

Image Processing - Exercise 2

Neriya Ben David, neriyaabd, 206698581

מבוא :

ראשית, מטרת התרגיל היא ניקוי רעשי רקע הקיימים בשני קבצי קול נתונים. הכלים המרכזיים בהם השתמשתי הינם

- DFT , $IDFT$ בכדי לקבל פירוק של התדרים לאורך כל הקטע קול ו- $STFT$, $ISTF$ לקבלת פירוק של התדרים לפי פרקי זמן שונים.
- הצגת המגניטודה של כל התדרים שהתקבלו מ- DFT ו- $STFT$ באמצעות גרפים וספקטרוגרמה.
- הסתכלות בנתונים על התדרים, ומניפולציה עליהם באמצעות הספריות $pandas$, $numpy$.

ההבדלים בין שני קבצי הקול נבעו מכך שבקטע קול הראשון היה רעש קבוע לאורך כל הקטע ואילו בקטע קול השני הרעש אינו קבוע והתפרש על פני זמנים שונים. רעש זה יכול להשפיע על הדרך לפתרון בכך שבחלק השני ייתכן שנאלץ להסתכל גם על קטעי זמן שונים על מנת לאבחן את מקור הרעש, ולנקות אותו, לעומת החלק הראשון בו הרעש כפי שנשמע, הוא ככל הנראה קבוע.

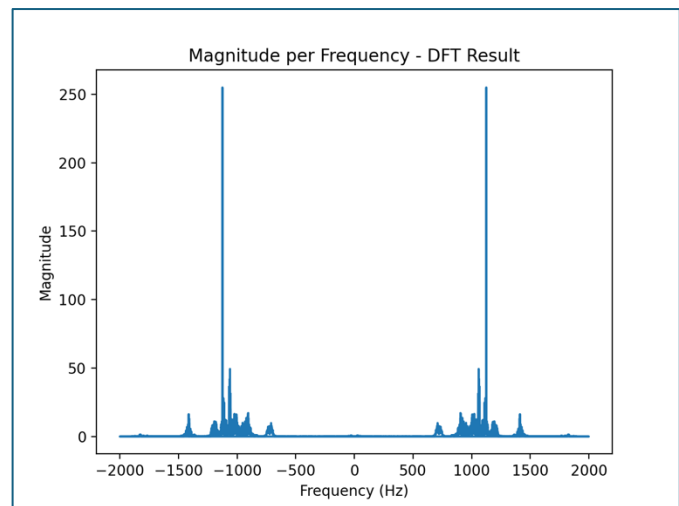
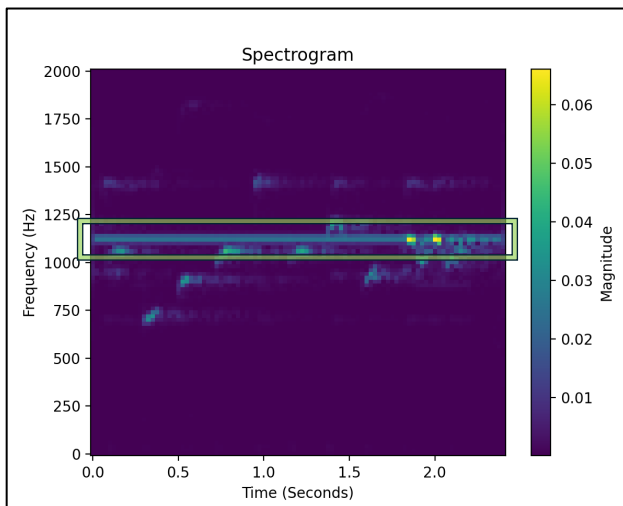
אלגוריתמים :

אלגוריתם – קטע קול 1 :

מחשבה וביסוס רעיון האלגוריתם : ראשית, משמיעת קטע הסאונד האינטואיציה שלי הייתה שהרעש בסאונד הינו קבוע לאורך כל הקטע.

כדי לקבל יותר מידע, יצרתי גרפים המייצגים את מגניטודות התדרים מה- DFT של כל קטע הקול. בנוסף יצרתי ספקטרוגרמה של תדרים לפי זמנים שונים בקטע.

