

HƯỚNG DẪN CHUẨN BỊ DỮ LIỆU

Tóm lược: Xử lý tín hiệu số là một trong các khâu rất quan trọng trong nhiều hệ thống phân tích, khai phá dữ liệu cũng như ứng dụng dữ liệu trong cuộc sống, chẳng hạn nhận dạng tiếng nói trong giao tiếp người máy, các bài toán thị giác máy tính. Tiếng nói nói riêng, âm thanh (audio) nói chung có thể dễ dàng cảm nhận bởi con người trở thành một nguyên liệu quan trọng cho việc minh họa các hiệu ứng, kết quả của các quá trình thực hiện xử lý tín hiệu số. Mục tiêu của bài tập phần này là chuẩn bị dữ liệu để làm dữ liệu đầu vào cho các bài tập trong phần tiếp theo trong môn học Xử lý tín hiệu số.

1. Mục tiêu và yêu cầu khi chuẩn bị dữ liệu

Mục tiêu:

- Sử dụng Audacity thực hiện phân đoạn tín hiệu tiếng nói thành các phân đoạn tương ứng với câu nói đơn, trọn nghĩa, chỉ do một người nói, và không bị can nhiễu (không lẫn nhạc nền), ... gọi chung là phân đoạn câu “sạch”
- Nghe và chuyển soạn phân đoạn câu “sạch” thành một văn bản (dạng text) kèm theo (gọi là quá trình chuyển văn bản)

Yêu cầu với dữ liệu kết quả:

- Dữ liệu âm thanh kết quả của việc phân đoạn chỉ ứng với **một câu đơn, trọn nghĩa, chỉ do một người nói và không bị can nhiễu, không ở trạng thái khó nghe** (do nhỏ quá, do méo, ...)
- Từ mỗi file dữ liệu gốc được giao, **tên các phân đoạn kết quả sẽ là tên file dữ liệu gốc và được tự động thêm vào thứ tự của các phân đoạn được lấy từ 001 đến hết**
Ví dụ: File âm thanh được giao có tên A120BRS20220911.mp3; giả sử file âm thanh này khi nghe thấy có chứa tiếng nói tương ứng của 25 câu nói; trong đó có 5 câu nói bị lẫn nhạc nền. Khi đó chỉ có thể lấy được 20 phân đoạn tương ứng với 20 câu nói “sạch”. Trong trường hợp này, các phân đoạn với các câu nói lấy được sẽ lần lượt là A120BRS20220911-001.wav đến A120BRS20220911-020.wav

Chú ý: Nếu sử dụng phần mềm Audacity và thực hiện chọn phân đoạn trước khi cắt tất cả trước khi xuất file (export) thì phần thứ tự segment sẽ được thêm tự động.

- **Mỗi phân đoạn câu nói sạch sẽ phải có một file văn bản ghi lời nói (gọi là file transcription) đi kèm. File transcription có cùng tên với file phân đoạn tiếng nói tương ứng nhưng có phần mở rộng là txt.** File transcription là file phiên âm, hay còn gọi là chuyển văn bản của tiếng nói từ file phân đoạn âm thanh tương ứng. File transcription phải được tạo bằng các trình soạn thảo đơn giản như Notepad, Notepad++, gõ tiếng Việt unicode và lưu ở dạng mã hóa UTF-8.

Chú ý 1: Quá trình chuyển tiếng nói thành văn bản có thể sử dụng phần mềm hỗ trợ được cung cấp. Việc tương tác với phần mềm hỗ trợ sẽ tiện lợi hơn: sẽ không phải

bạn tâm việc định dạng và tên file lưu trữ. Chỉ bạn nào sử dụng trực tiếp Notepad, Notepad++ thì cần để ý điểm này.

Chú ý 2: Việc chuyển văn bản chúng ta có thể tận dụng một số API để chuyển văn bản tự động. Tuy nhiên, cho đến thời điểm này không có nền tảng nào cung cấp API hỗ trợ miễn phí mà có độ chính xác 100%. Do đó, nếu chúng ta sử dụng các API tự động, yêu cầu bắt buộc là phải kiểm tra, soát lỗi trước khi lưu trữ. Độ chính xác của việc chuyển ngữ là một trong các tiêu chí quan trọng đánh giá mức độ hoàn thành nhiệm vụ.

Chú ý 3: Nếu sử dụng phần mềm hỗ trợ được cung cấp, chúng ta có thể sử dụng chức năng chuyển văn bản với Google Speech API đã được tích hợp để chuyển văn bản. Tuy nhiên, độ chính xác không cao nên cũng cần bắt buộc kiểm tra, soát lỗi trước khi lưu trữ

- Trong quá trình chuyển văn bản, phần đầu dòng văn bản tương ứng với lời câu nói cần thêm nội dung thông tin về: thể loại của phân đoạn, giới tính người nói, vùng miền của giọng nói, độ tuổi của người nói, cảm xúc của người nói, đặt trong cặp dấu ngoặc vuông []. Cụ thể có dạng [xyzmn], trong đó:
 - x: thông tin về thể loại của phân đoạn (theo bảng trong phần dưới); được xác định dựa trên phân đoạn lớn hơn mà phân đoạn câu nói được trích ra (phân đoạn lớn hơn hoặc file gốc mà câu nói được trích ra).
 - y: thông tin về giới tính người nói
 - z: thông tin vùng của giọng nói
 - m: thông tin cảm xúc của người nói
 - n: thông tin khoảng độ tuổi người nói

Chú ý: Một số thông tin (ví dụ tuổi) có thể mang tính tương đối dựa trên đánh giá chủ quan của mình khi nghe.

- Từ file dữ liệu gốc, không lấy các phân đoạn ứng với các câu là phần giới thiệu chương trình, câu chào hỏi đầu và cuối các đoạn chương trình. Ví dụ, không lấy các phân đoạn kiểu “Xin kính chào quý vị thính giả”, hoặc “Tôi là ABC, xin chào các bạn” lặp đi lặp lại trong các file dữ liệu gốc.
- ***Tổng thời lượng các phân đoạn câu nói “sạch” có được phải tối thiểu là 330 câu nói (330 phân đoạn câu sạch)***

Khái niệm câu đơn, trọn nghĩa, chỉ gồm một người nói:

- Câu đơn. Là câu đơn trong tiếng Việt hoặc câu ghép chính phụ; câu ghép đẳng lập. Với các câu ghép, tối đa gồm 02 câu thành phần. Trong trường hợp câu ghép dài hơn thì cắt tối đa đến 12s sao cho tương đối trọn nghĩa nhất.
- Trọn nghĩa: Trong trường hợp câu ghép, phần ngắt câu có nghĩa trọn vẹn
- Chỉ gồm một người nói: phân đoạn chỉ với giọng nói của một người

Một số yêu cầu đặc biệt trong phiên âm:

- Trong phân đoạn câu, nếu là số thì sẽ phiên âm dạng đọc chữ số. Ví dụ nghe thấy “Địa chỉ số bốn trăm bảy hai” sẽ phiên âm đúng là những gì đọc “Địa chỉ số bốn trăm bảy hai”
- Trong phân đoạn câu, nếu là ngày tháng năm thì phiên âm ngày tháng năm dạng đọc chữ số. Ví dụ nghe thấy “ngày mười hai tháng chín năm hai nghìn không trăm hai hai” thì phiên thành “ngày mười hai tháng chín năm hai nghìn không trăm hai hai”; nếu nghe thấy “ngày mười chín tháng hai” thì phiên âm thành “ngày mười chín tháng hai”
- Các tên riêng tiếng nước ngoài phiên âm các âm tiết và nối với nhau bằng dấu -. Các âm tiết phiên âm theo chuẩn (nên dùng internet tra trên các trang chính thống như VOV, báo nhân dân, ...)
- Các cụm từ phổ biến chỉ các cơ quan tổ chức trong nước thì cần phiên âm đầy đủ theo cách đọc. Ví dụ nghe thấy “Ủy ban nhân dân thành phố Hà Nội” thì ghi rõ là “Ủy ban nhân dân thành phố Hà Nội”, tuyệt đối không viết tắt thành UBND TP.
- Với các cụm viết tắt khác cho các cụm tiếng Việt ghi theo cách đọc. Nếu nghe thấy “Theo nghị quyết số ba tư en-nơ qui xê pê ngày ba tháng hai năm” thì ghi là “Theo nghị quyết số ba tư n q c p ngày ba tháng hai năm”
- Các cụm viết tắt các tổ chức quốc tế mà được đọc tắt thì ghi tắt, nếu đọc đầy đủ thì ghi đầy đủ theo cách đọc. Ví dụ nghe thấy “Tổ chức vê-kép-tê-ô vừa thông báo” thì ghi là “Tổ chức WTO vừa thông báo”; nếu nghe thấy “Tổ chức y tế thế giới” thì ghi rõ là “Tổ chức y tế thế giới”, không ghi “Tổ chức WHO” ; nhưng nếu đọc là “Tổ chức vê-kép-hát-ô” thì ghi là “Tổ chức WHO”
- Với các tên riêng tiếng Việt thì ghi đúng chuẩn tiếng Việt. Tên người sẽ viết hoa chữ cái đầu trong họ, tên đệm, tên (ví dụ ông Nguyễn Thành Vinh). Tên thành phố viết hoa chữ cái đầu trong các từ của tên (ví dụ Hà Nội).
- Nếu câu nói là câu kể có lời trích của nhân vật (trong các phần kể chuyện), không cần thêm các dấu trích (dấu nháy kép –“). Ví dụ nghe thấy “Anh Tú nói ‘Ngày mai tôi đi’” thì ghi “Anh Tú nói ngày mai tôi đi”

Bảng tiêu chí đánh giá kết quả:

Chú ý: Các khía cạnh dưới đây được coi là quan trọng trong việc đánh giá chất lượng và mức độ hoàn thành được giao. Bất cứ tiêu chí nào trong các tiêu chí chính được liệt kê dưới đây không đạt trên 97% sẽ ảnh hưởng đến đánh giá toàn bộ của kết quả

Tiêu chí	Mô tả	Baseline (lower bound)
Độ chính xác trong việc phân đoạn	Các phân đoạn đúng với yêu cầu là các câu đơn giản, trọn nghĩa, chỉ là giọng	98%

	một người nói, không bị nhiễu hoặc ở trạng thái khó nghe	
Độ chính xác trong việc chuyển văn bản	Các phân đoạn âm thanh câu đơn cắt được đều có phần chuyển văn bản (transcription). Văn bản tương ứng chính xác với tiếng nói	98%
Tổng thời lượng phân đoạn câu “sạch”	Tính tổng thời lượng phân đoạn câu sạch trên tổng số dữ liệu được giao	80%
Đúng định dạng yêu cầu	Các file phân đoạn câu “sạch” kết quả, các file *.json đánh dấu điểm đầu và cuối các phân đoạn; các file văn bản; tổ chức thư mục, ... theo đúng yêu cầu	100%
Thời gian hoàn thành	Với sự hỗ trợ của phần mềm cung cấp, thời gian hoàn thành công việc buồn tẻ sẽ giảm một nửa, do đó thời gian hoàn thành sẽ là 02 tuần kể từ ngày được giao.	100%
Sự hợp tác	Trong trường hợp các tiêu chuẩn trên chưa đạt và được yêu cầu bổ sung, chỉnh sửa, có thái độ hợp tác, tích cực để hoàn thành chất lượng yêu cầu	100%

Bảng quy ước thông tin phân đoạn:

Các thông tin về phân đoạn câu sạch là các thông tin đánh giá chủ quan của mình

- Thông tin phân đoạn thể loại đoạn chứa phân đoạn câu

0	Tin tức: từ đoạn tin tức
1	Phim, kịch: từ các đoạn phim, kịch
2	Truyện: từ các bài đọc truyện, kể chuyện, sách nói
3	Văn hóa: từ các bài về văn hóa
4	Du lịch: từ các bài về du lịch
5	Đời sống: những phần bình của các đoạn về cuộc sống, ẩm thực
6	Thể thao: Tin thể thao, bình luận thể thao
7	An ninh cuộc sống: trộm, cướp, ..
8	Giao thông: tin tức, an toàn giao thông
9	Sức khỏe: tư vấn sức khỏe, ...

- Thông tin giới tính người nói

M	giọng nam
F	giọng nữ

- Thông tin vùng của giọng nói

0	giọng miền Bắc (từ Thanh Hóa trở ra)
1	giọng miền Trung (từ Nghệ An – Bình Thuận)
2	giọng Tây nguyên (các tỉnh Gia Lai, Đăk-lăk, ...)
3	giọng miền Nam (các tỉnh miền Nam từ Bình Phước trở xuống)
4	giọng người dân tộc thiểu số

- Thông tin cảm xúc của người nói

N	Cảm xúc thông thường
J	Cảm xúc vui vẻ
A	Cảm xúc tức giận
S	Cảm xúc buồn bã
T	Cảm xúc mệt mỏi
D	Cảm xúc ghê
F	Cảm xúc sợ hãi
H	Cảm xúc hạnh phúc

- Thông tin về độ tuổi của người nói

C	giọng người nói là trẻ em, thiếu niên: dưới 16
Y	giọng thanh niên: từ 16-35
M	giọng trung niên: từ 35 - 55
O	giọng người già: từ 55 - 70
S	giọng người rất già: trên 70

2. Cách thức tiến hành

2.1. Thực hiện phân đoạn (cắt) câu nói “sạch”

Để hoàn thành nhiệm vụ này, chúng ta có thể có rất nhiều cách tiến hành. Trong phần này chúng ta giới thiệu cách tiếp cận sử dụng Audacity.

