

מעבדת עיבוד אות דיבור

פרוטוקול המעבדה 2021-2022

מטרות

- הקניית ידע בסיסי בתכונות אות דיבור.
- הכרת כלים לעיבוד אות דיבור.
- הכרת מודלים מתמטיים לייצוג אות דיבור.
- מימוש אלגוריתם סגמנטציה.
- בניית מערכת לומדת המסוגלת להבחין בין שני דוברים מוכרים.

מידע חשוב

- במסגרת מעבדה זו תיבנו באמצעות Matlab מערכת לומדת, אשר מסוגלת להבדיל בין שני דוברים מוכרים. דוברים אלה יהיו את/ם – צוות המעבדה.
- מעבדה זו איננה כוללת את המבנה הרגיל של כתיבת דו"ח מכין, וכתיבת דו"ח מסכם. מבנה הדוח וחלוקת הציון מפורטים במסמך "הנחיות לכתיבת דוחות למעבדות לא פרונטליות" הנמצא במודל.
- יש לבצע את ההוראות המפורטות במסמך זה לצורך בניית המערכת. בנוסף, לאורך התהליך תפיקו דו"ח הכולל מענה לשאלות, מימוש אלגוריתמים והפקת גרפים. יש לשים דגש על פירוט במענה לשאלות, בתיעוד האלגוריתמים (פירוט מתחת לכותרות הפונקציות לגבי תפקידה, משתני קלט/פלט וכו') ובהצגת המידע הגרפי (כותרות, יחידות וכו').
- בתאריך ה- 27.12.2021 נקבע שעות קבלה למענה על שאלות.
- בתאריך ה- 11.01.2022 ניפגש בזום. כל צוות יציג את המערכת שבנה ואת תוצאותיה בזיהוי חברי הצוות.
- מומלץ לעבוד על פי סדר הסעיפים המוצגים במסמך זה, שכן רוב הסעיפים מבוססים על הקודמים להם.

ספרות

1. Markel, John D. "The SIFT algorithm for fundamental frequency estimation." *Audio and Electroacoustics, IEEE Transactions on* 20.5 (1972): 367-377.
2. Deller Jr, John R., John G. Proakis, and John H. Hansen. *Discrete time processing of speech signals*. Prentice Hall PTR, 1993.
3. Fallside, Frank, and William A. Woods. *Computer speech processing*. Prentice Hall International (UK) Ltd., 1986.
4. Rabiner, Lawrence R., and Ronald W. Schafer. *Digital processing of speech signals*. Prentice Hall, 1978.
5. Sörnmo, Leif, and Pablo Laguna. *Bioelectrical signal processing in cardiac and neurological applications*. Vol. 8. Academic Press, 2005.

רקע תאורטי

תארו בקצרה את המערכת הפיזיולוגית האחראית ליצירת אות הדיבור, ואת איברי הדיבור המשתתפים בתהליך יצירת הדיבור.

באילו תדרים ניתן לדגום (להקליט באופן דיגיטלי) אות דיבור אנושי? מהם היתרונות ומהם החסרונות של תדרי הדגימה השונים?

מה תהיה ההשפעה של תדר דגימה נמוך מאד על אות הדיבור? כיצד ישמע הדיבור?

חלוקת הניקוד על כל חלק מפורטת בכותרת החלק, סה"כ המענה לשאלות הינו 80% מציון הדות.

1. עיבוד מקדים לאות הדיבור (5)

לפניכן/ם 'סכמת מלבנים' של עיבוד סיפרתי מקדים (Pre-processing) לאות הדיבור:



איור 1: עיבוד מקדים לאות דיבור

1.1. הסבירו כל שלב בסכמה זו – מדוע נחוץ וכיצד מבוצע. עבור שלב 'מסנן מדגיש' הציגו את נוסחת המסנן ומפת קטבים ואפסים. עבור שלב 'הכפלה בחלון' הציגו חלון כלשהו התואם את נימוקין/ם.

1.2. כתבו פונקציה שכותרתה:

```
function [ProcessedSig, FramedSig]=PreProcess (Signal, Fs, alpha, WindowLength, Overlap)
```

פונקציה זו תבצע את שלבי העיבוד המקדים כפי שהם מוצגים בסכמת המלבנים.

