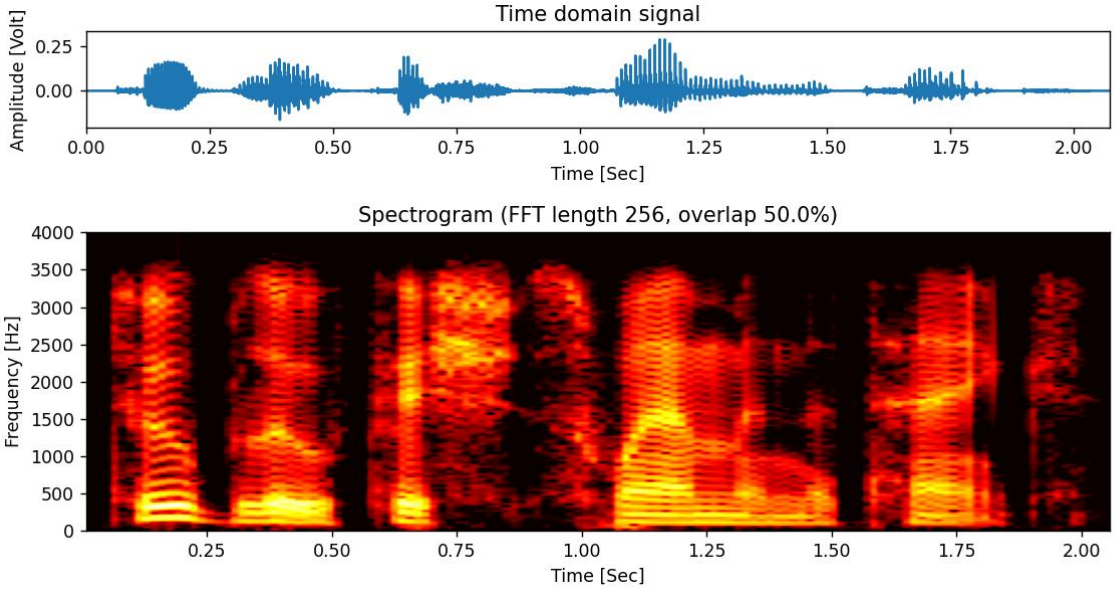
עידן פוגרבינסקי 325069565
אורי מירז 212641229
ספרה מזהה 5

תרגיל בית 3 – למידה עמוקה באותות דיבור

ספרה מזהה 5

שאלה 11:

סעיף א



הממדים של התמונה (כפי שמודפס בפונקציה) הם 128×129.

בחתך אנכי, ניתן לראות את הFFT בפריים קטן בזמן – כלומר את האמפליטודות של הסינוסים המרכיבים את הגל. מזה גם אפשר לראות האם יש צליל בזמן הזה (אם הפריים די שחור) ומה אופיו (דיבור או לא דיבור – לפי ההתפלגות של האמפליטודות).

בחתך אופקי אפשר לראות את השינוי בזמן בהופעתו של תדר מסויים לאורך ההקלטה, משתנה ברזולוציה של פריימים.

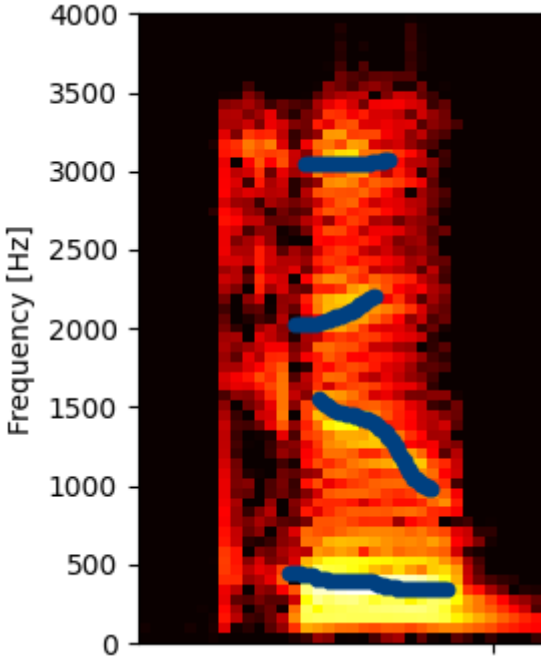
סעיף ב

א. 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.
ב. באיזורים "קוליים" יש דיבור ולכן נראה פורמנטים יחסית ברורים. כלומר, העוצמה של התדירויות לא תהיה יוניפורמית אלא תהיה בבירור חזקה יותר באיזור תדרים מסוימים.
ג. באיזורים "א-קוליים" אין דיבור ולכן העוצמה של התדירויות תהיה יחסית יוניפורמית – לא נראה פורמנטים ברורים אלא יותר "מריחה".

