

Image Processing - Exercise 2

Agam Hershko, id_214193831, 214193831

מבוא

התרגיל עוסק בעיבוד וניתוח של קבצי אודיו והוא מחולק לשלושה חלקים שונים. בחלק הראשון, הייתי צריך להוסיף Watermarks (אסביר את מטרתם בחלק המתאים) לקובץ אודיו בשתי דרכים בלתי תלויות- בדרך שתהיה ניתנת לזיהוי ובדרך שלא תהיה ניתנת לזיהוי. בחלק השני, הייתי צריך למיין תשעה קבצי אודיו לשלוש קבוצות שמייצגות Watermarks שונים ולזהות את הפונקציות שיצרו אותם.

בחלק השלישי, קיבלתי שני קבצי אודיו שעברו האצה/האטה- אחד מהם במרחב הזמן ואחד מהם במרחב התדר. כתוצאה מכך, משך הקטע השתנה בהתאם וחלק מהמאפיינים שלו (למשל שינוי בתדרים). המטרה שלנו בחלק זה היא לזהות מהו המרחב שבו כל קובץ אודיו עבר שינוי ומה הוא קבוע ההאצה/ההאטה שהשתמשו בו בשני הקבצים (מדובר באותו קבוע).

הטכניקות העיקריות מהחומר הנלמד שיישמתי בתרגיל זה הן ספקטרוגרמות והתמרת פורייה. השתמשתי בספקטרוגרמות כדי להציג בצורה ויזואלית את מרחבי התדר, הזמן ואת האמפליטודה של קבצי האודיו (המייצגים אות דיגיטלי של סאונד). זה יעזור בניתוח מידע ובהמחשת שינויים שנבצע במרחב התדר.

כמו כן, בחלקים 1-2 השתמשתי בהתמרת פורייה כדי למצוא את התדרים של אות הקול ובכך גם יכולתי להפעיל מניפולציות על מרחב התדר (למשל הוספת Watermarks). שמתי לב שבכל חלק משניהם קיים לפחות קובץ אודיו אחד שהאות בו הוא לא סטציונרי, כלומר שהתדר שלו לא מחזורי בזמן. לכן, בחרתי להפעיל בחלקים אלה STFT בהתאם לשיטה שלמדנו למקרה זה. כך יתאפשר לנו לבצע התמרת פורייה בסגמנטים שונים של האות שבכל אחד מהם הוא סטציונרי.

