פרויקט גמר -DSP

***נושא : עיבוד אות דיגיטלי***

**מנחי הקורס :**

ד"ר תום טריגנו   
  
מר עזריאל (עוזי) צסטר

**הרכב הצוות:**

יאיר אהרוני – 322572793

עדן סויסה (מילואים) – 206573875

ליאור טחן (מילואים) – 318995990

מיכאל לוגסי (מילואים) – 207277690

מבוא:

* עיבוד אותות קוליים (Audio Signal Processing) מהווה תחום מרכזי בעיבוד אותות דיגיטלי (DSP) עם יישומים מגוונים בתחומי זיהוי דיבור, סינון רעשים, תקשורת, מוסיקה ועוד. בפרויקט זה התמקדנו בזיהוי מאפיינים אקוסטיים מתוך אותות קוליים, תוך שימוש בשיטות ניתוח ספקטרליות וטכניקות זיהוי פעילות דיבור (Voice Activity Detection - VAD) וזיהוי גובה צליל (Pitch Detection). בנוסף, יושם תהליך של סינתוז אות פסאודו-דיבורי המדמה את מאפייני הדיבור המקוריים.

מטרת הפרוייקט:

* מטרת הפרויקט היא להעמיק את ההבנה של טכניקות עיבוד אותות קוליים, תוך שילוב ידע תיאורטי עם יישום מעשי בתכנות MATLAB. הפרויקט מבוסס על מודולים נפרדים אשר כל אחד מהם מבצע משימה ייחודית: טעינת קבצי אודיו, חלוקת האות לחלונות, ניתוח ספקטרלי, זיהוי פעילות דיבור, זיהוי גובה צליל וסינתוז אות דמוי דיבור.

מבנה הפרויקט:  
חלק א': זיהוי פעילות קולית (VAD)

* טעינת קובץ האודיו   
  תחילה בוצעה טעינת קובץ האודיו בפורמט WAV באמצעות הפונקציה audioread ב-MATLAB. נקבע תדר דגימה (Sample Rate) והאות הוכן לניתוח בהמשך.
* חלוקת האות לחלונות (Framing)  
  האות חולק לפריימים קצרים באורך 25[mSec] עם חפיפה של 30% בין פריים לפריים. לחלון הוחל פונקציית Hann להקטנת אפקטים של דליפת ספקטרום (Spectral Leakage).
* זיהוי פעילות דיבור (VAD)  
  בוצע חישוב של רמת RMS עבור כל פריים. הוגדר סף (Threshold) מבוסס ממוצע רמות ה-RMS לזיהוי פריימים עם פעילות דיבור. תהליך זה מסייע בהבחנה בין קטעי דיבור לבין רעשי רקע או שקט מוחלט.

חלק ב': זיהוי גובה הצליל (Pitch)

* ניתוח ספקטרלי וזיהוי גובה צליל (Pitch Detection)  
  בוצע ניתוח ספקטרלי באמצעות התמרת פורייה לטווח קצר (STFT), שסיפק תמונת ספקטרוגרמה של האות. בנוסף, זוהה גובה הצליל באמצעות חישוב Pitch בטווח תדרים שבין 60 ל-400 הרץ (טווח תדרי הדיבור האנושי). שיטה זו מאפשרת זיהוי תדירויות היסוד וההרמוניות הנלוות לדיבור.

