A white number on a grey background

Description automatically generated

**PHÂN TÍCH TÍN HIỆU TIẾNG NÓI**

**VÀ ỨNG DỤNG XỬ LÝ TIẾNG NÓI**

**Họ và tên SV báo cáo 1: MSSV:**

**Họ và tên SV báo cáo 2: MSSV:**

**Họ và tên SV báo cáo 3: MSSV:**

**Nhóm lớp Tiểu nhóm: Ngày thí nghiệm:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Điểm đánh giá** | | | | **CBGD nhận xét và ký tên** |
| **Chuẩn bị lý thuyết** | **Báo cáo và kết quả TN** | **Kiểm tra** | **Kết quả** |  |

# **CHUẨN BỊ THÍ NGHIỆM**

* 1. Bên cạnh nội dung tiếng nói (từ được phát âm), hãy liệt kê những thông tin khác có thể biết được khi người nghe nghe một đoạn tín hiệu tiếng nói.

* 1. Khi nghe một đoạn tiếng nói, người nghe có thể biết được tiếng nói này là của nữ hay nam không? Giải thích.

* 1. Tín hiệu tiếng nói khi được truyền trên kênh truyền điện thoại, băng thông của nó sẽ được giới hạn trong khoảng tần số nào? Cho biết lý do của việc giới hạn băng thông.

* 1. Việc thu tiếng nói trong Matlab có thể được thực hiện bằng lệnh y = wavrecord(N, Fs, DataType). Hãy cho biết ý nghĩa các thông số y, N, Fs, DataType? Vì sao cần phải xác định 3 thông số này trong quá trình thu tiếng nói.

* 1. Việc phát tiếng nói trong Matlab có thể được thực hiện bằng lệnh wavplay(y, Fs). Fs là tần số lấy mẫu của y. Hãy cho biết nếu nhập tần số lấy mẫu không đúng thì kết quả tiếng nói được phát âm sẽ như thế nào?

* 1. Tín hiệu tiếng nói y có tổng số 1000 mẫu. Biết rằng y được thu tại tốc độ lấy mẫu Fs=16000Hz. Cho biết tín hiệu y được thu trong bao nhiêu mili giây?...........................................................
  2. Tín hiệu tiếng nói y trong câu 6 được biết đổi Fourier rời rạc (Discrete Fourier Transform) tại N=1024 điểm. Biết rằng độ rộng giữa hai vạch phổ kế tiếp nhau là Δf. Tính Δf.

Δf = ……………….Hz

* 1. Vì sao việc triệt nhiễu trong tín hiệu tiếng nói là cần thiết? Cho biết một số cách có thể được dùng để đánh giá của hiệu quả của quá trình triệt nhiễu trong tiếng nói

* 1. Dựa vào đặc trưng nào để nhận biết đoạn tín hiệu thực sự có tiếng nói trong toàn bộ đoạn tín hiệu ghi âm được. Giải thích ngắn gọn.

# **TIẾN HÀNH THÍ NGHIỆM**

**2.1 Ước lượng băng thông và khảo sát sự phân bố năng lượng**

1. Sử dụng hàm wavread trong Matlab để đọc file tiếng nói ‘CleanSpeech.wav’, gán tín hiệu tiếng nói trong file này là biến sp. Xác định tần số lấy mẫu fs của tín hiệu tiếng nói lưu trong file này.

fs= …………………mẫu/giây

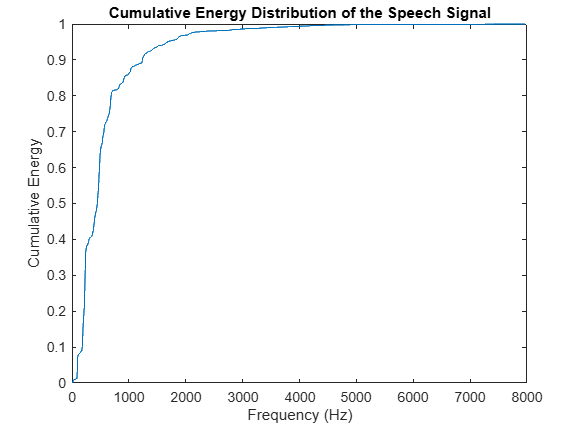
1. Biết rằng đoạn tiếng nói trên được thu trong khoảng thời gian t giây. Từ giá trị fs tìm được ở trên, hãy xác định t.

t= …………………giây

1. Sử dụng lệnh wavplay, phát và nghe lại đoạn tiếng nói trên tại ba tần số lấy mẫu khác nhau là f1=0.5fs, f2=fs, f3=2fs. Cho nhận xét về đoạn âm thanh nghe được trong mỗi trường hợp.