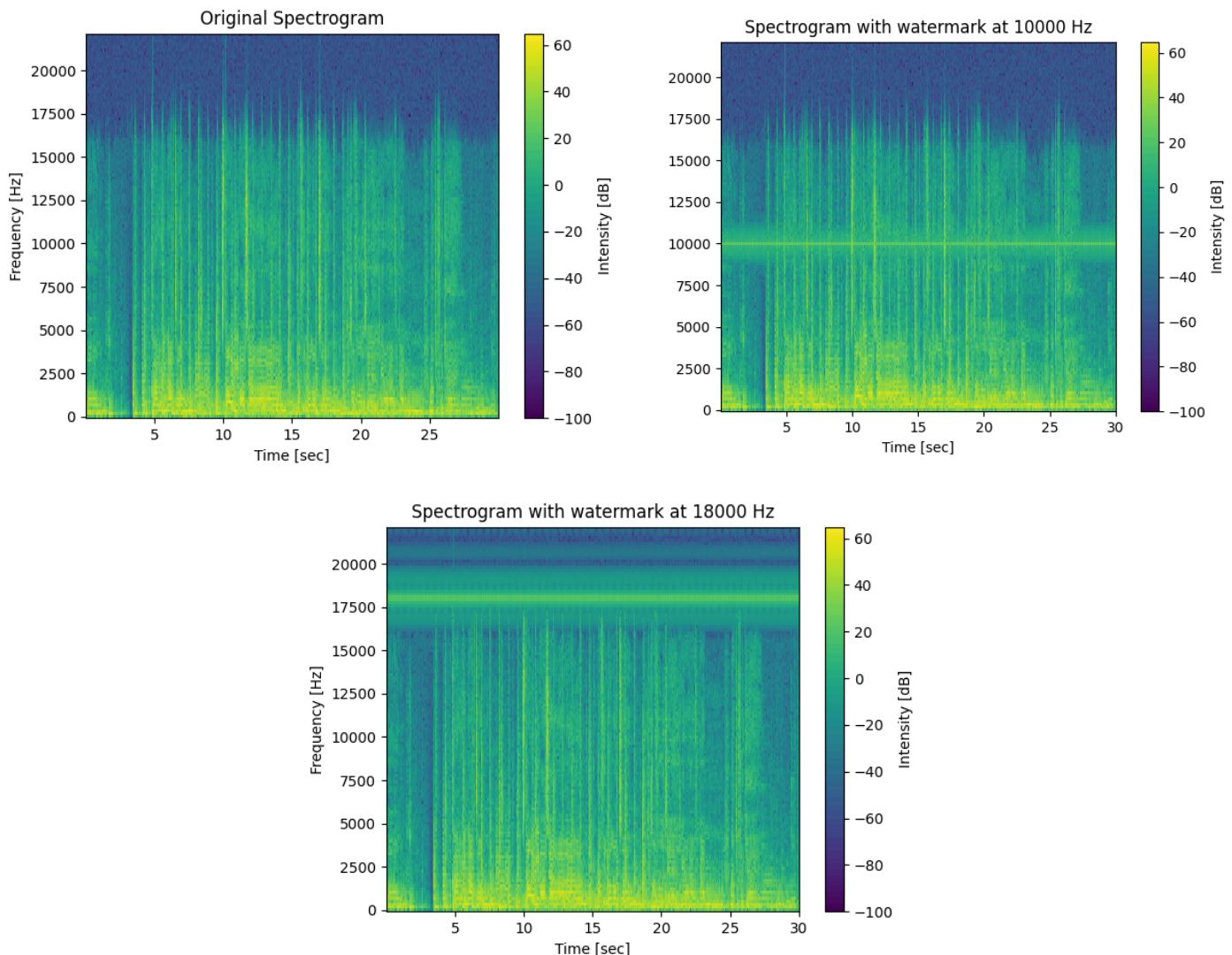
הוספת Watermarks

במהלך החלק הזה של התרגיל, הוספתי לקבצי האודיו watermarks. Watermarks הם מזהים דיגיטליים יחודיים ששולבו בקובץ האודיו ונועדו בעיקר למטרות של זכויות יוצרים. הם מתחלקים לשני סוגים של watermarks שהוספתי לקובץ באופן בלתי תלוי (יצרתי קבצים חדשים), good watermark, שלא ניתן להבחנה בשמיעה אנושית ו bad watermark שניתן לשמוע אותו בצורה בולטת.

להלן אתאר את האלגוריתם ליצירת watermarks. תחילה נקרא את הקובץ אודיו לתוך רשימה על ידי שימוש בפקודה wavefile.read. לאחר מכן, נקרא בנפרד לפונקציה שמוסיפה כל אחד מסוגי watermarks, הנבדלים על ידי התדר שבו נשנה את המגניטודה. כיוון שהאות לא סטציונרי, נפעיל התמרת stft על ידי הפקודה scipy.signal.stft ונקבל רשימה של תדרים. נמצא את האינדקס של התדר המתאים, בכך שנחזיר את הארגומנט שייתן את ההפרש המינימלי בערך מוחלט בין התדירות לתדירות

המתבקשת. נוסף למגניטודה בערך התדירויות המבוקש ערך של 200 ובכך ניצור watermark שיהיה ניתן להבחין בצורה בולטת בספקטרוגרמה. לאחר מכן, נבצע את התמרת פורייה ההפוכה על istft על המגניטודה באמצעות `scipy.signal.istft`. לבסוף, נשמור את התוצאה לקובץ אודיו באמצעות `wavefile.write`.

בחרתי בתדר 10000 הרץ בשביל להוסיף רעש (הגדלת מגניטודה) באופן שיהווה bad watermark, תדר זה מצוי בטווח השמיעה האנושי והקובץ החדש נשמע כמו הקודם רק עם זמזום בולט בזמן ההאזנה. בחרתי בתדר 18000 הרץ בשביל להוסיף רעש (הגדלת מגניטודה) באופן שיהווה good watermark, תדר זה הוא בקצה תחום השמיעה האנושי שמגיע עד ל 20000 הרץ ולכן לא ניתן להבחין בו בשמיעה.



