

Image Processing - Exercise 2

Ofek Avidan, ofek.avidan, 318879574

מבוא

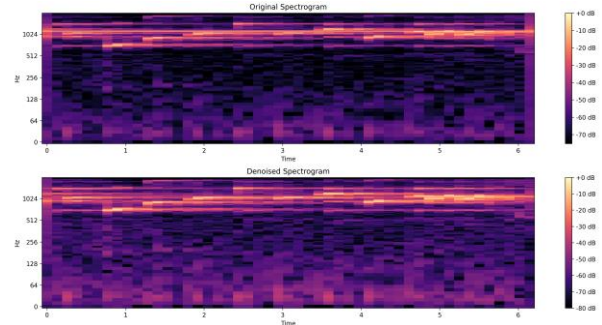
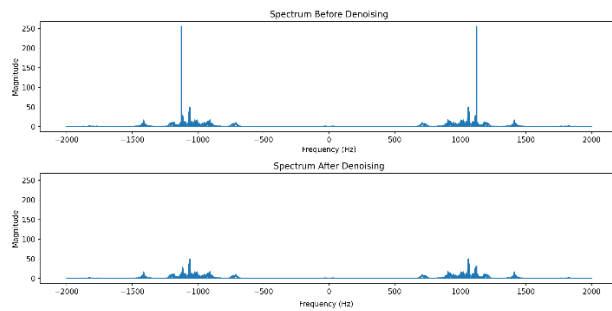
ראשית, המטרה העיקרית של תרגיל זה, היא ליישם טכניקות של זיהוי והסרת רעשים בקבצי אודיו. הכלים העיקריים המשמשים להסרת רעשים הם טרנספורם פורייה מהיר ואיפוס רכיב התדר עם העוצמה המקסימלית (עבור שאלה 1), או התמרת פורייה קצרה בחלונות (STFT) ואיפוס רכיב התדר לפי ספקטרוגרמה בה השתמשתי (והעין האנושית). העובדה שיכלנו להסתמך על כך שמדובר בקובץ האודיו הספציפי הנ"ל, עזרה לי לבצע ניסוי וטעייה (מחיקה והאזנה) על קובץ האודיו שהפונקציה יצרה. אפרט על כך בהמשך. כמובן שישנם כלים נוספים כמו חציון או קונבולוציה שיכלו לשמש אותנו אך בחרתי בכלים שנראים פשוטים יותר לטעמי.

בשאלה הראשונה הרעש "קבוע", או לפחות מתמשך לאורך כל קטע האודיו, וניתן יהיה לראות זאת בספקטרוגרמה שאצרף לאחר מכן בשאלת האלגוריתם. גם העוצמה שלו זהה לאורך כל קטע האודיו. לכן התמרה רגילה (אפשר לומר, חלון אחד גדול) תספיק לנו. בשאלה השניה הרעש לא קבוע, ולכן נרצה לחלק את האודיו לחלונות בהם יהיה קבוע (או כמה שיותר קבוע), בדיוק כפי שלמדנו בכיתה במצבים של אודיו לא קבוע וגלים לא סטציונריים. גם כאן אראה תמונה בהמשך.

אלגוריתם

אתחיל בשאלה הראשונה. ראשית, האלגוריתם לניקוי הרעש עבור קובץ הקלט הראשון הוא:

1. קריאת אודיו: טעינת קובץ האודיו באמצעות ספריית librosa, תוך ציון קצב דגימה של 4 קילו-הרץ.
2. ניתוח תדרים על ידי טרנספורם פורייה מהיר (FFT): החלת התמרת פורייה מהירה (FFT) על אות האודיו, ומעבר ממרחב הזמן למרחב התדר.
3. זיהוי רכיבי תדר כרעש: זיהוי רכיבי התדר הקשורים לרעש על ידי מציאת המדדים של הגודל המרבי ב-FFT.
4. שינוי רכיבי תדר: הגדרת רכיבי התדר שזיהינו בשלב 3 לאפס, ובכך למעשה הסרה של הרעש.
5. FFT הפוך: בצע את ה-FFT ההיפוך כדי להשיג את אות האודיו המעורער בתחום הזמן.
6. נרמול: נרמל את הסיגנל החדש כדי להבטיח שהוא נמצא בטווח של מספר שלם של 16 ביטים.



