SP COURSE PROJECT

Presented by Eliminators

OUR TEAM

Snehil

Roshan

Gautam

PART-1
BIRD RECOGNITION

INTRODUCTION

- Bird sounds are unique acoustic signatures that can help identify bird species.
- These sounds have specific time and frequency characteristics that can be analyzed.

GOAL

1. Analyze the sounds of three distinct bird species to identify distinguishing features.
2. Use these features to identify the specie of the given bird sounds.

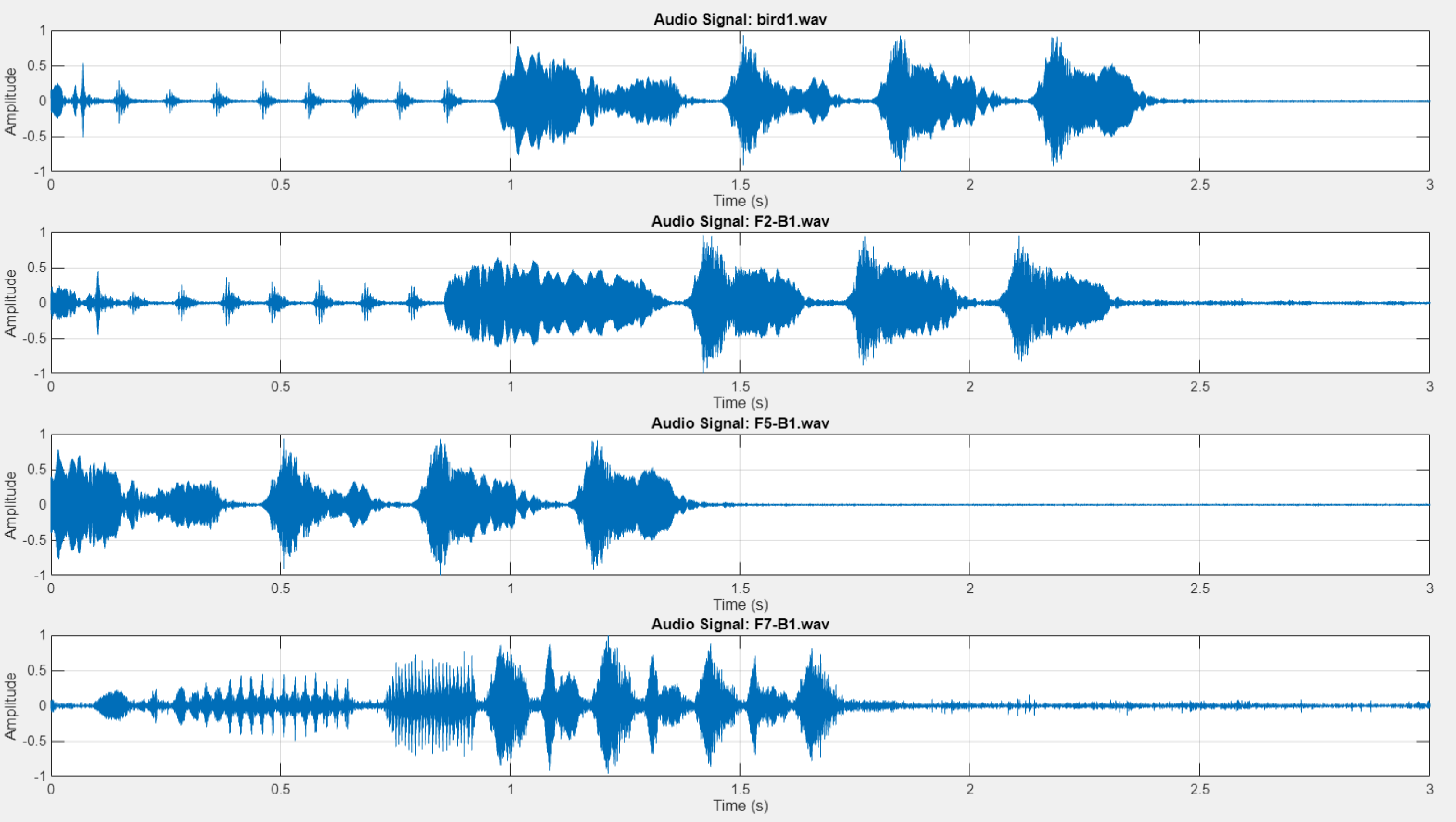
APPROACH AND METHODOLOGY

- We plotted waveforms of the three bird species to understand their time-domain patterns, and computed and analyzed the FFT magnitude and phase plots to observe frequency-domain characteristics.
- We observed that bird sounds of the same species have similar FFT magnitude plots. This consistency in frequency characteristics provides a unique signature for each bird species.
- Task bird sounds are compared with the reference bird sounds based on their FFT magnitude plots. This approach proved sufficient for distinguishing the bird species.

TIME-DOMAIN ANALYSIS

- We plot the waveforms of similar sounding bird files to visualize the time-domain patterns.
- We observe these plots to have similar shapes, but we will need a much complex algorithm to compare them in time domain to compute the bird species.

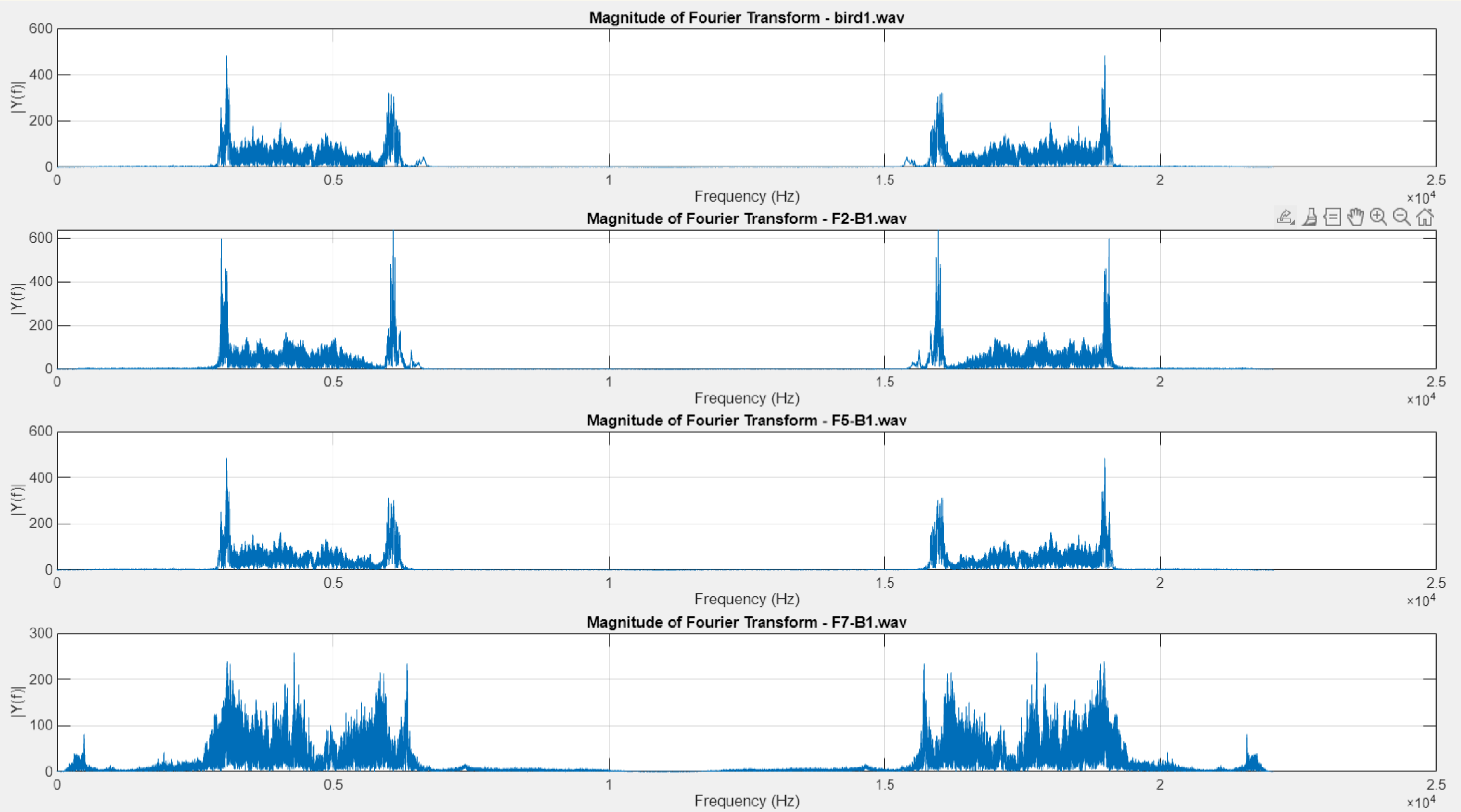
TIME-DOMAIN ANALYSIS



FREQUENCY-DOMAIN ANALYSIS

- We computed the FFTs of similar sounding bird files and plotted their magnitudes to visualize the frequency-domain patterns.
- We observe these plots to be very much similar, and thus we use this frequency domain analysis to compute the bird species.

FREQUENCY-DOMAIN ANALYSIS



FFT magnitudes of birds sounding similar to bird-1

IMPLEMENTATION OVERVIEW

1. **Reference Data**: FFT magnitude data of the three distinct bird species was precomputed and stored for comparison.
2. **Testing Process:**
   - A new bird sound is selected for testing.
   - The audio file is analyzed to extract its FFT magnitude.
3. **Comparison Logic:**
   - For each bird species, the sum of absolute differences between the test FFT magnitude and the reference FFT magnitude is computed.
   - These differences are calculated only within the selected frequency ranges to improve accuracy.
4. **Frequency Ranges:**
   - Specific frequency ranges are selected for comparison based on observed patterns in the reference data.
   - These ranges correspond to key frequency bands that are distinct for each bird species.
5. **Classification:**
   - The bird species corresponding to the minimum difference is identified as the most similar species.

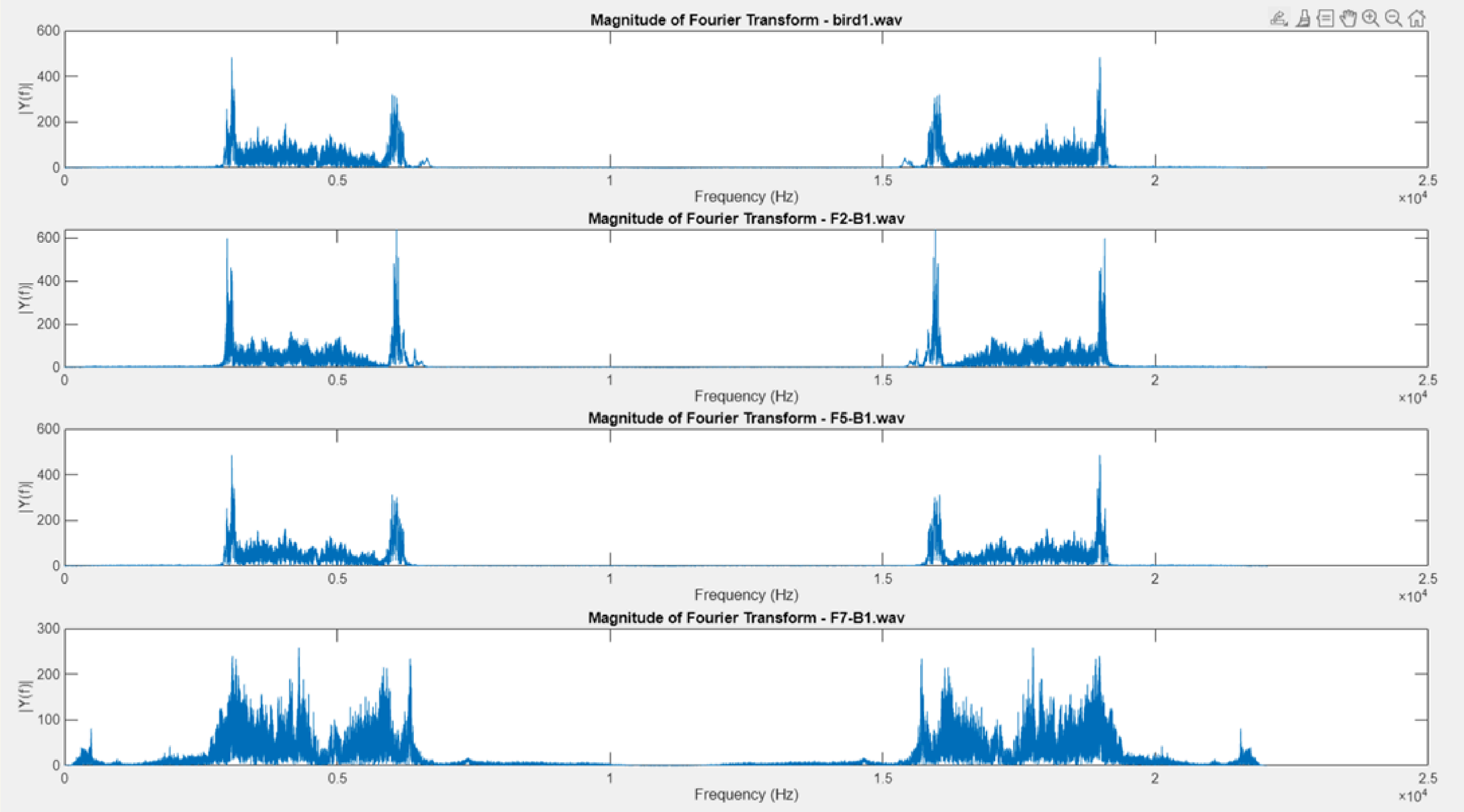
RESULTS

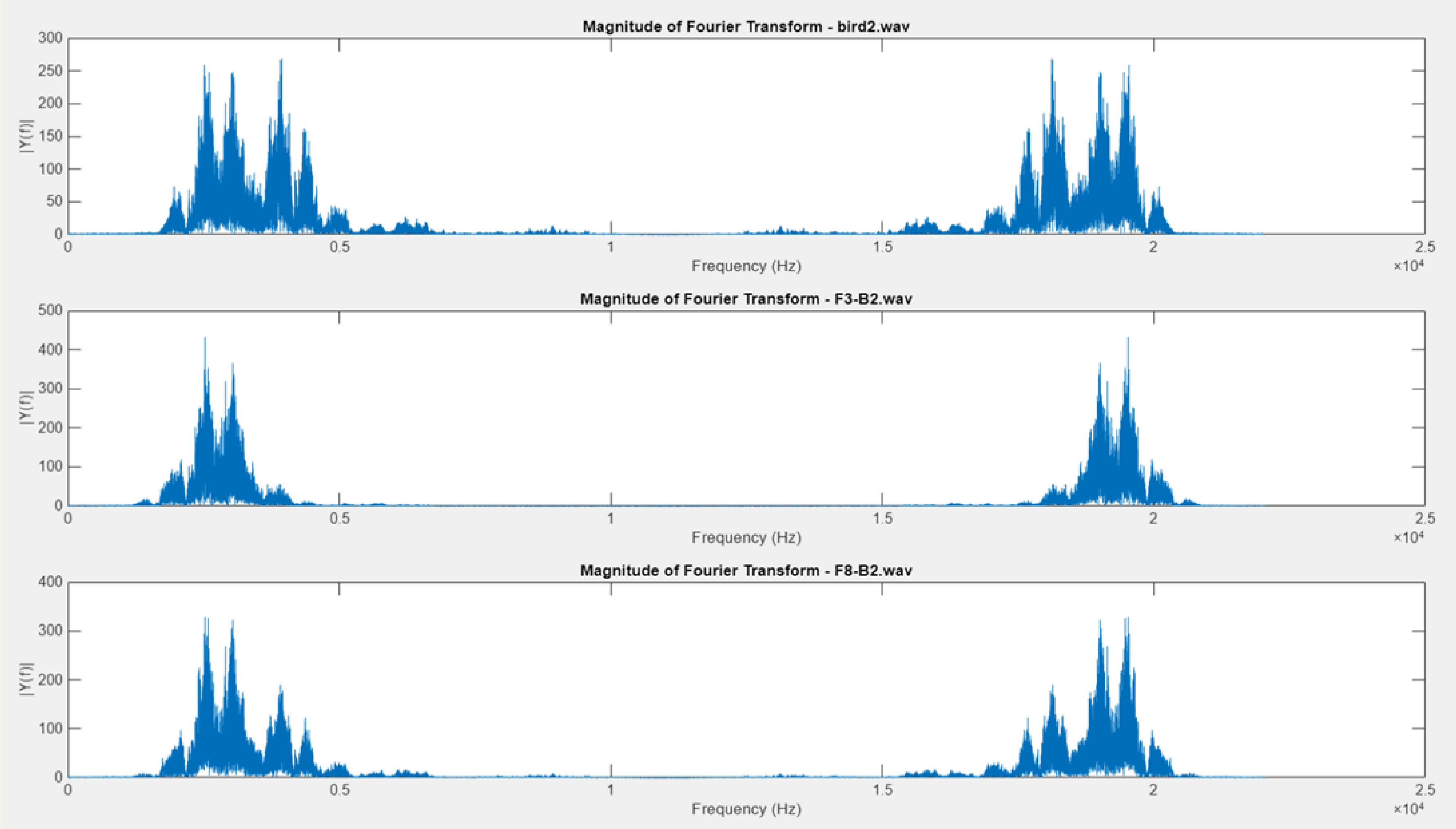
- Accuracy:
  - Correctly identified bird species using FFT magnitude analysis.
  - The selected frequency ranges ensured reliable classification.
- Strengths:
  - Simple and efficient method for species recognition.
  - Leverages distinct frequency characteristics to differentiate species.
- Challenges:
  - Assumes all bird sounds have clear frequency characteristics.
  - Noise in the test audio might affect accuracy.