2.1.1. Thiết lập công cụ

Download phần mềm Audacity tại:

<https://www.audacityteam.org/download/windows/>

Tùy thuộc hệ điều hành, trình duyệt và trang download sẽ tự động đề xuất phiên bản thích hợp. Trong hướng dẫn này, giả sử hệ điều hành là Windows 64-bit. File download bản mới nhất (tính tại thời điểm soạn bản hướng dẫn này) là audacity-win-3.1.3-64bit.exe.

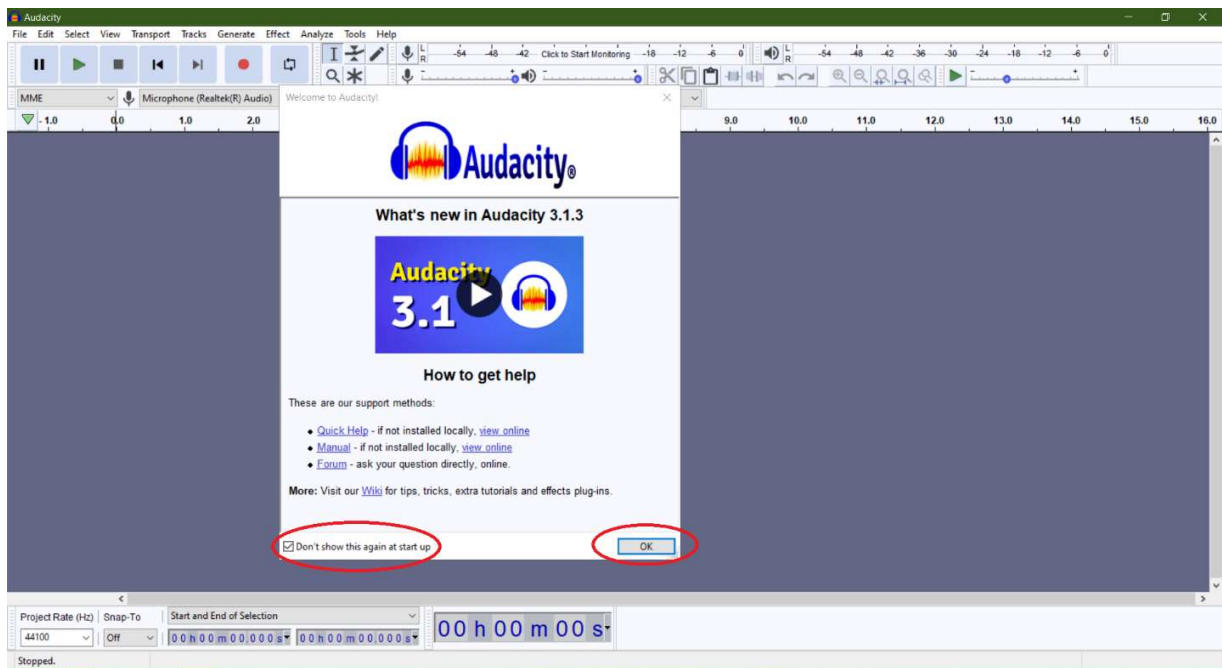
Sau khi download, nhấp đúp file audacity-win-3.1.3-64bit.exe để thực hiện cài đặt. Quá trình cài đặt đơn giản, chấp nhận các đề xuất thiết lập mặc định, chọn Next/OK cho đến khi hoàn tất.

Sau khi cài đặt thành công, trên Desktop sẽ xuất hiện biểu tượng của Audacity.



Hình 1: Icon sau khi cài đặt thành công Audacity

Mở ứng dụng Audacity lên, ta sẽ có giao diện như sau. Chọn “Don’t show this again at start up” sau đó nhấn “OK”.

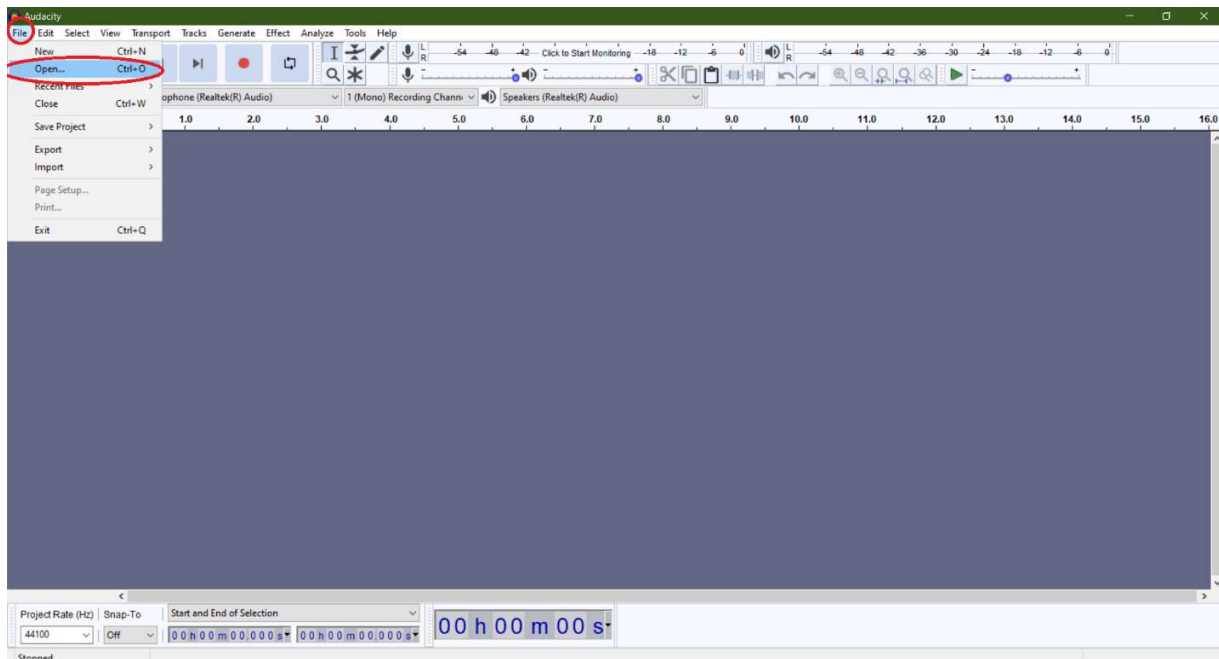


Hình 2: Minh họa giao diện Audacity khi mở ứng dụng

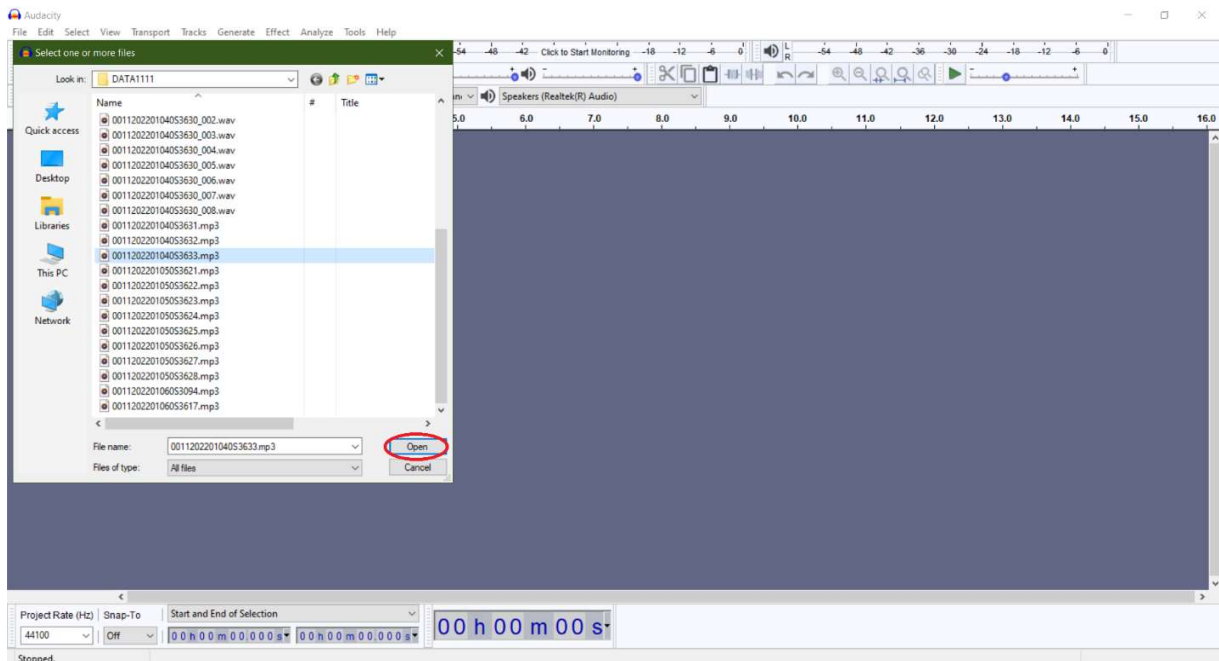
2.1.2. Cách thức thực hiện phân đoạn

a. Mở file âm thanh cần cắt

Từ menu của Audacity chọn: File -> Open (hoặc sử dụng tổ hợp phím tắt Ctrl+O). Chọn file cần mở và ấn Open.



Hình 3: Minh họa việc mở file



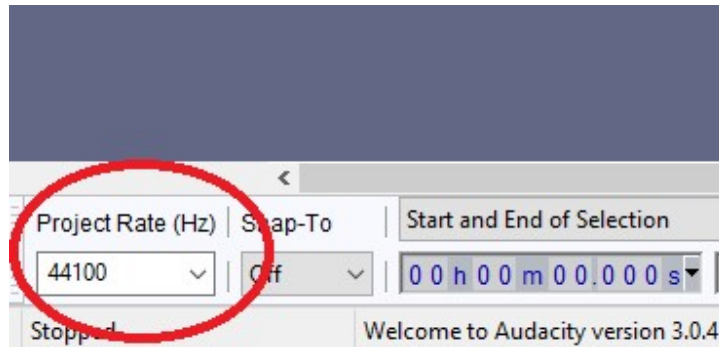
Thử mở một file âm thanh bất kỳ, bấm icon menu biểu tượng play để nghe thử và điều chỉnh âm lượng cho phù hợp. Điều này để đảm bảo điều kiện tốt cho việc nghe âm thanh và làm việc trong phần sau không có vấn đề gì trở ngại.

b. Thiết lập tham số làm việc

Tham số cho file dữ liệu kết quả phải có:

- Tần số lấy mẫu: 16kHz

Thông tin này sẽ được thiết lập cho project, nếu không chỉ cần đảm bảo thông số trước khi export ra file WAV. Thực hiện thiết lập theo cách sau:



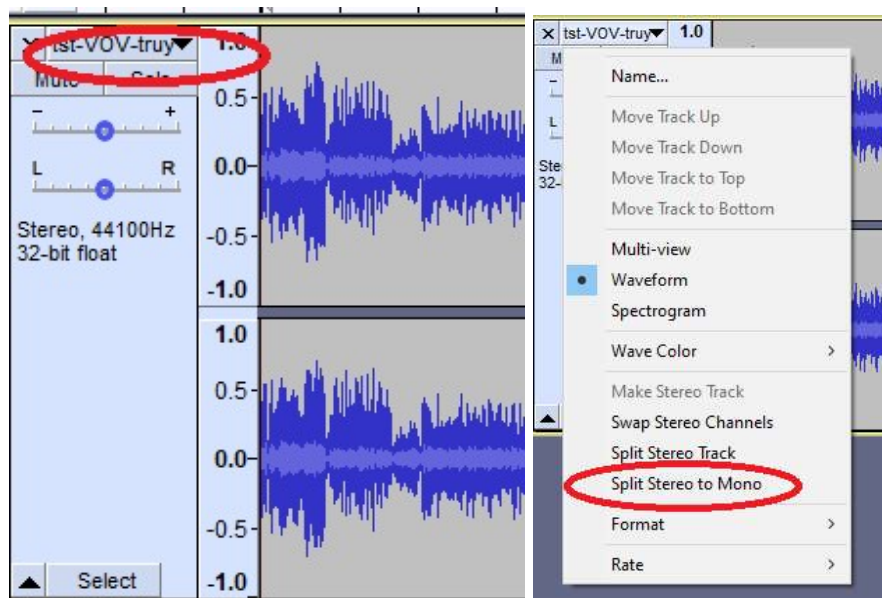
Hình 4: Cách thiết lập tần số lấy mẫu

- Số bit trên một mẫu: 16 bit

Xem trong phần export file WAV

- Số kênh: 1 (mono)

Để thiết lập âm thanh mono, thực hiện chọn dấu tam giác trên phần thiết lập track, chọn split Stereo to Mono



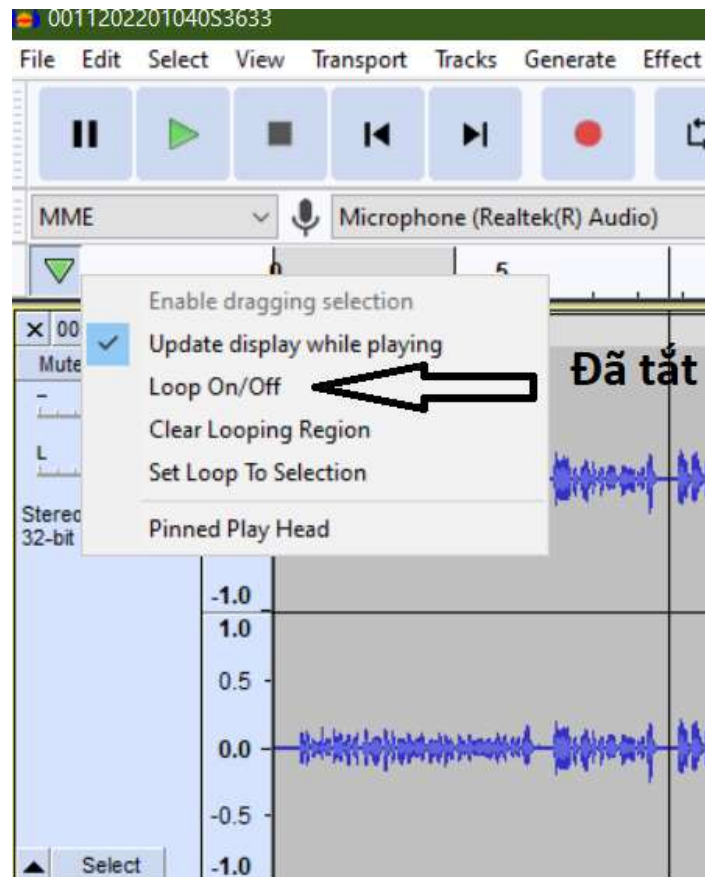
Hình 5: Minh họa tách chọn chuyển thành mono

