

חלק ב':

שאלה 1 – ספקטרוגרמה

ספקטרוגרמה (spectrogram) היא גרף ID-ממדי המורכב מאוסף התמרות DFT המבוצעות על קטעים קצרים ולעתים חופפים של אות. הציר האופקי מייצג את ציר הזמן והציר האנכי מייצג את ציר התדר. ערכי הצבע בתמונה מעידים על העוצמה: ככל שהצבע בהיר יותר כך העוצמה גדולה יותר. גרף מסוג זה הוא כלי שימושי לניתוח אותות בעלי מאפיינים המשתנים בזמן כמו רוב אותות השמע, ובפרט אותות דיבור.

הפונקציה `plot_spectrogram` הממומשת בקובץ `PART3.ipynb` מייצרת תמונה של האות בציר הזמן ומתחתיה את תמונת הספקטרוגרמה המתאימה לו. הפרמטר `winlen` מייצג את "אורך הקטע" ומשמש גם כאורך התמרת ה-DFT. הפרמטר `ovp` הוא מספר בין 0 ל-1 המייצג את רמת החפיפה בין הקטעים בזמן.

- א. טענו את אות הדיבור שאיתו עבדתם ע"י שימוש בפונקציה `read_wav_file` הממומשת בקובץ. לאחר מכן, הפעילו את הפונקציה על האות שאיתו עבדתם בחלק א' (זה שמתאים לספרה המזהה שלכם) כאשר `winlen=256` ו-`ovp=0.5`. **צרפו את התמונה המתקבלת ל-PDF.** מה הם הממדים של התמונה (אורך X רוחב)? מה המידע שניתן לראות ע"י התבוננות בחתך אנכי של תמונת הספקטרוגרמה? מה המידע שניתן לראות ע"י התבוננות בחתך אופקי של תמונת הספקטרוגרמה?
- ב. הסבירו את הקשר בין אות הדיבור לתמונת הספקטרוגרמה שלו:
- כיצד ניתן לזהות היכן מתחילה ונגמרת מילה בתמונת הספקטרוגרמה?
 - כיצד ניתן לזהות אזורים "קוליים" בתמונת הספקטרוגרמה?
 - איך נראים אזורים "א-קוליים" בתמונה?
- ג. איפה בא לידי ביטוי תדר ה-Pitch באזורים הקוליים בתמונה? צרפו ל-PDF תמונת "זום" הממוקדת לאזור של ההברה הקולית הראשונה וסמנו על ידה את תדר ה-pitch, כפי שבא לידי ביטוי בתמונה. האם תדר ה-pitch קבוע בכל האזורים הקוליים שמופיעים בתמונה?
- ד. הפעילו שוב את הפונקציה `spectrogram_plot`, כאשר הפעם נבדוק את השפעת הפרמטר `winlen`. הפעילו אותה פעם אחת עם אורך קטן פי שניים: `winlen=128`, ופעם נוספת עם אורך גדול פי שניים: `winlen=512`. השוו כל תמונה לתמונה המקורית שקיבלתם בסעיף א'. איך שינויי הערך משפיעים על התמונה (בנראות בציר הזמן, בנראות בציר התדר, בממדים?). הסבירו ממה נובעים ההבדלים.
- ה. הפעילו שוב את הפונקציה `plot_spectrogram`, כאשר הפעם נבדוק את השפעת הפרמטר `ovp`. החזירו את ערך `winlen` לערך המקורי שלו מסעיף א' (`winlen=256`). כעת הפעילו שוב את הפונקציה `plot_spectrogram`: פעם אחת עם ערך `ovp=0` ופעם נוספת עם ערך `ovp=0.95`. השוו כל תמונה לתמונה המקורית שקיבלתם בסעיף א'. איך שינויי הערך משפיעים על התמונה? הסבירו ממה נובעים ההבדלים.