בשתי התמונות ניתן לראות הספקטרוגרמה לפני ואחרי הסרת הרעש. ניתן לראות בתמונה השמאלית שהערך הגבוה ביותר בספקטרוגרמה עובר מ-250 ל-50, וניתן מעט לראות בתמונה הימנית את הסרת רעשי הרקע. **היישום** מסתמך במידה רבה על ספריית librosa עבור טעינת אודיו וחישוב FFT. השלבים הכוללים FFT, זיהוי תדר רעש ו-FFT הפוך מיושמים באמצעות NumPy.

מבחינת **היפרפרמטרים וספים**, הגדרתי כמה דברים:

ראשית, קצב דגימה: קצב הדגימה מוגדר ל-4 קילו-הרץ, התואם למפרטים של קבצי השמע שסופקו. סף רכיב תדר: לא נעשה שימוש בסף ספציפי; במקום זאת, האלגוריתם מזהה את רכיבי התדר עם העוצמה המקסימלית.

מבחינת **אתגרים ופתרונות**, מה שהיה קשה היה טיפול במספרים מורכבים: FFT מביא לתוצאה במספרים מרוכבים, ושינוי תדרים ספציפיים דורש התחשבות בגודל ובפאזה. האלגוריתם מטפל בכך על ידי שינוי הגודל תוך שימור הפאזה במהלך ה-FFT הפוך.

נורמליזציה: נורמליזציה של האות המעורפל היא חיונית כדי לשמור על עקביות ולמנוע גזירה. האלגוריתם מבטיח שהסיגנל החדש נמצא בטווח של מספר שלם של 16 ביטים.

אמשיך **לשאלה השניה: האלגוריתם שלה הוא כלדהלן:**

1. קריאת אודיו: נקרא את קובץ האודיו באמצעות מודול ה-wavfile מ-scipy, עם קצב דגימה של 4 קילו-הרץ.
2. STFT (טרנספורמציה של פורייה בזמן קצר): נחשב את התמרת פורייה בחלונות (STFT) כדי לנתח את תוכן תדר הזמן של אות השמע.
3. מיסוך ("מאסקינג") הרעש בתחום התדר: נזהה אזורים ספציפיים בתחום תדר הזמן המשויך לרעש ונגדיר את הערכים שלהם לאפס.
4. ISTFT (התמרת פורייה בחלונות הפוכה): נבצע את ה-STFT הפוכה כדי להשיג את אות האודיו ללא הרעש בחזרה למרחב הזמן.
5. נורמליזציה: ננרמל את הסיגנל החדש כדי להבטיח שהוא נמצא בטווח של מספר שלם של 16 ביטים.

היישום כולל שימוש בספריית scipy לקריאת קובץ השמע וחישוב ה-STFT. שלב 3 ו-Inverse STFT מיושמים באמצעות פונקציות stft ו-istft מ-signal-scipy.

מבחינת **היפרפרמטרים וספים**:

קצב דגימה: קצב הדגימה מוגדר ל-4 קילו-הרץ, בהתאם לקבצי האודיו שסופקו. פרמטרים של STFT:

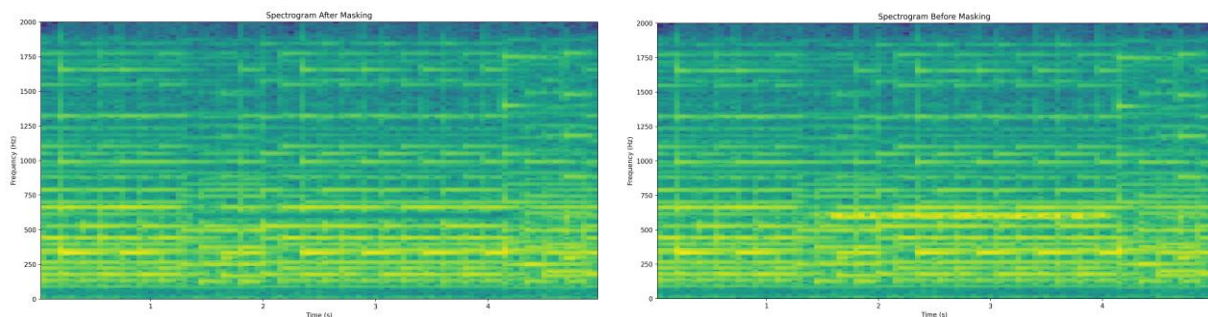
ביצעתי ניסוי וטעייה (וחיפוש באינטרנט) עד אשר הבנתי מה הם הפרמטרים הנדרשים (מפורט על כך בהמשך, באתגרים ופתרונות) מבחינת אתגרים ופתרונות:

בחירת פרמטרי STFT: בחירת פרמטרים מתאימים ל-STFT היא קריטית. האלגוריתם נותן מענה לאתגר זה באמצעות ניסויים ותצפית בתכונות הספקטרוגרמה.

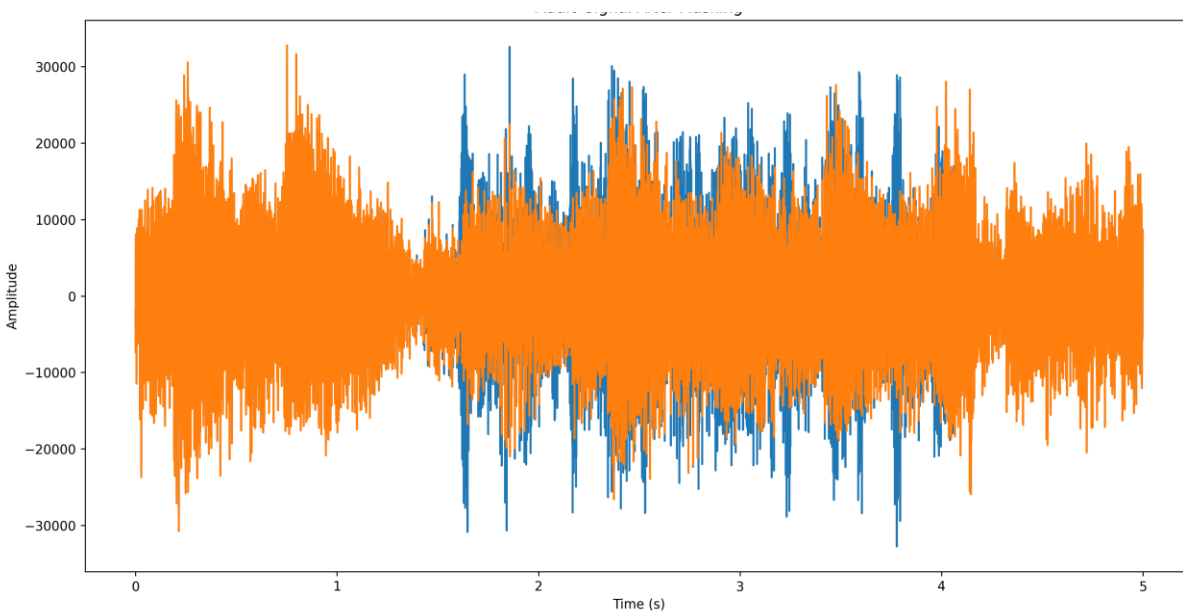
מיסוך תדר הרעש: זיהוי מדויק של האזורים שיש להסוות דורש ניתוח קפדני של הספקטרוגרמה. האלגוריתם בודק חזותית את הספקטרוגרמה כדי לקבוע אזורי מיסוך אופטימליים. זה לקח מעט זמן כי בכל פעם ניסיתי להסיר אזור אחר, וגם גודל המלבן שהסרתי היה נתון לשינוי (התחלתי עם מלבנים גדולים כדי לזהות איפה הרעש, ואז לאט לאט הקטנתי כדי להסיר את הרעש בלבד ולא צלילים נוספים).

אתגר נוסף היה לבחור את גודל החלון. בהקשר של עיבוד אודיו עבור $q2$, הבחירה בגודל החלון עבור טרנספורמציה פורייה קצרה (STFT) היא החלטה מכרעת שמשפיעה באופן משמעותי על הסרת הרעש. גודל החלון שנבחר של 250 לא היה שרירותי אלא תוצאה של שיקול זהיר והתאמה אסטרטגית עם המאפיינים של מערך השמע $q2$.