1. Sử dụng Matlab, vẽ dạng sóng tín hiệu tiếng nói sp trên với trục ngang biểu diễn theo đơn vị giây. Tính và vẽ phổ biên độ của sp trong khoảng tần số từ 0 đến fs/2 (sử dụng công cụ fft). Lưu ý: trục ngang của hình vẽ phổ biên độ phải được biểu diễn theo đơn vị Hz.

|  |
| --- |
| Dạng sóng tiếng nói |
| Phổ biên độ của tiếng nói |

1. Dựa trên hình vẽ phổ biên độ này, hãy ước lượng băng thông của tiếng nói (độ rộng của đoạn tần số tính từ 0Hz, chứa khoảng hơn 95% năng lượng của tín hiệu tiếng nói).

Băng thông của tín hiệu tiếng nói là: BW = Hz

1. Dựa trên kết quả tính fft trong câu 6.1, hãy tính phần trăm năng lượng của tiếng nói phân bố trên các vùng tần số khác nhau có cùng độ rộng là 1kHz và hoàn thành Bảng 1.

Bảng 1 : Phần trăm năng lượng tiếng nói trên các băng tần khác nhau

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Băng Tần | | | | | | | |
| Băng tần | 0-1kHz | 1-2kHz | 2-3kHz | 3-4kHz | 4-5kHz | 5-6kHz | 6-7kHz | 7-8kHz |
| % năng lượng |  |  |  |  |  |  |  |  |

Cho biết vùng tần số nào chứa nhiều năng lượng tiếng nói nhất?.......................

1. Dựa vào kết quả trong Bảng 1 hãy giải thích vì sao tín hiệu tiếng nói truyền trên đường truyền điện thoại bị giới hạn tần số dưới 3.5kHz?

**2.2 Xác định tần số lấy mẫu tối thiểu**

1. Từ kết quả trong câu 6.1, hãy ước lượng tần số lấy mẫu nhỏ nhất fsmin của tiếng nói để tín hiệu sau khi lấy mẫu không bị aliasing (chồng phổ).

fsmin=………………….mẫu/giây

1. Tín hiệu tiếng nói trong câu 6.1 có thể được giảm tần số lấy mẫu để tạo nên tín hiệu có tần số lấy mẫu nhỏ hơn theo dùng lệnh

spDownSampling=sp(1:N:end);

Trong đó N = 2, 3, 4…. là tốc độ giảm tần số lấy mẫu. Việc giảm tần số lấy mẫu của tín hiệu được thực hiện bằng cách trong mỗi đoạn N mẫu tín hiệu gốc liên tiếp nhau thì giữ lại một mẫu và bỏ đi N-1 mẫu còn lại để tạo thành tín hiệu mới.

1. Hãy cho biết lợi ích của việc giảm tần số lấy mẫu?

1. Hãy tiến hành giảm tần số lấy mẫu của tiếng nói gốc trong câu 6.1. Ứng với mỗi giá trị N, thực hiện các bước sau và hoàn thành Bảng 2.
2. Tính và vẽ phổ biên độ của tín hiệu tiếng nói spDownSampling trong khoảng tần số từ [0 fs/(2N)] Hz.
3. Vẽ dạng sóng tín hiệu tiếng nói trong 4 mili giây đầu tiên dùng lệnh stem(spDownSampling)
4. Nghe tín hiệu tiếng nói sau khi giảm tốc độ lấy mẫu theo lệnh sau: wavplay(spDownSampling,fs/N)

|  |  |
| --- | --- |
| Phổ biên độ tiếng nói với N=1 | Dạng sóng tiếng nói với N=1 |
| Phổ biên độ tiếng nói với N=2 | Dạng sóng tiếng nói với N=2 |
| Phổ biên độ tiếng nói với N=3 | Dạng sóng tiếng nói với N=3 |
| Phổ biên độ tiếng nói với N=4 | Dạng sóng tiếng nói với N=4 |
| Phổ biên độ tiếng nói với N=8 | Dạng sóng tiếng nói với N=8 |