תחילה ניסיתי לבחור תדר נמוך בשביל הוספת good watermark באופן זהה עם מגניטודה מתאימה, אך לא הצלחתי לבחור תדר מתאים כך שעדיין לא יהיה לשמוע שינוי באודיו אבל עדיין יהיה ניתן להבחין בשינוי ניכר בספקטרוגרמה (באופן שיעזור לזהות את ה watermark). כדי להוסיף Watermark בצורה טובה לתמונות, ניתן להפעיל על התמונה טרנספורם פורייה דו-ממדי (2D DFT), להטמיע את

ה-Watermark בתדרים מסוימים על ידי שינוי מקדמי התדר, ולאחר מכן להחזיר את התמונה באמצעות טרנספורם פורייה הפוך.

סיווג Watermarks

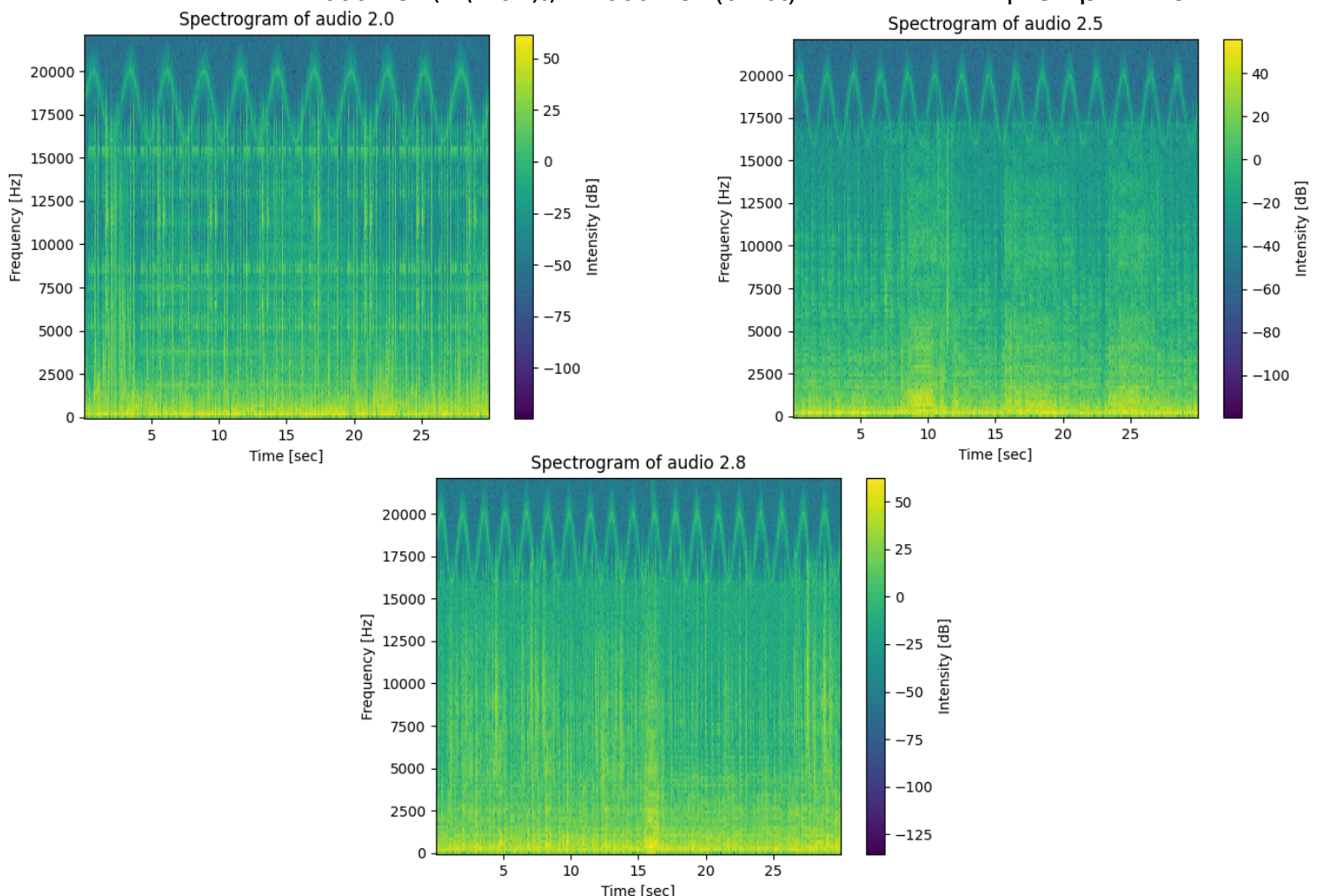
תחילה יצרתי ספקטרוגרמות עבור כל אודיו ושמתי לב שניתן להבחין בwatermark באופן בולט (אך לא ניתן לזהות בשמיעה)- מדובר בגל סינוסי (הזנחתי את הפאזה) באמפליטודה קבועה של 2000 (במרחב התדר- כאשר המקסימום נמצא ב20000 הרץ והמינימום נמצא ב16000 הרץ) עם זמן מחזור משתנה. למעשה ניתן להבחין בשלוש פונקציות סינוס עם שלושה זמני מחזור שונים.