קלט:

- Signal – the raw speech signal
- Fs – sampling frequency
- alpha – pre-emphasis filter parameter
- WindowLength – window length [seconds]
- Overlap – percentage of overlap between adjacent frames [0-100]

פלט:

- ProcessedSig – the preprocessed speech signal
- FramesSig – a matrix of the framed signal (each row is a frame)

* לצורך מימוש פונקציה זו ניתן, כמובן, לעשות שימוש בפונקציות מובנות של Matlab. בנוסף – מומלץ להשתמש בפונקציה המצורפת 'enframe.m' (שימו לב להבדלים שבין הפרמטר Overlap לבין הפרמטר המקביל, אותו מקבלת הפונקציה המצורפת).

2. סגמנטציה אוטומטית (15)

2.1. הסבירו כיצד פועלת שיטת הסגמנטציה האוטומטית המתוארת בספר הקורס 'עבוד אותות פיזיולוגיים' (מקור מס' 5 ברשימת הספרות במסמך זה), תחת הסעיפים:

3.5 EEG Segmentation, 3.5.1 Spectral Error Measure – The Periodogram Approach

2.2. על מנת לדייק את תהליך הסגמנטציה, עלינו 'לסמן' היכן באות נאמרה המילה. כלומר, נתעלם ממקטעי השקט בתחילת וסוף ההקלטה. כתבו פונקציה שכותרתה:

```
Idx = FindWordIdx(FramedSig, Fs, WindowLength, Overlap)
```

פונקציה זו תחשב עבור כל מסגרת את אנרגיית הסיגנל (ניתן להיעזר בפונ' מסעיף 5.2), ותיקח בחשבון רק את המסגרות שחצו סף אנרגטי מסוים ברצף (עליכם לקבוע סף נמוך, לדוגמה - 0.001). שימו לב כי את המסגרת הראשונה והאחרונה ברצף שזיהיתם, עליכם להמיר למספרי הדגימה המתאימים בסיגנל המקורי.
קלט:

- FramedSig – the framed speech signal after preprocessing.
- Fs – sampling frequency
- WindowLength – length of test and reference windows [seconds]
- Overlap – percentage of overlap between adjacent frames [0-100]

פלט:

- Idx – 2 integer vector: start and end indices of detected word.

2.3. כתבו פונקציה שכותרתה:

```
function [seg_ind, delta]=segmentation(signal, winlen, eta, dt, Fs, Idx)
```

פונקציה זו תממש את אלגוריתם הסגמנטציה האוטומטית שהסברתם בסעיף הקודם. אורכי החלונות יהיו קבועים לאורך התהליך, והחפיפה ביניהם תהיה מקסימלית.

קלט:

- signal – the speech signal after preprocessing
- winlen – length of test and reference windows [seconds]
- eta – threshold for Δ_1 spectral error measure
- dt – minimum time above threshold 'eta' [seconds]
- Fs – sampling frequency
- Idx- start & end indices of the word [samples]

פלט:

- Seg_ind – index for the beginning of each segment
- delta – spectral error measure Δ_1

* לצורך מימוש פונקציה זו ניתן, כמובן, לעשות שימוש בפונקציות מובנות של Matlab. בנוסף – מומלץ להשתמש בפונקציה המצורפת 'enframe.m'.

* שימו לב: משתנה Idx מחושב ע"י הפונקציה FindWordIdx והוא כולל וקטור של 2 משתנים: אינדקס התחלה ואינדקס סוף (בדגימות) של המילה. עליכם לחשב את הסגמנטציה רק עבור האות שנמצא בין שני האינדקסים הללו.

2.4. כעת נפעיל את אלגוריתם הסגמנטציה שמימשנו/ם על אות דיבור המכיל הקלטה של המילה 'shalom', במטרה להבחין בין הפונמות המרכיבות אותו. לצורך כך, הסבירו תחילה מהי פונמה (phoneme). דונו בהבדלים שבין פונמה קולית לפונמה א-קולית. לאחר מכן, בדקו את הפונקציה שכתבתם/ם בסעיף הקודם באמצעות ההקלטה המצורפת shalom_example.wav : הציגו גרף שבחלקו העליון את הדיבור לאחר הקיטוע לפונמות, ובחלקו התחתון ערכי מדד השוני הספקטרלי Δ_1 לאורך התהליך (בדומה לאיור 3.26 במקור המוזכר מעלה). שימו לב להתאים את אורך הווקטור של מדד השוני לסיגנל המקורי ע"י ריפוד מתאים באפסים במידת הצורך.

