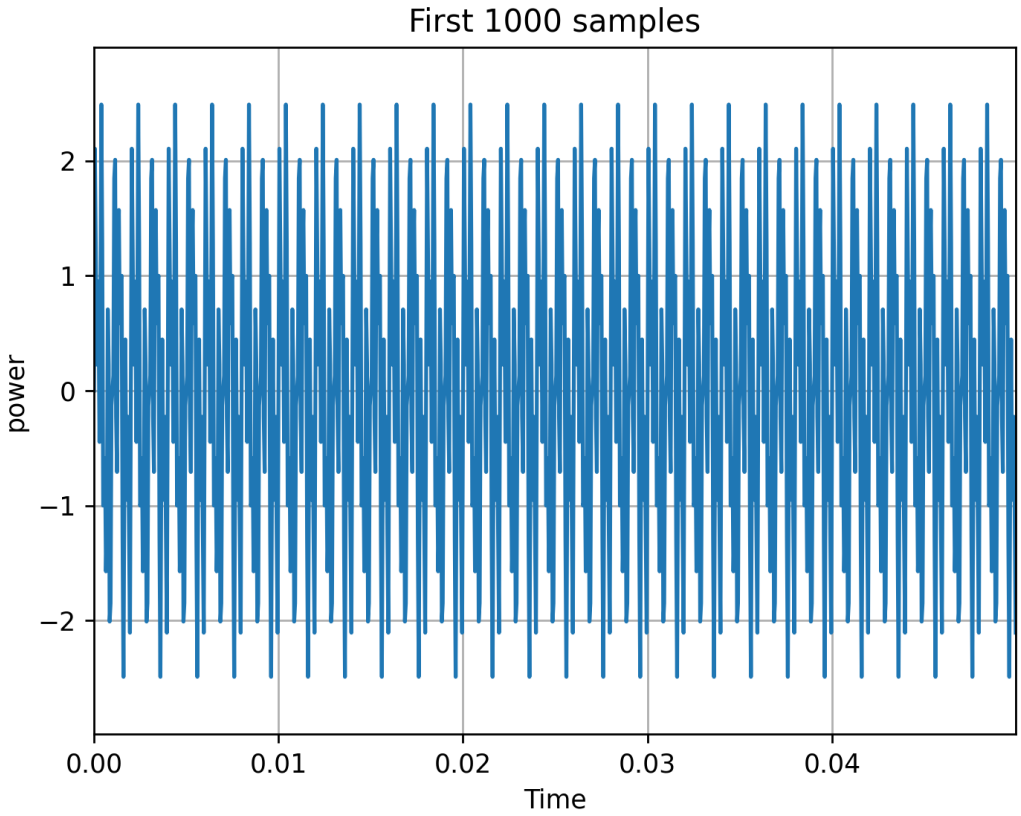
אורי מירז 212641229
עידן פוגרבינסקי 325069565
ספרה מזהה 5

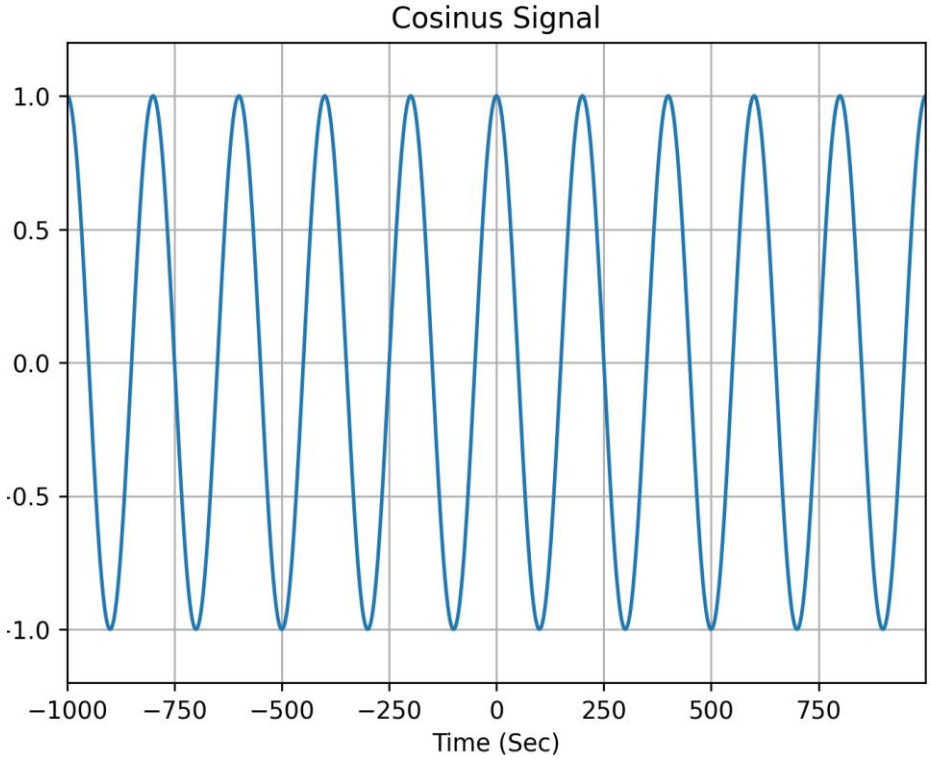
תרגיל בית 2 – למידה עמוקה באותות דיבור

ספרה מזהה 5

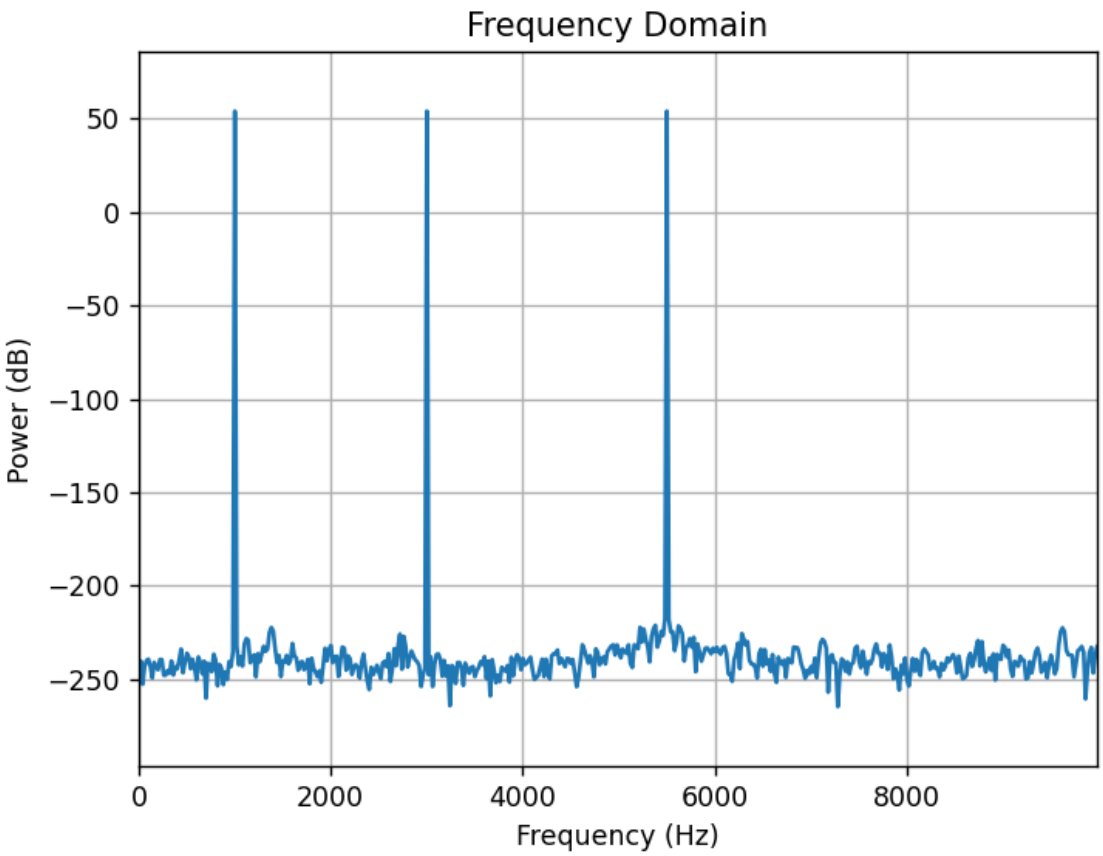
שאלה 1:

א.





