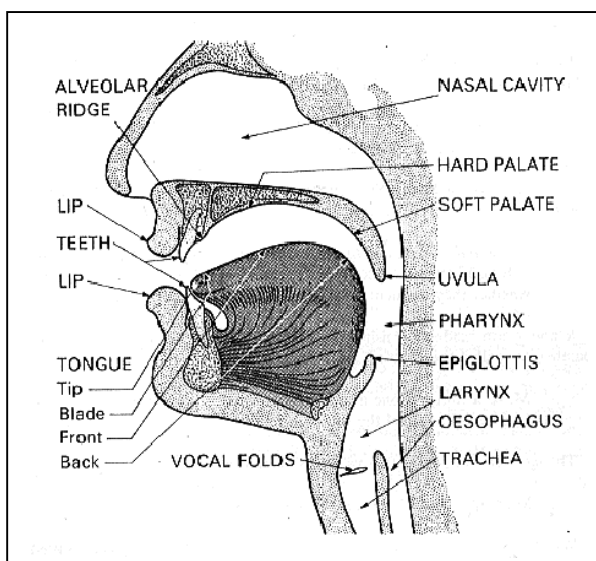
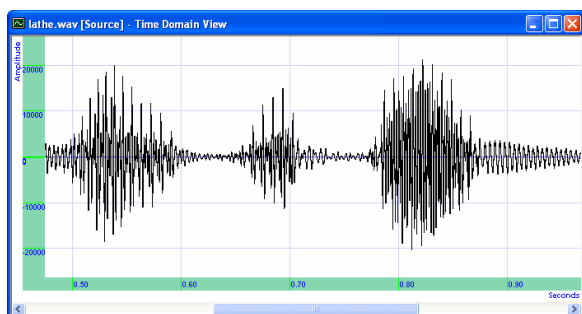


הטכניון – מכון טכנולוגי לישראל  
הפקולטה להנדסת חשמל  
המעבדה לעיבוד אותות ותמונות

ניסוי מעבדות 2-3 :

## עיבוד ספרתי של אותות דיבור Digital Processing of Speech Signals



נכתב ע"י: הדס אופיר, יאיר משה ינואר 2003  
עדכון: זוהר ברזילי, יאיר משה פברואר 2005  
עדכון: אלעד שטרר מרץ 2008  
עדכון: מיכל גנוסוב, הדס בן-איסטי: ספטמבר 2009, ספט' 2012  
עדכון: עידן רם, מרץ 2011  
עדכון: הדס בן-איסטי, אוגוסט 2013  
עדכון: אורי בריט והדס בן-איסטי, אוגוסט 2015

עדכונים לניסוי באתר LabAdmin

## תוכן עניינים

1.....	<b>הנחיות</b>
1.....	מהלך הניסוי
1.....	הגשת דו"ח מסכם
2.....	<b>רקע תיאורטי</b>
2.....	מטרת הניסוי
2.....	מבנה אות הדיבור ותכונותיו
4.....	שערוך תדר ה-PITCH
4.....	שיטת האוטוקורלציה
5.....	עיבוד מקדים
7.....	החלטה על קיום דיבור ועל VOICED/UNVOICED
11.....	חיזוי ליניארי
15.....	מדדים אובייקטיביים לאיכות דיבור
16.....	תוכנת SPDEMO
17.....	<b>שאלות הכנה לפגישת הניסוי הראשונה</b>
20.....	<b>מהלך הניסוי, פגישה ראשונה</b>
25.....	<b>שאלות הכנה לפגישת הניסוי השניה</b>
29.....	<b>מהלך הניסוי, פגישה שניה</b>
33.....	<b>מקורות לעיון נוסף</b>

## הנחיות

### **מהלך הניסוי**

- יש לקרוא את החלק התיאורטי במלואו ולענות על שאלות ההכנה לפני מועד המעבדה. התשובות לשאלות ההכנה יוגשו למדריך הניסוי כדו"ח מכין לפני המפגש (במערכת LabAdmin)
- במקרה ומתעוררת שאלה או בעיה במהלך הניסוי יש להיוועץ במדריך הניסוי.
- הניסוי יסתיים רק לאחר קבלת אישור ממדריך הניסוי.

### **הגשת דו"ח מסכם**

- הדו"ח המסכם יוגש כקובץ Word או כקובץ PDF במערכת LabAdmin למנחה הניסוי.
- לכל אחד מן החלקים יוגש דו"ח מסכם נפרד, לכל היותר שבועיים מתאריך הפגישה של כל חלק.
- הדו"ח המסכם צריך לכלול את כל תוכניות ה-Matlab שכתבתם.
- רצוי לכלול מספר גרפים בעמוד ע"י שימוש בפונקציה subplot.
- גרף המייצג תמונת תדר ייוצג ביחידות של dB, וציר התדר שלו יהיה ביחידות של Hz. יש להציג רק את החצי הראשון (החיובי:  $[0, \pi]$ ) של ציר התדר.
- על כל אחד מהגרפים יש לציין את היחידות של הצירים ולרשום כותרת מתאימה (בעזרת הפונקציות title, xlabel, ylabel, axis וכו').
- הצגת הגרפים תהיה בצירוף הסברים מתאימים, אפילו אם לא נאמר כך במפורש.

## רקע תיאורטי

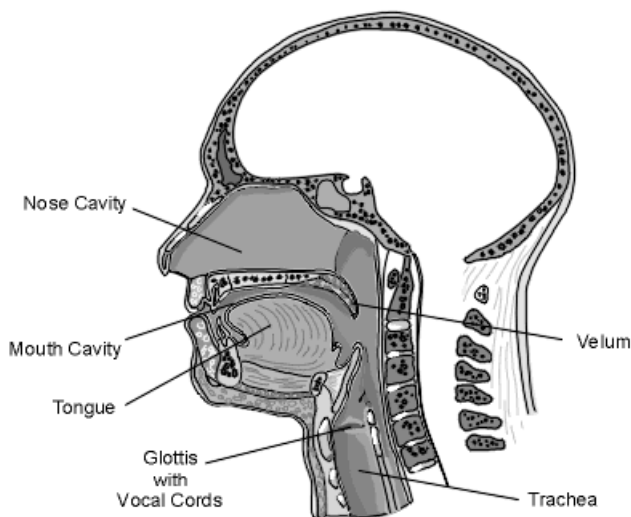
### מטרת הניסוי

קיימים אלגוריתמים רבים לעיבוד אות המתוכננים לפעול על אות דיבור. אלגוריתמים מסוג זה עשויים, למשל, לבצע דחיסה של אות הדיבור או שיפור שלו (ניקוי רעשים, הסרת הדהוד וכד'). כדי להשיג תוצאות טובות, אלגוריתמים אלה מסתמכים לרוב על התכונות הפיסיולוגיות של מערכת הדיבור האנושית.

מטרת הניסוי היא להכיר את מבנה אות הדיבור, את תכונותיו הסטטיסטיות ומודל פרמטרי לייצוגו. כמו כן, יוכרו שיטות המשמשות לשערוך פרמטרי של מודל זה ויודגם השימוש בו לשם ביצוע דחיסה יעילה של האות.

### מבנה אות הדיבור ותכונותיו

בציור שלפניכם ניתן לראות כיצד פועלת מערכת יצירת הקול.



ציור 1 : מערכת הקול האנושית

לחץ אוויר מהריאות עובר דרך קנה הנשימה (trachea) ומגיע למיתרי הקול (vocal cords) הנמצאים בבית הקול (glottis). מיתרי הקול, שהם שתי פיסות רקמה המוצמדות זו לזו, יכולים לאפשר לאוויר לעבור בחופשיות או שהם יכולים לרטוט, כלומר להיפתח ולהיסגר במהירות. פתיחה וסגירה זו יוצרת פולסי אוויר מחזוריים. המרווח המחזורי בין הפולסים, הנקרא מחזור ה-pitch, מושפע משינויים בלחץ האוויר ובמתיחות של מיתרי הקול.

זרם האוויר מתפשט ממיתרי הקול (הפתחים או הנפתחים-נסגרים לסירוגין) למעבר הקולי (vocal tract), כלומר, אל חלל הפה (mouth/oral cavity).

על פני הלשון, השיניים והשפתיים ואל חלל האף (nose cavity) דרך הנחיריים. החך הרך (velum) הוא שסתום הממוקם בפתח חלל האף ושולט על מעבר האוויר לתוך חלל זה.

אות העירור (excitation signal) למערכת הקול הוא זרם האוויר המגיע אל חללי הפה והאף. החך הרך, הלשון, השפתיים והלסת התחתונה יוצרים מעצורים שונים הגורמים למערבולות בזרם האוויר וכך יוצרים את חיתוך הדיבור (articulation). פונקציית התמסורת של חללי הפה והאף משתנה בזמן, מפני שצורתם של חללים אלה משתנה תוך כדי דיבור. פונקציית תמסורת זו מאפנת את פולסי האוויר ויוצרת את אות הדיבור הנפלט החוצה מהשפתיים ומהנחיריים.

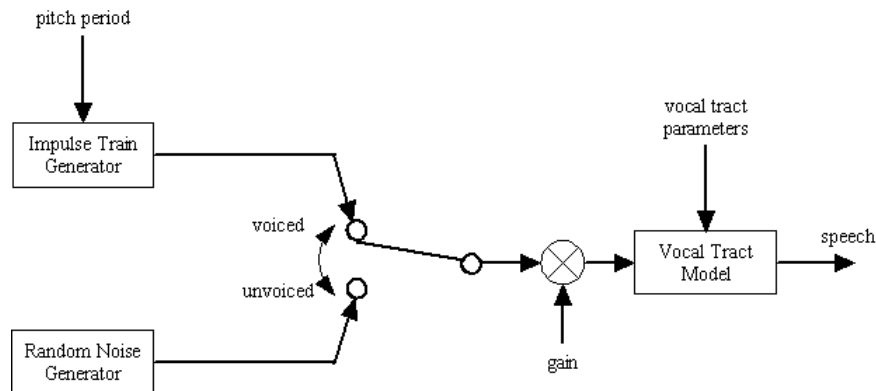
את אות הדיבור ניתן לחלק באופן גס לשני סוגים :

- אותות קוליים (voiced) – אותות מחזוריים הנוצרים מרעידה של מיתרי הקול. למשל, התנועות : a, e, o, u, i. נסו להניח את היד על הגרון בעת השמעת הברות אלה ותרגישו רטיטות.
- אותות א-קוליים (unvoiced) – אותות לא מחזוריים הנוצרים כאשר מיתרי הקול במצב פתוח ואינם רוטטים. המאפיינים של אותות מסוג זה דומים לרעש. למשל, התנועות : s, sh, ph. נסו להניח את היד על הגרון בעת השמעת הברות אלה ולא תרגישו רטיטות.

חללי האף והפה מהווים תיבת תהודה, כלומר, מסנן בעל תדרי תהודה (קטבים) שמשתנים לפי צורת החלל. תדרי תהודה אלו נקראים פורמנטים (formants). השילוב של תדרי פורמנטים שונים ושני סוגי הקולות יוצר את ההברות השונות.

המבנה הפיסיולוגי של מערכת הדיבור שונה מאדם לאדם והוא גורם לתדר pitch שונה ולתדרי פורמנטים שונים. תדר ה-pitch של דיבור אנושי נע בתחום שבין 50Hz (עבור גברים בעלי טון דיבור נמוך) ל-400Hz (עבור ילדים ונשים בעלות טון דיבור גבוה).