שאלה 12 – שיטת LPC

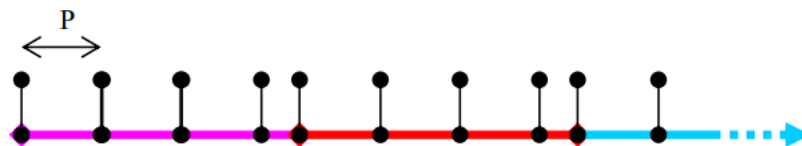
שיטת ה-LPC- שנמצאת בחומר הרקע בסעיף 2.6 מאפשרת לנו למצוא את האנרגיה של אות העירור שבעזרתו נוצר אות הדיבור, ע"י חישוב שגיאת החיזוי ביחס למסנן $A(z)$. נשתמש בערך זה במפענח, כדי ליצור אות עירור שהאנרגיה שלו תואמת לזו של אות העירור המקורי.

א. ממשו בקוד פונקציה בשם residual_energy המקבלת כקלט קטע של דגימות באורך כלשהו, ובנוסף וקטור מקדמי מסנן מסוג $A(z)$ (כפי שמוגדר בסעיף 2.6 של חומר הרקע). הפונקציה תחשב את האנרגיה הממוצעת של שגיאת החיזוי בקטע.

החישוב יתבצע באופן הבא: בשלב ראשון יש לסנן את המסגרת ע"י מקדמי המסנן $A(z)$ כאשר המסנן משמש כמסנן FIR (היעזרו בפונקציה filter). מתקבל אות המייצג את **שגיאת החיזוי**. כעת חשבו את האנרגיה הממוצעת של אות זה על פני חלון מלבני באורך המסגרת, לפי הנוסחה לאנרגיה בזמן קצר.
ב. במודל המקודד והמפענח שהצגנו בחלק המבוא התיאורטי, המקודד מוציא למפענח ארבעה פלטים. לאיזה מן הפלטים אחראית residual_energy? הסבירו.

שאלה 13 – LPC

שאלה זו תעסוק ביצירת אות עירור למסגרות שסווגו כ"קוליות" כדי לשחזרן במפענח. כזכור, אות העירור של מסגרת קולית הוא רכבת הלמים. לצורך השאלה נניח כי לתדר ה-Pitch ערך קבוע, אם כי בפועל הנחה זו אינה לגמרי נכונה: במקטעים קוליים עוקבים תדר ה-pitch משתנה קלות. מעוניינים לייצר אות עירור הדומה לאות המקורי בשאלה הקודמת: כלומר, רכבת הלמים הבנויה ממספר מקטעים עוקבים, כך שאם נשרשר אותם אחד לשני בסדר הנכון תתקבל רכבת הלמים רצופה. אורך כל מקטע הוא N דגימות, והמרווח בין ההלמים הוא מספר שלם כלשהו P ($P < N$) שאינו בהכרח מחלק של N (כלומר, תוצאת החלוקה: N/P אינה בהכרח מס' שלם). סטודנט חרוץ פתר את הבעיה ע"י לולאה, שבכל איטרציה שלה מייצרת מקטע אחד בלבד. לאחר שרשר המוקטעים, התקבלה התוצאה הבאה:



- א. מהי הטעות של הסטודנט?
- ב. מצאו נוסחה לחישוב תנאי ההתחלה (מיקום ההלם הראשון בכל מקטע) בהסתמך על מיקום ההלם האחרון במקטע הקודם כך שההלמים בשני מקטעים עוקבים ישמרו על המחזוריות הדרושה.
- ג. השלימו את הקוד החסר בפונקציה generate_impulse_train_frame שבקבצי העזר כך שהפונקציה תייצר בכל פעם מקטע אחד של רכבת ההלמים הנ"ל ותחשב את "השארית" (כמות הדגימות שנותרה ממיקום ההלם האחרון לשם שמירה על המחזוריות הדרושה) שנובעת מהמקטע הנוכחי. הפונקציה מקבלת כקלט את N , P ואת השארית שנותרה מהמקטע הקודם.

ד. בידקו את הפונקציה שיצרתם באמצעות הקוד שמופיע בקובץ PART3.ipynb : קוד זה מייצר רכבת הלמים הבנויה מארבעה קטעים עוקבים, שאורך כל אחד מהם הוא N, ושמאופיין באורך מחזור P. הקוד משתמש בפונקציה שכתבתם בסעיף ג' (המייצרת רכבת הלמים של מקטע בודד). הריצו את הפונקציה בעזרת הערכים $P=3$, $N=10$. הציגו את רכבת ההלמים שהתקבלה.

