

PHÂN ĐOẠN TÍN HIỆU TIẾNG NÓI SỬ DỤNG SHORT-TIME ENERGY AND ZERO CROSSING RATE

GVHD: Ninh Khánh Duy
SVTH : Nguyễn Hải Đăng
Nguyễn Vũ Anh Duy
Trần Nguyễn Hồng Sơn
Nguyễn Đình Hân
Nhóm: 12
Nhóm HP: 17N10



I.ĐẶT VẤN ĐỀ

II.LÝ THUYẾT XỬ LÝ TÍN HIỆU VÀ TIẾNG NÓI

III.THUẬT TOÁN

IV.KẾT QUẢ THỰC NGHIỆM

V.KẾT LUẬN



I.ĐẶT VẤN ĐỀ

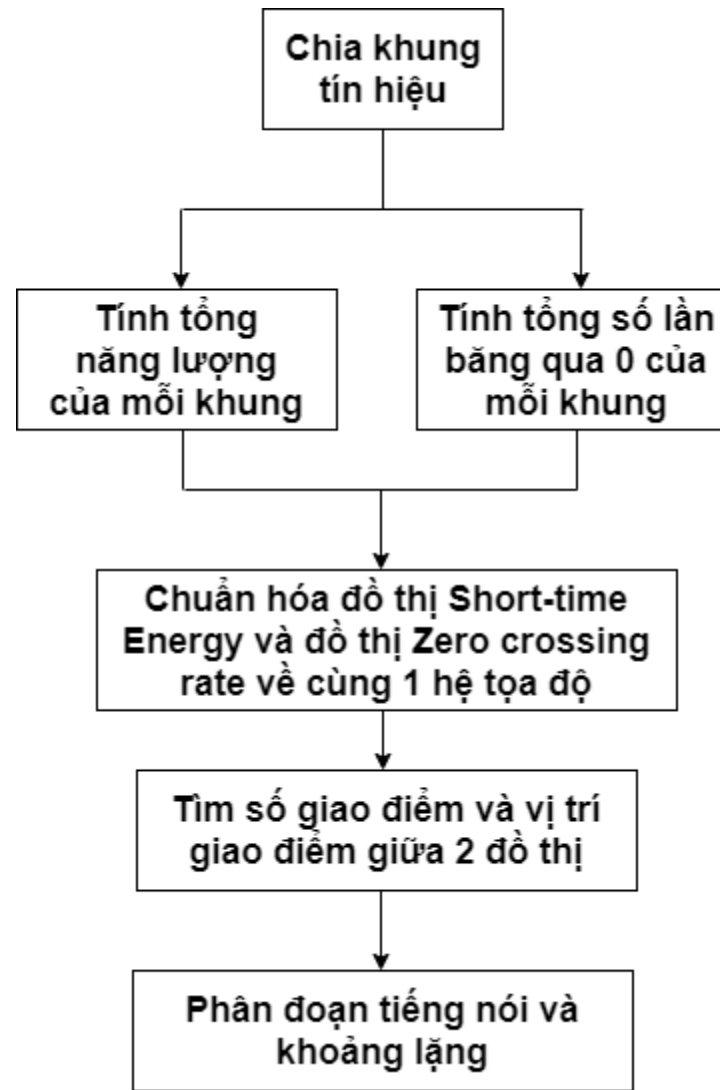
- Phân đoạn tín hiệu là gì?
- Tại sao phải phân đoạn tín hiệu?
- Phân đoạn cái gì?
- Phân đoạn như thế nào?

II. LÝ THUYẾT XỬ LÝ TÍN HIỆU VÀ TIẾNG NÓI

- Tiếng nói là phương tiện giao tiếp cơ bản của con người trong cuộc sống để trao đổi thông tin
- Xử lý tiếng nói là sự nghiên cứu dưới dạng tín hiệu , và tín hiệu tiếng nói thường thể hiện dạng số ,tức là được “số hóa”.Lĩnh vực xử lý tiếng nói rất rộng lớn nhưng trong phạm vi báo cáo này chỉ nghiên cứu phân đoạn tiếng nói