Khi đó cửa sổ dạng sóng sẽ chia làm 2, chọn một trong 2, xóa kênh còn lại bằng cách bấm vào dấu x



Hình 6: Minh họa xóa một trong hai kênh để lại tín hiệu mono

Để tránh khi nghe kiểm tra một phân đoạn chọn bị loop, đảm bảo rằng tính năng Loop on/off ở chế độ tắt (không có dấu tích bên cạnh). Nếu chưa thì thiết lập như sau:



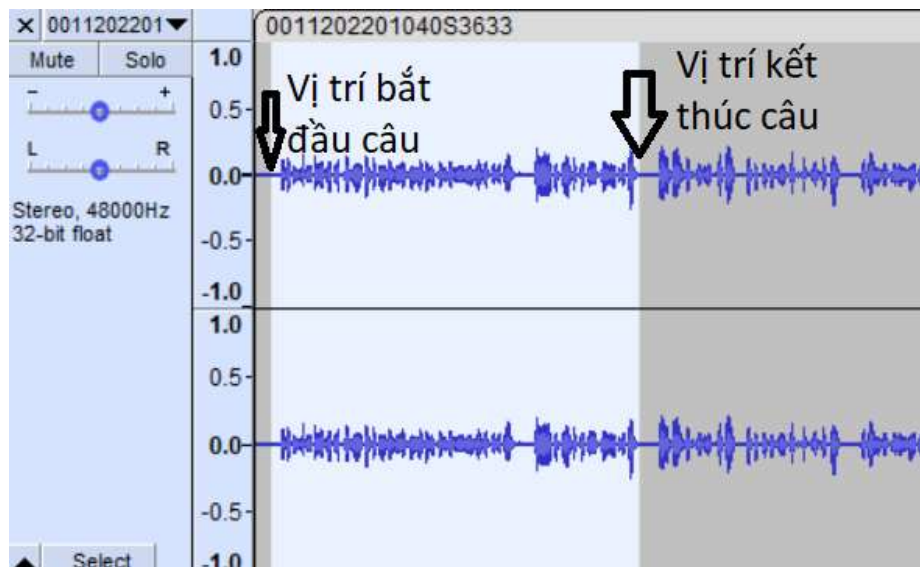
Hình 7: Tắt chế độ Loop

c. Thực hiện nghe và cắt

Sau khi mở file audio, sử dụng chuột di chuyển về điểm đầu tiên bên trái của đồ thị dạng sóng tín hiệu, bấm chuột phải để đặt con trỏ điểm bắt đầu nghe tại đầu file

Sử dụng chuột bấm nút play (icon menu) để nghe hoặc dừng nghe. Thao tác này có thể thực hiện bằng phím tắt là phím Space (phím dấu cách).

- Chọn một phân đoạn: (1) Sử dụng phím space để nghe/dừng nghe, khi biết điểm bắt đầu của một phân đoạn, sử dụng chuột di chuyển đến điểm tương ứng và bấm chuột trái; (2) Tiếp tục nghe, vừa nghe vừa quan sát sự dịch chuyển con trỏ (vạch) trên cửa sổ dạng sóng; (3) Khi nghe thấy kết thúc phân đoạn câu nói “sạch”, sử dụng space để dừng nghe; (4) Sử dụng chuột trái di chuyển về vạch đánh dấu điểm đầu, lúc này con trỏ chuột sẽ chuyển thành dạng bàn tay cầm bút, nhấn và giữ chuột trái, kéo đến điểm kết thúc, rồi bỏ chuột ra.



Hình 8: Minh họa phân đoạn được chọn

- Để đảm bảo tính chính xác của phân đoạn (không bị cắt, không lẫn phần tạp âm, ...), có thể sử dụng nhấn icon play để nghe lại phân đoạn đã chọn. Trong trường hợp một trong hai đầu bị thiếu hoặc thừa, có thể sử dụng chuột chọn đầu tương ứng (di chuyển con trỏ chuột về điểm đầu tương ứng, con trỏ chuột trở thành biểu tượng bàn tay), nhấn giữ chuột trái rồi kéo vào hoặc ra (tùy muốn bỏ hay thêm vào phân đoạn)
- Sau khi đã chắc chắn đoạn chọn, sử dụng tổ hợp Ctrl+B để đánh dấu đoạn chọn. Kết quả sẽ được hiển thị ngay phía dưới.



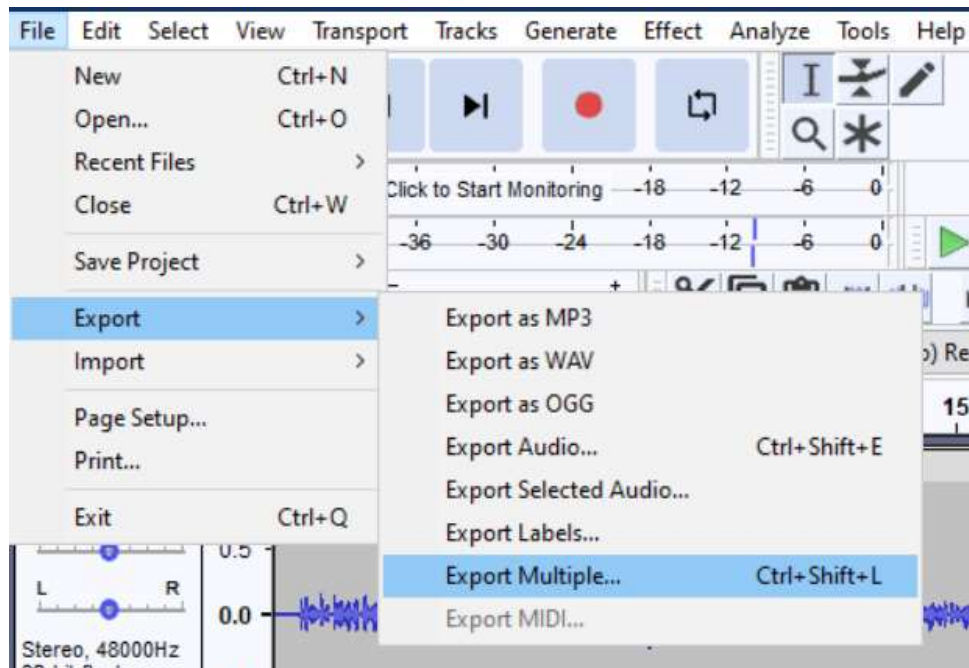
Hình 9: Minh họa kết quả phân đoạn cắt chọn sau khi bấm Ctrl+B

Để tiếp tục với phân đoạn tiếp theo. Bấm phím chuột trái tại vị trí tiếp theo và lặp lại các thao tác trên cho các phân đoạn còn lại của file dữ liệu gốc.

Trong quá trình thao tác, để thuận tiện quan sát dạng sóng, đảm bảo cho việc phân đoạn chính xác thì nên kết hợp các thao tác phóng to, thu nhỏ hiển thị dạng sóng (Ctrl+M hoặc Ctrl + Con lăn giữa của chuột); cũng như di chuyển trên cửa sổ dạng sóng (Shift+ Con lăn giữa của chuột).

Sau khi đã đánh dấu cắt xong tất cả các phân đoạn câu nói sạch. Tiến hành việc xuất ra các phân đoạn.

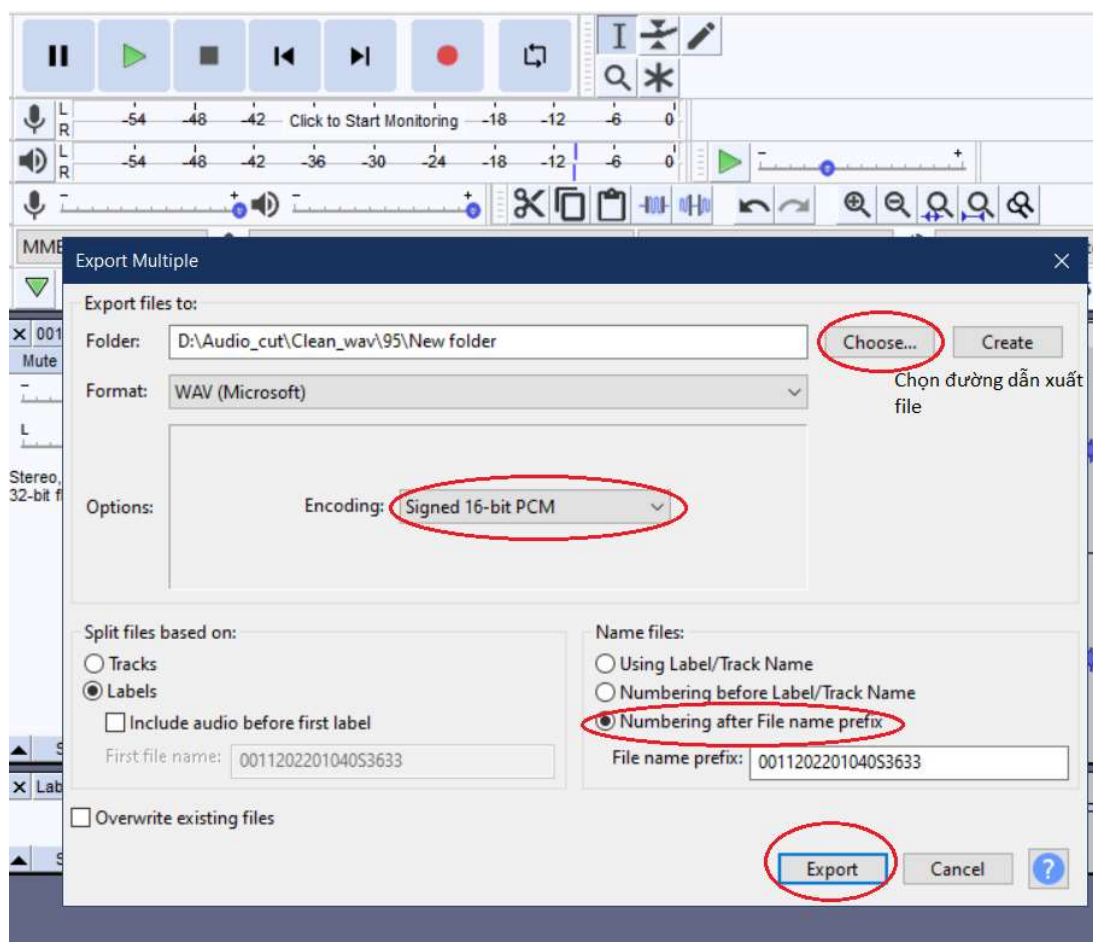
- Từ menu, chọn File -> Export -> Export Multiple ...



Hình 10: Thiết lập chuẩn bị export

- Trên cửa sổ thông tin xuất file, đảm bảo rằng:

- Chọn thư mục để lưu trữ các file đã trích xuất – nên tạo một thư mục mới trong thư mục làm việc và đặt tên thư mục trùng với tên file dữ liệu gốc
- Format: WAV (Microsoft)
- Encoding chọn Signed 16-bit PCM
- Name files chọn Numbering after Filename prefix
- Đảm bảo phần File name prefix trùng với tên file dữ liệu gốc.



Hình 11: Thiết lập các tham số cho trích xuất

Cuối cùng nhấn Export để thực hiện việc trích xuất.

Lặp lại các thao tác nghe và cắt trên cho tất cả các file dữ liệu được giao sao cho có được 330 phân đoạn câu nói sạch.

Nếu thiếu data để phân đoạn, cần phản hồi trong thread (sẽ được lập) trên nhóm lớp facebook để được bổ sung kịp thời.

2.2. Thực hiện chuyển văn bản cho phân đoạn câu nói sạch

2.2.1. Cách 01 - Phương thức sử dụng phần mềm hỗ trợ

2.2.1.1. Thiết lập công cụ

Để sử dụng bộ phần mềm hỗ trợ, chúng ta cần có Chương trình play đa phương tiện VLC (VLC Media Player) và phần mềm hỗ trợ.

a. Download và cài đặt Chương trình play đa phương tiện VLC

Tùy theo hệ điều hành trên máy tính chọn tương ứng bản VLC 64-bit hoặc 32-bit

Link:

- 64-bit: <https://get.videolan.org/vlc/3.0.17.4/win64/vlc-3.0.17.4-win64.exe>
- 32-bit: <https://get.videolan.org/vlc/3.0.17.4/win32/vlc-3.0.17.4-win32.exe>

Sau khi download, thực hiện chạy file vlc-3.0.17.4-winxx.exe để cài đặt.

b. Download và giải nén phần mềm hỗ trợ

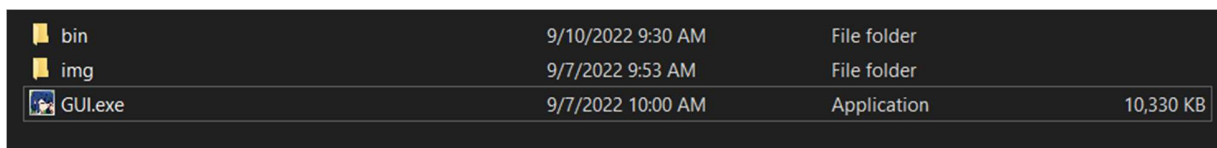
- Download bộ phần mềm hỗ trợ tại:

https://drive.google.com/drive/folders/1KtkOgZFv1c2_EX7gCqTRMhIoIDWpnNba?usp=sharing

Thực hiện giải nén vào thư mục mong muốn.