אורי מירז 212641229
עידן פוגרבינסקי 325069565
ספרה מזהה 5
ב.

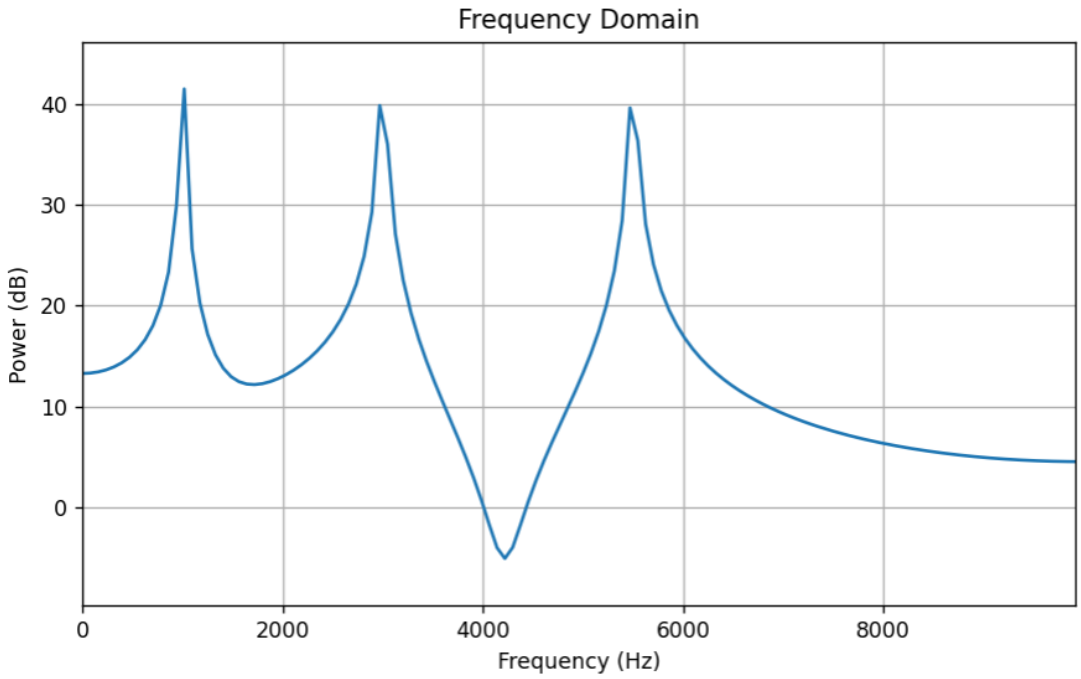


מחשבים log10 על העוצמה של כל תדר, כלומר על האמפליטודה של גל סינוס עם התדר הזה אשר מרכיב את האות. ה $signal.size = 1000$ ולכן גם גודל הFFT. תדירות הדגימה היא 20,000 אלף הרץ ולכן רזולוציית התדר היא $\frac{20,000}{1,000} = 20[Hz]$ לפי הגדרה.

תמונת התדר שהתקבלה מייצגת את האמפליטודה של הסינוס שמרכיב את התדר, לכל אחד מהתדרים המוצגים. ניתן לראות כי בערכים שהיינו רוצים לקבל – התדרים של הסינוסים שבאמת מרכיבים את הסיגנל, יש קפיצה משמעותית.

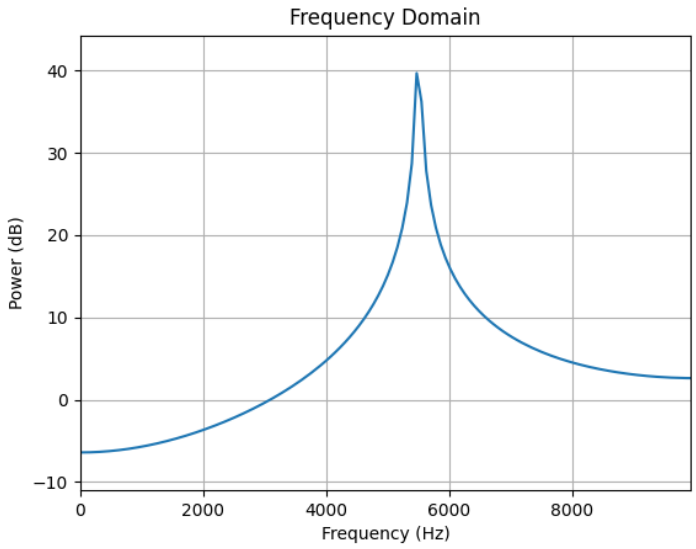
אורי מירז 212641229
עידן פוגרבינסקי 325069565
ספרה מזהה 5

ג.



השיאים מתקבלים ב1000, 3000 ו5500 הרץ. כעת התדריים הם בין 0 ל10,000 אך עכשיו יש 256 דגימות (אזי גודל הFFT הוא 256) ולכן, רזולוציית התדר היא $\frac{20,000}{256} \approx 78.125[Hz]$ תמונת התדר שמתקבלת היא כמה "פיקים" ב1000, 3000 ו5500 הרץ – שאלו אכן התדרים שהיינו אמורים לקבל – אלה התדרים של הסינוסים השונים אותם סכמנו. ניתן לראות גם כי הפיקים גם יותר מחודדים ויותר קרובים לדגימות האחרות.

ד.



פה ניתן לראות שהשיא ב19K לא קיים (למרות שיש סינוס בתדר הזה) והשיא ב1000 נעלם - כלומר רק מה שב5.5 נשאר.

