**HỌC VIỆN KỸ THUẬT QUÂN SỰ**

**KHOA CÔNG NGHỆ THÔNG TIN**



**BÀI TẬP LỚN**

**TƯƠNG TÁC NGƯỜI MÁY**

**Đề tài: Phần mềm tổng hợp giọng nói và trả về kết quả ở dạng văn bản.**

**Giáo viên hướng dẫn : TS. Ngô Thành Long**

**Sinh viên thực hiện: Trương Tiến Phúc – TH11A**

**Lê Văn Thứ – TH11A**

**Dương Hồ Minh Tú – TH11A**

**Phạm Hồng Thuận – TH11A**

***Hà Nội 05/2015*LỜI MỞ ĐẦU**

Ngày nay, công nghệ thông tin được sử dụng vào hầu hết mọi lĩnh vực trong đời sống con người. Một trong những lĩnh vực mà ngày nay là vấn đề nghiên cứu vô cùng quan trọng trong việc sử dụng máy tính đó là “Tương tác người máy (HCI)”, (HCI: Human Computer Interaction).

Vậy tương tác người máy là gì? Đó là sự nghiên cứu và phát triển các giao diện máy tính với mục đích làm cho người dùng dễ dàng sử dụng hệ thống máy tính đó hơn. HCI giao tiếp người máy không chỉ là thiết kế giao diện. HCI còn là lĩnh vực nghiên cứu liên quan đến:

- Nghiên cứu việc con người sử dụng giao diện thiết bị.

- Nghiên cứu phát triển các ứng dụng mới cho người dùng.

- Phát triển các thiết bị và công cụ mới cho người dùng.

HCI ngoài việc nghiên cứu các vấn đề trên còn liên quan tới các chuyên ngành khác như:

- Tâm lý học, xã hội học, triết học

- Sinh lý học, công thái học (là khoa học về việc thiết kế các máy móc, các công cụ, các máy tính và khu vực làm việc vật lý, sao cho mọi người dễ tìm thấy chúng và thoải mái sử dụng)

- Thiết kế đồ họa và công nghiệp, thiết kế âm thanh, điện ảnh

- Kỹ nghệ phần mềm

- Kỹ thuật điện, điện tử

Như vậy, HCI có một tầm rất quan trọng trong đời sống của chúng ta đặc biệt với xu thế công nghệ đang phát triển. Nếu thiết kế được một giao diện tương tác giữa người và máy tốt sẽ giảm chi phí do những trục trặc, tăng khả năng bán sản phẩm, giảm những lỗi nguy hiểm tính mạng con người, xét trên phương diện kinh tế tăng năng suất lao động, giảm chi phí đào tạo, giảm lỗi do người dùng, tạo ra sản phẩm chất lượng cao.

Hà Nội, tháng 6 năm 2015

1. **Giới thiệu bài toán**

Xử lý tiếng nói đang là vấn đề được quan tâm vì nó được ứng dụng rộng rãi trong nhiều lĩnh vực khác nhau... qua đó giúp quá trình tương tác người-máy trở nên hiệu quả và tự nhiên hơn.

Công nghệ nhận dạng giọng nói mới có thể giúp bạn tiết kiệm thời gian đáng kể nếu bạn nhập văn bản bằng cách sử dụng chế độ chính tả và nếu bạn kiểm soát menu bằng cách sử dụng lệnh thoại chế độ.

Ứng dụng của “nhận diện giọng nói”:

* + Phần mềm được chạy và phát triển trên điện thoại thông minh.
  + Phần mềm hỗ trợ việc giao tiếp giữa người bình thường (không biết ngôn ngữ cơ thể) với người bị khiếm thính và ngược lại.
  + Phần mềm có ứng dụng chuyển đổi từ một bài văn, bài thơ, lời bài hát từ giọng nói thành văn bản để in ra, lưu lại hoặc gửi đi rộng rãi thông qua SMS, mạng xã hội.
  + Phần mềm ghi lại cuộc nói chuyện, buổi thuyết trình, buổi đàm phán,... dưới dạng văn bản để in ấn, lưu truyền dễ dàng cho đông đảo mọi người.
  + Có thể ứng dụng phần mềm vào việc tìm kiếm bằng giọng nói, hoặc dịch sang ngôn ngữ khác
  + ….