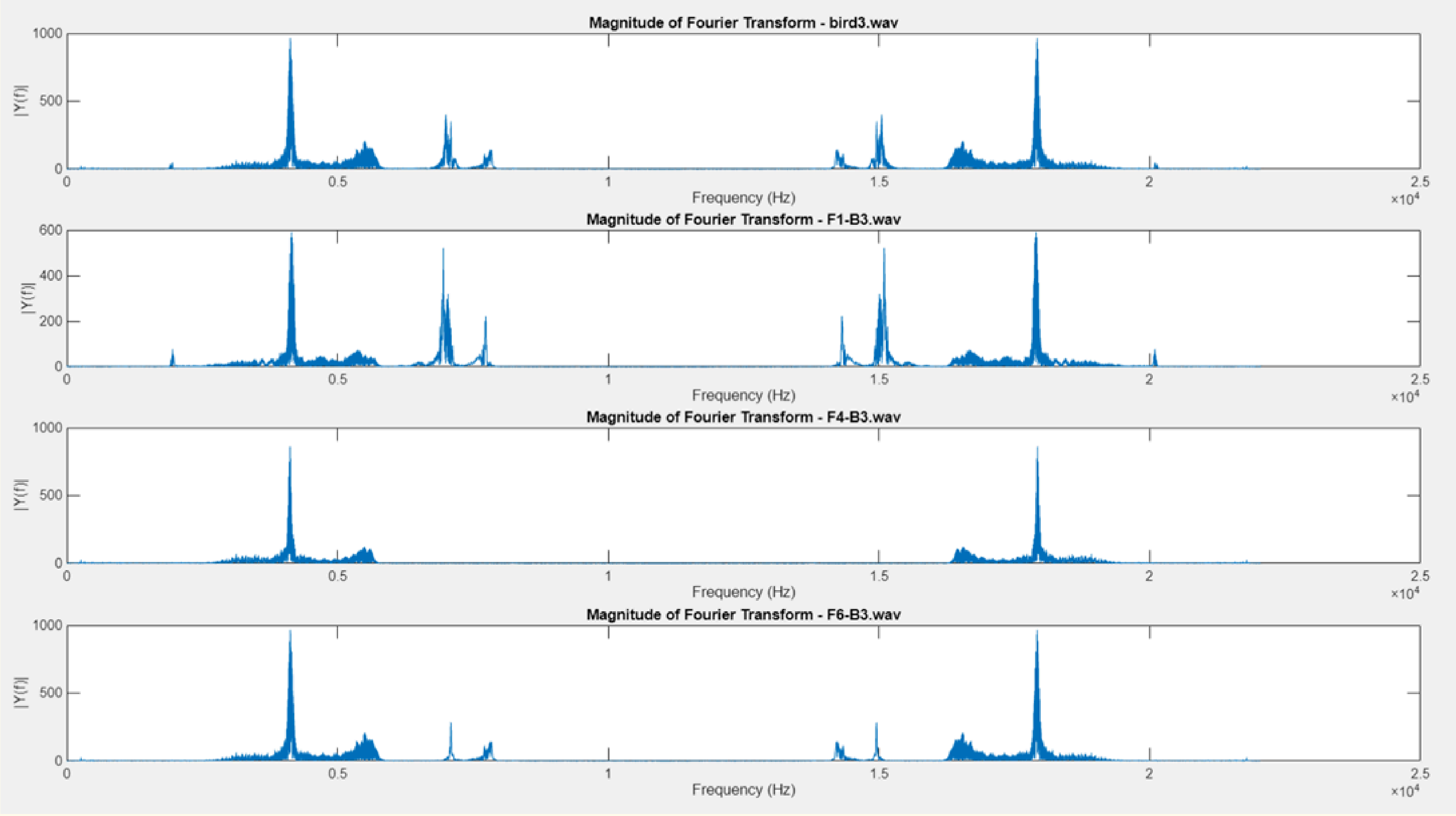
Task File	Identified Species
F1	3
F2	1
F3	2
F4	3
F5	1
F6	3
F7	1
F8	2

```
The given bird sound (F1-B3.wav) is similar to that of bird-3
The given bird sound (F2-B1.wav) is similar to that of bird-1
The given bird sound (F3-B2.wav) is similar to that of bird-2
The given bird sound (F4-B3.wav) is similar to that of bird-3
The given bird sound (F5-B1.wav) is similar to that of bird-1
The given bird sound (F6-B3.wav) is similar to that of bird-3
The given bird sound (F7-B1.wav) is similar to that of bird-1
The given bird sound (F8-B2.wav) is similar to that of bird-2
```

PLOTS


Magnitude of Fourier Transform - bird1.wav
Magnitude of Fourier Transform - F2-B1.wav
Magnitude of Fourier Transform - F5-B1.wav
Magnitude of Fourier Transform - F7-B1.wav
|Y(f)|
Frequency (Hz)
×10^4

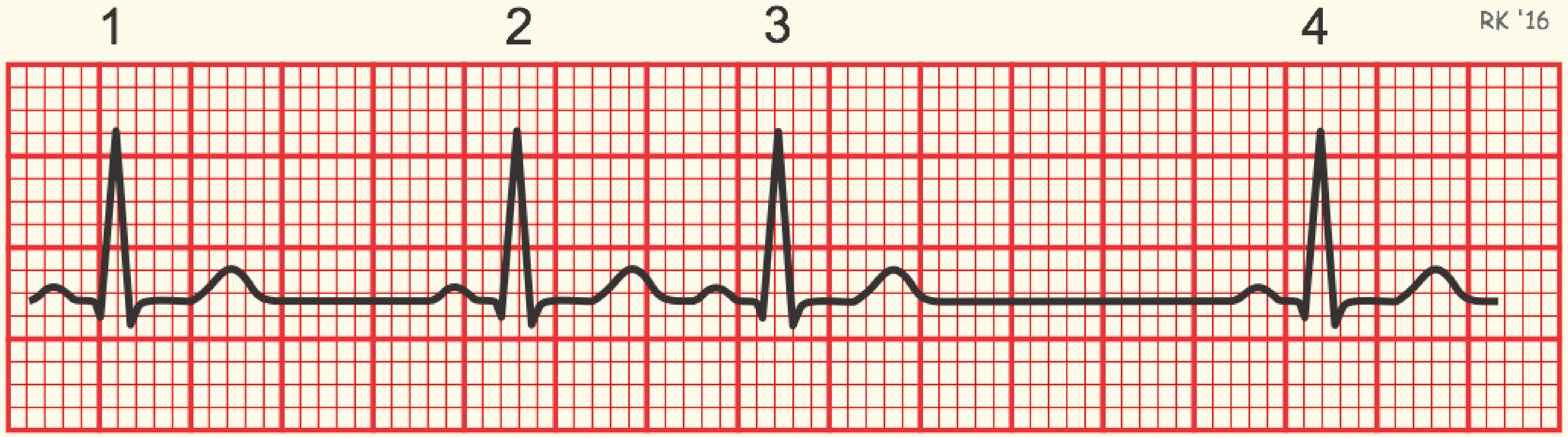

Magnitude of Fourier Transform - bird2.wav
Magnitude of Fourier Transform - F3-B2.wav
Magnitude of Fourier Transform - F8-B2.wav
|Y(f)|
Frequency (Hz)
×10^4


Magnitude of Fourier Transform - bird3.wav
Magnitude of Fourier Transform - F1-B3.wav
Magnitude of Fourier Transform - F4-B3.wav
Magnitude of Fourier Transform - F6-B3.wav
|Y(f)|
Frequency (Hz)
×10^4

PART-2
HEART RATE ESTIMATION

INTRODUCTION

This is a ECG

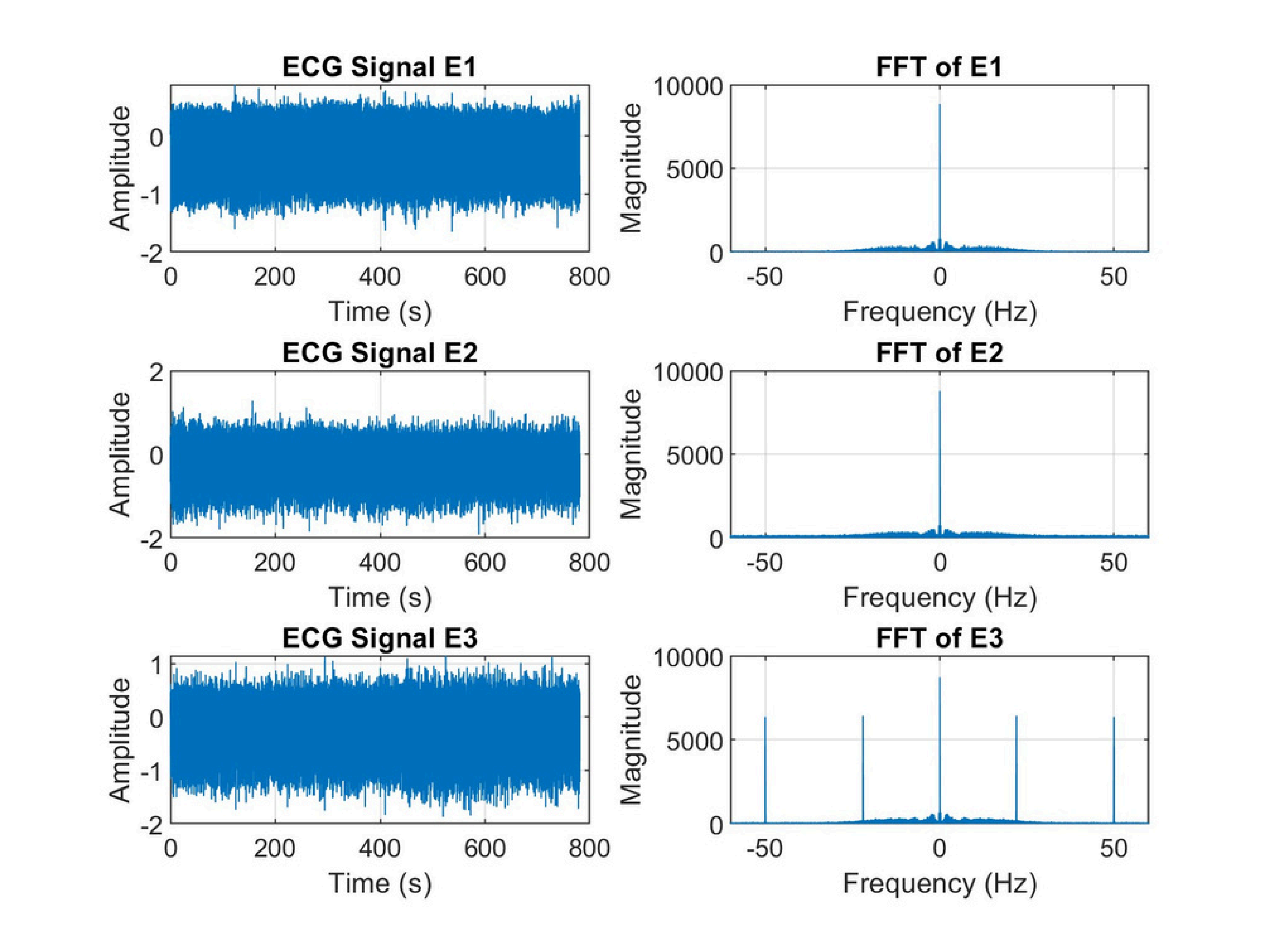


- Heart rate is measured between two consecutive R peaks
- HR(t)=60/r-r intervals
- Task 1 – Finding and Plotting the HR for E1 (Noise-Free ECG Signal)
- Task 2 – Noise Removal and repeating for E2 and E3 (Noisy ECG Signals):

PROBLEM

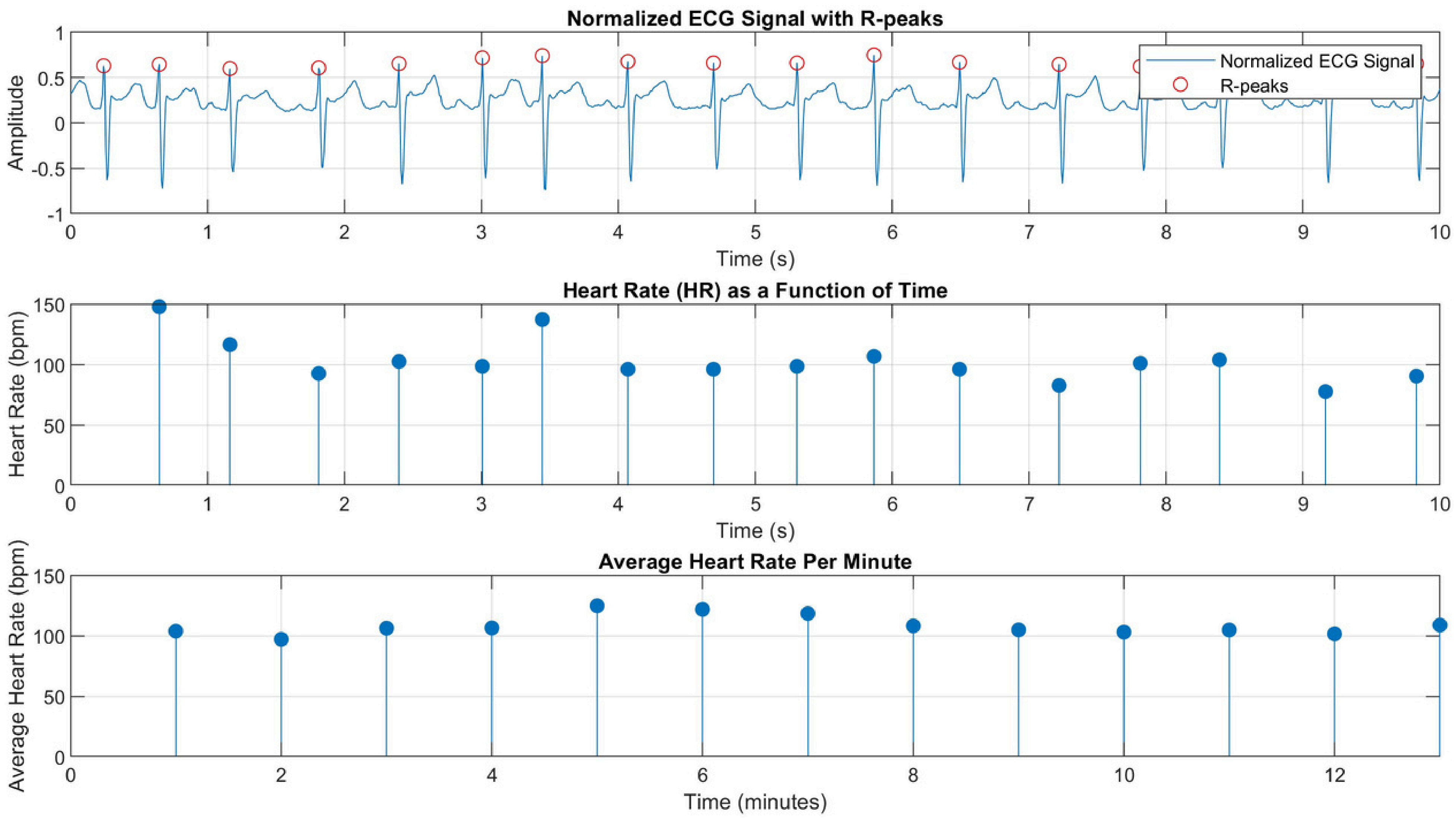
NOISE

- Given, E1 is ideal signal,Since E2 and E3 has frequency components whcih are unnecessary, we remove them by using LPF and bandstop filter(for E3)

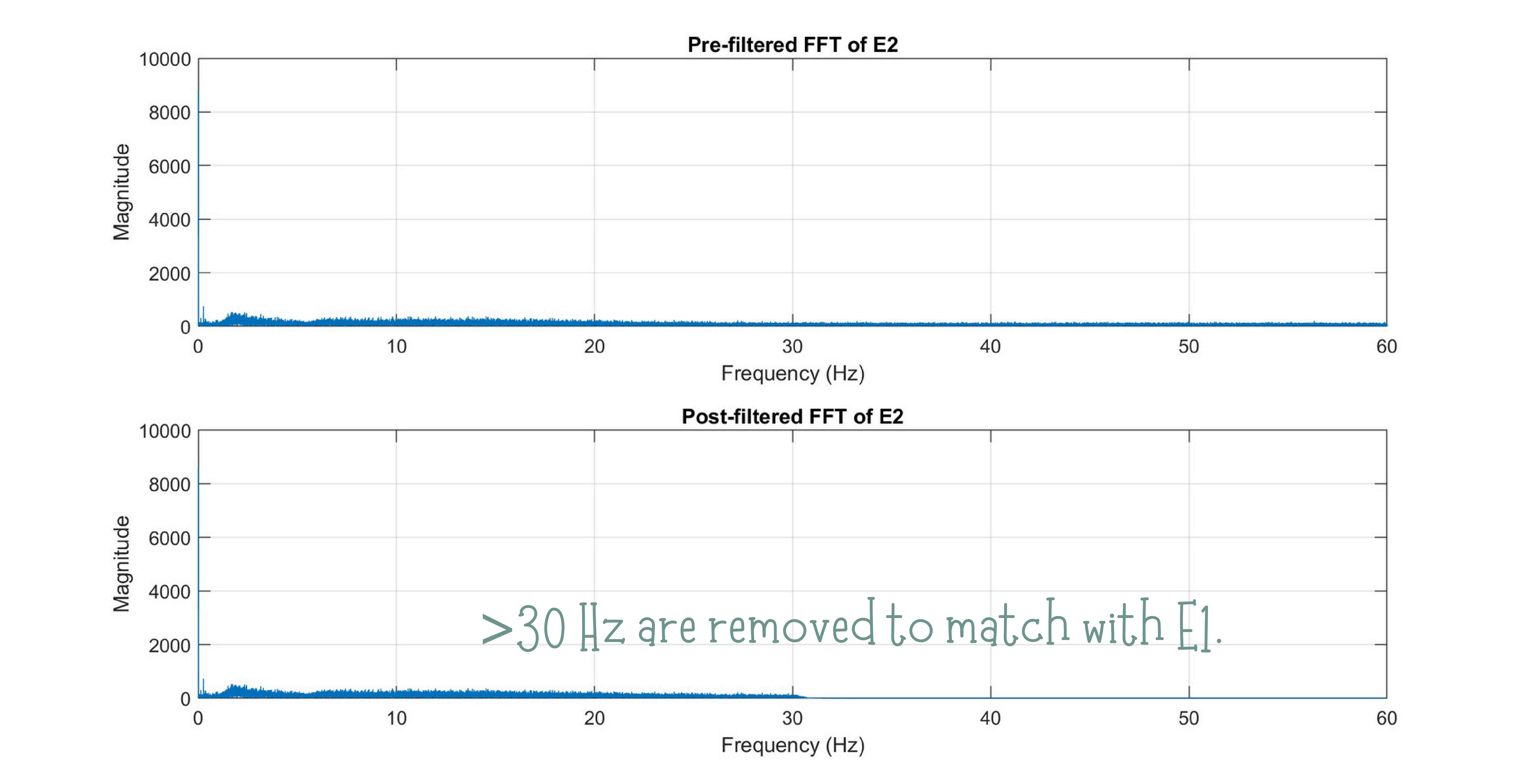


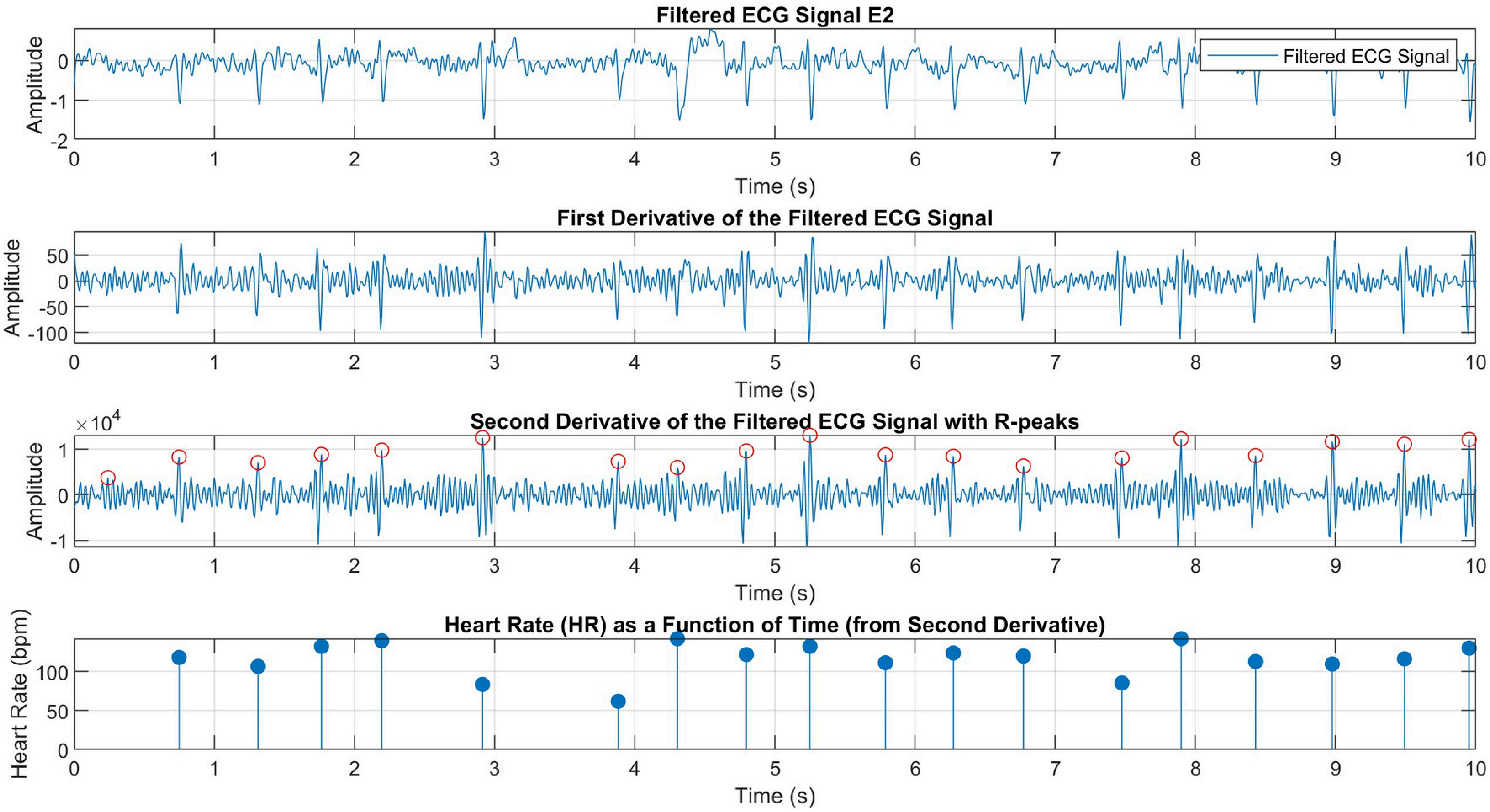
ESTIMATING HEART RATE FORM E1

- Since the signal is ideal, we use findpeak function and trace out all R peaks.
- We calculate the time between each peak.
- HR=60/T
- the second plot is heart rate every second and the third plot is per minute



E-2 PRE AND POST NOISE FFT




Filtered ECG Signal E2
Filtered ECG Signal
Amplitude
Time (s)
First Derivative of the Filtered ECG Signal
Amplitude
Time (s)
Second Derivative of the Filtered ECG Signal with R-peaks
×10^4
Amplitude
Time (s)
Heart Rate (HR) as a Function of Time (from Second Derivative)
Heart Rate (bpm)
Time (s)