הגישה הכללית שלי הייתה להפעיל על קובץ האודיו התמרת stft כיוון שחלק מהאותות אינם סטציונריים כדי למצוא את התדרים המתאימים לאודיו. לאחר מכן, מצאתי את הזמנים המתאימים לתדר 20000 הרץ וחילצתי מהם את זמן המחזור של הסינוס. בכך נוכל לסווג את הקבצים לקבוצות שונות ולמצוא את הפונקציות המתאימות לכל watermarks. חשוב לציין שהשתמשתי בכל שלושת קבצי האודיו בכל קבוצה כדי לחשב את זמן המחזור. להלן תוצג החלוקה לקבוצות והפונקציות המתאימות לכל watermark, שילוו בספקטרוגרמה מייצגת מכל קבוצה (פרטי המימוש יוצגו בהמשך).

הקבוצה הראשונה מתאימה לשלושת קבצי האודיו הראשונים (2.0-2.2), זמן המחזור בה הוא 2.16 שניות ולכן הפונקציה המתאימה היא $f = 2000 * \sin(1/(2.16)t) = 2000 * \sin(0.46t)$.

הקבוצה השנייה מתאימה לשלושה השניים של קבצי האודיו (2.3-2.5), זמן המחזור בה הוא 1.65 שניות ולכן הפונקציה המתאימה היא $f = 2000 * \sin(1/(1.65)t) = 2000 * \sin(0.61t)$.

הקבוצה השלישית מתאימה לשלושת קבצי האודיו הראשונים (2.6-2.8), זמן המחזור בה הוא 1.32 שניות ולכן הפונקציה המתאימה היא $f = 2000 * \sin(1/(1.32)t) = 2000 * \sin(0.76t)$.



בהמשך למה שהסברתי לעיל, כעת אעמיק את ההסבר בפרטי מימוש האלגוריתם. הספריות שהשתמשתי בהן הן `scipy` בשביל קריאה וכתובה של קבצי אודיו (`scipy.io`) והפעלת פעולות כגון פורייה על אותות אלה (`scipy.signal`) וכמובן `numpy` בשביל פעולות מתמטיות בין היתר. כמו כן, השתמשתי ב `matplotlib.pyplot` בשביל ספקטרוגרמות, אותן יצרתי באמצעות `scipy.signal.spectrogram`. בהתחלה השתמשתי בפקודה `wavefile.read` מ `scipy.io` כדי לקרוא את הנתב של קובץ האודיו לתוך רשימה `audio` וכדי לקבל את תדר הדגימה `sample_rate`. לאחר מכן נשתמש בפקודה `stft` מספריית `scipy.signal` כדי למצוא את התדרים של קובץ האודיו `f`, את הזמנים `t` ואת המגניטודה `sxx`. נמצא את האינדקסים שעבורם התדר הוא 20kHz ע"י מציאת הארגומנטים עבורם ההפרש בערך מוחלט מהתדר הרצוי הוא מינימלי. בחרתי בתדר זה כיוון שהוא מהווה ערך מקסימלי של פונקציית הסינוס ולכן מופיע פעם יחידה בכל מחזור (כך בתיאוריה, בפועל הוא מופיע יותר ולכן אפרט בהמשך איך התמודדתי עם כך). עבור האינדקסים הנתונים, נמצא את המגניטודות המתאימות (בערך מוחלט).

כדי למצוא את זמני המחזור של הסינוס בתדר, נרצה לקחת את התדרים שבהם המגניטודה גדולה יותר מערך סף על מנת שלא יתקבלו ערכים בלתי רצויים. רציתי שערך הסף יהיה חלק יחסי של המגניטודה המקסימלית, כלומר הערך המקסימלי כפול קבוע. הקבוע שבחרתי הוא 0.11 כיוון שהוא נתן לי את התוצאות המתאימות ביותר עבור החלוקה הברורה לקבוצות שהיה ניתן להבחין בה בספקטרוגרמות, שינוי קטן בערך זה שינה לי את התוצאות ושינוי גדול אשר הגדיל מדי את ערך הסף גרם לבעיות בתוכנית כיוון שלא היו מגניטודות אשר גדולות ממנו.