על מנת לנתח אותות דיבור בכלים של עיבוד אות יש צורך למצוא מודל שמתאר את מערכת הדיבור במושגים מתמטיים. מודל קלאסי כזה של מערכת המייצרת אות דיבור הוא :



ציור 2 : מודל יצירת אות דיבור

כאשר הדיבור הוא קולי (voiced), מיתרי הקול מייצרים פולסים מחזוריים המיוצגים במודל בעזרת מקור עירור המייצר רכבת הלמים, במרווחי זמן קצובים ביניהם. מרווחי זמן אלה הם משכי מחזור ה-pitch. המודל מניח כי קטעי דיבור קוליים הם קווי-מחזוריים, כלומר, הם משתנים באופן איטי יחסית לזמן המחזור שלהם כך שניתן לייצגם, למשך קטעי זמן קצרים, בעזרת מרווחי זמן קבועים. כאשר הדיבור הוא א-קולי (unvoiced), התהליך מיוצג ע"י מקור אות עירור דמוי רעש רחב סרט.

לפי מודל זה, מערכת הדיבור מפיקה בכל רגע אות קולי לחלוטין או אות א-קולי לחלוטין. זאת בניגוד למודל מורכב יותר, שעשוי לאפשר גם מצבים של הברות מעורבות, המשלבות את שני סוגי הקולות (למשל : z, v). הבחירה בין שני מקורות העירור נעשית ע"י מתג, ובשני המקרים מועבר אות העירור דרך מסנן משתנה בזמן המייצג את חללי הפה והאף. תגובת התדר של מסנן זה, המאופיינת בעיקר ע"י הקטבים (מן האפסים נוטים בדרך כלל להתעלם במקרה זה), מעצבת את צורת המעטפת הספקטראלית של אות העירור ומשנה את עוצמתו בהתאם. הפלט של המערכת הוא אות העירור המעוצב ספקטראלית.

אות דיבור הוא אות מוגבל סרט שרוב האנרגיה שלו מרוכזת בתדרים נמוכים. לרוב מעבירים אות זה דרך מסנן מעביר נמוכים בעל תדר קיטעון נמוך מעט מ-4KHz ודוגמים ב-8KHz. רוחב סרט זה שומר על תכונות המובנות, יכולת הזיהוי והטבעיות של אות הדיבור בצורה טובה.

## שערוך תדר ה- *pitch*

תדר ה-*pitch* הוא תכונה בסיסית של אות הדיבור. רגישות האוזן האנושית לשינויים בתדר זה היא גדולה יותר, בסדר גודל, מאשר הרגישות שלה לשינויים בפרמטרים אחרים של אות הדיבור. לכן, לשערוך תדר ה-*pitch* באופן מדויק יש חשיבות גדולה בעת יצירת אלגוריתם המשתמש במודל של אות הדיבור.

שערוך תדר ה-*pitch* באופן מדויק מאות דיבור דגום הוא בעייתי בשל מספר סיבות :

- תדר ה-*pitch* משתנה ממחזור למחזור לפי מיקום ההברה במשפט (שינוי זה יכול להגיע עד ל-10% בין שני מחזורים עוקבים).
- קיימים הבדלים באמפליטודת האות בין מחזור למחזור הנובעים מהדגשים שונים בתוך המשפט.
- בעת מעבר מאות קולי לאות א-קולי, או מאות א-קולי לאות קולי, עשוי להתווסף לאות המחזורי רכיב של רעש המקשה על גילוי המחזוריות.
- הרמוניות של תדר ה-*pitch* או פורמנטים עלולים לעיתים להיראות חזקים יותר מתדר ה-*pitch*.
- קצב הדגימה הנמוך יחסית והכימוי (=קוונטיזציה : ייצוג במספר סופי של רמות) הנגרמים בעת הדגימה של אות אנלוגי מוסיפים שגיאות ומקשים על חישוב מדויק של תדר ה-*pitch*.

קיימים אלגוריתמים רבים לשערוך תדר ה-*pitch*. בניסוי זה נכיר שתי שיטות : הראשונה והמרכזית היא שיטה המשמשת בסיס לרב אלגוריתמים אלה – שיטת האוטוקורלציה. השנייה היא שיטה קלאסית לשערוך תדר ה-*pitch* בעזרת מקדמי ה-*cepstrum*.

## שיטת האוטוקורלציה

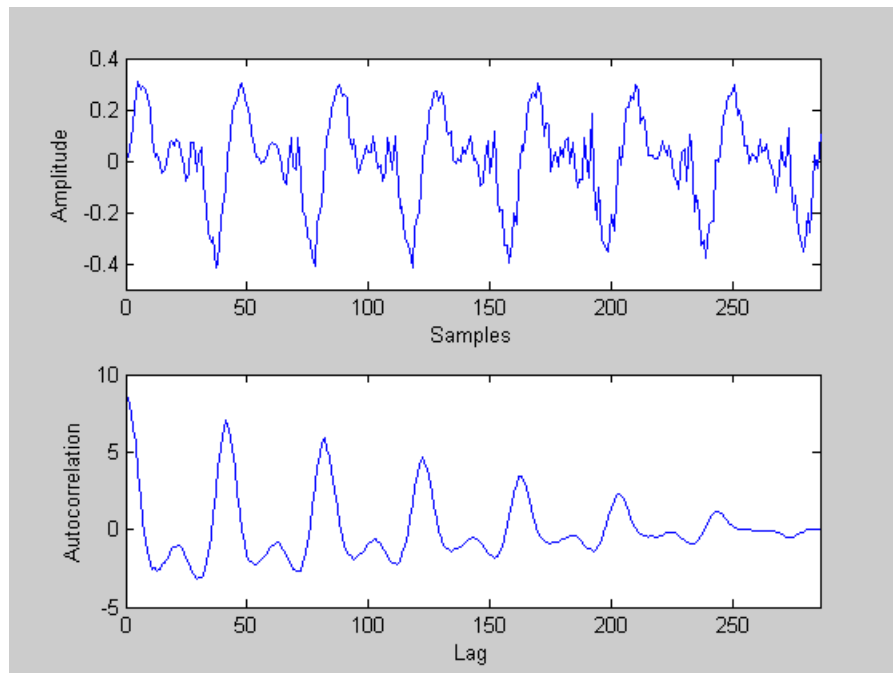
פונקציית הקורלציה מהווה מדד לדמיון בין שני אותות. פונקציית האוטוקורלציה מודדת את רמת הדמיון של האות לגרסה מוזזת בזמן של עצמו. לכן, המקסימום של פונקציית האוטוקורלציה של אות דיבור יתקבל במרווחי מחזור ה-*pitch* של האות המקורי.

נתון קטע אות דיבור  $x(n)$  באורך  $N$ . פונקציית האוטוקורלציה לזמן קצר של קטע אות זה מוגדרת באופן הבא :

$$r_k = \sum_{n=0}^{N-k-1} x_n x_{n+k} \quad k = 0, \dots, N-1$$

הפרמטר  $k$  הוא ההזזה יחסית לאות המקורי.

בשרטוט הבא אנו רואים קטע אות דיבור קולי ואת סדרת האוטוקורלציה שלו :



ציור 3 : קטע אות דיבור קולי ופונקציית האוטוקורלציה המתאימה

ניתן לראות בציור כי פונקציית האוטוקורלציה מקבלת מקסימום מקומי עבור ערכי  $k$  השווים למחזור ה-pitch ולכפולות שלמות שלו. מכאן ניתן לשערך את מחזור ה-pitch ע"י מציאת המקסימום המקומי הגדול ביותר של פונקציית האוטוקורלציה שאינו באפס.

לבחירת אורך החלון  $N$  יש שני שיקולים סותרים – מצד אחד, החלון צריך להיות ארוך מספיק כך שבכל מקרה יכיל לפחות שני מחזורי pitch שלמים. מצד שני, האות הוא קווי-סטציונרי ולכן רצוי לקחת חלון קצר מספיק כדי שהנחת הסטציונריות תתקיים. ערך מקובל לאותות דיבור, שעונה על שתי דרישות סותרות אלה הוא חלון באורך 20-40msec.

## עיבוד מקדים

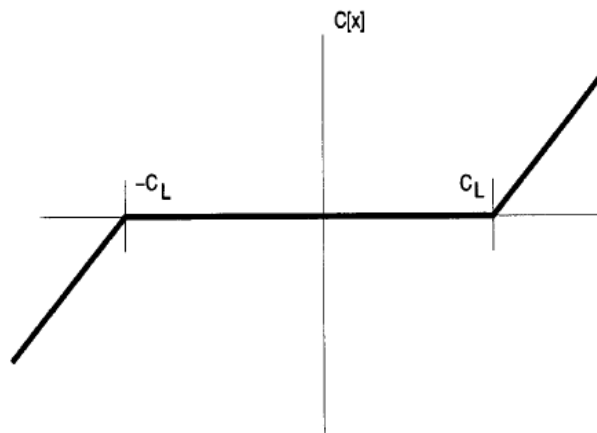
בפונקציית האוטוקורלציה ניתן להבחין, בנוסף לשיאים המופיעים בכפולות של זמן מחזור ה-pitch, גם בשיאים נוספים הנובעים מהשפעת הפורמנטים. שיאים נוספים אלה מקשים על זיהוי נכון של תדר ה-pitch. כדי להקל על פעולת הזיהוי, ניתן להעביר את האות דרך מסנן מעביר נמוכים. מסנן זה מנחית תדרים גבוהים הנמצאים מחוץ לתחום החוקי של תדר ה-pitch. בנוסף, ניתן לקטום את האות. תפקידה של פעולת הקטימה הוא להנחית שיאים הנובעים מפורמנטים גבוהים המקשים על זיהוי זמן מחזור ה-pitch.

קטימה (center clipping) היא הפעולה הלא-לינארית הבאה :

$$\text{clip}(x_n) = \begin{cases} x_n - C_L & x_n > C_L \\ 0 & |x_n| \leq C_L \\ x_n + C_L & x_n < -C_L \end{cases}$$

הפרמטר  $C_L$  הוא קבוע, הנקבע באופן אדפטיבי לכל חלון. ערכים אופייניים ל- $C_L$  הם בתחום 60%-80% מאנרגיית האות המקסימלית.

פעולת הקטימה נראית כך :

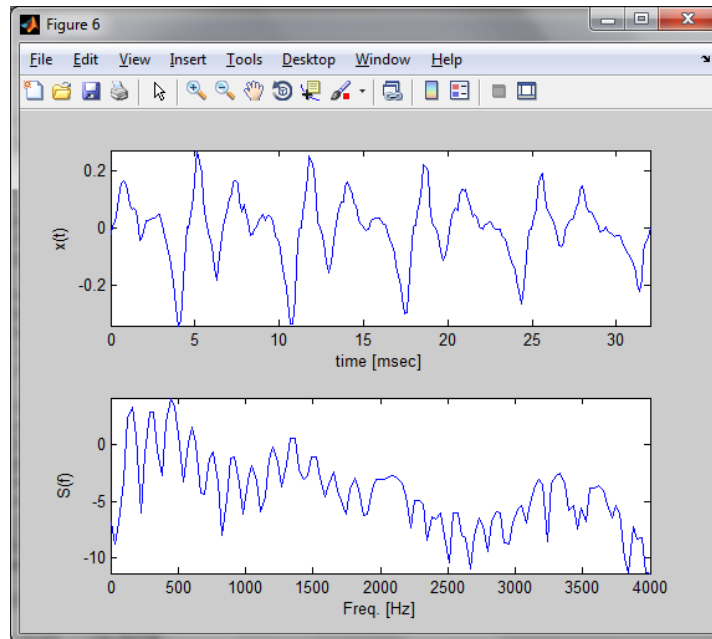


ציור 4 : פונקציית הקטימה



## שערוך תדר ה- *pitch* בעזרת מקדמי ה- *cepstrum*

נתבונן במסגרת בה קיים אות דיבור קולי בעל תדר *pitch* ממוצע של 147Hz :