Nên để trong thư mục sao cho trên đường dẫn từ ổ gốc không có dấu cách, không có tên tiếng Việt. Ví dụ nên để ở thư mục DSP trong ổ đĩa D (đường dẫn tương ứng sẽ là D:/DSP)

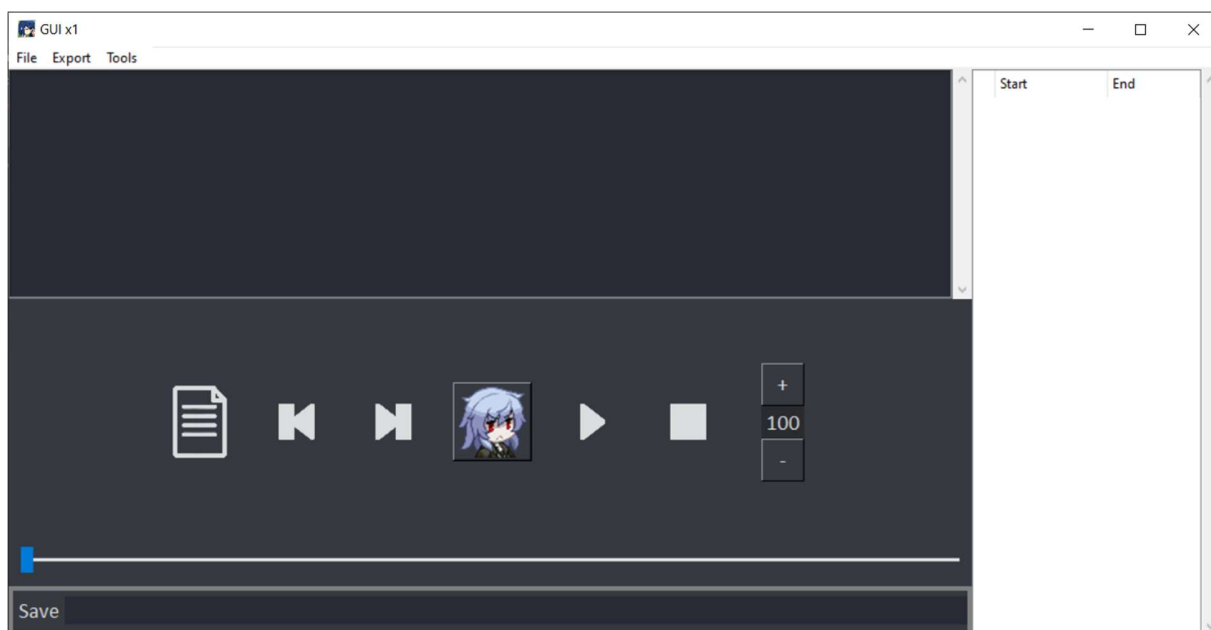
Kết quả thư mục giải nén sẽ có dạng như sau:



bin	9/10/2022 9:30 AM	File folder	
img	9/7/2022 9:53 AM	File folder	
GUI.exe	9/7/2022 10:00 AM	Application	10,330 KB

Hình 12: Kết quả giải nén phần mềm hỗ trợ

Thực hiện chạy thử chương trình bằng cách nhấn đúp vào GUI.exe; giao diện chương trình sẽ có dạng như sau:

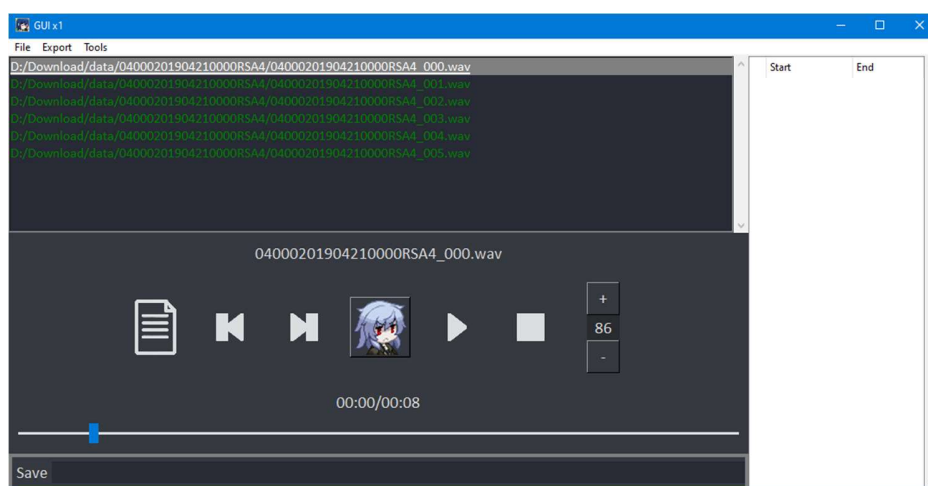


Hình 13: Giao diện chính của phần mềm hỗ trợ

2.2.1.2. Sử dụng công cụ cho phần việc chuyển văn bản

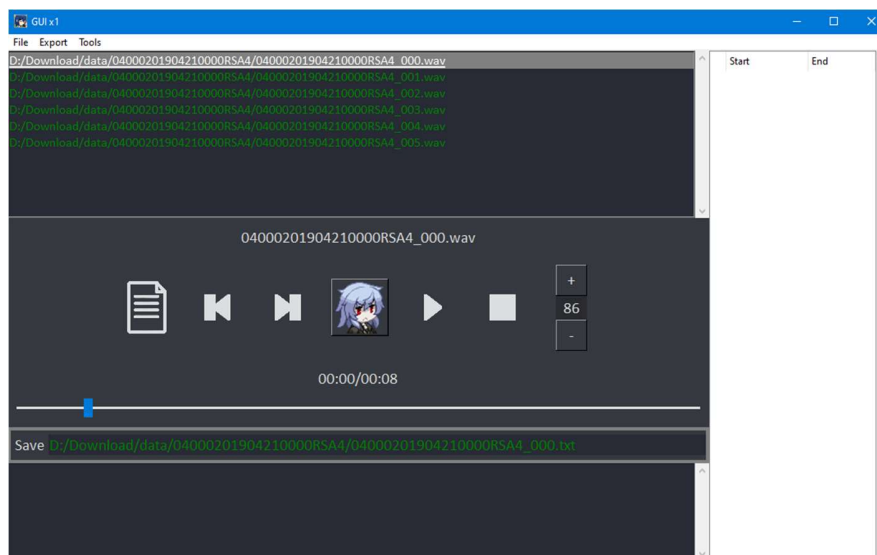
Giả sử việc phân đoạn (cắt) thành các câu nói “sạch” cho một file dữ liệu gốc hoặc cho tất cả các file dữ liệu gốc đã hoàn thành, bây giờ chúng ta tiếp tục thực hiện việc chuyển văn bản cho các phân đoạn.

Thêm các phân đoạn vào để làm việc với phần mềm hỗ trợ bằng cách từ menu, chọn File – Add folder (hoặc sử dụng tổ hợp phím tắt Ctrl+M). Sau khi chọn thư mục chứa các phân đoạn câu cần chuyển văn bản, tất cả các phân đoạn trong thư mục sẽ được thêm vào danh sách làm việc.



Hình 14: Thêm danh sách các phân đoạn từ một thư mục để thực hiện phiên âm

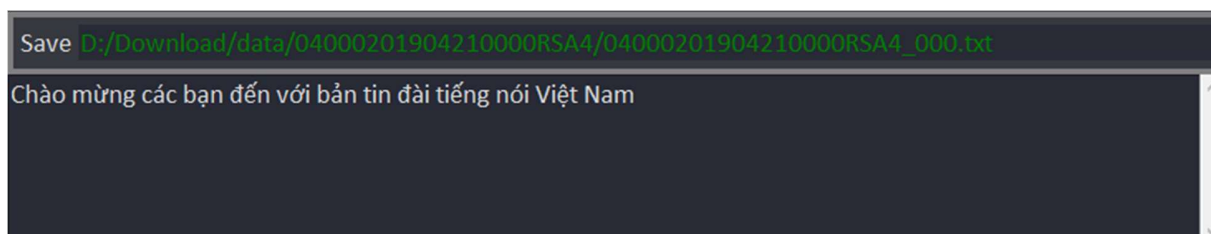
Tiếp tục từ menu, chọn Tools->Transcript để mở cửa sổ Transcript (phần cửa sổ nằm phía dưới cùng của giao diện)



Hình 15: Giao diện cửa sổ để thêm transcript

Chọn file phân đoạn cần chuyển văn bản bằng cách bấm đúp chuột vào tên file mong muốn trên danh sách file.

Thực hiện nghe và gõ trực tiếp viết phiên âm của file âm thanh vào cửa sổ Transcript (phần nằm dưới cùng của giao diện).



Hình 16: Giao diện phần nhập văn bản tương ứng với tiếng nói của phân đoạn

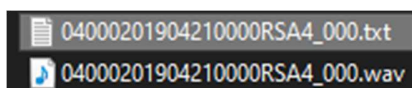
Trong quá trình nghe và gõ lại văn bản lời nói, chúng ta có thể thực hiện định vị vị trí điểm bắt đầu nghe hoặc nghe lại bằng cách điều khiển điểm con trỏ trên thanh trượt của player.

Trước khi lưu và chuyển sang phân đoạn câu tiếp theo, cần xác định nhanh các thông tin cơ bản của phân đoạn như: thể loại của phân đoạn, giới tính người nói, vùng miền của giọng nói, độ tuổi của người nói, cảm xúc của người nói, đặt trong cặp dấu ngoặc vuông theo cú pháp [xyzmn] (với x,y,z,m,n theo hướng dẫn ở trên). Các thông tin này là các đánh giá chủ quan của mình. Một số thông tin như độ tuổi chỉ mang tính tương đối dựa trên giọng. Sau đó gõ phần văn bản tương ứng với câu nói.

[0F0NY] Bà Hường ngồi trước sân chờ cho Hoa đi khuất hẳn

Hình 17: Minh họa một văn bản đầy đủ cùng thông tin đặc điểm phân đoạn

Sau khi xong, từ menu chọn File->Save (hoặc sử dụng tổ hợp phím tắt Ctrl+S; hoặc ấn nút Save trên màn hình) để lưu lại phiên âm của file âm thanh tương ứng. Phần mềm sẽ tự động lưu tên file phiên âm trùng với tên file âm thanh tương ứng ở định dạng đuôi .txt.



Hình 18: Minh họa kết quả sau khi hoàn thành chuyển văn bản cho phân đoạn câu sạch

Lặp lại thao tác trên với toàn bộ file phân đoạn âm thanh muốn phiên âm.

Chú ý:

Quá trình thực hiện thêm phiên âm có thể thực hiện tự động một phần **nếu có Internet**. Tuy nhiên **ĐỘ CHÍNH XÁC THẤP** nên cần chỉnh sửa sau khi chuyển đổi tự động.

Thực hiện các thao tác sau để chuyển văn bản tự động:

- Sau khi chọn file, từ menu, chọn Tools->Convert audio to text (hoặc sử dụng tổ hợp Ctrl+H) để chuyển phiên âm tự động.
- Sau khi chuyển tự động, cần nghe và kiểm tra, sử dụng chuột và bàn phím sửa lại phần phiên âm cũng như thêm thông tin đặc điểm phân đoạn (phần [xyzmn]) trước khi thực hiện lưu file phiên âm.

3. Cách thức nộp file dữ liệu

Tạo một thư mục có tên **Gxxmsv** trên máy cá nhân của mình

Trong đó xx sẽ là 04, 07, 08, hoặc 80; msv sẽ là mã sinh viên của mình (bao gồm cả phần chữ và phần số, phần chữ viết hoa).

Ví dụ sinh viên nhóm 04 có mã sinh viên là B21DCCN123 thì thư mục sẽ là **G04B21DCCN123**

Trong thư mục **Gxxmsv** sẽ gồm các thư mục con chứa các phân đoạn cắt được ở phần 2.1. và các file bản bản kèm theo. Mỗi thư mục con là sản phẩm các phân đoạn câu nói sạch và văn bản tương ứng của chúng ứng với một file dữ liệu gốc được giao. Tên thư mục trùng với file dữ liệu gốc được giao.

Ví dụ, sinh viên nhóm 01 có mã B21DCCN123 được giao 02 file dữ liệu có tên là MXN.mp3 và KBMD.mp3 thì tổ chức thư mục sẽ là

C01B21DCCN123

```
| --- MXN
|     |--- MXN-001.wav
|     |--- MXN-001.text
|     |--- ...
| --- KBMD
```

Khi nộp bài, mỗi sinh viên sẽ có một thư mục nộp bài riêng có tên thư mục trùng với mã sinh viên (thư mục này sẽ được thông báo sau). Thực hiện upload lần lượt từng thư mục kết quả lên thư mục drive **msv** (sẽ được thông báo sau).

Khi Upload, chọn upload folder và chọn thư mục kết quả thứ nhất. Lặp lại cho các thư mục kết quả còn lại.

Chú ý 1: KHÔNG upload file dữ liệu gốc - file dữ liệu được giao.