Dưới đây chúng em sẽ thiết kế phần mềm ***“Nhận dạng giọng nói trên điện thoại thông minh”*** để thể hiện tương tác người máy.

Tài liệu được đưa ra nhằm đặc tả các thiết kế giao tiếp người máy được sử dụng trong hệ thống, bao gồm giới thiệu một số sản phẩm tương tự, xác định và phân tích yêu cầu bài toán, các thiết kế giao diện đối với mỗi chức năng.

1. **Các sản phẩm tương tự**
2. **Dragon Search**
3. **Giới thiệu**

Dragon Search là phần mềm tìm kiếm bằng giọng nói, cho phép bạn tìm kiếm trên cả Google, Wikipedia, Youtube và một số kênh khác nữa. Dragon Search hiện tại chỉ hỗ trợ một số khu vực và khả năng nhận diện tiếng Anh cũng như tiếng Việt rất chính xác. Kết quả tìm kiếm được lấy trực tiếp từ các công cụ tìm kiếm chuyên nghiệp, do vậy không có gì phàn nàn về kết quả trả về.

1. **Ưu điểm**

* Với tính năng tìm kiếm bằng giọng nói, Dragon Search sẽ giúp người dùng iOS có thể ra lệnh thiết bị tìm kiếm thông tin mong muốn dễ dàng. Phương thức này được giới thiệu là nhanh gấp 5 lần so với việc gõ trên bàn phím. Dragon Search hỗ trợ tìm kiếm từ các công cụ Google, Yahoo, Twitter, iTunes, Wikipedia và YouTube với tốc độ khá nhanh.
* Giao diện đơn giản, hiệu quả, cho phép người dùng chỉ cần đọc truy vấn để tìm kiếm thông tin và câu trả lời trên web bằng điện thoại một cách dễ dàng và nhanh chóng
* Được cung cấp miễn phí.

1. **Nhược điểm**

* Nhận dạng chưa chính xác khi nói nhanh hoặc nới các từ nóng, từ địa phương.
* Chỉ hỗ trợ tìm kiếm, chưa hỗ trợ các chức năng nâng cao khác

1. **Dragon Dictation**
2. **Giới thiệu**

Ra mắt lần đầu tiên trên Apple App Store tại Mỹ vào tháng 12/2009, Dragon Dictation nhanh chóng trở thành ứng dụng được tải về hàng đầu, theo đại diện của Nuance thì tiếng Việt là ngôn ngữ thứ 38 được hãng hỗ trợ. Dragon Dictation tương thích với iOS 5, mang tính năng của một thanh công cụ pop-up với các biểu tượng ứng dụng giúp người dùng có thể truy cập email, tin nhắn SMS, cập nhật trạng thái trên Facebook và Twitter hoặc viết trên vùng đệm. Dictation còn được trang bị chức năng: Tự động lưu thuận tiện cho việc nhớ đoạn văn bản đã được chuyển từ giọng nói khi người dùng bị cuộc gọi đến làm gián đoạn.