ראשית, ההחלטה להשתמש בגודל חלון של 250 נובעת מתפקידו כמחלק של המספר הכולל של הדגימות במערך השמע $q2$, בדיוק 2000 דגימות. זה מבטיח שגודל החלון ישלים את המבנה של מערך הנתונים (דהיינו, שלא יהיה חלק באודיו שלא כיסינו עם חלון מסוים), גורם קריטי בשמירה על מהימנות במהלך שלבי העיבוד הבאים, שכן המשמעות של יישור גודל החלון עם המספר הכולל של דגימות מתבהרת בשלב היפוך פורייה בזמן קצר (ISTFT). בנוסף, הבטחת חלוקה במספר דגימות השמע היא שיקול מדויק שמטרתו למנוע הכנסת דגימות מיותרות. תשומת לב זו לפרטים חיונית לשמירה על שלמות אות האודיו והימנעות מחוסר עקביות באורך במהלך תהליך הסרת הרעש.



בגרף הנ"ל רואים את האיפוס שביצענו באיזור הרועש של הספקטרוגרמה. אשמור על כנות עם בודק קובץ זה ואגיד שהעובדה כי אנחנו בוודאות מקבלים קובץ מסויים עזרה לי מאוד, שכן כך יכלתי בעזרת ניסוי וטעייה לבדוק בכל פעם האם התדרים שאיפסתי גרמו לרעש או שסתם הזיקו לקובץ האודיו.



בגרף הנ"ל ניתן לראות את קובץ האודיו ביחס למרחב הזמן. בכחול – קובץ האודיו הישן, בכתום – קובץ האודיו החדש. יש בניהם חפיפה כמעט מושלמת, למעט כמובן רעשי הרקע. ניתן לראות בבירור כי מדובר ברעש לא קבוע – ובטח שלא מחזורי. במושגים שלנו, מחזוריות היא דבר שיכל להקל עלינו, שכן במידה והכל היה סטציונרי, לא היינו צריכים לחלק לחלונות. בקובץ האודיו q2.wav, אנחנו חייבים לחלק לחלונות כדי שנוכל למצוא את הרעש.

מסקנות

לסיכום, האלגוריתמים של ניקוי הרעשים שהוחלו על קבצי האודיו שסופקו הוכיחו דרגות שונות של הצלחה בהפחתת שיבושים לא רצויים. עבור q1, הופעלה גישת הניקוי של תחום תדר על ידי זיהוי והגדרת רכיבי תדר ספציפיים לאפס. התדר הדומיננטי ביותר לפני ניקוי הרעש זוהה ואושר חזותית באמצעות עמודות הספקטרוגרמה. הסיגנל הנקי שהתקבל הראה הפחתה משמעותית בתדר הרעש, ותרם לשיפור הכולל באיכות השמע. מצד שני q2, כללה טכניקת מיסוך תדר באמצעות Short-Time Fourier Transform (STFT). האזור הממוקד בתחום תדר הזמן הוגדר לאפס, ובכך למעשה דיכוינו רעש לא רצוי. ספקטרוגרמות הופעלו כדי לדמיון את ההשפעה של תהליך המיסוך על אות השמע. הגרסה המרופדת הפגינה הפחתה ברכיבי התדר שצוינו, מה שהוביל לפלט שמע נקי ומובן יותר. לאחר מכן לא אשקר כי ביצעתי ניסוי וטעייה והאזנתי לפלט עד אשר קיבלתי פלט נקי. בשני המקרים, האלגוריתמים של ניקוי הרעש השיגו את המטרה של הפחתת אלמנטים משבשים מבלי לפגוע בשלמות תוכן האודיו החיוני. הבחירה באסטרטגייה ובאלגוריתם הושפעה מאופי הרעש הקיים בכל קובץ שמע, מה שמדגיש את החשיבות של התאמת גישות הניקוי למאפייני רעש ספציפיים.

באופן כללי נהייתי לכתוב את התרגיל, למרות שהיה לי מעט קשה בלי ניסיון בnumpy או IML.