Bảng 2: Giảm tần số lấy mẫu

|  |  |  |  |
| --- | --- | --- | --- |
| **Tốc độ giảm tần số lấy mẫu** | **Tần số lấy mẫu mới (kHz)** | **Tổng số mẫu của spDownSampling** | **Mức độ aliasing** |
| N = 1 |  |  |  |
| N = 2 |  |  |  |
| N = 3 |  |  |  |
| N = 4 |  |  |  |
| N = 8 |  |  |  |

1. Trong các giá trị N đã chọn trong Bảng 2, bắt đầu từ giá trị nào thì hiện tượng aliasing xảy ra? Giải thích nguyên nhân.

**2.3 Triệt nhiễu nâng cao chất lượng tiếng nói**

Tín hiệu tiếng nói trong câu 6.1 bị ảnh hưởng của nhiễu cộng là một tín hiệu SIN tần số cao. Tiếng nói bị ảnh hưởng bởi nhiễu được lưu trong file ‘NoisySpeech.wav’.

Sử dụng hàm wavread trong Matlab để đọc file tiếng nói bị tác động bỡi nhiễu, gán tín hiệu tiếng nói trong file này là biến spNoisy. Sinh viên được yêu cầu ứng dụng bộ lọc để triệt nhiễu nhằm mục đích nâng cao chất lượng tiếng nói.

a. Tính và vẽ phổ biên độ của tín hiệu spNoisy trong khoảng tần số từ 0 đến fs/2. Lưu ý: trục ngang của hình vẽ phổ biên độ phải được biểu diễn theo đơn vị Hz. So

sánh với phổ của tín hiệu tiếng nói chưa bị tác động của nhiễu (sp) trong câu 6.1

để xác định tần số của tín hiệu nhiễu fNoise.

|  |
| --- |
| Phổ biên độ của tín hiệu tiếng nói bị nhiễu |

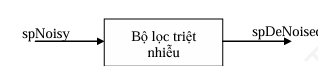
fNoise=……………..Hz

Tín hiệu spNoisy được truyền qua bộ lọc như Hình 8 nhằm triệt nhiễu để nâng cao chất lượng tiếng nói. Hãy cho biết để khử nhiễu hiệu quả, bộ lọc triệt nhiễu cần có đặc tính gì?

Loại bộ lọc (thông cao, thấp, dải, chắn dải)?.......................................................

Tần số cắt của bộ lọc:

Giải thích



**Hình 48. Bộ lọc triệt nhiễu**

b. Sử dụng 03 bộ lọc thông thấp Chebyshev loại 1 khác nhau để triệt nhiễu trong tín hiệu spNoisy. Cả ba bộ lọc này đều có tần số cắt fc=4kHz và độ gợn dải thông Rp=0.2dB. Bậc của bộ lọc thay đổi như trong Bảng 3.

Sử dụng lệnh cheby1 trong Matlab để xác định các hệ số B và A của các bộ lọc 1, 2 và 3. Sau đó dùng lệnh filter để thực hiện việc lọc nhiễu trong tín hiệu spNoisy. Hãy thực hiện các cách sau để đánh giá hiệu quả của quá trình triệt nhiễu và hoàn thành Bảng 3.

1. Nghe tín hiệu trước và sau khi triệt nhiễu.
2. Vẽ và so sánh dạng sóng của tín hiệu trước và sau khi triệt nhiễu.
3. Vẽ và so sánh phổ biên độ của tín hiệu tiếng nói trước và sau khi triệt nhiễu.
4. Tính thông số Tỷ số tín hiệu trên nhiễu được ký hiệu là SNR (Signal to Noise Ratio) theo công thức sau:



Trong đó: *SpeechPower* là công suất tín hiệu tiếng nói chưa bị ảnh hưởng bởi nhiễu (sp). *NoisePower* là công suất của nhiễu. Nhiễu được xác định là hiệu số của tín hiệu sau khi xử lý và tín hiệu trước khi xử lý (sp). Dễ hiểu là SNR càng lớn thì chất lượng tín hiệu tiếng nói càng tốt. Lưu ý: Công suất tín hiệu có thể được tính là tổng bình phương tất cá các mẫu của tín hiệu đó. . Trong đó s[k] là biên độ của mẫu tín hiệu thứ k trong tổng số L mẫu của tín hiệu s.