האות נדגם ב20 אלף הרץ ולכן, כל דגימה ניתן לתאר על ידי $n \in \mathbb{N}$ $t = \frac{n}{20000}$. אם נציב ב

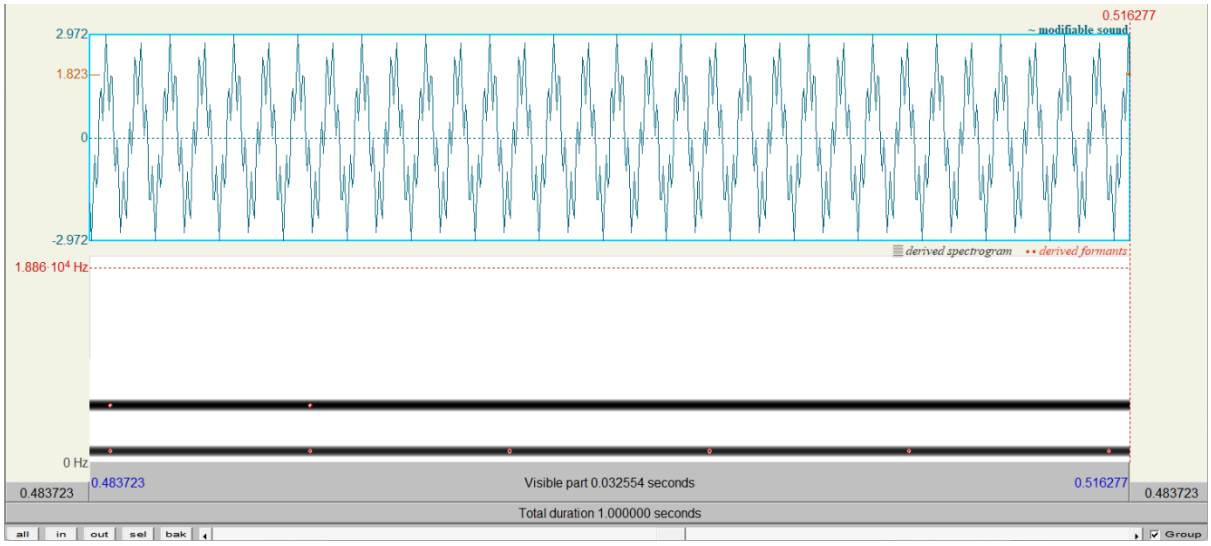
אורי מירז 212641229
עידן פוגרבינסקי 325069565
ספרה מזהה 5

$$\sin(2\pi\cdot 1000\,t)+\sin(2\pi\cdot 19000\cdot t)=\sin\left(2\pi\cdot\frac{1000}{20000}\,n\right)+\sin\left(2\pi\cdot\frac{19000}{20000}\cdot n\right)$$
$$=\sin\left(2\pi\cdot\frac{n}{20}\right)+\sin\left(2\pi\cdot\frac{19}{20}\cdot n\right)=\sin\left(2\pi\cdot\frac{n}{20}\right)+\sin\left(2\pi-2\pi\cdot\frac{n}{20}\right)$$
$$=\sin\left(2\pi\cdot\frac{n}{20}\right)-\sin\left(2\pi\cdot\frac{n}{20}\right)=0$$

כלומר, האותות מבטלים זה את זה.

באופן סכמתי, מאחר ואנו דוגמים ב20 אלף הרץ ויש תדר בתדירות 19 אלף הרץ שגדולה ב- $\frac{1}{2}$ 20,000 – קורית תופעה של aliasing. כלומר, הסיגנל בתדירות 19 אלף מבטל את הסיגנל בתדר של 1000 הרץ.

שאלה 2:



שאלה 3

א. לפי חומר העזר שניתן לתרגיל, תחום התדרים של pitch של דיבור אנושי נע בתחום $50Hz-400Hz$

ב. מאחר ותדירות היא הופכית לזמן המחזור, משך הזמן האופייני של מחזור pitch הוא בין $\frac{1}{400}[s]-\frac{1}{50}[s]$.

ג. האות נדגם בתדירות של fs ויש p דגימות בכל מחזור של pitch ולכן, הזמן שייקח לכל מחזור הוא $\frac{p}{fs}$ ומכאן התדירות היא $\frac{f_s}{p}$.

ד. כפי שהסברנו לעיל, זמן המחזור הוא $100\cdot\frac{1}{8000}=\frac{1}{80}[s]$ ולכן התדר המתקבל הוא $80[Hz]$ ואכן $50\le 80\le 400$ ולכן הוא בתחום.

שאלה 4

סעיף א

את רכבת ההלמים ניתן לבטא בתור $f(t)=\sum_{n=-\infty}^{\infty}\delta(t-nT)$ כאשר T הוא זמן המחזור של רכבת ההלמים. כעת, נבטא את רכבת ההלמים בתור טור פורייה. ניתן לעשות את זה מאחר והיא מחזורית.

$$f(t)=\sum_{n=-\infty}^{\infty}c_n e^{j\cdot 2\pi\cdot n\cdot\frac{1}{T}}$$

אורי מירז 212641229
עידן פוגרבינסקי 325069565
ספרה מזהה 5

$$c_n = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} f(t)e^{-j\cdot 2\pi\cdot n\cdot\frac{t}{T}}dt = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} \delta(t)e^{-j\cdot 2\pi\cdot n\cdot\frac{t}{T}}dt = \frac{1}{T}$$

מכאן נקבל כי

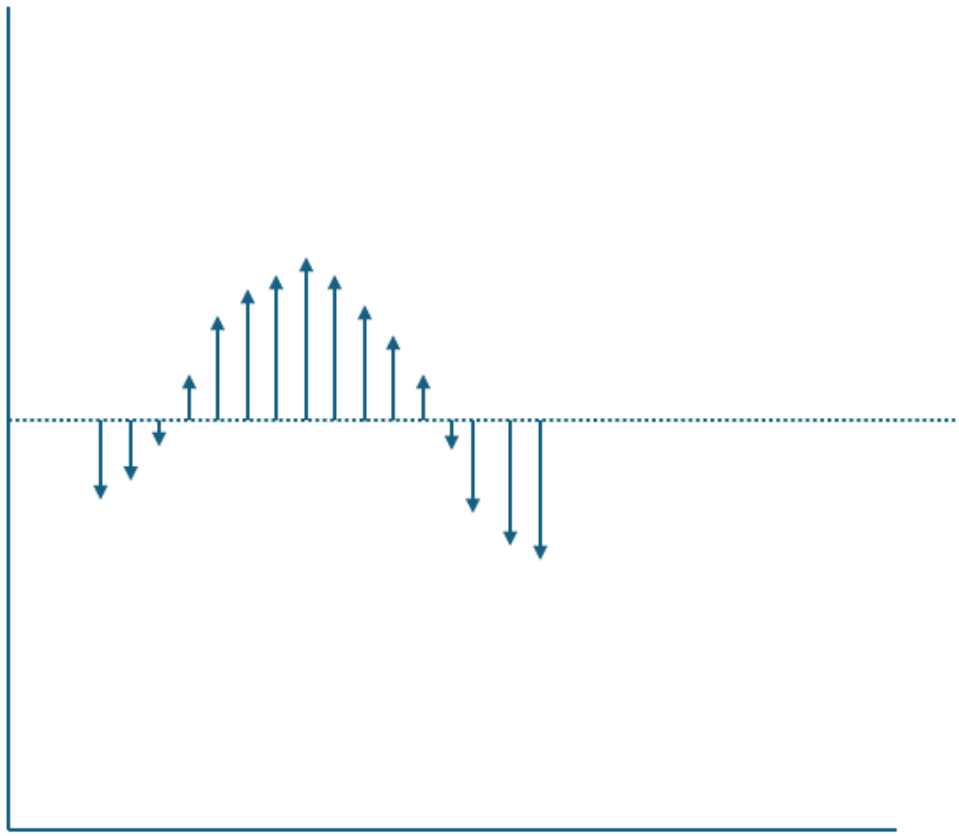
$$\mathcal{F}[f] = \mathcal{F}\left[\frac{1}{T}\sum_{n=-\infty}^{\infty} e^{j\cdot 2\pi\cdot n\cdot\frac{1}{T}}\right] = \frac{1}{T}\mathcal{F}\left[\sum_{n=-\infty}^{\infty} e^{j\cdot 2\pi\cdot n\cdot\frac{1}{T}}\right] = \frac{1}{T}\mathcal{F}\sum_{n=-\infty}^{\infty}\delta\left(\omega - n\cdot\frac{1}{T}\right)$$

קיבלנו כי ההתמרת פורייה של רכבת ההלמים היא רכבת הלמים בתדר עם "זמן מחזור" של $\frac{1}{T}$. המשמעות היא שהפלט של העברת סיגנל במסנן אחרי שעבר קונבולוציה עם רכבת הלמים (לדוגמא בדיבור שלנו) שקול למכפלה של הפלט של הסיגנל ברכבת הלמים בתדר (מאחר וקונבולוציה בזמן שקולה לכפל בתדר) – כלומר דגימה כאשר יש "הלם".