ציור 5 : מסגרת קולית בזמן ובתדר

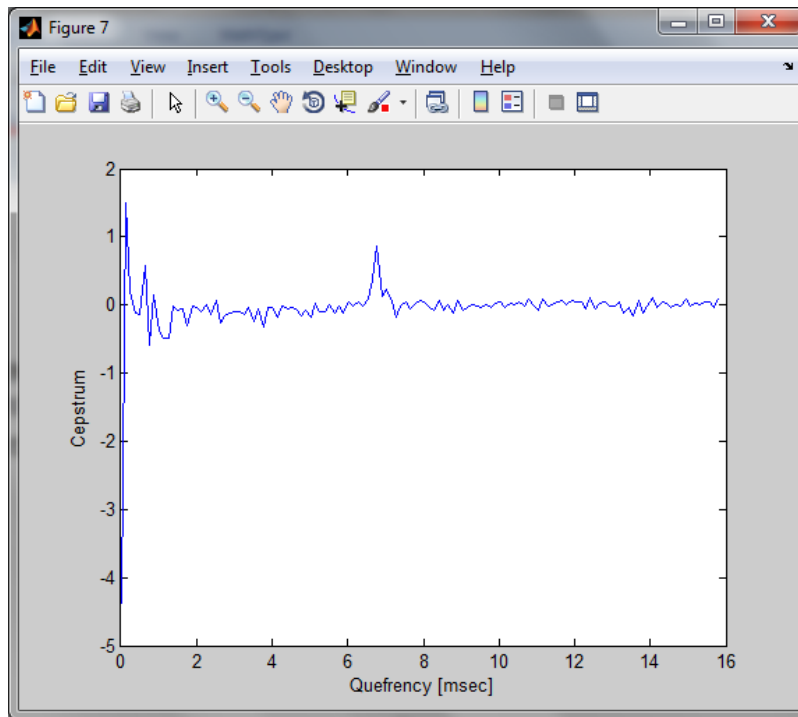
הספקטרום של האות המופיע בגרף התחתון מחושב מתוך האות הזמני  $x(t)$  (המוצג בגרף העליון) על ידי :

$$S(f) = 20 \cdot \log_{10} \left( \left| \text{FFT} \{ x(t) \} \right| \right)$$

הספקטרום מורכב מאות בתדר נמוך יחסית הנקרא מעטפת ספקטראלית ומאות מהיר שנובע מתדר ה-*pitch*.  
מקדמי ה-*cepstrum* (מבוטא : קפסטרום) מחושבים על ידי החלק הממשי של התמרת הספקטרום :

$$C(n) = \text{Re} \{ \text{IFFT} \{ S(f) \} \}$$

כיוון שמבצעים שתי התמרות, הציר בו מחושבים מקדמי הקפסטרום הוא בעל יחידות זמן ונקרא Quefrency.



ציור 6 : מקדמי הקפסטרום

ציור 6 מתאר את מקדמי הקפסטרום של המסגרת הקולית שהוצגה בציור 5. אכן ניתן לראות שני רכיבים: המקדמים הנמוכים מייצגים את התמרת המעטפת הספקטראלית והחלק המהיר שנובע מתדר ה-pitch, מיוצג על ידי השיא ב-6.8msec – בהתאמה לתדר ה-pitch שחושב מתוך האות הזמני.

**הערה:** מערכות רבות לעיבוד אותות דיבור כגון זיהוי דובר, זיהוי דיבור, text-to-speech, מבוססות על ייצוג המעטפת הספקטראלית על ידי מקדמי הקפסטרום (הנמוכים).

**החלטה על קיום דיבור ועל voiced/unvoiced**

ניתן לחלק כל אות דיבור דגום לקטעים בהם קיים דיבור ולקטעי שקט. לפי המודל שהוצג ליצירת אות דיבור, את הקטעים בהם קיים דיבור ניתן לחלק לקטעים המכילים אות קולי ולקטעים המכילים אות א-קולי.

לאלגוריתם המזהה קיום דיבור או קיום שקט נוהגים לקרוא VAD (Voice Activity Detector). בנייתו של VAD איכותי היא מלאכה מורכבת המערבת לרוב שקלול של פרמטרים שונים. הפרמטר החשוב ביותר הוא פונקציית האנרגיה לזמן קצר של האות (short-time energy function) המוגדרת באופן הבא:

$$E_n = \sum_{m=-\infty}^{\infty} [x_m w_{n-m}]^2$$

כאשר  $w_n$  היא פונקציית חלון המחליקה לאורך סדרת הכניסה. תפקידה הוא לבחור אינטרוול אשר יהיה מעורב בחישוב פונקציית האנרגיה לזמן קצר. ניתן לבחור בפונקציית חלון כלשהי, כגון חלון מלבני, חלון Hamming, וכו'. בניסוי זה נעבוד עם פונקציית חלון מלבנית:

$$w_n = \begin{cases} \frac{1}{\sqrt{N}} & 0 \leq n \leq N-1 \\ 0 & elsewhere \end{cases}$$

שימו לב שעבור חלון מלבני, פונקציית האנרגיה לזמן קצר היא הערך הממוצע של אנרגיית האות :

$$E_n = \frac{1}{N} \sum_{m=0}^{N-1} x_m^2$$

כפי שניתן להבין באופן אינטואיטיבי, קיים סיכוי גבוה כי הקטעים שבהם האנרגיה לזמן קצר של האות הדגום נמוכה מאוד הם קטעי שקט.

גם זיהוי קטעי אות קולי וקטעי אות א-קולי הוא פעולה עדינה המערבת שקלול של מספר פרמטרים. פרמטר המביא לרוב לתוצאות טובות הוא פונקציית האנרגיה לזמן קצר, שזה עתה הוצג. קטעים קוליים מאופיינים באנרגיה גבוהה, לעומת קטעים א-קוליים המאופיינים באנרגיה נמוכה.

פרמטר שימושי נוסף המבדיל בצורה טובה בין קטעים קוליים וקטעים א-קוליים הוא קצב חציות האפס (zero-crossing rate). קצב חציות האפס הוא מספר הפעמים בו דגימות עוקבות של האות משנות את סימןן האלגברי משלילי לחיובי או להפך. פונקציה זו מוגדרת באופן הבא :

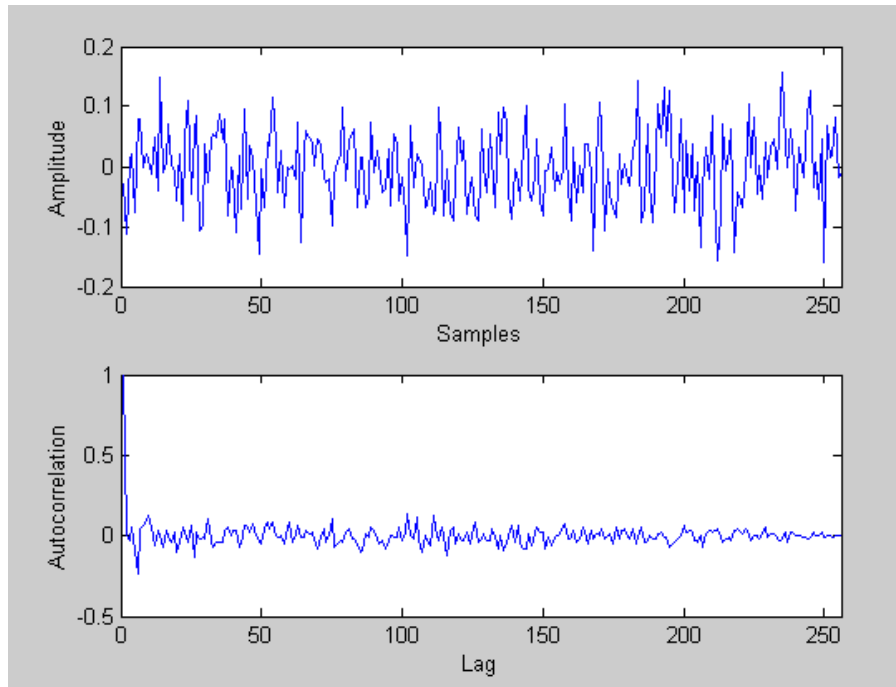
$$Z_n = \frac{1}{2} \sum_{m=n-N+1}^n |\text{sgn}(x_m) - \text{sgn}(x_{m-1})|$$

כאשר :

$$\text{sgn}(x) = \begin{cases} 1 & x_n \geq 0 \\ -1 & x_n < 0 \end{cases}$$

קצב חציות האפס הוא גבוה יותר בקטעים א-קוליים מאשר בקטעים קוליים.

אמצעי הבחנה נוסף בין קטעים קוליים לקטעים א-קוליים הוא צורת פונקציית האוטוקורלציה. לפונקציית האוטוקורלציה צורה שונה עבור אות קולי ועבור אות א-קולי מכיוון שבאות קולי קיימת קורלציה גבוהה בין מחזורי ה-pitch (כפי שניתן היה לראות בציור 3) ובאות א-קולי, שהוא דמוי רעש ובעל אקראיות גבוהה, הקורלציה בין אזורים שונים באות היא נמוכה (כפי שרואים בציור 5).



צירוף 7: קטע אות דיבור א-קולי ופונקציית האוטוקורלציה המתאימה

## חיזוי ליניארי

חיזוי ליניארי או LPC (Linear Predictive Coding) היא שיטה נפוצה לייצוג המעטפת הספקטרלית של אותות דיבור. שיטה זו עומדת בבסיסם של רוב מקודדי הדיבור המודרניים. היא משמשת במקודדים אלה לשערוך מקדמי המסנן המייצג את מערכת הקול במודל ליצירת אות הדיבור שהוצג. חיזוי ליניארי מאפשר לייצג את מעטפת הספקטרום של קטע דיבור קצר בעזרת מספר מועט של פרמטרים.

הנחה בסיסית הנובעת ממודל הדיבור היא שאות הדיבור ניתן לייצוג ע"י מודל אוטורגרסיבי מסדר  $p$  (AR(p) – autoregressive model of order p), כלומר, שכל דגימה שלו מקיימת:

$$x_n = \sum_{i=1}^p a_i x_{n-i} + u_n$$

כאשר  $\{a_i\}_{i=1}^p$  הוא סט של מקדמים קבועים המיוצג ע"י הווקטור  $\underline{a}$ , ו- $u_n$  הוא דגימה מתהליך רעש לבן בעל ממוצע אפס. לכן, ניתן להתייחס ל- $x_n$  כאל מוצא של מערכת ליניארית בעלת פונקציית תמסורת  $H(z)$  מסוג all-pole המעוררת ע"י רעש לבן:

$$H(z) = \frac{X(z)}{W(z)} = \frac{1}{A(z)}$$

כאשר:

$$A(z) \triangleq 1 - \sum_{i=1}^p a_i z^{-i}$$

אנו מעוניינים לשערך את הדגימה  $x_n$  מתוך סט של דגימות קודמות עוקבות:  $[x_{n-1}, x_{n-2}, \dots, x_{n-p}]$ , ע"י קומבינציה ליניארית שלהן:

$$x'_n = \sum_{i=1}^p a'_i x_{n-i}$$

כלומר, חוזים את ערכו של  $x_n$  ע"י חזאי ליניארי מסדר  $p$ . שגיאת החיזוי היא:

$$e(n) \triangleq x_n - x'_n = x_n - \sum_{i=1}^p a'_i x_{n-i}$$

כך שאם  $\underline{a}' = a$ , הרי ששגיאת החיזוי תהיה לבנה. בחירה נכונה של סדר המודל  $p$  תביא לקיום תכונה זו בקירוב.