לאחר מכן, מימשתי פונקציית עזר שמחשבת את זמן המחזור של הפונקציה. פונקציה זאת פועלת באופן הבא: היא מחשבת את הפרש הזמנים בין כל דגימה שנייה של הזמן (בתיאוריה כיוון ש20kHz הוא המקסימום, היה אמור להתקבל ערך יחיד אבל בפועל זה לא קרה- לדעתי זה קשור לקוונטיזציה של התדרים בזמן או לחוסר דיוק בתדר המקסימלי למרות שגם כאשר ניסיתי לשנות את התדר התוצאות לא השתפרו) ואם ההפרש גדול מערך סף מסוים (אני בחרתי 1- כדי להימנע מזמנים סמוכים עקב דגימה כל פרק זמן קטן), היא מוסיפה אותם לרשימה של הפרשים. לבסוף היא מחזירה את ממוצע ההפרשים ובכך חישבתי זמן מחזור.

לבסוף, עבור כל קובץ אודיו נעבור על כל הקבוצות ונוסיף את המספר של הקטע קול ונעדכן את זמן המחזור של הקבוצה (על ידי חישוב ממוצע מתאים) אם הוא מתאים לקבוצה. התאמה לקבוצה מוגדרת על ידי הפרש בערך מוחלט בין זמן המחזור של הקבוצה לזמן המחזור של קטע הקול שקטן שקטן ב 10% (ערך סף) מזמן המחזור של הקבוצה. אם לא מתקבלת התאמה לאף קבוצה, ניצור קבוצה חדשה.

כדי להסיר Watermark מקובץ קול, יש לנתח את הספקטרוגרמה שלו, לזהות את תחום התדרים של ה-Watermark (נשתמש ב `stft`) ולהחליש את העוצמה שלו תוך שמירה על תדרים קרובים. התהליך דורש כוונן מדויק כדי למנוע פגיעה באיכות האודיו המקורי, במיוחד אם תדרים קרובים ל-Watermark שייכים לאות המקורי.

זיהוי שיטת האצה

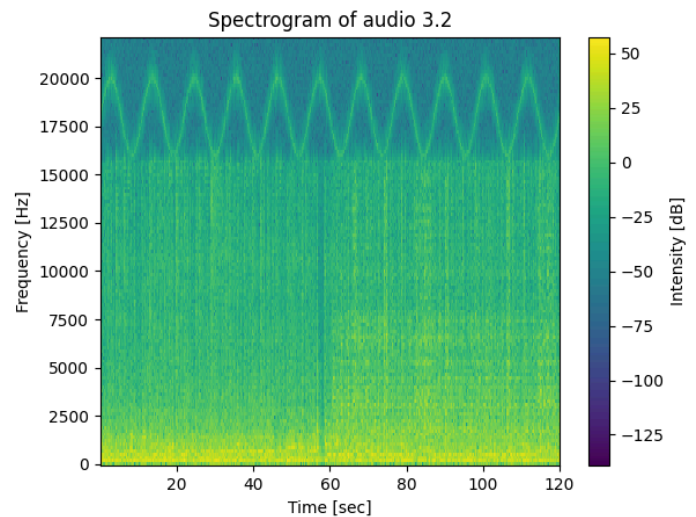
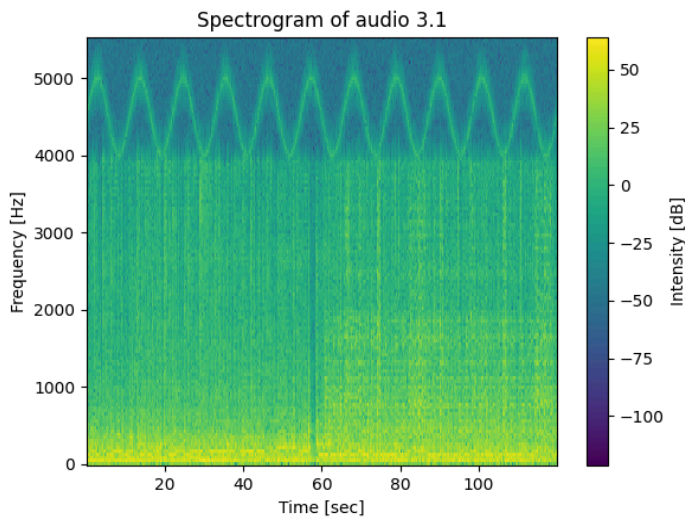
כעת ננתח שני קבצים שעברו האצה (או האטה), אחד מהם במרחב התדר ואחד מהם במרחב הזמן. בעת האזנה לשני הקטעים, ניתן להבחין שהקטע הראשון הוא קטע שעבר מניפולציה והתדרים בו נמוכים. זאת לעומת הקטע השני, שנשמע סטנדרטי ולא ניתן להבחין בו במשהו מיוחד.