כעת נבנה מערכת מקודד-מפענח פשוטה המבוססת על חיזוי ליניארי של מודל אוטוגרסיבי. LPC כפי שצוין בחומר הרקע, ברצוננו לייצג את אות הדיבור באמצעות כמות סיביות נמוכה. המערכת שאנו בונים תאפשר לנו לקודד את אות הדיבור כך שיצרוך נפח זיכרון נמוך משמעותית, אך עדיין יישמע דומה למקור.

בחלק הראשון בתרגיל הבית בנינו את הבלוק המסווג מסגרת אות דיבור כ Voiced / Unvoiced - ואת הבלוק המשערך את תדר ה Pitch - של המסגרת. כעת נסיים את מימוש שאר חלקי המקודד: בלוק שערך האנרגיה של אות העירור והבלוק שמשערך את מקדמי החזאי הלינארי (מקדמי ה LPC). כמו כן, נממש את המפענח - בתרגיל זה נמשיך לעבוד עם אותו קובץ השמע מהתרגיל הקודם (לדוגמה speech1.wav) ועם אותן מסגרות דוגמה קולית וא- קולית שבחרתם.

שאלה 14 – ניתוח האות לפי סגמנטים

1. בדיקת `vu classify` על האות כולו

- א. השלימו את הקוד הרלוונטי בקובץ PART4.ipynb עבור הפעולות הבאות :
- a. עדכנו את הספרה המזהה שלכם וקראו את קובץ הקלט המתאים ואת תדר הדגימה.
- b. רפדו את האות באפסים כדי שאורכו יהיה כפולה שלמה של מסגרות.
- c. הגדירו מחדש את "מסגרת הדוגמה הקולית" ואת "מסגרת הדוגמה הא-קולית" שבחרתם בחלק הקודם של תרגיל בית זה.
- ב. כתבו לולאה שתריץ את הפונקציה `vu_classify` על כל המסגרות של האות המרופד. עבור המסגרות שסווגו כקוליות - מצאו את ערך מחזור ה-pitch של כל מסגרת קולית ע"י הפעלת הפונקציה `pitch_detect_corr` על המסגרת. צרו וקטור המכיל את ערכי מחזור ה-pitch של כל המסגרות. מסגרות שאינן קוליות יקבלו ערך אפס. הריצו את הלולאה על האות שלכם, כאשר הוא מרופד באפסים.
- ג. השתמשו בפונקציה `plot_pitch` שממומשת בקובץ PART4.ipynb. כדי להציג את האות בזמן ומתחתיו את ערכי מחזור ה-pitch ביח' של Hz. האם התוצאה המתקבלת תואמת את ציפיותיכם? פרטו והסבירו.

שאלה 15 – שערך מקדמי החזאי הליניארי

1. שערך מקדמי החזאי

- שערך מקדמי ה-LPC יתבצע תוך שימוש בפונקציה `lpc` הממומשת בקובץ PART4.ipynb. פונקציה זו מקבלת את סדרת הקלט ואת סדר מקדמי החזאי, p , ומחזירה וקטור המכיל את $p+1$ מקדמי מסנן החזאי הליניארי עבור סדרת הקלט. המקדם הראשון בו וקטור יהיה תמיד שווה ל-1.
- א. הפעילו את הפונקציה כך שתחזיר את מקדמי החזאי מסדר 10 של מסגרת הדוגמה הקולית.

- ב. חשבו את שונות שגיאת החיזוי בעזרת הפונק' residual_energy שכתבתם בשאלה 12.
- ג. את האמפליטודה של המעטפת הספקטרלית של מסגרת הדוגמה הקולית ניתן לקבל ע"י שימוש בפונקציה freqz באופן הבא:

$$H = \text{freqz}([1], \text{LPC_vec}, \text{frameLen}/2)$$

כאשר LPC_vec הוא וקטור מקדמי ה-LPC שהתקבל בסעיף הקודם, המשמש כאן כמסנן IIR, ו-H היא החלק החיובי של ספקטרום המסנן.

