The Andrew and Erna Viterbi Faculty of ELECTRICAL & COMPUTER ENGINEERING

SIPL SIGNAL AND IMAGE PROCESSING LAB

TECHNION Israel Institute of Technology

דו"ח סיכום פרויקט: ב'

קריוקי למידה עמוקה

Karaoke Maker

מבצע:

איזאר חסון **Izar Hasson**

מנחה: **הדס אופיר** **Hadas Ofir**

סמסטר רישום: חורף תשפ"ג

תאריך הגשה: פברואר, 2025

P 7065-1-25

אני מקדיש דו"ח מסכם זה למשפחתי האהובה.

תודות

אני מודה מקרב לב למנחה היקרה הדס אופיר, עזרה ותמיכה והבנה מלאה לכל אורך הפרויקט.

אני מודה לארטיום בודובסקי ורועי מיטרני שעזרו במהירות ודחיפות ובטוב לב בכל תקלה שהייתה.

אני מודה גם ליאיר משה שכתב את תבנית זו לדו"ח מסכם ובכך עזר לי לכתוב דו"ח מסכם בהצלחה ☺

תוכן עניינים

רשימת איורים

תקציר

הפרויקט עוסק בפיתוח מודל מבוסס למידה עמוקה להפרדת אודיו, תחום המספק פתרונות טכנולוגיים מתקדמים להפרדת מקורות קול שונים מתוך מיקס אודיו. המטרה המרכזית של הפרויקט היא ליצור מודל המסוגל להפריד את הרכיבים השונים של שיר (כגון שירה, תופים, ובס) תוך שימוש במידע ספקטרלי ובטכניקות מתקדמות כמו טרנספורמרים ולמידה רציפה.

במהלך הפרויקט הותאם המודל Open-Unmix [1] לשימוש במבני רשת חדשים, כולל שילוב של שכבות קונבולוציה וטרנספורמרים לשיפור הדיוק והביצועים. האתגרים המרכזיים כללו מציאה ובנייה של הארכיטקטורות המתאימות, אופטימיזציה של זיכרון GPU , עבודה עם מסדי נתונים מורכבים והתאמת המודל לעבודה על קטעי אודיו באורכים משתנים.

ההישגים המרכזיים הם שילוב מוצלח של מבני רשת מתקדמים, שיפור ביצועים על ידי התאמת היפר-פרמטרים ותכנון אופטימלי של רשתות עצביו והתמודדות עם מסדי נתונים מורכבים ואופטימיזציה של תהליך האימון.

המודל שהתקבל מציג ביצועים מרשימים, והפרויקט תורם להבנת טכנולוגיות חדשניות בתחום עיבוד האודיו ומדגים את הפוטנציאל של למידה עמוקה ליצירת פתרונות חכמים ומדויקים.

Abstract

The project focuses on developing an advanced deep learning model for audio source separation, providing innovative solutions for isolating different sound components from an audio mix. The primary goal of the project is to create a model capable of separating the various elements of a song (e.g., vocals, drums, and bass) using spectral information and advanced techniques such as transformers and sequential learning.

As part of the project, the Open-Unmix model was adapted to include new network architectures, incorporating convolutional layers and transformers to enhance accuracy and performance. The project tackled significant technical challenges, including identifying suitable architectures, optimizing GPU memory, working with complex datasets, and adapting the model to handle audio segments of varying lengths.

Key achievements include Successful integration of advanced network structures, Improved model performance through hyperparameter tuning and optimal neural network design and Effective handling of complex datasets and optimization of the training process.

The developed model demonstrates impressive performance, contributing to the understanding of innovative technologies in audio processing and showcasing the vast potential of deep learning to create intelligent and precise solutions.

1. מבוא

1.1. רקע כללי

בעידן הנוכחי, בו טכנולוגיות למידה עמוקה מכתיבות את קצב הפיתוח בתחומים מגוונים, תחום עיבוד האודיו הפך להיות אחד מהתחומים החדשניים ביותר. יישומים כגון זיהוי דיבור, סינתזה קולית, והפרדת מקורות קול הפכו להיות קריטיים בתעשיות רבות, לרבות תעשיית המוזיקה, בידור, וחוויות משתמש במערכות קוליות. הפרדת אודיו (Audio Source Separation) מתמקדת במשימה של פירוק אות מיקס למרכיבים הבודדים שלו, כמו שירה, כלי נגינה ודיבור. משימה זו מהווה אתגר בשל המורכבות הספקטרלית של אותות קוליים, חפיפה בין תדרים, רעשים חיצוניים, איכות ההקלטה והגיוון הרב שיש במוזיקה.