III. THUẬT TOÁN

1. Sơ đồ khối thuật toán



Hình 1. Sơ đồ thuật toán phân đoạn tín hiệu

2. Hàm năng lượng ngắn hạn (Short-time Energy)

$$E_m = \sum_{n=N_{1m}}^{N_{2m}} (x[n])^2$$

Trong đó

E_m : tổng năng lượng tại khung m

N_{1m}, N_{2m} : các chỉ số bắt đầu và kết thúc
của khung m

$x[n]$: biên độ tín hiệu tại vị trí n

3. Hàm ZCR ngắn hạn (Zero crossing rate)

$$ZCR_m = \text{số lần } \{ x[n] * x[n + 1] < 0 \}$$

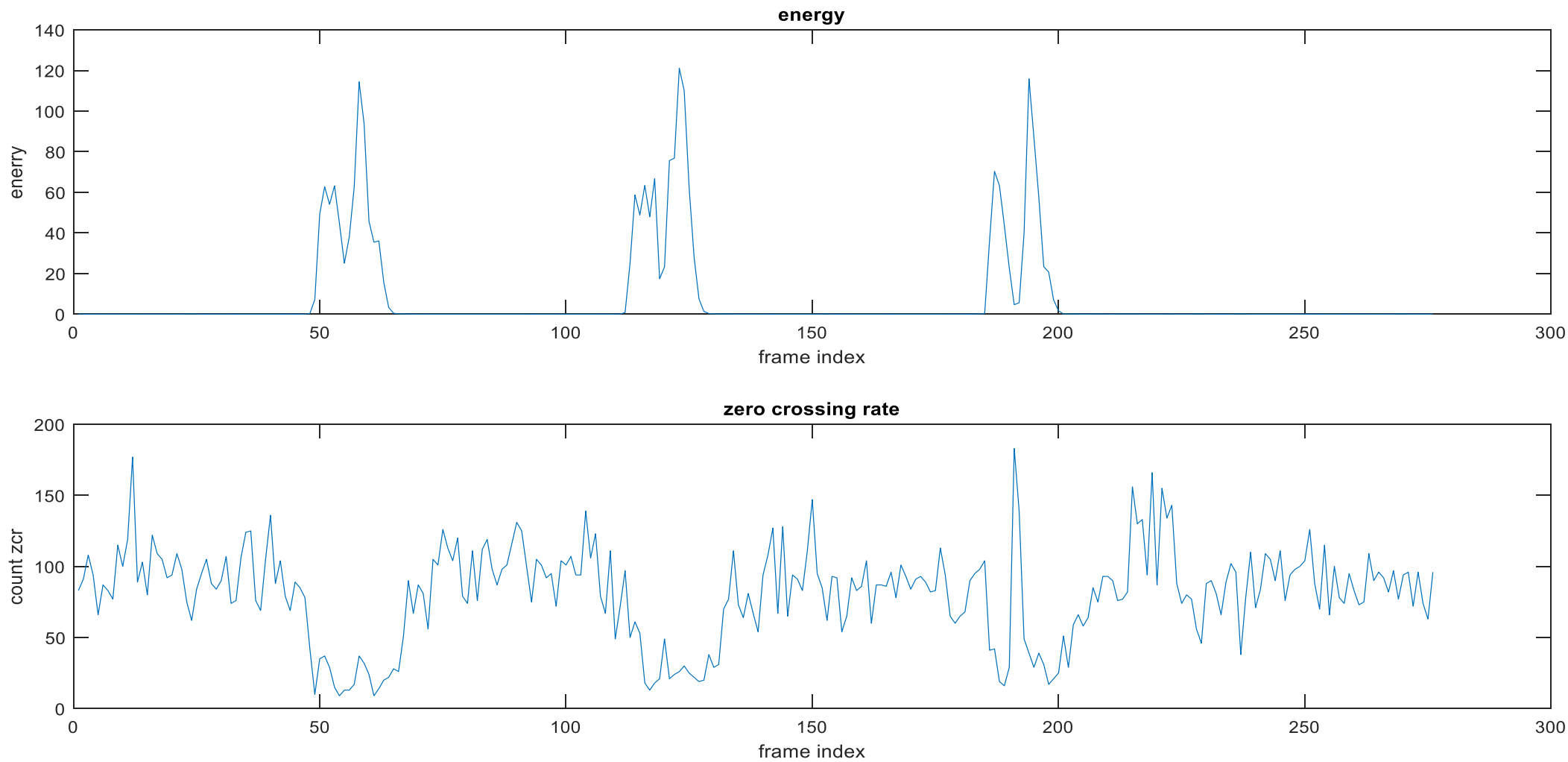
Với $n \in [N_{1m}, N_{2m}]$

Trong đó

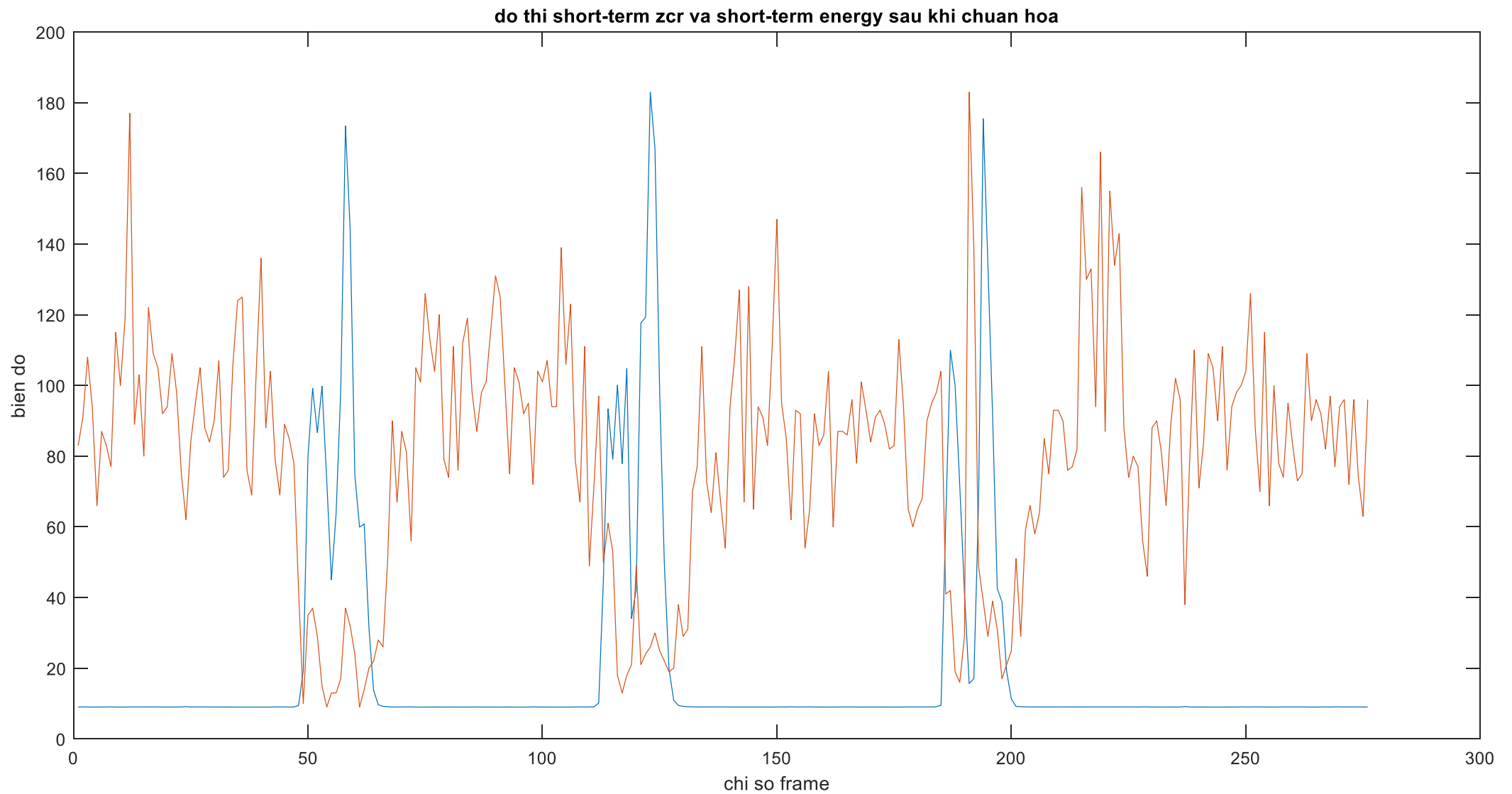
ZCR_m : tổng số lần biên độ tín hiệu băng qua 0
của khung m