הציגו ביחידות של dB גרף המכיל את ספקטרום מסגרת הדוגמה הקולית (ע"י ביצוע התמרת DFT והצגת החלק החיובי בלבד של ציר התדר), ויחד איתו את צורת המעטפת הספקטרלית המתקבלת ממקדמי ה-LPC (יש להמיר גם אותה ליח' של dB). לשם ההצגה ניתן לשתמש בפונקציה $\text{plot_frame_and_filter}$ הממומשת בקובץ. הסבירו את התוצאה שבגרף.

2. שערון עבור סדרים שונים של החזאי

חיזרו על הסעיפים הקודמים כאשר סדר מסנן החזאי הוא 5 וכאשר מסנן החזאי הוא 50. כיצד משפיעה הגדלת/הקטנת סדר החזאי על איכות השערון? כיצד היא משפיעה על שגיאת החיזוי? מהו ה-off-trade מבחינת יחס דחיסה ומבחינת טיב השערון שמתקבל בסדרים שונים של מסנן החזאי?

3. קביעת הסדר של החזאי:

א. הסתכלו על שלושת הגרפים שהתקבלו: איזה סדר חזאי משערך בצורה הכי טובה את התמרת התדר של מסגרת הדוגמה הקולית? איזה סדר חזאי משערך בצורה הכי טובה את צורת המעטפת הספקטרלית של מסנן מערכת הדיבור? האם מדובר באותו סדר חזאי? נמקו.

ב. באיזה סדר מסנן חזאי כדאי להשתמש לדעתכם במערכת המקודד-מפענח שנבנה?

שאלה 18 – מימוש המקודד והמפענח

קעת נבנה את מערכת המקודד-מפענח המלאה, כפי שהוצגה בפרק המבוא התיאורטי. תחילה נחלץ מאות הדיבור את ארבעת הפרמטרים המאפיינים אותו במודל ה-LPC. לאחר מכן "נשדר" פרמטרים אלו למפענח. המפענח, בהסתמך על מודל יצירת הדיבור, ישתמש בהם על מנת ליצור את אות הדיבור המשוחזר.

1. מימוש המקודד

על פונקציית המקודד לבצע את הפעולות הבאות: עליה לקבל כקלט את האות המקורי ולרפד אותו באפסים כך שאורכו יהיה כפולה שלמה של 256. לאחר מכן, תופעל לולאה שתבצע את הפעולות הבאות על כל מסגרת:

- קביעת אופי המסגרת - קולית או א-קולית ע"י שימוש בפונקציה vu_classify .
- קביעת ערך מחזור ה-pitch – אם נקבע כי המסגרת קולית, מחזור ה-pitch ישוערף ע"י שימוש ב- pitch_detect_corr . אם נקבע כי המסגרת אינה קולית, ערך מחזור ה-pitch יוגדר כאפס.
- שערך מקדמי ה-LPC המתאימים למסגרת ע"י שימוש בפונקציה lpc , בהתאם לסדר שקבעתם בשאלה הקודמת.
- שיערוך האנרגיה הממוצעת של אות העירור - ע"י שימוש בפונקציה residual_energy :

הפונקציה תקבל את המסגרת ואת וקטור מקדמי ה-LPC שחושבו בסעיף הקודם.

פונקציית ה מקודד תוציא ארבעה פלטים:

N – מספר המסגרות שמכיל האות

p - וקטור באורך N המכיל את ערכי ה-pitch בכל מסגרת (שימו לב - ערכי pitch הם מספרים שלמים המייצגים כמות של דגימות, ולא ביחידות זמן או תדר).

lp - מטריצה בגודל $(LPC\ order + 1) \times N$ המכילה את וקטורי ה-LPC של כל מסגרת.

e - וקטור באורך N המכיל את שורש ערכי האנרגיה הממוצעת של אות העירור בכל מסגרת (שורש ריבועי של המוצא של הפנוקציה residual_energy).

a. השלימו את הקוד של המקודד (encoder) הנתון בקובץ PART4.ipynb על מנת שיעבוד כהלכה.

b. הסבירו למה אחראי כל אחד מארבעת הפרמטרים הנ"ל במודל יצירת אות הדיבור.

c. חשבו את נפח הזיכרון שצורך קובץ אות הדיבור המקורי (טרם הריפוד באפסים): בהנחה שכל דגימה מיוצגת ע"י שני בתים (bytes).

d. חשבו את נפח הזיכרון שצורך קובץ אות הדיבור המקודד: בהנחה שכל אחד מהערכים המספריים המופיע בקובץ (N, ערכי הוקטור p, ערכי המטריצה lp וערכי הוקטור e) מיוצג ע"י שני בתים (bytes).

e. נגדיר את יחס הדחיסה: $Compression\ ratio = \frac{Original\ File\ Size}{Encoded\ File\ Size}$

חשבו מהו יחס הדחיסה במערכת שלנו?