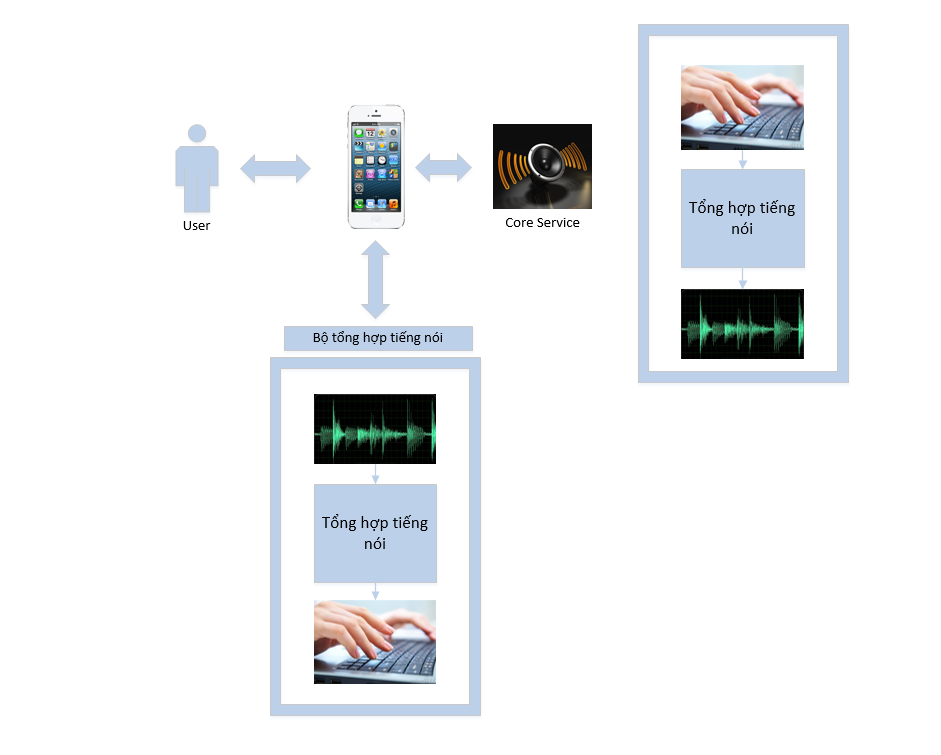
1. **Ưu điểm**

* Giao diện Dragon Dictation vô cùng đơn giản nhưng hiệu quả.
* Có thể nhận diện được một số cụm từ hoặc câu ngắn phổ biến khá chính xác.
* Hỗ trợ nhiều tiện ích như truy cập email, tin nhắn SMS, cập nhật trạng thái trên Facebook và Twitter…
* Được cung cấp miễn phí.

1. **Nhược điểm**

* Nhận dạng chưa chính xác khi nói nhanh hoặc nới các từ nóng, từ địa phương.
* Chưa hỗ trợ tự động nhận dạng ngôn ngữ nói bất kì.

1. **Xác định, phân tích yêu cầu bài toán**
2. **Mô tả hệ thống**
3. **Tổng quan**