$x[n]$: biên độ tín hiệu tại vị trí n

$x[n + 1]$: biên độ tín hiệu tại vị trí n + 1

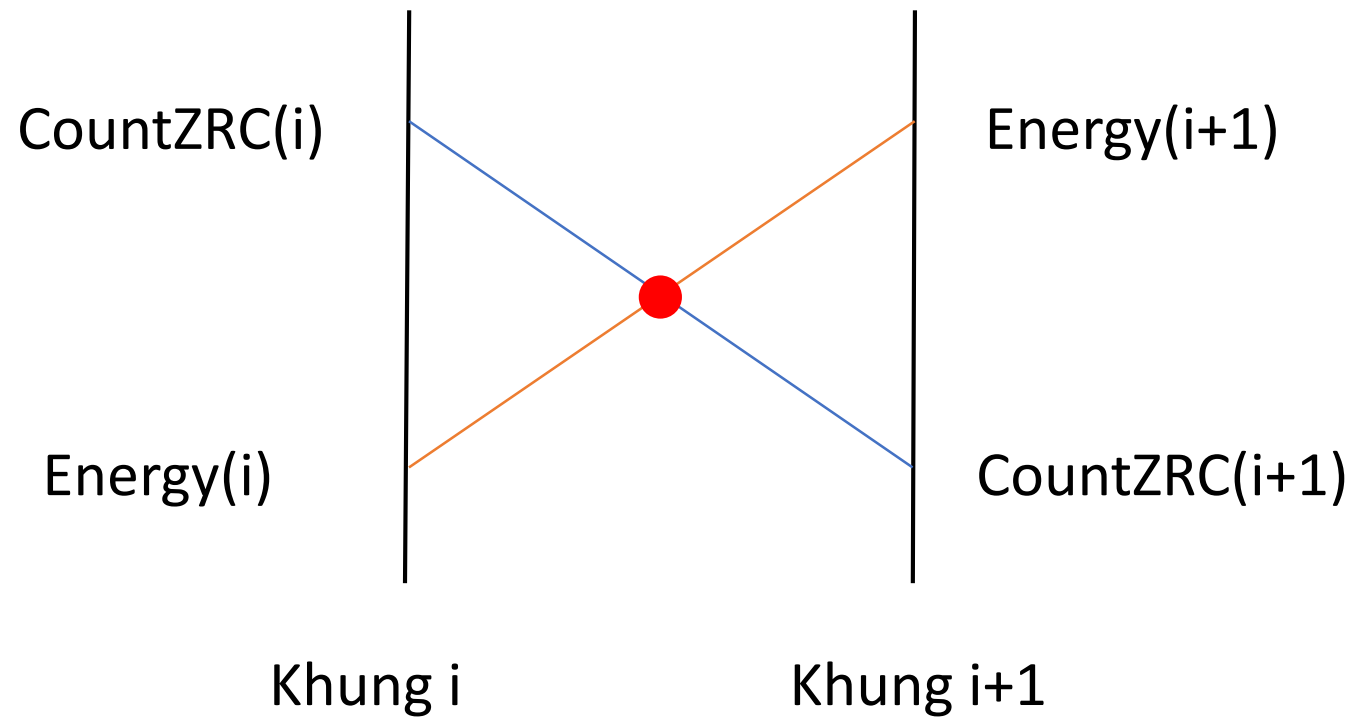


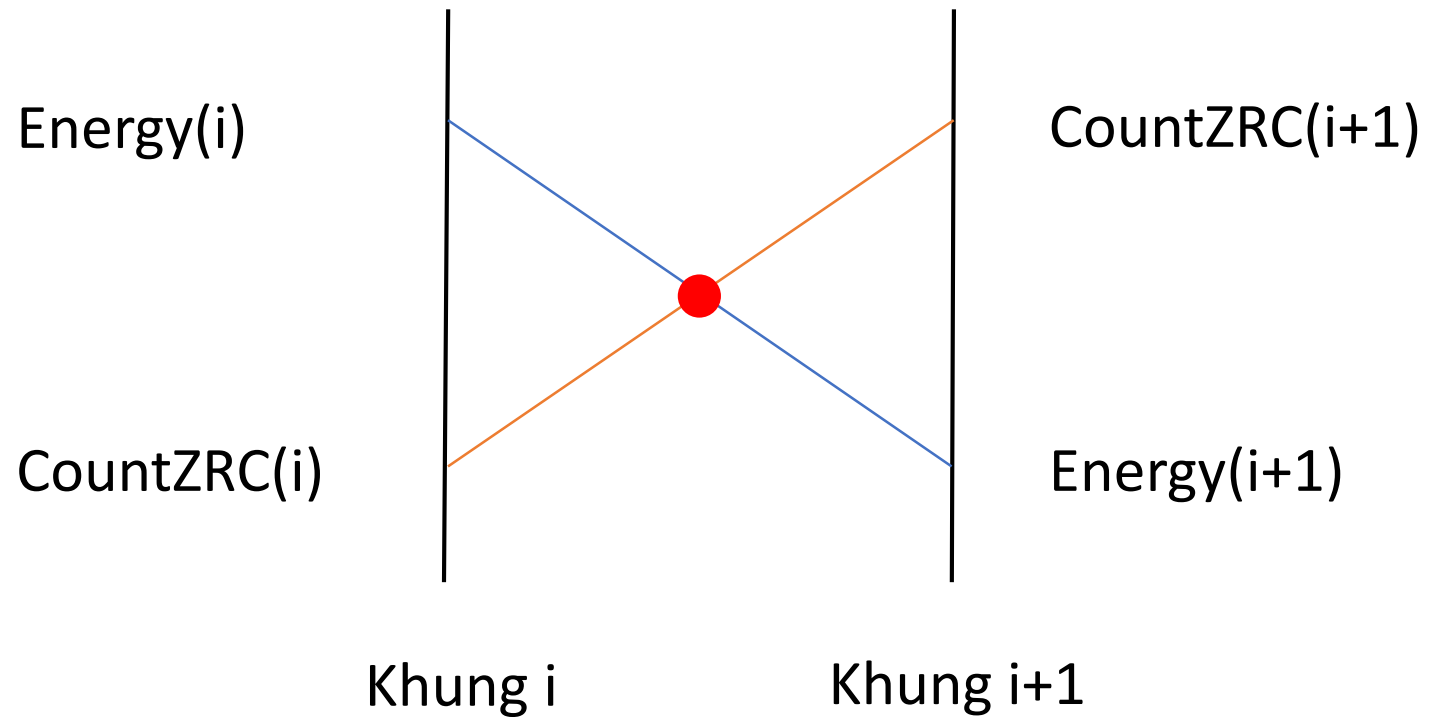
Hình 2. hình ảnh minh họa đồ thị short-time energy và zero-crossing rate khi chưa chuẩn hóa



