**TÌM TẦN SỐ CƠ BẢN CỦA TÍN HIỆU TIẾNG NÓI**

**DÙNG HÀM TỰ TƯƠNG QUAN**

**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 6, 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— Việc tìm tần số cơ bản là công việc quan trọng trong việc nhận diện giọng nói, xác định ngữ điệu của người nói, là cơ sở để tái tạo lại âm thanh giống với tiếng nói tự nhiên của con người. Việc dùng hàm tự tương quan để tìm tần số cơ bản trên miền thời gian là cách đơn giản và phổ biến, thông dụng để tìm chu kỳ cơ bản của tín hiệu giọng nói, và cũng là cách được sử dụng trong thuật toán được trình bày sau đây. Cụ thể: Thuật toán này chia tín hiệu đầu vào thành các đoạn tiếng nói-khoảng lặng. Sau đó, với mọi đoạn khoảng lặng thì đều có tần số cơ bản (F0) là không xác định. Ngược lại, với mỗi đoạn tiếng nói, thuật toán thực hiện chia tín hiệu thành các khung, sau đó tiến hành tính hàm tự tương quan cho khung tín hiệu, rồi tìm độ trễ của giá trị cực đại địa phương, sau đó là kiểm tra độ trễ có thỏa mãn các điều kiện đặc biệt của thuật toán hay không. Nếu thỏa mãn thì thuật toán tìm được F0 cho khung tín hiệu, nếu không thì khung tín hiệu có F0 không xác định. Kết quả thử nghiệm: (cập nhật sau)

Từ khóa— xử lý tín hiệu số, tự động đo tần số cơ bản, tự động tìm cao độ giọng nói, phân biệt tiếng nói-khoảng lặng, hàm tự tương quan, tín hiệu tiếng nói.

Mục lục

[I. ĐẶT VẤN ĐỀ 3](#_Toc55335758)

[II. CƠ SỞ LÝ THUYẾT VÀ CÁC THUẬT TOÁN 3](#_Toc55335759)

[A. Sơ đồ khối thuật toán 3](#_Toc55335760)

[B. Thuật toán tính năng lượng ngắn hạn 3](#_Toc55335761)

[C. Thuật toán chuẩn hóa 3](#_Toc55335762)

[*1.* Chuẩn hóa sử dụng phân phối chuẩn tắc 3](#_Toc55335763)

[*2.* Chuẩn hóa về thang giá trị [0;1] 3](#_Toc55335764)

[D. Thuật toán tìm giá trị biên 3](#_Toc55335765)

[E. Lỗi biên ảo 3](#_Toc55335766)

[*1.* Vấn đề 3](#_Toc55335767)

[*2.* Giải pháp đề xuất và kết quả 3](#_Toc55335768)

[III. MÃ CHƯƠNG TRÌNH CÀI ĐẶT CÁC THUẬT TOÁN 3](#_Toc55335769)

[A. Hàm tính năng lượng mỗi frame: 3](#_Toc55335770)

[B. Hàm chuẩn hoá tín hiệu X sử dụng phân phối chuẩn tắc: 3](#_Toc55335771)

[C. Hàm chuẩn hoá tín hiệu X về dạng [0;1]: 3](#_Toc55335772)

[D. Hàm tính các giá trị biên: 3](#_Toc55335773)

[E. Hàm lọc biên ảo 3](#_Toc55335774)

[F. Hàm tổng hợp (sử dụng phương pháp chuẩn hóa dùng phân phối chuẩn tắc) 3](#_Toc55335775)

[IV. KẾT QUẢ THỰC NGHIỆM 3](#_Toc55335776)

[A. Dữ liệu mẫu 3](#_Toc55335777)

[B. Kết quả thu được 3](#_Toc55335778)

[C. Sử dụng thuật toán chuẩn hóa về thang giá trị [0;1] 3](#_Toc55335779)

[V. KẾT LUẬN 3](#_Toc55335780)

[VI. TÀI LIỆU THAM KHẢO 3](#_Toc55335781)