עידן פוגרבינסקי 325069565
אורי מירז 212641229
ספרה מזהה 5

סעיף ג:

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

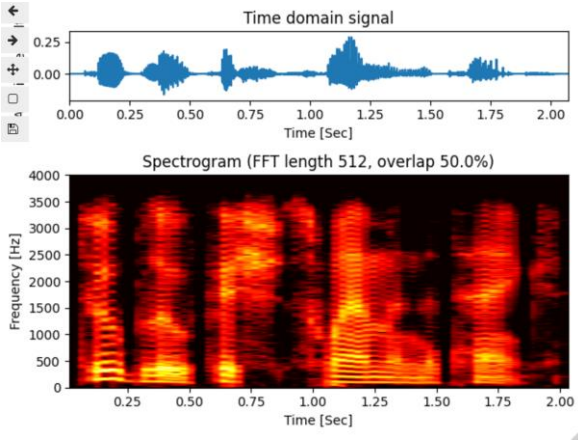


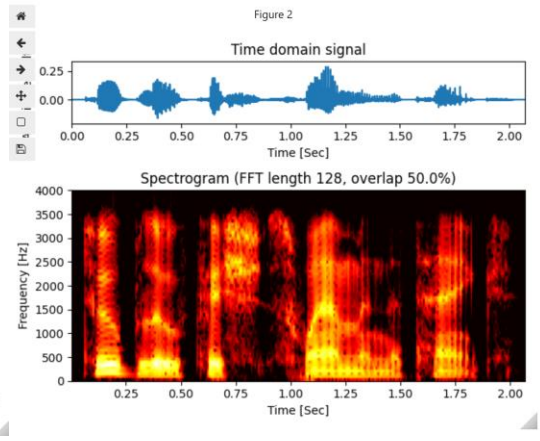
בתמונה ניתן לראות את הפורמנטים השונים (מסומנים ידנית, ה pitch הוא הקו התחתון). מכאן אנו יודעים כי הפורמנט הראשון (ולכן גם הpitch) הוא בערך בתדירות של 330 הרץ. הסתכלנו עם העכבר על הגרף וראינו כי הגובה של הפורמנט הראשון הוא 330.

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. הדבר שגורם לכך הוא ה tradeoff בין הרזולוציה של הדגימה בציר הזמן, לעומת הדיוק שלנו במימד בתדר. ניתן לראות זאת באמצעות הסתכלות על מימדי הספקטוגרמה – עם 128 הן 258×65 ועם 512 הן 63×257.



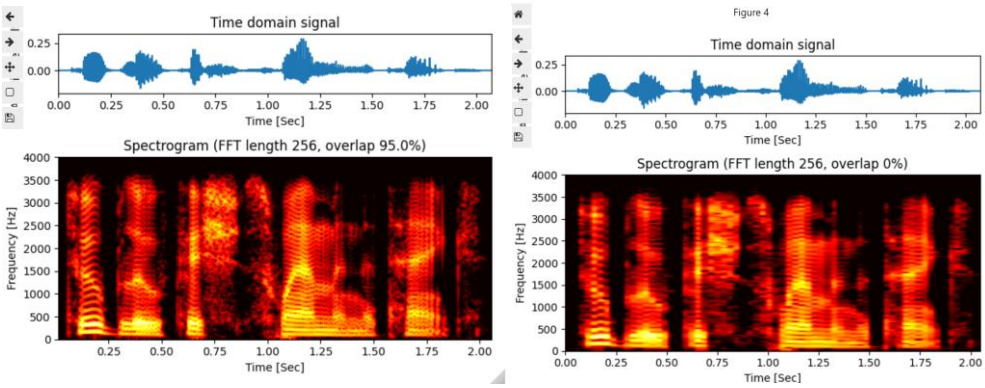


עידן פוגרבינסקי 325069565
אורי מירז 212641229
ספרה מזהה 5

סעיף ה:

ניתן לראות שכאשר ה$ovp = 0$, אנו מקבלים ספקטוגרמה יותר coarse ורואים את הפורמנטים בצורה הרבה פחות מדויקת מאשר עם $ovp = 0.5$. במקרה ש$ovp = 0.95$, 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. (ניתן לראות בתמונות)



