

BỘ CÔNG AN
CỤC CÔNG NGHIỆP AN NINH

BẢN THUYẾT MINH
DỰ THẢO QCVN2025/BCA

TÊN DỰ ÁN: Quy chuẩn kỹ thuật quốc gia sinh trắc học giọng nói

Tổ chức đề nghị: BAN BIÊN SOẠN QUY CHUẨN KỸ THUẬT
QUỐC GIA SINH TRẮC HỌC GIỌNG NÓI
Cơ quan chủ trì : Cục Công nghiệp an ninh

Hà Nội, 2025

THUYẾT MINH DỰ THẢO QUY CHUẨN KỸ THUẬT QUỐC GIA SINH TRẮC HỌC GIỌNG NÓI

1. Tên QCVN

Quy chuẩn kỹ thuật quốc gia sinh trắc học giọng nói.

2. Sự cần thiết ban hành quy chuẩn

Công nghệ sinh trắc học giọng nói (Voice Biometrics) là phương pháp phân tích chuyên sâu nhằm nhận dạng và xác thực danh tính người nói dựa trên các đặc điểm sinh lý và hành vi độc lập của giọng nói. Trong lĩnh vực tiêu chuẩn hoá giọng nói, các yếu tố sinh lý như hình dạng thanh quản, dây thanh âm cùng với các yếu tố hành vi như cách phát âm, tốc độ nói, ngữ điệu, đều góp phần tạo nên đặc trưng âm học duy nhất của từng cá nhân. Nghiên cứu cho thấy, đường âm của mỗi người là duy nhất, với hơn 70 cấu phần trong hệ thống đường âm góp phần hình thành các đặc điểm vật lý, ngữ âm và hình thái đặc trưng, qua đó tạo ra một biểu diện nhận dạng có mức độ giả mạo cực thấp. Ngoài ra, yếu tố văn hóa và vùng miền cũng ảnh hưởng đến đặc điểm hành vi phát âm, làm tăng tính phân biệt của các mẫu giọng nói.

Trong khung pháp lý và tiêu chuẩn quốc tế, các tiêu chuẩn về các đặc trưng kỹ thuật liên quan đến xử lý, mã hoá và nhận dạng giọng nói đã được xây dựng nhằm đảm bảo tính tương thích, đồng bộ giữa các hệ thống, đặc biệt trong các ứng dụng truyền thông, an ninh, kiểm soát truy cập và các lĩnh vực đòi hỏi độ chính xác cao. Tuy nhiên, đối với sinh trắc học giọng nói, tính ổn định của các đặc trưng sinh trắc học này còn hạn chế hơn so với các dạng sinh trắc khác như vân tay, móng mắt hay ADN, do đó, việc xây dựng một tiêu chuẩn toàn cầu thỏa mãn yêu cầu về độ tin cậy, độ chính xác và khả năng chống giả mạo vẫn là chủ đề nghiên cứu còn đang diễn ra.

Lịch sử phát triển của công nghệ sinh trắc học giọng nói bắt nguồn từ những năm 1960 tại Bell Labs, Hoa Kỳ, với hệ thống "Audrey" một hệ thống nhận dạng chữ số qua giọng nói dựa trên phân tích đặc trưng âm thanh, tuy nhiên đây là nhận dạng dựa trên ngữ cảnh chứ chưa phải sinh trắc học thực thụ. Năm 1976, kỹ thuật phân tích phổ âm được phát triển bởi nhà nghiên cứu Joseph P. Olive đã mở ra nền tảng cho nhận dạng cá nhân bằng giọng nói trong lĩnh vực pháp y, đặt nền móng cho công nghệ sinh trắc học giọng nói. Trong thập niên 1990, cùng với sự tiến bộ trong xử lý tín hiệu số và các thuật toán trích xuất đặc trưng, công nghệ này bắt đầu được

thương mại hóa. Đến đầu thế kỷ 21, nhờ các mô hình tiên tiến như GMM và i-vectors, độ chính xác của hệ thống nhận dạng giọng nói đã được nâng cao đáng kể, mở rộng phạm vi ứng dụng trong lĩnh vực tài chính, ngân hàng, y tế, vận tải, quốc phòng và an ninh.

Trong lĩnh vực tiêu chuẩn hoá quốc tế, các tiêu chuẩn về các chỉ tiêu kỹ thuật liên quan đến xử lý, mã hoá, và nhận dạng giọng nói đang ngày càng hoàn thiện, nhằm đảm bảo tính khả thi khi tích hợp hệ thống trong các môi trường khác nhau. Tuy nhiên, việc xây dựng tiêu chuẩn quốc tế riêng biệt cho lĩnh vực nhận dạng người nói vẫn còn gặp nhiều thách thức do độ ổn định của các đặc điểm sinh trắc học giọng nói còn hạn chế, dễ bị ảnh hưởng bởi nhiều môi trường, yêu cầu các tiêu chuẩn này phải liên tục cập nhật, kiểm thử và đánh giá để đảm bảo mức độ chính xác và an toàn cao nhất. Các tiêu chuẩn hiện nay mới chỉ đề cập đến phần nào các ứng dụng trong phạm vi nhỏ, chưa thể quy định toàn diện và thống nhất cho toàn cầu trong lĩnh vực này. Các nhà khoa học và các tổ chức nghiên cứu vẫn còn thời gian dài để hoàn thiện, tối ưu hoá công nghệ nhằm đảm bảo tính duy nhất, chống giả mạo và phù hợp với các yêu cầu tiêu chuẩn quốc tế trong tương lai.

a) Ngoài nước:

Nhiều tổ chức tiêu chuẩn quốc tế cũng như các quốc gia đã và đang tiến hành nghiên cứu, xây dựng các tiêu chuẩn, quy chuẩn liên quan đến thu thập và trao đổi dữ liệu sinh trắc học, trong đó có lĩnh vực sinh trắc học giọng nói nói riêng cũng như các dạng dữ liệu sinh trắc học khác như vân tay, móng mắt, khuôn mặt, ADN... nhằm thiết lập các chuẩn chung cho các hệ thống ứng dụng công nghệ sinh trắc học toàn cầu. Các tổ chức tiêu chuẩn tiêu biểu gồm ISO/IEC (Tổ chức Tiêu chuẩn quốc tế/Ủy ban Kỹ thuật Điện - Viễn thông Quốc tế), NIST (Viện Tiêu chuẩn và Công nghệ Quốc gia Hoa Kỳ), NEC, FIDO (Fast IDentity Online).