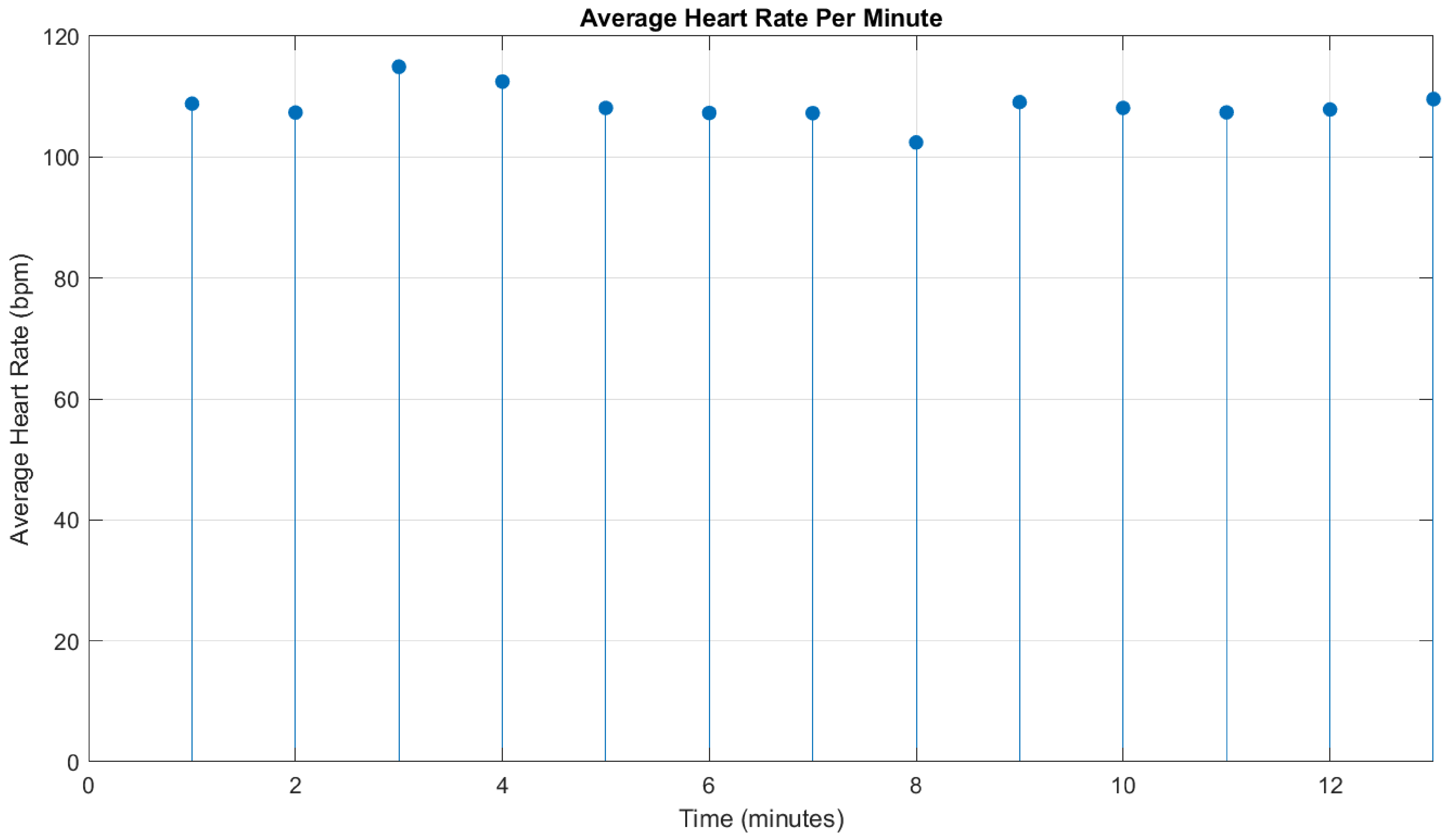
EXPLANATION E2

- The noise is still threre in filtered signal and we cannot use findpeak fn as it will detect nosiy peaks
- Inorder to solve this, we differentiate the signal twice and then use findpeak function on the second differential.

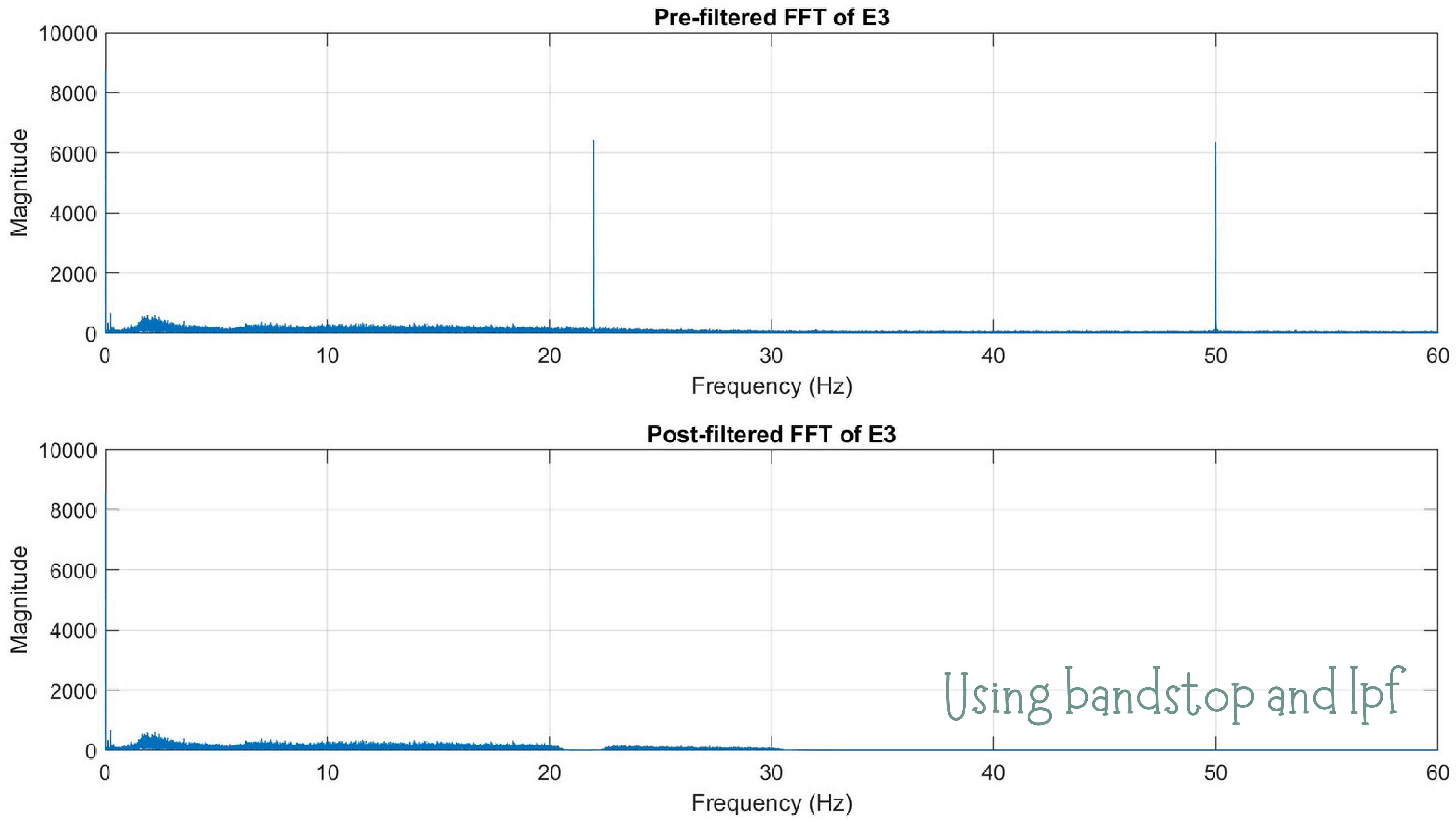
WHY 2ND DIFFERENTIAL

- The second derivative detects the rate of change of the slope. It is particularly sensitive to the points where the signal undergoes rapid changes, such as at the peak of the R-wave. At the R-peak, the ECG waveform typically transitions from a steep rise to a rapid fall

E2 HEART RATE PER MIN



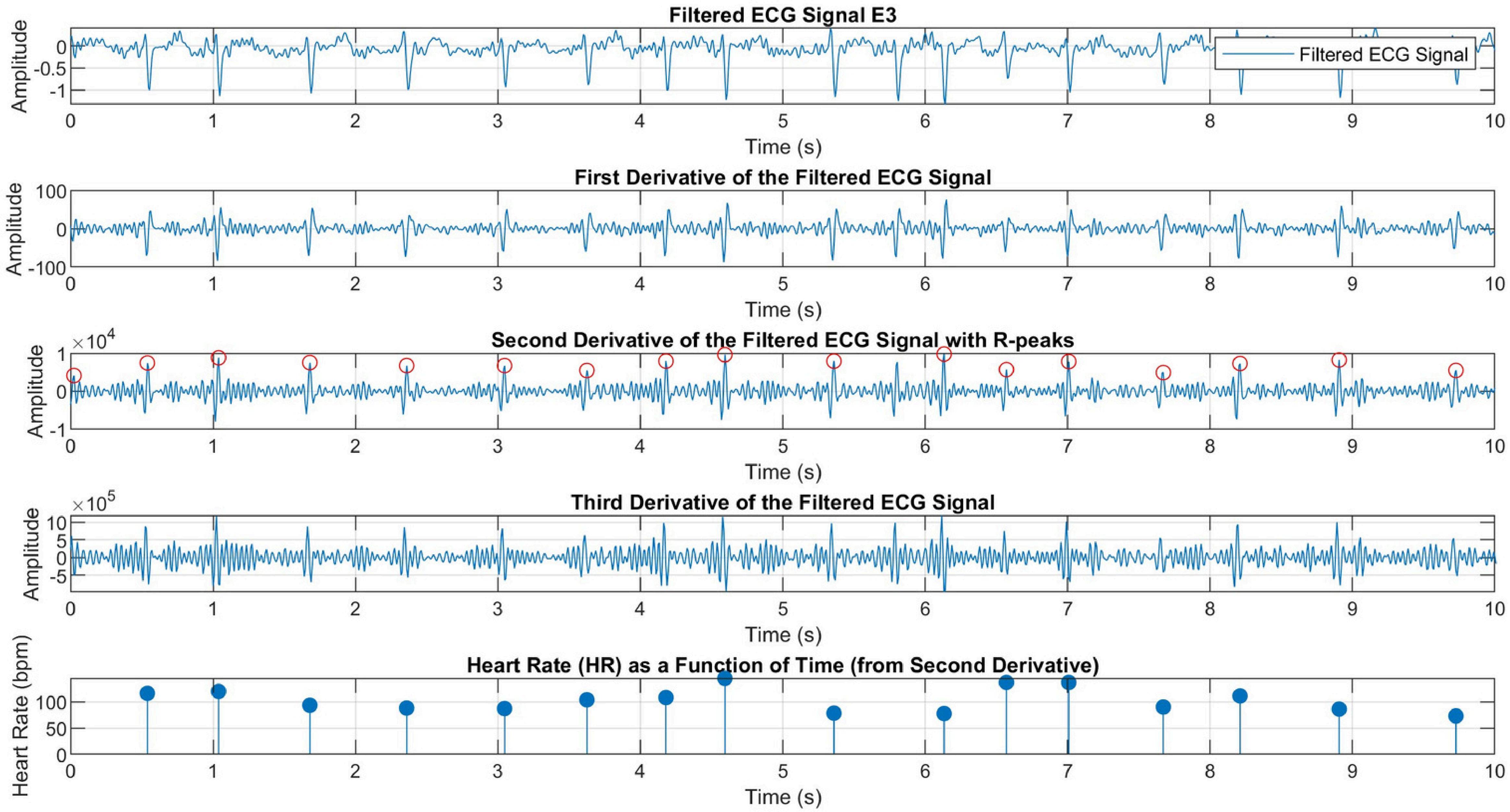
E3 NOISE



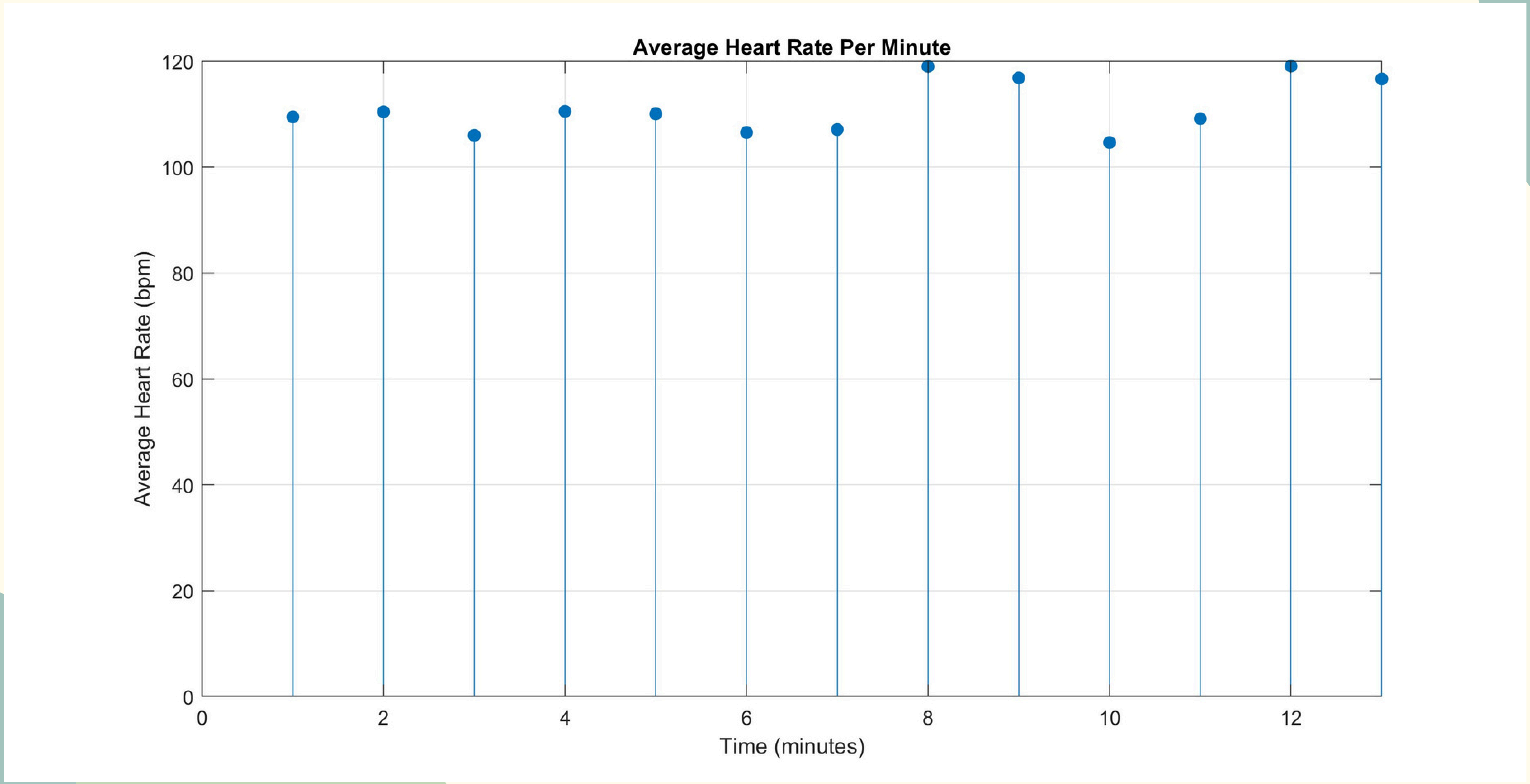
Using bandstop and lpf

E3 ANALYSIS

- the same goes here. we cant use findpeak fn directly
- we use it in second derivative



E3 HEART RATE



PART-3
LOUDNESS SEGMENTATION

INTRODUCTION

Speech is a natural mode of communication to convey the feelings and intentions. These intentions will be passed by changing loudness of the words spoken in one's speech. Often, the words in speech have strong perceptual boundaries however similar strong indications can't be observed in the speech signal.

PROBLEMS

You are given different speech samples and text files. The text files contain information of spoken words, its start and end time and markings indicate louder (1) or not (0) for each speech sample.

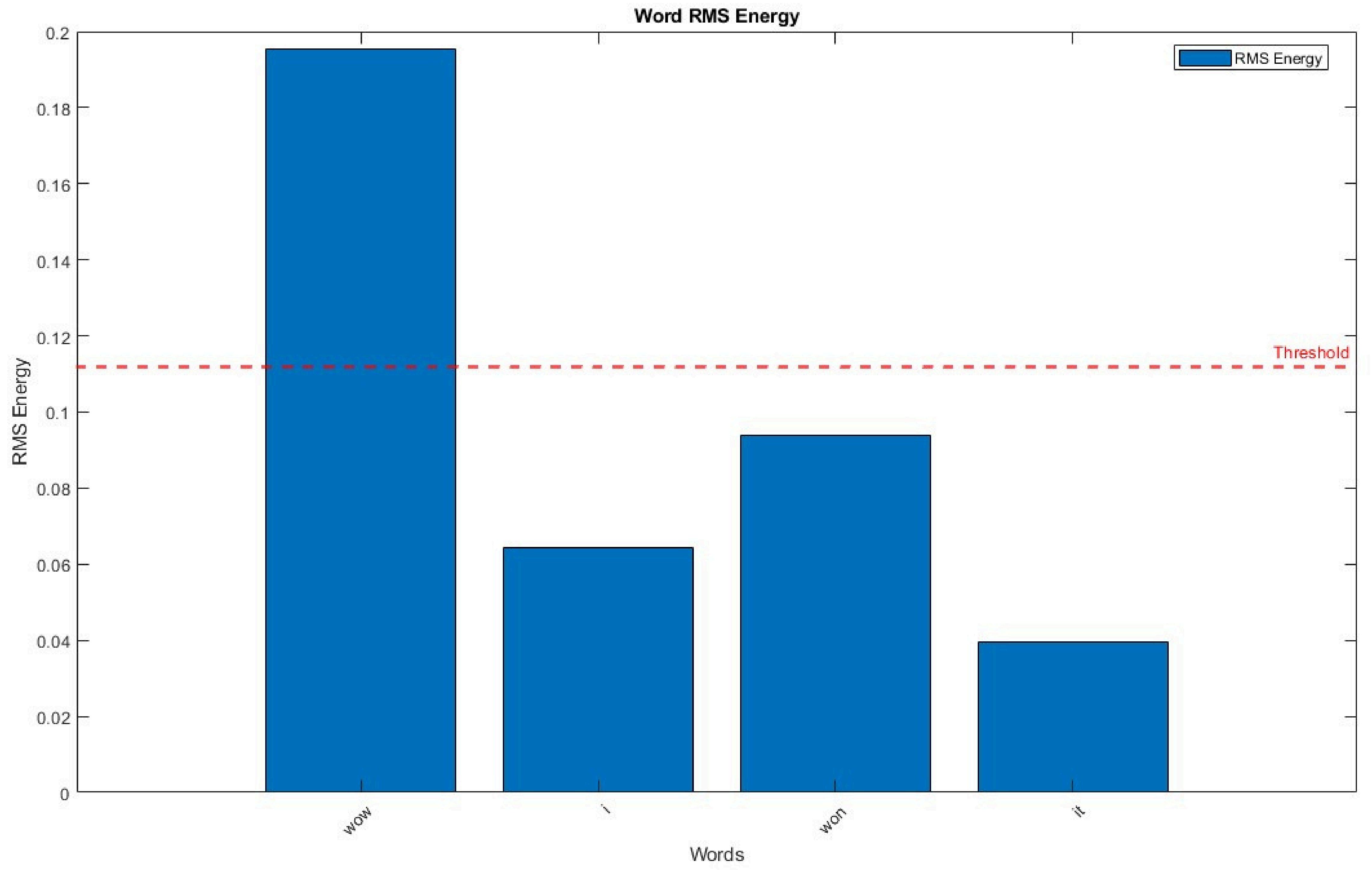
1. For given speech samples, find the louder words using its respective start and end times.

