**TỰ ĐỘNG PHÂN ĐOẠN TÍN HIỆU THÀNH TIẾNG NÓI VÀ KHOẢNG LẶNG**

**DỰA VÀO NĂNG LƯỢNG NGẮN HẠN CỦA TÍN HIỆU**

**Nguyễn Nhật Tùng, Ninh Hải Hoàng**

Khoa Công nghệ thông tin, Trường Đại học Bách Khoa, Đại học Đà Nẵng

nhattungnguyen.2kgl@gmail.com, ninhhaihoang@gmail.com

**Nhóm X, lớp HP: 18N15**

|  |  |  |  |
| --- | --- | --- | --- |
| **Điểm** | **Bảng phân công nhiệm vụ** | | **Chữ ký của SV** |
|  | Nguyễn Nhật Tùng (nhóm trưởng) | Viết báo cáo (tr. 1, 3), viết báo cáo kết quả thực nghiệm và kết luận (tr. A-b). |  |
|  | Ninh Hải Hoàng | Viết báo cáo (tr. 1, 3), đọc tài liệu, viết báo cáo về cơ sở lý thuyết,mã cài đặt thuật toán và tài liệu tham khảo (tr. X-y) |  |

Lời cam đoan: Chúng tôi, gồm các sinh viên có chữ ký ở trên, cam đoan rằng báo cáo này là do chúng tôi tự viết dựa trên các tài liệu tham khảo liệt kê ở cuối báo cáo. Các số liệu thực nghiệm và mã nguồn chương trình nếu không chỉ dẫn nguồn tham khảo đều do chúng tôi tự làm. Nếu vi phạm thì chúng tôi xin chịu trách nhiệm và tuân theo xử lý của giáo viên hướng dẫn.

TÓM TẮT— Tự động phân đoạn tín hiệu thành tiếng nói và khoảng lặng là công việc cơ bản mà quan trọng trong lĩnh vực xử lý tín hiệu số, đặc biệt là trong việc nhận diện giọng nói, giúp xác định các khoảng có tiếng nói và khoảng không có tiếng nói. Thuật toán được trình bày sau đây chia tín hiệu âm thanh đầu vào thành các đoạn ngắn, sau đó dựa vào năng lượng của tín hiệu trên từng đoạn đem so sánh với giá trị năng lượng làm mốc xác định để phân đoạn tín hiệu đầu vào thành các đoạn tiếng nói và khoảng lặng. Kết quả thử nghiệm với 4 mẫu tín hiệu cho thấy các biên được tìm tự động có giá trị gần đúng với giá trị biên tìm bằng phương pháp thủ công, với sai số trung bình là: 0.0533 giây, hoặc 0.0420 giây (khi dùng phương pháp chuẩn hóa về thang giá trị [0;1]).

Từ khóa— xử lý tín hiệu số, phân biệt tiếng nói và khoảng lặng, năng lượng ngắn hạn.

Mục lục

[I. ĐẶT VẤN ĐỀ 3](#_Toc55232972)

[II. LÝ THUYẾT XỬ LÝ TÍN HIỆU TIẾNG NÓI VÀ CÁC THUẬT TOÁN 3](#_Toc55232973)

[III. MÃ CHƯƠNG TRÌNH CÀI ĐẶT CÁC THUẬT TOÁN 3](#_Toc55232974)

[IV. KẾT QUẢ THỰC NGHIỆM 3](#_Toc55232975)

[A. Dữ liệu mẫu 3](#_Toc55232976)

[B. Kết quả thu được 3](#_Toc55232977)

[C. Sử dụng thuật toán chuẩn hóa về thang giá trị [0;1] 6](#_Toc55232978)

[V. KẾT LUẬN 9](#_Toc55232979)

[VI. TÀI LIỆU THAM KHẢO 9](#_Toc55232980)

# ĐẶT VẤN ĐỀ

Kể từ khi các kỹ sư có thể dựa vào các đặc trưng âm thanh để phân biệt các âm tiết, người ta đã nảy ra nhiều ý tưởng để nhận biết các loại âm tiết khác nhau, từ các âm tiết cơ bản tới các âm tiết phức tạp, nhằm mục đích nhận biết các từ ngữ dựa vào tín hiệu âm thanh được ghi âm.

Tuy nhiên để nhận biết được chính xác hơn, hệ thống nhận biết từ ngữ đó cần phải sử dụng các đoạn âm thanh chỉ chứa tiếng nói (hay các tín hiệu âm thanh có giá trị). Vì vậy việc tự động phân đoạn tín hiệu thành tiếng nói và khoảng lặng giúp xác định các khoảng có tiếng nói và khoảng không có tiếng nói từ đó giúp hệ thống khác trích xuất tín hiệu giọng nói và dùng cho việc nhận diện giọng nói sau đó.

Thuật toán được trình bày sau đây thực hiện việc tự động phân đoạn tín hiệu thành tiếng nói và khoảng lặng dựa vào năng lượng ngắn hạn (short-time energy) của tín hiệu âm thanh. Cụ thể: Thuật toán này chia tín hiệu âm thanh đầu vào thành các đoạn ngắn, sau đó dựa vào năng lượng của tín hiệu trên từng đoạn đem so sánh với giá trị năng lượng làm mốc xác định để phân đoạn tín hiệu đầu vào thành các đoạn tiếng nói và khoảng lặng.

