

Phát biểu bài toán: Tìm tần số cơ bản (F0) của tín hiệu chỉ chứa nguyên âm và khoảng lặng.

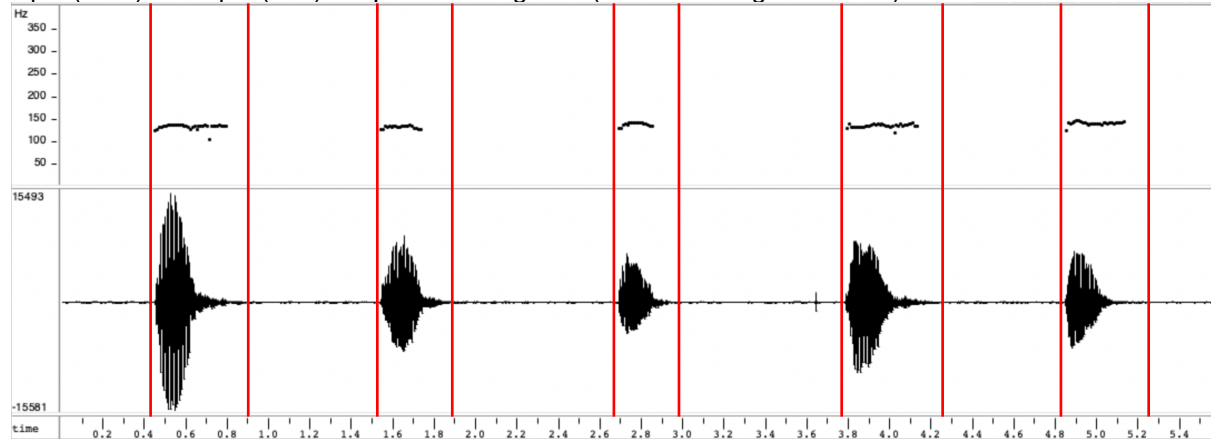
Input: Tín hiệu chỉ chứa nguyên âm và khoảng lặng (file .wav).

Output: Các biên thời gian phân biệt nguyên âm/khoảng lặng và đường F0 của cả tín hiệu.

Yêu cầu cài đặt 2 thuật toán tuần tự:

- Thuật toán 1: tìm biên thời gian giữa nguyên âm (vowel) và khoảng lặng (silence) dùng đặc trưng trên miền thời gian (STE, MA, ZCR,...) (3 điểm).
- Thuật toán 2: tìm F0 của một khung hữu thanh dựa trên phổ biên độ (magnitude spectrum) của khung đó (7 điểm).

Input (dưới) và output (trên) ví dụ của chương trình (các biên thời gian màu đỏ):



Chú ý:

- Tín hiệu của nguyên âm gồm **đa số** là các khung hữu thanh vì tín hiệu có dạng tuần hoàn (có 1 vài khung ở phần cuối nguyên âm có thể tuần hoàn yếu do lẫn với hơi thở → F0 ko tồn tại hoặc ko phát hiện được F0). Ngược lại, tín hiệu của khoảng lặng ko tồn tại F0 vì không tuần hoàn.
- Trong vd trên, tín hiệu gồm 5 nguyên âm và 6 khoảng lặng.
- Dải giá trị F0 của người trưởng thành: từ 70 đến 400 Hz (dùng dải này để thu hẹp phạm vi tìm kiếm F0).

Các TLTK:

- Luận văn cao học “Luận văn_Tran Van Tam_2019.pdf”: Chương 2 và 3.
- CS425 Audio and Speech Processing_Hodgkinson_2012: Chapter 2 & 3.
- textbook “Digital Speech_Kondoz_2004”: 6.2.2 Frequency-Domain PDAs (pitch determination algorithms).
- Các nguồn khác trên internet.

Các yêu cầu cài đặt:

- Độ dịch khung (frame shift – khoảng cách giữa 2 mẫu ở tâm của 2 khung liên tiếp) = 10 ms → tăng độ phân giải thời gian.
- Độ dài khung (frame length): 20-30 ms
- Các hàm thư viện được dùng: findpeaks(), fft(). Còn lại SV tự code.
- Xuất hình vẽ kết quả trung gian (đồ thị biểu diễn các đặc trưng thời gian/tần số của **tín hiệu/khung tín hiệu**) trên giao diện kết quả.
- Chú ý khảo sát ảnh hưởng các tham số của thuật toán lên kết quả, ví dụ:
 - o Ngưỡng xác định khung tín hiệu nguyên âm/khoảng lặng.
 - o Số điểm tính FFT (số điểm lấy mẫu trên miền tần số).
 - o Ngưỡng xác định mức độ tuần hoàn của khung để ra quyết định voiced/unvoiced (tuỳ thuật toán có dùng tham số này hay ko). Nếu dùng thì SV tự đưa ra giá trị ngưỡng phù hợp vì GV ko gán nhãn voiced/unvoiced trong file *.lab.
- Các tín hiệu huấn luyện (training data) đóng vai trò giúp xác định bộ tham số tối ưu của thuật toán.
- Hiệu suất của thuật toán đề xuất được kiểm chứng trên tập tín hiệu kiểm thử (test data) bằng cách thống kê sai số.
- Để tiết kiệm thời gian demo, SV cài đặt CT duyệt qua 4 file tín hiệu kiểm thử và xuất ra 4 figure thể hiện input & output & KQ trung gian (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 trên 2 phép đo:

- sai số tuyệt đối (tính bằng ms) giữa các biên thời gian tìm được và các biên chuẩn.
- sai số tương đối (%) giữa các F0_mean và F0_std tìm được và các F0_mean và F0_std chuẩn.

- Ngoài ra, kết quả đường F0 (pitch contour) của phần mềm WaveSurfer được dùng làm tham chiếu để so sánh với kết quả của thuật toán do SV cài đặt. SV đưa hình vẽ đường F0 của WaveSurfer lên SLIDE để so sánh.

Lịch hỏi/đáp về thi CK trên Team riêng cả mỗi nhóm (từ 19h30-22h00 mỗi ngày):

- 1912: 6/12
- 1914: 7/12
- 1910: 8/12
- 1913: 9/12
- 1911: 10/12
- 1915: 11/12