3. שערוך ספקטרי ומציאת פורמנטות (15)

- 3.1. שערוך שאינו פרמטרי: והציגו דוגמה גרפית באמצעות פרידוגרמה שחושבה מתוך הפונמה 'a' בהקלטה 'shalom'. שימו לב לקחת חלק מההקלטה שאתם בטוחים ששייך לפונמה זו. ניתן גם לבצע הקלטה משלכם לפונמה זו, ולעבוד איתה.
- 3.2. שערוך פרמטרי: הסבירו בקצרה על LPC (Linear Predictive Coding) ושרטטו את מודל החיזוי הלינארי לייצוג מנגנון יצירת הדיבור. בכמה מקדמים נהוג להשתמש עבור אות דיבור?
- 3.3. הוסיפו לגרף שיצרתם בסעיף 4.1 את שערוך הספקטרום הפרמטרי המחושב מתוך מקדמי ה-LPC. דונו בהבדלים בין השיטות.
- 3.4. הסבירו מהן פורמנטות. שערכו את שלוש הפורמנטות הראשונות ($F1-F3$) של הפונמה 'a' מתוך שיאי הספקטרום המשוערך (מתוך התבוננות בגרף שיצרתם).
- 3.5. הציגו מפת קטבים ואפסים של מודל ה-LPC עבור הפונמה 'a', וחשבו מתוך מפת הקטבים והאפסים את תדרי הפורמנטות (הסבירו איזה קטבים נבחר מבין כל קטבי המודל בכדי לייצג את שלוש הפורמנטות הראשונות). ערכו השוואה בין הערכים המשוערך (בשתי השיטות) לבין הערכים המצויים בספרות.
- 3.6. כתבו פונקציה שכותרתה:

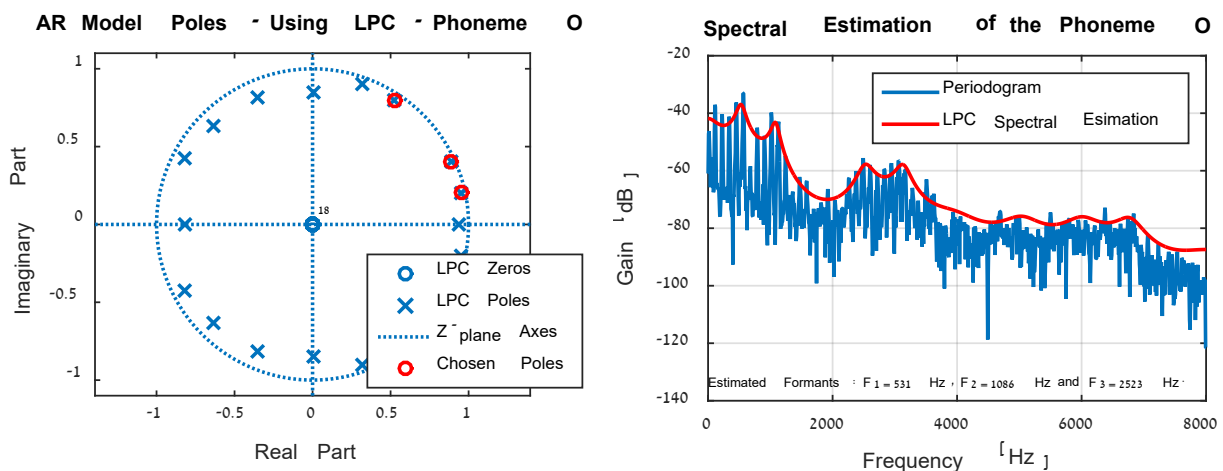
```
function [h1,h2]=estimatePhonemeFormants(PhonemeSig,Fs,phonemeName)
```

פונקציה זו תציג שני גרפים. הגרף הראשון יכיל פרידוגרמה ושערוך ספקטרום פרמטרי (LPC), כפי שביצע/תם/ בסעיפים 4.1-4.3. הגרף השני יציג את מפת הקטבים והאפסים של מודל ה-LPC, כפי שביצע/תם/ בסעיף 4.5. בנוסף, חשבו מתוך מפת הקטבים והאפסים את תדרי שלוש הפורמנטות הראשונות (במידה וקיימות), והציגו תדרים אלה בתחתית הגרף הראשון. איור 2 מציג את הגרפים שעל פונקציה זו לייצר.