Hình 3. hình ảnh minh họa đồ thị short-time energy và zero-crossing rate khi đã chuẩn hóa

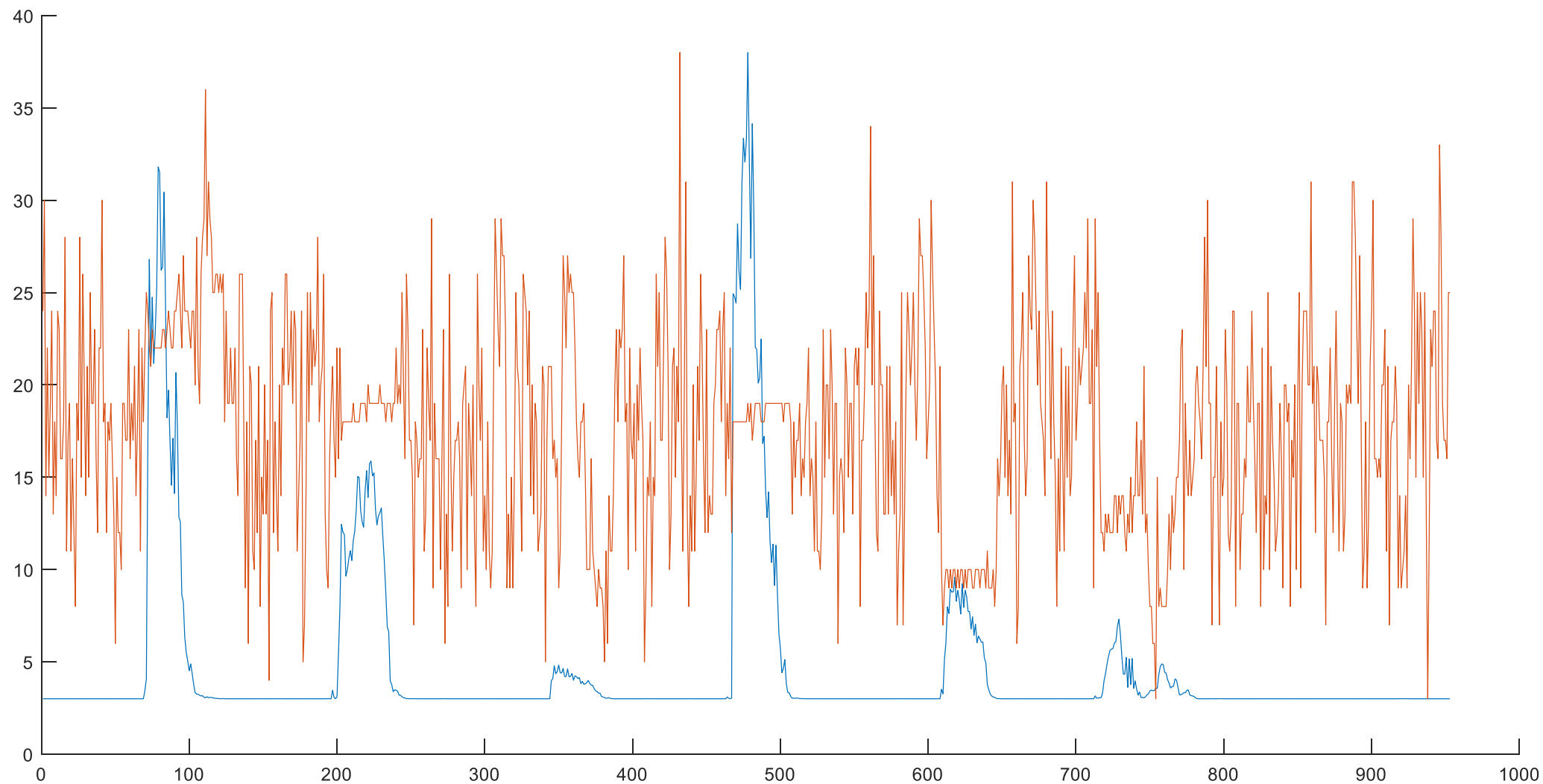
Các trường hợp giao điểm giữa các khung





4. Vấn đề của thuật toán

Sau khi chuẩn hóa 2 đồ thị Energy và ZCR về cùng 1 trục thì tiến hành tìm giao điểm. Lúc đó sẽ có thể có những đoạn tín hiệu không có điểm giao làm cho sau khi xuất tín hiệu đã được phân đoạn sẽ bị mất một vài âm thanh.