# ĐẶT VẤN ĐỀ

Tần số cơ bản của tín hiệu tiếng nói đặc trưng cho cao độ của tiếng nói. Trong lĩnh vực xử lý tín hiệu số, nó là nghịch đảo của chu kỳ cơ bản. Trong lĩnh vực âm học, nó là tốc độ rung của dây thanh trong bộ máy phát âm của con người, là các âm tiết cơ bản trong thang thanh nhạc ( !hỏi wikipedia), và còn là một đặc trưng cơ bản cho ngữ điệu của tiếng nói. Việc tìm tần số cơ bản giúp xác định ngữ điệu của người nói, giúp nhận diện giọng nói; là cơ sở để tái tạo lại âm thanh giống với tiếng nói tự nhiên của con người, hỗ trợ việc giao tiếp tự nhiên giữa máy tính với con người và giữa con người với con người.

Việc dùng hàm tự tương quan để tìm tần số cơ bản trên miền thời gian là cách đơn giản và phổ biến, thông dụng để tìm chu kỳ cơ bản của tín hiệu giọng nói, và cũng là cách được sử dụng trong thuật toán được trình bày sau đây. Cụ thể: Thuật toán này chia tín hiệu đầu vào thành các đoạn tiếng nói-khoảng lặng. Sau đó, với mọi đoạn khoảng lặng thì đều có tần số cơ bản (F0) là không xác định. Ngược lại, với mỗi đoạn tiếng nói, thuật toán thực hiện chia tín hiệu thành các khung với độ dài tối đa là 20ms. Sau đó tiến hành tính hàm tự tương quan cho khung tín hiệu, rồi tìm độ trễ của giá trị cực đại địa phương, sau đó là kiểm tra độ trễ có thỏa mãn các điều kiện đặc biệt của thuật toán hay không. Nếu thỏa mãn thì thuật toán tìm được F0 cho khung tín hiệu, nếu không thì khung tín hiệu có F0 không xác định.

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 hiện tại với thuật toán khi không thực hiện chia tiếng nói-khoảng lặng. Cuối cùng là kết luận rút ra được trình bày ở phần V.

# CƠ SỞ LÝ THUYẾT VÀ CÁC THUẬT TOÁN

## Sơ đồ khối thuật toán

## Thuật toán phân biệt tiếng nói-khoảng lặng

## Hàm tự tương quan

## Âm vô thanh và âm hữu thanh

## Thuật toán tìm tần số cơ bản của tín hiệu

## Vấn đề phát sinh

## Giải pháp đề xuất và kết quả

# MÃ CHƯƠNG TRÌNH CÀI ĐẶT CÁC THUẬT TOÁN

Mã chương trình ở đây viết trên ngôn ngữ Matlab.

# 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 lấy mẫu là 16 kHz, độ dài trung bình là 8 giây.



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

## Kết quả thu được

Áp dụng thuật toán cho từng dữ liệu mẫu với ngưỡng năng lượng chuẩn E0=0.4, ta thu được kết quả như trong các hình bên dưới.



Hình 5. 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 6. So sánh biên tìm được với biên chuẩn cho tín hiệu mẫu “lab-female”.



Hình 7. So sánh biên tìm được với biên chuẩn cho tín hiệu mẫu “studio-male”.



Hình 8. So sánh biên tìm được với biên chuẩn cho tín hiệu mẫu “studio-female”.

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



Hình 9. 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).

Các tham số ảnh hưởng đến sai số của thuật toán: Mức độ chênh lệch về năng lượng giữa tiếng ồn từ môi trường và tiếng nói trong tín hiệu, và ngưỡng năng lượng chuẩn E0 được chọn làm mốc. Cụ thể, sai số của thuật toán càng thấp khi mức độ chênh lệch năng lượng giữa tiếng ồn môi trường và tiếng nói càng cao (như khi áp dụng thuật toán cho tín hiệu “lab-female”), hoặc khi ngưỡng năng lượng chuẩn E0 có giá trị càng nhỏ sao cho E0 lớn hơn năng lượng tối đa của tiếng ồn môi trường.

## So sánh với thuật toán không thực hiện chia tiếng nói-khoảng lặng

# 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