תרגיל בית 3 – למידה עמוקה באותות דיבור

# שאלה 11:

## סעיף א

A screenshot of a screen

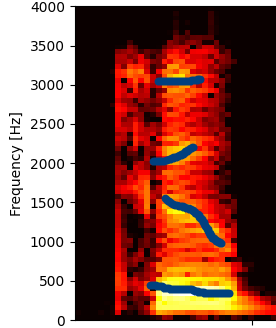
Description automatically generated

הממדים של התמונה (כפי שמודפס בפונקציה) הם

## סעיף ב

1. בדרך כלל בין מילים יש שקט, משמע אין אות של קול בזמן הזה. ניתן לראות זאת בספקטוגרמה בזמנים שבהם יש חושך יחסי בכל או רוב של התדירויות.
2. באיזורים "קוליים" יש דיבור ולכן נראה פורמנטים יחסית ברורים. כלומר, העוצמה של התדירויות לא תהיה יוניפורמית אלא תהיה בבירור חזקה יותר באיזור תדרים מסוימים.
3. באיזורים "א-קוליים" אין דיבור ולכן העוצמה של התדירויות תהיה יחסית יוניפורמית – לא נראה פורמנטים ברורים אלא יותר "מריחה".

## סעיף ג:

אנו יודעים כי הpitch הוא הפורמנט הראשון (כלומר עם התדר הכי נמוך) .   


בתמונה ניתן לראות את הפורמנטים השונים (מסומנים ידנית, הpitch הוא הקו התחתון). מכאן אנו יודעים כי הפורמנט הראשון (ולכן גם הpitch) הוא בערך בתדירות של 330 הרץ. הסתכלנו עם העכבר על הגרף וראינו כי הגובה של הפורמנט הראשון הוא 330.

תדר הpitch לא קבוע בכל האזורים הקוליים בתמונה. זה הגיוני כי מבטאים הבהרות שונות ולכן הסיגנלים שונים בין החלקים השונים.

## סעיף ד:

ניתן לשים לב לכמה הבדלים עיקריים. תחילה נבין מה השינוי אמור בכלל, דגימות קצרות יותר אומרות שנפעיל FFT על מקטעים קצרים יותר מהשמע, שזה גורר FFT פחות מדוייק באותות, אבל בתמורה אנחנו מקבלים רזולוציה טובה יותר בציר הזמן, אנו נתפוס שינויים שהיו בקול מהר יותר, לעומת תדירות הדגימה הארוכה יותר, שתפספס שינוים אם קרו כמה שינויים באותו חלון FFT. כלומר יש פה tradeoff בין הרזולוציה של הדגימה בציר הזמן, לעומת הדיוק שלנו במימד בתדר. ניתן לראות זאת באמצעות הסתכלות על מימדי הספקטוגרמה – עם 128 הן ועם 512 הן .

## סעיף ה:

ניתן לראות שכאשר ה, אנו מקבלים ספקטוגרמה יותר coarse ורואים את הפורמנטים בצורה הרבה פחות מדויקת מאשר עם . במקרה ש, אנו רואים את ההפך, הספקטוגרמה הרבה יותר מדויקת.

פרמטר הovp מהווה אחוז חפיפה בין הדגימות כדי לחשב את הFFT. כאשר הovp הוא גדול, בין כל דגימה לדגימה הבאה לא יהיה הבדל משמעותי בFFT מאחר ורק הוספנו והורדנו חלק קטן מהדגימה. כך, כאשר הovp גדל, נקבל ספקטוגרמה חלקה יותר ועם כמות גדולה יותר משמעותית של דגימות – אך עדיין אם אותה כמות תדרים. בנוסף, כאשר הovp מספיק גדול, אנחנו מרוויחים יציבות בעבור תדרים "שנפלו" בין דגימות במקרה בלי הoverlap.

# שאלה 12

1. מומש בקוד
2. הפלט השלישי הוא זה שאחראי על הresidual energy. ראינו בהרצאה כי העוצמה של אות הערעור היא זו שגורמת לכך שסיגנל הדיבור לא מתאים בדיוק למידול של הLPC.

# שאלה 13

1. הטעות של הסטודנט היא שההלם הראשון בכל מקטע התחיל בתחילת המקטע וכך יצא שאחרי כל 3 מרווחים, הרווח הבא היה קטן יותר מP.
2. נסמן את המיקום של ההלם האחרון במקטע בתור . כמות הדגימות שנותרה עד סוף המקטע היא . אנחנו נרצה שיהיה מרווח של P דגימות, נסמן באת מיקום ההלם הראשון במקטע הבא, אז אנחנו נרצה שיתקיים ולכן .
3. נכתב בקוד
4. A graph with blue dots and a red line

   Description automatically generated

# שאלה 14

1. נעשה בקוד
2. נעשה בקוד
3. A diagram of a sound wave

   Description automatically generated  
   התוצאה המתקבלת אכן תואמת את הציפיות שלנו. יש קשר חזק בין המקומות שבהם אנו מזהים בעין דיבור לבין מקומות שהמודל חזה דיבור. יש כמה טעויות, בעיקר בסוף ובהתחלת קטעי דיבור אך מאחר והפונקציה היא פונקציה יחסית פשוטה ויוריסטית, זה די הגיוני. לעומת זאת, לפעמים אנו רואים שהpitch הוא לא בין 50 ל400 הרץ מה שלא תואם את הציפיות שלנו אך ייתכן וזו שגיאה של האלגוריתם.

# שאלה 15