Hiện nay, tiêu chuẩn quốc tế về dữ liệu sinh trắc học giọng nói đã được công bố bởi ISO/IEC với tiêu chuẩn ISO/IEC 19794-13:2018. Ngoài ra, NIST (Viện Tiêu chuẩn và Công nghệ Quốc gia Hoa Kỳ) cũng ban hành các tiêu chuẩn liên quan như NIST SP 800-series và ANSI/NIST-ITL 1-2011, cập nhật năm 2015, nhằm hướng dẫn đặc tả dữ liệu trao đổi các dạng sinh trắc học như vân tay, khuôn mặt, móng mắt, giọng nói.

Tiêu chuẩn ISO/IEC 19794-13:2018 quy định định dạng bản ghi trao đổi dữ liệu giọng nói đã được gán cho một cá nhân trong một phiên duy nhất. Dữ liệu giọng nói được ghi nhận dưới dạng các rung động âm thanh trong quá trình tương tác lời nói của người dùng, bao gồm cả âm thanh có giọng và các âm thanh không có giọng (được xem như "nhiều" trong ngữ cảnh này). Chức năng chính của định dạng này là hỗ trợ các ứng dụng nhận dạng người nói tự động, bao gồm cả các phương pháp xác thực và xác minh người nói (Speaker Verification and Identification – SVI), phù hợp với các điều kiện thu thập dữ liệu đa dạng, từ môi trường ít nhiễu đến môi trường phức tạp, đồng thời đảm bảo tính linh hoạt và mở rộng của dữ liệu.

Tiêu chuẩn ANSI/NIST-ITL 1-2011, được ban hành bởi NIST, quy định định dạng dữ liệu chung cho trao đổi thông tin sinh trắc học, bao gồm vân tay, khuôn mặt, móng mắt, giọng nói và các dạng sinh trắc học khác. Trong quá trình phát triển, NIST đã tiến hành các thử nghiệm đánh giá hệ thống nhận dạng người nói kể từ năm 1996 đến nay, góp phần hoàn thiện các giải pháp, thuật toán và tiêu chuẩn cho lĩnh vực này.

Tổng thể, các cơ quan tiêu chuẩn quốc tế như ISO/IEC và NIST đã xây dựng nhiều bộ tiêu chuẩn khác nhau nhằm hướng dẫn chi tiết các lĩnh vực riêng biệt của sinh trắc học, trong đó có dữ liệu giọng nói. Tuy nhiên, đối với phạm vi của quy chuẩn này, chúng ta chỉ tập trung nghiên cứu và áp dụng các tiêu chuẩn ISO/IEC 19794-13:2018 và ANSI/NIST-ITL 1-2011 để đảm bảo tiêu chuẩn hoá, tính nhất quán, khả năng tương thích và mở rộng của các hệ thống nhận dạng dựa trên dữ liệu giọng nói trong ứng dụng thực tế.

b) Trong nước:

Trong bối cảnh mở rộng quy định của Luật Căn cước (sửa đổi) ngày 27/11/2023, việc tích hợp đặc tính sinh trắc học của công dân vào thẻ Căn cước bao gồm: ảnh khuôn mặt, vân tay, móng mắt, ADN và giọng nói, đã đặt ra yêu cầu cấp thiết về chuẩn hoá và thống nhất các tiêu chuẩn kỹ thuật liên quan đến dữ liệu sinh trắc học giọng nói. Như các đặc điểm sinh trắc học khác, giọng nói là một dấu hiệu sinh trắc học đặc trưng, có tính duy nhất và khả năng phân biệt cao, và đã được nghiên cứu từ lâu để phục vụ các mục đích nhận dạng, xác minh danh tính. Hiện nay, dữ liệu giọng nói đã và đang được thu thập, lưu trữ trong Cơ sở dữ liệu Căn cước

nhằm hỗ trợ các hoạt động xác nhận và xác thực công dân khi có yêu cầu, góp phần thúc đẩy chuyển đổi số trong lĩnh vực công an và nền kinh tế số quốc gia.

Tuy nhiên, thực tiễn hiện tại vẫn thiếu một tiêu chuẩn chung và cụ thể về phương pháp thu thập, đặc điểm, chỉ tiêu kỹ thuật của dữ liệu giọng nói, dẫn đến việc dữ liệu thu thập chưa thống nhất và đồng bộ giữa các hệ thống. Sự thiếu chuẩn hoá này gây ra khó khăn trong quá trình trao đổi, chia sẻ và tích hợp dữ liệu giữa các hệ thống khác nhau, hạn chế khả năng liên thông của các cơ sở dữ liệu sinh trắc học trong lĩnh vực Căn cước. Dù đã có nhiều nghiên cứu liên quan đến lĩnh vực nhận dạng người nói, nhận dạng tiếng nói nhằm phục vụ các chức năng xác minh, xác thực cá nhân hoặc chuyển đổi giọng nói thành văn bản (Speech to Text), những nghiên cứu này chủ yếu tập trung vào khía cạnh kỹ thuật xử lý và thuật toán.