2. מימוש המפענח

על פונקציית המפענח לקבל כקלט את ארבעת הפלטים של פונקציית המקודד ולבצע עבור כל מסגרת את הפעולות הבאות, בהתאם לתרשים המופיע בפרק המבוא התיאורטי:

א. יצירת אות העירור:

- אם המסגרת היא א-קולית (ערך pitch שווה לאפס), אות העירור יהיה רעש גאوسی לבן באורך 256 דגימות (השתמשו בפונקציה randn).

- אם המסגרת היא קולית (ערך pitch שונה מאפס), אות העירור יהיה רכבת הלמים בעוצמה 1 ובאורך 256 דגימות כאשר המרווחים בין ההלמים הם מרווחי ה-pitch ששוערכו במקודד. יש לשמור על המחזוריות של ההלמים בין שתי מסגרות קוליות עוקבות.

ב. הגדרת עוצמה (gain) מתאימה לאות העירור:

נוודא שהאנרגיה של אות העירור המסוננת תהיה זהה לאנרגיה של אות העירור המקורי בכל מסגרת. נעשה זאת באופן הבא:

- ננרמל את אות העירור המסוננת לערך יחידה (כלומר, האנרגיה לזמן קצר של כל מקטע תהיה שווה 1). שימו לב: בעת פעולת הנרמול של אות העירור, יש להשתמש בשורש גורם הנרמול.
- נכפול את אות העירור המסוננת והמנורמל בערך e_i , כאשר e_i הוא שורש האנרגיה הממוצעת ששוערכה במקודד עבור מסגרת זו. כך נקבל שהאנרגיה עבור מקטע בודד של אות העירור המקורי (האות שייצר את אות הדיבור לאחר שחלף במסנן $\frac{1}{A(z)}$) שווה לאנרגיה עבור מקטע בודד של אות העירור המסוננת.

א. סינון אות העירור:

- ניצור את אות הדיבור על ידי סינון של אות העירור עם מקדמי החזאי הלינארי ששוערכו למסגרת. נעשה זאת תוך שימוש בפונקציה lfilter עם תנאי התחלה (כפי שראיתם בשאלת הכנה 4), כאשר המסנן משמש הפעם כמסנן IIR מסוג "Pole-All". אתחלו את ווקטור תנאי ההתחלה באפסים.

ב. שרשרת המסגרות:

- על המפענח לשרשר את המסגרות המשוחזרות זו לזו, אחת אחרי השנייה, לקבלת אות מסוננת שאורכו זהה לאורך האות המקורי (לאחר הריפוד באפסים).

- a. העתיקו לקובץ PART4.ipynb את הפונקציה `generate_impulse_train_frame` שהשלמתם ובדקתם. לאחר מכן, השלימו את הקוד של המפענח (decoder) הנתון בהמשך הקובץ על מנת שיעבוד כהלכה.
- b. הריצו את מערכת המקודד- מפענח על האות שלכם, והציגו את התוצאות: הציגו את האות המקורי ואת האות המשוחזר זה מעל זה ע"י שימוש בפונקציה `basic_plot`. האם הם נראים דומה? במה הם דומים ובמה הם שונים?
- c. השמיעו את האות המקורי ואת האות המשוחזר בעזרת הפונקציה `Audio`. האם האות המשוחזר נשמע דומה לאות המקורי? מה ההבדלים העיקריים ביניהם?
- d. שימרו את האות המשוחזר כקובץ wav בעזרת הפונקציה `write_wav_file` שמופיעה בקובץ, לטובת הגשת PDFה

בהצלחה 😊