Chú ý 2: Tùy tình trạng của mạng, nếu mạng chậm và kém ổn định thì nên upload lần lượt từng thư mục kết quả. Tránh tình trạng upload đồng thời và tất cả các thư mục đều bị crashed

HƯỚNG DẪN CHUẨN BỊ DỮ LIỆU

Tóm lược: Xử lý tín hiệu số là một trong các khâu rất quan trọng trong nhiều hệ thống phân tích, khai phá dữ liệu cũng như ứng dụng dữ liệu trong cuộc sống, chẳng hạn nhận dạng tiếng nói trong giao tiếp người máy, các bài toán thị giác máy tính. Tiếng nói nói riêng, âm thanh (audio) nói chung có thể dễ dàng cảm nhận bởi con người trở thành một nguyên liệu quan trọng cho việc minh họa các hiệu ứng, kết quả của các quá trình thực hiện xử lý tín hiệu số. Mục tiêu của bài tập phần này là chuẩn bị dữ liệu để làm dữ liệu đầu vào cho các bài tập trong phần tiếp theo trong môn học Xử lý tín hiệu số.

1. Mục tiêu và yêu cầu khi chuẩn bị dữ liệu

Mục tiêu:

- Sử dụng Audacity thực hiện phân đoạn tín hiệu tiếng nói thành các phân đoạn tương ứng với câu nói đơn, trọn nghĩa, chỉ do một người nói, và không bị can nhiễu (không lẫn nhạc nền), ... gọi chung là phân đoạn câu “sạch”
- Nghe và chuyển soạn phân đoạn câu “sạch” thành một văn bản (dạng text) kèm theo (gọi là quá trình chuyển văn bản)

Yêu cầu với dữ liệu kết quả:

- Dữ liệu âm thanh kết quả của việc phân đoạn chỉ ứng với **một câu đơn, trọn nghĩa, chỉ do một người nói và không bị can nhiễu, không ở trạng thái khó nghe** (do nhỏ quá, do méo, ...)
- Từ mỗi file dữ liệu gốc được giao, **tên các phân đoạn kết quả sẽ là tên file dữ liệu gốc và được tự động thêm vào thứ tự của các phân đoạn được lấy từ 001 đến hết**
Ví dụ: File âm thanh được giao có tên A120BRS20220911.mp3; giả sử file âm thanh này khi nghe thấy có chứa tiếng nói tương ứng của 25 câu nói; trong đó có 5 câu nói bị lẫn nhạc nền. Khi đó chỉ có thể lấy được 20 phân đoạn tương ứng với 20 câu nói “sạch”. Trong trường hợp này, các phân đoạn với các câu nói lấy được sẽ lần lượt là A120BRS20220911-001.wav đến A120BRS20220911-020.wav

Chú ý: Nếu sử dụng phần mềm Audacity và thực hiện chọn phân đoạn trước khi cắt tất cả trước khi xuất file (export) thì phần thứ tự segment sẽ được thêm tự động.

- **Mỗi phân đoạn câu nói sạch sẽ phải có một file văn bản ghi lời nói (gọi là file transcription) đi kèm. File transcription có cùng tên với file phân đoạn tiếng nói tương ứng nhưng có phần mở rộng là txt.** File transcription là file phiên âm, hay còn gọi là chuyển văn bản của tiếng nói từ file phân đoạn âm thanh tương ứng. File transcription phải được tạo bằng các trình soạn thảo đơn giản như Notepad, Notepad++, gõ tiếng Việt unicode và lưu ở dạng mã hóa UTF-8.

Chú ý 1: Quá trình chuyển tiếng nói thành văn bản có thể sử dụng phần mềm hỗ trợ được cung cấp. Việc tương tác với phần mềm hỗ trợ sẽ tiện lợi hơn: sẽ không phải

bạn tâm việc định dạng và tên file lưu trữ. Chỉ bạn nào sử dụng trực tiếp Notepad, Notepad++ thì cần để ý điểm này.

Chú ý 2: Việc chuyển văn bản chúng ta có thể tận dụng một số API để chuyển văn bản tự động. Tuy nhiên, cho đến thời điểm này không có nền tảng nào cung cấp API hỗ trợ miễn phí mà có độ chính xác 100%. Do đó, nếu chúng ta sử dụng các API tự động, yêu cầu bắt buộc là phải kiểm tra, soát lỗi trước khi lưu trữ. Độ chính xác của việc chuyển ngữ là một trong các tiêu chí quan trọng đánh giá mức độ hoàn thành nhiệm vụ.

Chú ý 3: Nếu sử dụng phần mềm hỗ trợ được cung cấp, chúng ta có thể sử dụng chức năng chuyển văn bản với Google Speech API đã được tích hợp để chuyển văn bản. Tuy nhiên, độ chính xác không cao nên cũng cần bắt buộc kiểm tra, soát lỗi trước khi lưu trữ

- Trong quá trình chuyển văn bản, phần đầu dòng văn bản tương ứng với lời câu nói cần thêm nội dung thông tin về: thể loại của phân đoạn, giới tính người nói, vùng miền của giọng nói, độ tuổi của người nói, cảm xúc của người nói, đặt trong cặp dấu ngoặc vuông []. Cụ thể có dạng [xyzmn], trong đó:
 - x: thông tin về thể loại của phân đoạn (theo bảng trong phần dưới); được xác định dựa trên phân đoạn lớn hơn mà phân đoạn câu nói được trích ra (phân đoạn lớn hơn hoặc file gốc mà câu nói được trích ra).
 - y: thông tin về giới tính người nói
 - z: thông tin vùng của giọng nói
 - m: thông tin cảm xúc của người nói
 - n: thông tin khoảng độ tuổi người nói

Chú ý: Một số thông tin (ví dụ tuổi) có thể mang tính tương đối dựa trên đánh giá chủ quan của mình khi nghe.

- Từ file dữ liệu gốc, không lấy các phân đoạn ứng với các câu là phần giới thiệu chương trình, câu chào hỏi đầu và cuối các đoạn chương trình. Ví dụ, không lấy các phân đoạn kiểu “Xin kính chào quý vị thính giả”, hoặc “Tôi là ABC, xin chào các bạn” lặp đi lặp lại trong các file dữ liệu gốc.
- ***Tổng thời lượng các phân đoạn câu nói “sạch” có được phải tối thiểu là 330 câu nói (330 phân đoạn câu sạch)***

Khái niệm câu đơn, trọn nghĩa, chỉ gồm một người nói:

- Câu đơn. Là câu đơn trong tiếng Việt hoặc câu ghép chính phụ; câu ghép đẳng lập. Với các câu ghép, tối đa gồm 02 câu thành phần. Trong trường hợp câu ghép dài hơn thì cắt tối đa đến 12s sao cho tương đối trọn nghĩa nhất.
- Trọn nghĩa: Trong trường hợp câu ghép, phần ngắt câu có nghĩa trọn vẹn
- Chỉ gồm một người nói: phân đoạn chỉ với giọng nói của một người

Một số yêu cầu đặc biệt trong phiên âm:

- Trong phân đoạn câu, nếu là số thì sẽ phiên âm dạng đọc chữ số. Ví dụ nghe thấy “Địa chỉ số bốn trăm bảy hai” sẽ phiên âm đúng là những gì đọc “Địa chỉ số bốn trăm bảy hai”
- Trong phân đoạn câu, nếu là ngày tháng năm thì phiên âm ngày tháng năm dạng đọc chữ số. Ví dụ nghe thấy “ngày mười hai tháng chín năm hai nghìn không trăm hai hai” thì phiên thành “ngày mười hai tháng chín năm hai nghìn không trăm hai hai”; nếu nghe thấy “ngày mười chín tháng hai” thì phiên âm thành “ngày mười chín tháng hai”
- Các tên riêng tiếng nước ngoài phiên âm các âm tiết và nối với nhau bằng dấu -. Các âm tiết phiên âm theo chuẩn (nên dùng internet tra trên các trang chính thống như VOV, báo nhân dân, ...)
- Các cụm từ phổ biến chỉ các cơ quan tổ chức trong nước thì cần phiên âm đầy đủ theo cách đọc. Ví dụ nghe thấy “Ủy ban nhân dân thành phố Hà Nội” thì ghi rõ là “Ủy ban nhân dân thành phố Hà Nội”, tuyệt đối không viết tắt thành UBND TP.
- Với các cụm viết tắt khác cho các cụm tiếng Việt ghi theo cách đọc. Nếu nghe thấy “Theo nghị quyết số ba tư en-nơ qui xê pê ngày ba tháng hai năm” thì ghi là “Theo nghị quyết số ba tư n q c p ngày ba tháng hai năm”
- Các cụm viết tắt các tổ chức quốc tế mà được đọc tắt thì ghi tắt, nếu đọc đầy đủ thì ghi đầy đủ theo cách đọc. Ví dụ nghe thấy “Tổ chức vê-kép-tê-ô vừa thông báo” thì ghi là “Tổ chức WTO vừa thông báo”; nếu nghe thấy “Tổ chức y tế thế giới” thì ghi rõ là “Tổ chức y tế thế giới”, không ghi “Tổ chức WHO” ; nhưng nếu đọc là “Tổ chức vê-kép-hát-ô” thì ghi là “Tổ chức WHO”
- Với các tên riêng tiếng Việt thì ghi đúng chuẩn tiếng Việt. Tên người sẽ viết hoa chữ cái đầu trong họ, tên đệm, tên (ví dụ ông Nguyễn Thành Vinh). Tên thành phố viết hoa chữ cái đầu trong các từ của tên (ví dụ Hà Nội).
- Nếu câu nói là câu kể có lời trích của nhân vật (trong các phần kể chuyện), không cần thêm các dấu trích (dấu nháy kép –“). Ví dụ nghe thấy “Anh Tú nói ‘Ngày mai tôi đi’” thì ghi “Anh Tú nói ngày mai tôi đi”

Bảng tiêu chí đánh giá kết quả:

Chú ý: Các khía cạnh dưới đây được coi là quan trọng trong việc đánh giá chất lượng và mức độ hoàn thành được giao. Bất cứ tiêu chí nào trong các tiêu chí chính được liệt kê dưới đây không đạt trên 97% sẽ ảnh hưởng đến đánh giá toàn bộ của kết quả

Tiêu chí	Mô tả	Baseline (lower bound)
Độ chính xác trong việc phân đoạn	Các phân đoạn đúng với yêu cầu là các câu đơn giản, trọn nghĩa, chỉ là giọng	98%

	một người nói, không bị nhiễu hoặc ở trạng thái khó nghe	
Độ chính xác trong việc chuyển văn bản	Các phân đoạn âm thanh câu đơn cắt được đều có phần chuyển văn bản (transcription). Văn bản tương ứng chính xác với tiếng nói	98%
Tổng thời lượng phân đoạn câu “sạch”	Tính tổng thời lượng phân đoạn câu sạch trên tổng số dữ liệu được giao	80%
Đúng định dạng yêu cầu	Các file phân đoạn câu “sạch” kết quả, các file *.json đánh dấu điểm đầu và cuối các phân đoạn; các file văn bản; tổ chức thư mục, ... theo đúng yêu cầu	100%
Thời gian hoàn thành	Với sự hỗ trợ của phần mềm cung cấp, thời gian hoàn thành công việc buồn tẻ sẽ giảm một nửa, do đó thời gian hoàn thành sẽ là 02 tuần kể từ ngày được giao.	100%
Sự hợp tác	Trong trường hợp các tiêu chuẩn trên chưa đạt và được yêu cầu bổ sung, chỉnh sửa, có thái độ hợp tác, tích cực để hoàn thành chất lượng yêu cầu	100%

Bảng quy ước thông tin phân đoạn:

Các thông tin về phân đoạn câu sạch là các thông tin đánh giá chủ quan của mình

- Thông tin phân đoạn thể loại đoạn chứa phân đoạn câu

0	Tin tức: từ đoạn tin tức
1	Phim, kịch: từ các đoạn phim, kịch
2	Truyện: từ các bài đọc truyện, kể chuyện, sách nói
3	Văn hóa: từ các bài về văn hóa
4	Du lịch: từ các bài về du lịch
5	Đời sống: những phân bình của các đoạn về cuộc sống, ẩm thực
6	Thể thao: Tin thể thao, bình luận thể thao
7	An ninh cuộc sống: trộm, cướp, ..
8	Giao thông: tin tức, an toàn giao thông
9	Sức khỏe: tư vấn sức khỏe, ...

- Thông tin giới tính người nói

M	giọng nam
F	giọng nữ

- Thông tin vùng của giọng nói

0	giọng miền Bắc (từ Thanh Hóa trở ra)
1	giọng miền Trung (từ Nghệ An – Bình Thuận)
2	giọng Tây nguyên (các tỉnh Gia Lai, Đăk-lăk, ...)
3	giọng miền Nam (các tỉnh miền Nam từ Bình Phước trở xuống)
4	giọng người dân tộc thiểu số

- Thông tin cảm xúc của người nói

N	Cảm xúc thông thường
J	Cảm xúc vui vẻ
A	Cảm xúc tức giận
S	Cảm xúc buồn bã
T	Cảm xúc mệt mỏi
D	Cảm xúc ghê
F	Cảm xúc sợ hãi
H	Cảm xúc hạnh phúc

- Thông tin về độ tuổi của người nói

C	giọng người nói là trẻ em, thiếu niên: dưới 16
Y	giọng thanh niên: từ 16-35
M	giọng trung niên: từ 35 - 55
O	giọng người già: từ 55 - 70
S	giọng người rất già: trên 70

2. Cách thức tiến hành

2.1. Thực hiện phân đoạn (cắt) câu nói “sạch”

Để hoàn thành nhiệm vụ này, chúng ta có thể có rất nhiều cách tiến hành. Trong phần này chúng ta giới thiệu cách tiếp cận sử dụng Audacity.