1.2. מטרת הפרויקט

מטרת הפרויקט היא לפתח מודל מבוסס למידה עמוקה המסוגל להפריד רכיבי אודיו באופן מדויק, תוך שיפור ביצועים קיימים בשיטות אחרות. הפרויקט מתמקד בשילוב טכניקות חדשניות, כגון שימוש בטרנספורמרים ומודלים היברידיים הכוללים שכבות קונבולוציה, על מנת להתמודד עם מורכבות המשימה.

באמצעות מודל זה, ניתן לשפר את האופן בו מנותחים אותות אודיו וליצור כלים מתקדמים עבור מפיקים מוזיקליים, אפליקציות חכמות, ואמצעים טכנולוגיים המיועדים לשיפור נגישות לקהל רחב.

1.3. מבנה הפרויקט

הפרויקט מחולק למספר שלבים:

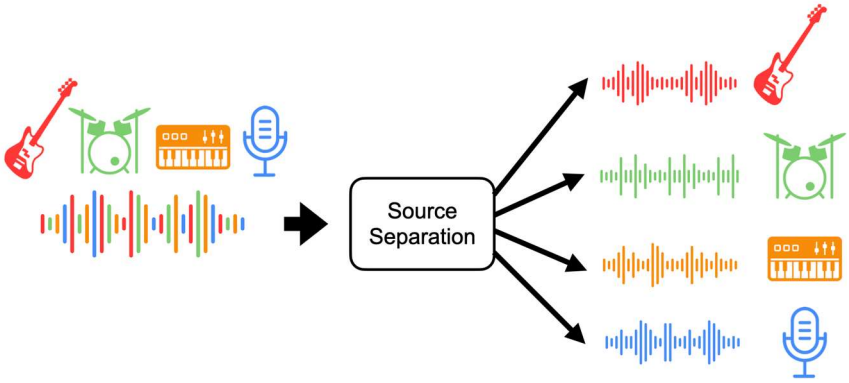
- בשלב הראשון בוצע סקר ספרות על שיטות קיימות להפרדת אודיו, תוך התמקדות במודלים כמו Open-Unmix [1], Hybrid Demucs [2], Band split RNN [3] ועוד.
- בשלב השני הוגדרו מבנים חדשים למודל, תוך שילוב שכבות מתקדמות כמו שכבות LSTM, קונבולוציה.
- בשלב השלישי הוטמעו השינויים במודל והוגדרו פרמטרים אופטימליים.
- בשלב האחרון בוצע אימון המודל, ולאחר מכן נבחנו הביצועים.

בפרקים הבאים נעמיק בסקר הספרות, נפרט את תהליך הפיתוח של המודל, ונציג את תוצאותיו.

2. סקר ספרות

2.1. סקירה כללית על הפרדת שמע

הפרדת שמע היא תחום חשוב בעיבוד אותות, שמטרתו לפצל אותות קוליים מעורבים לרכיבים נפרדים. משימה זו נפוצה בתעשיית המוזיקה, פיתוח מערכות חכמות, ושיפור חוויות משתמש במערכות אודיו. לדוגמה, כמתואר באיור 1:: הפרדת שירה מכלים מוזיקליים, מאפשרת יצירת פלייבקים, רמיקסים ושיפור איכות הקלטות.



איור 1: הפרדת מקורות בשיר

שיטות קלאסיות להפרדת אודיו, כמו Independent Component Analysis (ICA) ו - Non-negative Matrix Factorization (NMF) מתמקדות בשימוש במודלים מתמטיים ואופטימיזציה. עם זאת, מגבלותיהן טמונות ביכולתן להתמודד עם תדרים חופפים ומערכות רועשות.

2.2. שימוש בלמידה עמוקה להפרדת שמע

בשנים האחרונות, למידה עמוקה שינתה את אופן ההתמודדות עם בעיית הפרדת האודיו. מודלים נוירוניים מסוגלים ללמוד מאפיינים מורכבים של אותות קוליים מתוך מסדי נתונים גדולים, ולהשיג תוצאות מדויקות יותר משיטות קלאסיות.

דוגמאות לשיטות מבוססות למידה עמוקה כוללות:

Open-Unmix: מודל פופולרי המשלב שכבות LSTM לעיבוד מידע ספקטרלי, תוך שימוש בידע על זמן ותדר.