סעיף ב

אנו יודעים שקונבולוציה בזמן שקול למכפלה בתדר ולכן, אנחנו רק נכפיל את התגובה ברכבת הלמים נוספת (אנו יודעים מסעיף א שבאמת נצטרך להכפיל ברכבת הלמים)

לכן, הגרף יהיה בערך:



כלומר דגימות של הגרף המקורי במחזורים שונים.

אורי מירז 212641229
עידן פוגרבינסקי 325069565
ספרה מזהה 5

שאלה 5

א:

$$r(-k) = \sum_{n=0}^{N+k-1} x_n x_{n-k} = \sum_{n=0}^{k-1} x_n \underbrace{x_{n-k}}_{0} + \sum_{n=k}^{N-1} x_n x_{n-k} + \sum_{n=N}^{N+k-1} \underbrace{x_n}_{0} x_{n-k}$$
$$= \sum_{n=k}^{N-1} x_n x_{n-k} \underbrace{=}_{n'=n-k} \sum_{n'=0}^{N-k-1} x_{n'+k} \cdot x_{n'} = r(k)$$

ב:

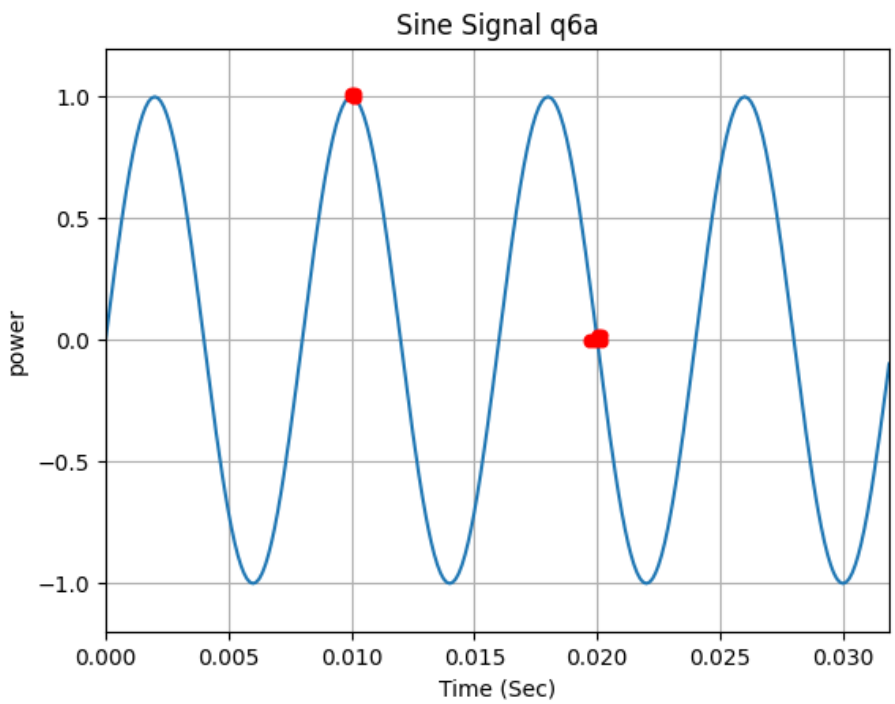
a) האוטוקורלציה היא מכפלה פנימית של וקטור עם עצמו (עם הזחה של k וריפוד עם אפסים בצדדים). עבור k=0, זוהי פשוט מכפלה פנימית בין הוקטור לעצמו ואנחנו יודעים שהמקסימום האפשרי של מכפלה פנימית של וקטור עם וקטור אחר היא עם עצמו. המכפלה הפנימית עם עצמו היא הנורמה בריבוע וזו אנרגיית האות.

b) אם האות מחזורי במחזור של P, אז בהזחה של $k = \pm n \cdot P$ כאשר $n \in \mathbb{N}$ המכפלה שבסכום היא מכפלה של איבר בעצמו. אם נסתכל על הזחה קטנה באחד או גדולה באחד מnP, נקבל שכמות האיברים היא אותה כמות אך במקום מכפלה פנימית עם עצמו, מכפילים את הוקטור עם וקטור אחר.

c) כאשר אנו מתרחקים מ$k = 0$, ככה יש יותר ריפוד של אפסים שמאפס את המכפלה וכך עבור כל מחזור שנתרחק מ0 נאבד אנרגיה של מחזור שלם.