הגרפים שהתקבלו הינם :



מהספקטרוגרמה ניתן לראות שישנו תדרים קבועים בסביבות 1100 Hz לאורך כל הקטע קול (מסומן במלבן), מהגרף שמציג את המגניטודות ניתן להבחין כי ישנו תדר / תדרים סביב 1100 Hz , -1100 Hz שהמגניטודה שלהם חזקה משמעותית, בסביבת 250, ולכן הסקתי שזה מקור הרעש (מצפים לרעש אחיד בסביבת תדר זה). כדי למצוא את התדר / תדרים אלו, מיינתי את כל התדרים בסדר יורד לפי המגניטודה שלהם וקיבלתי את הטבלה הבאה:

Frequency	-1125.00	1125.00	-1060.416667	1060.416667	-1062.083333	1062.083333	1063.333333	-1063.333333
Magnitude	254.94	254.94	49.330000	49.330000	41.980000	41.980000	41.640000	41.640000

מהטבלה ניתן להבחין שיש תדר סביב 1100 Hz שהמגניטודה שלו נמצאת באזור 250, התדר 1125 Hz (ומגניטודת התדר 1125 Hz גם כן, כמו שניתן להבחין בגרף ולצפות מתכונת השיקוף של המגניטודה בהתמרת פורייה). לכן, האלגוריתם שאבחר לממש יאפס את מקור הרעש, התדרים 1125 Hz , -1125 Hz .

a. תיאור שלבי האלגוריתם:

1. קריאת הקובץ קול.
2. הפעלת DFT על מערך הדגימות שהתקבל (באמצעות שימוש באלגוריתם FFT).
3. מציאת התדרים בעלי המגניטודה הכי גבוהה ואיפוסם.
4. ביצוע $IDFT$ באמצעות שימוש באלגוריתם $IFFT$.
5. החזרת החלק הממשי שהתקבל כתוצאה מאלגוריתם $IFFT$.

b. מימוש ופרטי מימוש של האלגוריתם:

- קריאת הקול באמצעות פונ' `wavfile.read` מספריית `scipy.io`.
- ביצוע התמרת פורייה על קלט הקול ע"י שימוש בפונ' `fft` מספריית `numpy` על מנת לקבל את הפירוק ה- DFT של קטע הקול.
- למציאת התדרים בעלי ערך המגניטודה הגבוהה ביותר הפעלתי ערכים מוחלטים על כל הערכים המרוכבים שהתקבלו מהתמרת פורייה באמצעות פונ' `abs` מספריית `numpy`. לאחר מכן השתמשתי בפונ' `argmax` על מנת לגשת לאפס את התדרים בעלי המגניטודה הגבוהה ביותר (תדרים בקואר' $-argmax$, $argmax$).
- שימוש בפונ' `fft.ifft` מספריית `numpy` על מנת להחזיר את הדגימות לממד הזמן.
- החזרת החלק הממשי כיוון שהוא מייצג את החלק בעל המשמעות בקטע קול.

הסבר על `parameters - hyper thresholds` ובחירות נוספות:

- לא השתמשתי ב-`hyper parameters` וב-`thresholds` באלגוריתם זה.

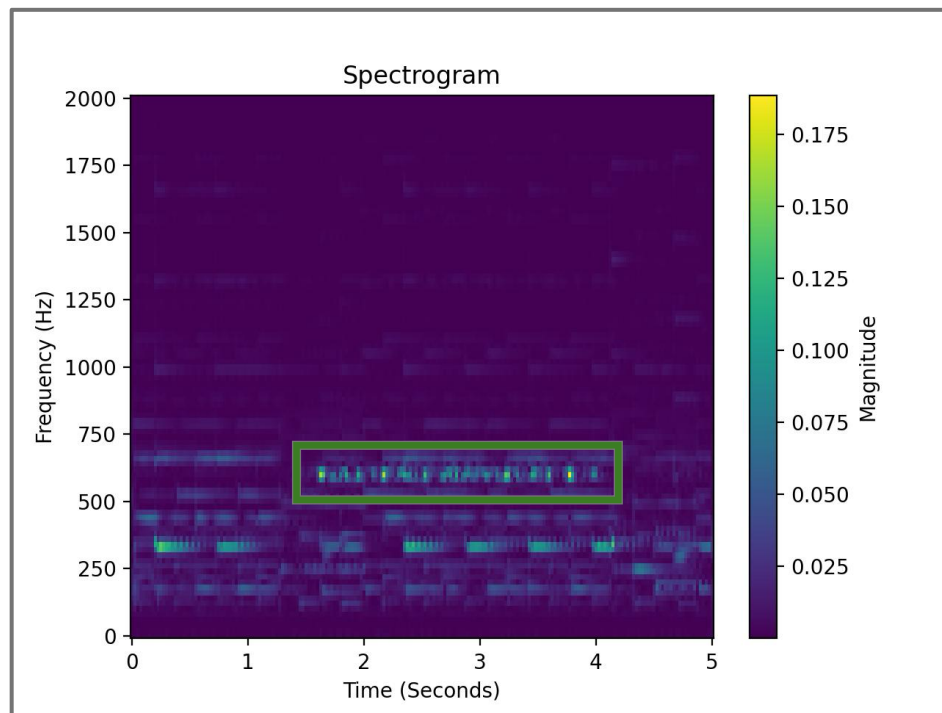
אתגרים שעלו ממיושם האלגוריתם:

- אתגרים שליוו אותי במימוש האלגוריתם הם הבנת ייצוג ה- DFT ב-`numpy` ושימוש בספריית `scipy.io`, התמודדתי עם אתגרים אלו באמצעות קריאת הדוקו' של הפונ' באינטרנט, ודוג' שלקחתי מ-`copilot` על קלטים קטנים והתנסות איתם.