אטען שהקטע הראשון עבר האטה במרחב הזמן (או האצה בקבוע שקטן מ1 ושווה לאחד חלקי קבוע ההאטה) והקטע השני עבר האצה במרחב התדר. נשים לב שהאטה בקבוע x פירושה היא שמשך הקטע התארך פי x . לפי מה שלמדנו בהרצאה, כאשר נעשה האצה במרחב הזמן (על ידי דגימה של חלק מהנתונים), טווח התדרים יגדל בהתאם וגובה הקול יגדל כתוצאה מכך וכן האמפליטודה תקטן. לכן, כאן נקבל את המקרה הנגדי בו אות הקול הואט (על ידי ריפוד באפסים למשל), טווח התדרים (וכתוצאה מכך גובה הצליל) קטן והאמפליטודה גדלה. נבסס זאת לפי הנוסחה ל `scaling` בזמן כמו שלמדנו בתרגול:

$$\text{Scaling: if } f(x) \xrightarrow{\text{Fourier}} F(u)$$

$$\text{then } f(ax) \xrightarrow{\text{Fourier}} \frac{1}{|a|} \cdot F\left(\frac{u}{a}\right)$$

A הוא קבוע ההאצה (שקטן מ1 ושווה לאחד חלקי קבוע ההאטה x). הצבתו מתאימה לכך שקטע הקול הוארך בזמן, התכווץ בטווח התדרים שלו והאמפליטודה שלו גדלה. לעומת זאת, למדנו שהאצה במרחב התדר (על ידי דגימה של תדרים/הספקטרוגרמה) תקצר את משך האודיו מבלי לשנות את מאפייניו. בהתאמה, האטה במרחב התדר תאריך את משך האודיו מבלי לשנות את מאפייניו. כעת אציג את הספקטרוגרמות ואסיק מהן מסקנות.



ניתן להבחין שטווח התדרים בספקטרוגרמה הראשונה קצר יותר מטווח התדרים של הספקטרוגרמה השנייה כמו שצפינו. בנוסף לכך, אמנם קשה להבחין בכך אך גם שינוי האפליטודה בספקטרוגרמה (dB) מתאים למרות שזה לא המאפיין העיקרי וניתן שלא להתייחס לכך. ניתן להבחין בwatermark בספקטרוגרמה השנייה שהפונקציה שלו זהה לאחת הפונקציות בחלק הקודם של התרגיל, מדובר בפונקציית סינוס עם אמפליטודה של 2kHz שהמקסימום שלו מתקבל ב20kHz והמינימום שלו מתקבל ב16kHz. קיים watermark דומה בספקטרוגרמה הראשונה והוא פונקציית סינוס עם אמפליטודה של 0.5kHz כאשר המקסימום מתקבל ב5kHz והמינימום מתקבל ב4kHz. נמצא את קבוע ההאטה x על ידי השוואת האמפליטודה בתדר (לא קשור לcolorbar שיחידותיו dB) של watermark בזמן. בקטע השני, מרחב התדרים לא השתנה ובקטע הראשון הוא קטן פי x (קבוע ההאטה) מהקטע השני. לכן נסיק מהאמפליטודות שציינתי שקבוע ההאטה הוא $x=4$.

מסקנות

1. ניתוח ספקטרוגרמות והפעלת STFT מאפשר זיהוי והתמצאות במרחב התדר של האודיו, ומשמש כלי מרכזי לסיווג ותכנון מחדש של התדרים.
2. זיהינו בתרגיל את המורכבות הנדרשת בטיפול באותות לא-סטציונריים ולכן השתמשנו בSTFT.
3. ההבנה כיצד מרחב הזמן ומרחב התדר משפיעים זה על זה אפשרה זיהוי ושינוי מדויק של קבצי אודיו שעברו האצה או האטה.
4. הוספת Watermarks בקבצי אודיו הדגיש את החשיבות של איזון בין שמירה על איכות האודיו לעמידות בפני גילוי.