שאלה 6

א

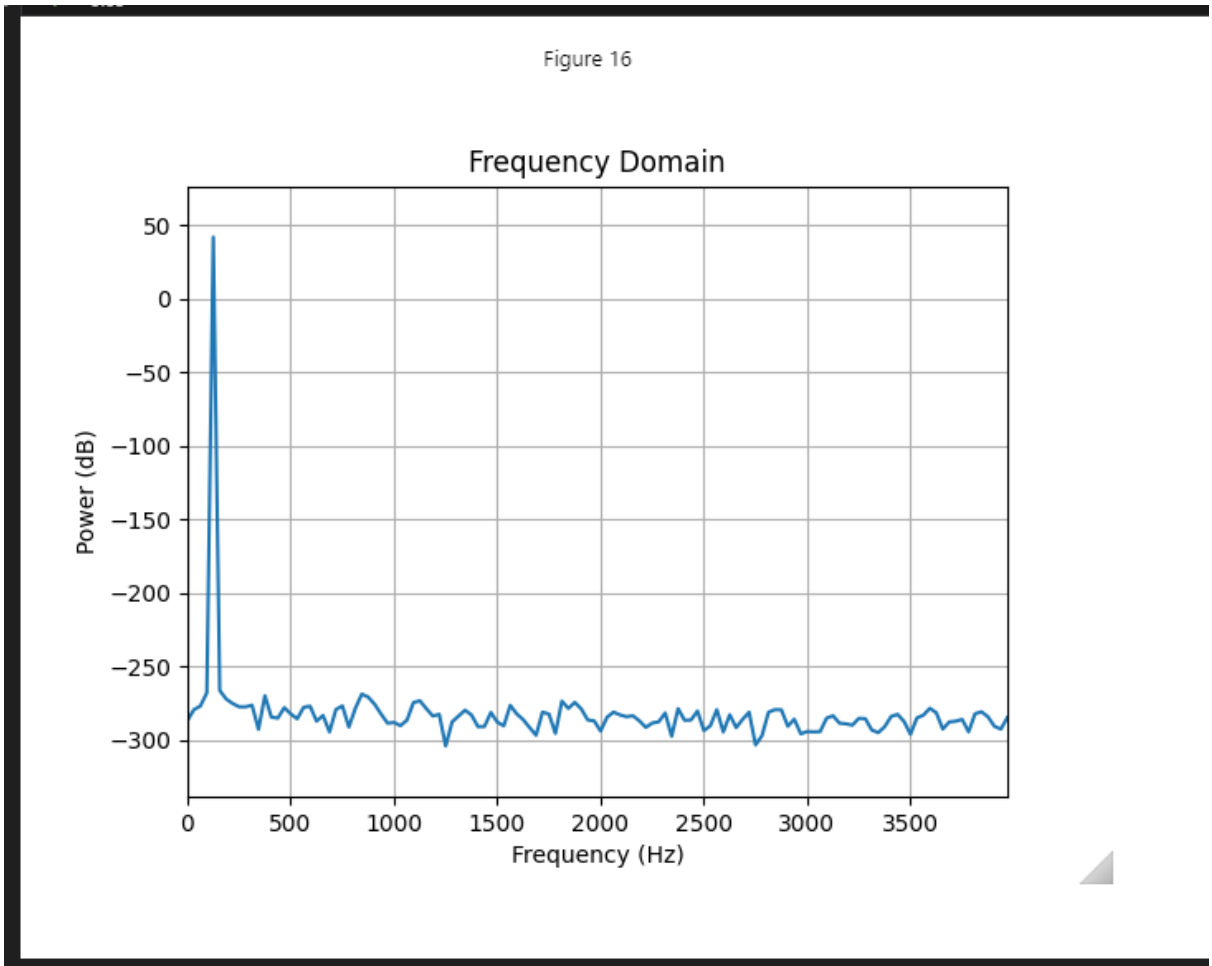


נסתכל על שתי נקודות בגרף אשר נוגעות ב grid שקל לחשב בעזרתן את זמן המחזור.
נשים לב כי ההפרש בין שתי הנקודות ה$\frac{5}{4}$ זמן מחזור. ומבחינת זמן הוא שווה ל0.010 sec.

ולכן נציב במשוואה ונפתור. $\frac{5}{4}T = 0.01\,\text{sec} \rightarrow T = 0.008$. מכאן התדר הוא $125[Hz]$ – כמו שציפינו.

אורי מירז 212641229
עידן פוגרבינסקי 325069565
ספרה מזהה 5

ב

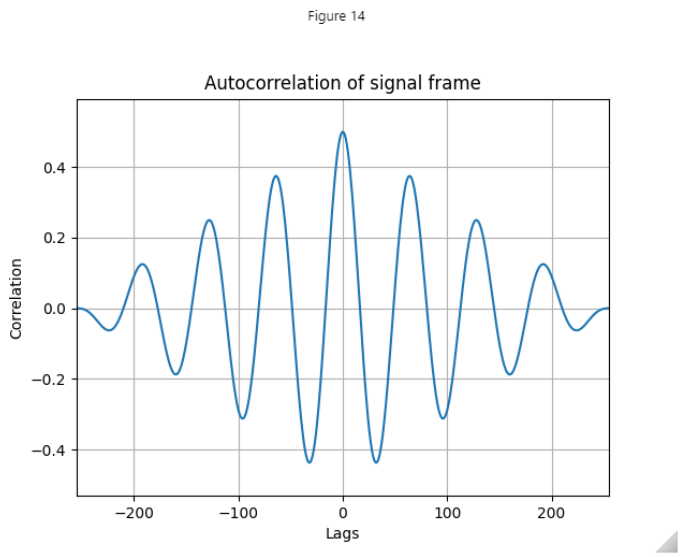


We can see a spike at 125, as we expected to see for our frequency, (can be seen by zooming in)

לפי התמונה ניתן לראות (באמצע למעלה) שהתדר של ה*pitch* הוא 125.5 הרץ. זה לא בדיוק אותו מספר אך מאוד קרוב וניתן להסביר זאת בגלל חוסר דיוק שלנו בסימון ה*peaks* של הגל.

אורי מירז 212641229
עידן פוגרבינסקי 325069565
ספרה מזהה 5

ג



The first peak is after at $k = 64$, so we know that (according to Q5), $p = 64$. The sampling frequency is 8K [Hz] and so, the pitch is $\frac{8K}{64} = 125[Hz]$

סעיפים ד' ה' נמצאים בקוד ואכן קיבלנו אישור על התקינות.

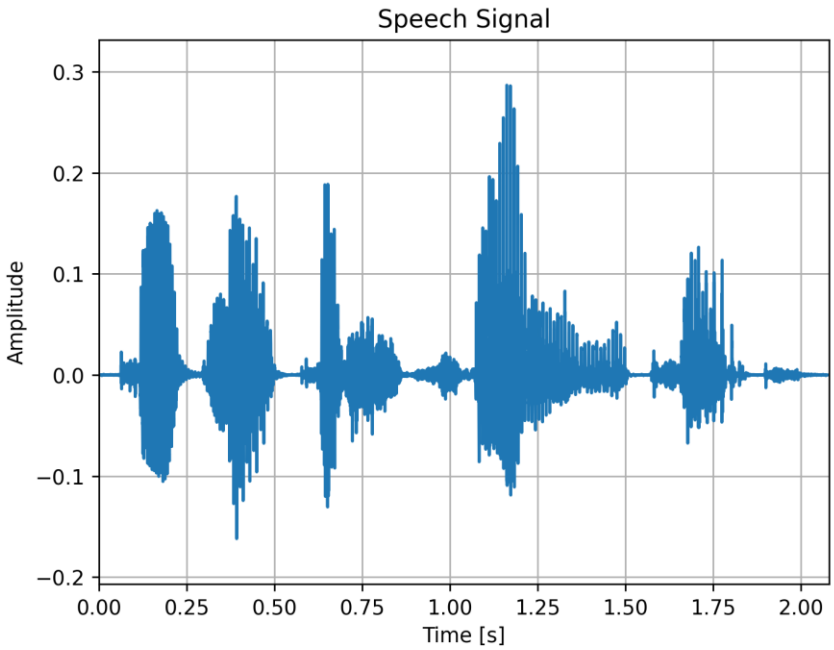
שאלה 7

1

א. בקוד
ב. בקוד
ג. 2 blue fish swam in the tank

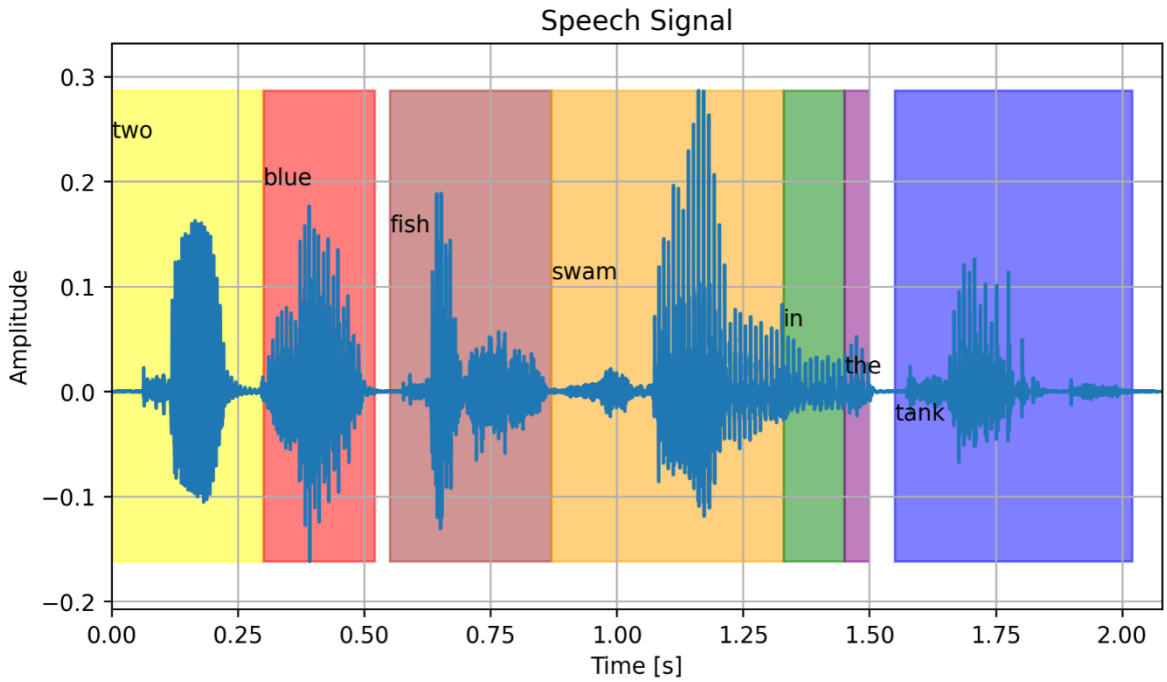
2

א.