Bài viết có bố cục như sau: Phần II trình bày tổng quan về cơ sở lý thuyết liên quan tới tín hiệu âm thanh, nguyên lý của các thuật toán, những vấn đề phát sinh trong thuật toán và cách khắc phục. Phần III ghi mã nguồn cách cài đặt thuật toán bằng Matlab. Phần IV trình bày kết quả thu được và các đánh giá kết quả đó khi áp dụng thuật toán lên các dữ liệu mẫu; so sánh thuật toán chuẩn hóa hiện tại với thuật toán chuẩn hóa khác (!). Cuối cùng là kết luận rút ra được trình bày ở phần V.

# LÝ THUYẾT XỬ LÝ TÍN HIỆU TIẾNG NÓI VÀ CÁC THUẬT TOÁN

# MÃ CHƯƠNG TRÌNH CÀI ĐẶT CÁC THUẬT TOÁN

# KẾT QUẢ THỰC NGHIỆM

## Dữ liệu mẫu

Dữ liệu mẫu được sử dụng để đánh giá thuật toán là 4 tín hiệu giọng nói được thu âm bởi 2 người khác nhau trong 2 môi trường khác nhau, được lấy mẫu với tần số 16 kHz, độ dài trung bình 8 giây.



Hình 1. 4 tín hiệu mẫu được sử dụng để đánh giá thuật toán.

## Kết quả thu được

Áp dụng thuật toán cho từng dữ liệu mẫu, ta thu được kết quả như trong các hình bên dưới.



Hình 2. So sánh biên tìm được với biên chuẩn cho tín hiệu mẫu “lab-male”.

C:\Users\nhatt\AppData\Local\Microsoft\Windows\INetCache\Content.Word\result_lab-female_std.eps

Hình 3. So sánh biên tìm được với biên chuẩn với tín hiệu mẫu “lab-female”.



Hình 4. So sánh biên tìm được với biên chuẩn với tín hiệu mẫu “studio-male”.



Hình 5. So sánh biên tìm được với biên chuẩn với tín hiệu mẫu “studio-female”.

Sai số trung bình (Root mean square error – RMSE) của giữa biên chuẩn và biên tìm bằng thuật toán cho từng dữ liệu mẫu được thể hiện trong hình dưới đây.



Hình 6. Sai số trung bình (RMSE) của thuật toán cho từng dữ liệu mẫu.

Như vậy sai số trung bình của thuật toán trong 4 dữ liệu mẫu là 5.3267 khung (0.0533 giây).

## Sử dụng thuật toán chuẩn hóa về thang giá trị [0;1]

Bằng cách sử dụng thuật toán chuẩn hóa khác, đưa năng lượng ngắn hạn về thang giá trị [0;1], ta cũng thu được kết quả tương đối chính xác như các hình bên dưới.



Hình 7. So sánh biên tìm được với biên chuẩn với tín hiệu mẫu “lab-male” (chuẩn hóa về [0;1]).



Hình 8. So sánh biên tìm được với biên chuẩn với tín hiệu mẫu “lab-female” (chuẩn hóa về [0;1]).



Hình 9. So sánh biên tìm được với biên chuẩn với tín hiệu mẫu “studio-male” (chuẩn hóa về [0;1]).



Hình 10. So sánh biên tìm được với biên chuẩn với tín hiệu mẫu “studio-female” (chuẩn hóa về [0;1]).

Sai số của thuật toán sử dụng cách chuẩn hóa này được thể hiện trong hình sau.



Hình 11. Sai số trung bình (RMSE) của thuật toán cho từng dữ liệu mẫu (sử dụng phương pháp chuẩn hóa về [0;1]).

Như vậy sai số trung bình là 4.2041 khung (0.0420 giây).

Qua 4 dữ liệu mẫu, ta thấy sử dụng thuật toán chuẩn hóa này đem lại sai số nhỏ hơn so với cách chuẩn hóa đưa về phân phối chuẩn tắc. Sai số của thuật toán phân đoạn tiếng nói-khoảng lặng sử dụng 2 phương pháp chuẩn hóa khác nhau được thể hiện trong hình sau đây.



Hình 12. So sánh sai số trung bình (RMSE) của thuật toán cho từng dữ liệu mẫu khi dùng từng cách chuẩn hóa.

# KẾT LUẬN

Nhóm đã cài đặt thành công thuật toán tự động phân đoạn tiếng nói và khoảng lặng dựa vào năng lượng của tín hiệu theo 2 phương pháp chuẩn hóa khác nhau là đưa về phân phối chuẩn tắc và chuẩn hóa về thang giá trị [0;1].

Kết quả thử nghiệm với 4 mẫu tín hiệu cho thấy các biên được tìm tự động có giá trị gần đúng với giá trị biên tìm bằng phương pháp thủ công, với sai số trung bình là: 0.0533 giây, hoặc 0.0420 giây (khi dùng phương pháp chuẩn hóa về thang giá trị [0;1]). Kết quả đó cũng cho thấy sử dụng phương pháp chuẩn hóa đưa về thang giá trị [0;1] giúp giảm sai số của thuật toán.

Trong tương lai nhóm sẽ thử nghiệm thuật toán với nhiều mẫu dữ liệu khác để tìm ra ngưỡng năng lượng chuẩn chính xác hơn giúp giảm sai số của thuật toán.

# TÀI LIỆU THAM KHẢO