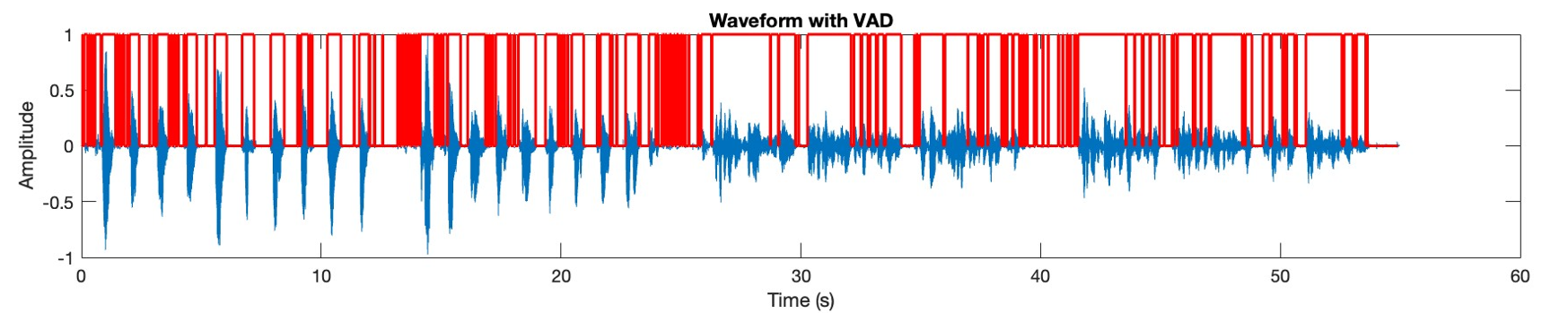
חלק ג': סינתוז אות דמוי דיבור

* בהתבסס על תוצאות ה-VAD וה-Pitch, בוצע סינתוז של אות פסאודו-דיבורי. התהליך כלל יצירת סדרות הרמוניות המבוססות על גובה הצליל שזוהה, ושיחזור האות במרחב הזמן באמצעות התמרת פורייה הופכית (IFFT).

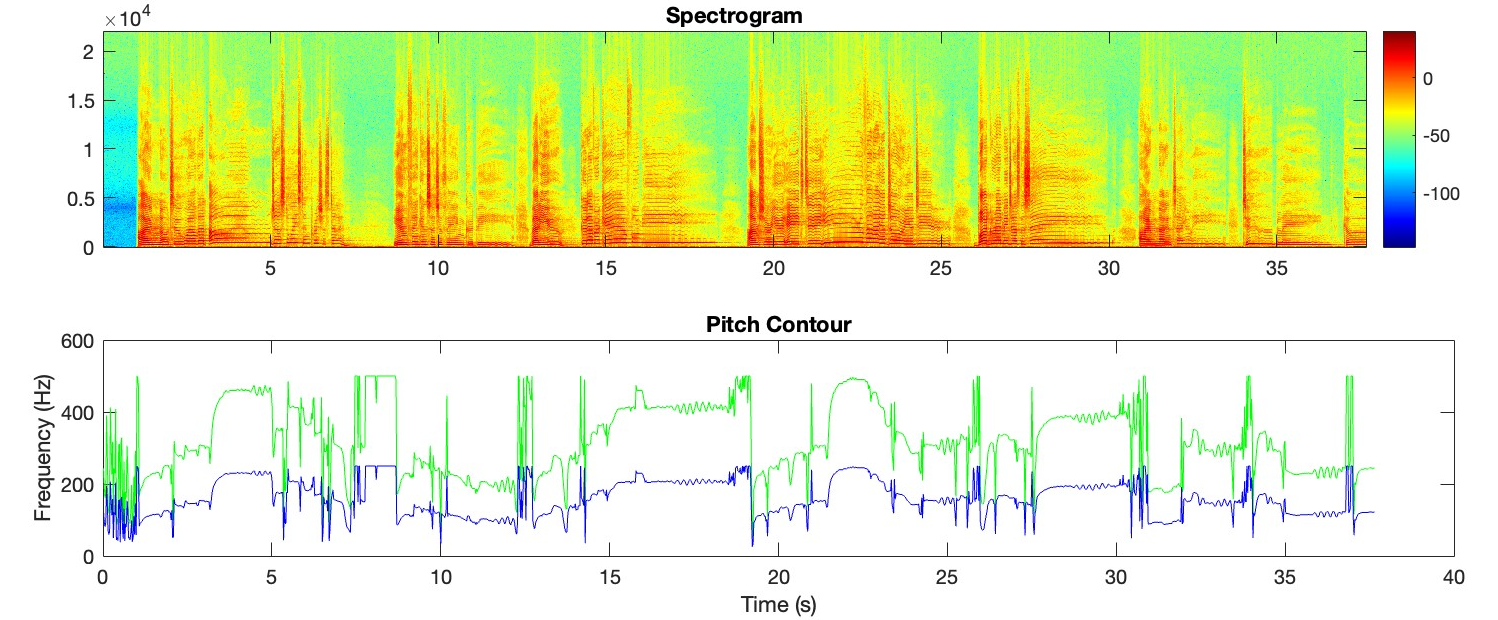
# פירוט הקוד:

## חלק א': זיהוי פעילות קולית (VAD)

* זיהוי פעילות דיבור (Voice Activity Detection - VAD) הוא שלב קריטי בכל מערכת לעיבוד דיבור, שכן הוא מאפשר לסנן מידע לא רצוי ולהתמקד בקטעי דיבור משמעותיים בלבד. השימוש בזיהוי פעילות דיבור נפוץ במערכות זיהוי קולי, תקשורת מבוססת קול ודחיסת דיבור. בפרויקט זה, יישמנו גישה המבוססת על חישוב אנרגיית האות בכל פריים וקביעת סף אנרגיה כדי לסווג פריימים עם דיבור ופריימים של שקט או רעש רקע.  
  שלבי העבודה:
* טעינת אות הדיבור - קריאה של קובץ WAV באמצעות audioread, המאפשרת גישה לערכי הדגימה ותדר הדגימה, עפ"י עקרון Nyquist .
* חלוקת האות לפריימים - חלוקה לחלונות של 25[mSec] עם חפיפה של 25% כדי להבטיח מעבר חלק בין הפריימים.
* חישוב האנרגיה - חישוב האנרגיה של כל פריים על ידי סכימה של הערכים המרובעים של הדגימות , ערכים אלו מעידים על אינטנסיביות של האות לאורך הזמן.
* קביעת סף פעילות – חישוב האנרגיה (RMS) של כל פריים וקביעת סף פעילות המבוסס על 30% מהממוצע של ערכי ה-RMS של כל הפריימים. סף הפעילות קובע את גבולות המסגרת עבור אות דיבור.
* יצירת מסכת VAD - הגדרת מערך בינארי שבו 1 מציין פעילות דיבור ו-0 מציין שקט או רעש.
* הצגת התוצאות - יצירת גרפים להשוואת האות המקורי מול מסכת ה-VAD.

נקודות תורפה:   
אלגוריתם VAD מתבסס על אנרגיה של האות , הסקת המסקנות והדרך לתוצאה מבוססת על האנרגיה של האות קיימים מספר מקרים בהם האלגוריתם עשוי להוביל לתוצאות שגויות והם:  
התחלה וסיום של מילים – שינוי חד באנרגיה של האות כלומר בהגייה של המילה האנרגיה עולה ויורדת , תופעה שמקשה על האלגוריתם להבדיל בין דיבור לשקט. כנ"ל במידה והדיבור הוא בלחישה האנרגיה חלשה ועלולה להתפרש כשקט.  
הפסקות דיבור – הפסקת הדיבור במהלך משפט עלולה להוביל לחפיפת חלונות שיתפרשו כשקט במידה והם ארוכות לדוגמא נשימה בין משפט למשפט או בין מילים.  
צלילים – עיצורים של מילים וצלילי האותיות עלולים להתפרש כקטעי שקט בעקבות האנרגיה הנמוכה בהגייה של האותיות/מילים במקרים בהם יש רעשי רקע , דוגמא לאותיות ועיצורים: "פ","ק","ש","ט","ב","ס","ח","ת".

חלק ב׳ : זיהוי גובה הצליל (pitch)

* עיבוד אות קולי וזיהוי גובה הצליל (Pitch) באמצעות MATLAB
* גובה הצליל (Pitch) הוא מאפיין מרכזי של אותות קוליים, במיוחד בדיבור ובמוזיקה. הוא משקף את התדירות הבסיסית (F0) של האות, אשר קובעת אם הצליל נשמע גבוה או נמוך. תדירות זו נוצרת כתוצאה מהתנודות המחזוריות באות הקולי.
* כדי לשפר את דיוק הזיהוי, נעשה שימוש באלגוריתם VAD (Voice Activity Detection), שתפקידו לזהות מקטעים שבהם קיים אות קולי ולהפרידם מאזורים שאינם קוליים. כך נמנעת השפעה של רעשים או מקטעים שקטים, והתהליך מתבצע רק באזורים רלוונטיים.
* לצורך ניתוח התוכן התדרי של האות לאורך הזמן, השתמשנו בהתמרת פורייה קצרה (STFT), אשר מפרקת את האות לחלונות חופפים ומבצעת FFT לכל חלון בנפרד. השיטה מאפשרת זיהוי תבניות הרמוניות ואיתור התדר הבסיסי של האות.
* שלבי העיבוד:
* טעינת אות הקול – השתמשנו באות הקולי "easy to love", שמרנו אותו במשתנה signal וביצענו עליו STFT. כתוצאה מכך, קיבלנו את הספקטרום (בערכים מרוכבים, ולכן לקחנו את הערך המוחלט), ציר התדרים וציר הזמן של כל דגימה.
* זיהוי תדרים בסיסיים – הפעלנו לולאה העוברת על כל פריים בספקטרום. כאשר נמצא אות קולי (לפי ה-VAD), חיפשנו את התדרים הבסיסיים של הפריים בעזרת הפונקציה findpeaks. הסף שנקבע לזיהוי היה 1% מהמגניטודה המרבית בפריים.
* ניתוח תדר הבסיס – הצגנו את התדרים הבסיסיים עבור כל פריים. יש לזכור שהתדר הבסיסי הוא ההרמוניה הראשונה, כלומר התדירות הנמוכה ביותר בגל מחזורי. עם זאת, היא לא תמיד תהיה בעלת המגניטודה הגבוהה ביותר בספקטרום.
* הצגת תוצאות – לבסוף, הצגנו גרף STFT במבט דו-ממדי, המציג את ההתמרה לאורך ציר הזמן והתדר. כמו כן, יצרנו pitch contour, המתאר את השתנות גובה הצליל לאורך הזמן.