קלט:

- PhonemeSig – one phoneme (after pre-processing)
- Fs – sampling frequency
- phonemeName – a string with the phoneme name

פלט:



- h1,h2 – handles for the two generated graphs
- איור 2: שני גרפים הנוצרים בפונקציה שבסעיף 4.6

3.7. הסבירו מהו Pitch. תארו את הנקודות העיקריות באלגוריתם SIFT לחישוב תדר ה-Pitch.

4. מיצוי מאפיינים (15)

עבור כל פונמה נרצה לייצר וקטור מאפיינים. וקטור מאפיינים זה יהווה ייצוג קומפקטי של הפונמה, ויאפשר לנו להבדיל בהמשך הדרך בין הדוברים השונים.

4.1. הציגו נוסחה לחישוב אנרגיה עבור מסגרות אות דיבור. כיצד יעזור לנו מאפיין זה להפריד בין פונמות קוליות לפונמות א-קוליות?

4.2. כתבו פונקציה שכותרתה:

```
function EnergySignal=calcNRG(framedSignal)
```

פונקציה זו תחשב את סיגנל האנרגיה עבור מסגרות אות דיבור באמצעות הנוסחה שהצגתם בסעיף הקודם.

קלט:

- framedSignal – a matrix of the framed signal, after preprocessing

פלט:

- EnergySignal – a column vector of the energy values of the signal

4.3. הציגו נוסחה לחישוב קצת חציית אפס עבור מסגרות אות דיבור (Zero Crossing Rate, ZCR). כיצד יעזור לנו מאפיין זה להפריד בין פונמות קוליות לפונמות א-קוליות?

4.4. כתבו פונקציה שכותרתה:

```
function ZeroCrossingSignal=calcZCR(framedSignal)
```

פונקציה זו תחשב את קצב חציית האפס עבור מסגרות אות דיבור באמצעות הנוסחה שהצגתם בסעיף הקודם. שימו לב שיש לחסר את הממוצע מכל מסגרת בטרם ביצוע החישוב.

קלט:

- framedSignal – a matrix of the framed signal, after preprocessing

פלט:

- ZeroCrossingSignal – a column vector of the zero-crossing values of the signal

4.5. שרטטו שוב את גרף הסגמנטציה של המילה 'shalom', כאשר הפעם בחלקו התחתון של הגרף יופיעו מדדי האנרגיה וה-ZCR. הסבירו האם התוצאה תואמת את הציפיות.

4.6. כתבו פונקציה שכותרתה:

```
function [FeatsVector,Feat_title]=FeatExt(Phoneme,Fs,framedPhoneme)
```

פונקציה זו תבנה וקטור בעל 24 מאפיינים בעבור פונמה אחת מאות הדיבור שהוקלט (פירוט המאפיינים בהמשך).

קלט:

- Phoneme – one phoneme (after pre-processing)
- Fs – sampling frequency
- framedPhoneme – the phoneme after framing

פלט:

- FeatsVector – 1X24 vector of features of the analyzed phoneme
- Feat_title – 1X24 cell array of the names of the calculated features

* לצורך מימוש פונקציה זו ניתן ורצוי לעשות שימוש בפונקציות מובנות של Matlab. בנוסף – מומלץ להשתמש בפונקציות המצורפות 'sift.m' ו-'formants.m'.

סדר המאפיינים בוקטור יהיה:

- א. אנרגיה ממוצעת.
- ב. ZCR ממוצע.
- ג. Pitch.
- ד. 18 מקדמי LPC.
- ה. שלוש הפורמנטות הראשונות F1-F3.

ממשו את הפונקציה שבניתם על כל אחת מהפונמות שמצאתם קודם לכן במילה 'שלום'.