אלגוריתם – קטע קול 2:

מחשבה וביסוס רעיון האלגוריתם: מההאזנה לקטע קול זה, ישנו רעש שמופיע החל מהשנייה הראשונה, עד תחילת – אמצע השנייה הרביעית. על מנת להבין יותר על התדרים הקיימים בכל יחידת זמן לפי זמנים שונים. השתמשתי, כפי שראינו בכיתה ב- $STFT$, וייצגתי אותו באמצעות ספקטרוגרמה. ב- $STFT$ בחרתי 200 דגימות עבור כל יחידת זמן, שהן מהוות כ-0.025 שניות.

הספקטרוגרמה שהתקבלה:



נזכור כי אנו מחפשים רעשי רקע שמתחילים במהלך השנייה הראשונה ומסתיימים במהלך השנייה הרביעית. ניתן לראות בגרף כי ישנם תדרים בין $550 - 650 \text{ Hz}$ באים והולכים בפרק זמן זה (מסומן בירוק), שהמגניטודה שלהם גדלה וקטנה.

כדי לבחון באופן יותר טוב את המידע שהתקבל מהתבוננות בספקטרוגרמה, הסתכלתי על מגניטודות התדרים בין $550 - 650 \text{ Hz}$ מ-1 שניות עד 4.5 שניות הללו.

מהתבוננות אבחנתי כי ישנה עליות וירידות חדות במגניטודה בתדרים $580 - 640 \text{ Hz}$ החל מ-1.6 שניות ועד 4.025 שניות. מצורפים זמני ההתחלה והסיום של עליית המגניטודות של התדרים הנ"ל.

		Time (Sec)													
Frequency (Hz)		÷ 1.55	÷ 1.575	÷ 1.6	÷ 1.625	÷ 1.65	÷ 1.675	÷ 1.7	÷ 1.725	÷ 1.75	÷ 1.775	÷ 1.8	÷ 1.825	÷ 1.85	÷ 1.875
	540	0.01689	0.01245	0.01165	0.01184	0.00912	0.00874	0.00556	0.00744	0.00518	0.00426	0.00299	0.00234	0.00629	0.00505
	560	0.00340	0.00300	0.00780	0.00945	0.01033	0.00309	0.00824	0.01077	0.01165	0.01138	0.00950	0.01190	0.00794	0.00983
	580	0.00337	0.00563	0.02678	0.08958	0.06850	0.01017	0.03310	0.05954	0.05845	0.04016	0.05090	0.01877	0.06559	0.04603
	600	0.01399	0.01036	0.03824	0.16600	0.13282	0.03763	0.03556	0.06633	0.06782	0.01745	0.07598	0.04413	0.11202	0.03871
	620	0.01683	0.01593	0.04218	0.09936	0.08280	0.03412	0.02468	0.01876	0.02450	0.01561	0.03579	0.06180	0.06877	0.00353
	640	0.00615	0.00690	0.01064	0.01163	0.01505	0.00937	0.00365	0.00749	0.01063	0.01180	0.01245	0.00327	0.01278	0.01607
	660	0.00210	0.00273	0.00554	0.00567	0.01387	0.01593	0.01809	0.01924	0.01888	0.01815	0.02015	0.01844	0.01291	0.01476
	680	0.00217	0.00177	0.00183	0.00222	0.00916	0.00887	0.01003	0.01044	0.01107	0.01110	0.00672	0.00659	0.00791	0.00880
		Time (Sec)													
Frequency (Hz)		÷ 3.75	÷ 3.775	÷ 3.8	÷ 3.825	÷ 3.85	÷ 3.875	÷ 3.9	÷ 3.925	÷ 3.95	÷ 3.975	÷ 4.0	÷ 4.025	÷ 4.05	÷ 4.075
	540	0.01991	0.02493	0.02308	0.02072	0.01990	0.01861	0.01839	0.01906	0.02221	0.01881	0.01668	0.01488	0.01470	0.01226
	560	0.01482	0.01440	0.01526	0.00977	0.00568	0.00686	0.00335	0.00469	0.00183	0.00116	0.00227	0.00388	0.00308	0.00265
	580	0.03589	0.11003	0.07964	0.03159	0.01204	0.02037	0.01843	0.00736	0.02632	0.03635	0.03379	0.01938	0.01686	0.00968
	600	0.04310	0.18847	0.10230	0.03404	0.00842	0.01239	0.03352	0.01792	0.03318	0.08892	0.07562	0.01250	0.02363	0.00974
	620	0.05295	0.10943	0.06277	0.02221	0.01767	0.00756	0.02382	0.01732	0.02532	0.05527	0.05017	0.01337	0.01134	0.00105
	640	0.01631	0.02405	0.02010	0.00782	0.00641	0.00909	0.01474	0.01590	0.00913	0.01573	0.01157	0.01799	0.01692	0.01596
	660	0.01713	0.00893	0.01620	0.02057	0.02553	0.02544	0.02555	0.02573	0.02757	0.03148	0.02856	0.02900	0.03576	0.03629
	680	0.00967	0.00669	0.01049	0.01124	0.01631	0.01749	0.01872	0.01759	0.01614	0.01326	0.01811	0.01901	0.02014	0.02038