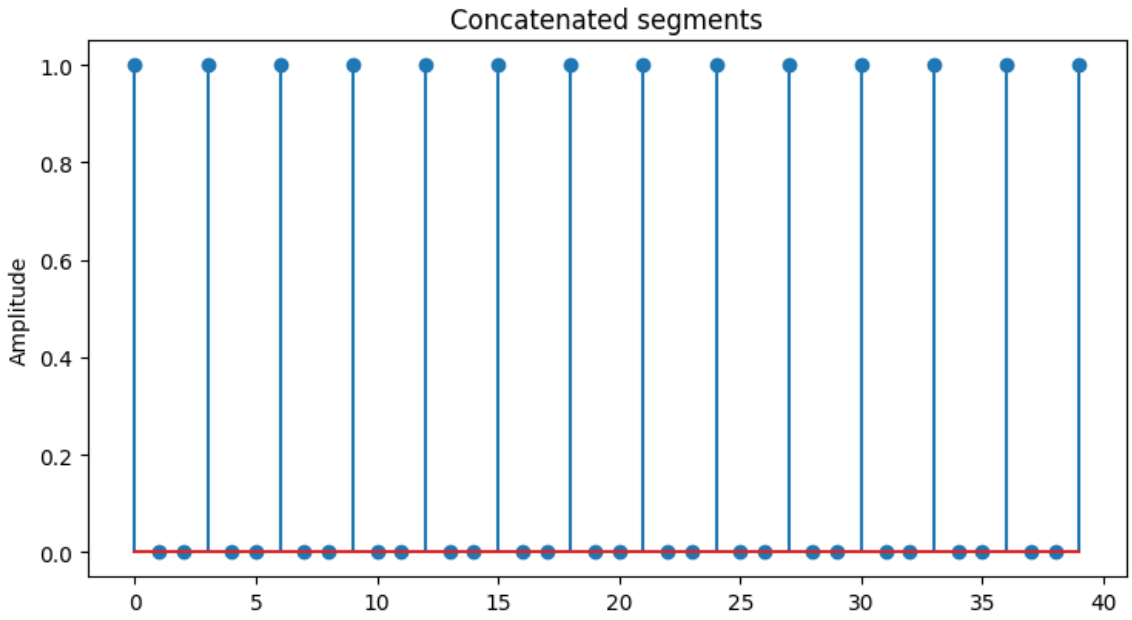
שאלה 12

א. מומש בקוד
ב. 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

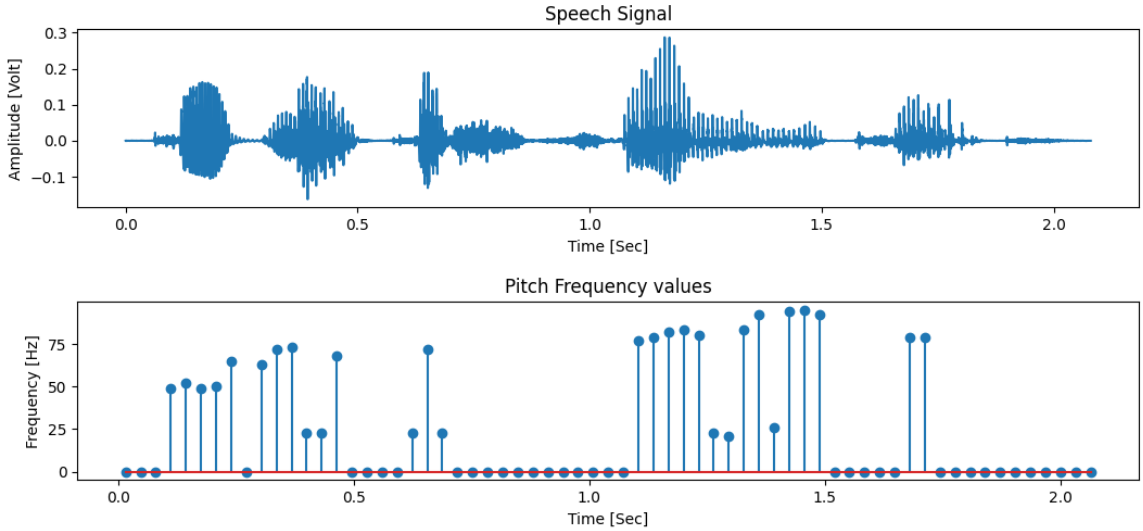
א. הטעות של הסטודנט היא שההלם הראשון בכל מקטע התחיל בתחילת המקטע וכך יצא שאחרי כל 3 מרווחים, הרווח הבא היה קטן יותר מP.
ב. נסמן את המיקום של ההלם האחרון במקטע בתור m. כמות הדגימות שנותרה עד סוף המקטע היא $mod_P(N-m)$. אנחנו נרצה שיהיה מרווח של P דגימות, נסמן ב x את מיקום ההלם הראשון במקטע הבא, אז אנחנו נרצה שיתקיים $N-n+x=P$ ולכן $x=P-N+n$.
ג. נכתב בקוד

עידן פוגרבינסקי 325069565
אורי מירז 212641229
ספרה מזהה 5
ד.