2.1.1. Thiết lập công cụ

Download phần mềm Audacity tại:

<https://www.audacityteam.org/download/windows/>

Tùy thuộc hệ điều hành, trình duyệt và trang download sẽ tự động đề xuất phiên bản thích hợp. Trong hướng dẫn này, giả sử hệ điều hành là Windows 64-bit. File download bản mới nhất (tính tại thời điểm soạn bản hướng dẫn này) là audacity-win-3.1.3-64bit.exe.

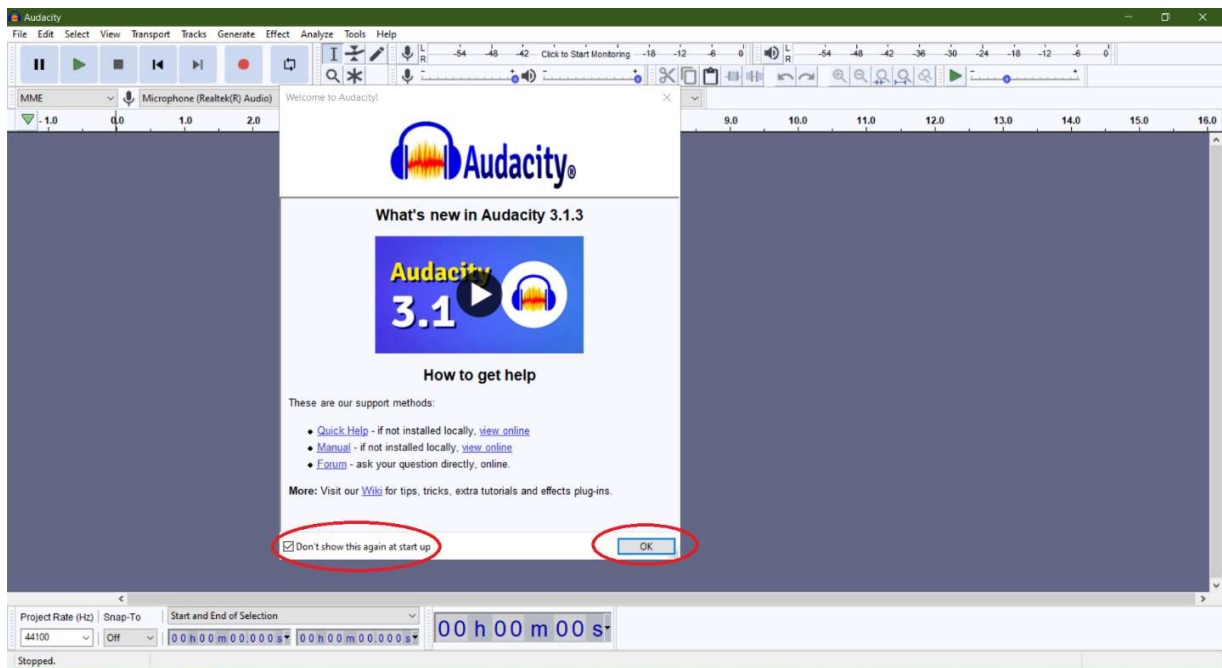
Sau khi download, nhấp đúp file audacity-win-3.1.3-64bit.exe để thực hiện cài đặt. Quá trình cài đặt đơn giản, chấp nhận các đề xuất thiết lập mặc định, chọn Next/OK cho đến khi hoàn tất.

Sau khi cài đặt thành công, trên Desktop sẽ xuất hiện biểu tượng của Audacity.



Hình 1: Icon sau khi cài đặt thành công Audacity

Mở ứng dụng Audacity lên, ta sẽ có giao diện như sau. Chọn “Don’t show this again at start up” sau đó nhấn “OK”.

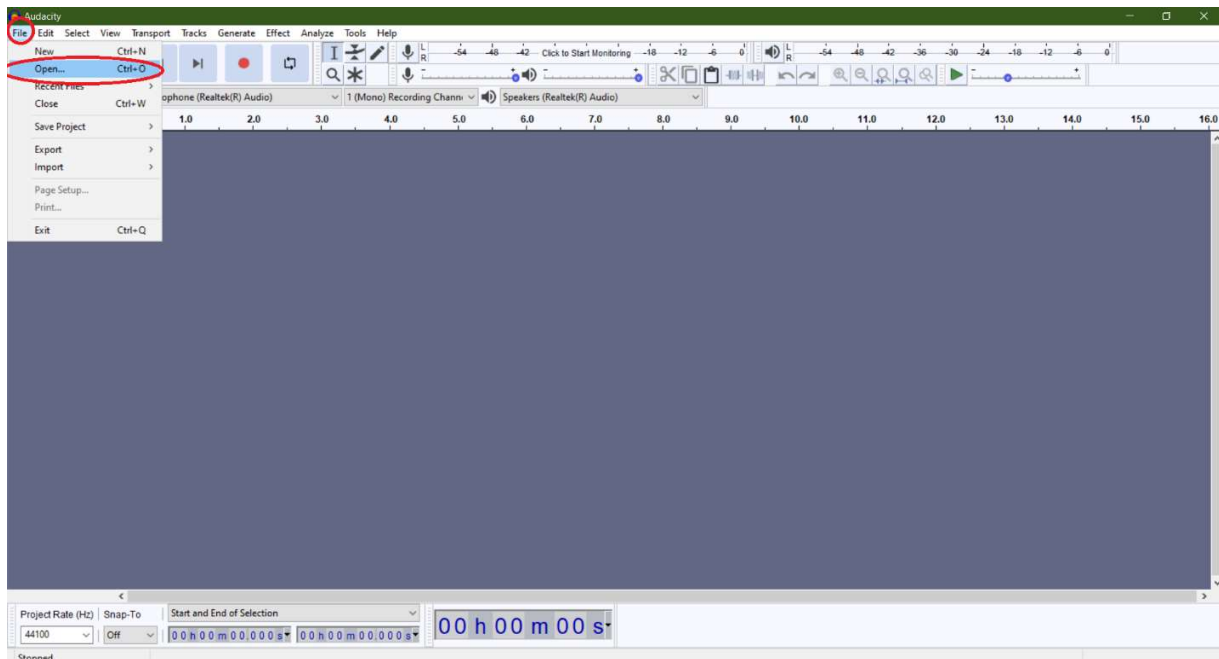


Hình 2: Minh họa giao diện Audacity khi mở ứng dụng

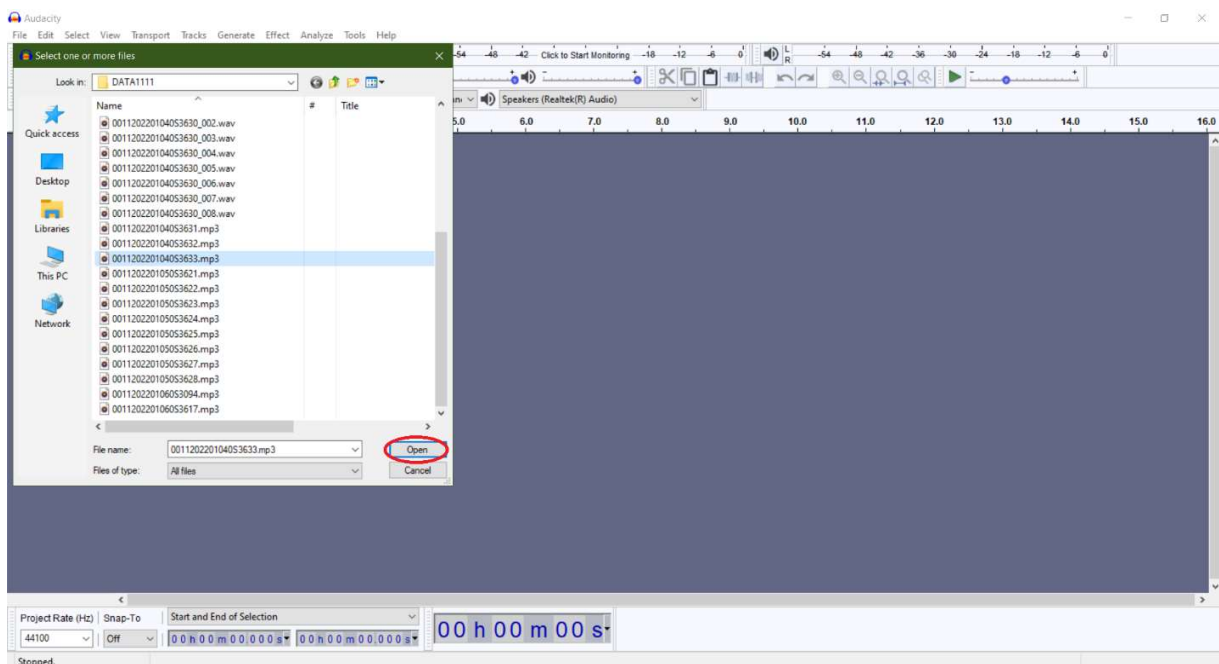
2.1.2. Cách thức thực hiện phân đoạn

a. Mở file âm thanh cần cắt

Từ menu của Audacity chọn: File -> Open (hoặc sử dụng tổ hợp phím tắt Ctrl+O). Chọn file cần mở và ấn Open.



Hình 3: Minh họa việc mở file



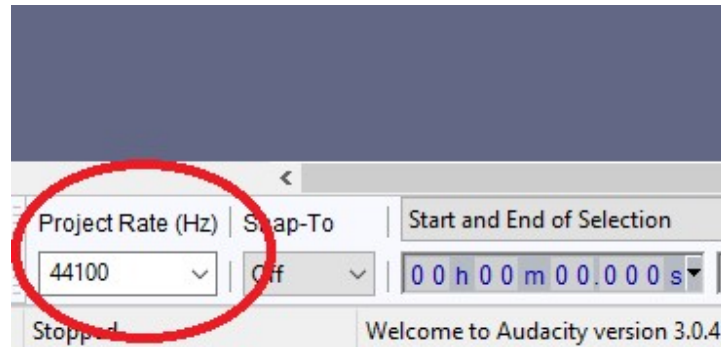
Thử mở một file âm thanh bất kỳ, bấm icon menu biểu tượng play để nghe thử và điều chỉnh âm lượng cho phù hợp. Điều này để đảm bảo điều kiện tốt cho việc nghe âm thanh và làm việc trong phần sau không có vấn đề gì trở ngại.

b. Thiết lập tham số làm việc

Tham số cho file dữ liệu kết quả phải có:

- Tần số lấy mẫu: 16kHz

Thông tin này sẽ được thiết lập cho project, nếu không chỉ cần đảm bảo thông số trước khi export ra file WAV. Thực hiện thiết lập theo cách sau:



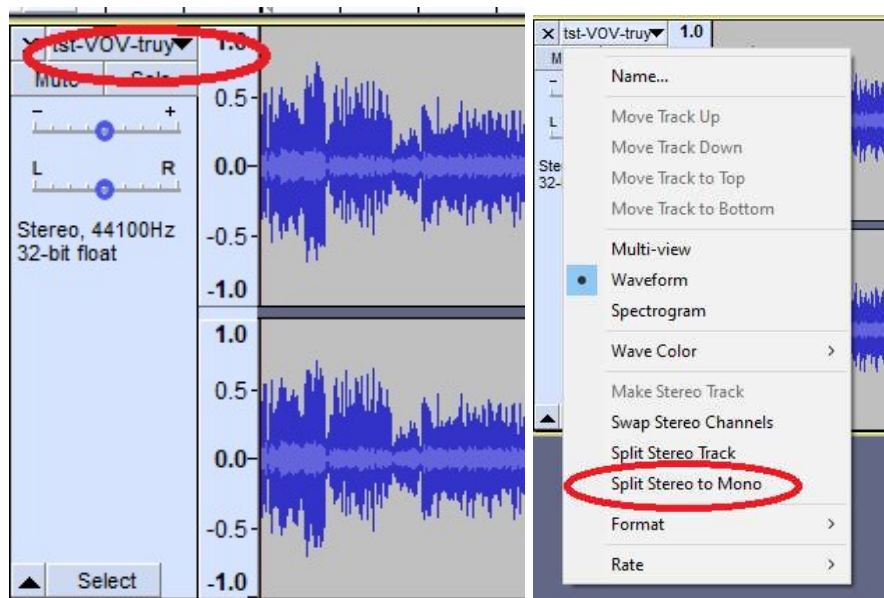
Hình 4: Cách thiết lập tần số lấy mẫu

- Số bit trên một mẫu: 16 bit

Xem trong phần export file WAV

- Số kênh: 1 (mono)

Để thiết lập âm thanh mono, thực hiện chọn dấu tam giác trên phần thiết lập track, chọn split Stereo to Mono



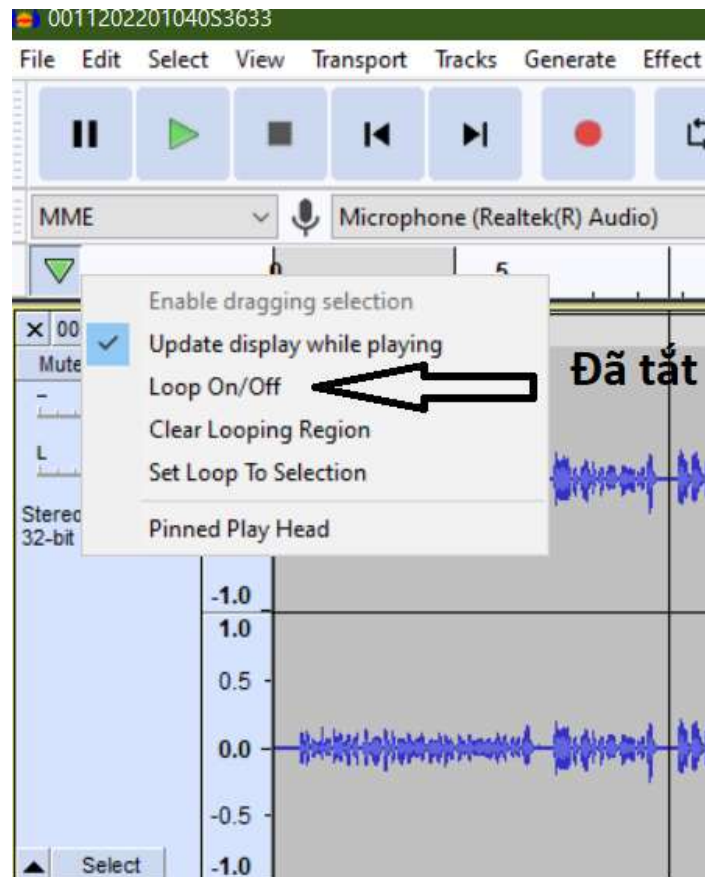
Hình 5: Minh họa tách chọn chuyển thành mono