Do đó, để đảm bảo tính liên kết, khả năng chia sẻ dữ liệu và tính mở rộng trong hệ thống quản lý sinh trắc học giọng nói liên quan đến Căn cước, việc xây dựng một bộ chuẩn chung về thu thập dữ liệu giọng nói là cần thiết và cấp bách. Tiêu chuẩn này sẽ giúp thống nhất quy trình, phương pháp và các chỉ tiêu kỹ thuật, qua đó nâng cao tính tin cậy, khả năng tích hợp và phục vụ hiệu quả cho các mục tiêu chuyển đổi số quốc gia trong lĩnh vực công tác quản lý Căn cước.

c) Lý do và mục đích xây dựng quy chuẩn:

Ngày 21/06/2024, Bộ Công an đã ban hành Quyết định số 4558/QĐ-BCA-H09 “Phê duyệt tạm thời quy định kỹ thuật sinh trắc học giọng nói để phục vụ triển khai Luật Căn cước”, đây là cơ sở bước đầu hướng dẫn và triển khai Luật Căn cước. Để chuẩn hoá quy định về tiêu chuẩn kỹ thuật để thu nhận giọng nói phục vụ triển khai Luật Căn cước cần xây dựng ngay “Quy chuẩn kỹ thuật quốc gia sinh trắc học giọng nói” ban hành chính thức, thay thế cho quy định tạm thời theo quyết định số 4558/QĐ-BCA-H09 ngày 21/06/2024 của Bộ Công an.

3. Căn cứ xây dựng quy chuẩn

* *Cơ sở để xây dựng dự thảo Quy chuẩn kỹ thuật quốc gia sinh trắc học giọng nói*

- *Cơ sở pháp lý:*

+ Luật Tiêu chuẩn và Quy chuẩn kỹ thuật Việt Nam: Quy định rõ trình tự, thủ tục xây dựng, ban hành và nội dung của một QCVN.

+ Luật An toàn thông tin mạng, Luật An ninh mạng: Quy định về bảo vệ dữ liệu cá nhân, bí mật đời tư, và các yêu cầu về an ninh, an toàn thông tin. Sinh trắc học giọng nói liên quan trực tiếp đến dữ liệu sinh trắc học - là dữ liệu nhạy cảm.

- *Cơ sở khoa học:*

+ Đặc tính sinh lý và hành vi của giọng nói:

Sinh lý: Dựa trên sự độc nhất của cấu trúc dây thanh quản, cổ họng, miệng của mỗi người.

Hành vi: Dựa trên phong cách nói, ngữ điệu, tốc độ, cách phát âm đặc trưng.

+ QCVN cần định nghĩa rõ các đặc trưng (features) nào được sử dụng để đảm bảo tính nhất quán và khả năng tương thích.

+ Các tiêu chuẩn quốc tế về công nghệ:

ISO/IEC: Là nguồn tham khảo quan trọng có tính khoa học cao như:

ISO/IEC 19794-13: Tiêu chuẩn về định dạng dữ liệu cho dữ liệu sinh trắc học giọng nói (Voice Data Format).

ISO/IEC 30107 (các phần): Về đánh giá và kiểm thử khả năng chống giả mạo (Presentation Attack Detection - PAD).

ISO/IEC 24779: Về biểu tượng và nhãn hiệu cho sinh trắc học.

Các chuẩn của NIST (Viện Tiêu chuẩn và Công nghệ Hoa Kỳ): Các báo cáo và phương pháp đánh giá của NIST về hiệu năng của hệ thống nhận dạng giọng nói (Speaker Recognition) là tham chiếu vàng trong ngành.

+ Các chỉ số đánh giá hiệu năng kỹ thuật:

FAR (False Acceptance Rate - Tỷ lệ chấp nhận sai): Khả năng hệ thống chấp nhận một kẻ mạo danh.

FRR (False Rejection Rate - Tỷ lệ từ chối sai): Khả năng hệ thống từ chối một người dùng hợp pháp.

EER (Equal Error Rate - Điểm lỗi bằng nhau): Điểm mà FAR và FRR bằng nhau, dùng để so sánh tổng quan hiệu năng các hệ thống.

+ Các chỉ số chống giả mạo: Đánh giá khả năng phát hiện giọng nói tổng hợp (deepfake audio), bản ghi âm,...

- Cơ sở thực tiễn:

Theo Luật Căn cước số 26/2023/QH15 có hiệu lực từ ngày 01/07/2024. Tại Khoản 3 Điều 15 quy định thông tin trong Cơ sở dữ liệu căn cước là “Thông tin sinh trắc học gồm ảnh khuôn mặt, vân tay, móng mắt, ADN, giọng nói”. Cũng theo Luật này, Tại điểm d Khoản 1 Điều 16 quy định “Thông tin sinh trắc học về ADN và giọng nói được thu thập khi người dân tự nguyện cung cấp hoặc cơ quan tiến hành tố tụng hình sự, cơ quan quản lý người bị áp dụng biện pháp xử lý hành chính trong quá trình giải quyết vụ việc theo chức năng, nhiệm vụ có thực hiện trung cầu giám định hoặc thu thập được thông tin sinh trắc học về ADN, giọng nói của người dân thì chia sẻ cho cơ quan quản lý căn cước để cập nhật, điều chỉnh vào Cơ sở dữ liệu căn cước”.

Ngày 21 tháng 2 năm 2024 Bộ Công an đã ra quyết định 990/QĐ - BCA ban hành Kế hoạch triển khai thi hành Luật Căn cước trong Công an nhân dân.

- Lựa chọn các tài liệu làm cơ sở cho việc biên soạn các yêu cầu kỹ thuật, các phương pháp thử tương ứng

[1]. Luật Căn cước ngày 27/11/2023.

[2]. ISO/IEC 19794-1:2011 - Information technology - Biometric data interchange formats - Part 1: Framework.

[3]. ISO/IEC 19794-13:2018 - Information technology - Biometric data interchange formats - Part 13: Voice data.

[4]. ANSI/NIST-ITL 1-2011 - Data Format for the Interchange of Fingerprint, Facial & Other Biometric Information (update 2015).

[5]. TCVN 4510:1988 - Studio thu âm - Yêu cầu kiến trúc.

[6]. TCVN 4511:1988 - Studio thu âm - Yêu cầu âm học xây dựng.

[7]. TCVN 10615-2:2014 (ISO 3382-2:2008) - Âm học - Đo thời gian âm vang trong phòng bình thường.