שאלה 14

א. נעשה בקוד
ב. נעשה בקוד
ג.



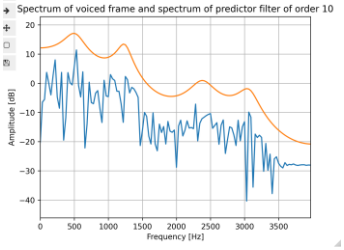
התוצאה המתקבלת אכן תואמת את הציפיות שלנו. יש קשר חזק בין המקומות שבהם אנו מזהים בעין דיבור לבין מקומות שהמודל חזה דיבור. יש כמה טעויות, בעיקר בסוף ובהתחלת קטעי דיבור אך מאחר והפונקציה היא פונקציה יחסית פשוטה ויוריסטית, זה די הגיוני. לעומת זאת, לפעמים אנו רואים שהpitch הוא לא בין 50 ל400 הרץ מה שלא תואם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

עידן פוגרבינסקי 325069565
אורי מירז 212641229
ספרה מזהה 5

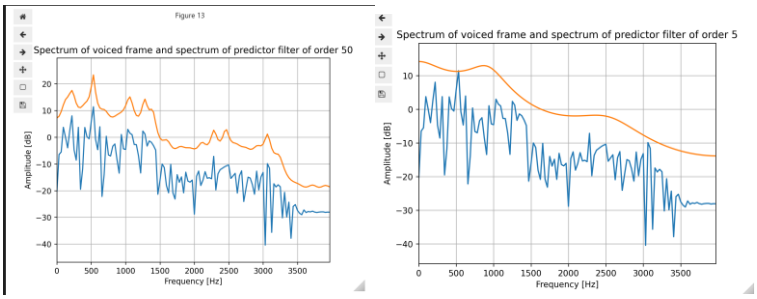
שאלה 15

1.

א. מומש בקוד
ב. מומש בקוד
ג.



הגרף מראה את המעטפת הספקטרלית, ניתן לראות כי המעטפת של החזאי לא מושלמת, ניתן להניח שזה נובע מחוסר דיוק ושגיאה שנוצרת.



2. ניתן לראות בבירור כי ככל שיש יותר מקדמים איכות השערוך משתפרת(ניתן גם לראות מספרית בקוד כי שגיאת החיזוי קטנה), אך זה דבר שאינו מפתיע אותנו. ככל שיש יותר מקדמים, כך תקטן שגיאת השערות ונתקרב יותר לקול המקורי, אך נקבל deminishing returns החל משלב מסויים ולכן נרצה למצוא את האיזון.

3. א. ניתן לראות כי הגרף שמשערך הכי טוב את התמרת התדר הוא הגרף המשוייך לסדר 50, כפי שיש לו יותר מקדמים והוא יכול לעשות fit טוב יותר לנתונים שלנו, אך אם נסתכל על מה החזאי שחוזה את המעטפת הספקטרלית בצורה האופטימלית, נגיד שזה החזאי בעל עשר המקדמים מאחר שהוא פחות נוטה לoverfit על הדאטא ומוצא את המעטפת הכללית בצורה נקייה.
ב. לפי הטיעונים בסעיף הקודם, נבחר בחזאי בעל עשרה פרמטרים, נעדיף זאת מאחר שאנחנו צריכים את הצורה הכללית של המעטפת ולא את הפרטים הקטנים (רעש) שיפריעו לחלקות של המודל שלנו ופחות מתארים את המעטפת.

שאלה 18

סעיף 1

a. נעשה בקוד
b.

N– קובע כמה פריימים יש, מה ה"רזולוציה" בעבורה אנחנו שולחים מקטעים.