חלק ג׳ :סינטוז אות דיבור

סינתזת אות דיבור היא תהליך שבו נבנה אות קולי מלאכותי המדמה את המאפיינים האקוסטיים של דיבור אנושי. בתהליך זה, הדיבור מנותח ומפורק לרכיבים בסיסיים – תדר הבסיס (Pitch) וההרמוניות הנלוות לו, ולאחר מכן משוחזר מחדש באמצעות טכניקות עיבוד אותות דיגיטליים.

בפרויקט זה, הסינתזה בוצעה על בסיס תוצאות שלבי הזיהוי הקולי (VAD) וזיהוי גובה הצליל (Pitch) מהחלקים הקודמים. השלבים המרכזיים כללו:

* זיהוי חלונות המכילים פעילות קולית בעזרת וקטור ה-VAD.
* חישוב ההרמוניות המבוססות על תדר הבסיס שנמצא בכל חלון פעיל (עד 5 הרמוניות).
* בניית ספקטרום הרמוני פשוט (בינארי) ושימוש באלגוריתם IFFT לשחזור האות בזמן.
* שילוב הפריימים לאות רציף, תוך ניהול נכון של החפיפות (Overlap) כדי לשמור על זרימה חלקה של הצליל.

לצורך דיוק מירבי בשחזור הקול, בכל חלון קול פעיל חילצנו את תדר הבסיס (Pitch) והפקנו ממנו עד 5 תדרים הרמוניים – כפולות של תדר הבסיס עד לתדר הגבוה ביותר המותר (תנאי נייקויסט). ההרמוניות הללו הוכנסו לספקטרום בינארי (כל תדר שמזוהה מקבל ערך 1, והשאר 0), ולאחר מכן בוצע IFFT לשחזור האות במרחב הזמן.

מגבלות ושיקולים:

* סף האנרגיה לזיהוי פעילות קולית (VAD) הוגדר ל-30% מהממוצע של ערכי ה-RMS. סף זה עובד היטב ברוב המקרים, אך ייתכן שחלק ממקטעי דיבור חלשים במיוחד לא יזוהו כראוי, מה שעלול להוביל לחסרים קלים בשחזור.
* זיהינו כי חלק מהמקטעים עם עוצמה נמוכה עלולים להיעלם עקב סף זה, במיוחד בהברות חלשות או הפסקות נשימה קצרות. יחד עם זאת, שמירה על סף זה מאפשרת להפחית רעשי רקע שאינם קשורים לדיבור במקרה שאינו אקפלה.

סעיף הבונוס: שיפור אות הדיבור

בחלק זה הוספנו רעש ורוד לאות הדיבור כדי לדמות סביבה רועשת, ולאחר מכן ניסינו לשפר את איכות הדיבור על ידי סינון הרעש באמצעות עיבוד של הספקטרום (FFT ו-IFFT). המטרה הייתה לבדוק איך אפשר לשחזר את האות , בסיגינל המקורי.

הוספת רעש ורוד

* + בחרנו רעש ורוד כי הוא דומה לרעשים טבעיים שקיימים בסביבה יומיומית.
  + שילבנו אותו עם אות הדיבור בעוצמה נמוכה (בערך 5% מעוצמת האות) כדי ליצור אות "מלוכלך".