Spleeter: כלי מבוסס ספריית TensorFlow המספק פתרונות מהירים להפרדת אודיו באמצעות מודלים מתוכננים מראש.

Hybrid Transformer Demucs: מודל מתקדם המשלב טרנספורמרים לעיבוד תלות גלובלית וקונבולוציה לאחיזת פרטים מקומיים.

2.3. אתגרים בהפרדת אותות שמע

האתגרים המרכזיים בתחום כוללים:

- **מורכבות ספקטרלית:** מקורות האודיו מכילים לעיתים קרובות רכיבים ספקטרליים חופפים, מה שמקשה על זיהוי ברור של הרכיבים השייכים לכל מקור בנפרד. תדרים משותפים בין כלי נגינה או קולות שונים דורשים יכולת עיבוד גבוהה לזיהוי דפוסים דקים באות.
- **שונות במקורות האודיו:** תנאי הקלט של האודיו משתנים בהתאם לאיכות ההקלטה, לסוגי הכלים או הקולות, ולמאפיינים הייחודיים של כל מקור. שונות זו מצריכה מהמודל יכולת הכללה רחבה להתמודדות עם מגוון רחב של תרחישים.
- **רעשי רקע:** סיגנלים מעורבים כוללים לעיתים קרובות רעשי סביבה אשר אינם שייכים למקורות הרצויים, רעשים אלו מקשים על תהליך הבידוד ועלולים להוביל לתוצאות לא מדויקות.

בפרויקט, התמודדנו עם אתגרים אלו באמצעות שימוש במודולים חדשניים כגון Band-Split RNN , המאפשרים עיבוד מדויק ויעיל של הרכיבים הספקטרליים ונטרול משמעותי של רעשי רקע באופן מותאם.

3. פתרון מוצע

הפרק הבא יתאר את הפתרון שפותח במסגרת הפרויקט, כולל מבנה המודל, תהליך האימון, ושיטות האופטימיזציה שנבחרו. המודל משלב גישות חדשניות להפרדת אודיו באמצעות למידה עמוקה, תוך שימוש במודולים כגון Band-Split RNN ומודול Mask Estimation להתמודדות עם האתגרים שתוארו בפרק הקודם.

3.1. מאגר הנתונים: MUSDB18

המאגר [4] מכיל 150 קטעי מוזיקה באורך של כ-10 דקות כל אחד, המחולקים לז'אנרים שונים כמו רוק, פופ, ג'אז ועוד.

המידע בו מחולק לתתי-קטגוריות, כך שניתן לאמן מודל להפריד בין המרכיבים השונים של מיקס האודיו.

אתגרים במאגר:

- **גיוון מוזיקלי** :המאגר כולל שירים בז'אנרים שונים, מה שמצריך יכולת הסתגלות של המודל למגוון רחב של סגנונות.
- **קולות חופפים** :לעיתים, תדרים שונים של רכיבים מוזיקליים חופפים זה לזה, מה שמקשה על תהליך ההפרדה.

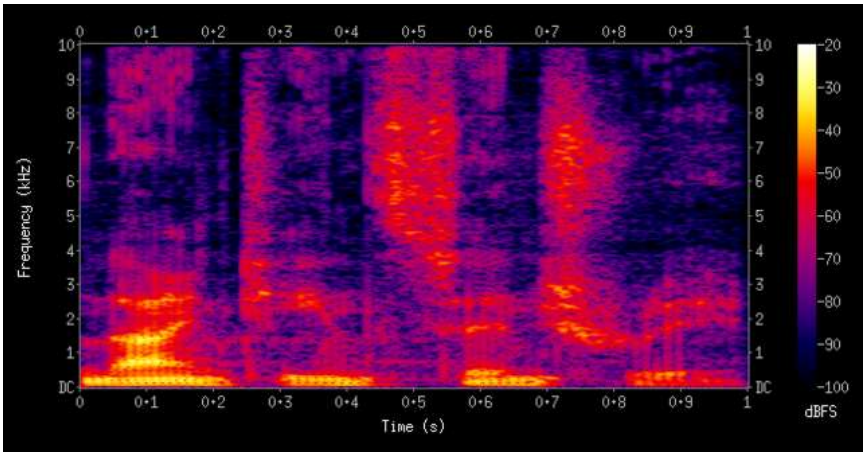
- **מספר נמוך של קטעים**: 100 קטעים לאימון, 50 קטעים לבדיקה בלבד, מספרים נמוכים לבעיה מורכבת זו.