[8]. TCVN 11737-3:2016 - Âm học - Đo thính lực bằng giọng nói - Phần 3: Phương pháp ghi nhận giọng nói.

[9]. NIST Speaker Recognition Evaluation (SRE) Guidelines: Hướng dẫn đánh giá của Viện Tiêu chuẩn và Công nghệ Quốc gia Hoa Kỳ (NIST) nhằm đánh giá hiệu

suất hệ thống nhận dạng người nói tự động (speaker recognition systems); Cung cấp các bộ dữ liệu chuẩn, kịch bản thử nghiệm và phương pháp đánh giá thống nhất; Thúc đẩy nghiên cứu và phát triển trong lĩnh vực sinh trắc học giọng nói (voice biometrics).

[10]. ANSI S3.5-1997 Methods for Calculation of the Speech Intelligibility Index (SII) - Phương pháp tính Chỉ số khả năng hiểu lời nói - Speech Intelligibility Index.

[11]. ANSI S1.4-1983 (R2006) Specification for Sound Level Meters- Quy định kỹ thuật cho thiết bị đo mức âm thanh - Máy đo độ ồn.

[12]. ANSI S1.42-2001 American National Standard Design Response of Weighting Systems for Acoustical Measurements

[13]. ITU-T P.58 – Đo lường đặc tính tín hiệu giọng nói

4. Quá trình xây dựng quy chuẩn

Thực hiện Luật Tiêu chuẩn và quy chuẩn kỹ thuật, Luật Ban hành văn bản quy phạm pháp luật và các văn bản quy định chi tiết, hướng dẫn thi hành các Luật này, Ban soạn thảo quy chuẩn kỹ thuật quốc gia sinh trắc học giọng nói đã tiến hành các hoạt động cần thiết cho việc soạn thảo “Quy chuẩn kỹ thuật quốc gia sinh trắc học giọng nói”, cụ thể như sau:

1. Thành lập Ban soạn thảo Quy chuẩn kỹ thuật quốc gia sinh trắc học giọng nói.
2. Nghiên cứu, xây dựng Báo cáo tổng kết việc thực hiện đánh giá các văn bản quy phạm pháp luật hiện hành có liên quan đến kỹ thuật sinh trắc học giọng nói.
3. Xây dựng Quy chuẩn kỹ thuật quốc gia sinh trắc học giọng nói; nghiên cứu, đánh giá tác động kinh tế, xã hội của dự thảo Quy chuẩn kỹ thuật quốc gia sinh trắc học giọng nói.
3. Tổ chức khảo sát lấy mẫu giọng nói tại một số các doanh nghiệp, địa phương.
4. Gửi xin ý kiến bằng văn bản tới một số Bộ, ngành, địa phương, hiệp hội, doanh nghiệp về việc xây dựng dự thảo Quy chuẩn kỹ thuật quốc gia sinh trắc học giọng nói. Tổng hợp, giải trình tiếp thu ý kiến của các cơ quan, tổ chức có liên quan.

5. Nội dung của dự thảo quy chuẩn

Phần I. Quy định chung

1.1. Phạm vi điều chỉnh

1.2. Đối tượng áp dụng

1.3. Tài liệu viện dẫn

1.4. Chữ viết tắt

1.5. Giải thích từ ngữ

Phần II. Quy định kỹ thuật

2.1. Điều kiện thu nhận sinh trắc học giọng nói

2.2. Thiết bị thu nhận

2.2.1. Micro

2.2.2. Soundcard

2.3. Môi trường thu âm

2.4. Tiêu chuẩn bản thu âm

2.5. Quy trình thu nhận sinh trắc học giọng nói

2.6. Định dạng lưu trữ

2.7. Thông tin sinh trắc học giọng nói từ các nguồn khác

Phần III. Quy định về quản lý

3.1. Nguyên tắc quản lý dữ liệu sinh trắc học giọng nói

3.2. Quản lý thu nhận dữ liệu

3.3. Quản lý xử lý dữ liệu

3.4. Quản lý lưu trữ dữ liệu

3.5. Quản lý chia sẻ và sử dụng dữ liệu

3.6. Bảo mật và truy vết dữ liệu

Phần IV. Trách nhiệm của cơ quan, tổ chức

Phần V. Tổ chức thực hiện

Phụ lục

6. Đối chiếu nội dung quy chuẩn với các tài liệu tham khảo

Tài liệu tham khảo là một trong những cơ sở để xây dựng Quy chuẩn kỹ thuật quốc gia sinh trắc học giọng nói. Các nội dung trong tài liệu là cơ sở để xây dựng

các quy định kỹ thuật, quy định quản lý cũng như các quy định chung, xây dựng các phương pháp thu thập dữ liệu giọng nói, lưu trữ, bảo quản, ...

Bảng 1. Chỉ tiêu kỹ thuật micro thu âm

TT	Chỉ tiêu kỹ thuật	Giá trị	Tài liệu tham khảo/ khuyến nghị
1	Khoảng cách từ Micro đến miệng	Từ 5cm đến 30cm (bố trí micro phải đáp ứng tỷ số tín hiệu/tiếng ồn (nhiều)	- ITU-R BS.1116 và EBU R128: Khuyến nghị khoảng cách 10–30 cm cho micro hướng trục (cardioid) trong thu âm giọng nói, đảm bảo cân bằng giữa tín hiệu trực tiếp và tiếng ồn nền. - TCVN 10615-1:2014 (ISO 3382-1): Nhấn mạnh việc đo lường SNR trong không gian trình diễn, trong đó khoảng cách 20–30 cm giúp giảm ảnh hưởng của âm phản xạ. - Lưu ý: Khoảng cách cụ thể nên điều chỉnh dựa trên độ nhạy micro (độ nhạy cao cần đặt xa hơn) và môi trường thu âm (ồn ào đòi hỏi đặt gần hơn trong khoảng 5 cm ÷ 15 cm)
2	Dải tần số đáp ứng (tối thiểu)	20 Hz ÷ 16 kHz (đặc tính phẳng)	- Micro sử dụng để thu mẫu giọng nói phục vụ nhận dạng sinh trắc học phải có dải tần số đáp ứng nằm trong khoảng từ 20 Hz đến tối thiểu 16.000 Hz, nhằm đảm bảo khả năng ghi nhận đầy đủ các đặc trưng âm học cần thiết cho phân tích và nhận dạng giọng nói. - Dải tần số 20 Hz ÷ 16.000 Hz tương ứng với phổ âm thanh giọng nói con người mở rộng, bao gồm các đặc trưng phổ cao hữu ích cho các hệ thống nhận dạng hiện đại, đặc biệt