שימו לב שניתן לקבל את שגיאת החיזוי ע"י סינון של האות במסנן  $A(z)$ :

$$e(n) \triangleq x_n - x'_n = x_n * \{1, -a'_1, \dots, -a'_p\}$$

כלומר, המודל שאנו עובדים איתו הוא:

$$excitation \rightarrow \boxed{\frac{1}{A(z)}} \rightarrow Speech$$

$$Speech \rightarrow \boxed{A(z)} \rightarrow excitation = e[n]$$

כך שאם נדע לשערך בקירוב טוב את מקדמי  $A(z)$ , האנרגיה של שגיאת החיזוי  $e(n)$  לאחר שנעביר את אות הדיבור במסנן  $A(z)$ , תהיה זהה לאנרגיה של אות העירור.

הבעיה שעומדת לפנינו היא לשערך את וקטור הפרמטרים  $\underline{a}'$ , ונרצה למצוא את וקטור המקדמים  $\underline{a}'$  האופטימלי במובן של מינימום שגיאה ריבועית ממוצעת (MMSE – Minimum Mean Squared Error), כלומר:

$$\underline{a}'_{opt} = \min_{\underline{a}'} \varepsilon^2 = \min_{\underline{a}'} E\{e_n^2\}$$

מטרתנו היא לשערך את וקטור המקדמים כך ששוונות שגיאת החיזוי  $\varepsilon^2$  תהיה מינימלית. לאחר גזירת הביטוי המתקבל עבור  $\varepsilon^2$  לפי כל אחד מהפרמטרים והשוואה לאפס נקבל את הסט הבא של משוואות לחישוב  $\underline{a}$  (בהנחה ש-  $e(n)$  לבנה):

$$\sum_{i=1}^p a_i \cdot r_{j-i} = r_j, \quad j=1,2,\dots,p$$

כאשר:

$$r_j \triangleq R_{yy}(j) = E\{x_n x_{n-j}\}$$

מערכת משוואות זו נקראת משוואות Yule-Walker או משוואות נורמליות.  
ברישום מטריצי:

$$R_p \underline{a} = \underline{r}$$

כאשר:

$$\underline{r} \triangleq [r_1, r_2, \dots, r_p]^T$$

ו-

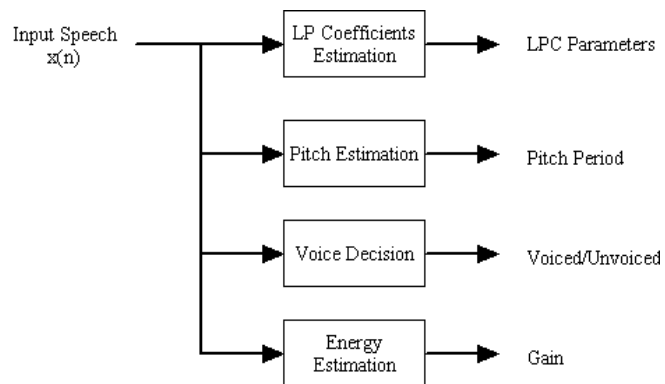
$$R_p \triangleq \begin{bmatrix} r_0 & r_1 & r_2 & \dots & r_{p-1} \\ r_1 & r_0 & r_1 & \dots & r_{p-2} \\ r_2 & r_1 & r_0 & \dots & \\ \vdots & & & & \vdots \\ r_{p-2} & & & r_0 & r_1 \\ r_{p-1}, r_{p-2}, & \dots & r_1 & r_0 \end{bmatrix}$$

דרך אחת לפתרון היא ע"י היפוך המטריצה -

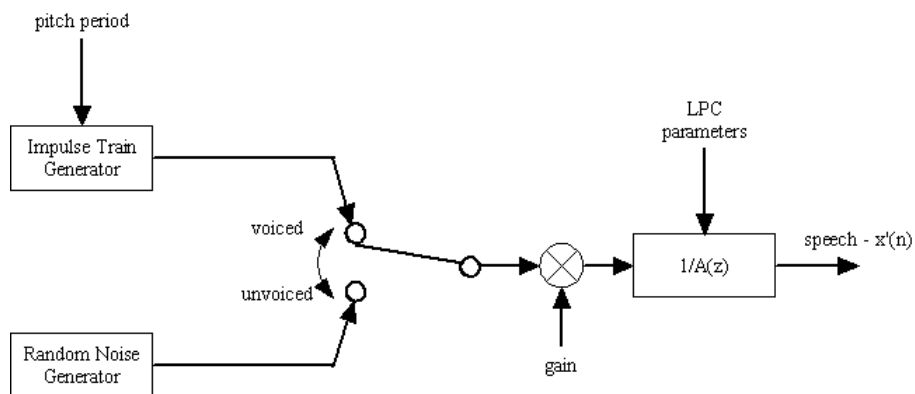
$$\underline{a} = R_p^{-1} \underline{r}$$

לפתרון מערכת המשוואות ללא היפוך ישיר של המטריצה משתמשים באלגוריתם חישוב רקורסיבי הנקרא אלגוריתם Levinson-Durbin (פונק' lpc ב-Matlab משתמשת באלגוריתם זה) אלגוריתם זה הוא מעבר להיקף המעבדה, אך ניתן למצוא עליו מידע במקורות לעיון נוסף המופיעים בסוף החוברת.

לסיכום הדיון, נוכל לבנות סכמה כללית של מקודד ומפענח מבוססי LPC. שימו לב שהמקודד מוציא כפלט ארבעה פרמטרים הנדרשים למפענח בשלבים שונים של יצירת אות הדיבור המשוחזר.



צור 8 : מקודד LPC



צור 9 : מפענח LPC



## מדדים אובייקטיביים לאיכות דיבור

כאשר מתכננים אלגוריתם לעיבוד דיבור קיים צורך להעריך את איכות הדיבור לאחר העיבוד לעומת איכותו לפני העיבוד. לא קיימת נוסחה המסוגלת לתת הערכה מושלמת של איכות דיבור משום שאיכות זו מושפעת מגורמים פסיכולוגיים וסביבתיים רבים וקשורה באופן הדוק לתפיסה הסובייקטיבית של כל אדם ואדם. לכן, במקרים רבים משתמשים לצורך הערכת איכות דיבור במדדים סובייקטיביים המבוססים על ניתוח תוצאות האזנה מבוקרת ודירוג של אנשים שונים.

עם זאת, קיים מאמץ בלתי פוסק למציאת נוסחה שתיתן תוצאות קרובות ככל האפשר לבדיקה סובייקטיבית. נציין בהקשר זה את מדד ה-PESQ (Perceptual Evaluation of Speech Quality) עפ"י תקן ITU P.862 הנותן תוצאות קרובות מאוד למדדים הסובייקטיביים בתחום  $[-0.5, 4.5]$ . בנוסף ל-PESQ, במסגרת מעבדה זו נעזר בשני מדדים אובייקטיביים קלאסיים ופשוטים לחישוב – SNR ו-Segmental SNR.

מדד ה-SNR (Signal to Noise Ratio) מחשב את היחס בין האנרגיה של האות המקורי לבין האנרגיה של הרעש (ההפרש בין האות המעובד לאות המקורי):

$$SNR = 10 \log_{10} \left( \frac{\sum_{n=0}^{N-1} x_n^2}{\sum_{n=0}^{N-1} (x_n - x'_n)^2} \right)$$

כאשר  $x$  הוא האות המקורי ו- $x'$  הוא האות המשוחזר.  $N$  הוא אורך מקטע האות שעליו מתבצע החישוב.

מדד ה-SNR נותן לאזורים באות בעלי אמפליטודה גדולה משקל גדול יותר מאשר לאזורים בעלי אמפליטודה נמוכה. עם זאת, שגיאות קטנות באזורים בעלי אמפליטודה נמוכה עלולות להיות חשובות. מדד Segmental SNR פותר בעיה זו ע"י חלוקת האות למקטעים קצרים וחישוב ה-SNR בכל אחד מהם בנפרד. התוצאה הסופית היא ממוצע ה-SNR בכל המקטעים. באופן זה, השגיאה מחושבת באופן יחסי לאנרגיה בכל מקטע וכך מקטעים בעלי אנרגיה נמוכה תורמים באופן שווה למקטעים בעלי אנרגיה גבוהה. החישוב מתבצע לפי הנוסחה הבאה:

$$SegmentalSNR = \frac{10}{M} \sum_{m=0}^{M-1} \log_{10} \left( \frac{\sum_{l=0}^{L-1} x_{mL+l}^2}{\sum_{l=0}^{L-1} (x_{mL+l} - x'_{mL+l})^2} \right)$$

כאשר  $x$  הוא האות המקורי,  $x'$  הוא האות המשוחזר ו- $ML$  הוא אורך האות. האות מחולק ל- $M$  מקטעים באורך  $L$  כל אחד. ערכים אופייניים ל- $L$  הם בין 10ms ל-20ms.

לעתים מגבילים את הסכימה רק למקטעים בהם עוצמת האות לא נופלת מערך מסוים, וזאת כדי למנוע ארגומנטים קטנים של פונקציית הלוגריתם.

## תוכנת SPDemo

SPDemo היא תוכנה חופשית המפותחת במעבדה לעיבוד אותות ותמונות ומהווה כלי עזר להוראה וללימוד עצמי של אלגוריתמים וטכניקות שונות בעיבוד דיבור וקול. התוכנה מאפשרת לפתוח קובץ WAV כלשהו, להפעיל עליו אלגוריתמים שונים תוך קביעת פרמטרים רלוונטיים ולצפות בתוצאות. ניתן להציג תוצאות ביניים וכן תוצאות סופיות, לצפות בהן גרפית או טקסטואלית ואף להאזין להם. כל זאת כדי לעודד התנסות ולהקל על ההבנה של אופן פעולת האלגוריתמים השונים.

מכיוון ש-SPDemo היא תוכנה קלה לשימוש, לא נפרט כאן הוראות להפעלתה (במידת הצורך, ניתן להיעזר במדריך הניסוי). נציין כאן רק מספר יכולות שלה שהן רלוונטיות לניסוי זה:

- ניתן לצפות באות או במספר אותות ברמות הגדלה שונות עד רמת הדגימה הבודדת.
- ניתן לראות ספקטרוגרמה של האות (מושג זה יוסבר בחלק ב' של הניסוי) תוך קביעת פרמטרי ספקטרוגרמה שונים.
- המודול SPFTR מאפשר לבצע פעולות בסיסיות על אות דיבור, כגון אוטוקורלציה, קטימה, חישוב אנרגיה לזמן קצר...
- המודול Q-EVAL מאפשר לבצע השוואה אובייקטיבית בין אותות שונים בשיטת SNR, Segmental PESQ, SNR ובשיטות נוספות.

בהמשך הניסוי מומלץ להשתמש ב-SPDemo לצורך אנליזה (בדיקת איכות התוצאות) באמצעות השוואה לתוצאות שתקבלו.

הערה: תוכנת SPDemo מותקנת על מחשבי הניסוי. כמו כן, ניתן להוריד אותה מאתר המעבדה ולהתקין בבית.

## שאלות הכנה לפגישת הניסוי הראשונה

לפני המפגש במעבדה, יש להגיש את הדו"ח המכין (במערכת LabAdmin!) ( המכיל תשובות לשאלות הכנה אלו, תוכניות ה-Matlab שתכתבו וגרפים, במידת הצורך.

- כדאי להביא את קבצי פתרונות ה-Matlab גם בצורה אלקטרונית.
- הקפידו על דו"ח מסודר וברור.