2. For given speech samples, find the louder words without using start and end time information.

TASK-1

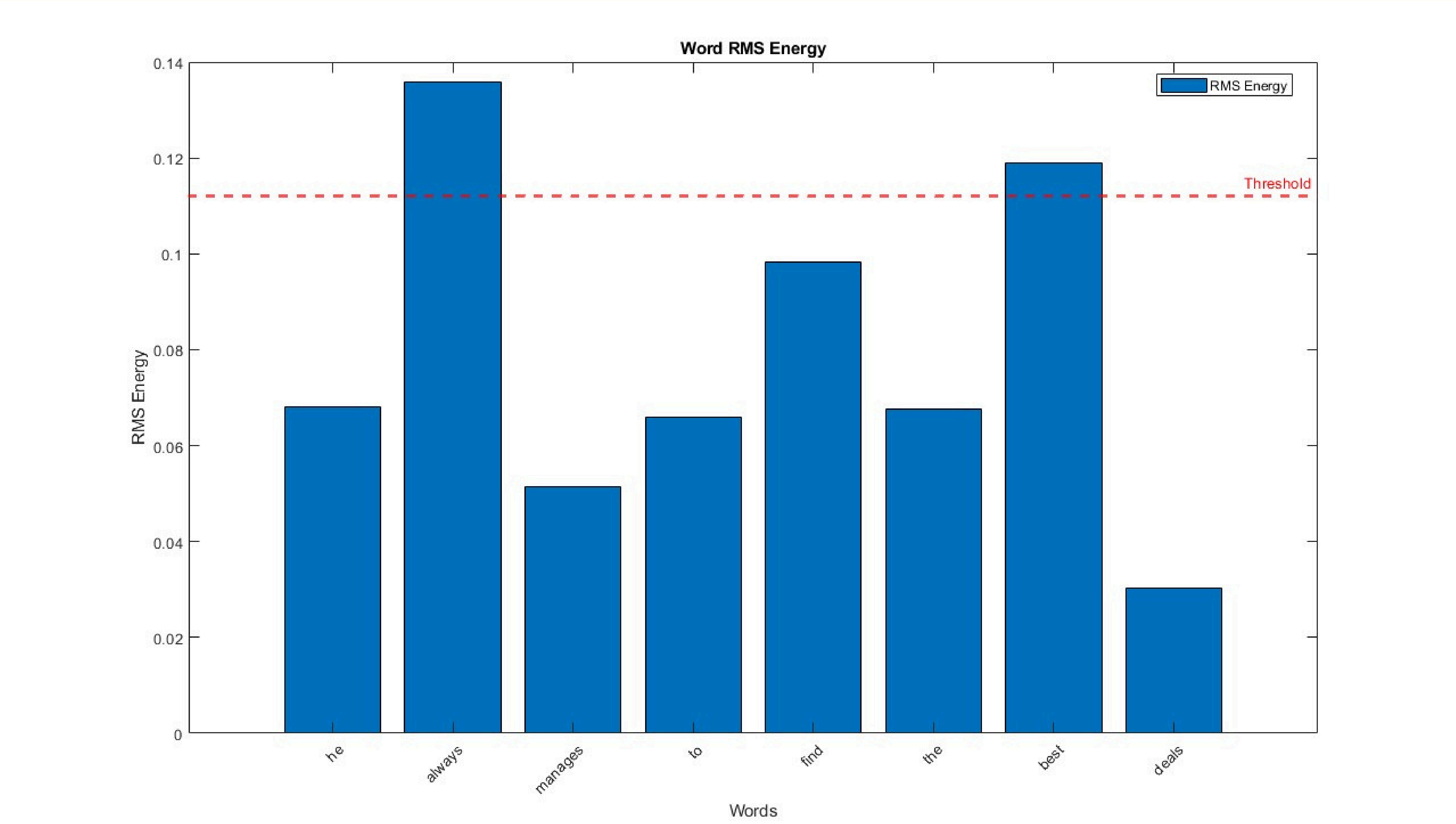
For task-1, we use the time stamps to get the segments from the audio signal where the words are located. Then we calculate the rms energy of all the segments and compare them. We see that the emphasised words have much higher energy than the softer words. Hence we set an appropriate threshold to get the louder words

AUDIO-1



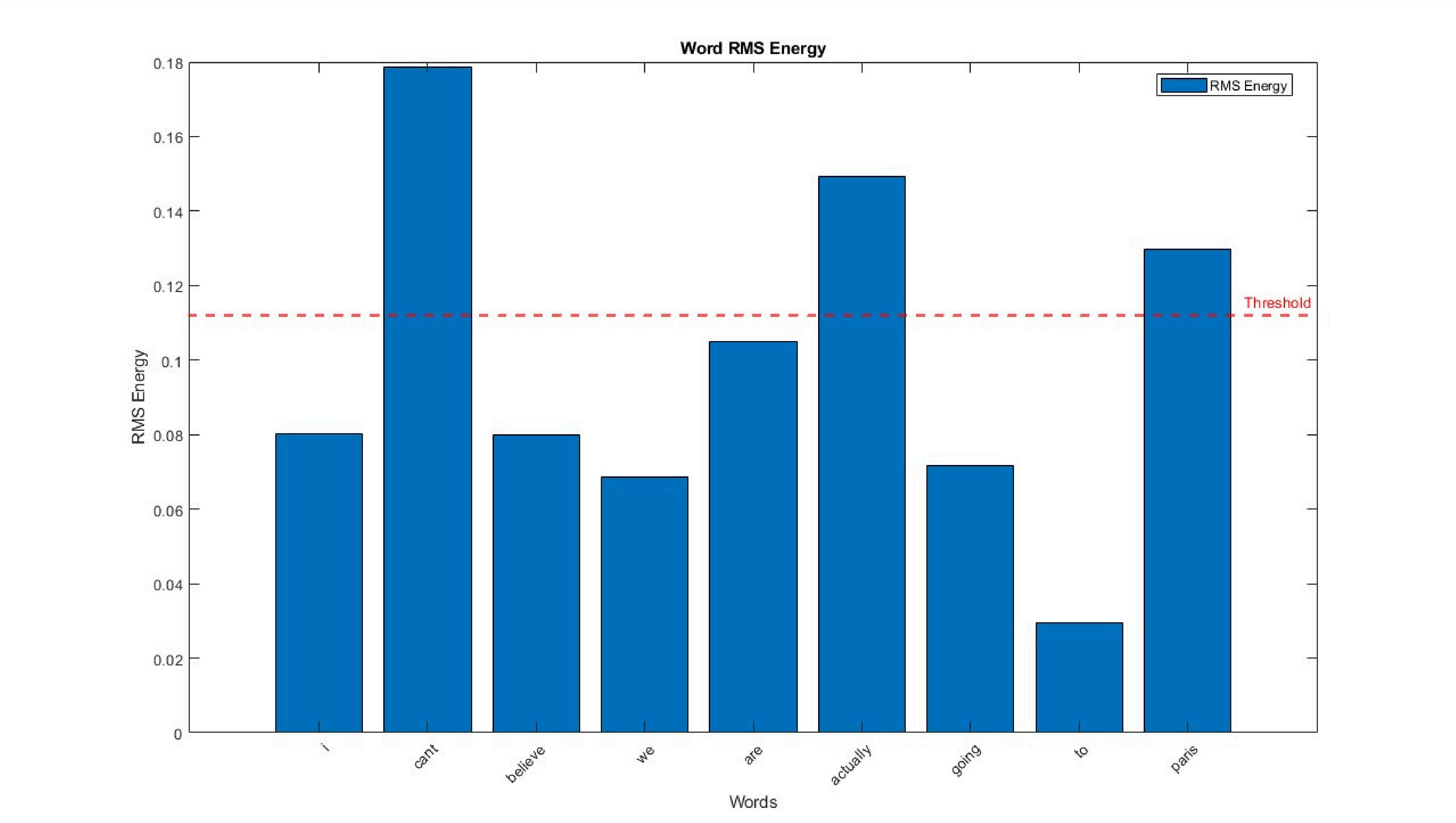
Word RMS Energy
RMS Energy
Threshold
RMS Energy
Words
wow
i
won
it
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0

AUDIO-2



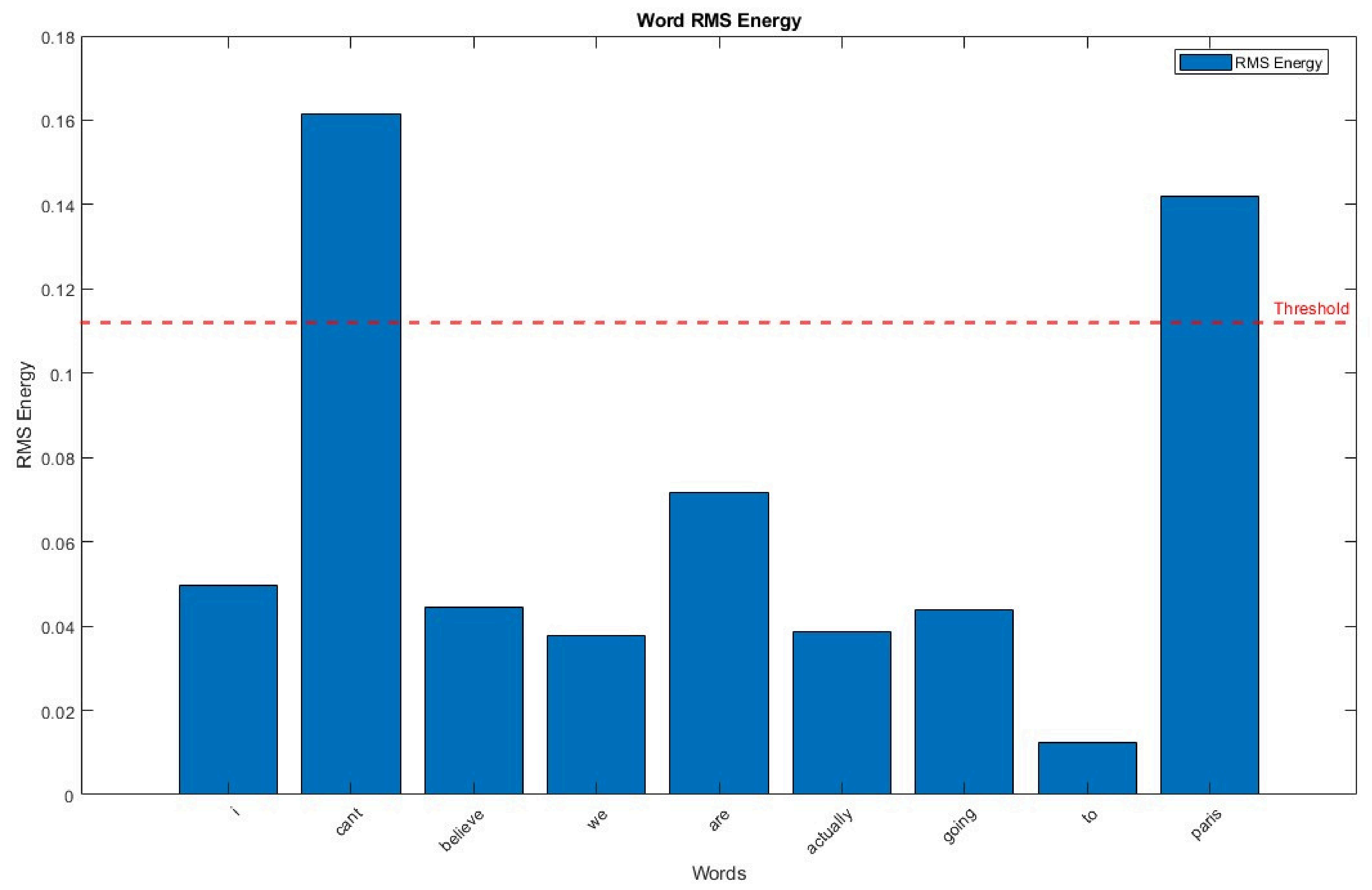
Word RMS Energy
RMS Energy
Threshold
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
RMS Energy
he
always
manages
to
find
the
best
deals
Words

AUDIO-3



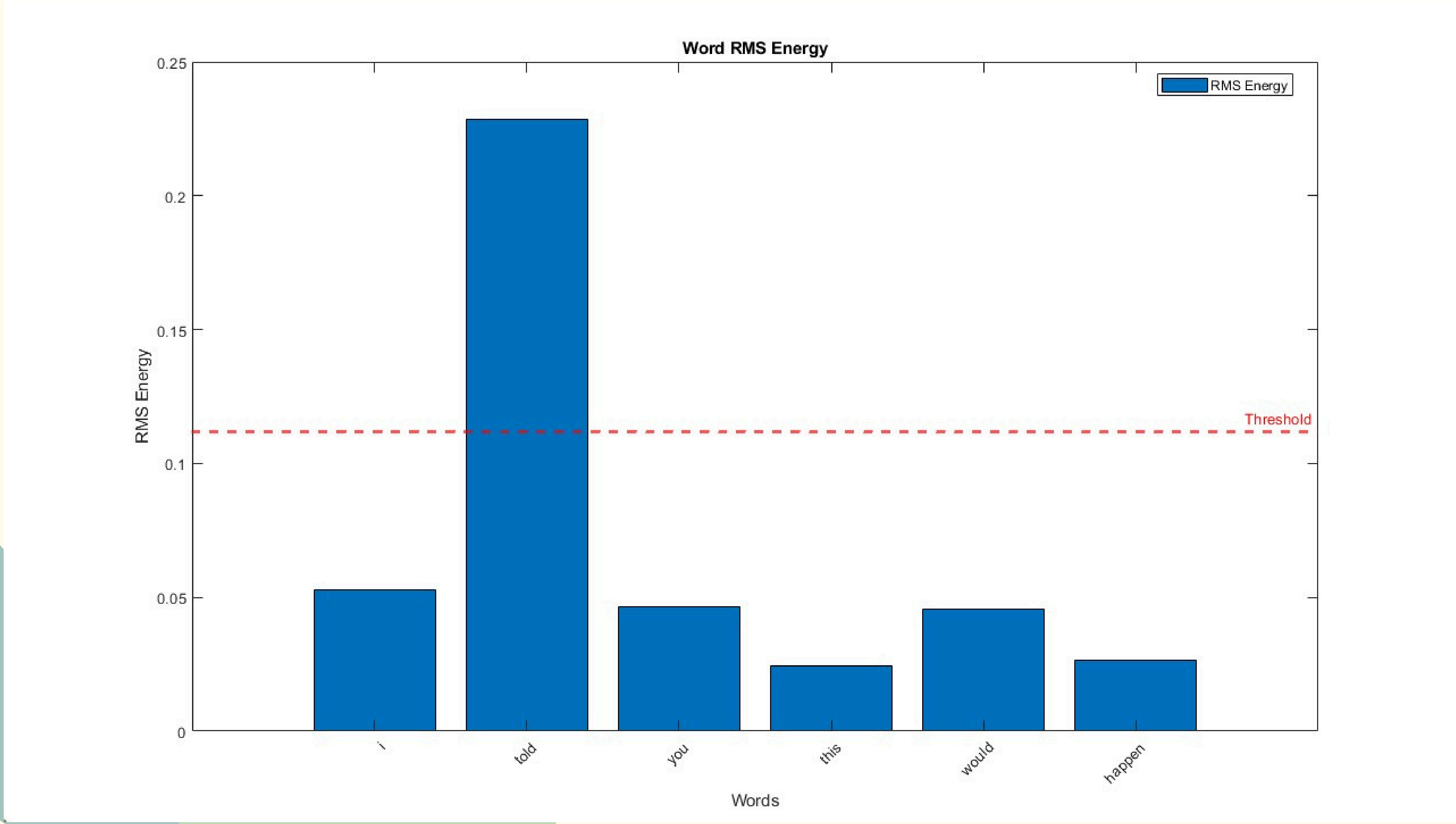
Word RMS Energy
RMS Energy
Threshold
RMS Energy
Words
0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
0.16
0.18
i
cant
believe
we
are
actually
going
to
paris

AUDIO-4



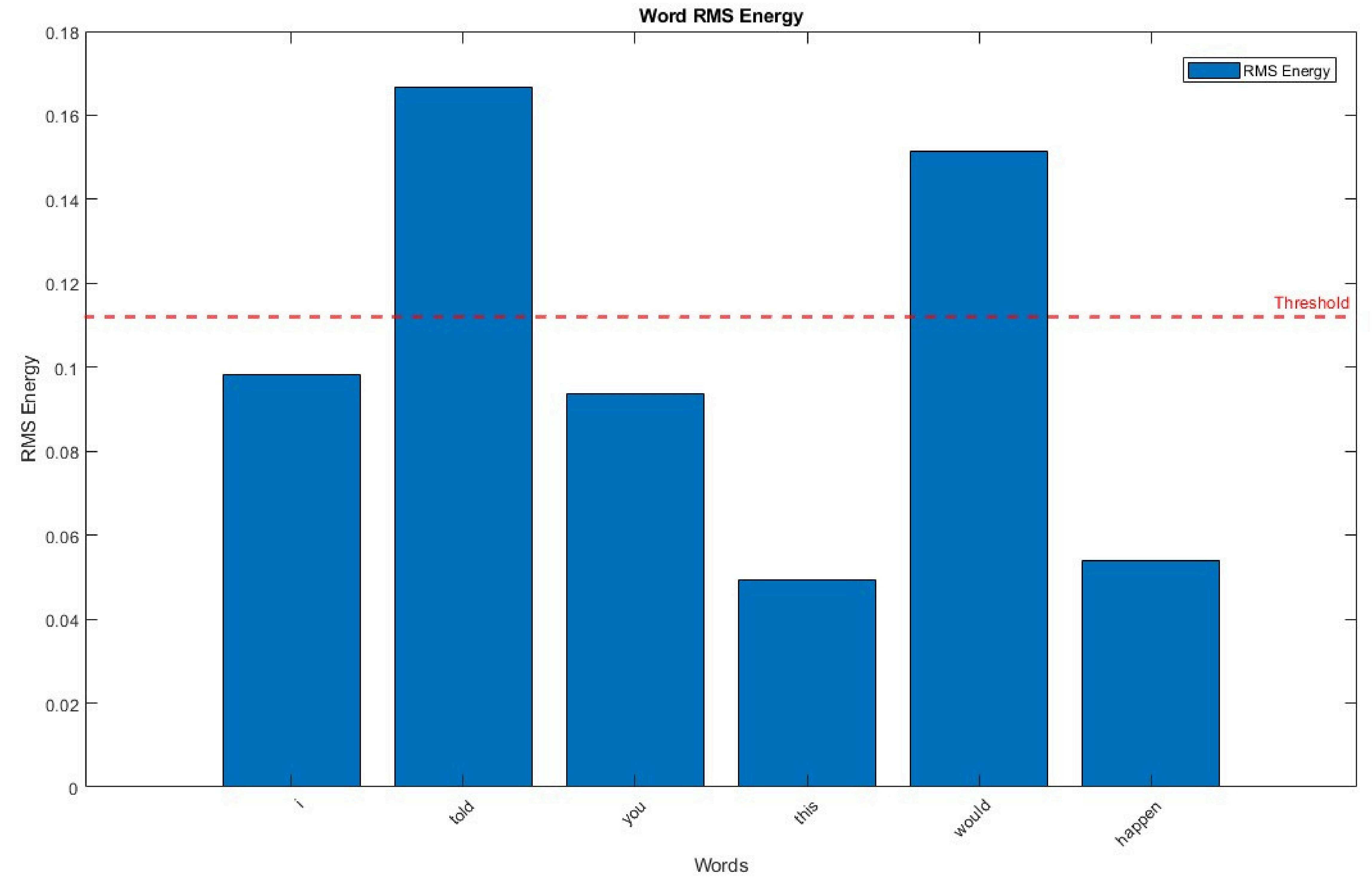
Word RMS Energy
RMS Energy
Threshold
RMS Energy
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
i
cant
believe
we
are
actually
going
to
paris
Words