******

*Hình 1: Các thành phần và tác nhân của hệ thống*

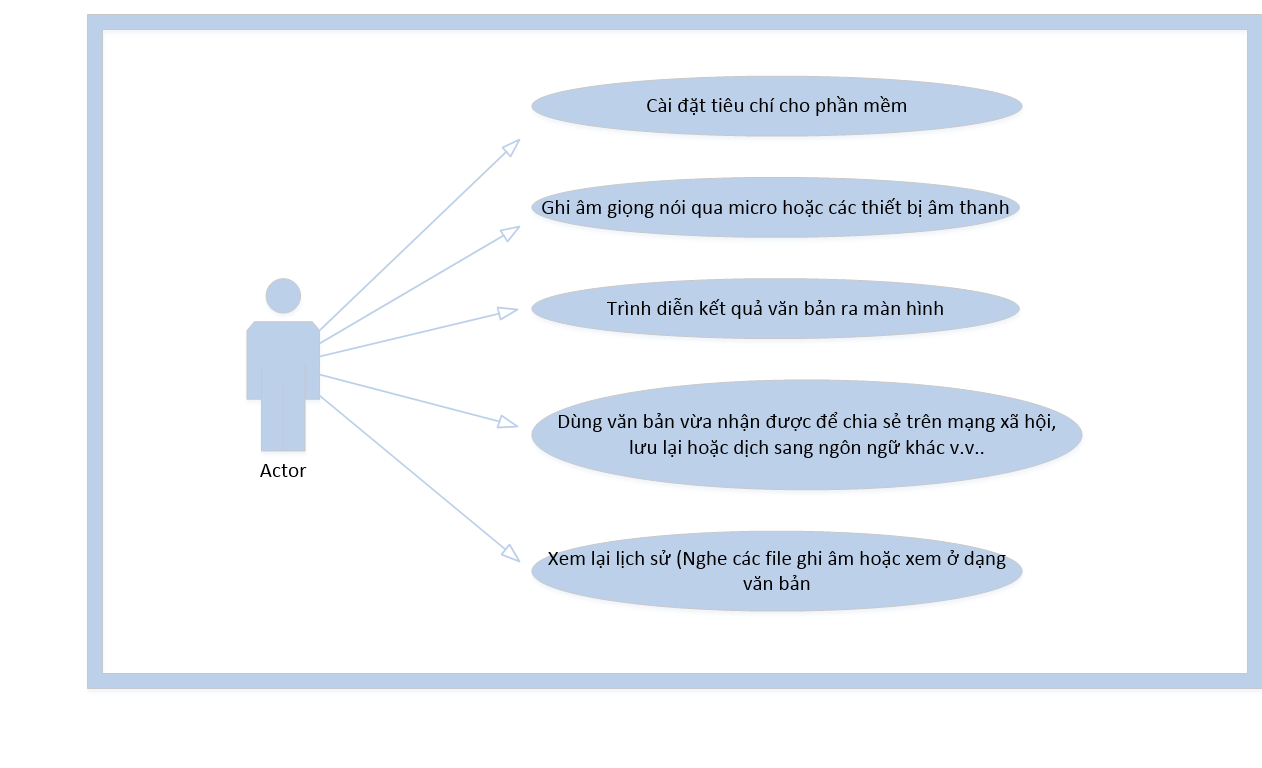
1. **Phạm vi**

**Đối tượng:** Công nghệ xử lý tiếng nói.

**Phạm vi:** Chỉ xử lý và nhận dạng ngôn ngữ nói.

1. **Yêu cầu chức năng**

**Mô hình Use Case:**



1. **Khảo sát yêu cầu tương tác của người dùng**
2. **Yêu cầu chung**

Nhận dạng tiếng nói cho phép chuyển đổi từ giọng nói sang văn bản. Có thể xử lý được ngôn ngữ của nhiều quốc gia. Có thể sử dụng văn bản thu được để phục vụ các tiện ích khác như gửi tin nhắn, chia sẻ trên mạng xã hội, dịch sang ngôn ngữ khác …

Đối với người sử dụng, người sẽ thao tác chủ yếu với thành phần của hệ thống, thông tin nhập cần nhanh chóng chính xác.

1. **Các nhóm đối tượng sử dụng**
   1. **Khách hàng:**
      1. **Đặc điểm tâm lý:**

* Thái độ và động cơ:
* Sự kiên nhẫn: Trung bình, cần tốc độ và sự chính xác, ít chú ý chi tiết
* Mong muốn: Văn bản đưa ra cần chính xác, nhanh.
  + 1. **Đặc điểm vật lý**
    - Tuổi: Trên 5.
    - Giới tính: Tất cả mọi người.
    - Tay thuận: Không.
    - Khuyết tật: Có thể bị khiếm thích (điếc) hoặc không.
    1. **Tốc độ tương tác:**
    - Tốc độ đọc: Phụ thuộc vào tuổi, với tuổi trẻ thì đọc nhanh lướt; tuổi trung niên thì đọc chậm, tỉ mỉ.
    1. **Kiếm thức và kinh nghiệm**
    - Kiến thức về máy tính: Trung bình
    - Kiến thức về hệ thống: Trung bình
    - Kiến thức về ứng dụng phần mềm: Khá
    - Trình độ giáo dục: Đa phần là trên phổ thông.
  1. **Người quản trị hệ thống (Nhà phát triển phần mềm)**
     1. **Đặc điểm tâm lý:**
* Sự kiên nhẫn: Khá, cần tốc độ và sự chính xác, chú ý chi tiết, tỉ mỉ.
* Mong muốn: Cần một hệ thống có tính an toàn cao, chính xác đáp ứng được các yêu cầu người dùng.
  + 1. **Đặc điểm vật lý:**
* Tuổi: Thường là trẻ tuổi.
* Giới tính: Tất cả mọi người.
* Tay thuận: Không.
  + 1. **Tốc độ tương tác**
* Tốc độ đọc: Tốt, có khả năng tổng hợp thông tin.
* Tốc độ đánh máy: Tốt.
  + 1. **Trình độ kinh nghiệm**
* Kiến thức về máy tính: Tốt
* Kiến thức về hệ thống: Tốt
* Kiến thức về ứng dụng phần mềm: Tốt
* Kiến thức chuyên môn nghiệp vụ: Tốt
* Trình độ giáo dục: Trên cao đẳng.

1. **Các yêu cầu về chức năng**
   * + - Đầu vào: Giọng nói (ngôn ngữ tự nhiên).

Đầu ra: Hiển thị dưới dạng văn bản (text) và âm thanh.