### שאלה 1

- א. נתון אות בזמן רציף המורכב מסכום של שלושה סינוסים בעלי אמפליטודות יחידה ותדרים של 1KHz, 2KHz ו-3KHz. האות נדגם בתדר דגימה של 10KHz ונשמרים 256 הדגמים הראשונים שלו. הציגו את ספקטרום האות (ליתר דיוק, התמרת פורייה שלו) ע"י הכפלה בחלון Hamming וביצוע DFT. הציגו את ציר התדר ביחידות של Hz (היזכרו במשפט הדגימה ובקשר בין נקודות הדגימה של ה DFT לבין ציר התדר). הסבירו את התוצאה המתקבלת.
- ב. כעת נדגם באותו קצב דגימה אות רציף המורכב מסכום של שלושה אותות סינוסים בעלי אמפליטודות יחידה ותדרים של 1KHz, 2KHz ו-9KHz. הציגו באותו אופן את ספקטרום האות. הסבירו את התוצאה המתקבלת באופן אנליטי ותארו סכמתית במישור התדר את תהליך הדגימה שהביא לתוצאה זו.

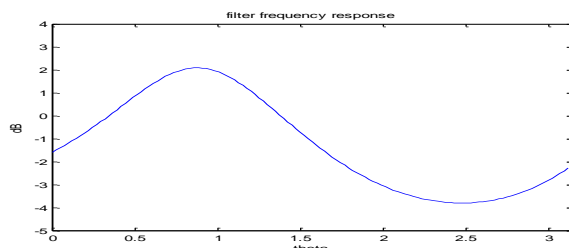
### שאלה 2

- א. מה נקבל אם נבצע התמרת פורייה על רכבת ההלמים הבאה :

$$\sum_{n=-\infty}^{\infty} \delta(t-nT) \xrightarrow{F} ?$$

הוכיחו את תשובתכם.

- ב. נתון מסנן H שצורת תגובת התדר שלו היא :



כאשר נעביר את רכבת ההלמים דרך מסנן זה, מה תהיה צורת אות המוצא בתדר? ציירו והסבירו איכותית.

$$\sum_{n=-\infty}^{\infty} \delta(t-nT) \rightarrow [H] \rightarrow ?$$

שאלה 3

- א. רשמו ביטוי כללי למסנן FIR בזמן, וביטוי תואם בתדר. מהי תלות מסנן FIR בזמן עבר? האם התלות היא בכניסות או ביציאות?
- ב. רשמו ביטוי כללי למסנן IIR בזמן, וביטוי תואם בתדר. מהי תלות מסנן IIR בזמן עבר? האם התלות היא בכניסות או ביציאות?
- ג. הסתכלו בתיעוד של הפונקציה freqz ב-Matlab וענו על הסעיפים הבאים:
1. מה עושה הפונקציה?
  2. מה משמעות הפרמטרים a ו-b לפונקציה?
  3. היזכרו בקשר בין תצורת הקטבים/אפסים של מסנן במישור המרוכב לבין תגובת התדר שלו. ציירו והסבירו איכותית.

שאלה 4

בהתייחס לפונקציית האוטוקורלציה:

$$r_k = \sum_{n=0}^{N-k-1} x_n x_{n+k} \quad k = 0, \dots, N-1$$

ובהנחה כי האות מתאפס מחוץ לתחום שבין 0 ל-(N-1), הוכיחו כי  $r(k)$  סימטרית, כלומר, הוכיחו כי:  $r_k = r_{-k} \quad k = 0, 1, 2, \dots, N-1$ .

שאלה 5

- הסבירו (אין צורך בהוכחה פורמלית) מדוע פונקציית האוטוקורלציה מקיימת את התכונות הבאות:
- המקסימום הגלובלי שלה מתקבל ב  $k=0$  והוא שווה לאנרגיית האות.
  - אם האות  $x(n)$  מחזורי במחזור  $P$ , מתקבלות מקסימום מקומיות ב-  $0, \pm P, \pm 2P, \pm 3P, \dots$ .

שאלה 6

ממשו ב-Matlab פונקציה בשם pitch\_detect\_corr המקבלת כקלט מסגרת באורך 256 דגימות שנדגמה ב-8KHz ומחזירה כפלט את ערך מחזור ה-pitch (במספר דגימות) המאפיין את המסגרת. השתמשו בשיטת האוטוקורלציה לשערוך מחזור ה-pitch.

הנחיה 1: בעת חישוב מחזור ה-pitch יש להתחשב רק בתחום התדרים התקין עבור התוצאה, כלומר, יש לאפס ערכים שאינם אפשריים מבחינה פיזיולוגית.

הנחיה 2: כדי לחשב את פונקציית האוטוקורלציה של הקטע, היעזרו בפונקציה xcorr. כדי למצוא את מיקום השיא, היעזרו בפונקציה find.

שאלה 7

- ממשו ב-Matlab פונקציה בשם clip המקבלת כקלט מסגרת באורך 256 דגימות שנדגמה ב-8KHz ומבצעת עליה קטימה לא-לינארית, כפי שהיא מופיעה בפרק הרקע התיאורטי. ערכו של  $C_L$  ייקבע להיות 65% מהערך המוחלט המקסימלי בקטע.
- הסבירו את פעולתה של הפונקציה ע"י ציור גרף איכותי של אות סינוסי לפני ואחרי פעולת הקטימה. סמנו את  $C_L$  ואת  $-C_L$  כקווים בגרף.

שאלה 8

תכננו מסנן FIR מסוג מעביר-נמוכים. על המסנן לפעול על אותות בקצב דגימה של 8KHz. דרישות המסנן הן הבאות:

- תדר הקיטעון (ירידה של 3dB) יהיה ב-900Hz.
- על המסנן לקבל ניחות של לפחות 13dB - בתדר 1000Hz.

- מצאו מסנן העומד בדרישות הנ"ל בעל פחות מ-60 מקדמים. ניתן להיעזר בכלים sptool או fdatool.
- צרפו לדו"ח המכין גרף המתאר את צורת הספקטרום של המסנן שתכנתם ואת עמידתו בדרישות (ניתן להיעזר בפונקציה freqz).
- שימרו את מקדמי המסנן שקיבלתם בקובץ (ב-fdatool – ע"י file-export), על מנת שתוכלו להשתמש בו במהלך הניסוי.

שאלה 9

ממשו ב-Matlab פונקציה בשם zero\_cross המקבלת כקלט מסגרת באורך 256 דגימות ומחשבת את מספר חציות האפס שהתרחשו במסגרת.

שאלה 10

ממשו ב-Matlab פונקציה בשם pitch\_detect\_ceps המקבלת כקלט מסגרת באורך 256 דגימות שנדגמה ב-8KHz ומחזירה כפלט את ערך מחזור ה-pitch (במספר דגימות) המאפיין את המסגרת. השתמשו בשיטת הקפסטרום לשערוך מחזור ה-pitch.

**הנחיה 1:** בעת חישוב מחזור ה-pitch יש להתחשב רק בתחום התדרים התקין עבור התוצאה, כלומר, יש לאפס ערכים שאינם אפשריים מבחינה פיזיולוגית.

**הנחיה 2:** כדי למצוא את מיקום השיא, היעזרו בפונקציה find.

## מהלך הניסוי, פגישה ראשונה

בניסוי זה נבנה מקודד דיבור פשוט המבוסס על חיזוי ליניארי מסדר 10. לשם כך נכיר את מאפייני אות הדיבור ונממש מספר בלוקים בסיסיים שירכיבו את המקודד המלא. בלוקים אלו הם:

- סיווג אות כ- Voiced / Unvoiced.
- שיעור מחזור ה-pitch.
- שיעור האנרגיה הממוצעת של אות העירור.
- שיעור מקדמי החזאי הלינארי (מקדמי ה-LPC).

בפגישת הניסוי הראשונה נעסוק בנושאים כלליים בעיבוד אות ובהכרת המאפיינים העיקריים של אות דיבור. לאחר מכן נבנה את שני הבלוקים הראשונים. את שאר הבלוקים, ואת הרכבת מערכת המקודד-מפענח נבצע בפגישת הניסוי השנייה, וגם נבדוק את ביצועיה עם אותות דיבור "אמיתיים".

מספר הנחיות לגבי הצגת גרפים (גם בדו"ח המסכם):

- גרף המייצג תמונת תדר (ספקטרום) ייוצג ביחידות של dB, וציר התדר שלו יהיה ביחידות של Hz. יש להציג רק את החצי הראשון (החיובי:  $[0, \pi]$ ) של ציר התדר.
- על כל אחד מהגרפים יש לציין את היחידות של הצירים ולרשום כותרת מתאימה (בעזרת הפונקציות title, xlabel, ylabel, axis).
- הצגת הגרפים תהיה בצירוף הסברים מתאימים, אפילו אם לא נאמר כך במפורש.

## 1. עיבוד אותות בסיסי

- 1.1 שימוש ב-DFT לצורך שיעור ספקטרום  
הגדירו שני אותות סינוסיים: הראשון בתדר 256Hz והשני בתדר 250Hz. האותות נדגמו בתדר דגימה של 8000Hz, ואורכם 256 דגמים כל אחד.  
  - הציגו את ערכה המוחלט של התמרת ה-DFT של כל אחד מהאותות בנפרד. השתמשו בפונקציה stem (אין צורך לעבור ליחידות של dB).
  - מה ההבדלים בין שתי תמונות התדר ומה מקורם?
  - הציגו את אותן ההתמרות בגרף רציף (פונקציית plot) וביחידות של dB. האם פונקציית הדלתא עדיין נראית בברור?

- 1.2 עיצוב ספקטרום של רעש לבן ע"י סינון  
צרו סדרה אקראית, מפולגת נורמלית עם תוחלת אפס ושונות יחידה (היעזרו בפונקציה randn) באורך 1000 דגימות.  
  - מהו הספקטרום התיאורטי של אות כזה?
  - שערכו בעזרת שיטת הפריודוגרמה (ר' הסבר) את הספקטרום המתקבל מתוך 1000 דגימות הרעש שיוצרו והציגו אותן על גרף לוגריתמי.

הסבר – ספקטרום של אות אקראי (צפיפות הספק ספקטרלית) מחושב באופן תאורטי באמצעות התמרת פורייה של פונק' האוטוקורלציה. פונק' האוטוקורלציה  $R_X(\tau)$  של תהליך אקראי  $X(n)$ , סטציונרי במובן הרחב מוגדרת באופן הבא:

$$R_X(\tau) = E[X(n) \cdot X(n-\tau)]$$

תהליך רעש לבן הוא תהליך סטציונרי במובן הרחב וערכיו בזמנים שונים בת"ס ביניהן, ז"א:

$$\forall n \neq k, E[X(n)X(k)] = E[X(n)] \cdot E[X(k)]$$

כאשר אין בידינו את המודל של האות האקראי מקובל לחשב מתוך דגימות האות את הספקטרום. קיימות מספר שיטות שהפשוטה שבהן (אך לא המדויקת שבהן) היא שיטת הפריודוגרמה שמחושבת באופן הבא :

$$\text{Periodogram\_spectrum} = (\text{abs}(\text{fft}(\text{signal})))^2 / \text{signal\_length}$$

- אם תרצו לקבל תוצאות משופרות ניתן להשתמש בשיטת Welch לשערוך הספקטרום. בשיטה זו מבצעים מיצוע על מספר פריודוגרמות.
- כעת נעביר את **הסדרה האקראית** דרך המסנן שיצרתם בשאלת הכנה מס' 8 (בעזרת הפונקציה filter).
- מה צריך להיות הספקטרום התיאורטי של אות המוצא?
  - שערכו את ספקטרום האות במוצא המסנן בשיטת הפריודוגרמה.
  - האם התוצאה שהתקבלה מתאימה לתוצאה התאורטית?
  - היכן תבוא תוצאה זו לידי ביטוי במודל עיבוד הדיבור שהצגנו?