AUDIO-5



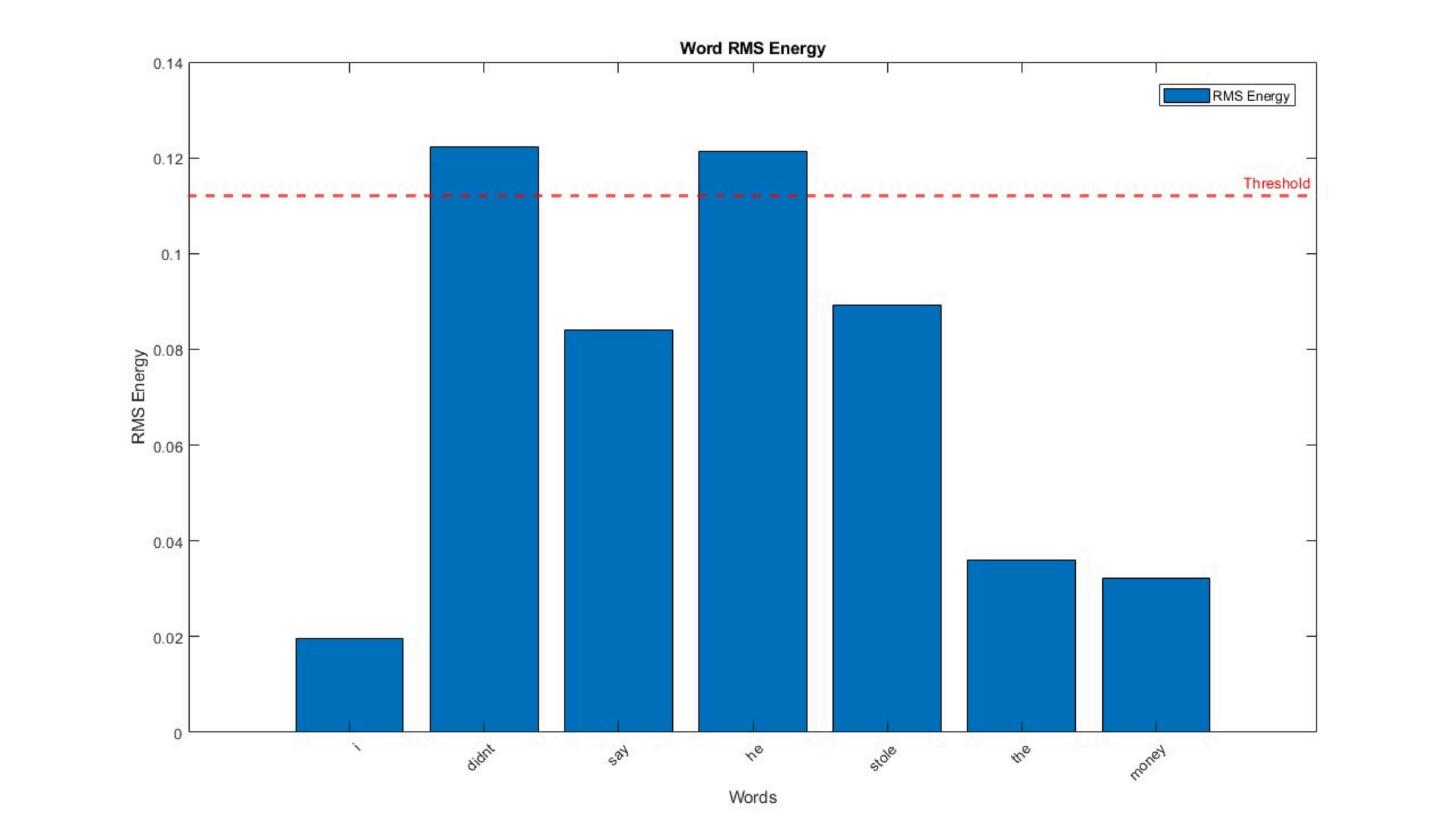
Word RMS Energy
RMS Energy
Threshold
0.25
0.2
0.15
0.1
0.05
0
RMS Energy
i
told
you
this
would
happen
Words

AUDIO-6



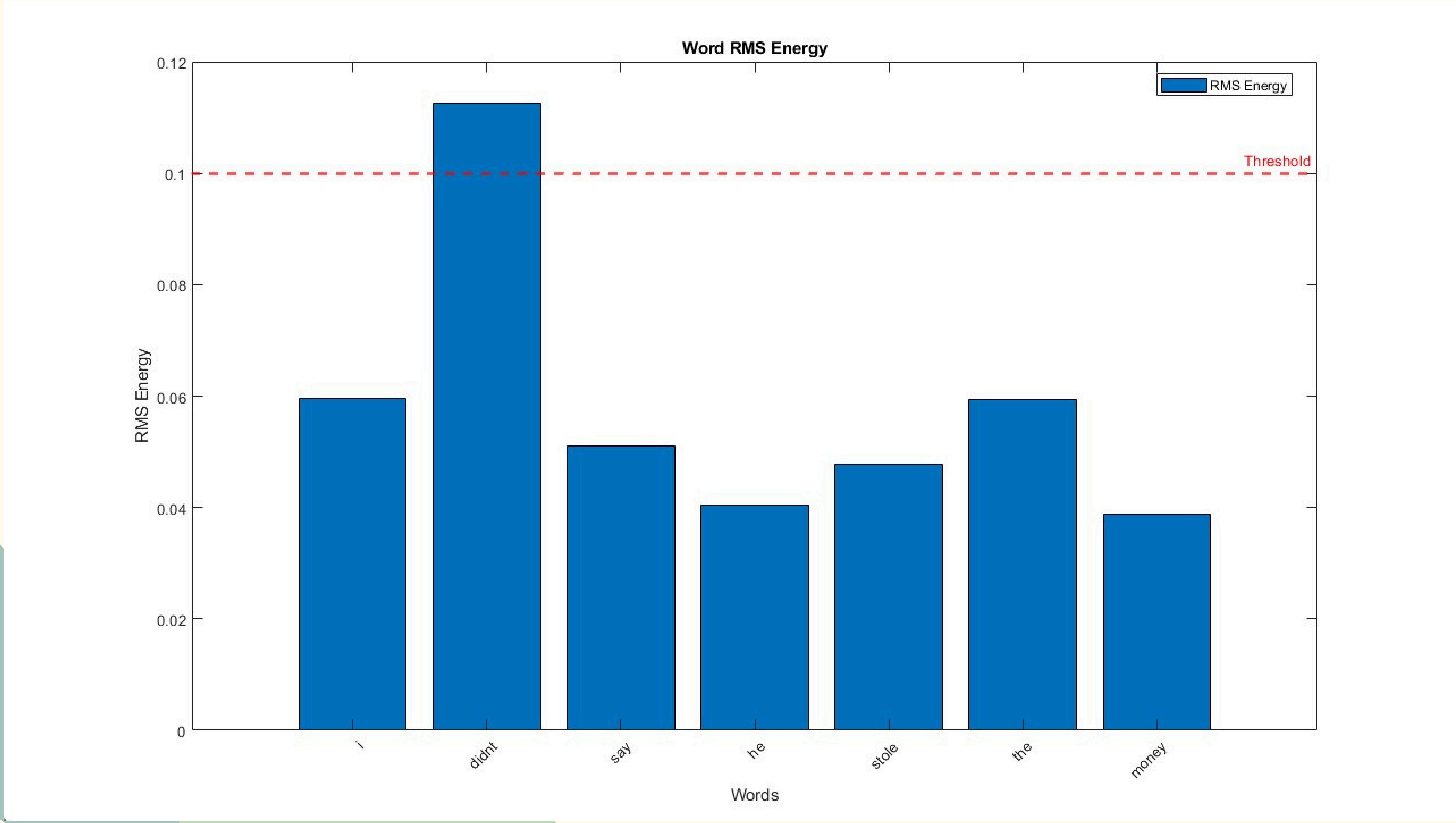
Word RMS Energy
RMS Energy
Threshold
RMS Energy
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
i
told
you
this
would
happen
Words

AUDIO-7



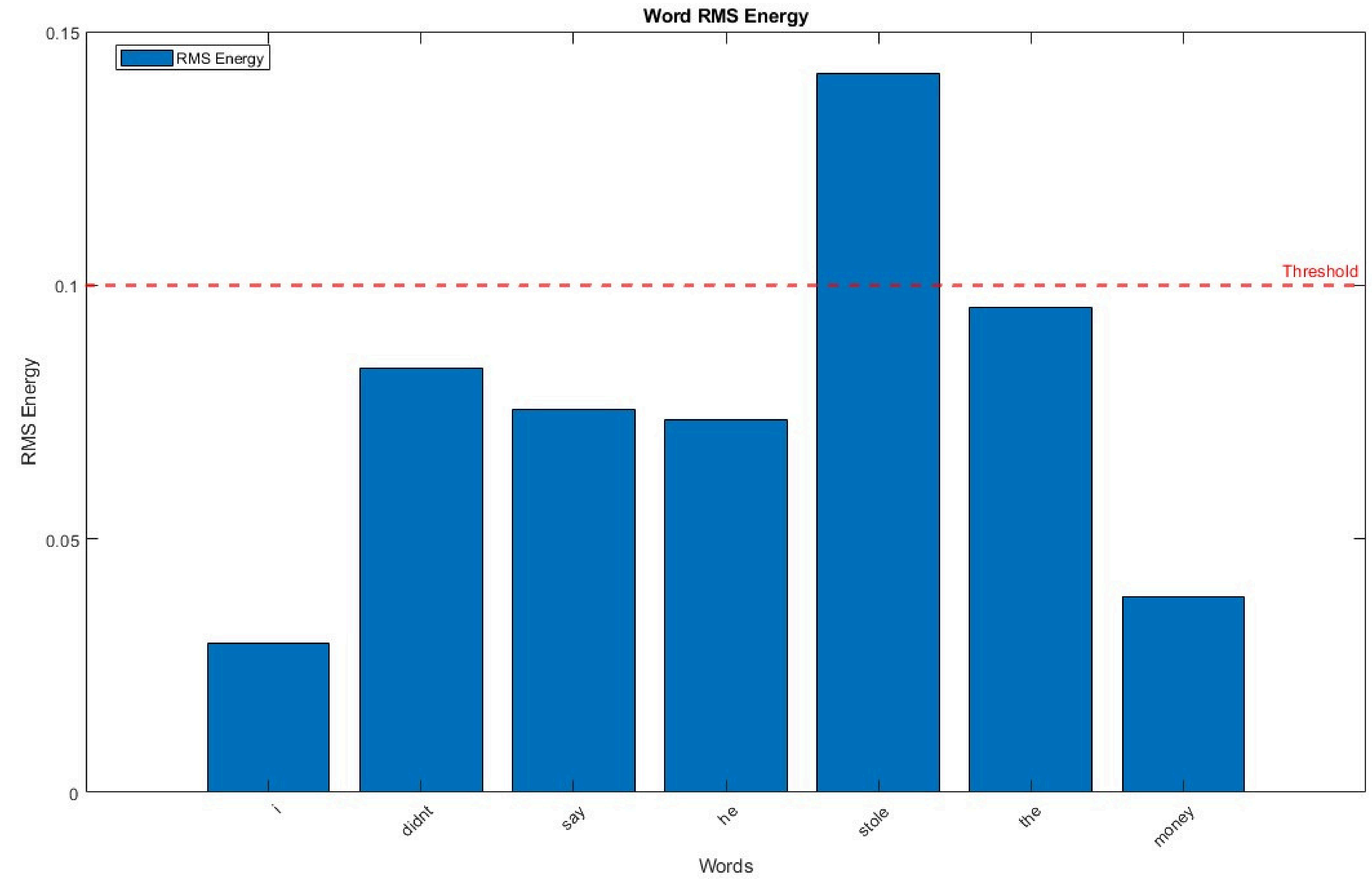
Word RMS Energy
RMS Energy
Threshold
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
RMS Energy
i
didnt
say
he
stole
the
money
Words

AUDIO-8



Word RMS Energy
RMS Energy
Threshold
0.12
0.1
0.08
0.06
0.04
0.02
0
RMS Energy
i
didnt
say
he
stole
the
money
Words

AUDIO-9

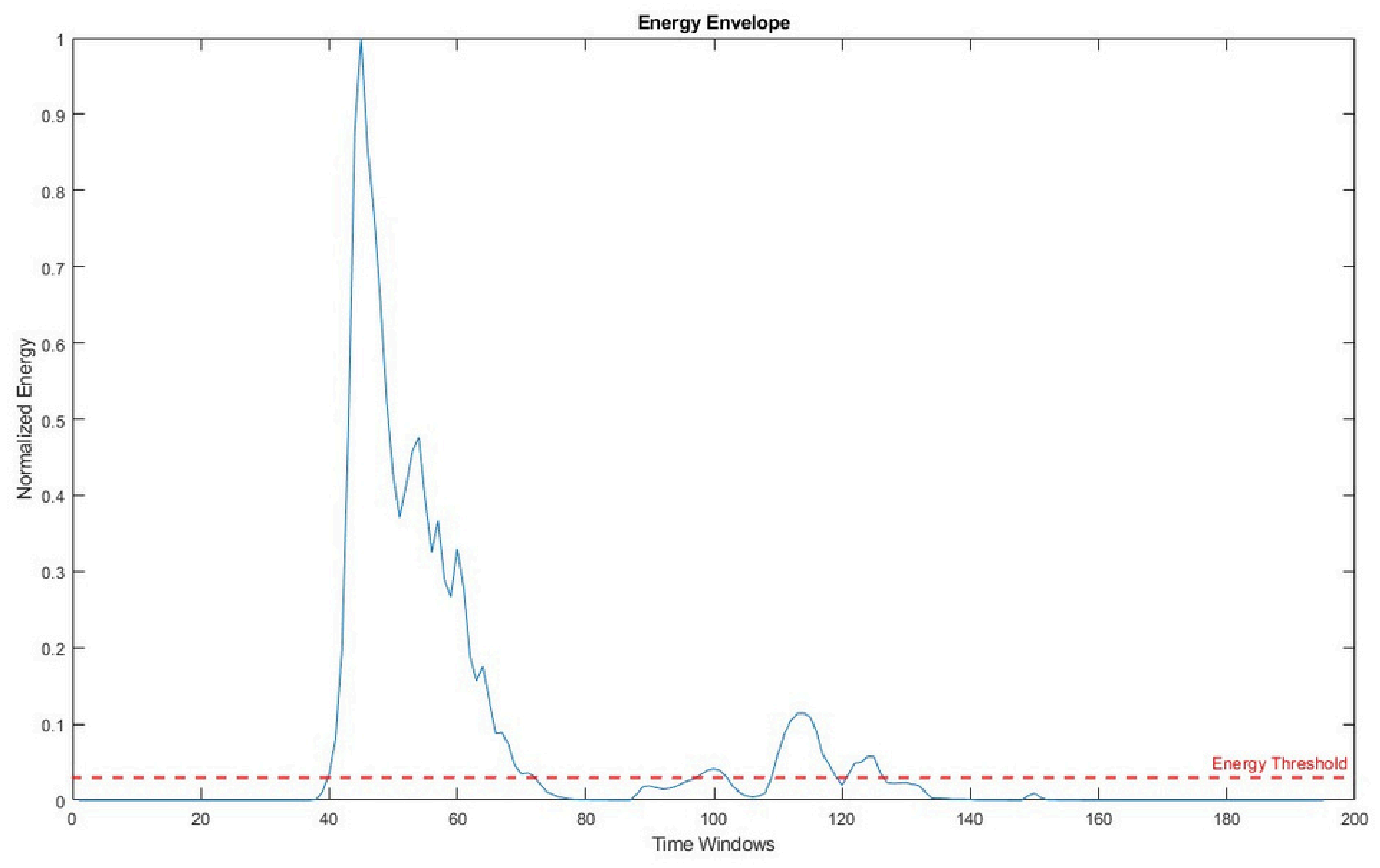


Word RMS Energy
RMS Energy
0.15
0.1
0.05
0
Threshold
i
didnt
say
he
stole
the
money
Words

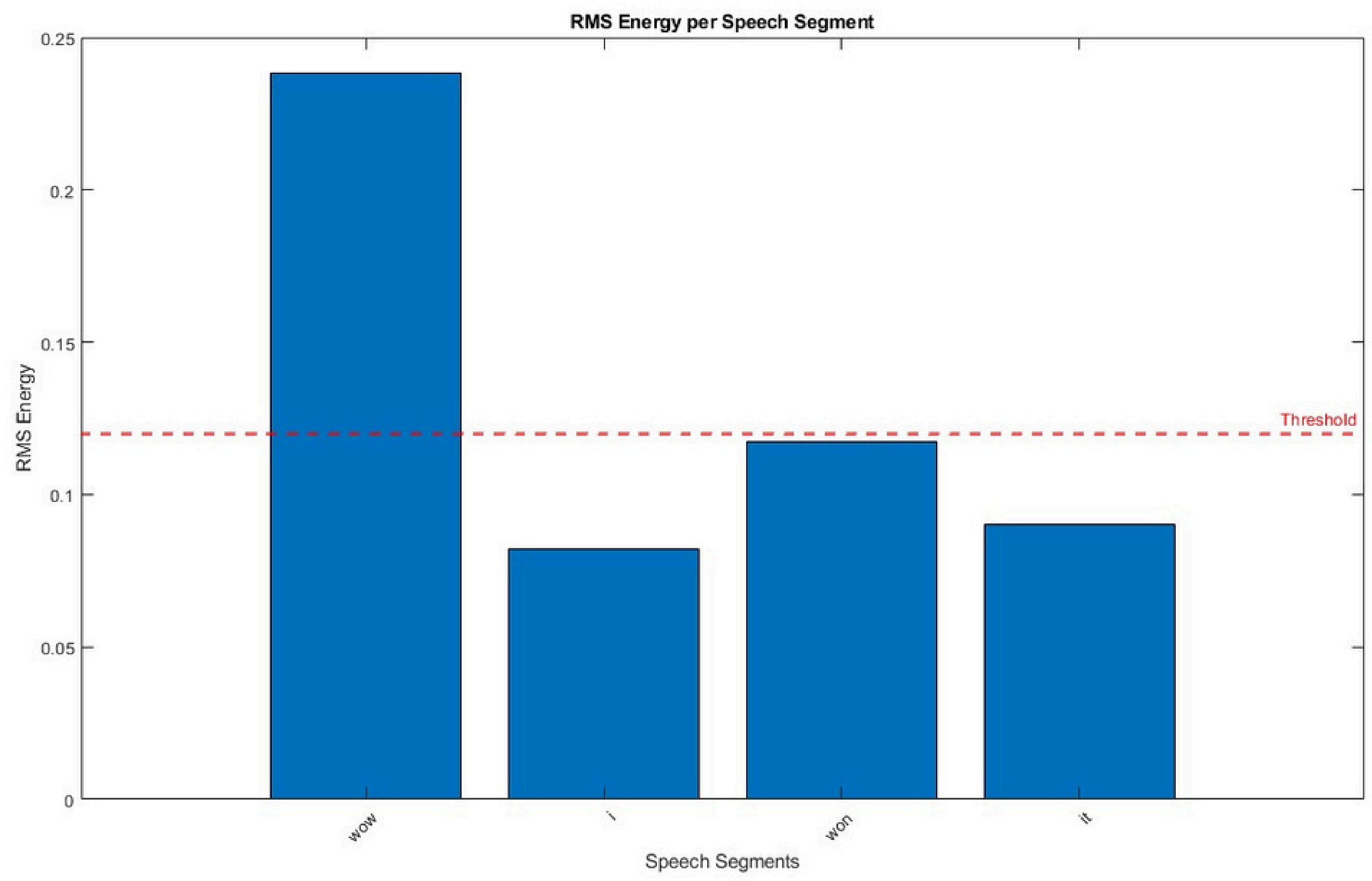
TASK-2

For task-2, we don't have the time stamps. So we calculate the energy of the signal by taking small 25ms windows and then set a low threshold to distinguish between speech region and silence region. Then we separate these segments and calculate their rms energies and repeat the steps from task-1 again