5. בניית המודל – יצירת בסיס הנתונים (5)

- כעת תבצעו הקלטות, בכדי ליצור את בסיס הנתונים שלכם. במהלך סעיף זה תארו בדו"ח בקצרה, במילים, את שעשיתם/ם. אין צורך בחלוקה לתת-סעיפים.
- 5.1. החליטו מי מכן/ם 'דובר A' ומי 'דובר B'.
 - 5.2. שימרו את הפונקציות שבניתם/ם בסעיפים הקודמים בתיקיה Train GUI and model. בתיקיה זו מצוי קובץ ה-Train GUI, אשר בעזרתו תבצעו את ההקלטות. פתחו אותו.
 - 5.3. ודאו כי קיים מיקרופון מחובר למחשב, ובצעו הקלטה דרך הכפתור Record. (לחצו Stop לעצירה).
 - 5.4. לחיצה על כפתור Segmentation תקרא לפונקציות העיבוד המקדים והסגמנטציה שכתבתם, והתוצר יוצג בפניכם תוך זמן קצר. שימו לב שניתן לשנות את הפרמטרים של העיבוד המקדים ושל הסגמנטציה ולראות כיצד הם משפיעים על התוצר. חשוב- נסו להתקבע על פרמטרים המתאימים לכן/ם ולהישאר איתם מכאן ואילך, ללא שינוי, לכל אורך התהליך של בניית ואימון המודל, אך לא על חשבון סגמנטציה תקינה. (רשמו אותם בדו"ח).
 - 5.5. בצעו 10 הקלטות של המילה 'shalom' עבור כל דובר/ת, והקפידו להוסיף למודל עם תיוג מתאים של הדובר. שימו לב כי במידה ולא בוצעה סגמנטציה תקינה מבחינה כמותית עבור ההקלטה (חמש פונמות במילה 'shalom'), לא יתאפשר לכן/ם להוסיף את ההקלטה לבסיס הנתונים של המודל.
 - 5.6. להחצן Add to model מחשב את המאפיינים עבור כל הפונמות בהקלטה, ומשרשר אותם לכדי סופר-וקטור יחיד לכל הקלטה. שימו לב להוסיף כל הקלטה תחת הדובר המתאים ('דובר A' או 'דובר B').
 - 5.7. המודל ישמר אוטומטית בכל לחיצה על Add to model בתיקיית העבודה, תחת השם Model.mat. כאשר תפתחו את ה-GUI בהמשך הוא יתריע בפניכם כי מתבצעת עבודה עם מודל קיים (הוספת הקלטות). במידה וברצונכם לעבוד על מודל חדש, ניתן למחוק את הקובץ או לשנות את שמו (בצורה חיצונית ל-GUI). המודל הנוכחי ב-Train GUI תמיד יהיה Model.mat.

6. ניתוח STFT (ספקטרוגרמה) (5)

- בעבודה עם אות דיבור נהוג להשתמש בשיטת STFT (Short Time Fourier Transform).
- 6.1. הסבירו מדוע משתמשים בשיטה זו עבור אותות דיבור.
 - 6.2. הציגו ספקטרוגרמה של הקלטת דיבור כלשהי. הסבירו את בחירתכם/ם לערכי הפרמטרים היוצרים ספקטרוגרמה זו. שנו את ערכי הפרמטרים (אין חובה לשנות את כולם), הציגו גרף נוסף והסבירו מה השתנה.

7. סיווג דובר באמצעות המערכת (20)

- כעת נבצע הקלטות נוספות באמצעות ה-GUI שנמצא בתיקייה Test GUI, ונבדוק האם הוא יודע לסווג כל הקלטה לדובר הנכון על-פי בסיס הנתונים אותו אספנו והמאפיינים שחישבנו. אין צורך להתייחס בדו"ח לסעיפים 8.1-8.4 ולסעיף 8.7.
- 7.1. פתחו את ה-GUI, ובצעו הקלטה של 'דובר A'. שימו לב שגם כאן, כמובן, על הדובר/ת לומר את המילה 'shalom'.
 - 7.2. לנוחיותכם/ם, ב-GUI זה ניתן לצפות בספקטרוגרמה של האות שהוקלט. קיימת גם אפשרות לשנות את ערכי הפרמטרים היוצרים ספקטרוגרמה זו.