* + - * Phân tích tổng hợp giọng nói nhanh, ổn định, chuẩn xác.
      * Đối với đầu vào: Tổng hợp phân tích được các giọng nói, các từ ngữ một cách chính xác => có chức năng ghi âm.
      * Đối với đầu ra: Có chức năng hiển thị những gì đã ghi âm; tự động lưu trữ lại file ghi âm khi người dùng tạm dừng hoặc dừng quá trinh ghi âm (giới hạn lưu trữ trong 1 tuần).
      * Phần mềm có khả năng nhận diện đa ngôn ngữ.
  + Khi ghi âm có nhiều giọng nói xung quanh cần lọc ra được giọng nói chính cần ghi âm.
    - * Tiện ích:
    - Chức năng kết nối chia sẻ với cộng đồng thông qua các ứng dụng được cài đặt trên thiết bị như tin nhắn, Facebook, Twitter....
    - Chức năng dịch sang ngôn ngữ khác.
    - Chức năng lưu lại file văn bản, file âm thanh.
  + Các thông báo đưa ra cần chính xác, xúc tích.
  + Cài đặt: Bao gồm các tùy chọn về lựa chọn ngôn ngữ nhận dạng, lựa chọn ngôn ngữ phần mềm, lựa chọn font chữ, tải lại cài đặt của ứng dụng …
  + Xem lịch sử: Cho phép người dùng xem danh sách các file ghi âm được tạo trong 1 tuần gần nhất. Và có tùy chọn xóa nếu như người dùng có nhu cầu.
  + Tạm dừng ghi âm và đưa ra thông báo khi bộ nhớ đầy.
  + Tự động tạm dừng ghi âm khi người dùng bị cuộc gọi đến làm gián đoạn.

1. **Yêu cầu phi chức năng**
   1. **Nhằm lợi ích người sử dụng**

* Yêu cầu tính tiện dụng: Giao diện thân thiện với người sử dụng.
* Yêu cầu tính ổn định:

+ Không làm treo hệ thống.

+ Có tính mềm dẻo và nâng cấp với xác nhận từ khách hàng.

* Yêu cầu về hiệu năng:

+ Đáp ứng khối lượng thông tin lưu trữ lớn.

+ Tốc độ xử lí tốt.

* Tính sáng tạo: Hệ thống có cơ chế nhận dạng giọng nói một cách tự động, xử lý theo thời gian thực.
  1. **Nhằm lợi ích phát triển của phần mềm**
* Yêu cầu tính dễ phát triển và hoàn thiện: có thể mở rộng và tăng cường thêm một số chức năng một cách dễ dàng.
* Yêu cầu tính modun: có sự độc lập chức năng của các thành phần trong hệ thống.
* Yêu cầu sao lưu và phục hồi: Quản lý phiên bản tốt, dễ sao lưu, phục hồi khi hệ thống gặp sự cố.
* Yêu cầu về tính hỗ trợ: Hệ thống hỗ trợ một cách dễ dàng việc thực hiện các chức năng.
* Cơ sở dữ liệu đảm bảo quản trị tốt, hỗ trợ cho người quản trị chỉnh sửa thông tin tối ưu.

1. **Đặc tả chức năng chính**
2. **Chức năng nhận dạng giọng nói** 
   1. **Ghi âm:**

Cần một micro để nhận âm thanh. Tốc độ nhận dạng kém hơn đáng kể hoặc có thể không được chấp nhận nếu bạn sử dụng một micro kém. Chất lượng âm thanh đầu vào quyết định tính chính xác của dữ liệu đầu ra.

Ghi âm giọng nói được bật khi bạn chạm vào biểu tượng “Micro” trên giao diện chính.

Để dừng nhận dạng giọng nói, bạn click chọn biểu tượng “Micro” trên giao diện chính.

**Lưu ý:** Quá trình ghi âm sẽ luôn được thực hiện cho đến khi người dùng lựa chọn tạm dừng, dừng hoặc khi ứng dụng không nhận được âm thanh trong quãng thời gian chờ (5s). Trong quá trình sử dụng, nếu như người dùng đột ngột thoát khỏi ứng dụng thì dữ liệu vẫn sẽ được tự động lưu.

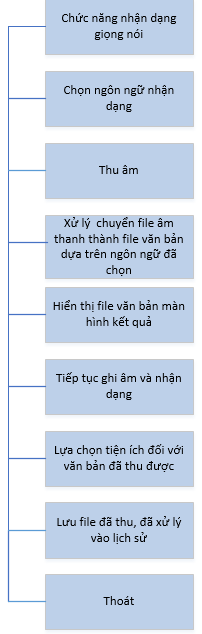
* 1. **Hiển thị văn bản**

Đây cũng là một trong những chức năng chính của phần mềm. Chức năng này được thực hiện trong khi người dùng đang ghi âm, có nghĩa là người dùng nói đến đâu thì khung màn hình hiển thị sẽ hiển thị văn bản luôn đến đó.