AUDIO-1

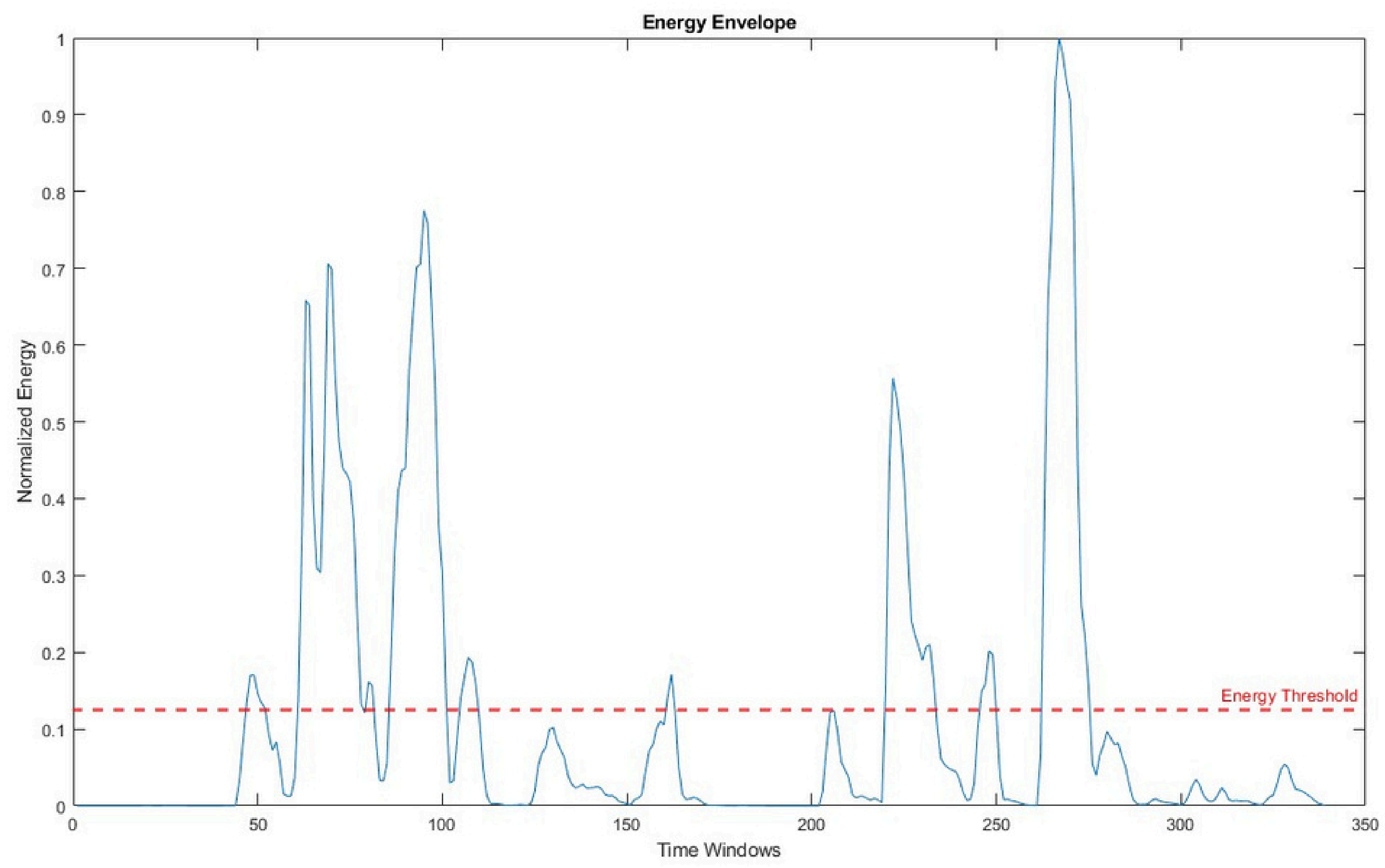


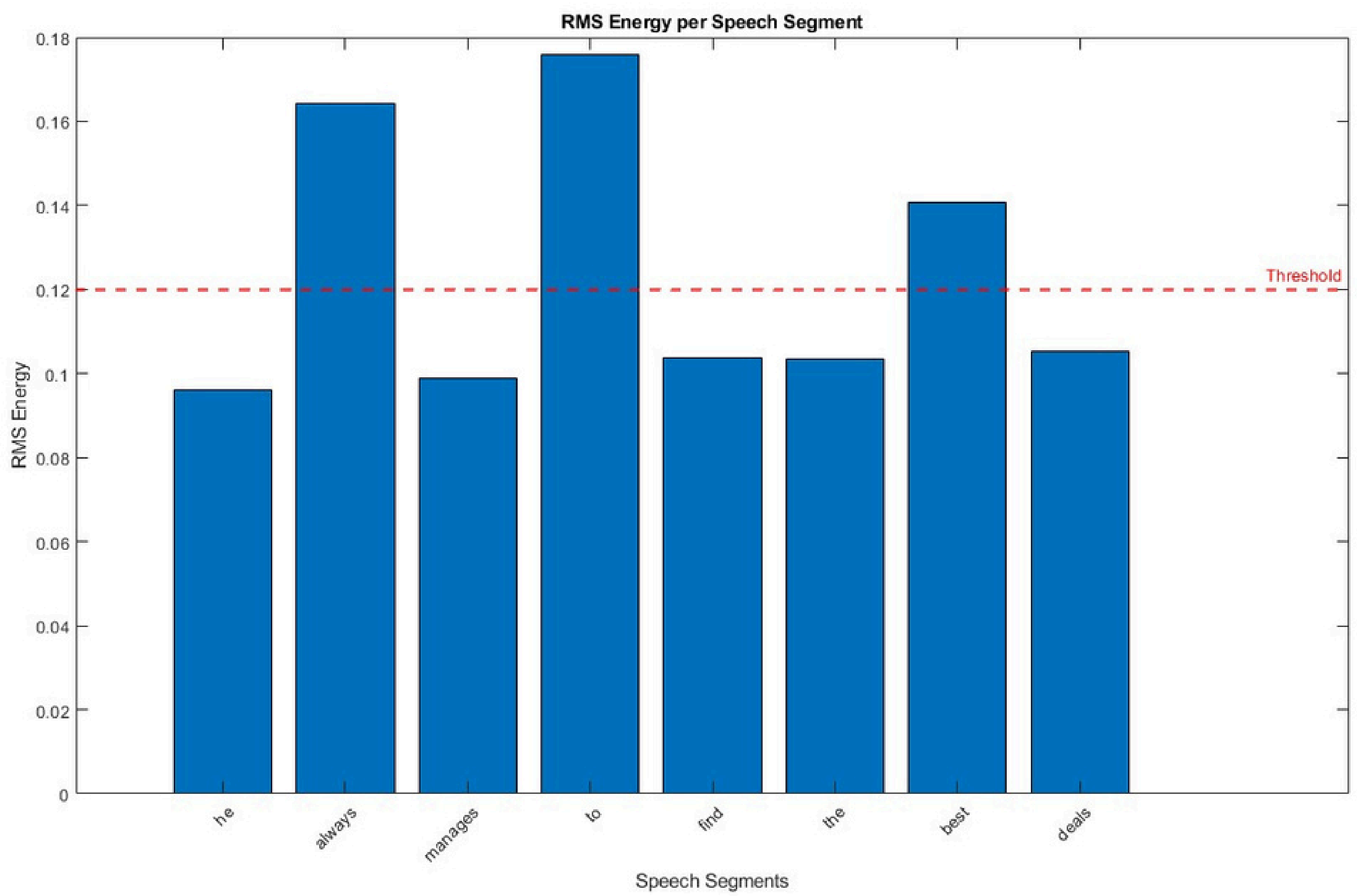
Energy Envelope
Normalized Energy
Time Windows
Energy Threshold
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
20
40
60
80
100
120
140
160
180
200



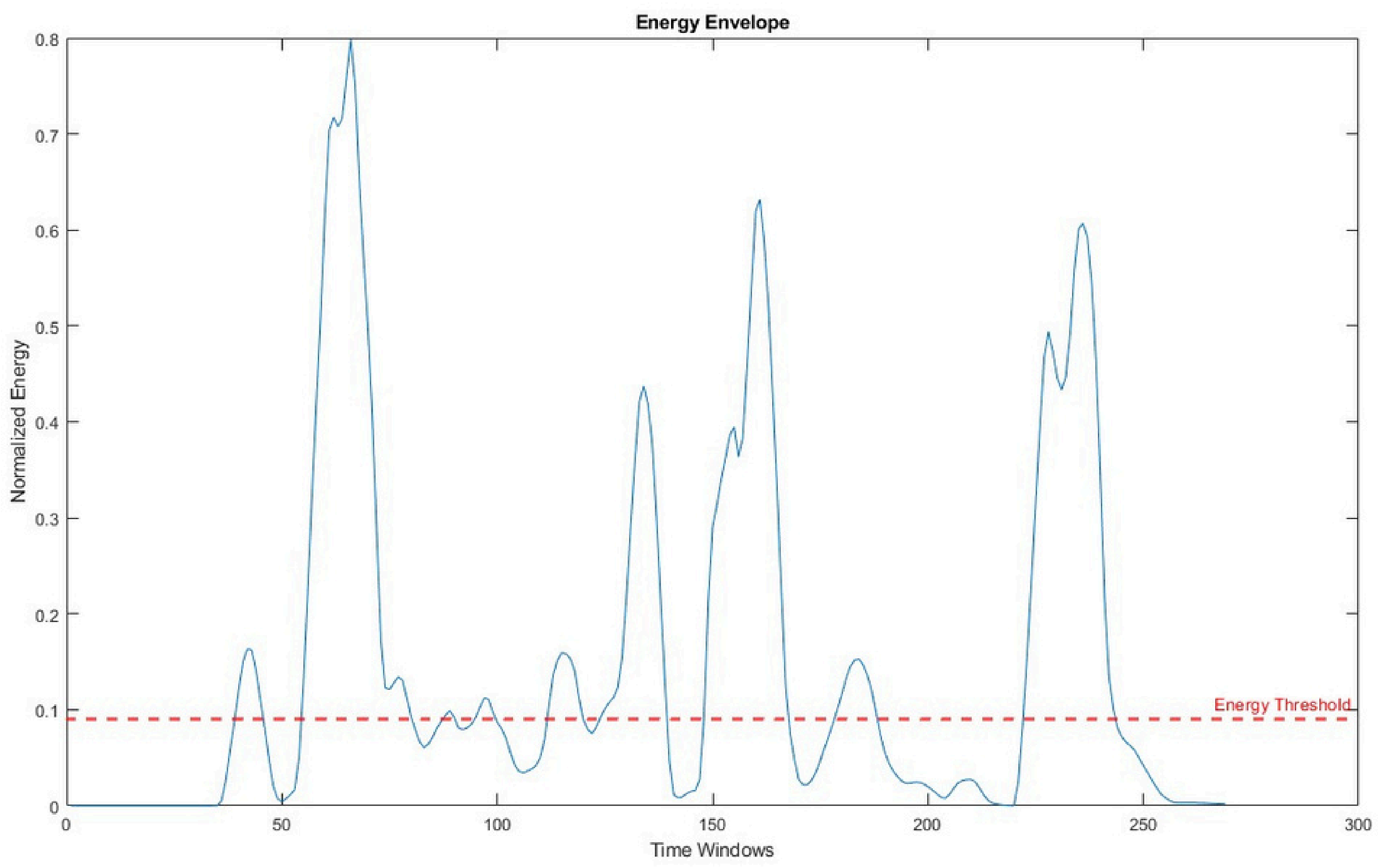
RMS Energy per Speech Segment
RMS Energy
Speech Segments
Threshold
0
0.05
0.1
0.15
0.2
0.25
wow
i
won
it

AUDIO-2

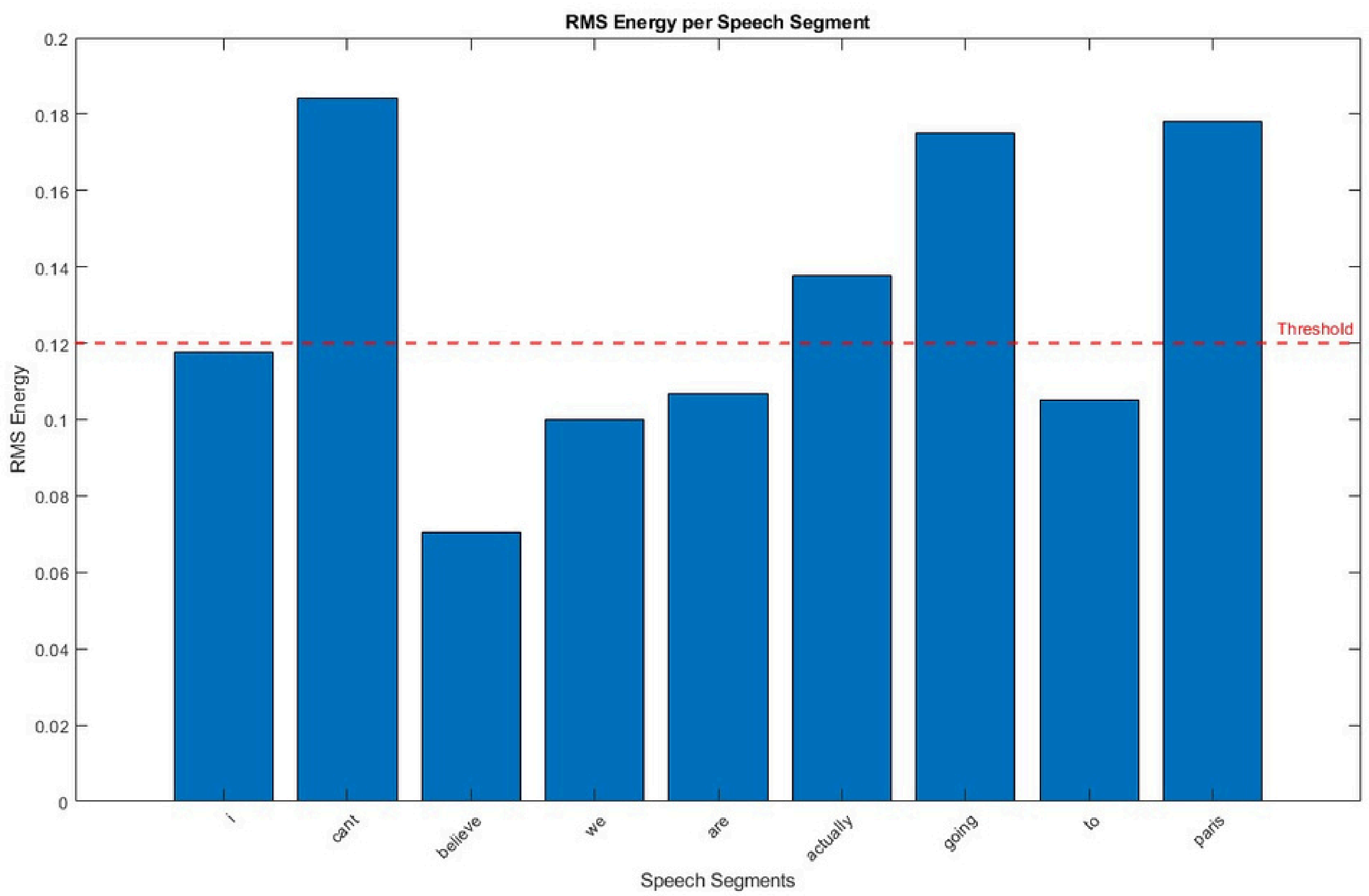

Energy Envelope
Normalized Energy
Time Windows
Energy Threshold
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
50
100
150
200
250
300
350


RMS Energy per Speech Segment
RMS Energy
Speech Segments
Threshold
0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
0.16
0.18
he
always
manages
to
find
the
best
deals

AUDIO-3

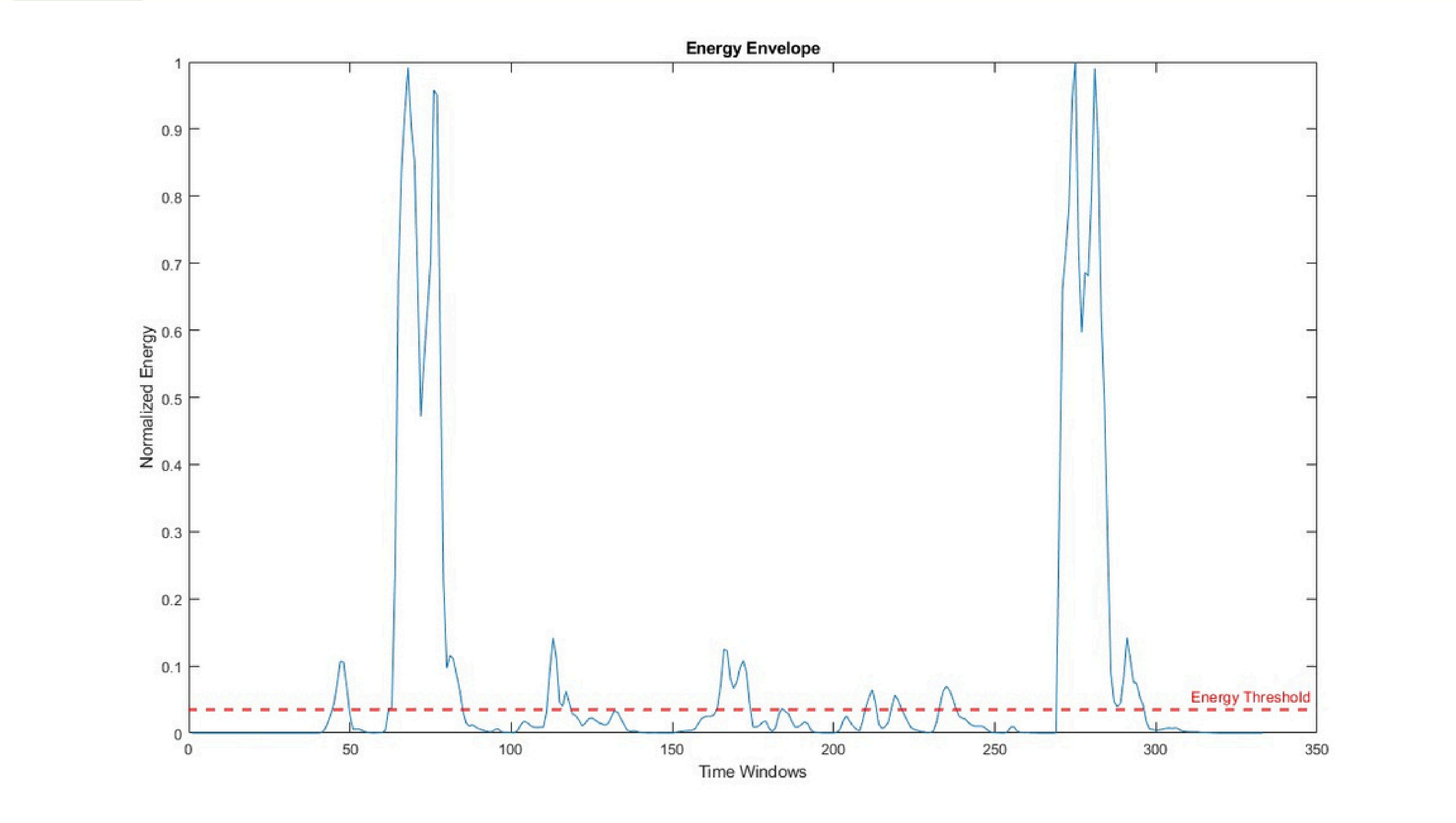


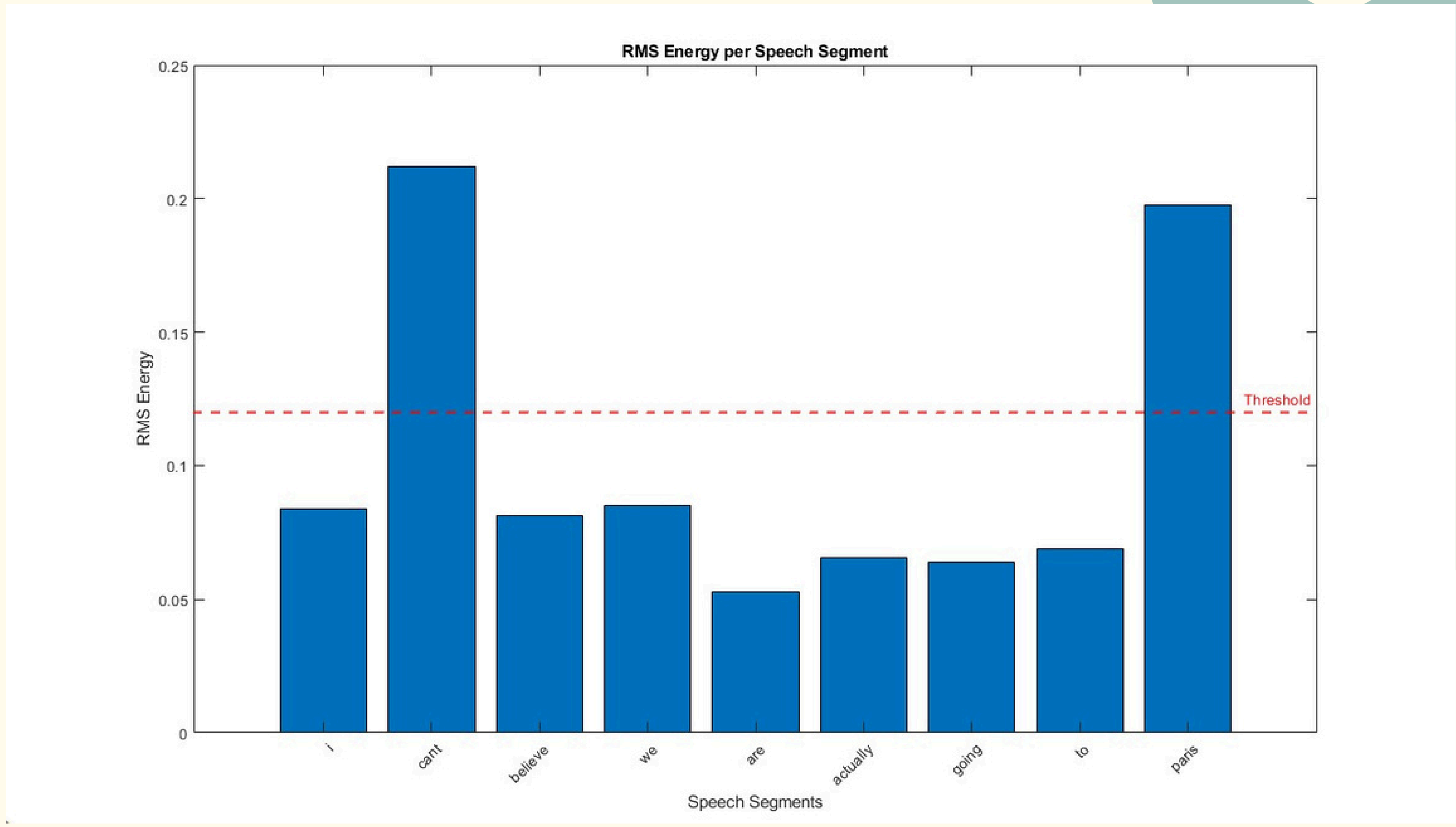
Energy Envelope
Normalized Energy
Time Windows
Energy Threshold
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
50
100
150
200
250
300