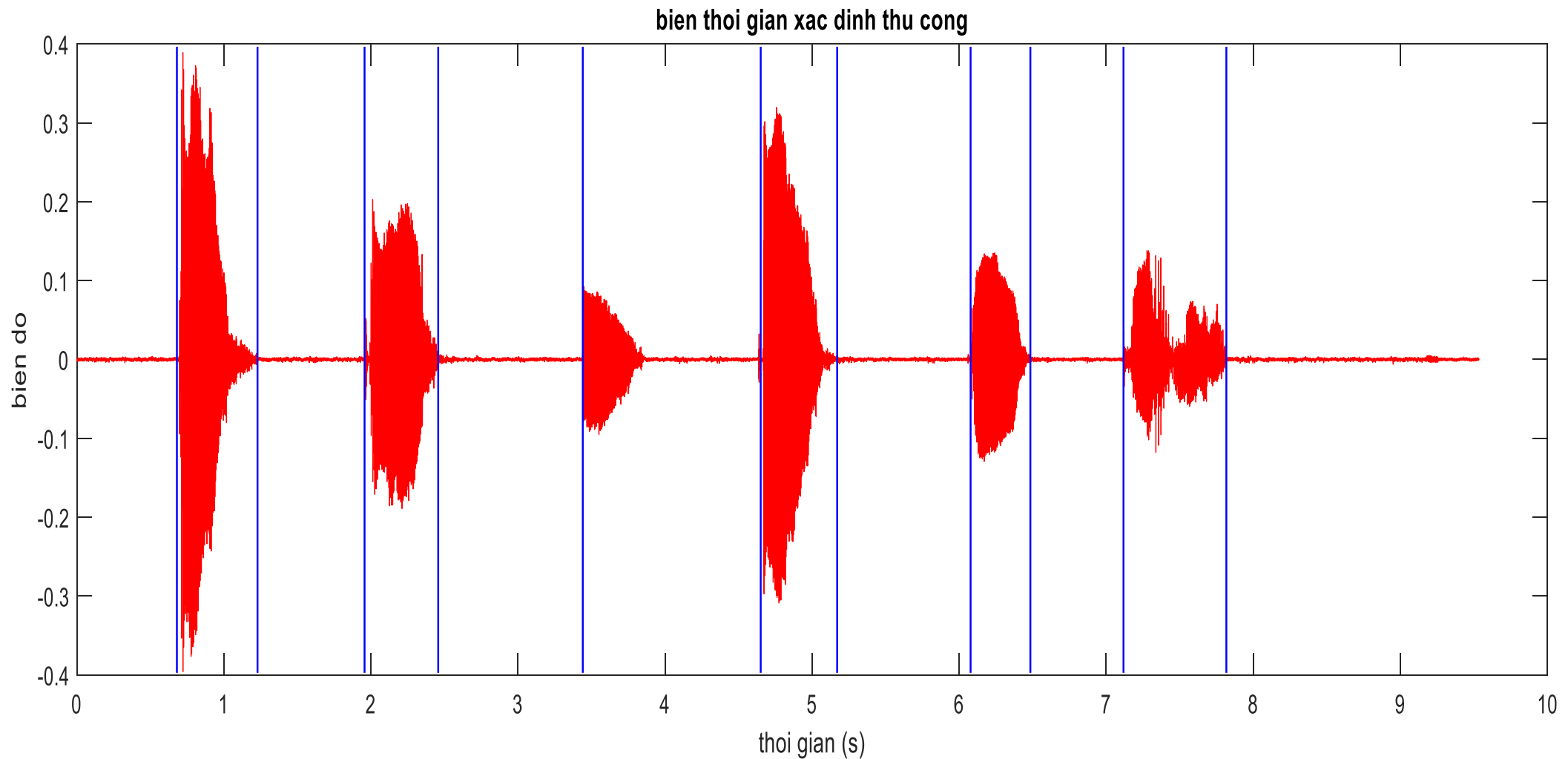
Hình 4. hình ảnh minh họa đồ thị Energy và ZRC không có những vị trí không giao nhau

5. Giải pháp và hiệu quả

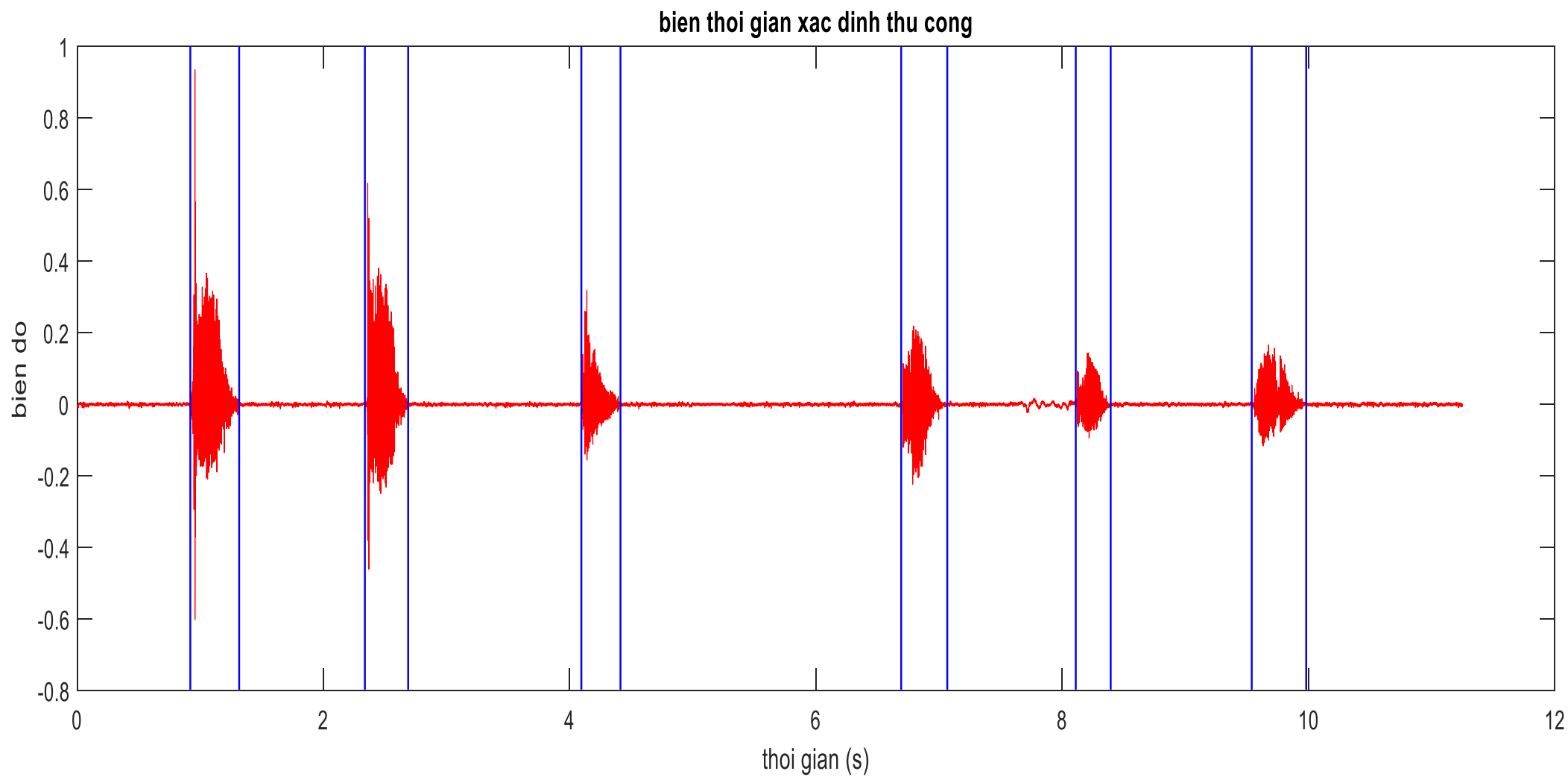
Trong những trường hợp không tìm thấy giao điểm giữa 2 đồ thị Energy và ZCR thì chúng em khảo sát sau khi vẽ các file tín hiệu nhận thấy rằng những đoạn có âm thanh thì có biên độ cao hơn khoảng 0.03 nên đã áp dụng xét ngưỡng để tách được khoảng lặng và âm thanh.