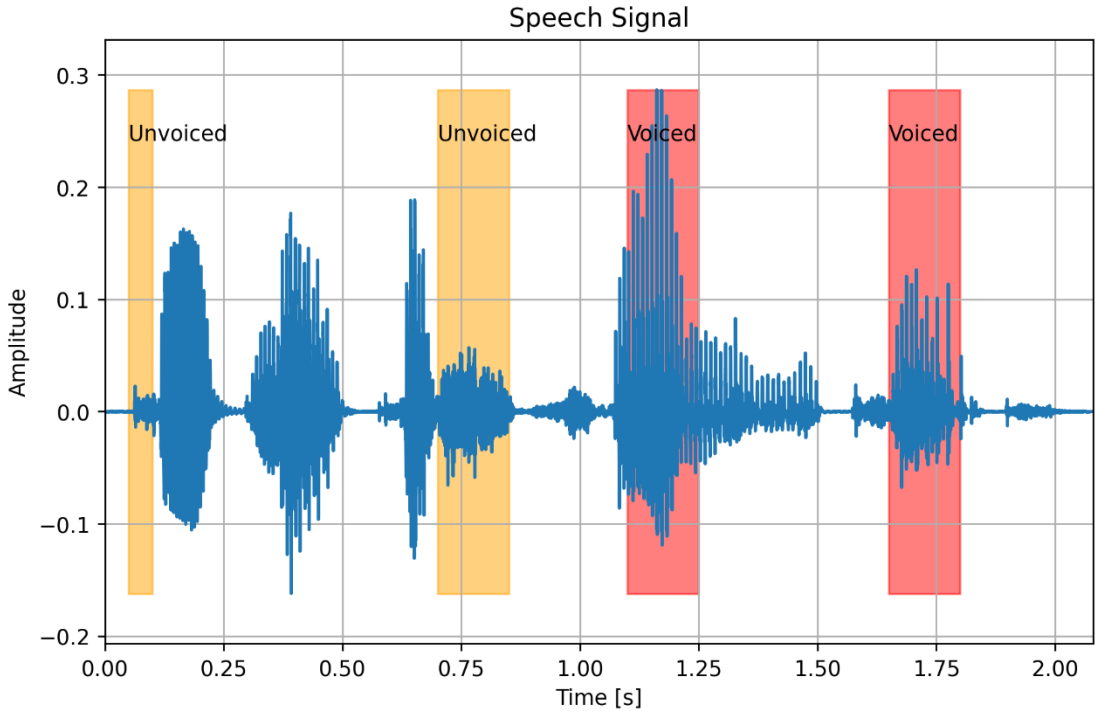


ב.

אורי מירז 212641229
עידן פוגרבינסקי 325069565
ספרה מזהה 5



3



Voiced, are parts of the speech where the vocal cords are used, it can be seen by noting a pitch in the wave form, unlike areas where the is no prominent pitch, but just a “broad band” noise like sound, it is when the vocal cords aren’t used.

אורי מירז 212641229
עידן פוגרבינסקי 325069565
ספרה מזהה 5

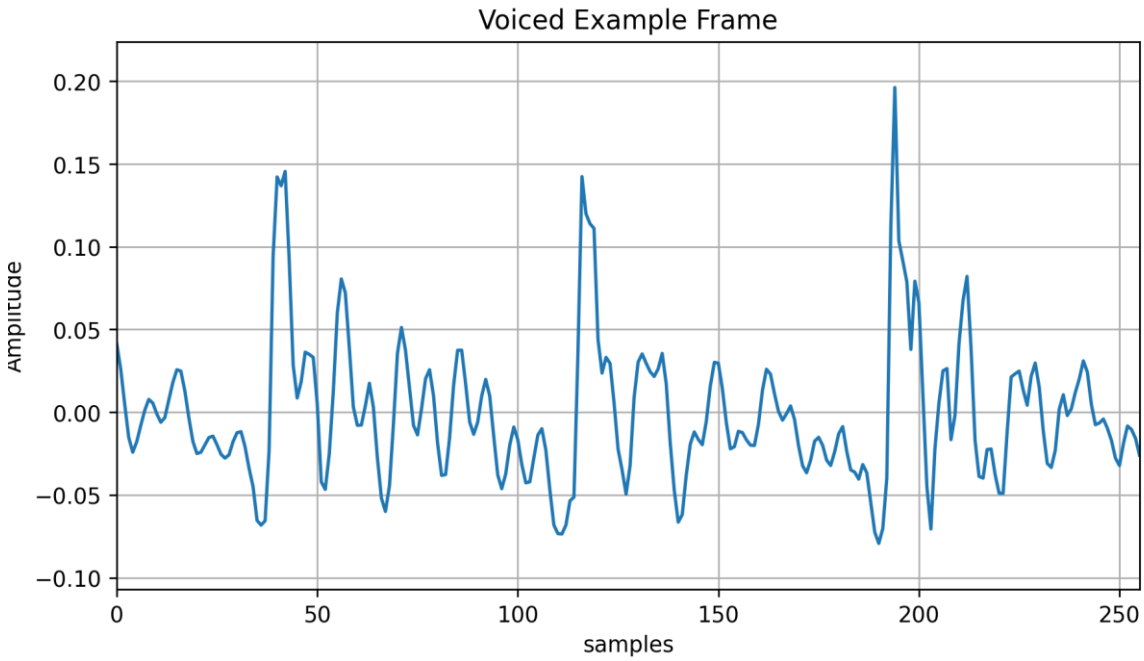
4

א. בקוד
ב. אינדקס המסגרת הוא 34 וההבהרה היא A מתוך המילה swam.
ג. אינדקס המסגרת הוא 2 וההבהרה היא T מתוך המילה two.