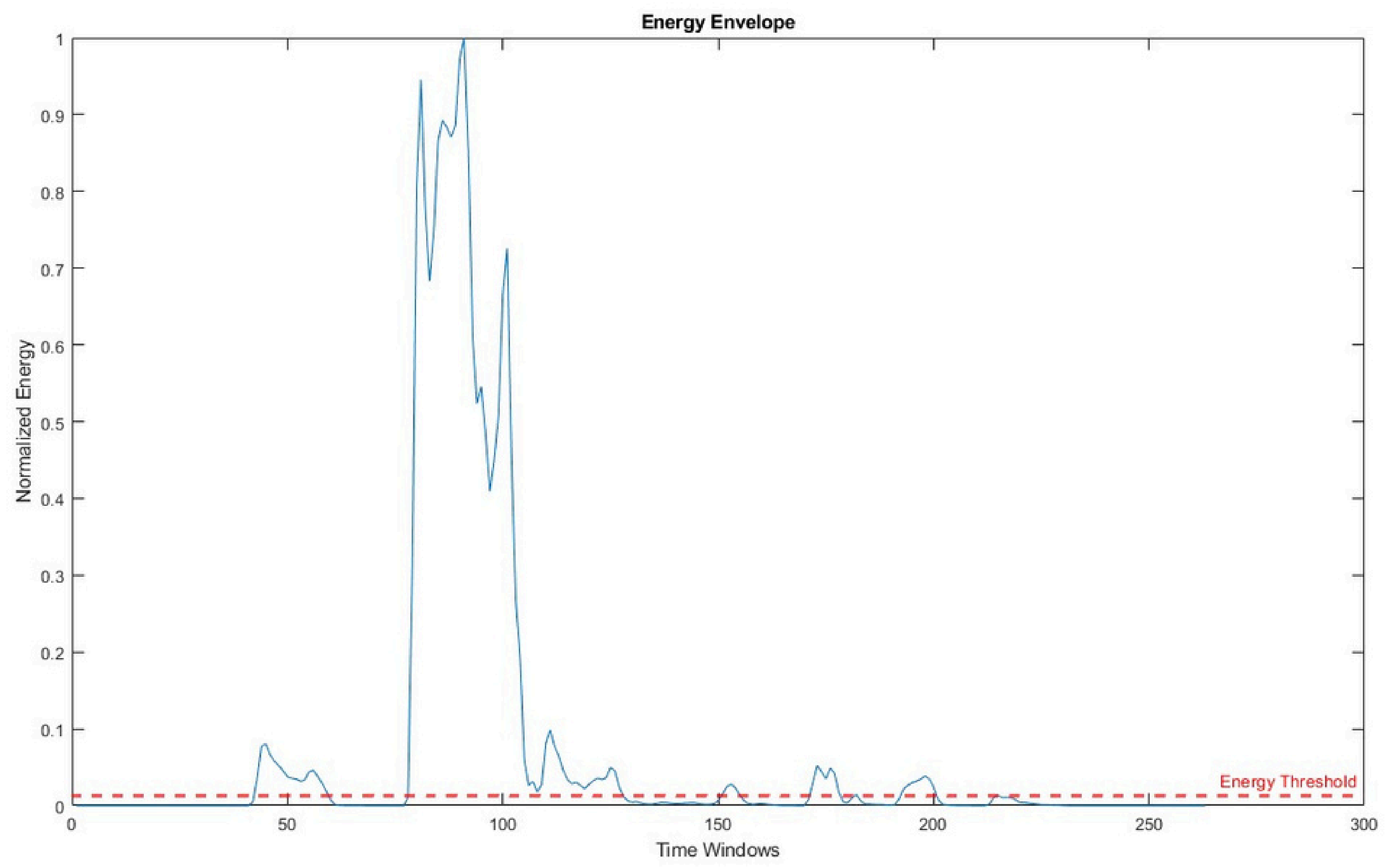
RMS Energy per Speech Segment
RMS Energy
Speech Segments
Threshold
0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
0.16
0.18
0.2
i
cant
believe
we
are
actually
going
to
paris

AUDIO-4

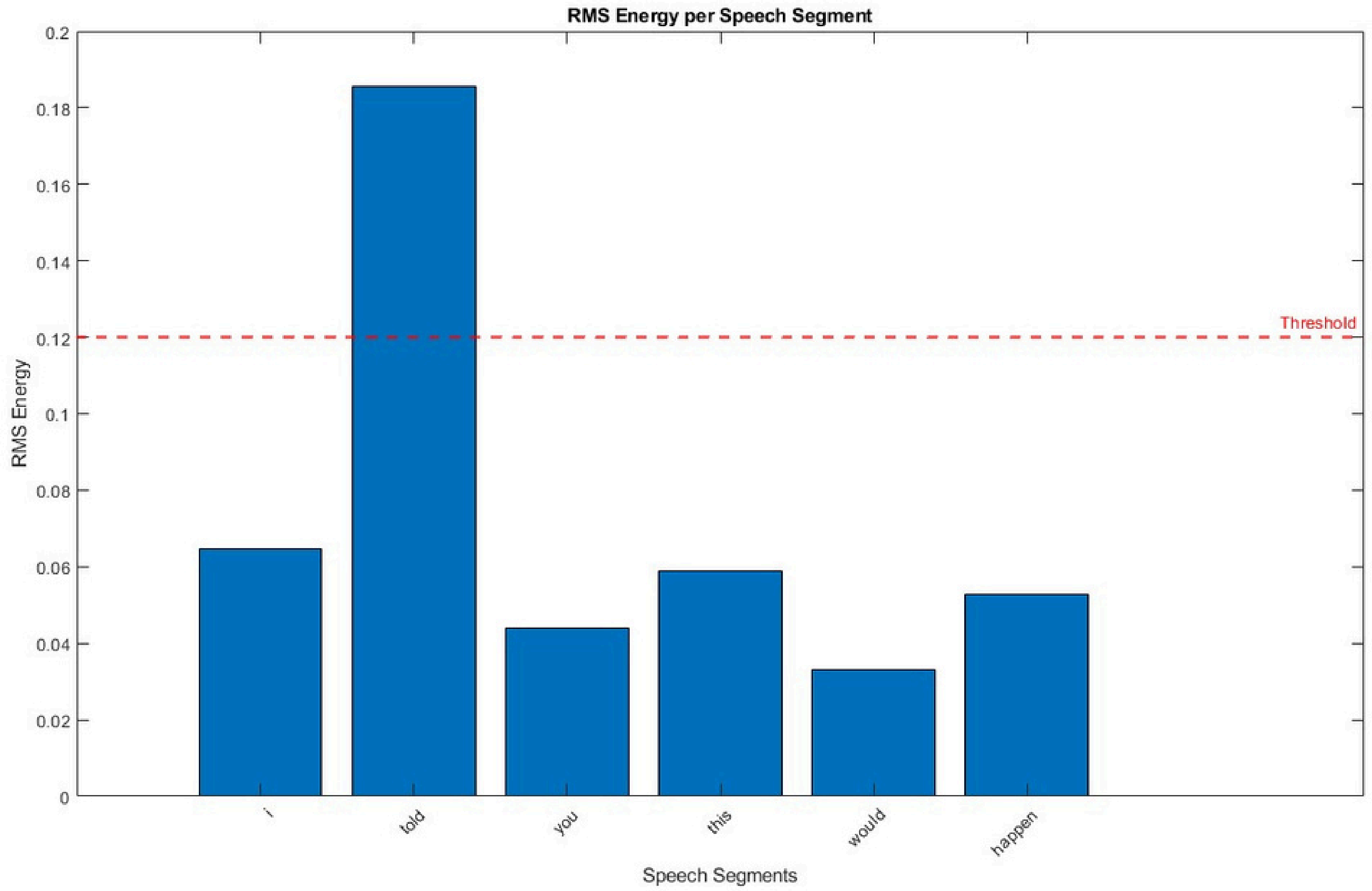

Energy Envelope
Normalized Energy
Time Windows
Energy Threshold
0
50
100
150
200
250
300
350
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


RMS Energy per Speech Segment
RMS Energy
Speech Segments
Threshold
0
0.05
0.1
0.15
0.2
0.25
i
cant
believe
we
are
actually
going
to
paris

AUDIO-5

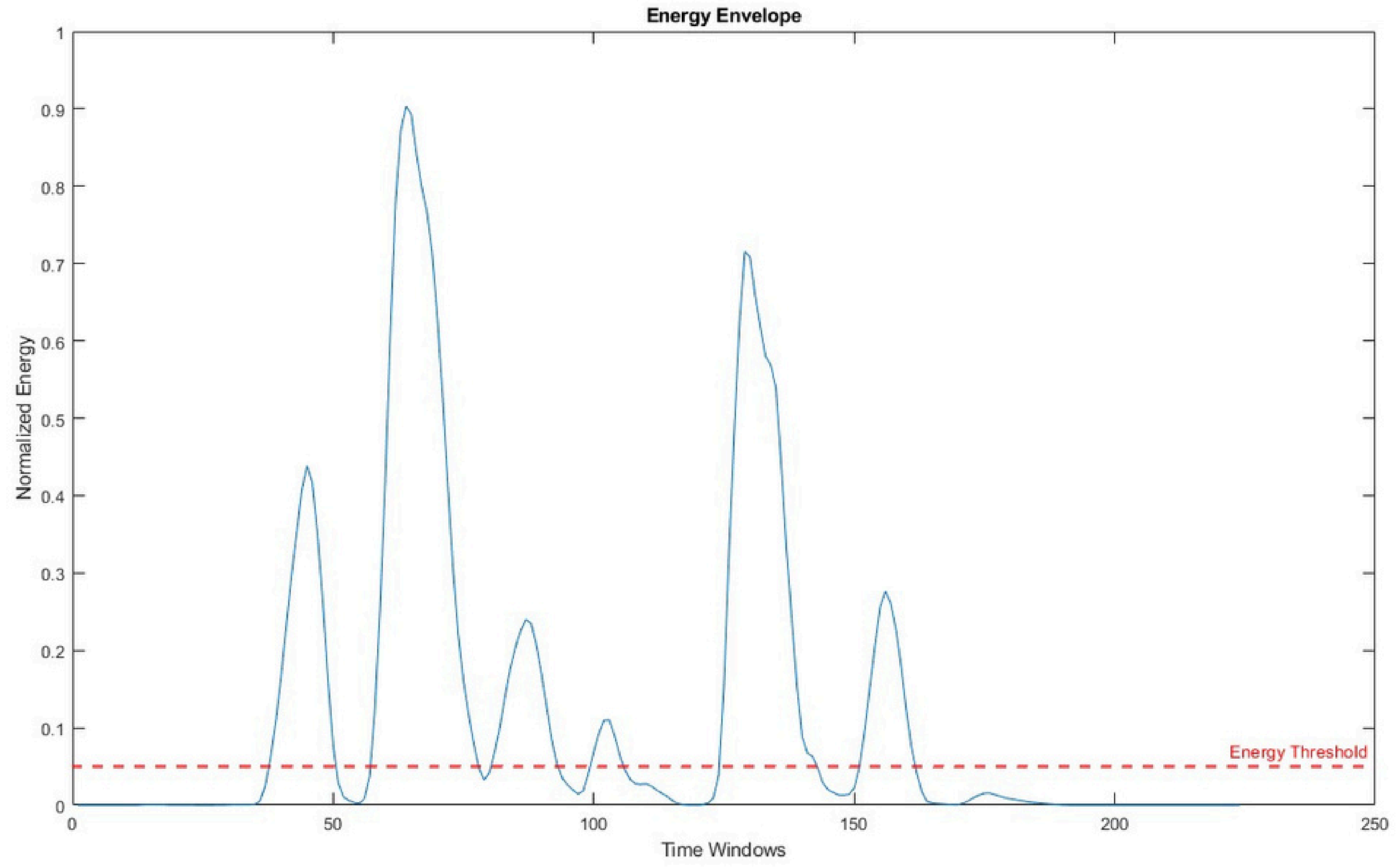


Energy Envelope
Normalized Energy
Time Windows
Energy Threshold
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
50
100
150
200
250
300

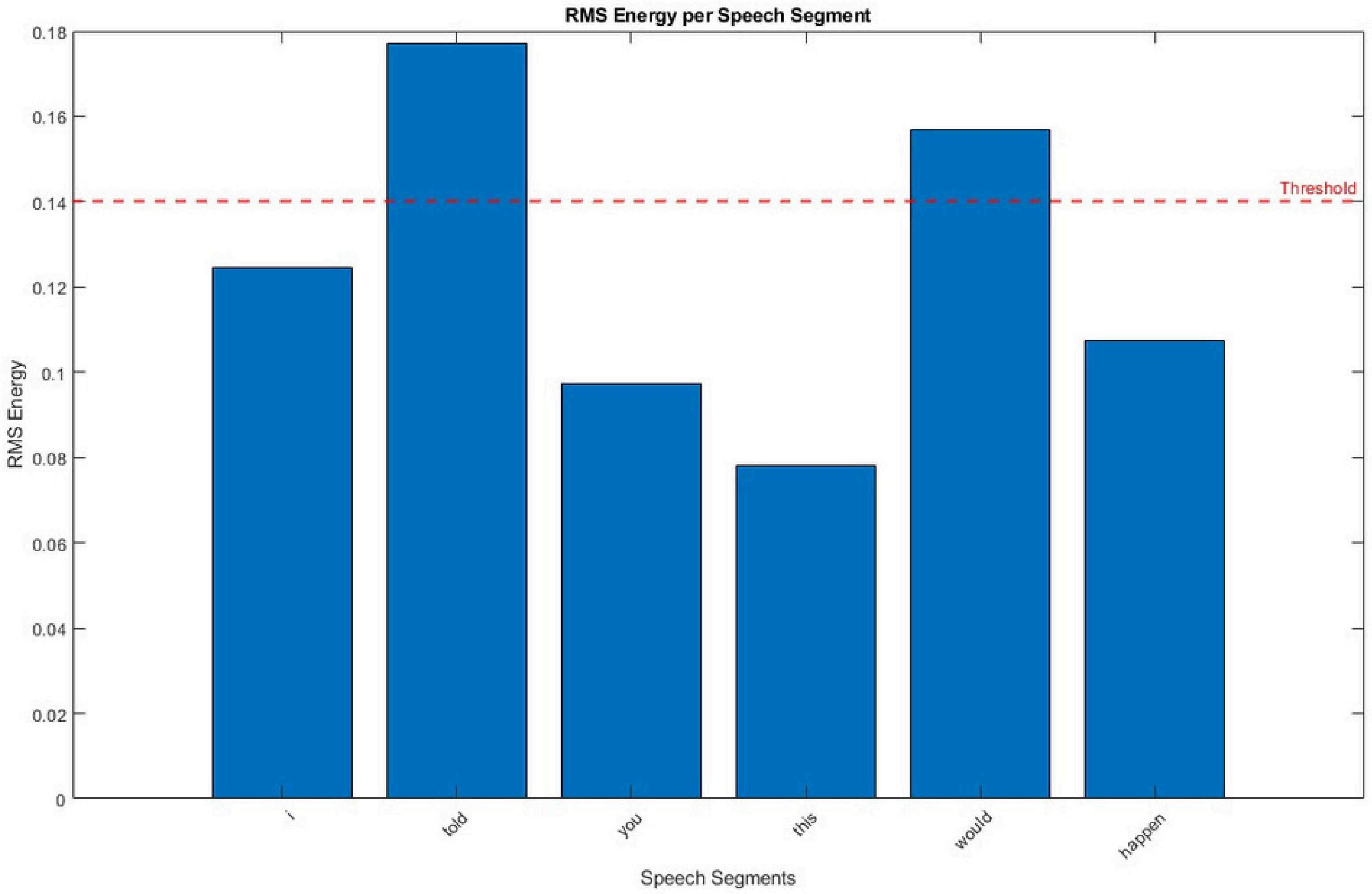


RMS Energy per Speech Segment
RMS Energy
Speech Segments
Threshold
0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
0.16
0.18
0.2
i
told
you
this
would
happen