7.3. לחיצה על כפתור Segmentation תקרא לפונקציות העיבוד המקדים והסגמנטציה שכתבתם, והתוצר יוצג בפניכם תוך זמן קצר. שימו לב שניתן לשנות את הפרמטרים של העיבוד המקדים ושל הסגמנטציה ולראות כיצד הם משפיעים על התוצר.

7.4. המודל שנבחר אוטומטית לצורך סיווג הדובר הוא המודל Model.mat, אשר נמצא בתיקייה Train GUI and model במידה וברצונכם לעשות שימוש במודל אחר, ניתן לעשות זאת באמצעות כפתור Select model.

7.5. בכדי לסווג את ההקלטה שערכתם/לדובר המתאים לחצו על Analyze. החלטת המערכת שיצרתם/תופיע מתחת לכפתור זה. האם הסיווג נכון? בצעו 5 הקלטות לכל דובר/ת ורשמו כמה פעמים צדקה המערכת. שימו לב שבכל פעם יש לבצע הקלטה, סגמנטציה תקינה, ואנליזה.

7.6. הסבירו מה מבצעת הפונקציה classifySpeaker. לפי איזה מודל היא פועלת? הסבירו באמצעות נוסחאות מתאימות. 7.7. פעולות נוספות שהתרחשו בעת הלחיצה על Analyze הן הצגת גרף האנרגיה וקצב חציית האפס, והופעת אפשרות לבחירת פונמה והצגת גרפים עבודה. מדובר בפונקציות שאתם כתבתם. ודאו כי הן 'מתמשקות' כראוי ל-GUI.

7.8. כתבו פונקציה שכותרתה:

```
function exportGraphs(folder_name, Signal, Fs, phon, seg_ind, STFTwinLength, ...
STFToverlap, STFTnfft, STFTcmin, STFTcmax, NRG, ZCR, Flag)
```

פונקציה זו תציג ותשמור את 13 הגרפים הבאים בתיקייה המבוקשת: גרף הסיגנל הזמני לאחר עיבוד מקדים וסגמנטציה, גרף ספקטרוגרמה, גרף אנרגיה וקצב חציית אפס, גרף שערך ספקטרלי (פרמטרי ולא פרמטרי) לכל פונמה וגרף מפת קטבים ואפסים כולל סימון שלושת הקטבים הקשורים לשלוש הפורמנטות הראשונות לכל פונמה. איורים 2 ו-3 מציגים את הגרפים שעליכם לייצר. במידה ו-Flag=0 ייווצרו וישמרו רק שני הגרפים הראשונים ברשימה זו.