## 2. הכרת אות הדיבור ותכונותיו

2.1. בחירת קובץ הקלט המתאים  
חברו את שתי הספרות האחרונות של מספרי הסטודנט שלכם. את המספר המתקבל חלקו ב-4 ומצאו את השארית. קובץ הקלט שלכם ייקבע לפי השארית. לדוגמה: עבור שארית 1, קובץ הקלט יהיה speech1.wav. קובץ זה נדגם בקצב של 8KHz. טענו את קובץ הקלט שלכם ל-Matlab ע"י שימוש בפונקציה audioread והשמיעו אותו ע"י שימוש בפונקציה soundsc.

2.2. ריפוד באפסים

- בהנחה שכל דגימה מיוצגת ע"י 16 סיביות (2 בתים), כמה מקום צורך כל הקטע?
- רפדו את קטע האות באפסים כך שאורכו יהיה כפולה שלמה של 256.

2.3. התבוננות באות הדיבור

- הציגו את קטע האות על המסך. האם אתם מזהים בעין את האזורים הנראים כקוליים? סמנו אותם בגרף. האם אתם מזהים בעין אזורים הנראים כא-קוליים? סמנו גם אותם בגרף.
- הנחיה: רצוי להיעזר בשלב זה בתוכנת SPDemo.
- נסו לקבוע מהסתכלות מספר תכונות המאפיינות אותות קוליים ותכונות המאפיינות אותות א-קוליים (צורת האות, עוצמת האות וכד').

2.4. בחירת מסגרת קולית ומסגרת א-קולית

- חלקו את האות למסגרות באורך 256 דגימות כל אחת (ניתן להשתמש בפונקציה reshape, שימו לב שהיא תמיד מסדרת את האיברים לאורך העמודות, ולא לאורך השורות).
- בחרו מסגרת של 256 דגימות מתוך אחד מהאזורים הקוליים שסימנתם בסעיף 2.3 והציגו אותה. ציינו מאיזו מילה ומאיזו הברה נלקחה המסגרת. מסגרת זו תיקרא בהמשך מסגרת הדוגמה הקולית.
- בחרו מסגרת של 256 דגימות מתוך אחד מהאזורים הא-קוליים שסימנתם בסעיף 2.3 והציגו אותה. ציינו מאיזו מילה ומאיזו הברה נלקחה המסגרת. מסגרת זו תיקרא בהמשך מסגרת הדוגמה הא-קולית.

← **וודאו בשלב זה שמדריך הניסוי מאשר את בחירתכם!** בחירה לא נכונה תשפיע על המשך הניסוי ועלולה לעלות לכם בזמן יקר...

2.5. שערך פרמטרים של מסגרת קולית במישור התדר

- בצעו התמרת DFT למסגרת הדוגמה הקולית והציגו אותה בתחום הזמן ובתחום התדר. האם ניתן להבחין במעטפת ספקטרלית בתמונת ההתמרה? (היזכרו בשאלת הכנה 2. היכן אנו רואים כעת את המעטפת הספקטרלית של האות לאחר שעבר במסך?) סמנו אותה על גבי הגרף.
- שערכו את מחזור ה-pitch מתוך הסתכלות במסגרת הדוגמה הקולית (השתמשו באופציית zoom על מנת לקבל שערך טוב יותר). ציינו את אורך מחזור ה-pitch בדגימות וחשבו גם את התדר שלו.
- שערכו את מחזור ה-pitch מתוך התמרת ה-DFT של מסגרת הדוגמה הקולית. על מנת להקל על השערוך, כדאי להכפיל בחלון ולרפד באפסים לפני ביצוע ה-DFT.
- הסבירו כיצד הכפלה בחלון וריפוד באפסים מקלה על השערוך.
- בדקו האם קיימת התאמה בין תדרי ה-pitch בשני אופני החישוב הנ"ל.



## 2.6. עיבוד מקדים

- סננו את קטע האות המקורי (כולו) ע"י המסנן שתוכנן בשאלת הכנה 8.
- בודדו שוב את מסגרת הדוגמה הקולית מהאות המסונן (אותה מסגרת שנבחרה קודם) והשתמשו במסגרת המסוננת כקלט לפונקציה clip שכתבתם בבית. כעת חזרו על הפעולות שביצעתם בסעיף 2.5 והסבירו את ההבדלים בתוצאות.

## 2.7. שערך פרמטרים של מסגרת א-קולית במישור התדר

- בצעו התמרת DFT למסגרת הדוגמה הא-קולית והציגו אותה בתחום הזמן ובתחום התדר.
- האם גם כאן ניתן להבחין במחזור pitch? הסבירו את תשובתכם.

## 2.8. שערך מחזור ה-pitch עבור מסגרת קולית במישור הזמן

- (1) הציגו את פונקציית האוטוקורלציה של המסגרת (היעזרו בפונקציה xcorr) ונסו לזהות בעין את אורך מחזור ה-pitch על פי גרף הפונקציה.
- (2) הפעילו את הפונקציה pitch\_detect\_corr שכתבתם בבית על מסגרת הדוגמה הקולית המקורית (ללא עיבוד מקדים) והשוו את התוצאה המתקבלת לתוצאות שקיבלתם מהסתכלות.
- (3) כעת יש לכם את אורך מחזור ה-pitch בדגימות. חשבו מתוך ערך זה את תדר ה-pitch והשוו לתוצאות שקיבלתם בסעיף 2.5.
- (4) הציגו גרף של פונקציית האוטוקורלציה, הפעם על המסגרת לאחר שעברה עיבוד מקדים (סינון וקטימה). השוו לגרף שקיבלתם עבור המסגרת הרגילה.
- (5) הפעילו שוב את הפונקציה pitch\_detect\_corr, הפעם על המסגרת לאחר שעברה עיבוד מקדים. האם מתקבלת אותה התוצאה?
- (6) כיצד תורם העיבוד המקדים לאיכות השערך של מחזור ה-pitch?
- (7) חזרו על סעיפים (2)-(6) עבור הפונקציה pitch\_detect\_ceps.
- (8) השוו בין הביצועים של שתי השיטות.

## 2.9. פונקציית אוטוקורלציה של מסגרת א-קולית

- הציגו את פונקציית האוטוקורלציה של מסגרת הדוגמה הא-קולית.
- מהם ההבדלים העיקריים בין הגרף המתקבל, לגרפים שהתקבלו עבור מסגרת הדוגמה הקולית?

## 3. סיווג לאותות קוליים/א-קוליים

## 3.1. מימוש הפונקציה vu\_classify

ממשו ב-Matlab פונקציה בשם vu\_classify המקבלת כקלט שתי מסגרות באורך 256 דגימות כל אחת - האחת היא המסגרת של האות המקורי, והשנייה היא אותה המסגרת לאחר שעברה עיבוד מקדים הכולל סינון וקטימה, כפי שביצעתם בפגישת הניסוי הראשונה בסעיף 2.8. הפונקציה תחזיר כפלט ערך בוליאני:

1 – המסגרת היא קולית

0 – המסגרת אינה קולית (א-קולית או שקט)

בשלב ראשון הפונקציה תבדוק את ערך האנרגיה הממוצעת של המסגרת על פני חלון מלבני באורך המסגרת (לפי הנוסחה לאנרגיה בזמן קצר).

אם ערך זה קטן מ- $10^{-5}$  נניח כי מדובר במסגרת המכילה "שקט" ונחזיר ערך '0'. אחרת, הפונקציה תחליט שהמסגרת היא א-קולית אם מתקיימים **לפחות שניים** מתוך ארבעת התנאים הבאים:

1. מספר חציות האפס במסגרת גדול מ-128 (50% מאורך המסגרת).
2. ערך האנרגיה הממוצעת במסגרת (ששוערך קודם) קטן מאחוז אחד מהערך המוחלט המקסימלי במסגרת.
3. בחישוב פונקציית האוטוקורלציה, השיא המקומי הגדול ביותר שאינו ממוקם באפס, ממוקם באזור שאינו יכול להכיל ערכי pitch חוקיים.
4. בחישוב פונקציית האוטוקורלציה, היחס בין ערך השיא המקומי הגדול ביותר שאינו ממוקם באפס לבין ערך הפונקציה באפס, קטן מ-0.35.

**הנחיה 1:** את מספר חציות האפס ואת האנרגיה הממוצעת במסגרת יש לחשב על המסגרת המקורית, ואילו את פונקציית האוטוקורלציה יש לחשב על המסגרת שעברה עיבוד מקדים.

**הנחיה 2:** את התנאי השלישי ניתן למצוא באופן הבא: חפשו את המינימום המקומי הראשונה של פונקציית האוטוקורלציה (**מכיוון שפונקציית האוטוקורלציה סימטרית, קצצו את צידה השמאלי, והביטו רק בימני, החל מ-0 ואילך**), והחל מימינה חפשו את המקסימה.

- במימוש של `vu_classify` לא ניתן להשתמש בפונקציה `pitch_detect_corr` כפונקציית עזר. מדוע?
- בתנאי השלישי, מדוע אנו מחפשים את השיא המקומי הגדול ביותר שאינו ממוקם באפס?

3.2. ההבדלים בין אות קולי לאות א-קולי

- חשבו את האנרגיה הממוצעת של כל אחת ממסגרות הדוגמה. למי אנרגיה ממוצעת גדולה יותר?
- בדקו את הפונקציה `zero_cross` שכתבתם בבית על שתי מסגרות הדוגמה. למי קצב חציות אפס גדול יותר?
- בדקו את הפונקציה `vu_classify` שכתבתם על שתי מסגרות הדוגמה.

## שאלות הכנה לפגישת הניסוי השניה

יש להגיש את הדו"ח המכין **במערכת LABADMIN** לפני פגישת הניסוי השניה. בדוח יהיו תשובות לשאלות הכנה אלו, תוכניות ה-Matlab שתכתבו וגרפים, במידת הצורך.

- כדאי להביא את קבצי פתרונות ה-Matlab גם בצורה אלקטרונית.
- הקפידו על דו"ח מסודר וברור.

חלק זה הוא המשך ישיר של חלקו הראשון של הניסוי, ולכן מומלץ לרענן את הזיכרון לפני הפגישה השנייה ע"י קריאה חוזרת של התיאוריה ושל מהלך הניסוי בפגישה הראשונה. כמו כן, יש להביא את קוד ה-Matlab שכתבתם בפגישה הראשונה, מכיוון שהוא ישמש אתכם גם בפגישה זו.

מספר הנחיות לגבי הצגת גרפים (גם בדו"ח המסכם):

- גרף המייצג תמונת תדר (ספקטרום) ייוצג ביחידות של dB, וציר התדר שלו יהיה ביחידות של Hz. יש להציג רק את החצי הראשון (החיובי:  $[0, \pi]$ ) של ציר התדר.
- על כל אחד מהגרפים יש לציין את היחידות של הצירים ולרשום כותרת מתאימה (בעזרת הפונקציות title, xlabel, ylabel, axis וכו').
- הצגת הגרפים תהיה בצירוף הסברים מתאימים, אפילו אם לא נאמר כך במפורש.

### שאלה 1

כיצד לדעתכם יישמע הפלט של מערכת מקודד-מפענח דיבור אם נכניס למערכת זו כקלט אות שאיננו אות דיבור? מדוע? האם צפוי להיות שוני באיכות הפלט עבור קלט של מוסיקה קלאסית, למשל?