AUDIO-6

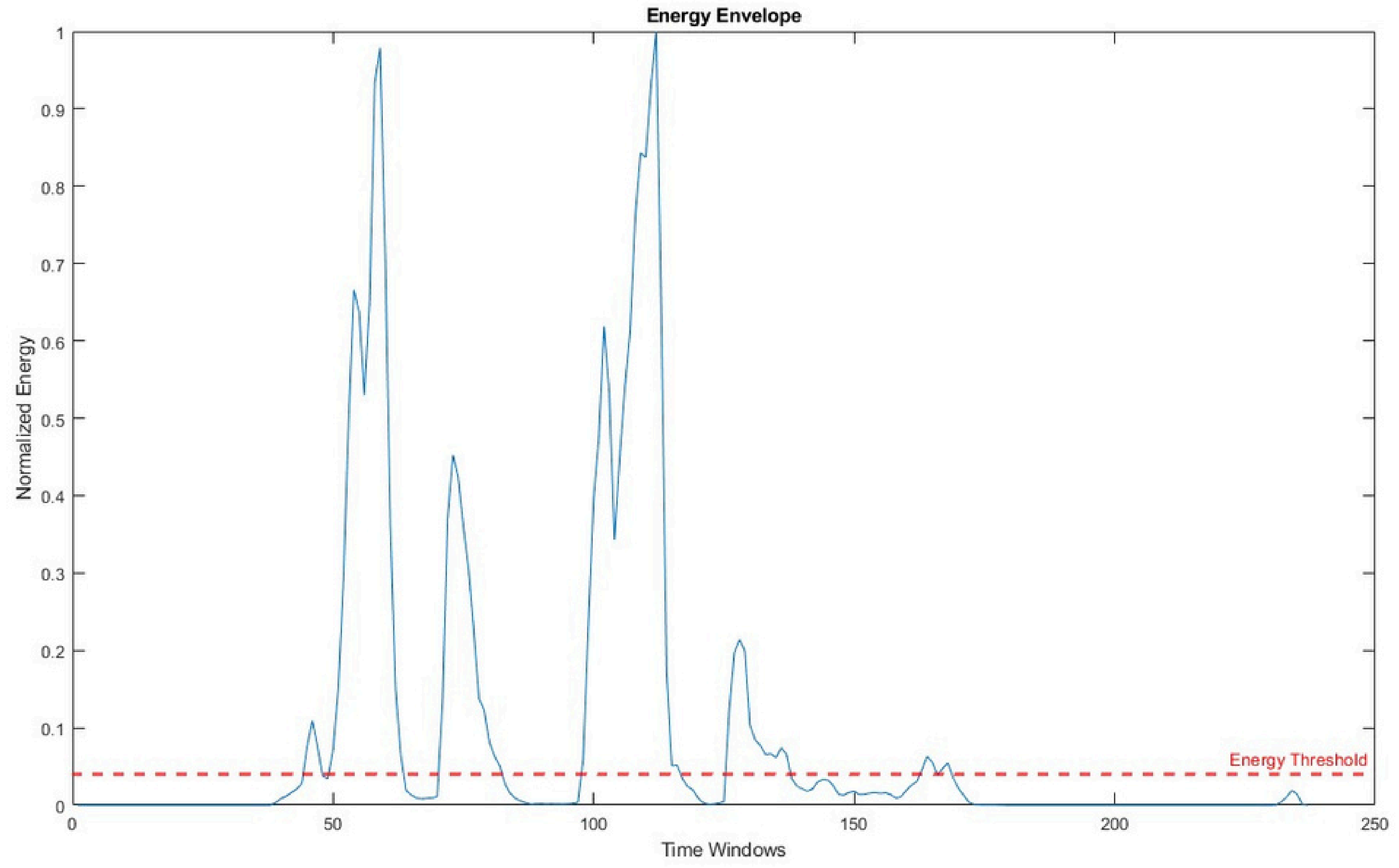


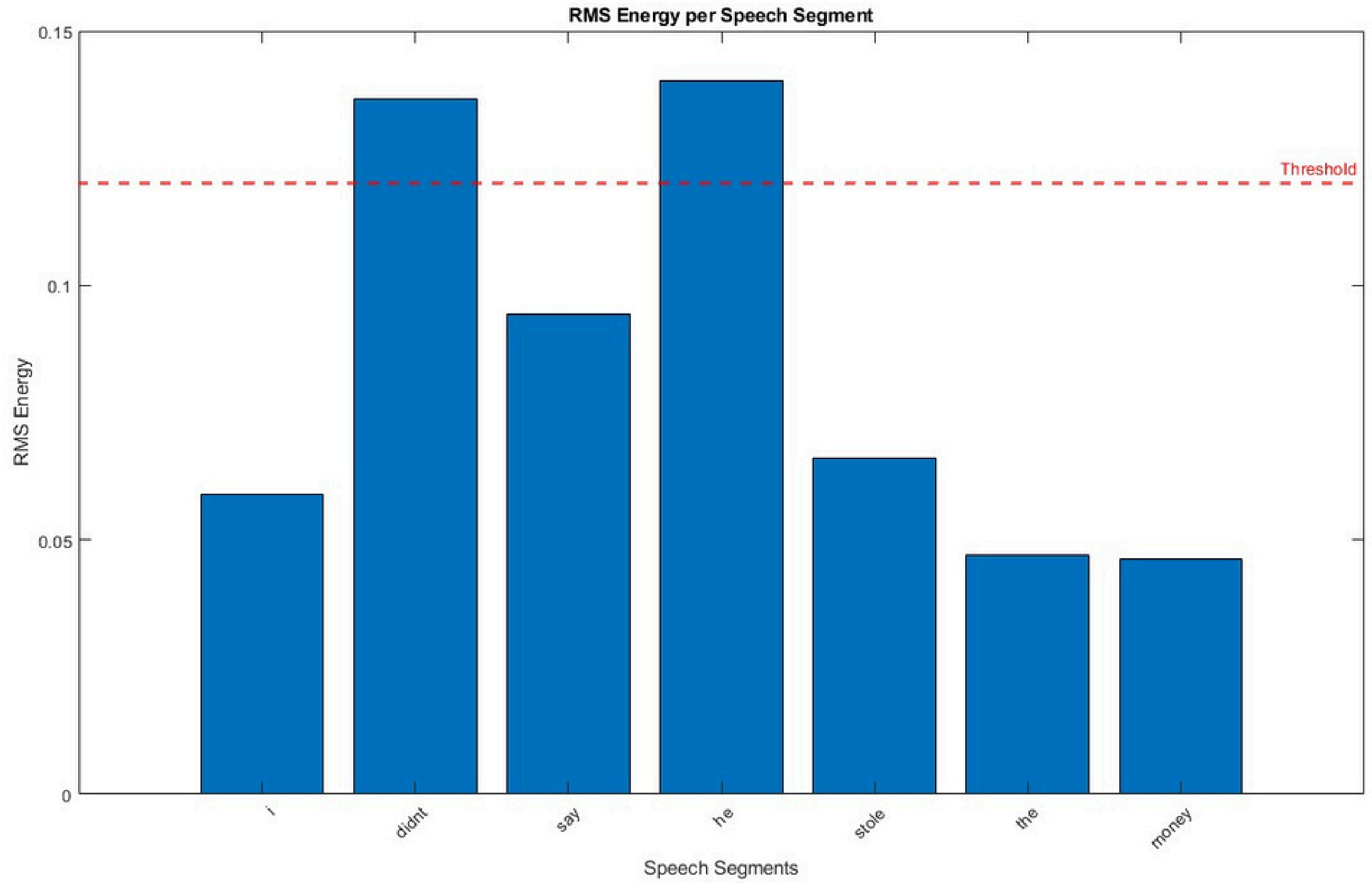
Energy Envelope
Normalized Energy
Time Windows
Energy Threshold
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
50
100
150
200
250



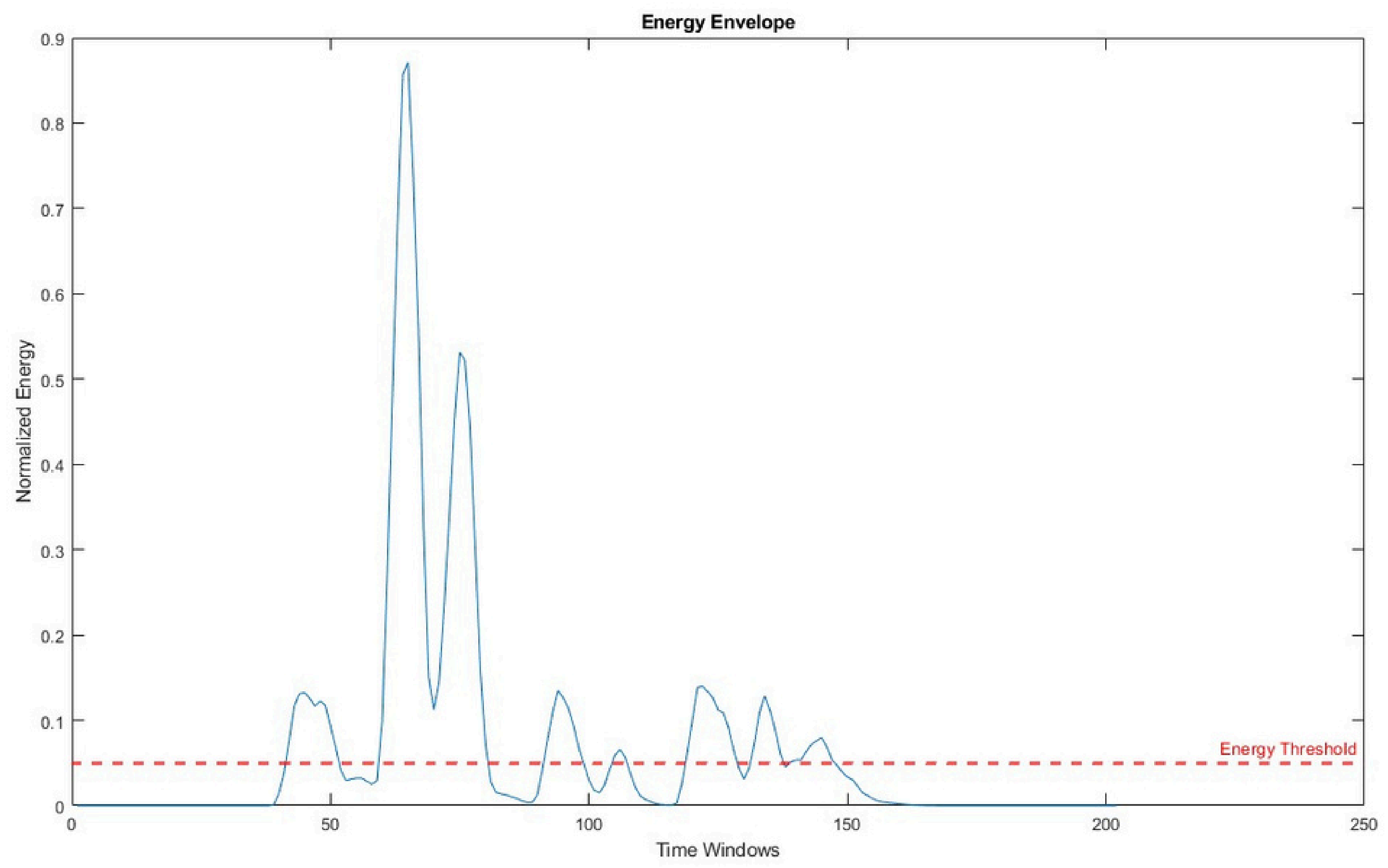
RMS Energy per Speech Segment
RMS Energy
Speech Segments
Threshold
0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
0.16
0.18
i
told
you
this
would
happen

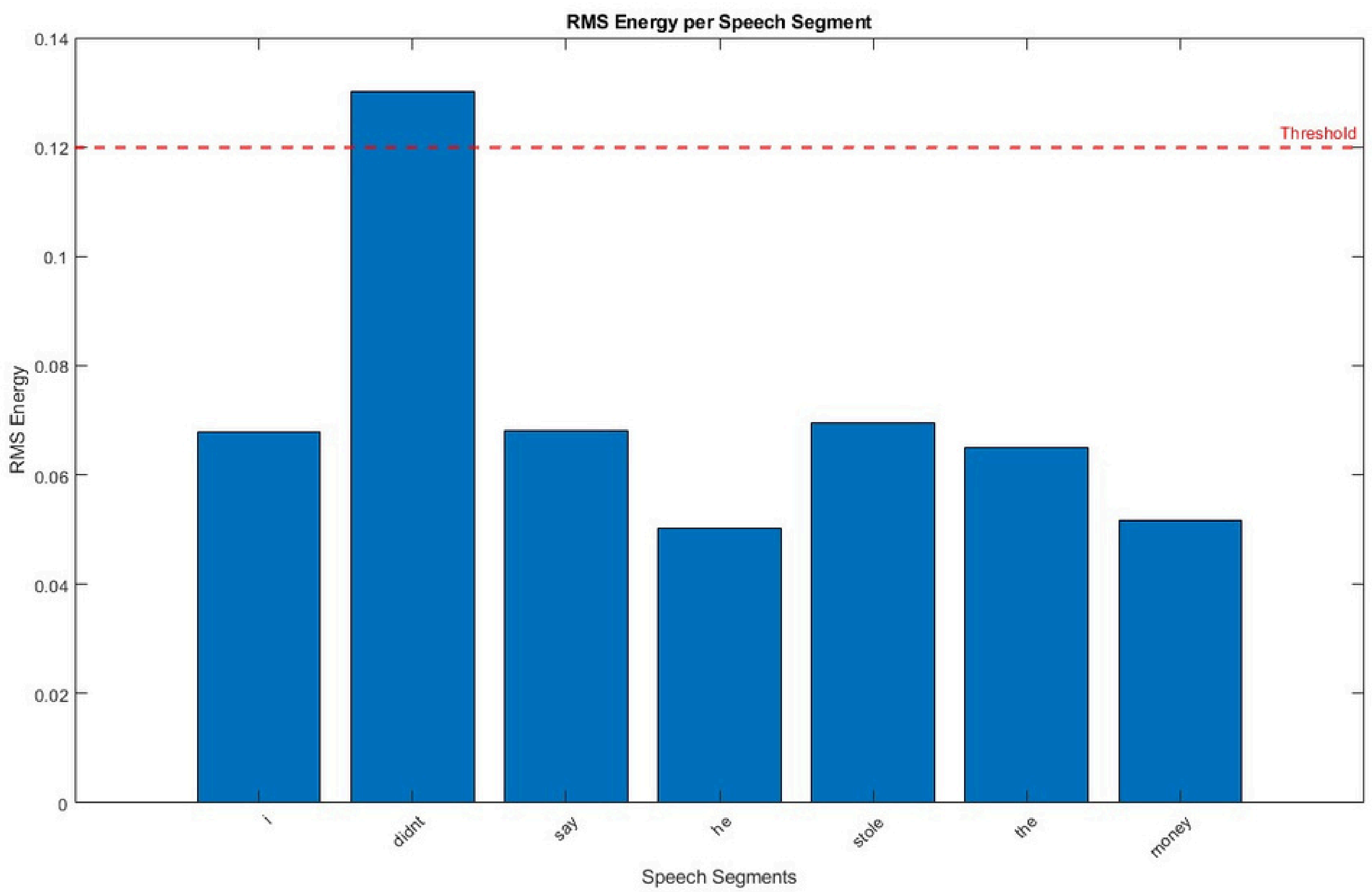
AUDIO-7


Energy Envelope
Normalized Energy
Time Windows
Energy Threshold
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
50
100
150
200
250

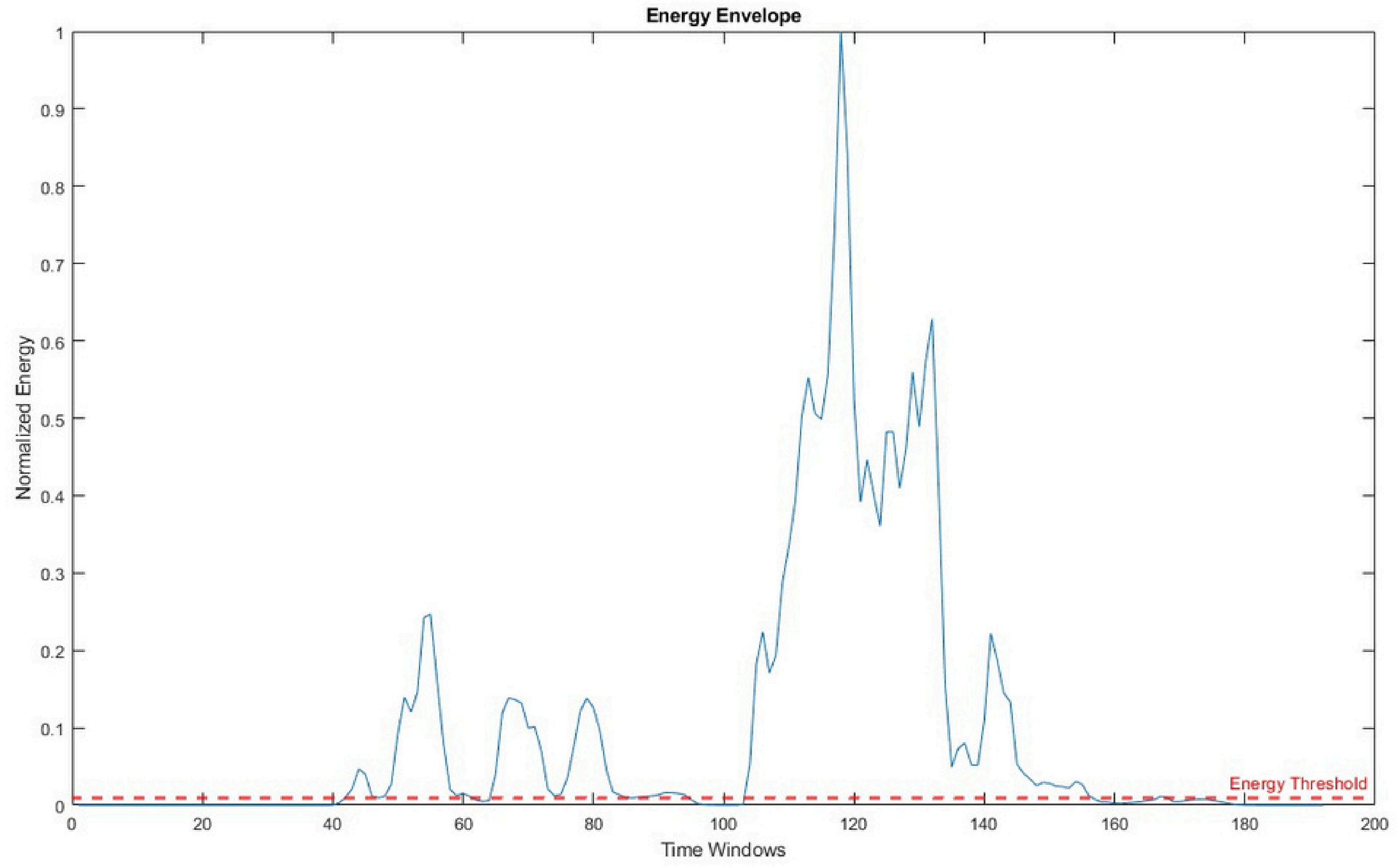

RMS Energy per Speech Segment
RMS Energy
Speech Segments
Threshold
0
0.05
0.1
0.15
i
didnt
say
he
stole
the
money

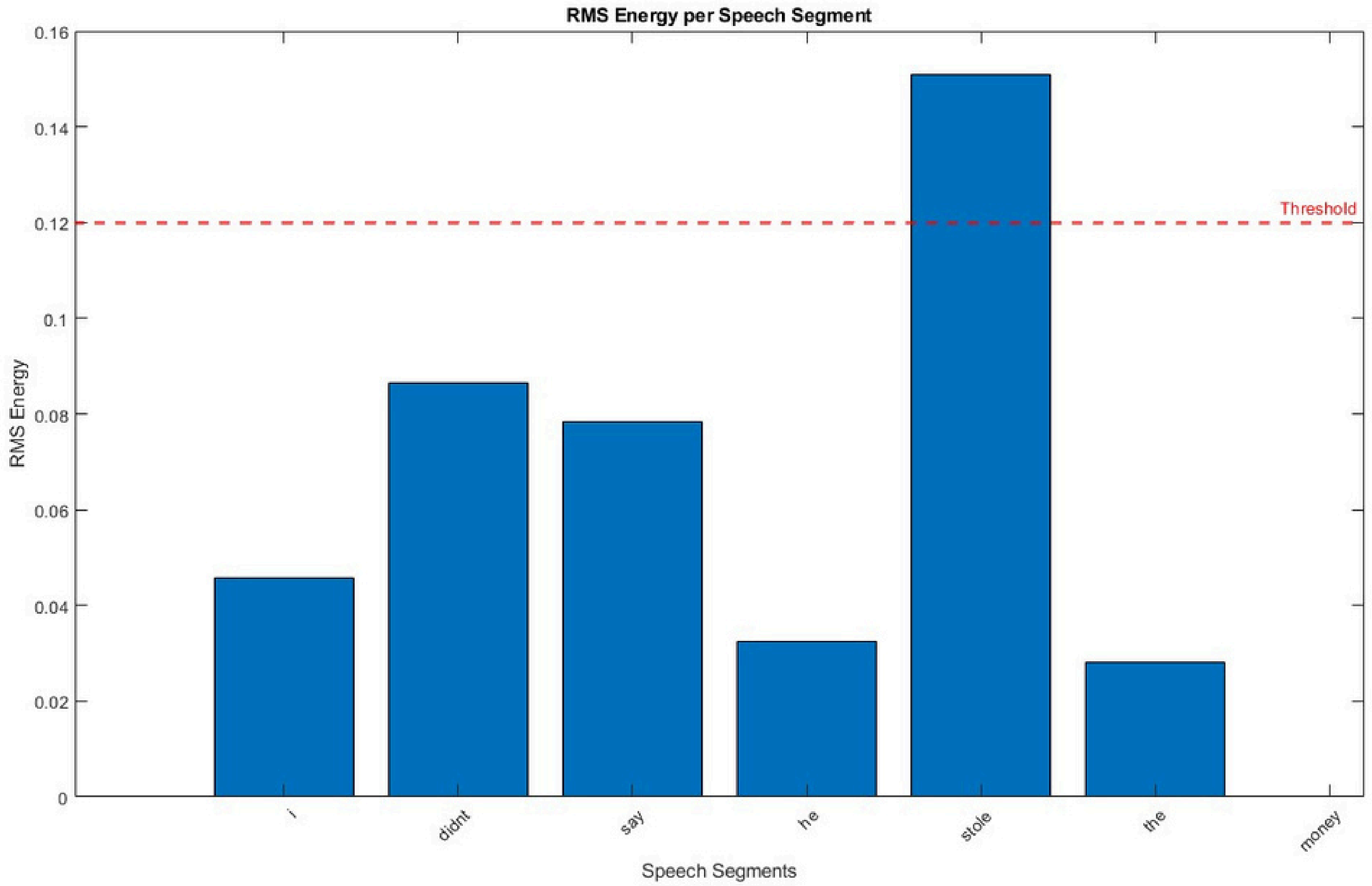
AUDIO-8


Energy Envelope
Normalized Energy
Time Windows
Energy Threshold
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
50
100
150
200
250


RMS Energy per Speech Segment
RMS Energy
Speech Segments
Threshold
0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
i
didnt
say
he
stole
the
money

AUDIO-9


Energy Envelope
Normalized Energy
Time Windows
Energy Threshold
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
20
40
60
80
100
120
140
160
180
200


RMS Energy per Speech Segment
RMS Energy
Speech Segments
Threshold
0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
0.16
i
didnt
say
he
stole
the
money

LIMITATIONS

1. Fixed Thresholds

- Problem: The method uses fixed thresholds for energy (energy_threshold) and loudness (rms_threshold), which are not adaptable to varying background noise levels or audio characteristics.
- Impact: This could lead to:
  - False Positives: Background noise exceeding the energy threshold is detected as speech.
  - False Negatives: Low-energy speech or words may not be detected.

2. Overlap Between Speech and Silence

- Problem: Energy-based methods struggle to handle overlapping regions where speech transitions into silence or vice versa.
- Impact: Results in inaccurate segmentation boundaries, leading to split or combined segments.

3. Lack of Spectral Information

- Problem: The method uses energy, which is a time-domain feature, and ignores spectral characteristics of the signal.
- Impact:
  - Cannot distinguish between speech and similar energy non-speech sounds (e.g., music, claps, or noise).
  - Sensitive to variations in speaker pitch or tonal differences.

THANK YOU