3.2. מבנה המודל

מודל Open-Unmix שהותאם בפרויקט משלב ארכיטקטורות מבוססות RNN ו FC, המאורגנות במודולים עיקריים:

1. קלט המודל

הקלט למודל הוא ייצוג ספקטרלי של אות האודיו כפי שרואים באיור 2: ייצוג ספקטרלי של אות, שמתקבל לאחר ביצוע Short-Time Fourier Transform (STFT).



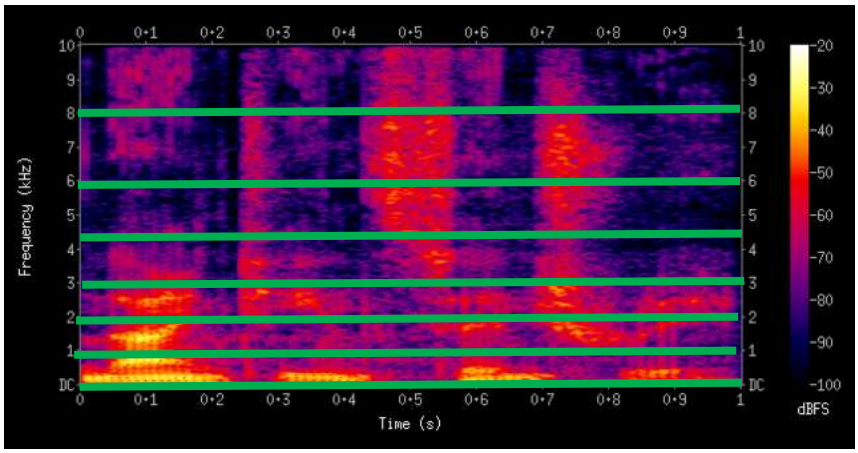
איור 2: ייצוג ספקטרלי של אות שמע.

- **צורת הקלט :**טנזור בגודל [B,C,F,T] שבו :
  - B: מספר הדגימות (Batch Size) .
  - C: מספר הערוצים (לרוב 2 עבור אודיו סטריאו).
  - F: מספר תדרים (bins), שנגזר מגודל ה-FFT.
  - T: מספר המסגרות (frames) בזמן.

הקלט מנורמל כך שממוצע הספקטרום יהיה אפס, וסטיית התקן תהיה 1, כדי לשפר את יציבות האימון.

2. מודל Band Split :

מודל זה מחלק את הספקטרום לתת-תחומים תדריים (**Subbands**) , בדומה לדוגמה באיור 3: דוגמה לחלוקת תדרים.



איור 3: דוגמה לחלוקת תדרים

- **מטרות:**

1. **התאמה לאופי המידע הספקטרלי**:
   - בתחומים נמוכים יש צורך בדיוק רב יותר, ולכן הרצועות שם צרות יותר.
   - בתדרים גבוהים, שבהם יש פחות מידע משמעותי, ניתן להרחיב את הרצועות.
2. **שיפור עיבוד מקבילי**:
   - החלוקה לרצועות מאפשרת עיבוד עצמאי של כל תחום תדר, מה שמפחית את התלות בין רכיבי הקול.
3. **יעילות חישובית**:
   - רוחב הרצועות מותאם כך שצפיפות המידע הספקטרלי בכל תחום תדר תישאר עקבית, מה שמפחית עומס על המודל.
4. **שמירה על דיוק בהפרדה**:
   - בתחומים נמוכים ובינוניים, שבהם יש חפיפה גבוהה בין מקורות, ישנו צורך ברצועות צרות ומדויקות יותר.

- **מימוש :**
  - הספקטרום מחולק לתחומים מוגדרים מראש, כאשר כל תחום מייצג רצועת תדר מסוימת.
  - כל תת-תחום עובר נרמול והתאמה מקומית (LayerNorm).
  - **פלט :**טנזור בגודל [B,K,D,T] , שבו:

    K הוא מספר תתי-התחומים ו D-הוא ממד פיצ'רים קבוע עבור כל תת-תחום.