שאלה 8

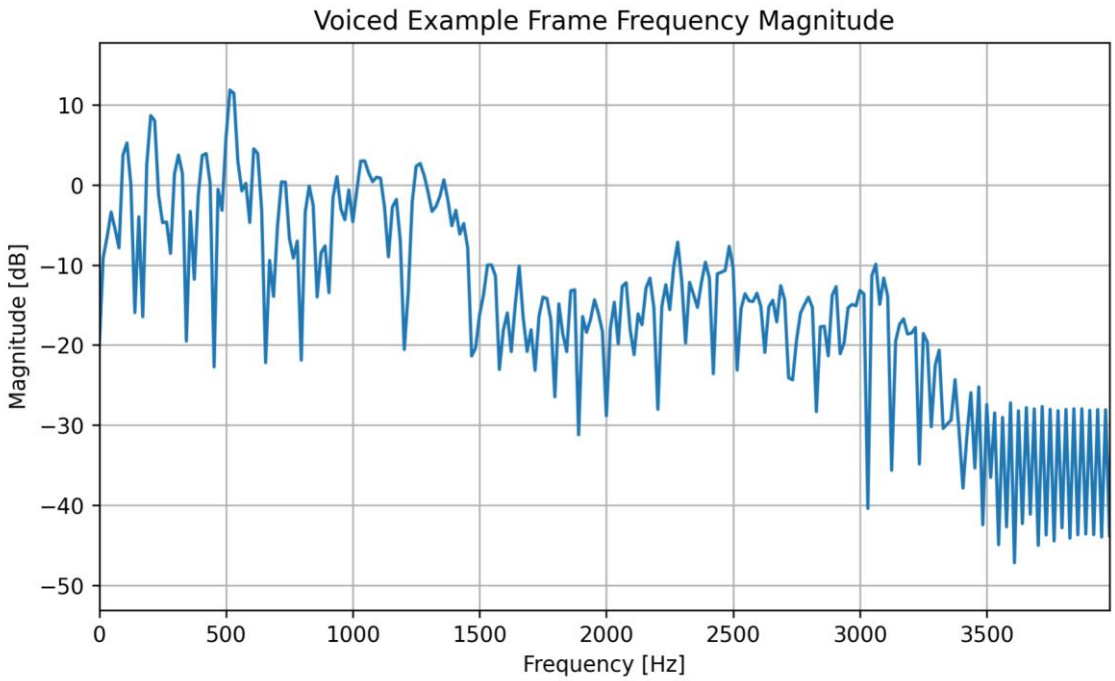
1

א.



ניתן לראות את הפיקים של רכבת ההלמים בערך ב40 וב115 ולכן, אורך המחזור בדגימות הוא 75.
האות נדגם בתדירות של $8k[Hz]$ ולכן, כמו שעשינו בתחילת התרגיל, התדר של pitch הוא $= \frac{8k}{75}$
$106.66[Hz]$ ובשניות הוא $\frac{75}{8k} = 0.0094\ [s]$.

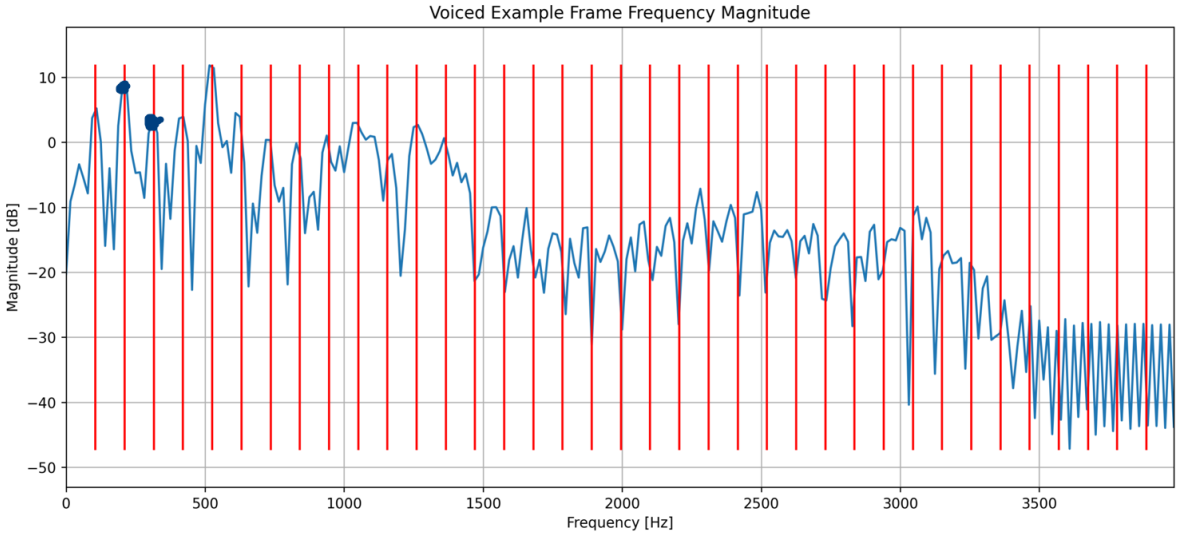
אורי מירז 212641229
עידן פוגרבינסקי 325069565
ספרה מזהה 5
ב.



יש 512 דגימות (כולל הpadding) ומכאן גודל הFFT הוא 512. תדירות הדגימה היא 8000 ולכן רזולוציית התדר היא $\frac{8000}{512} = 15.625$.

ניתן לראות כמה פורמנטים שונים (העוצמה שלהם גבוהה יותר משאר התדרים) והם בערך 105, 205, 515. אנחנו יודעים שהpitch הוא הפורמנט הקטן ביותר ולכן, מהתמונה אנחנו יכולים להעריך שהpitch הוא 105 וזה אכן תואם את הסעיף הקודם.

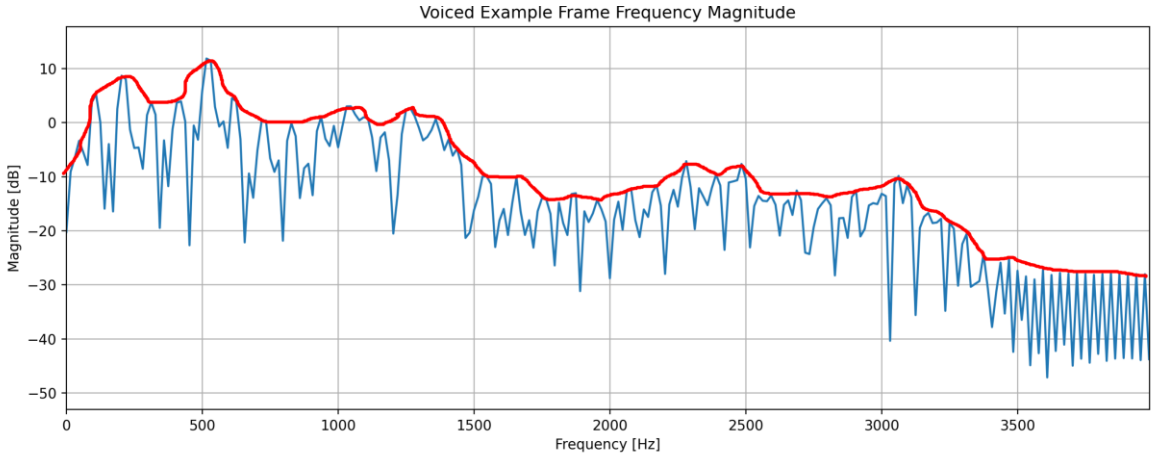
ג. ניזכר כי ההרמוניה הK היא תהיה בעלת תדירות K כפול התדר המרכזי, כאשר התדר המרכזי הוא הpitch, ולכן אם נסתכל על גרף ונבדוק עבור כל peak שנראה, אם הוא נמצא בכפולה טבעית של התדר המרכזי, נדע לזהות את ההרמוניות. עשינו plot בו סימנו כפולות טבעיות של הpitch ונספור רק את הדוגמאות בהן הפסגה מסתדרת עם הקו האדום שסימנו.



הצלחנו לספור 13 הרמוניות.