Khi đó cửa sổ dạng sóng sẽ chia làm 2, chọn một trong 2, xóa kênh còn lại bằng cách bấm vào dấu x



Hình 6: Minh họa xóa một trong hai kênh để lại tín hiệu mono

Để tránh khi nghe kiểm tra một phân đoạn chọn bị loop, đảm bảo rằng tính năng Loop on/off ở chế độ tắt (không có dấu tích bên cạnh). Nếu chưa thì thiết lập như sau:



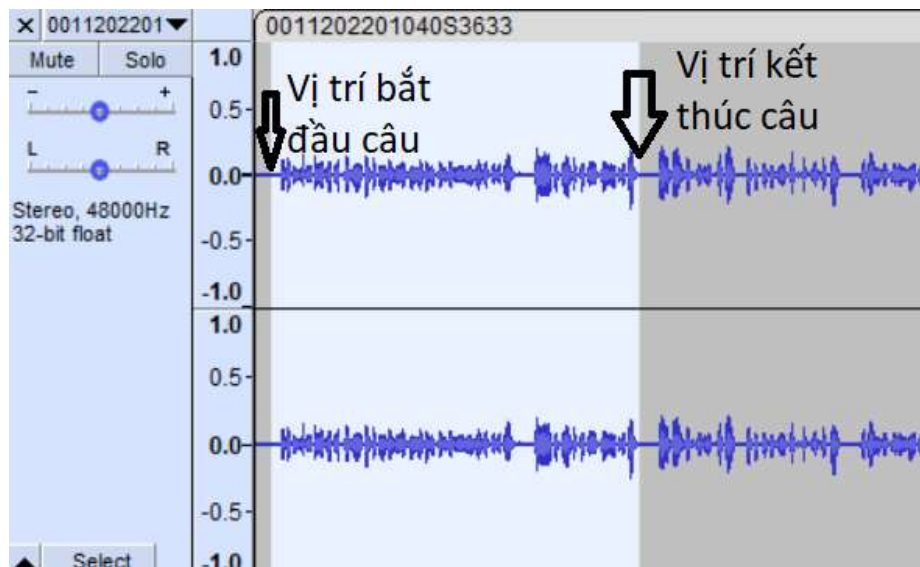
Hình 7: Tắt chế độ Loop

c. Thực hiện nghe và cắt

Sau khi mở file audio, sử dụng chuột di chuyển về điểm đầu tiên bên trái của đồ thị dạng sóng tín hiệu, bấm chuột phải để đặt con trỏ điểm bắt đầu nghe tại đầu file

Sử dụng chuột bấm nút play (icon menu) để nghe hoặc dừng nghe. Thao tác này có thể thực hiện bằng phím tắt là phím Space (phím dấu cách).

- Chọn một phân đoạn: (1) Sử dụng phím space để nghe/dừng nghe, khi biết điểm bắt đầu của một phân đoạn, sử dụng chuột di chuyển đến điểm tương ứng và bấm chuột trái; (2) Tiếp tục nghe, vừa nghe vừa quan sát sự dịch chuyển con trỏ (vạch) trên cửa sổ dạng sóng; (3) Khi nghe thấy kết thúc phân đoạn câu nói “sạch”, sử dụng space để dừng nghe; (4) Sử dụng chuột trái di chuyển về vạch đánh dấu điểm đầu, lúc này con trỏ chuột sẽ chuyển thành dạng bàn tay cầm bút, nhấn và giữ chuột trái, kéo đến điểm kết thúc, rồi bỏ chuột ra.



Hình 8: Minh họa phân đoạn được chọn

- Để đảm bảo tính chính xác của phân đoạn (không bị cắt, không lẫn phần tạp âm, ...), có thể sử dụng nhấn icon play để nghe lại phân đoạn đã chọn. Trong trường hợp một trong hai đầu bị thiếu hoặc thừa, có thể sử dụng chuột chọn đầu tương ứng (di chuyển con trỏ chuột về điểm đầu tương ứng, con trỏ chuột trở thành biểu tượng bàn tay), nhấn giữ chuột trái rồi kéo vào hoặc ra (tùy muốn bỏ hay thêm vào phân đoạn)
- Sau khi đã chắc chắn đoạn chọn, sử dụng tổ hợp Ctrl+B để đánh dấu đoạn chọn. Kết quả sẽ được hiển thị ngay phía dưới.



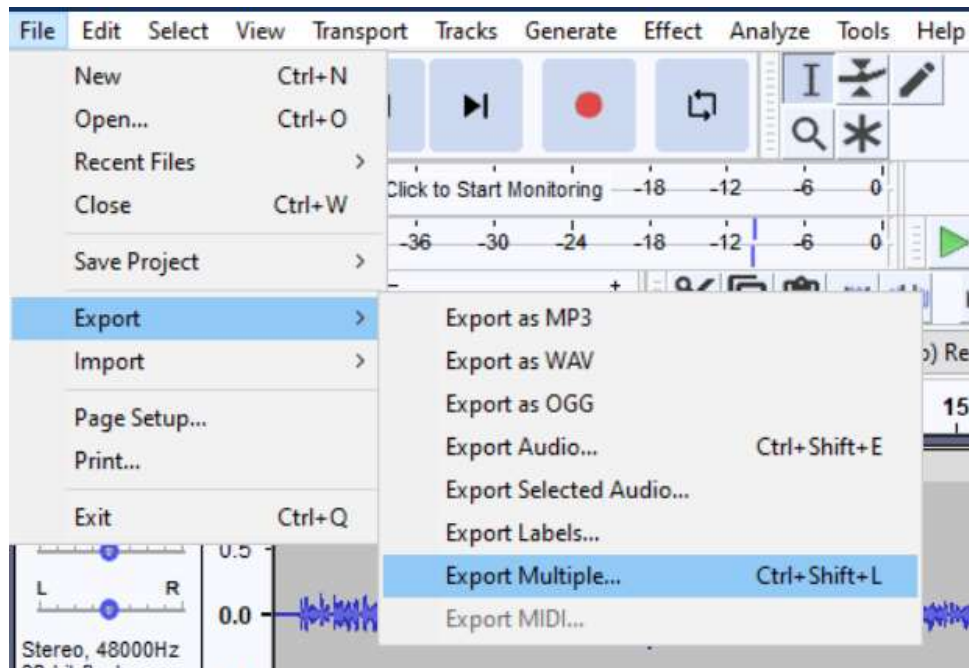
Hình 9: Minh họa kết quả phân đoạn cắt chọn sau khi bấm Ctrl+B

Để tiếp tục với phân đoạn tiếp theo. Bấm phím chuột trái tại vị trí tiếp theo và lặp lại các thao tác trên cho các phân đoạn còn lại của file dữ liệu gốc.

Trong quá trình thao tác, để thuận tiện quan sát dạng sóng, đảm bảo cho việc phân đoạn chính xác thì nên kết hợp các thao tác phóng to, thu nhỏ hiển thị dạng sóng (Ctrl+M hoặc Ctrl + Con lăn giữa của chuột); cũng như di chuyển trên cửa sổ dạng sóng (Shift+ Con lăn giữa của chuột).

Sau khi đã đánh dấu cắt xong tất cả các phân đoạn câu nói sạch. Tiến hành việc xuất ra các phân đoạn.

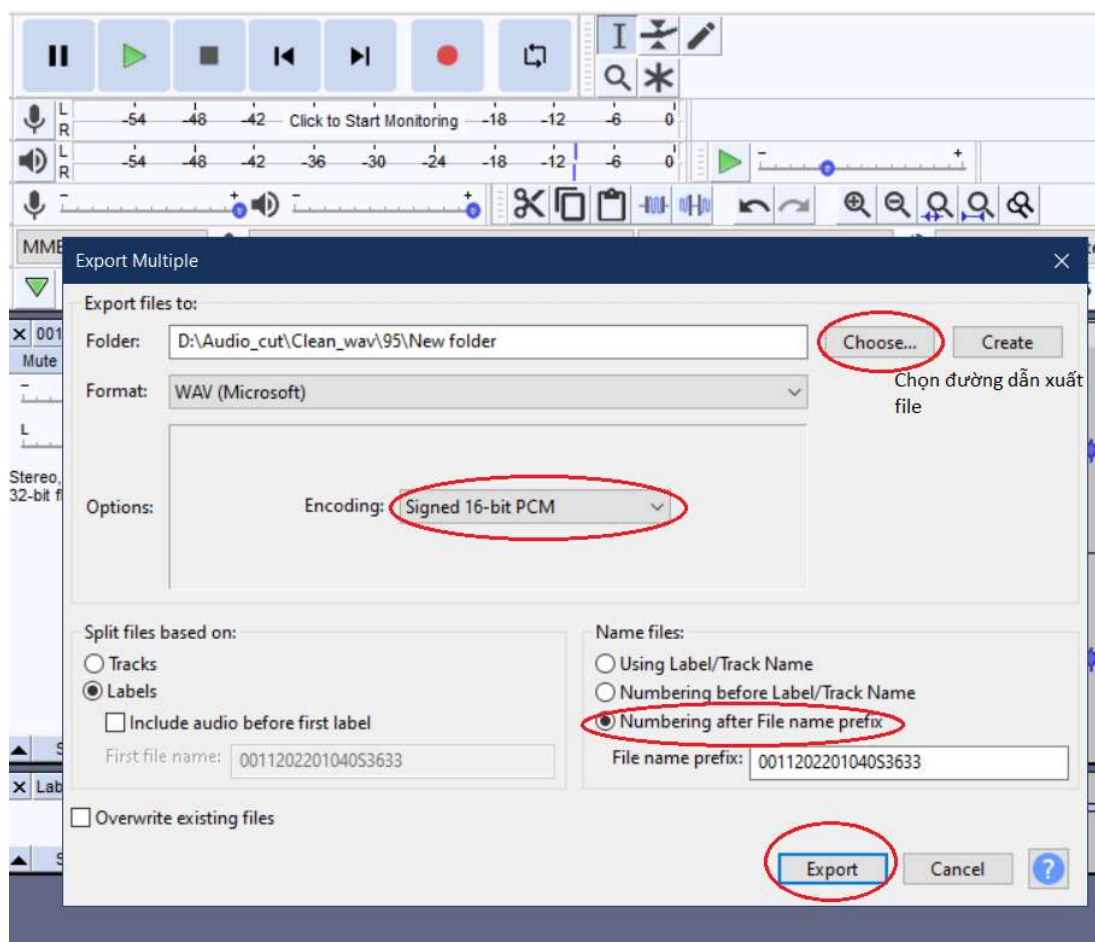
- Từ menu, chọn File -> Export -> Export Multiple ...



Hình 10: Thiết lập chuẩn bị export

- Trên cửa sổ thông tin xuất file, đảm bảo rằng:

- Chọn thư mục để lưu trữ các file đã trích xuất – nên tạo một thư mục mới trong thư mục làm việc và đặt tên thư mục trùng với tên file dữ liệu gốc
- Format: WAV (Microsoft)
- Encoding chọn Signed 16-bit PCM
- Name files chọn Numbering after Filename prefix
- Đảm bảo phần File name prefix trùng với tên file dữ liệu gốc.



Hình 11: Thiết lập các tham số cho trích xuất

Cuối cùng nhấn Export để thực hiện việc trích xuất.

Lặp lại các thao tác nghe và cắt trên cho tất cả các file dữ liệu được giao sao cho có được 330 phân đoạn câu nói sạch.

Nếu thiếu data để phân đoạn, cần phản hồi trong thread (sẽ được lập) trên nhóm lớp facebook để được bổ sung kịp thời.

2.2. Thực hiện chuyển văn bản cho phân đoạn câu nói sạch

2.2.1. Cách 01 - Phương thức sử dụng phần mềm hỗ trợ

2.2.1.1. Thiết lập công cụ

Để sử dụng bộ phần mềm hỗ trợ, chúng ta cần có Chương trình play đa phương tiện VLC (VLC Media Player) và phần mềm hỗ trợ.

a. Download và cài đặt Chương trình play đa phương tiện VLC

Tùy theo hệ điều hành trên máy tính chọn tương ứng bản VLC 64-bit hoặc 32-bit

Link:

- 64-bit: <https://get.videolan.org/vlc/3.0.17.4/win64/vlc-3.0.17.4-win64.exe>
- 32-bit: <https://get.videolan.org/vlc/3.0.17.4/win32/vlc-3.0.17.4-win32.exe>

Sau khi download, thực hiện chạy file vlc-3.0.17.4-winxx.exe để cài đặt.

b. Download và giải nén phần mềm hỗ trợ

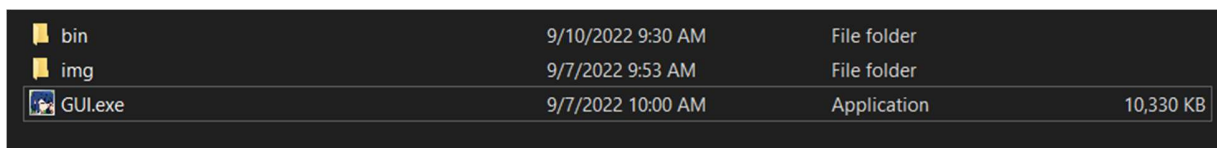
- Download bộ phần mềm hỗ trợ tại:

https://drive.google.com/drive/folders/1KtkOgZFv1c2_EX7gCqTRMhIoIDWpnNba?usp=sharing

Thực hiện giải nén vào thư mục mong muốn.