IV. KẾT QUẢ THỰC NGHIỆM

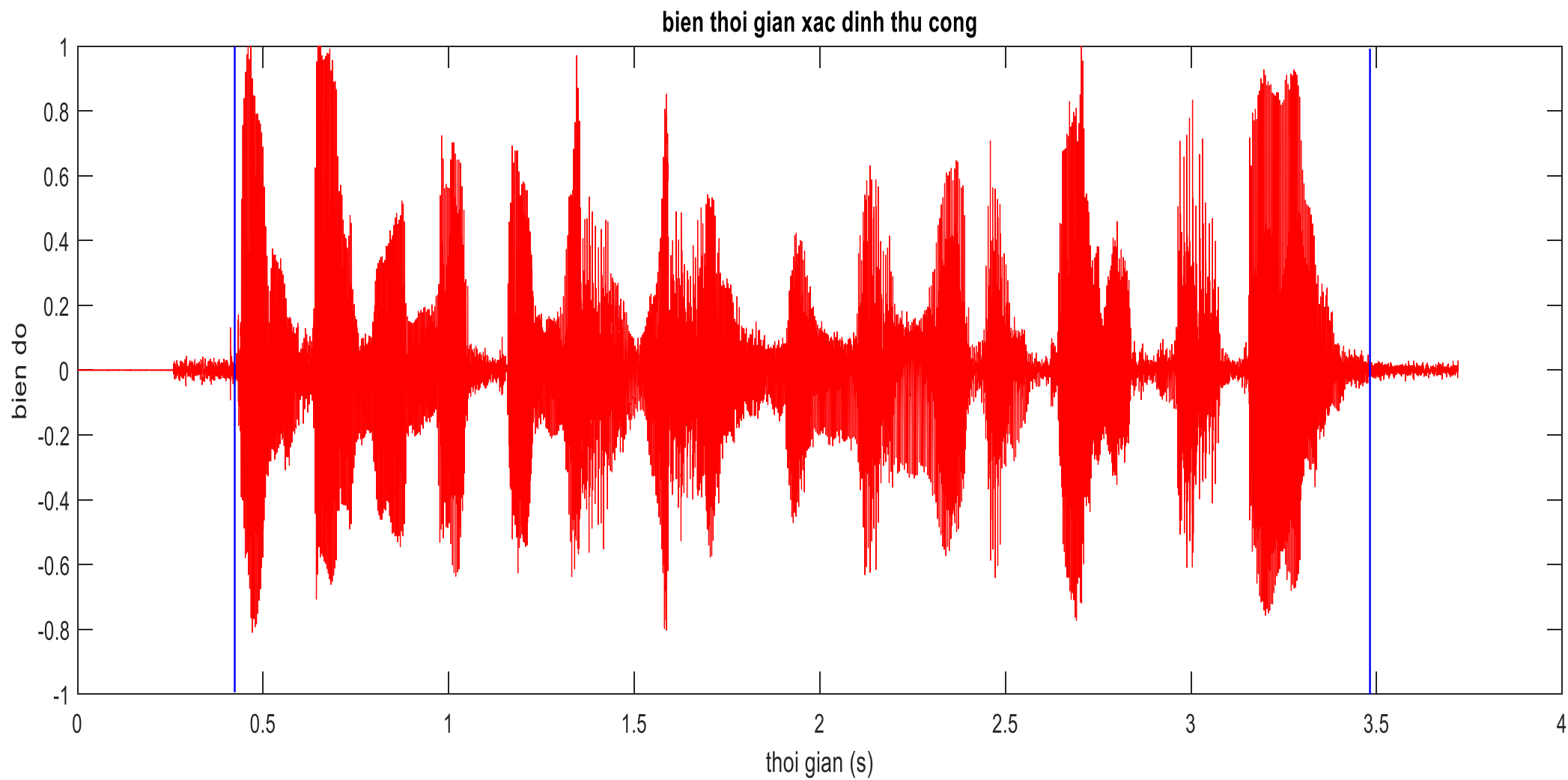
1. Kết quả định tính



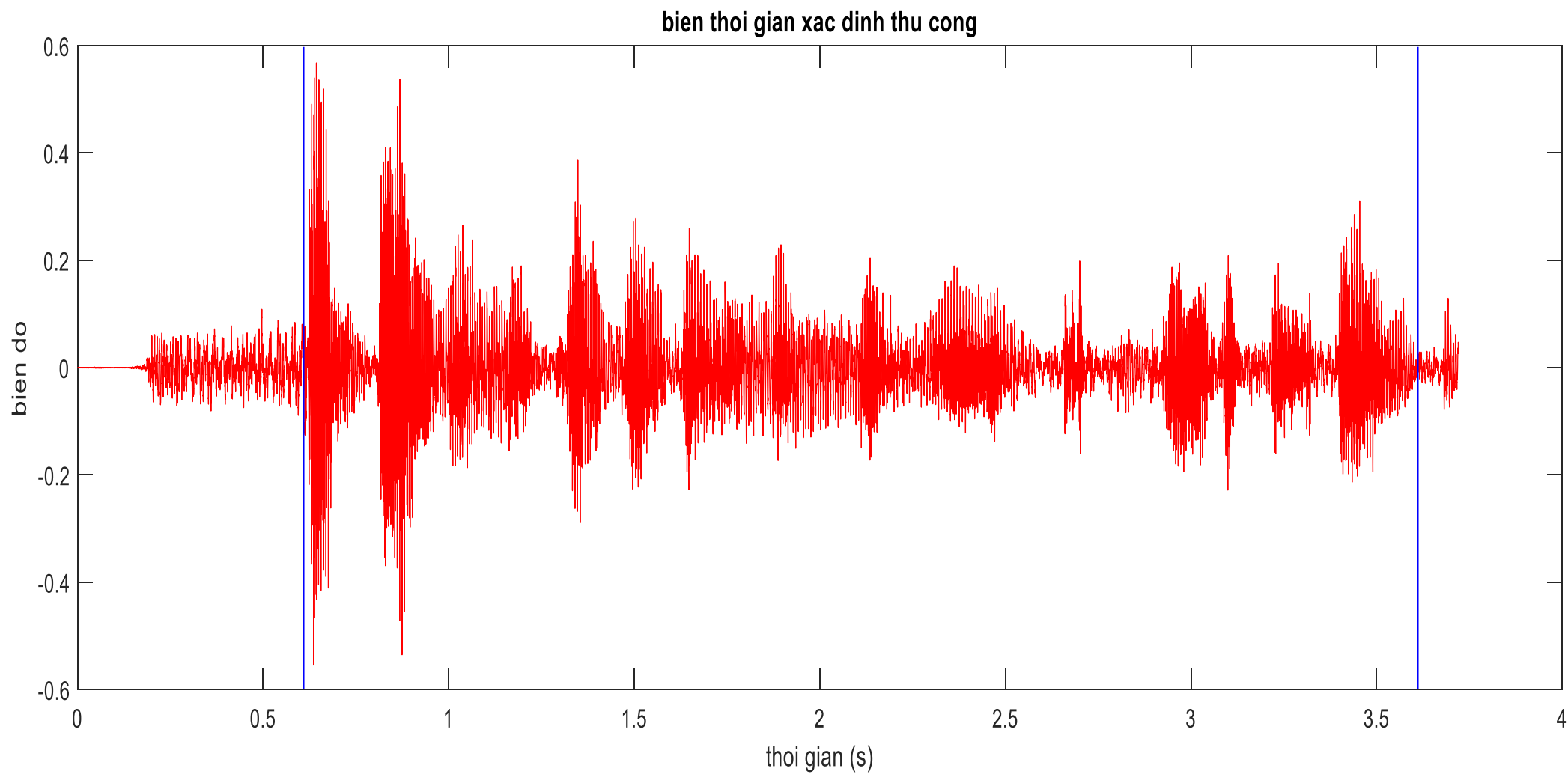
Hình 5. Biên thời gian được