* 1. **Chức năng tiếp tục nhận dạng**

Chức năng này được thực hiện khi người dùng chọn nút “Ghi âm” có biểu tượng “Micro” trên giao diện tạm ngừng. Khi đó, hệ thông tiếp tục ghi âm và văn bản được xử lí sẽ hiển thị nối tiếp phần văn bản của phiên làm việc trước đó.

***Lưu đồ mô tả chức năng:***



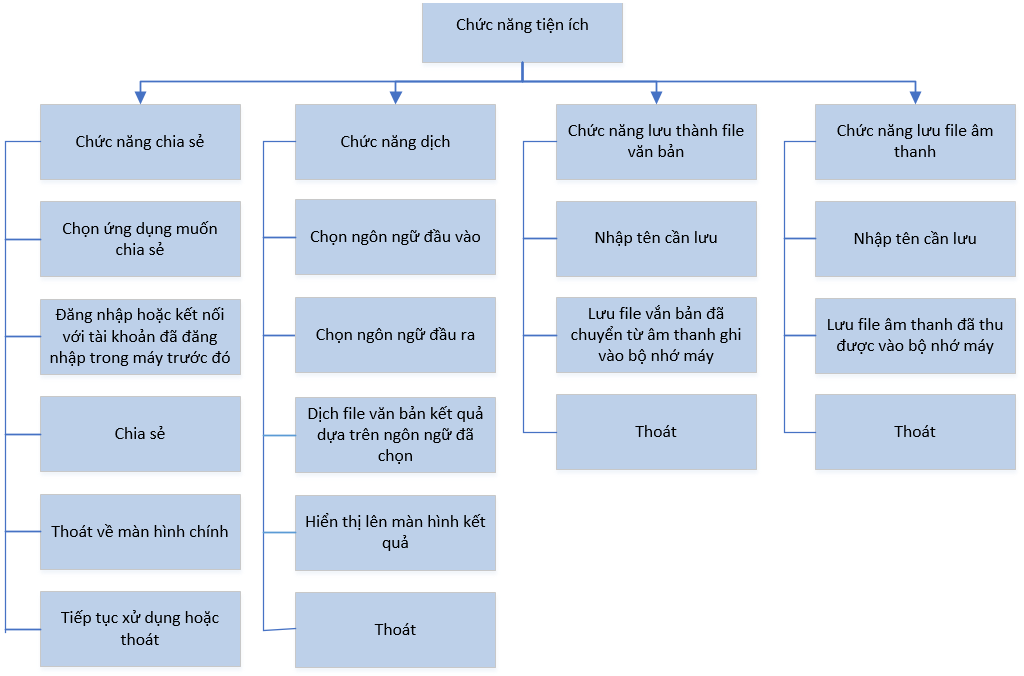
1. **Chức năng “Tiện ích”**

Sau khi nhận được văn bản hiển thị trên màn hình, nếu bạn muốn lưu lại giọng nói bạn có thể lưu lại file âm thanh hay file văn bản.

Bạn cũng có thể chia sẻ văn bản đi cho mọi người qua ứng dụng được cài đặt trong thiết bị…

Tiện ích cũng hỗ trợ dịch văn bản thu được sang ngôn ngữ khác. Với chức năng dịch, người dùng sẽ lựa chọn ngôn ngữ đầu vào và đầu ra.

***Lưu đồ mô tả chức năng:***

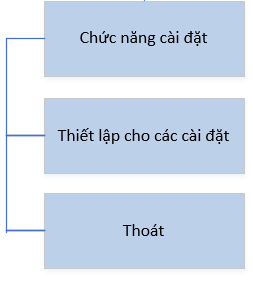


1. **Chức năng cài đặt**

Với mục cài đặt, bạn có thể lựa chọn cài đặt cho các tiêu chí của phần mềm.

* Lựa chọn ngôn ngữ phần mềm: Mục này cho phép bạn lựa chọn ngôn ngữ hiển thị của phần mềm.
* Lựa chọn ngôn ngữ nhận dạng: Mục này cho phép bạn lựa chọn ngôn ngữ nhận dạng khi bạn ghi âm (Với giá trị là tự động hoặc 1 ngôn ngữ cụ thể nào đó)
* Lựa chọn font chữ: Mục này cho phép bạn lựa chọn font chữ với kích thước to hoặc nhỏ…
* Lựa chọn reset phần mềm: Chức năng này đưa các mục cài đặt về mặc định, và xóa lịch sử ghi âm.