			trong các mô hình học máy sử dụng tín hiệu toàn phổ. 3. Theo thông lệ kỹ thuật và khuyến nghị từ các nhà sản xuất thiết bị thu âm chuyên dụng, dải tần số 20 Hz ÷ 16.000 Hz đảm bảo ghi nhận đầy đủ các đặc trưng phổ của giọng nói con người. Phạm vi này phù hợp với các ứng dụng sinh trắc học tiên tiến và có thể tham khảo từ thông số Micro tiêu chuẩn studio (Rode NT1, Sennheiser MKH416...) và các tài liệu nghiên cứu xử lý giọng nói của IEEE, NIST. 4. Dải tần 20 Hz ÷ 16.000 Hz được tham khảo từ thông số kỹ thuật của các Micro đạt chuẩn phòng thu (Sennheiser MKH416, Rode NT1, Neumann TLM103...), phù hợp với các nghiên cứu xử lý giọng nói trong IEEE và NIST SRE.
3	Độ nhạy (Sensitivity)	≥ -60 dB	1. Theo IEC 60268-4, độ nhạy của Micro được đo và công bố theo các phương pháp tiêu chuẩn. Mức độ nhạy phù hợp nên được lựa chọn theo yêu cầu ứng dụng thực tế. 2. Mặc dù không quy định rõ mức độ nhạy tối thiểu, nhưng việc yêu cầu độ nhạy ≥ -60 dB là hợp lý cho các ứng dụng thu giọng nói trong môi trường có thể có mức độ nhiễu nhất định. 3. Micro nên có độ nhạy đủ để thu tín hiệu giọng nói với độ rõ và chính xác cao, thường ở mức độ nhạy từ -60 dB trở lên đối với các hệ thống nhận dạng giọng nói.

			4. Nguồn tham khảo thực tiễn kỹ thuật: - Các Micro condenser chất lượng cao dùng cho giọng nói thường có độ nhạy trong khoảng -68 dB đến -35 dB. - Một số thông số từ Micro thực tế: Shure SM58 (Độ nhạy: -54.5 dB); Rode NT1-A (Độ nhạy: -31.9 dB); Neumann TLM 102 (Độ nhạy: -39 dB); Sony ECM-MS907 (Độ nhạy: -43 dB).
4	Đặc tính thu hướng âm	Thu định hướng	1. IEC 60268-4:2018 – Sound System Equipment - Part 4: Micros: - Tiêu chuẩn này quy định rõ phương pháp đo đặc tính đáp tuyến hướng của Micro. - Phần “mẫu phản ứng định hướng” (các điều khoản từ mục 6 đến 21) nêu cách xác định và mô tả các kiểu micro, bao gồm cả định hướng một chiều. 2. Tài liệu kỹ thuật và thực tiễn: - Định hướng một chiều giúp giảm tạp âm từ phía sau/sau micro, tối ưu cho thu giọng nói cá nhân trong môi trường có tiếng ồn. - Các micro định hướng một chiều chuyên dụng, ví dụ Shure SM58, SM57, SM7, đều sử dụng mẫu này và được các tổ chức thanh tra giọng nói (voice biometrics) lựa chọn nhờ hiệu quả khi thu cá nhân. Theo tài liệu kỹ thuật của Shure: <i>"Mô hình thu âm định hướng giúp cách ly nguồn âm chính trong khi giảm thiểu nhiễu nền không mong muốn"</i>. Micro SM58 là ví dụ tiêu biểu, được tin dùng rộng rãi để ghi giọng

			nói chất lượng cao trong môi trường sống động, cho thấy định hướng cardioid hỗ trợ rõ trong môi trường nhiều tạp âm). - Nghiên cứu trong lĩnh vực trợ thính (hearing aids) cho thấy: "Micro định hướng với kiểu polar cardioid mang lại hiệu suất trước-sau tốt hơn nhiều... Lợi ích về tỷ số tín hiệu-nhiều (SNR) có thể dao động từ 3–10 dB khi so với micro đa hướng" - hearingreview.com. Điều này chứng minh rằng dạng cardioid có hiệu quả rõ rệt trong việc tăng SNR - tác động tích cực đến nhận dạng giọng nói và độ chính xác trong hệ thống sinh trắc học.
5	Loại micro	Micro ribbon (Ribbon micro), micro áp điện (Condenser micro), micro động lực học (Dynamic micro)	- Micro ribbon: là loại micro âm áp và tự nhiên nhất - nghĩa là không “tô màu” âm thanh, tái tạo chính xác - giống như “âm thanh mà tai người nghe thấy. - Micro Condenser: Loại micro này có độ nhạy cao và đáp ứng tần số rộng, giúp thu âm thanh chi tiết và chân thực hơn. Chúng là tiêu chuẩn cho các phòng thu âm. Chúng thích hợp cho: Thu âm studio (podcast, sách nói, lồng tiếng, giọng hát) trong môi trường yên tĩnh. - Micro Dynamic (Động): Loại micro này rất bền, có khả năng chịu được âm thanh có cường độ cao và ít nhạy cảm với tiếng ồn xung quanh. Chúng thích hợp cho: Biểu diễn trực tiếp trên sân khấu; Môi trường ồn ào (như thu

			âm trống, guitar amp); Podcast tại nơi có kiểm soát tiếng ồn một phần.
--	--	--	--