שימוש בFFT

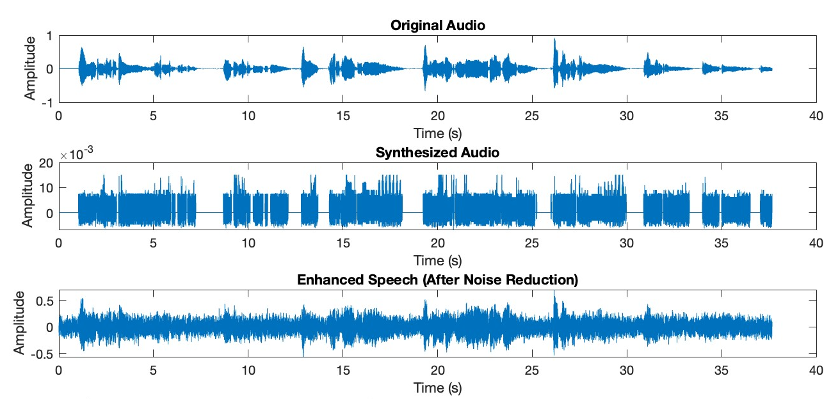
* + ביצענו FFT על האות עם הרעש כדי להעביר אותו למישור התדר.
  + חישבנו את עוצמת התדרים (Magnitude) כדי להבין איפה נמצא הרעש.

ניקוי הרעש

* + חיסרנו את ממוצע עוצמת הרעש מספקטרום האות כדי לנקות את הרעש.
  + דאגנו שכל הערכים יהיו חיוביים, כך שערכים שליליים הוגדרו כאפס.

שחזור האות בעזרת IFFT

* + אחרי שהפחתנו את הרעש, החזרנו את האות למרחב הזמן באמצעות IFFT.
  + השתמשנו במידע הפאזי המקורי של האות כדי לשמור על צליל טבעי יותר.

השיטה הצליחה להפחית את הרעש ולשפר את איכות הדיבור בצורה טובה יחסית. כך ניתן להשתמש בכלים של עיבוד אותות דיגיטליים (DSP) כדי לשפר איכות דיבור במערכות כמו תקשורת, זיהוי קולי ויישומים אודיו נוספים.

סיכום

הפרויקט שלנו מתמקד בעיבוד אותות דיבור בשלושה שלבים מרכזיים: זיהוי פעילות דיבור (VAD), זיהוי גובה הצליל (Pitch Detection) וסינתזת אות דיבור.

בשלב הראשון, אנו מחלצים דגימות קול מקובץ WAV, מחלקים אותן לפריימים חופפים עם חפיפה של כ-25%, מחשבים את האנרגיה של כל פריים, וקובעים על פיה האם מדובר בפעילות דיבור. שלב זה מאפשר סינון רעשי רקע והתמקדות במקטעים רלוונטיים בלבד.

בשלב השני, אנו מבצעים זיהוי גובה הצליל על ידי ניתוח ספקטרלי באמצעות STFT, דרכו אנו מזהים את התדר הבסיסי (Fundamental Frequency) של אות הדיבור. בשלב זה אנו משווים בין פריימים עם פעילות דיבור לזיהוי הפריימים שבהם גובה הצליל מזוהה בצורה הברורה ביותר.

לבסוף, בשלב סינתזת אות הדיבור, אנו משתמשים ב-IFFT לשחזור ההרמוניות של האות וליצירת אות דיבור מלאכותי. לכל פריים שעבר טרנספורמציה אנו מחלצים את ההרמוניות שלו, ולבסוף מחברים את כל החלקים לכדי אות דיבור מלא.

במהלך העבודה שילבנו גם סעיף בונוס של שיפור אות הדיבור (Speech Enhancement), בו הוספנו רעש ורוד לאות וביצענו סינון רעש באמצעות עיבוד ספקטרלי (FFT ו-IFFT), במטרה להמחיש כיצד ניתן לשפר איכות דיבור בסביבות רועשות.

את כל החישובים, הניתוחים והסינתזה ביצענו בעזרת MATLAB, שסיפק לנו את הכלים הנדרשים להתמודד עם עיבוד האותות בצורה מדויקת ויעילה. הפרויקט אפשר לנו להעמיק בהבנת עקרונות עיבוד אותות דיגיטליים ולהתנסות ביישום טכניקות מתקדמות בתחום עיבוד הדיבור.

### מקורות

1. MathWorks. "Voice Activity Detection (VAD)." https://www.mathworks.com/help/signal/ref/vad.html
2. MathWorks. "Pitch Detection." https://www.mathworks.com/help/audio/ref/pitch.html
3. MathWorks. "Short-Time Fourier Transform (STFT)." https://www.mathworks.com/help/signal/ref/stft.html
4. MathWorks. "Inverse Fast Fourier Transform (IFFT)." https://www.mathworks.com/help/matlab/ref/ifft.html
5. MathWorks. "Noise Reduction Using Spectral Subtraction." https://www.mathworks.com/help/audio/noise-reduction-using-spectral-subtraction.html