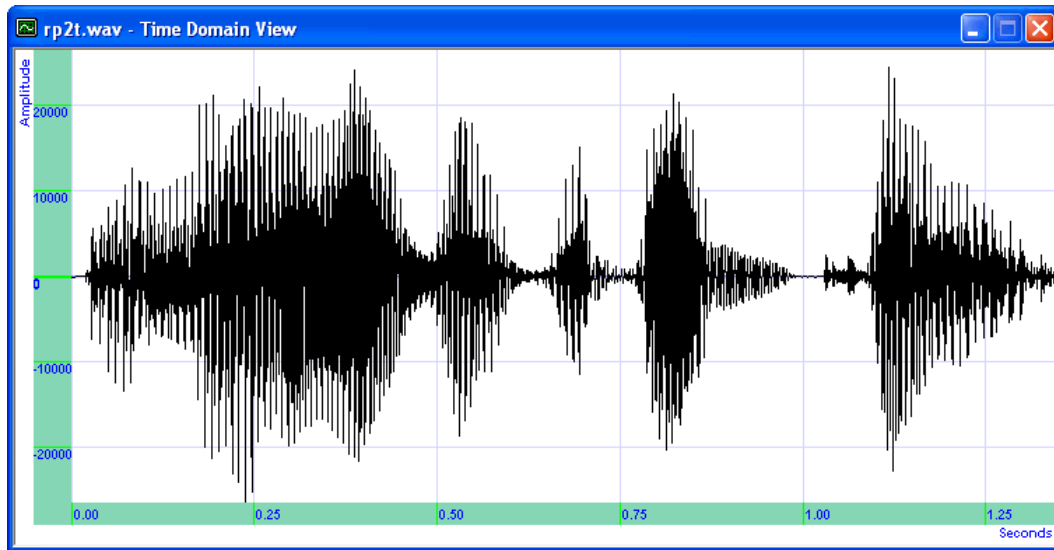
### שאלה 2

אלגוריתם דחיסה בטכניקת MELP (Mixed Excitation Linear Predication) לא מבצע החלטה אחת על עירור קולי או עירור א-קולי עבור כל קטע זמן, אלא משתמש במודל המשלב את שניהם. האלגוריתם מחלק את אות הדיבור למספר פסי תדר, ובכל פס תדר קובע האם העירור יהיה קולי או א-קולי. אות העירור הסופי יתקבל משילוב כל אותות העירור בפסי התדר השונים. מדוע מאפשר מודל זה לייצר דיבור באיכות טובה יותר מאשר המודל שהוצג בפרק התיאוריה?

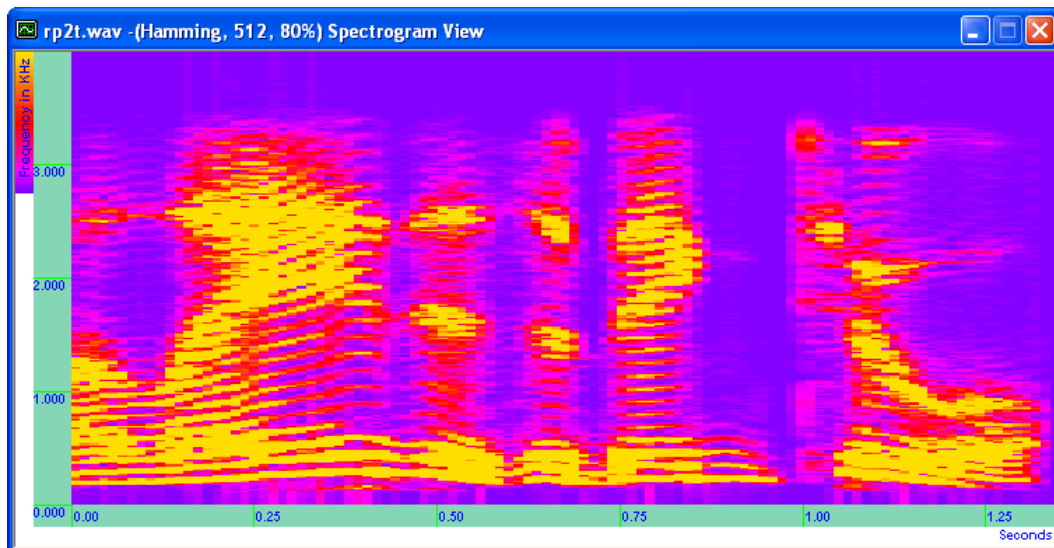
### שאלה 3

ספקטרוגרמה (spectrogram) הוא גרף דו-ממדי המורכב מאוסף התמרות DFT המבוצעות על קטעים קצרים ולעתים חופפים של אות. הציר האופקי הוא ציר הזמן והציר האנכי הוא ציר התדר. ערכי הספקטרום השונים מסומנים בגוונים שונים (רמות אפור או רמות צבע). גרף מסוג זה הוא כלי שימושי לניתוח מאפייני אות דיבור.

הגרף הבא מתאר את המשפט "A lathe is a big tool" הנאמר ע"י דובר זכר:



הספקטרוגרמה הבאה מתארת בגוויי סגול כהה (נמוך) עד צהוב בהיר (גבוה) את אותו משפט :



ספקטרוגרמה זו נוצרה ע"י DFT עם חלונות Hamming בגודל 512 דגימות כל אחד עם חפיפה של 80% בין חלון לחלון.

- מה המידע שניתן לראות ע"י התבוננות בחתך אנכי של הספקטרוגרמה? מה המידע שניתן לראות ע"י התבוננות בחתך אופקי של הספקטרוגרמה?
- מהם היתרונות והחסרונות בהגדלת/הקטנת גודל החלון המשמש לציור ספקטרוגרמה?
- סמנו על הציור האופקי של הספקטרוגרמה הנ"ל את המילים במשפט הנאמר.
- סמנו על הציור האנכי של הספקטרוגרמה את תדר ה-pitch בזמן  $t=0$ .

הערה: את תמונת הספקטרוגרמה הנ"ל ניתן לייצר בעזרת SPDemo ולהעתיק אותה לתוך מסמך התשובות שלכם.

שאלה 4

ע"י מודל החיזוי הלינארי שהצגנו, ע"י קירוב המסנן  $A(z)$ , מתאפשר לנו למצוא את האנרגיה של אות העירור שהוכנס למערכת. כך, כשנממש את המפענח יהיה באפשרותנו ליצור אות עירור שהאנרגיה שלו זהה לאנרגיה של אות העירור המקורי. בשאלה זו נממש את הפונקציה אשר בהינתן המסנן  $A(z)$  ואות הדיבור, מוצאת את האנרגיה של אות העירור.

ממשו ב-Matlab פונקציה בשם residual\_energy המקבלת כקלט מסגרת באורך 256 דגימות ווקטור מקדמי מסנן מסוג  $A(z)$  (כפי שמוגדר בחלק המבוא התיאורטי) ומחשבת את האנרגיה הממוצעת של **שגיאת החיזוי** במסגרת. אנו נשתמש בערך זה כדי לקבוע את עוצמת אות העירור, על מנת שהאות המשוחרר יהיה באותה עוצמה כמו האות המקורי.

החישוב יתבצע באופן הבא: בשלב ראשון יש לסנן את המסגרת ע"י המסנן הנתון 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).

- במודל המקודד והמפענח שהצגנו בחלק המבוא התיאורטי, המקודד מוציא למפענח ארבעה פלטים. לאיזה מן הפלטים אחראית residual\_energy? הסבירו.
- היכן נשתמש בפונקציה זו? במקודד או במפענח?
- מדוע אנו משתמשים במסנן הנתון כמסנן FIR?

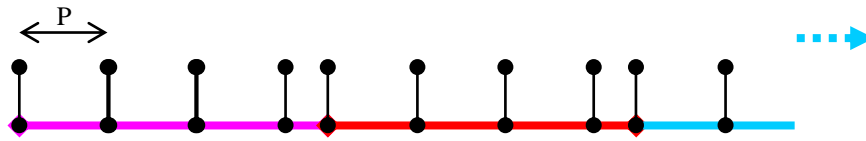
שאלה 5

מעוניינים לבצע סינון של רכבת הלמים באורך 120 דגימות:  $1 \leq n \leq 120$ ,  $x[n] = \delta(n \bmod 10)$ . ע"י מסנן IIR שפולינום המונה שלו הוא:  $P(z)=1$  ופולינום המכנה שלו הוא:  $Q(z)=1-0.8z^{-1}$ . סדרת הדגימות מתקבלת בשלושה מקטעים של 40 דגימות בכל פעם.

- בצעו סינון של שלושת המקטעים ע"י שימוש בפונקציה filter ושרשרו את הקטעים המסוננים זה לזה לקבלת סדרה ארוכה בת 120 דגימות. הציגו גרף של סדרה זו מול גרף של הסדרה המתקבלת ע"י סינון של כל 120 הדגימות בבת אחת. האם התוצאות זהות? מדוע?
- על מנת לפתור את הבעיה שנוצרה צריך "להעביר מידע" ממקטע אחד לשני. כדי לבצע זאת נפעיל את הפונקציה filter עם תנאי התחלה בפורמט הבא:  $[y, zi] = \text{filter}(P, Q, X, zi)$ . כאשר  $zi$  יאותחל לערך אפס, ולאחר סינון כל מקטע נבצע  $zi=zf$ . בצעו סינון של שלושת המקטעים באופן הזה והשוו שוב לגרף המתקבל מסינון כל 120 הדגימות בבת אחת. האם התוצאות כעת זהות? מהו המידע המועבר בתנאי ההתחלה? הסבירו את השיפור שהתבצע באופן איכותי (ניתן להיעזר בתיעוד של Matlab). אם המסנן הוא IIR, מהו המידע שיש להעביר (כניסות/יציאות קודמות)? ומה עבור מסנן FIR?
- כעת מעוניינים לבצע את פעולת הסינון על סדרת דגימות ארוכה המחולקת למקטעים של 256 דגימות עם מסנן במכנה באורך L (כאשר  $L < 256$ ), ומסנן מונה  $P(z)=1$ . כמה ערכים נצטרך לשרשר בתחילת כל מקטע? כמה ערכים נצטרך לשרשר עבור מסנן מונה באורך L-1 ומסנן מכנה באורך 1. העזרו בתיעוד עבור פונק' filter.
- היכן בבניית המקודד והמפענח נעשה שימוש במסנן IIR? הסבירו.

שאלה 6

מעוניינים לייצר אות הדומה לאות המקורי בשאלה 5. 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 שרשרו המקטעים, 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. כתבו פונקציית Matlab שתייצר את המקטעים הדרושים, כפונקציה של  $N$  ו- $P$ .

### שאלה 7

- בעת חישוב מדד Segmental SNR נוצרת בעיית חישוב כאשר שני האותות המשווים זהים במקטע אחד או יותר מבין  $L$  המקטעים שעליהם מחושב הממוצע. מוצעות שתי דרכים לפתרון הבעיה:
1. המקטעים בהם מתקיים השוויון לא יכנסו לשקלול ומדד השגיאה יבוצע אך ורק על מקטעים שונים עליהם ניתן לחשב את ה-Segmenal SNR.
  2. להשתמש בנוסחה

$$SegmenalSNR = \frac{10}{M} \sum_{m=0}^{M-1} \log_{10} \left( \frac{\sum_{l=0}^{L-1} x_{mL+l}^2}{\sum_{l=0}^{L-1} (x_{mL+l} - x'_{mL+l})^2 + \sigma^2} \right)$$

כאשר התוספת  $\sigma^2$  במכנה קטנה כרצוננו ובאה למנוע בעיות נומריות. איזו הצעה נכונה יותר? הסבר מדוע.

## מהלך הניסוי, פגישה שניה

המשיכו עם אותו קובץ דגימות מהפגישה הקודמת ועם אותן מסגרות דוגמה (קולית וא-קולית).