    מבנה החלוקה שהשתמשנו בו:

1. **0–1000Hz רוחב רצועה 100Hz:**
   - מייצגת את התחום הנמוך ביותר.
   - תחום קריטי הכולל שירה, כלי נגינה מרכזיים, והמידע העיקרי של מרבית האודיו
   - סך הכל 10 רצועות
2. **1000Hz–4000Hz רוחב רצועה 250Hz:**
   - תחום קריטי הכולל שירה, כלי נגינה מרכזיים, והמידע העיקרי של מרבית האודיו.
   - הרוחב (250Hz) משקף את הצפיפות הגבוהה של מידע בתדרים אלו.
   - סך הכל 12 רצועות
3. **4000Hz–8000Hz רוחב רצועה 500Hz:**
   - תחום הכולל תדרים גבוהים יותר כמו גיטרות, אפקטים מוזיקליים, ורכיבים חדים יותר של השירה.
   - הרוחב הרחב יותר (500Hz) מאפשר לכלול את המידע בצורה מרוכזת.
   - סך הכל 8 רצועות
4. **8000Hz–16000Hz רוחב רצועה 1000Hz:**
   - תדרים גבוהים יותר שמיוצגים על ידי כלי נגינה חדים (כמו מצילות) ורכיבים רקעיים של שירה.
   - הרוחב כאן גדל אף יותר, כדי להתמודד עם צמצום המידע בתדרים גבוהים.
   - סך הכל 8 רצועות
5. **16000Hz–20000Hz רוחב רצועה 2000Hz:**
   - התחום הגבוה ביותר, כולל צלילים חדים מאוד.
   - רוחב הרצועה הוא הרחב ביותר ,(2000Hz) כיוון שתדרים אלו מכילים פחות מידע משמעותי, אך נדרשים עבור תחושת צליל מלאה.
   - סך הכל 2 רצועות

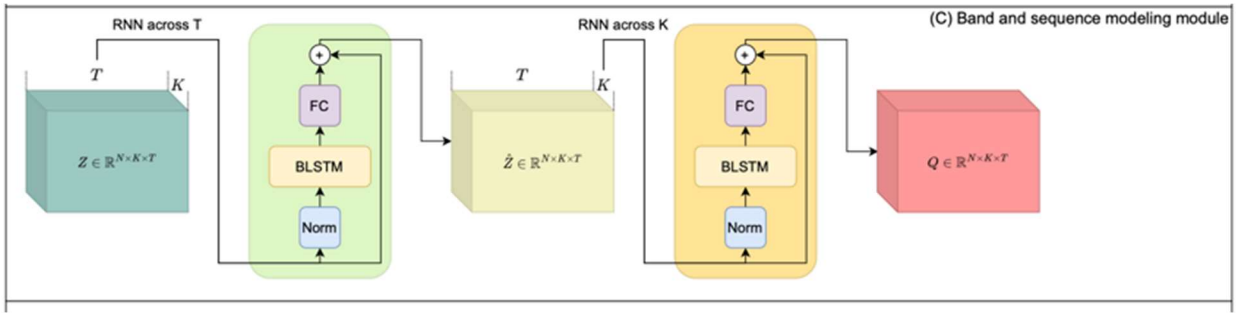
3. Band Sequence RNN:

מודול ה-BandSequence הוא חלק מרכזי במודל שמטרתו לעבד את המידע המתקבל לאחר חלוקת התדרים (BandSplit). המודול משתמש בשכבות RNN , ובפרט- BiLSTMs, כדי לנתח את הדינמיקה של האודיו בשני ממדים:

1. **ממד הזמן (Time Dimension)**
2. **ממד רצועות התדר (Subband Dimension)**

המטרה היא ללמוד קשרים ותלויות הן לאורך הזמן והן בין רצועות התדר השונות כדי לשפר את הביצועים בהפרדת האודיו.

מבנה ה-Band Sequence RNN מודול



איור 4: מבנה מודל ה *Band Sequence RNN*

1. RNNModule

בירוק וכתום באיור 4: מבנה מודל ה Band Sequence RNN, תת-מודול זה מיישם:

- **Group Normalization**: משמש לנרמול המידע בממד ה-Input Features.
- **BiLSTM (RNN) [5]:** עיבוד המידע באופן דו-כיווני (קדימה ואחורה) כדי ללכוד תלות גלובלית.
- **Fully Connected Layer:** מתפקדת כ"שכבת הפלט" שמחזירה את המידע לגודל המקורי (Input Dimension).