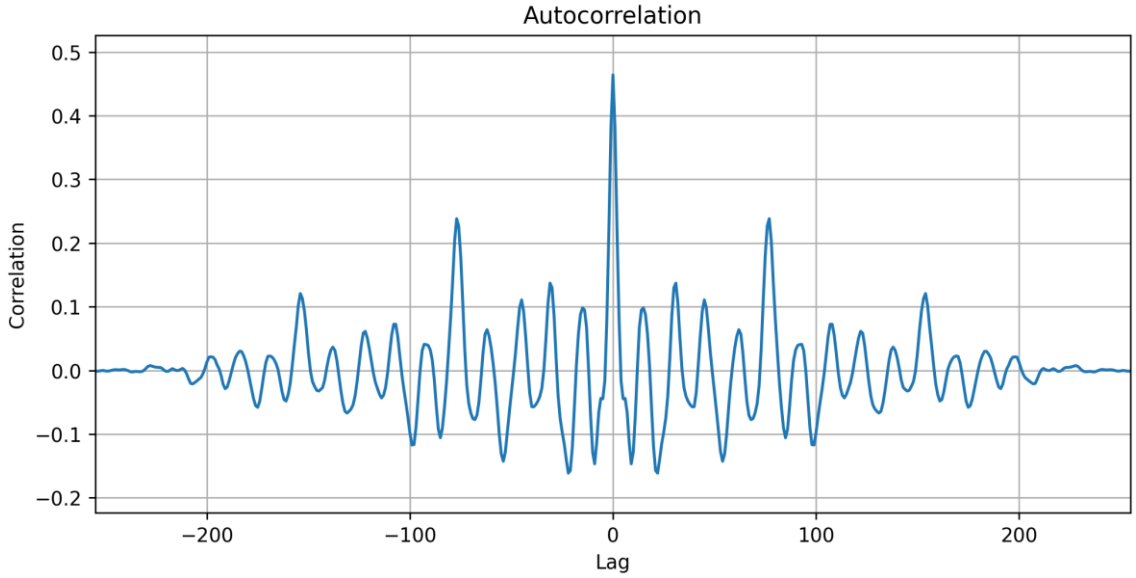
אורי מירז 212641229
עידן פוגרבינסקי 325069565
ספרה מזהה 5

ד. המעטפת מסומנת באדום



2

א.



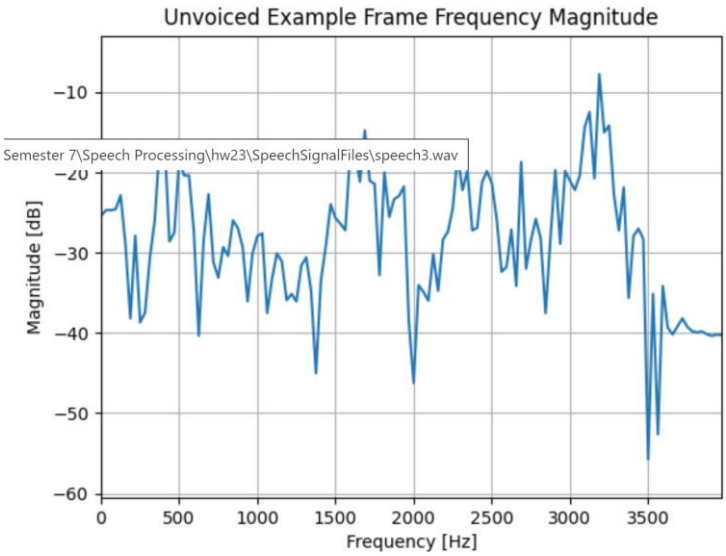
ניתן לראות כי המחזור של pitch הוא אחרי 76 דגימות (מאחר והפיקים הקטנים יותר הם הרמוניות ולא pitch) ולכן, הכמות זמן למחזור היא $\frac{76}{8000}$ שניות והתדר הוא $\frac{8000}{76} = 105.26\ [Hz]$.

ב. בעזרת הפונקציה pitch_detect_corr, מצאנו כי הכמות דגימות במחזור של pitch הוא 77 דגימות והתדר הוא 103.9 הרץ.

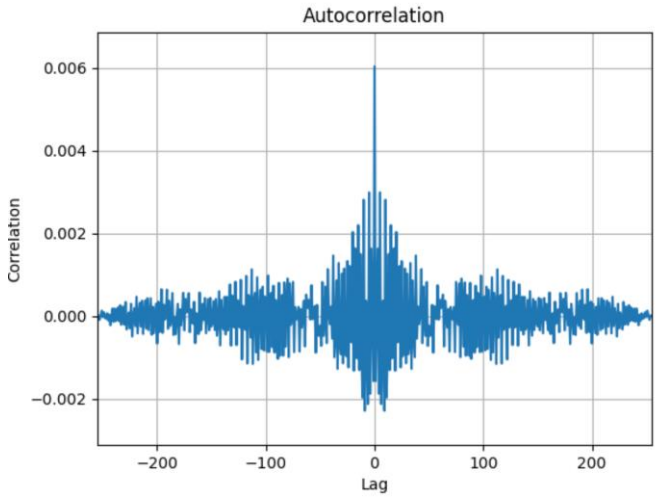
אורי מירז 212641229
עידן פוגרבינסקי 325069565
ספרה מזהה 5

3

א. לא, לאחר הסתכלות בהתמרת הפוריה, לא ניתן לראות מחזור pitch, האות נראה רועש וההפרש בין הpeaks אינו קבוע ואין שום חוקיות שניתנת לזיהוי בעין.



ב. ההבדל העיקרי שניתן לשים לב אליו בהסתכלות על שני הגרפים, והוא שבגרף הקולי, פונקציית האוטוקורלציה נראית "חלקה", לעומת הפונקציה של הגרף הלא קולי, שנראית מחוספסת ומלאה ברעש, בלי שום הרמוניות או קפיצות קבועות.



שאלה 9

1

א. מומש בקוד

ב. ניתן לראות שכמות הzero crossings משמעותית יותר גדולה בעבור הרעש (123) מאשר הסיגנל של הסינוס (6.5).

2

מומש

3

אכן קיבלנו שהפונקציה חזתה שהvoiced הוא voiced והunvoiced הוא unvoiced.

אורי מירז 212641229
עידן פוגרבינסקי 325069565
ספרה מזהה 5

4

א.

```
if __name__ == "__main__":

    print(f'Average energy of voiced: {short_time_energy(voiced_example_frame):.7f}')
    print(f'Average energy of unvoiced: {short_time_energy(unvoiced_example_frame):.7f}')
✓ 0.0s

Average energy of voiced: 0.0018121
Average energy of unvoiced: 0.0000236
```

ניתן לראות שהאנרגיה הממוצעת של הvoiced היא יותר גדולה.

ב.

```
if __name__ == "__main__":

    print(f'Zero crossing of voiced: {zero_cross(voiced_example_frame)}')
    print(f'Zero crossing of unvoiced: {zero_cross(unvoiced_example_frame)}')
✓ 0.0s

Zero crossing of voiced: 41.0
Zero crossing of unvoiced: 153.0
```

קצב חציית ה0 של הunvoiced יותר גדולה.

ג.

```
if __name__ == "__main__":

    print(f'Classification result of voiced: {vu_classify(voiced_example_frame, fs)}')
    print(f'Classification result of unvoiced: {vu_classify(unvoiced_example_frame, fs)}')
✓ 0.0s

Classification result of voiced: True
Classification result of unvoiced: False
```

הפונקציה אכן סיווגה נכון את הדוגמאות.