xác định thủ công của file lab_female.wav



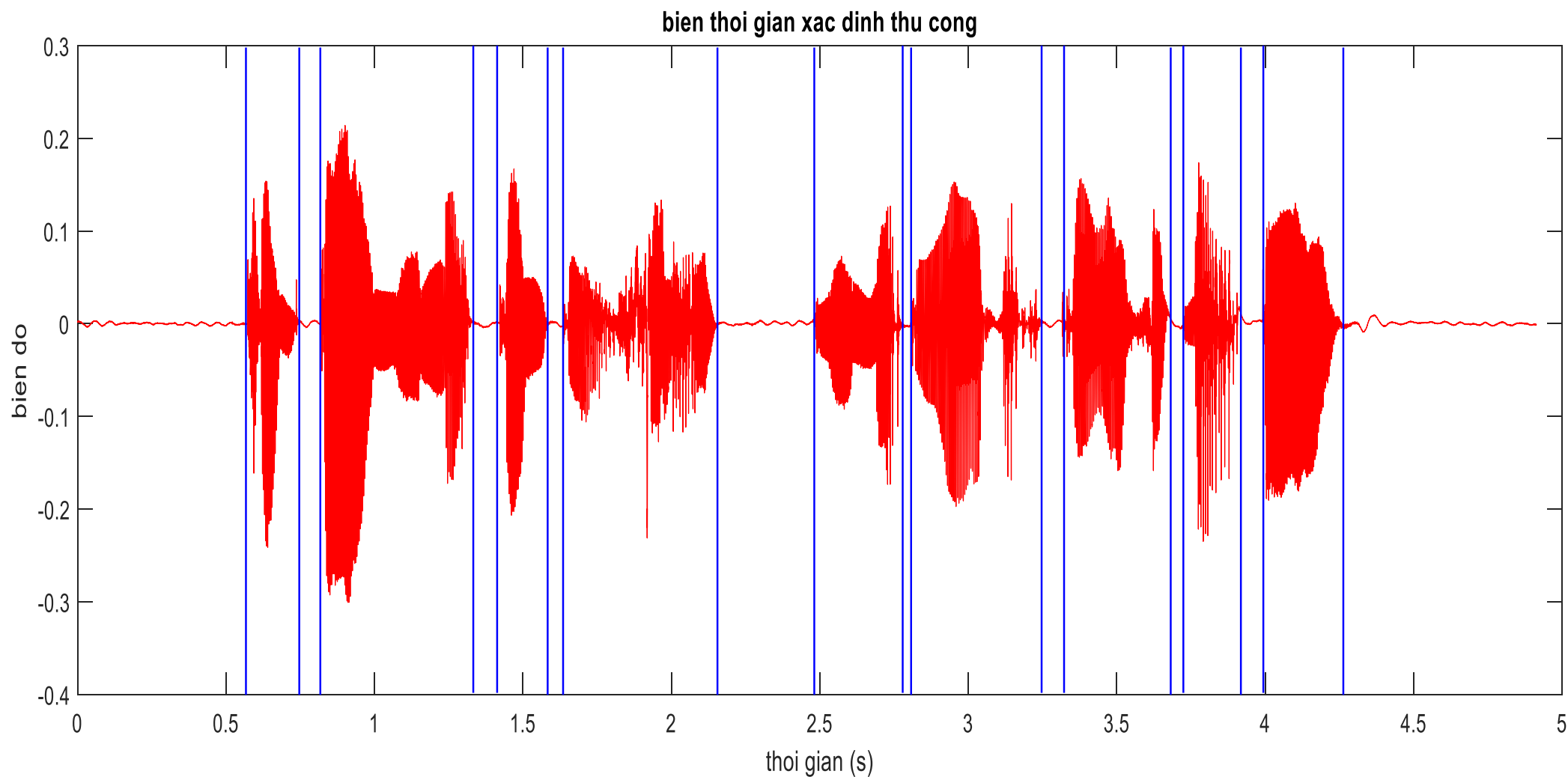
Hình 6. Biên thời gian được xác định thủ công của file lab_male.wav