שלבי החישוב:

- **קלט** : טנסור בממדים [B, K, T, N] כאשר :
  - B: גודל המיני-באטץ'.
  - K: מספר רצועות תדר (Subbands).
  - T: מספר פריימים בזמן.
  - N: גודל התכונות (Features).
- **חישוב נורמליזציה:** Group Normalization על ממד התכונות.
- **עיבוד עם BiLSTM:** לימוד תלות רציפה.
- **שכבת Fully Connected:** מתאימה את התכונות לגודל המתאים.
- **Residual Connection [6]:** משמרת מידע מקורי ע"י הוספה של הקלט המקורי לפלט (x + out).
- **שינוי פרמוטציה**: מעבר בין ממדים שונים (T <-> K).

2. BandSequenceModelModule :

המודול השלם מורכב מרשימת שכבות RNNModule שמבצעות:

- עיבוד לאורך ממד הזמן (Time).
- עיבוד לאורך ממד רצועות התדר (Subbands) .

שלבי החישוב:

- מבוצע **עיבוד כפול** בכל שכבה :
  1. BiLSTM לאורך ממד הזמן (Time).

2. BiLSTM לאורך ממד רצועות התדר (Subbands).

- תהליך זה חוזר על עצמו num_layers פעמים (לדוגמה, 3 פעמים בקוד שלנו).

רציונל לחלוקת העיבוד (Time & Sub-bands)

1. **עיבוד לאורך הזמן (Time)**:
   - מסייע להבין כיצד מידע האודיו משתנה לאורך פריימים בזמן.
   - מאפשר למודל ללמוד דפוסים דינמיים כמו חזרתיות במנגינה.
2. **עיבוד לאורך רצועות התדר (Sub-bands)**:
   - מאפשר למודל ללכוד קשרים בין רצועות תדר שונות, שבהן עשויה להיות חפיפה ספקטרלית בין מקורות.
   - לדוגמה, קול של אדם ותדרים גבוהים של תופים עשויים להופיע באותם תחומי תדר, והעיבוד הזה מסייע בהפרדתם.

יתרונות המודול

- **לימוד רב-ממדי:** מבנה המודול מאפשר ללמוד קשרים בממדים שונים (זמן ותדר) בצורה מקבילית.
- **רשת דו-כיוונית (Bidirectional):** תורמת ללכידת תלות גלובלית ולא רק מקומית.
- **Residual Connections:** מסייעות לשמור מידע רלוונטי ומונעות איבוד נתונים חשובים במהלך השכבות.

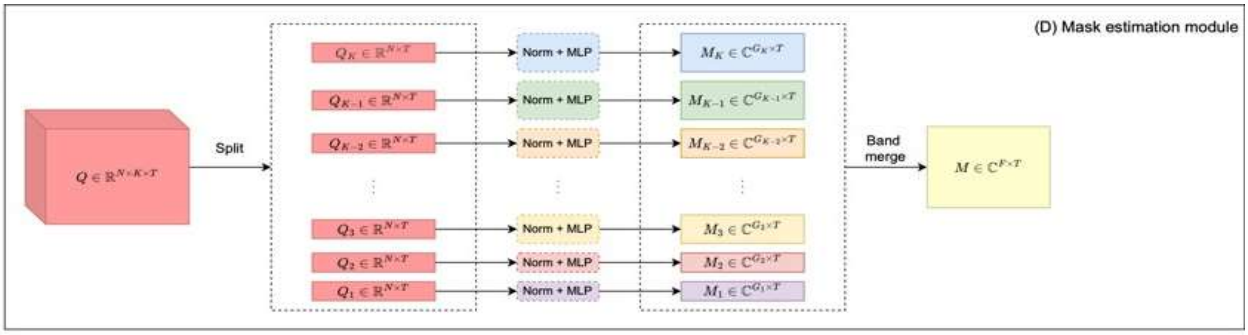
שימוש במודל

בקוד שלנו:

- גודל הקלט (Input) הוא [4, 41, 259, 128]:
  - 4 דוגמאות במיני-באטץ.'
  - 41 רצועות תדר.
  - 259 פריימים בזמן.
  - 128 מאפיינים לכל פריים.
- גודל הפלט (Output) הוא זהה ([4, 41, 259, 128]) , אך המידע בתוכו עבר עיבוד מעמיק ע"י השכבות.

3. Mask Estimation:

רכיב זה מעריך את ה"מסכה" שתשמש לשחזור רכיבי האודיו.



איור 5: מבנה מודול ה- *Mask Estimation*.