עידן פוגרבינסקי 325069565
אורי מירז 212641229
ספרה מזהה 5

p -קובע את הpitch שיהיה בכל פריים. זה התדר של רכבת ההלמים בקטע שאנו מסווגים אותו בתור "קולי:

lp – המקדמים של המשוואות שלנו, שמייצג את הצורה של המסגרות שלנו.

e – הערעור (gain)

c. יש 16578 דגימות וכל דגימה היא 2 בתים ולכן, מיוצגת על ידי $2 \cdot 16578 = 33456$ בתים.
d. נחשב כמה ערכים מספריים יש לנו:
N הוא ערך מספרי אחד, בp יש N ערכים מספריים (אחד לכל מסגרת), בlp יש 11N ערכים ובe יש N ערכים. סה"כ הנפח הוא $2 \cdot (1 + N + 11N + N) = 2 \cdot (13N + 1) = 2 \cdot 846 = 1692$ בתים.
e. $\frac{33156}{1692} \approx 19.6$

סעיף 2

a. נעשה בקוד

עידן פוגרבינסקי 325069565
אורי מירז 212641229
ספרה מזהה 5
.b

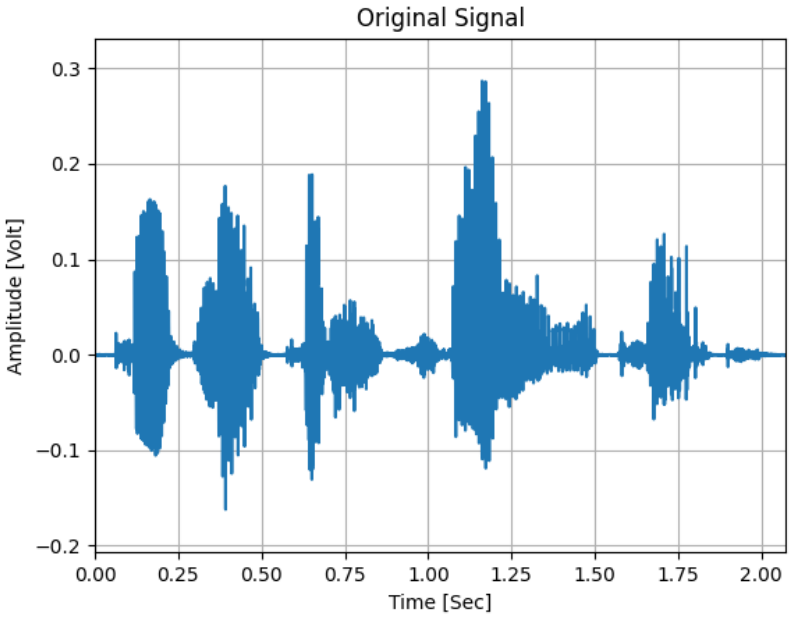
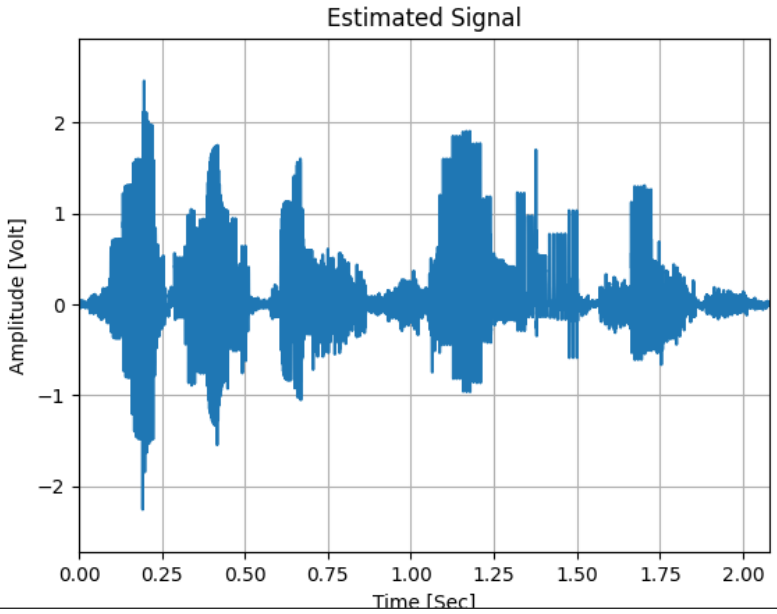


Figure 6



האותות נראים דומים מבחינת ההפרדות בין המילים וההבהרות וגם בין אזורים של קול לבין אזורים של אות א-קולי. ההבדל העיקרי הוא שבאות המשוחזר, האמפליטודות יותר יוניפורמיות ועם פחות "פיקים" קיצוניים.

c. האות נשמע די דומה ואפשר בקלות להבין את הנאמר אך האות המשוחזר נשמע יותר רובוטי ועם רעש סטטי.
d. נשמר.