tình huống lần lượt là cảnh báo nhầm, phi công đọc lại đúng (không có cảnh báo) và cảnh báo đúng.



Hình 3: Kết quả mô phỏng phát hiện lỗi đọc lại

Từ kết quả mô phỏng trên, ta thấy rằng:

- Giá trị recall (tỷ lệ phát hiện đúng lỗi sai trên tổng các lỗi sai) cao tuyệt đối (=1) cho thấy rằng tất cả lỗi đọc lại đều được nhận dạng một cách triệt để.

- Số lượng cảnh báo sai vẫn còn tồn tại. Mặc dù nó không tác động trực tiếp đến an toàn hàng không nhưng có thể là nhân tố gây mất sự chú ý, tạo nên áp lực cho kiểm soát viên.

Sau đây là một số đề xuất mà nhóm nghiên cứu đưa ra từ các thử nghiệm trong quá trình ứng dụng hệ thống ASR vào hỗ trợ KSVKL trong công tác điều hành bay:

- Cần cải thiện cũng như nâng cao cơ sở vật chất, đặc biệt là chất lượng của các thiết bị thu phát, chuyển mạch thoại, có những biện pháp nhằm giảm tạp âm trong môi trường làm việc của kiểm soát viên.

- Cần đảm bảo, nhắc nhở và kiểm tra thường xuyên phát âm quy chuẩn đối với kiểm soát viên.

- Xây dựng nguồn dữ liệu đảm bảo và đầy đủ hơn về cả số lượng và chất lượng nhằm tăng tính thuyết phục và dễ thấy rõ tiềm năng ứng dụng của hệ thống ASR.

5. KẾT LUẬN

Với những phân tích và kết quả thu được thí nghiệm trên, hệ thống ASR đã cho thấy khả năng ứng dụng cho việc khai thác và triển khai rộng rãi hơn nữa trong tương lai. Đây cũng là nền tảng làm tiền đề cho các nghiên cứu tiếp theo với quy mô lớn hơn, nhằm mục đích xây dựng hệ thống nhận dạng giọng nói phục vụ cho hàng không Việt Nam.

TÀI LIỆU THAM KHẢO

- [1]. ICAO Secretariat. Outlook for air transport to the year 2025. Report No. Cir, 313, 2007.
- [2]. Thomas Pellegrini, Jerome Farinas, Estelle Delpech, and Francois Lancelot, “The Airbus Air Traffic Control speech recognition 2018 challenge: towards ATC automatic transcription and call sign detection”, Airbus, 2018.
- [3]. Y. LIN, M. RUAN, K. CAI, D. LI, Z. ZENG, F.LI, B. YANG, “Identifying and managing risks of AI-driven operations: A case study of automatic speech recognition for improving air traffic safety”, *Chinese Journal of Aeronautics*, Vol. 36, Issue 4, pp. 336-386, 2023.
- [4]. VTV, "Hành khách đường hàng không tăng hơn 90% trong hai tháng đầu năm", 22/02/2023. [Trực tuyến]. Địa chỉ: <https://vtv.vn/kinh-te/hanh-khach-duong-hang-khong-tang-hon-90-trong-hai-thang-dau-nam-20230222143619948.htm> [Truy cập 8/6/2023].
- [5]. X. Huang, Al. Acero, H. W. Hon, and R. F. Reddy. Spoken language processing: A guide to theory, algorithm, and system development. *Prentice Hall PTR*, 2001.
- [6]. H.D. Kopald, A. Chanen, Shuo Chen, E.C. Smith, and R.M. Tarakan. Applying automatic speech recognition technology to air traffic management. *In Digital Avionics Systems Conference (DASC), 2013 IEEE/AIAA 32nd*, pp. 6C3-1-6C3-15, Oct 2013.
- [7]. V. N. Nguyen and H. Holone, “Possibilities, Challenges and the State of the Art of Automatic Speech Recognition in Air Traffic Control”, *Engineering and Technology International Journal of Computer and Information Engineering*, Vol. 9, No. 8, pp. 1933-1940, 2015.
- [8]. V. Radha and C. Vimala. A review on speech recognition challenges and approaches. *doaj. org*, 2(1): pp.1-7, 2012.
- [8]. L. V. Vũ, H. L. U. Thục, “ỨNG DỤNG KỸ THUẬT NHẬN DẠNG TIẾNG NÓI TỰ ĐỘNG VÀO HỖ TRỢ HOẠT ĐỘNG KSKL”, *Kỷ yếu Hội nghị KHCN Quốc gia lần thứ XIII về Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin (FAIR)*, Nha Trang, ngày 8-9/10/2020. DOI: 10.15625/vap.2020.00223.