- **מטרה:** לקבוע אילו תדרים שייכים לכל מקור אודיו נפרד.
- **מימוש:**
  - כמתואר באיור 5: מבנה מודול ה- Mask Estimation. השימוש בנתוני ה-LSTM משלב קודם, יחד עם מידע ספקטרלי נוסף, משמש להזנת רשת רב-שכבתית (MLP).
  - הפלט של ה- MLP הוא המסכה, בגודל [B, C, F, T] שבה:
    - B: גודל המיני-באטץ'.
    - C: מספר הערוצים (למשל, שירה, תופים, וכו').
    - F: תדרים.
    - T: מסגרות זמן.

4. שילוב ופלט

- המסכה שמחושבת בשלב הקודם מוכפלת בנתוני הספקטרום המקוריים.
- הפעולה הזו יוצרת ספקטרום משוחזר לכל מקור אודיו בנפרד.
- לבסוף, מבוצעת Inverse STFT (ISTFT) כדי להחזיר את האודיו לתחום הזמן.

התאמות ייחודיות

- **מודול Band Split דינמי:** הגדרת תתי-התחומים גמישה, בהתאם לתדרים שבהם מתמקדים המקורות השונים.
- **התמודדות עם אורכים משתנים:** המודל הותאם לטפל בקטעי אודיו באורכים שונים על ידי חלוקה למקטעים קטנים.

יתרונות המבנה

1. **עיבוד תדרים ממוקד:** חלוקה לתת-תחומים מאפשרת עיבוד יעיל ומדויק יותר.
2. **למידת תלות בזמן:** השימוש ב-LSTM משפר את הבנת ההקשרים המבניים באודיו.
3. **שיפור הפרדה:** הערכת המסכה על ידי שילוב MLP ונתונים ספקטרליים נוספים מביאה לשיפור באיכות ההפרדה.

המודל הזה הוא גמיש ומסוגל להתמודד עם אתגרי הפרדת אודיו מורכבים, תוך ניצול ארכיטקטורות מתקדמות ועיבוד נתונים מותאם.

3.3. תהליך האימון

המודל אומן על סט נתונים גדול של שירים, תוך שימוש בהפסד מסוג L1 ולמידת פרמטרים בעזרת אלגוריתם Adam. ערך למידת הקצב (Learning Rate) הותאם ל-0.003, והופעל מנגנון Decay המקטין את ערך הקצב באופן מדורג לאורך האימון.

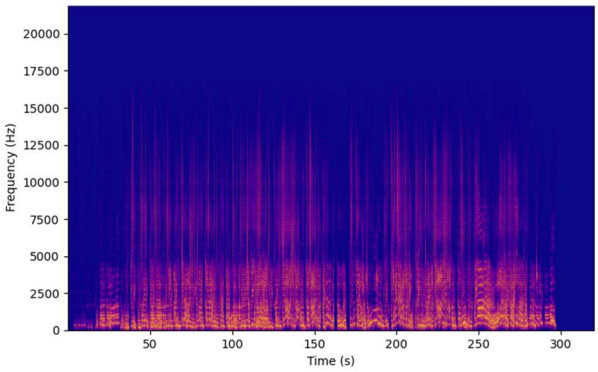
- **עיבוד מקדים:**
  1. קטעי האודיו הומרו לייצוג זמן-תדר באמצעות Short-Time Fourier Transform (STFT) .
  2. מעבר למשרעת (magnitude) הספקטרום, הוזנו למודל גם נתוני פאזה (phase), שהם מרכיב חשוב לצורך שחזור מדויק של האודיו.
  3. שימוש בנרמול, שבו הנתונים מועברים כך שממוצעם יהיה אפס וסטיית התקן תהיה קבועה. פעולה זו מסייעת לשיפור יציבות המודל במהלך האימון.
  4. הוספת נתונים ממוצעים (mean) או טווחי ערכים (scale) המותאמים לנתוני האודיו הספציפיים שקיבלנו מהאימון.
  5. חלוקת הספקטרום לתתי-תחומים (sub bands) כשלב מקדים לפני הכנסתו למודל. זהו חלק מרכזי במודל שמקל על עיבוד התדרים בנפרד.