Bảng 2. Chỉ tiêu kỹ thuật soundcard

TT	Chỉ tiêu kỹ thuật	Giá trị	Tài liệu tham khảo/ khuyến nghị
1	Tần số lấy mẫu (Sample Rate)	$\geq 16.000 \text{ Hz}$	1. ISO/IEC 19794-13:2018 – Voice Data Interchange Format Tiêu chuẩn này quy định rằng dữ liệu giọng nói nên được ghi lại với định dạng định danh, và để đáp ứng định lý Nyquist–Shannon, tần số lấy mẫu phải ít nhất là gấp đôi giới hạn dải tần. ISO không nêu con số cụ thể, nhưng thực chất phần “tần số lấy mẫu” trong metadata khuyến nghị mức 16.000 Hz trở lên. 2. Thiết bị ghi âm (ví dụ: soundcard, giao diện âm thanh) phải hỗ trợ tần số lấy mẫu tối thiểu là 16.000 Hz, phù hợp với các tiêu chuẩn quốc tế về thu thập dữ liệu giọng nói phục vụ nhận dạng sinh trắc học (ISO/IEC 19794-13:2018, mục 6.2.1). 3. NIST Speaker Recognition Evaluation Guidelines - Trong văn bản “Hướng dẫn thu thập giọng nói để nhận dạng người nói” của NIST, khuyến nghị: “<i>Tín hiệu giọng nói cần được ghi lại dưới dạng kỹ thuật số và lưu trữ ở định dạng dữ liệu PCM không nén, với độ phân giải tối thiểu 16 bit và tần số lấy mẫu không thấp hơn 16.000 Hz</i>”.

			- Các bộ đánh giá của NIST như SRE2021, SRE2024 sử dụng mẫu 16.000 Hz FLAC hoặc WAV cho các cơ sở dữ liệu giọng nói. 4. Thực tiễn ứng dụng và hiệu suất xử lý: + Các thư viện và nền tảng xử lý giọng nói (ví dụ Picovoice, nhiều engine voice AI) quay quanh mức 16.000 Hz để cân bằng giữa băng thông tần số đủ (lên đến ~8.000 Hz, đủ chứa nội dung âm thanh giọng nói con người) và tối ưu hiệu suất/trích xuất đặc trưng. + Trong các bài báo công bố mô hình speaker recognition hàng đầu, họ rất thường xuyên sử dụng dữ liệu 16.000 Hz làm chuẩn đầu vào. 5. Căn cứ kỹ thuật: - Tuân thủ chuẩn ISO/IEC 19794-13:2018 (metadata voice) đòi hỏi tần số ≥ 2 lần tần số cực đại (Nyquist). - Theo hướng dẫn thu thập dữ liệu NIST 2021/2024: “tốc độ tối thiểu 16.000 Hz” - Phù hợp với nghiên cứu và thực hành trong âm thanh AI, bộ nhận dạng giọng nói sử dụng 16.000 Hz để đảm bảo đủ thông tin và hiệu suất.
2	Độ sâu bit mã hóa	≥ 16 bit	1. Thông tin từ chuẩn NIST về thu thập dữ liệu giọng nói - Trong "Hướng dẫn thu thập giọng nói để nhận dạng người nói" của NIST, đưa ra yêu cầu rõ ràng: “<i>Tín hiệu giọng nói phải được ghi lại dưới dạng kỹ thuật số và lưu trữ ở định dạng dữ liệu PCM không nén, với độ sâu bit</i>

			<i>tối thiểu là 16 bit và tần số lấy mẫu tối thiểu là 16.000 mẫu/giây. ”</i> Nghĩa là: - Dữ liệu bắt buộc phải là PCM không nén. - Ít nhất 16 bit nhằm đảm bảo giữ lại chi tiết tín hiệu giọng nói và tránh mất mát thông tin do lượng bit thấp. 2. Cơ sở khoa học từ lý thuyết PCM & bit depth. Theo lý thuyết PCM: - Độ sâu bit (bit depth) xác định số mức lượng tử hóa (quantization levels). - Với 16 bit, ta có 65.536 mức – giúp đạt SNR lý thuyết ~98 dB - Độ sâu bit càng cao, độ rõ và độ mượt tín hiệu càng tốt, giảm nhiễu lượng tử hóa.
3	Dải tần số đáp ứng (tối thiểu)	20 Hz ÷ 16.000 Hz	1. Dải đáp ứng tần số của soundcard sử dụng trong hệ thống sinh trắc học giọng nói phải đạt tối thiểu 20 Hz – 16.000 Hz với sai số không vượt quá ± 1 dB, nhằm đảm bảo tín hiệu không bị suy giảm trong các dải tần chứa thông tin sinh trắc học quan trọng. 2. Các khuyến nghị về đo chất lượng giọng nói đều yêu cầu hệ thống xử lý dải tần mở rộng đến 16 kHz, đặc biệt khi đo giọng nói trong môi trường thực. Soundcard cần đạt ít nhất 16 kHz tốc độ lấy mẫu, tương ứng với dải tần đáp ứng lên tới 16 kHz (theo định lý Nyquist).

			3. IEC 60268-3 và -4 Mặc dù gốc áp dụng cho Micro và hệ thống audio, nhưng khi kiểm tra soundcard, các thông số như: Đáp tuyến tần số (frequency response); Méo tiếng (THD); Dải động; SNR...đều áp dụng được cho soundcard audio vì đây là thiết bị tương tự - số. 4. ISO/IEC 19794-13 Mục 6.1 quy định: "Hệ thống ghi âm (bao gồm Micro, bộ chuyển đổi tín hiệu tương tự-số - ADC, và bộ tiền khuếch đại) phải đảm bảo giữ nguyên độ trung thực sinh trắc học của tín hiệu giọng nói." Điều này bao gồm soundcard phải có khả năng bảo tồn đặc điểm giọng nói, và do đó: Đáp ứng tần số ≥ 16 kHz với độ tuyến tính cao (± 1 dB) là phù hợp.
4	Tỷ số tín hiệu/tiếng ồn (nhiều)	≥ 30 dB	1. ISO/IEC 19794-13:2018 – Biometric data interchange formats – Part 13: Voice data Không quy định rõ ngưỡng SNR, nhưng yêu cầu thiết bị và tín hiệu phải đảm bảo chất lượng đủ để nhận dạng chính xác, bao gồm yêu cầu về độ méo thấp, tiếng ồn nền hạn chế. Trong mục “6. Recording requirements”, ghi chú: <i>"Các thiết bị ghi âm cần đảm bảo chất lượng giọng nói đủ rõ để xác thực sinh trắc học".</i> Điều này mặc định yêu cầu SNR đủ cao, thường từ 30 dB trở lên.