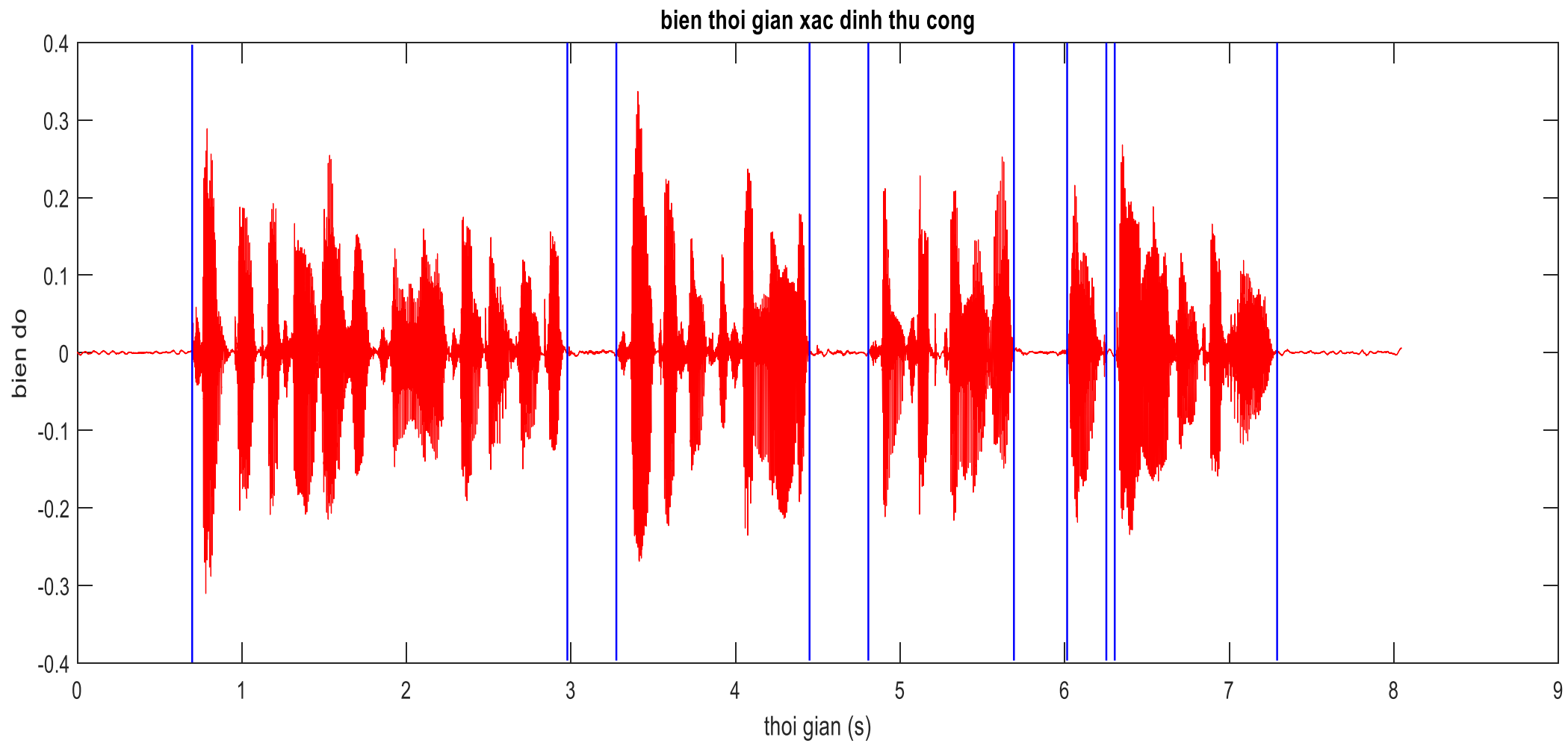
Hình 7. Biên thời gian được xác định thủ công của file phone_female.wav



Hình 8. Biên thời gian được xác định thủ công của file phone_male.wav



Hình 9. Biên thời gian được xác định thủ công của file studio_female.wav



Hình 10. Biên thời gian được xác định thủ công của file studio_male.wav

2.Kết quả định lượng

	Fs	Độ dài file (s)	Tổng số mẫu (mẫu)	Độ dài 1 frame(s)	Số mẫu / 1 frame (mẫu)	Tổng số frame
lab_female.wav	16000	9.5331	152529	0.01	160	953
lab_male.wav	16000	11.2491	179985	0.01	160	1124
phone_female.wav	16000	3.72	59520	0.01	160	372
phone_male.wav	16000	3.72	59520	0.01	160	372
studio_female.wav	16000	4.9136	78618	0.01	160	491
studio_male.wav	16000	8.0460	128736	0.01	160	804

V.KẾT LUẬN

- Bài báo cáo này thực hiện việc phân đoạn tín hiệu âm thanh, chúng em đã phân đoạn được một vài tín hiệu nhưng cũng có những trường hợp không đúng kết quả mong đợi, và vẫn chưa xác định được các đoạn nhiễu có trong tín hiệu.
- Trong tương lai chúng em sẽ cải thiện thuật toán để đạt được tín hiệu mong muốn

NHÓM MÌNH XIN CẢM ƠN
MỌI NGƯỜI ĐÃ LẮNG NGHE !