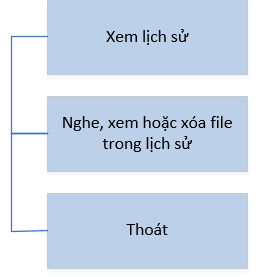
***Lưu đồ mô tả chức năng:***



1. **Chức năng xem lịch sử**

Với tùy chọn xem lịch sử, màn hình sẽ hiển thị lên 1 danh sách các file thu âm được trong 1 tuần trở lại. Bạn có thể tùy chọn nghe hay xem file văn bản tương ứng. Bạn cũng có thể xóa file âm thanh nào đó khi click chọn biểu tượng xóa trên file âm thanh tương ứng.

***Lưu đồ mô tả chức năng:***



1. **Chức năng lọc tiếng ồn**

Phần mềm có thể tích ứng thêm một số công nghệ lọc tiếng ồn, lọc âm thanh để nhận dạng tiếng nói một cách chính xác nhất.

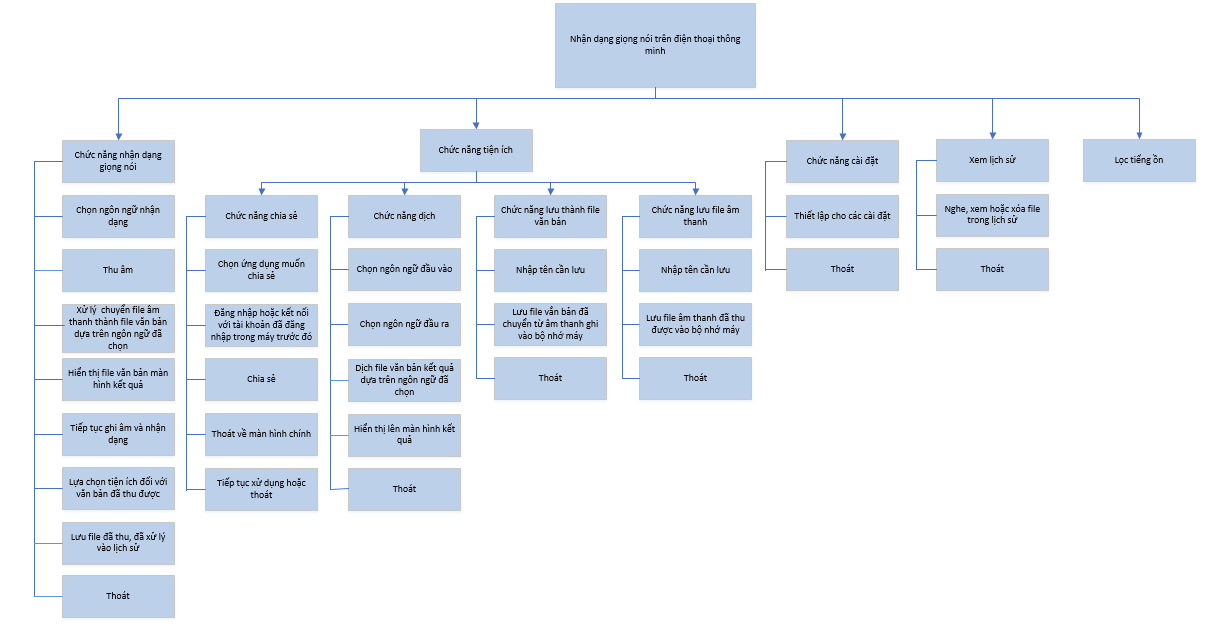
1. **Chức năng tạm dừng khi có cuộc gọi đến**

Tự động dừng ghi âm và lưu khi người dùng bị cuộc gọi đến làm gián đoạn.

1. **Chức năng thoát.**

Sau khi người dùng không muốn nhận dạng nữa có thể click vào biểu tượng “Home” trên thiết thị để tắt ứng dụng

***Lưu đồ mô tả toàn chức năng của toàn bộ hệ thống:***



1. **Thiết kế giao diện người dùng.**

…..