Xác định SNR của tín hiệu tiếng nói trước khi triệt nhiễu spNoisy

= ……………….dB

Bảng 3: So sánh hiệu quả các bộ lọc triệt nhiễu

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Bậc bộ lọc** | **SNR sau khi triệt nhiễu (dB)** | **Chất lượng tiếng nói sau khi triệt nhiễu** |
| Bộ lọc 1 | N = 2 |  |  |
| Bộ lọc 2 | N = 4 |  |  |
| Bộ lọc 3 | N = 8 |  |  |

c. Hãy vẽ dạng sóng và phổ biên độ của sp, spNoisy và spDeNoised hình vẽ sau.

|  |  |
| --- | --- |
| Dạng sóng tiếng nói gốc (chưa bị nhiễu) | Phổ biên độ tiếng nói gốc |
| Dạng sóng tiếng nói bị nhiễu | Phổ biên độ tiếng nói bị nhiễu |
| Dạng sóng tiếng nói sau triệt nhiễu dùng Bộ lọc 1 | Phổ biên độ tiếng nói sau triệt nhiễu dung Bộ lọc 1 |
| Dạng sóng tiếng nói sau triệt nhiễu dùng Bộ lọc 2 | Phổ biên độ tiếng nói sau triệt nhiễu dung Bộ lọc 2 |
| Dạng sóng tiếng nói sau triệt nhiễu dùng Bộ lọc 3 | Phổ biên độ tiếng nói sau triệt nhiễu dung Bộ lọc 3 |

d. Cho nhận xét và giải thích về hiệu quả triệt nhiễu của các bộ lọc 1, 2, 3.

**2.4 Tách tiếng nói ra khỏi đoạn tín hiệu thu được**

a. Hãy phân đoạn đoạn tín hiệu tiếng nói sp trong mục 6.1 theo hướng dẫn trong Mục 4 và hãy cho biết tổng số đoạn tiếng nói nhận được là ……………………..đoạn.

b. Viết một đoạn chương trình tính Năng lượng từng đoạn tiếng nói và vẽ thông số này trên cùng màn hình với dạng sóng tiếng nói. Lưu ý: Hình vẽ dạng sóng tiếng nói và Năng lượng phải được sắp xếp thẳng hàng. Trục ngang hình vẽ dạng sóng tiếng nói biểu diễn theo chỉ số mẫu, trục ngang của hình vẽ năng lượng biểu diễn theo chỉ số đoạn.

|  |
| --- |
| Dạng sóng tiếng nói |
| Năng lượng tiếng nói |

Từ kết quả nhận được hãy cho biết sự thay đổi của Năng lượng ứng với các vùng tiếng nói khác nhau

c. Tiếng nói sau khi thu được sẽ bao gồm hai phần, phần có tín hiệu tiếng nói và phần không có tín hiệu tiếng nói (tương ứng với khoảng thời gian yên lặng trong quá trình phát âm). Việc trích đặc trưng trong các ứng dụng xử lý tiếng nói được thực hiện trong đoạn có tiếng nói. Vì vậy, việc nhận biết và tách đoạn tiếng nói ra khỏi đoạn tín hiệu thu được là cần thiết. Quá trình này được gọi là phát hiện tiếng nói (Voice Activity Detection VAD).

Từ sự quan sát về sự thay đổi của Năng lượng trong phần 6.4b, hãy đề nghị một giải thuật tách đoạn tín hiệu tiếng nói ra khỏi đoạn tín hiệu thu được sử dụng Năng lượng. Viết đoạn chương trình thực hiện giải thuật này bằng Matlab.

Giải thuật tách đoạn tín hiệu tiếng nói ra khỏi đoạn tín hiệu ghi âm được:

Đoạn chương trình thực hiện VAD bằng Matlab

Kết quả thực hiện

|  |
| --- |
| Dạng sóng của toàn bộ tín hiệu ghi âm |
| Dạng song tiếng nói tách được |

**2.5 Nhận dạng giới tính**

Lệnh pitch\_rapt được dùng để trích F0 trên tất cả các đoạn tiếng nói (chiều dài 25ms và chồng lấn 15ms giữa hai đoạn kế tiếp). Lệnh này được dùng như sau: [mf0]=pitch\_rapt(s,fs) trong đó mf0 là vector chứa giá trị F0 trên các đoạn tiếng nói của trong tín hiệu s, fs là tần số lấy mẫu của s.