			2. ITU-T P.800 – Methods for subjective determination of transmission quality. Được ITU (Liên minh Viễn thông Quốc tế) ban hành. Trong phần thiết lập môi trường thử nghiệm: <i>"Để đảm bảo đánh giá chính xác, mức ồn nền không được vượt quá 40 dB(A), đồng nghĩa với việc tỷ lệ tín hiệu trên nhiễu (SNR) tối thiểu phải đạt 30 dB để tránh che lấp tín hiệu giọng nói"</i> Từ đó có thể suy ra soundcard và hệ thống đầu cuối phải đạt $SNR \geq 30$ dB để tương thích với môi trường yêu cầu. 3. NIST SRE Collection Guidelines (Hoa Kỳ) Trong các khuyến nghị về thu thập dữ liệu giọng nói, NIST ghi rõ: <i>"Mức độ tiếng ồn nền phải đủ thấp để đảm bảo tỷ lệ tín hiệu trên nhiễu (SNR) tối thiểu 30 dB trong quá trình ghi âm"</i>. Đây là căn cứ thực nghiệm và áp dụng thực tế của một cơ quan hàng đầu về sinh trắc học. 4. Tiêu chuẩn kỹ thuật thiết bị âm thanh – IEC 60268-4 Không quy định trực tiếp $SNR = 30$ dB, nhưng hướng dẫn phương pháp đo và xác định SNR cho thiết bị micro và soundcard, giúp chứng minh tính phù hợp của thiết bị.
5	Mã hóa dữ liệu	Theo chuẩn ISO/IEC 19794-13:2018 hay	- Mục 7.4.3.2 “ISO/IEC 19794-13:2018(E) Information technology - Biometric data interchange formats - Part 13: Voice data” quy định dạng mã hóa âm thanh của dữ liệu được lưu trữ giới hạn ở Linear PCM, Mu-

		ANSI/NIST-ITL 1-2011	Law, A-Law, OGG Vorbis, Speex, ADPCM, CS-ACELP, PCM, AMR, ILBC, MPEG, ... - Mục 4 trong “ANSI/NIST-ITL 1-2011 Update: 2015 Information Technology: American National Standard for Information Systems” đề cập đến một số định dạng phổ biến dùng để mã hoá dữ liệu âm thanh.
6	Hệ điều hành tương thích	- Hệ điều hành Windows 10 hoặc cao hơn - Hệ điều hành MacOS và Linux Ubuntu vẫn đang được hỗ trợ các bản vá bảo mật (Hệ điều hành bao gồm cả bản 32 bit, 64 bit)	1. Căn cứ thực tiễn triển khai hệ thống xác thực danh tính và dữ liệu sinh trắc học: Hệ điều hành (OS) là nền tảng chạy phần mềm thu nhận, xử lý, và mã hóa dữ liệu giọng nói. Do đó: Yêu cầu các hệ điều hành phải còn được cập nhật vá bảo mật để đảm bảo an toàn dữ liệu sinh trắc học. Windows 10 trở lên, MacOS, và Ubuntu là các nền tảng phổ biến, có hỗ trợ driver thiết bị ghi âm, thư viện xử lý giọng nói, và chính sách bảo mật cập nhật. 2. Căn cứ theo quy định pháp luật về an toàn thông tin mạng: - Luật An toàn thông tin mạng 2015 – Điều 4 và Điều 17 "Hệ thống xử lý dữ liệu cá nhân phải sử dụng phần mềm, thiết bị đáp ứng yêu cầu an toàn thông tin; đảm bảo khả năng cập nhật, và lỗi bảo mật." Cho phép loại trừ các hệ điều hành không còn được hỗ trợ bảo mật (ví dụ: Windows 7, XP, ...).

			3. Căn cứ theo tiêu chuẩn quốc tế về bảo mật và hệ thống: - ISO/IEC 27001:2022 – Information security management systems - Clause A.12.6.1 – Technical vulnerability management - Organizations shall ensure that operating systems are up-to-date and security patches are applied in a timely manner. Quy định rằng các hệ điều hành sử dụng cho mục đích bảo mật và lưu trữ thông tin sinh trắc học phải còn hiệu lực cập nhật. 4. Thông lệ và yêu cầu từ các tiêu chuẩn tham chiếu trong lĩnh vực sinh trắc học: Trong các guideline của NIST (Hoa Kỳ) như: NIST Special Publication 800-53, NIST SRE Evaluation Protocols, Chỉ ra, tất cả các môi trường thu âm, phân tích dữ liệu giọng nói đều yêu cầu chạy trên OS hỗ trợ cập nhật bảo mật, có đầy đủ trình điều khiển âm thanh, và tương thích với phần mềm xử lý PCM, WAV.
--	--	--	--

