SV cần làm các việc và báo cáo (nói trên slide và viết bằng file MS-Word) theo trình tự sau:

**1. Thu thập tín hiệu (signal acquisition):**

Nhóm SV dùng 06 files tín hiệu đã được thu âm sẵn trong folder “TinHieuHuanLuyen” để thử nghiệm và viết báo cáo.

**2. Phân tích tín hiệu thủ công (manual signal analysis):**

Dùng phần mềm WaveSurfer (PowerPlot pane).

**3. Phân tích tín hiệu tự động (automatic signal analysis):**

Tìm hiểu lý thuyết và cài đặt 02 thuật toán: phân đoạn tín hiệu thu âm thành tiếng nói và khoảng lặng [1], phân đoạn khung tiếng nói thành khung hữu thanh và khung vô thanh [2].

Chú ý: Độ dài tối thiểu của 1 khoảng lặng là 300 ms (dùng điều kiện này để loại bỏ các khoảng lặng “ảo” có chiều dài quá ngắn).

**Các TLTK:**

[1] CS425 Audio and Speech Processing\_Hodgkinson\_2012:

2.1 Energy-based Speech/Silence discrimination (phân biệt tiếng nói & khoảng lặng)

2.2 Voiced/Unvoiced discrimination (phân biệt tiếng nói hữu thanh/vô thanh)

[2] Digital Speech\_Kondoz\_2004:

6.3 Voiced–Unvoiced Classification

[3] Luận văn cao học “Luận văn\_Tran Van Tam\_2019.pdf”: Chương 1 Tổng quan về xử lý tiếng nói

[4] Chuẩn hoá dữ liệu: <https://www.statisticshowto.com/probability-and-statistics/normal-distributions/normalized-data-normalization/> (áp dụng để chuẩn hoá hàm năng lượng ngắn hạn hoặc hàm logE để dễ tìm ngưỡng chung cho nhiều tín hiệu)

**4. Kiểm tra tiến độ (Buổi học Tuần 11, ngày 11-16/10/2021):**

* Mỗi SV cài đặt và demo mỗi thuật toán nêu trên.
* SV xuất hình vẽ kết quả hàm STE/MA ngắn hạn (hoặc logE/logMA).
* Chú ý khảo sát ảnh hưởng các tham số của thuật toán lên kết quả, ví dụ:
  + Các giá trị ngưỡng nhằm phân đoạn tín hiệu.
  + Mức nhiễu nền của môi trường thu âm.
* Các tín hiệu huấn luyện (training data) đóng vai trò giúp xác định bộ tham số tối ưu của thuật toán:

Muốn xác định ngưỡng V/U tin cậy dựa trên dữ liệu, SV cần khảo sát tất cả các voiced frames trong dữ liệu huấn luyện (timestamps lấy từ \*.lab) và thống kê xem normalized STE có phân bố dữ liệu ntn (meanV=?, stdV=?). Làm tương tự với các unvoiced frames để có (meanU=?, stdU=?). Dựa vào 2 phân bố dữ liệu này để tìm ra ngưỡng STE giúp tách biệt 2 phân bố này (giả sử dữ liệu theo phân bố chuẩn).

Áp dụng cách thức tương tự để tìm ngưỡng ZCR giúp tách biệt 2 phân bố của V và U.

Nếu ko dựa trên thống kê chỉ có cách làm mò mẫm (tự đưa ra 1 giá trị ngưỡng nào đó) -> ko thuyết phục và khó đúng với nhiều tín hiệu.

Để xác định ngưỡng speech/silence cũng cần làm thống kê dựa trên dữ liệu huấn luyện như vậy, trừ khi SV áp dụng các thuật toán trong CS425.

* Hiệu suất của thuật toán đề xuất sẽ được kiểm chứng thực sự trên tập tín hiệu kiểm thử (test data) (GV sẽ upload sau 02 tuần nữa)
* SV nào ko demo được chương trình do chính mình viết trong buổi kiểm tra tiến độ này thì bị loại khỏi nhóm và nhận 0 điểm thi giữa kỳ. Để tiết kiệm thời gian, mỗi SV chạy script cài đặt riêng task của mình, duyệt qua 4 file tín hiệu huấn luyện và xuất ra 4 figure thể hiện input & output (mỗi figure cho 1 file tín hiệu) trong 01 lần chạy CT duy nhất để GV kiểm tra kết quả.
* Kết quả chuẩn của mỗi file tín hiệu \*.wav được chứa trong file \*.lab tương ứng. Định dạng của file .lab được mô tả trong file README. SV dùng dữ liệu chuẩn này để đưa ra đánh giá về độ chính xác của thuật toán như sau: vẽ các biên do thuật toán xuất ra (đường dọc màu xanh) và các biên chuẩn (đường dọc màu đỏ) trên đồ thị tín hiệu.