### 1. סיווג לאותות קוליים/א-קוליים

- 1.1 בדיקת `vu_classify` על האות כולו
- צרו גרסה מסוננת של האות המקורי, כפי שביצעתם בפגישת הניסוי הראשונה בסעיף 2.6 וחלקו את האות המקורי ואת האות המסונן למסגרות של 256 דגימות כל אחד.
- כתבו לולאה שתריץ את הפונקציה `vu_classify` שכתבתם בפגישת הניסוי הראשונה, על כל המסגרות. בתוך הלולאה, מצאו את ערך מחזור ה-`pitch` של כל מסגרת קולית ע"י הפעלת הפונקציה `pitch_detect_corr` על המסגרת, לאחר שעברה סינון וקטימה.
  - צרו וקטור המכיל את ערכי מחזור ה-`pitch` של כל המסגרות כאשר מסגרות שאינן קוליות יקבלו ערך אפס. הציגו את האות המקורי ומתחתיו את ערכי הווקטור שיצרתם **על אותו הגרף** ע"י שימוש בפונקציה `subplot`. האם התוצאה המתקבלת מתיישבת עם ציפיותיכם?
  - האם ערכי מחזור ה-`pitch` של הקטעים הקוליים זהים לאורך כל הקטע? הסבירו.

### 2. שערך מקדמי החזאי הליניארי

- 2.1 שערך מקדמי החזאי
- שערך מקדמי ה-LPC יתבצע תוך שימוש בפונקציה `lpc` של Matlab. 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 בווקטור יהיה תמיד שווה ל-1.
- הפעילו את הפונקציה כך שתחזיר את מקדמי החזאי מסדר 10 של מסגרת הדוגמה הקולית.
- הנחיה:** הפונקציה `lpc` מחזירה במקרה זה מספרים מרוכבים שהחלק המדומה שלהם הוא אפס. מכיוון שהמסנן שלנו הוא מסנן ממשי, יש לקחת רק את החלק הממשי של התוצאה בעזרת הפונקציה `real`.
- פונק' `lpc` יכולה גם לחשב את שונות שגיאת החיזוי (בדומה לפונק' שכתבנו בתרגיל ההכנה).** יש להשוות עבור מקטע בודד את הערך המתקבל מהפונק' `lpc` מול הערך המתקבל מהפונק' `residual_energy`.

- 2.2 הצגת ספקטרום האות לעומת ספקטרום החיזוי
- את האמפליטודה של המעטפת הספקטרלית של מסגרת הדוגמה הקולית ניתן לקבל ע"י שימוש בפונקציה `freqz` באופן הבא:

```
[H, W] = freqz([1], LPC_vec, 128);
```

כאשר `LPC_vec` הוא וקטור מקדמי ה-LPC שהתקבל בסעיף הקודם, המשמש כאן כמסנן IIR, ו-`H` היא תגובת התדר שלו.

- הציגו ביחידות של dB גרף המכיל את ספקטרום מסגרת הדוגמה הקולית (ע"י ביצוע התמרת DFT והצגת החלק החיובי בלבד של ציר התדר) יחד עם צורת המעטפת הספקטרלית המתקבלת ממקדמי ה-LPC.
- הסבירו את התוצאה הנראית בגרף.

- 2.3. בדיקה עבור סדרים שונים של החזאי
- חיזרו על שני הסעיפים האחרונים, כאשר סדר מסנן החזאי הוא 50 וכאשר מסנן החזאי הוא 5. כיצד השפיעה הגדלת/הקטנת סדר החזאי על צורת המעטפת?
  - מהו ה-trade off שמציגים שלושת הגרפים? באיזה סדר מסנן חזאי נעדיף להשתמש לדעתכם במערכת המקודד-מפענח שנבנה?

### 3. מימוש המקודד והמפענח

כעת נבנה מערכת מקודד-מפענח מבוססת LPC, 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.

- 3.1. 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. בהמשך, על הפונקציה ליצור גרסה מסוננת של האות המקורי ולחלק את שני האותות (המקורי והמסונן) למסגרות של 256 דגימות כל אחת. לאחר מכן, הפעולות הבאות תתבצענה על כל מסגרת:
- א. קביעת אופי המסגרת - קולית או א-קולית ע"י שימוש בפונקציה `vu_classify`: הפונקציה תקבל את המסגרת המקורית, וגרסה מסוננת וקטומה של המסגרת.
  - ב. קביעת ערך מחזור ה-pitch - אם נקבע כי המסגרת קולית, מחזור ה-pitch ישוערך ע"י שימוש בפונקציות `clip` ו-`pitch_detect_corr`. אם נקבע כי המסגרת אינה קולית, ערך מחזור ה-pitch יהיה אפס.
  - ג. שיערוך מקדמי ה-LPC מסדר 10 המתאימים למסגרת - ע"י שימוש בפונקציה `lpc`.
  - ד. שיערוך האנרגיה הממוצעת של אות העירור - ע"י שימוש בפונקציה `residual_energy`: הפונקציה תקבל את המסגרת ואת וקטור מקדמי ה-LPC שחושבו בסעיף הקודם.

המקודד יוציא ארבעה פלטים:

- א.  $N$  - מספר המסגרות באות
- ב.  $p$  - וקטור באורך  $N$  המכיל את ערכי ה-pitch בכל מסגרת
- ג.  $lp$  - מטריצה בגודל  $11 \times N$  המכילה את וקטורי ה-LPC של כל מסגרת
- ד.  $e$  - וקטור באורך  $N$  המכיל את שורש ערכי האנרגיה הממוצעת של אות העירור בכל מסגרת (שורש ריבועי של המוצא של הפונקציה `residual_energy`)

- כתבו פונקציה בשם `encoder` המממשת את המקודד הנ"ל.
- למה אחראי כל אחד מארבעת הפרמטרים הנ"ל במודל יצירת אות הדיבור?
- בהנחה שכל אחד מהערכים של  $p$ ,  $lp$  ו- $e$  מיוצג ע"י 16 סיביות (שני בתים), כמה מקום נדרש לשמירת כל הפרמטרים? השוו לכוונת שגורש לשמירת האות בצורתו המקורית (חושב בפגישת הניסוי הראשונה). מהו יחס הדחיסה?

### 3.2. 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:
- יצירת אות העירור:
    - אם המסגרת היא א-קולית (ערך `pitch` שווה לאפס), אות העירור יהיה רעש גאוסני 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 תנאי ההתחלה כדי שההלמים בשתי מסגרות קוליות עוקבות ישמרו על המחזוריות של אות העירור (בהתאם לשאלת הכנה מס' 6).
- מתן אנרגיה (gain) מתאימה למסגרת העירור :  
ננרמל את אות העירור כך שהאנרגיה שלו תהיה זהה לאנרגיה של אות העירור המקורי. נעשה זאת באופן הבא :
    - ננרמל את אות העירור המסונטז לערך יחידה (האנרגיה לזמן קצר של כל מקטע תהיה שווה 1)
    - נכפול את אות העירור המסונטז 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 המסונטז.
  - סינון אות העירור ע"י שימוש במקדמי החזאי הלינארי ששוערכו למסגרת על מנת ליצור את אות הדיבור. זאת נעשה תוך שימוש בפונקציה filter עם תנאי התחלה (כפי שראיתם בשאלת הכנה מס' 5), כאשר המסנן משמש כמסנן IIR :  

$$[new\_frame, final\_cond] = filter([1], lp, signal\_to\_be\_filtered, final\_cond);$$
 כאשר הווקטור final\_cond יאותחל באפסים.
- על המפענח לשרשר את המסגרות המשוחררות אחת לשנייה לקבלת אות שאורכו זהה לאורך האות המקורי.
- כתבו פונקציה בשם decoder המממשת את המפענח הנ"ל.

#### 4. בדיקת איכות מערכת המקודד-מפענח

- 4.1. בדיקה עם האות המקורי  
הריצו את מערכת המקודד-מפענח על האות שלכם.
- השמיעו את האות המקורי ואת האות המשוחרר בעזרת הפונקציה soundsc. האם האות המשוחרר נשמע דומה לאות המקורי? האם ניתן לחוש בהבדלים?
  - הציגו את האות המקורי ואת האות המשוחרר אחד מעל השני ע"י שימוש בפונקציה subplot. מה ההבדלים העיקריים ביניהם?
- 4.2. בדיקה עם הקלטה עצמית  
בסעיף זה תקליטו את עצמכם ותפעילו את מערכת המקודד-מפענח על האות שהקלטתם. לשם כך נשתמש ברשמקול של מערכת ההפעלה (נמצא תחת accessories).
- קראו את ההקלטה לסביבת מטלב בעזרת audioread.
- הקשיבו להקלטה שביצעתם בעזרת הפונקציה soundsc וודאו שהיא תקינה ולא רועשת מאוד. אם ההקלטה אינה ברמה מספקת, בצעו אותה שוב.
- שימו לב שאם קצב הדגימה שונה מ-8 קילו הרץ, יש לשנות אותו בעזרת פונקציה resample, לדוגמא :  
 $[X, FS] = audioread(FILENAME);$   
 $Y = resample(X, 8e3, FS);$
- בצעו transpose להקלטה שיצרתם (c0), הפעילו עליה את מערכת המקודד-מפענח וחזרו על סעיף 4.1 עבודה.

- 4.3. הפעלת מדדים אובייקטיביים לאיכות מערכת המקודד-מפענח
- בסעיף זה נבדוק את טיב מדד ה-Segmental SNR ואת טיב מדד ה-PESQ כמדדים לאיכות של אות דיבור ע"י השוואה בין המדדים שהתקבלו עבור שני אותות שונים :
- שימרו את האות המשוחזר מסעיף 4.1 בעזרת הפונקציה audiowrite
  - הכינו אות נוסף לבדיקה : הרעישו את האות המקורי ע"י הוספת רעש לבן בעל שונות  $4 \cdot 10^{-6}$  ולאחר מכן איפוס כל 1024 דגימות עוקבות לסירוגין (1024 יושארו כפי שהיו באות המקורי ו-1024 יאופסו, וחוזר חלילה). שימרו גם אות זה בעזרת הפונקציה audiowrite.
  - הפעילו את תוכנת ה-SPDemo והשוו את האות המקורי לשני האותות הנ"ל בעזרת מדד Segmental SNR על מקטעים באורך 30msec ללא חפיפה ובעזרת מדד PESQ. היעזרו במודול בשם Q-Eval. ההשוואה בין 2 אותות נעשית ע"י סימון האות המקורי עם העכבר, והגדרת האות המשווה אליו כ-Test signal. שימו לב ששני האותות צריכים להיות בעלי אותו האורך !!
  - הקשיבו לאות המקורי, לאות המשוחזר ולאות הנוסף שהכנתם. מי מביניהם מובן יותר, לדעתכם? למי מביניהם נתן מדד Segmental SNR ציון גבוה יותר? לפי מביניהם נתן מדד PESQ ציון גבוה יותר? מהי המסקנה?

## מקורות לעיון נוסף

1. דוד מלאך, שלום רז, מבוא לעיבוד ספרתי של אותות – שקפים להרצאות, תשנ"ט, הטכניון
2. דוד מלאך, קידוד אותות – שקפים להרצאות. ניתן להוריד את השקפים מאתר הקורס : [/http://www.ee.technion.ac.il/courses/048955](http://www.ee.technion.ac.il/courses/048955), פרק 6, עמ' 1-10 מציגים את העקרונות המועברים בניסוי.
3. דוד מלאך, עיבוד ספרתי של אותות – שקפים להרצאות. ניתן להוריד את השקפים מאתר הקורס : [/http://www.ee.technion.ac.il/courses/046745](http://www.ee.technion.ac.il/courses/046745)
4. Rabiner Lawrence R., Schafer Ronald W., Digital Processing of Speech Signals, Prentice Hall, 1978.
5. Goldberg Randy, Riek Lance, A Practical Handbook of Speech Coders, CRC Press, 2000.
6. Barnwell Thomas P., Nayebi Kambiz, Richardson Craig H., Speech Coding – A Computer Laboratory Textbook, Wiley, 1996.
7. A. Michael Noll, Cepstrum Pitch Determination, Journal of the Acoustical Society of America, Vol. 41, No. 2, pp. 293-309, 1967.