

Full name: Nguyen Nhat Hoang - 20520516

Team: 7 – Uranus

Other team members: Nguyen Minh Thang – 21521433

Title: Homework 3

Question: Speech Analysis Techniques?

1. Fast Fourier Transform (FFT)

Giới thiệu phương pháp: FFT là thuật toán hiệu quả để tính toán nhanh phép biến đổi Fourier rời rạc (Discrete Fourier Transform - DFT), dùng để chuyển đổi tín hiệu từ miền thời gian sang miền tần số.

Giải thích cơ chế: FFT phân rã một tín hiệu thời gian phức tạp thành tổng các tín hiệu sin và cosin ở các tần số khác nhau, giúp phát hiện cấu trúc điều hòa (harmonic structure) của tín hiệu.

Phân tích lý thuyết: FFT dựa trên cơ sở lý thuyết biến đổi Fourier, trong đó bất kỳ tín hiệu tuần hoàn nào cũng có thể biểu diễn dưới dạng chuỗi các sóng sin và cosin với các tần số riêng biệt.

Ưu điểm:

- Tính toán nhanh và hiệu quả.
- Rõ ràng về mặt lý thuyết, dễ hiểu, dễ triển khai.
- Phổ biến, nhiều công cụ hỗ trợ sẵn.

Nhược điểm:

- Kết quả nhạy cảm với nhiễu.
- Không phân biệt rõ ràng cấu trúc bao phổ và formant.

Ứng dụng thực tiễn:

- Phân tích phổ tín hiệu âm thanh và âm nhạc.
- Nhận diện tần số cơ bản trong tín hiệu giọng nói.

...

2. Cepstral Analysis

Giới thiệu phương pháp: Cepstral Analysis là kỹ thuật dùng để tách cấu trúc bao phổ (spectral envelope) ra khỏi tín hiệu âm thanh, giúp phân biệt rõ đặc điểm cộng hưởng của giọng nói.

Giải thích cơ chế: Cepstral Analysis thực hiện biến đổi Fourier trên phổ logarit của tín hiệu để thu được "cepstrum", một miền cho phép tách biệt thông tin nguồn và bộ lọc trong tín hiệu giọng nói.

Phân tích lý thuyết: Cepstral dựa trên khái niệm "homomorphic signal processing" (xử lý tín hiệu đồng hình), trong đó phổ tín hiệu được coi như kết quả của phép nhân giữa tín hiệu nguồn và đáp ứng hệ thống (vocal tract).

Ưu điểm:

- Tách rõ cấu trúc bao phổ.
- Giảm thiểu sự ảnh hưởng của các thành phần nhiễu và cộng hưởng không mong muốn.

Nhược điểm:

- Khá phức tạp về mặt lý thuyết.
- Yêu cầu nhiều tính toán hơn FFT cơ bản.

Ứng dụng thực tiễn:

- Nhận dạng và tổng hợp giọng nói.
- Phân tích tiếng nói để chẩn đoán bệnh lý giọng nói.

...

3. Linear Predictive Coding (LPC)

Giới thiệu phương pháp: LPC là kỹ thuật dùng mô hình tuyến tính để dự đoán mẫu tín hiệu hiện tại dựa vào các mẫu trước đó, được sử dụng rộng rãi trong phân tích formant và nén tín hiệu giọng nói.

Giải thích cơ chế: LPC mô hình hóa giọng nói như là đầu ra của một bộ lọc tuyến tính được kích thích bởi tín hiệu nguồn, qua đó xác định các tham số bộ lọc để biểu diễn tín hiệu một cách hiệu quả.

Phân tích lý thuyết: Phương pháp LPC dựa trên việc giảm tối đa lỗi dự đoán (prediction error) giữa mẫu tín hiệu thực tế và mẫu dự đoán từ các mẫu trước đó bằng cách giải một hệ phương trình tuyến tính.

Ưu điểm:

- Hiệu quả cao trong việc biểu diễn và nén tín hiệu giọng nói.
- Xác định rõ vị trí các formant (đỉnh phổ).

Nhược điểm:

- Nhạy cảm với nhiễu, đặc biệt trong môi trường ồn ào.
- Độ chính xác phụ thuộc vào lựa chọn bậc mô hình tuyến tính.

Ứng dụng thực tiễn:

- Nén dữ liệu trong truyền thông giọng nói.
- Nhận dạng và phân loại âm thanh, âm vị học trong xử lý tiếng nói.