Nên để trong thư mục sao cho trên đường dẫn từ ổ gốc không có dấu cách, không có tên tiếng Việt. Ví dụ nên để ở thư mục DSP trong ổ đĩa D (đường dẫn tương ứng sẽ là D:/DSP)

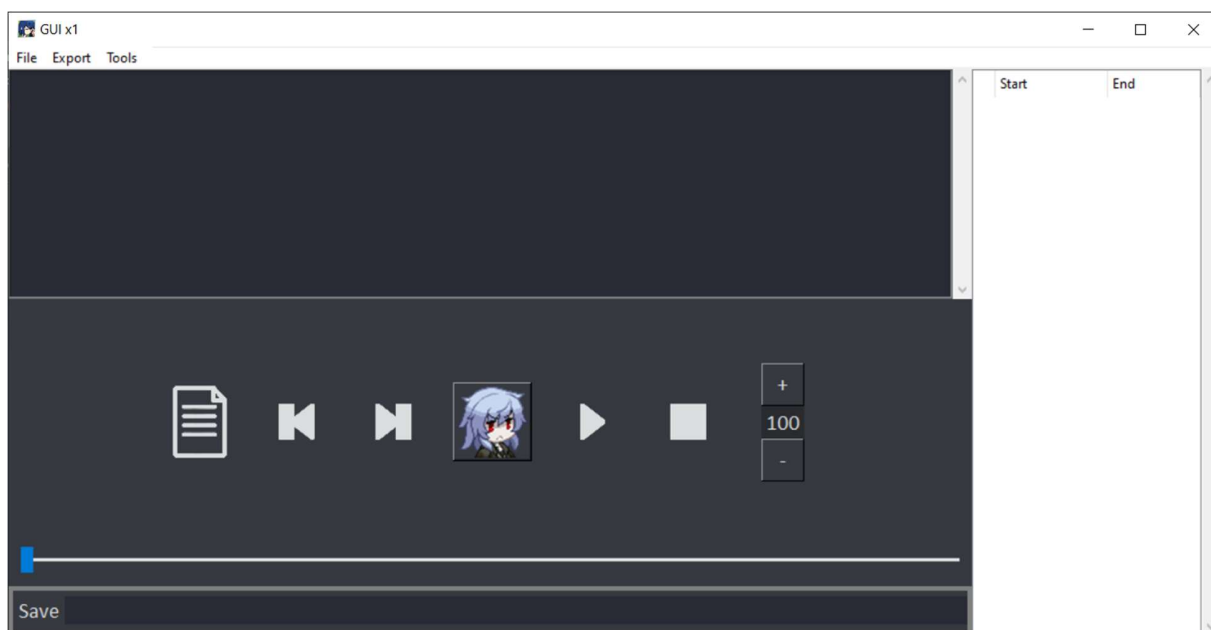
Kết quả thư mục giải nén sẽ có dạng như sau:



bin	9/10/2022 9:30 AM	File folder	
img	9/7/2022 9:53 AM	File folder	
GUI.exe	9/7/2022 10:00 AM	Application	10,330 KB

Hình 12: Kết quả giải nén phần mềm hỗ trợ

Thực hiện chạy thử chương trình bằng cách nhấn đúp vào GUI.exe; giao diện chương trình sẽ có dạng như sau:

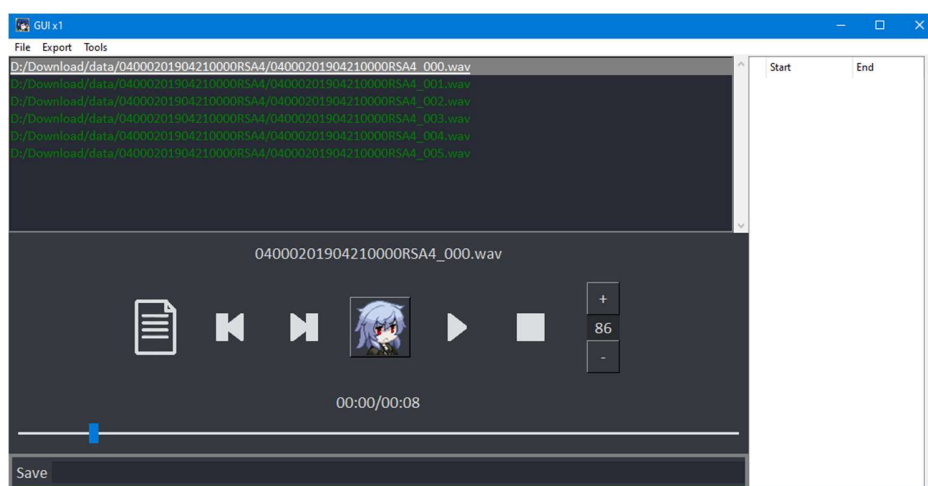


Hình 13: Giao diện chính của phần mềm hỗ trợ

2.2.1.2. Sử dụng công cụ cho phần việc chuyển văn bản

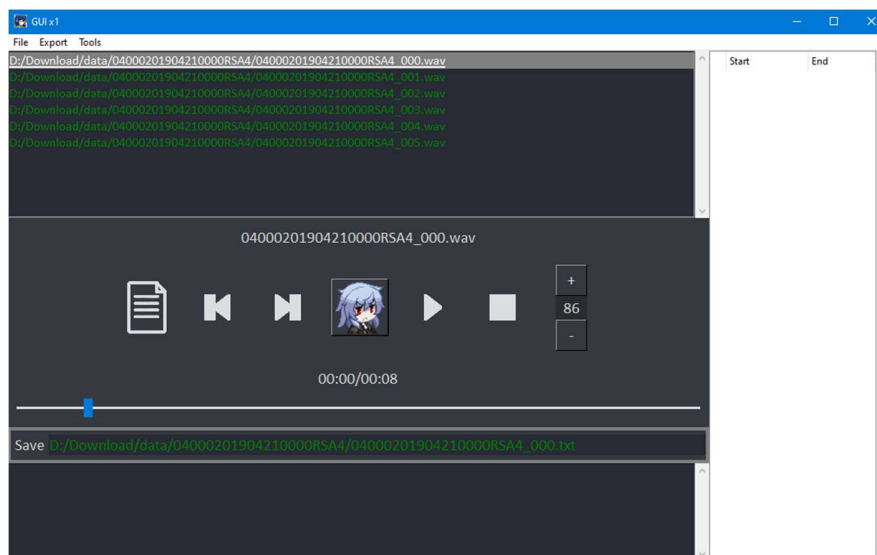
Giả sử việc phân đoạn (cắt) thành các câu nói “sạch” cho một file dữ liệu gốc hoặc cho tất cả các file dữ liệu gốc đã hoàn thành, bây giờ chúng ta tiếp tục thực hiện việc chuyển văn bản cho các phân đoạn.

Thêm các phân đoạn vào để làm việc với phần mềm hỗ trợ bằng cách từ menu, chọn File – Add folder (hoặc sử dụng tổ hợp phím tắt Ctrl+M). Sau khi chọn thư mục chứa các phân đoạn câu cần chuyển văn bản, tất cả các phân đoạn trong thư mục sẽ được thêm vào danh sách làm việc.



Hình 14: Thêm danh sách các phân đoạn từ một thư mục để thực hiện phiên âm

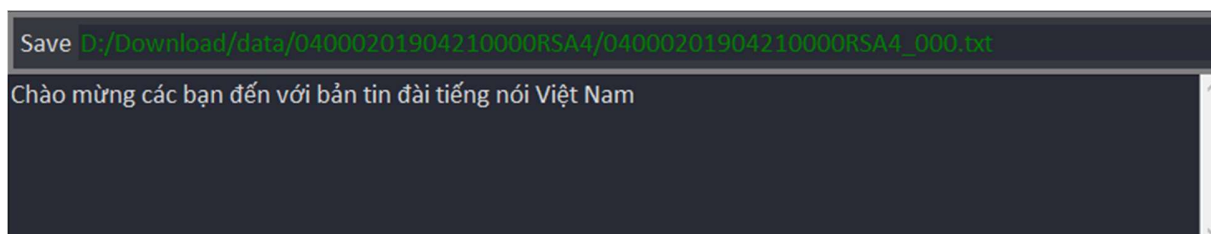
Tiếp tục từ menu, chọn Tools->Transcript để mở cửa sổ Transcript (phần cửa sổ nằm phía dưới cùng của giao diện)



Hình 15: Giao diện cửa sổ để thêm transcript

Chọn file phân đoạn cần chuyển văn bản bằng cách bấm đúp chuột vào tên file mong muốn trên danh sách file.

Thực hiện nghe và gõ trực tiếp viết phiên âm của file âm thanh vào cửa sổ Transcript (phần nằm dưới cùng của giao diện).



Hình 16: Giao diện phần nhập văn bản tương ứng với tiếng nói của phân đoạn

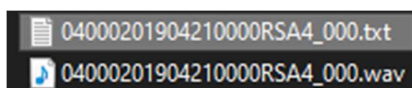
Trong quá trình nghe và gõ lại văn bản lời nói, chúng ta có thể thực hiện định vị vị trí điểm bắt đầu nghe hoặc nghe lại bằng cách điều khiển điểm con trỏ trên thanh trượt của player.

Trước khi lưu và chuyển sang phân đoạn câu tiếp theo, cần xác định nhanh các thông tin cơ bản của phân đoạn như: thể loại của phân đoạn, giới tính người nói, vùng miền của giọng nói, độ tuổi của người nói, cảm xúc của người nói, đặt trong cặp dấu ngoặc vuông theo cú pháp [xyzmn] (với x,y,z,m,n theo hướng dẫn ở trên). Các thông tin này là các đánh giá chủ quan của mình. Một số thông tin như độ tuổi chỉ mang tính tương đối dựa trên giọng. Sau đó gõ phần văn bản tương ứng với câu nói.

[0F0NY] Bà Hường ngồi trước sân chờ cho Hoa đi khuất hẳn

Hình 17: Minh họa một văn bản đầy đủ cùng thông tin đặc điểm phân đoạn

Sau khi xong, từ menu chọn File->Save (hoặc sử dụng tổ hợp phím tắt Ctrl+S; hoặc ấn nút Save trên màn hình) để lưu lại phiên âm của file âm thanh tương ứng. Phần mềm sẽ tự động lưu tên file phiên âm trùng với tên file âm thanh tương ứng ở định dạng đuôi .txt.



Hình 18: Minh họa kết quả sau khi hoàn thành chuyển văn bản cho phân đoạn câu sạch

Lặp lại thao tác trên với toàn bộ file phân đoạn âm thanh muốn phiên âm.

Chú ý:

Quá trình thực hiện thêm phiên âm có thể thực hiện tự động một phần **nếu có Internet**. Tuy nhiên **ĐỘ CHÍNH XÁC THẤP** nên cần chỉnh sửa sau khi chuyển đổi tự động.

Thực hiện các thao tác sau để chuyển văn bản tự động:

- Sau khi chọn file, từ menu, chọn Tools->Convert audio to text (hoặc sử dụng tổ hợp Ctrl+H) để chuyển phiên âm tự động.
- Sau khi chuyển tự động, cần nghe và kiểm tra, sử dụng chuột và bàn phím sửa lại phần phiên âm cũng như thêm thông tin đặc điểm phân đoạn (phần [xyzmn]) trước khi thực hiện lưu file phiên âm.

3. Cách thức nộp file dữ liệu

Tạo một thư mục có tên **Gxxmsv** trên máy cá nhân của mình

Trong đó xx sẽ là 04, 07, 08, hoặc 80; msv sẽ là mã sinh viên của mình (bao gồm cả phần chữ và phần số, phần chữ viết hoa).

Ví dụ sinh viên nhóm 04 có mã sinh viên là B21DCCN123 thì thư mục sẽ là **G04B21DCCN123**

Trong thư mục **Gxxmsv** sẽ gồm các thư mục con chứa các phân đoạn cắt được ở phần 2.1. và các file bản bản kèm theo. Mỗi thư mục con là sản phẩm các phân đoạn câu nói sạch và văn bản tương ứng của chúng ứng với một file dữ liệu gốc được giao. Tên thư mục trùng với file dữ liệu gốc được giao.

Ví dụ, sinh viên nhóm 01 có mã B21DCCN123 được giao 02 file dữ liệu có tên là MXN.mp3 và KBMD.mp3 thì tổ chức thư mục sẽ là

C01B21DCCN123

```
| --- MXN
|     |--- MXN-001.wav
|     |--- MXN-001.text
|     |--- ...
| --- KBMD
```

Khi nộp bài, mỗi sinh viên sẽ có một thư mục nộp bài riêng có tên thư mục trùng với mã sinh viên (thư mục này sẽ được thông báo sau). Thực hiện upload lần lượt từng thư mục kết quả lên thư mục drive **msv** (sẽ được thông báo sau).

Khi Upload, chọn upload folder và chọn thư mục kết quả thứ nhất. Lặp lại cho các thư mục kết quả còn lại.

Chú ý 1: KHÔNG upload file dữ liệu gốc - file dữ liệu được giao.

Chú ý 2: Tùy tình trạng của mạng, nếu mạng chậm và kém ổn định thì nên upload lần lượt từng thư mục kết quả. Tránh tình trạng upload đồng thời và tất cả các thư mục đều bị crashed