כפי שניתן לראות, העלייה במגניטודות של התדרים מתמתנת אחרי 4.025 שניות ולכן נסיק כי העליות החדות במגניטודות בתדרים שזיהינו בספקטרוגרמה נגמרות ב-4.025 שניות.

אבחנה: נשים לב כי הזמנים לא מדויקים בדיוק מאחר ואנחנו עובדים על זמן דיסקרטי וכל ההתייחסות לזמנים בכתובתי היא לפי זמנים אלו.

לכן, האלגוריתם אותו אממש, יאפס את התדרים בהם מצאנו מגניטודות גדולות וקטנות ביחידות הזמן הרלוונטיות כפי שתואר.

a. תיאור שלבי האלגוריתם:

1. קריאת הקובץ קול.
2. הפעלת $STFT$ על מערך הדגימות.
3. איפוס התדרים $660\text{ Hz} - 580$ בזמנים $4.025 - 1.6$ שניות על ה- $STFT$ שהתקבל.
4. ביצוע $ISFT$ על ה- $STFT$ לאחר האיפוס והחזרתו.

מימוש ופרטי מימוש של האלגוריתם:

- קריאת קובץ הקול באמצעות פונ' `wavfile.read` מספריית `scipy.io`.
- ביצוע $STFT$ על המערך ההתקבל מקריאת הקובץ ע"י שימוש בפונ' `stft` מספריית `scipy.signal`.
- העברת ה- $STFT$ ל-`DataFrame` בספריית `pandas` והוספת קואו' בשמות לשורות והעמודות לפי תדרים וזמן בהתאמה.
- איפוס התדרים בזמנים שצוינו לעיל ע"י שימוש ב-`DataFrame.loc` ב-`pandas`.
- שימוש בפונ' `istft` מספריית `scipy.signal` על מנת להחזיר את הדגימות לממד הזמן והחזרת המערך שהתקבל מפונ' זו.

הסבר על `parameters - hyper`, `thresholds` ובחירות נוספות:

- השתמשתי ב-`hyper parameter` של 200 דגימות בכל יחידת זמן ב- $STFT$, בחירה זו הייתה כיוון ש-200 דגימות מתחלק במספר הדגימות הכולל (כדי שכל החלונות בגודל שווה) וגם זה נותן ייצוג טוב בספקטרוגרמה לתדרים (על ידי ניסוי וטעיה של מגוון גדלי חלון).
- לא השתמשתי ב-`thresholds` באלגוריתם זה.

אתגרים שעלו ממימוש האלגוריתם:

- אתגר שליווה אותי במימוש האלגוריתם היה כיצד לסנכרן בין הזמן והתדר לאינדקסים ב- $STFT$ כדי להסתכל על הקואורדינטות הנכונות במערך הדו - ממדי. כדי להתמודד עם האתגר השתמשתי בספריית `DataFrame` של ספריית `pandas` על מנת לתת לכל קואו' ייצוג לפי התדר והזמן אותו היא מייצגת כדי שאוכל לגשת ולהסתכל על התאים הרלוונטיים בקלות.

מסקנות:

בתרגיל זה ראיתי לעומק את השימוש בהתמרת פורייה על קבצי סאונד על מנת לערוך אותו ובפרט במקרה זה, להסיר רעשים.

במהלך התרגיל התמודדתי עם הורדת רעש יציב ורעש שאינו יציב, הן באמצעות DFT על כל קטע הסאונד ביחד (בקובץ הראשון) והן באמצעות ביצוע DFT בפרקי זמן שונים ע"י שימוש ב- $STFT$ (בקובץ השני). בתרגיל זה הבנתי לעומק את האופן בו ניתן להשתמש בממד התדר המתקבל מהתמרת פורייה על מנת לבצע מניפולציות על קטעי קול, ואת המשמעות והיתרונות של התמרת פורייה המאפשרת ביעילות לערוך סאונד ע"י שינוי התדרים באופן מהיר ופשוט. בנוסף, לאחר בדיקה של קבצי הסאונד שייצאתי, ווידאתי כי אכן הרעש הוסר משני הקבצים.