a. Dùng lệnh pitch\_rapt hãy trích và vẽ F0 của 02 file tiếng nói, 1 file tiếng nói của nữ (‘Female.wav’), 1 file tiếng nói của nam (‘Male.wav’). Tính F0 trung bình của tiếng nói người nữ Mean\_F0nữ và người nam Mean\_F0nam. Lưu ý rằng F0 của những đoạn không có tiếng nói không xác định được và pitch\_rapt sẽ gán giá trị F0=0 cho những đoạn này. F0 trung bình chỉ tính cho những đoạn có F0>0.

|  |
| --- |
| Hình vẽ F0 của nữ, file ‘Female.wav’ |
| Hình vẽ F0 của nam, file ‘Male.wav’ |

Nhận xét về khoảng dao động và giá trị trung bình của F0 người nữ và người nam.

b. Sinh viên được cung cấp 10 file tiếng nói của 10 người nói (speaker) lưu dưới dạng file .wav, thực hiện các bước sau và hoàn thành Bảng 4.

i. Hãy nghe 10 file tiếng nói này và sau khi nghe xác định giới tính người nói là nam hay nữ. Điền kết quả vào hàng (1).

ii. Sử dụng lệnh pitch\_rapt hãy xác định viết một đoạn chương trình để tính F0 trung bình của từng file tiếng nói. Từ đó so sánh với ngưỡng 150Hz để xác định giới tính người phát âm. Điền kết quả vào hàng (3) và (4).

Bảng 4: Kết quả nhận dạng giới tính

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Speaker | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1. Nghe |  |  |  |  |  |  |  |  |  |  |
| 1. Kết luận |  |  |  |  |  |  |  |  |  |  |
| 1. F0 (Hz) |  |  |  |  |  |  |  |  |  |  |
| 1. Chương trình |  |  |  |  |  |  |  |  |  |  |
| 1. Kết luận |  |  |  |  |  |  |  |  |  |  |

Đối chiếu kết quả nhận được với Giáo viên hướng dẫn để biết kết quả nhận dạng theo mỗi phương pháp là đúng hay sai và ghi vào ô tương ứng trong hàng (2)&(5) Bảng 4. Độ chính xác của quá trình nhận dạng theo phần trăm được tính như sau:

*Độ chính xác (%) = (Số lần nhận dạng đúng/Tổng số lần nhận dạng)\*100.*

c. Tính độ chính xác việc nhận dạng giới tính 10 file cho sẵn bằng cách lắng nghe ………….%

Tính độ chính xác việc nhận dạng giới tính bằng chương trình Matlab………….%

d. Sử dụng wavread, đọc và thu vào máy tính dòng chữ: “Bộ môn Viễn thông, Khoa Điện, Trường Đại học Bách khoa Thành phố Hồ Chí Minh”. Tốc độ lấy mẫu fs=16kHz, datatype=’double’, gán tín hiệu thu được vào biến s trong Matlab. Dùng pitch\_rapt hãy tính F0 trung bình của chuỗi tiếng nói này. Hãy lặp lại việc này 5 lần và ghi nhận kết quả vào phần sau:

Giới tính của bạn…………………

Lần 1: F01=………………Hz

Lần 2: F02=………………Hz

Lần 3: F03=………………Hz

Lần 4: F04=………………Hz

Lần 5: F05=………………Hz

F0TrungBình=………………Hz

**2.6 Phần mở rộng**

Lưu ý: Phần này không bắt buộc sinh viên thực hiện. Nếu hoàn thành đúng phần này, sinh viên được cộng 2 điểm vào bài báo cáo.

Sử dụng Matlab, hãy viết một đoạn chương trình để khi người đọc đọc vào máy tính một đoạn tiếng nói (chậm và rời rạc), chương trình sẽ thực hiện một số yêu cầu sau:

a. Xác định tổng số từ trong đoạn tiếng nói thu vào.

b. Tách tiếng nói của mỗi từ riêng biệt và lưu vào các file ‘.wav’. Tên file sẽ là số thứ tự của từ tương ứng trong đoạn tiếng nói. Ví dụ ‘1.wav’, ‘2.wav’…

c. Tính chiều dài của tiếng nói mỗi từ theo đơn vị mili giây và chiều dài trung bình của tất cả các từ đọc vào