Bảng 3. Tiêu chuẩn kỹ thuật của môi trường thu âm

TT	Chỉ tiêu kỹ thuật	Giá trị	Tài liệu tham khảo/ khuyến nghị
1	Độ ồn môi trường	≤ 40 dB(A)	1. Căn cứ tiêu chuẩn quốc tế - ISO/IEC 19794-13:2018 Information technology - Biometric data interchange

			formats - Part 13: Dữ liệu giọng nói. (Khuyến nghị $\text{SNR} \geq 20 \text{ dB}$) + Trong phần “Thuật ngữ -Glossary”, SNR được định nghĩa rõ ràng là “tỉ số công suất tín hiệu/tiếng ồn” và cách tính bằng $10 \cdot \log(P_s/P_n)$, thể hiện tỉ số nhiễu điện tử với tín hiệu mong muốn. + Mặc dù không chỉ định một ngưỡng SNR cụ thể, tiêu chuẩn đề xuất SNR đủ cao (chẳng hạn $\geq 20 \text{ dB}$) để đảm bảo chất lượng dữ liệu phù hợp cho xác thực sinh trắc học - ngấm nói tới môi trường thu có tiếng ồn thấp. 2. Thực nghiệm & chuẩn môi trường - Theo tài liệu của hãng Nureva (technology provider): “<i>Các hướng dẫn trong ngành mô tả rằng tỷ số tín hiệu trên nhiễu (SNR) tối thiểu đối với không gian phòng họp nên đạt mức 20 dB. Do đó, nếu một người nói ở mức âm lượng 60 dB(A), thì lý tưởng là phòng nên có mức tiếng ồn nên không vượt quá 40 dB(A).</i>” Điều này cho thấy, để đạt $\text{SNR} \geq 20 \text{ dB}$ với giọng nói ở 60 dBA, mức tiếng ồn nên phải $\leq 40 \text{ dBA}$ - khớp hoàn toàn với giới hạn bạn đề xuất. - ITU-T P.800 – Căn cứ cho mức ồn môi trường $\leq 40 \text{ dB(A)}$ (<i>Không vượt quá 40 dB(A) để đảm bảo đánh giá chủ quan chính xác</i>). + Trích dẫn tiêu chuẩn Trong Phụ lục A (A.1.1.2.1) của ITU-T P.800 (Methods for subjective determination of
--	--	--	---

			transmission quality), nội dung quy định về độ ồn nền như sau: <i>"Mức độ ồn môi trường (trong điều kiện không cố ý đưa thêm tiếng ồn từ môi trường bên ngoài) cần được duy trì ở mức thấp nhất có thể. Vì lý do thực tiễn, giới hạn trên được khuyến nghị là NC-25 hoặc NR-25. Các giá trị này tương đương với mức tiếng ồn điển hình trong các không gian như phòng ngủ tại gia đình, bệnh viện hoặc thư viện."</i> *NC-25/NR-25 tương đương ≤ 40 dB(A), theo bảng tương ứng trong ISO 1996. *NC-25 là giới hạn âm thanh nền được chấp nhận trong phòng trị liệu, bệnh viện, thư viện... - RT60 môi trường thu âm giọng nói: Không vượt quá 0,5 giây tại dải tần 500–1000 Hz để đảm bảo môi trường “thấp âm vang”, bảo toàn độ rõ thoại, tránh méo đặc trưng giọng nói.
2	Thời gian âm vang	≤ 0.5 giây	1. TCVN 10615-2:2014 (ISO 3382-2:2008) - Thời gian âm vang của phòng làm việc/ghi âm nhỏ. <i>"Âm học – Đo các thông số âm thanh phòng – Phần 2: Thời gian âm vang trong phòng thông thường"</i>, mặc dù tiêu chuẩn không quy định mức RT60 tối đa cứng, nhưng mô tả thời gian âm vang điển hình trong phòng làm việc và phòng thu nhỏ dao động từ 0,2 đến 0,6 giây ở dải tần 500–1000 Hz

			2. ISO 3382-1, ISO 3382-2 – Tiêu chuẩn RT60 tối đa khuyến nghị. -Thông số $RT60 = 0,2-0,6$ s (500–1000 Hz) được ngầm đề xuất hoặc minh họa trong các phần sau của ISO 3382-2:2008: Giới thiệu: Trình bày mục đích đo RT60 trong các phòng bình thường và lý do đo ở dải 500–1000 Hz. Nơi thường xuất hiện thông tin định hướng như 0,2–0,6 s. Khoản 1 - Phạm vi: Mô tả phạm vi áp dụng và đặc trưng âm học cần đánh giá. Khoản 5 - Quy trình đo: Đề cập cách đo RT60, xác định dải tần sử dụng (bao gồm 500 Hz ÷ 1000 Hz). Phụ lục A - Độ không chắc chắn của phép đo: Có thể nêu ảnh hưởng của thời gian âm vang đến độ chính xác kết quả. 3. ITU-T P.800 - Phòng thử nghiệm thoại yêu cầu $RT60 \leq 0,5$ s để giữ chất lượng đánh giá.
--	--	--	--

7. Hiệu quả dự kiến của việc áp dụng quy chuẩn

Quy chuẩn kỹ thuật quốc gia sinh trắc học giọng nói sau khi được ban hành sẽ trở thành căn cứ để Cục Cảnh sát quản lý hành chính về trật tự xã hội; Cục Pháp chế và cải cách hành chính, tư pháp; Công an các đơn vị địa phương; ... triển khai thực hiện Luật Căn cước.

Chuẩn hóa toàn quốc hoạt động thu thập, xử lý và trao đổi dữ liệu giọng nói phục vụ định danh cá nhân.

Tăng cường tính bảo mật, phòng chống giả mạo, nâng cao hiệu quả các hệ thống xác thực sinh trắc học.

Góp phần phát triển các ứng dụng chính phủ điện tử, định danh điện tử, giám sát an ninh, chăm sóc sức khỏe và dịch vụ tài chính.

Đảm bảo tương thích với xu hướng quốc tế, hỗ trợ chuyển đổi số quốc gia.

8. Các ý kiến tham gia và tiếp thu, giải trình

8.1. Giải trình tiếp thu ý kiến bộ, ngành, địa phương, doanh nghiệp

Bổ sung sau khi lấy ý kiến các cơ quan liên quan.

8.2. Giải trình tiếp thu ý kiến thẩm định

Bổ sung sau khi có ý kiến thẩm định.