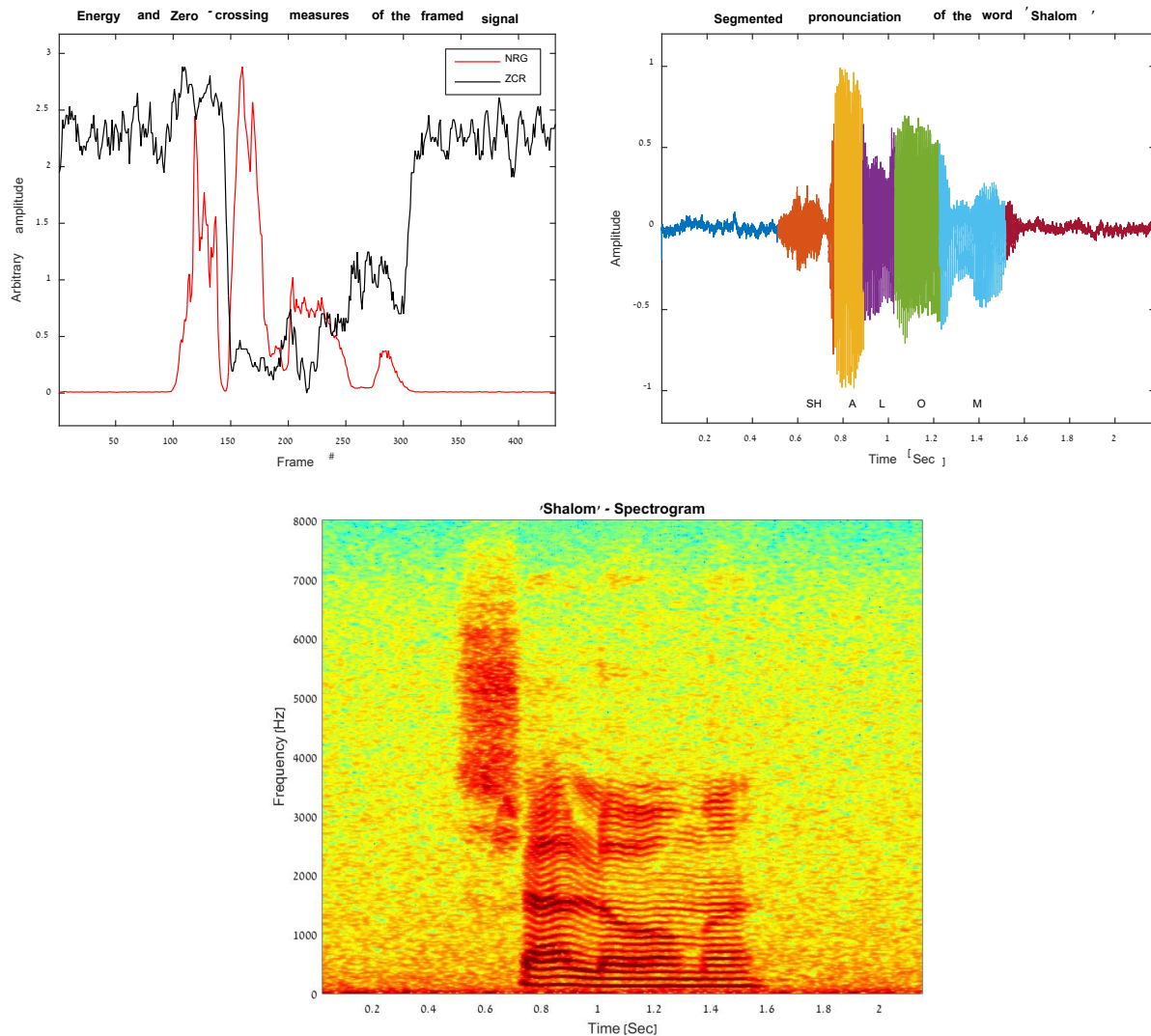
קלט:

- folder_name – full address of the folder to save the files in
- Signal – the recorded signal
- Fs – sampling frequency
- phon – an array of chars, representing the phonemes
- seg_ind – result of the segmentation process
- STFTwinLength – window length for the STFT (in samples)
- STFToverlap – overlap length for the STFT (in samples)
- STFTnfft – nfft length (in samples)
- STFTcmin – minimum value for color axis scaling
- STFTcmax – maximum value for color axis scaling
- NRG – energy signal
- ZCR – ZCR signal
- Flag – indicates which graphs to generate.

* כל הגרפים ישמרו בתיקייה המבוקשת בשני פורמטים: fig-1.jpg.

* וודאו כי הפונקציה עובדת כחלק מ-Test_GUI, בלחיצה על Export graphs, פעם אחת לפני לחיצה על Analyze (Flag=0), ופעם שניה לאחר מכן (Flag=1).

איור 3: גרפים המוצגים ב-Test_GUI ומיוצאים (ונשמרים) בעת לחיצה על Export graphs



איור 3 : דוגמה של תוצאות לעיבוד אות הדיבור של המילה 'שלום'

7.9. הציגו בדו"ח את הגרפים שנוצרו בסעיף הקודם.

7.10. בונוס: הוסיפו לספקטרוגרמה ולגרף האנרגיה וקצב חציית האפס את סימוני המקטעים של הפונמות כפי שזוהו ע"י אלגוריתם הסגמנטציה, כך שניתן יהיה להבין ביתר קלות איזה אזור בגרף שייך לכל פונמה.

7.11. בונוס: הוסיפו לגרף הספקטרוגרמה את סימוני הפורמנטות F1-F3 היכן שזה רלוונטי. הסבירו היכן הדבר מיותר. (דוגמה לתוצר ניתן לראות במסמך ההקניה למעבדה).

בהצלחה!