- **חלוקה:**

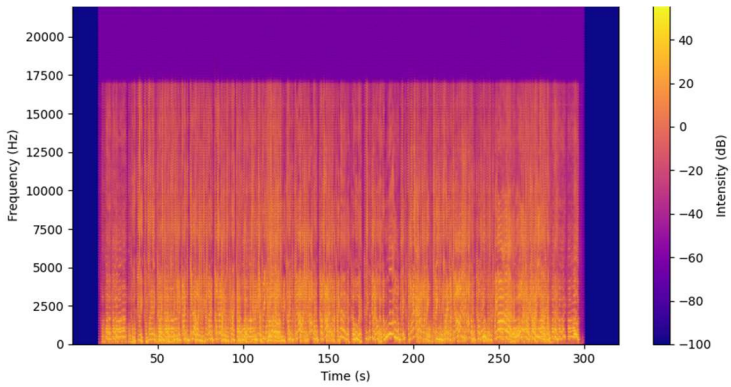
  כדי להתמודד עם קטעי אודיו באורכים שונים, קטעי אימון ואימות חולקו לחלונות באורך קבוע של 6 שניות, שעליהם מתאמן ורץ המודל.

4. תוצאות

התוצאות טובות לאוזן האנושית ויש הפרדה ברורה ואיכותית ברוב הזמן, משום שמאגר המידע שלנו מורכב רק מ-150 שירים בסך הכל, מספר נמוך יחסית למשימה מורכבת זו, ישנם שירים וז'אנרים בהם ההפרדה פחות טובה. למשל אם השיר הוקלט באיכות ירודה, בז'אנר שלא הופיע כלל או מעט מאוד במאגר הנתונים, כנראה שההפרדה לא תהיה איכותית גם לאוזן האנושית, אך בשירים שהוקלטו באיכות גבוהה, ההפרדה ברורה ואיכותית.



איור 6: תוצאת המודל שלנו



איור 7: תוצאת אמת

- ההבדלים העיקריים נמצאים בתדרים הגבוהים, שהאוזן האנושית אינה יכולה להבחין בהם.
- התוצאות ברובן נעימות לאוזן האנושית.
- התוצאות מצביעות על שגיאת ממוצע ריבועים (MSE) של 0.875 על מאגר נתוני הבדיקה.
- לשם השוואה, שגיאת ממוצע הריבועים של מודול Open-Unmix המקורי היא 0.913.

5.סיכום

מטרת הפרויקט הייתה לפתח מודל מתקדם להפרדת אודיו המסוגל לזהות ולהפריד מקורות קול שונים מתוך מיקס אודיו, כגון שירה, תופים ובס. השגת המטרה התבססה על גישות חדשניות כמו חלוקת תדרים, שימוש ברשתות BiLSTM וטרנספורמרים, וניצול מידע ספקטרלי רב-ממדי לשיפור הדיוק.

כיצד הושגה המטרה

1. **חלוקת תדרים (Bandsplits) :** האודיו חולק לרצועות תדר נפרדות, מה שאפשר למודל להתמקד בכל רצועה בנפרד ולשפר את איכות ההפרדה.
2. **מודול BandSequence :** שימוש בעיבוד לאורך ממד הזמן ורצועות התדר במקביל באמצעות BiLSTM , כדי ללמוד יחסים מורכבים בין תדרים ותנועת קול בזמן.
3. **שילוב מידע ספקטרלי נוסף:** הכנת הנתונים כללה נרמול ושימוש בתכונות נוספות לשיפור האימון.
4. **טכניקות אופטימיזציה :** התאמת היפר-פרמטרים ושיפור ניהול זיכרון אפשרו אימון יעיל של המודל גם עם נתונים מורכבים.

מסקנות

- **דיוק גבוה בהפרדה :** שילוב טכניקות כמו חלוקת תדרים ועיבוד כפול במודול BandSequence תרמו משמעותית לשיפור ביצועי המודל.
- **חשיבות ההכנה המקדימה :** תהליך הנרמול ושימוש במידע ספקטרלי נוסף היו קריטיים להצלחת האימון והמודל הסופי.

- **אתגרי שונות :** המודל הצליח להתמודד היטב עם שונות באיכות ההקלטות ורעשי רקע, אך תוצאות ההפרדה היו תלויות לעיתים ברמת השונות בנתונים.
- **עבודה עם קטעי אודיו ארוכים :** שימוש בחלוקה לקטעים קצרים אפשר למודל להתמודד גם עם קטעי אודיו ארוכים, מה שהפך אותו לשימושי במגוון רחב של תרחישים.

הפרויקט הציג פתרון חדשני ויעיל להפרדת אודיו באמצעות למידה עמוקה. הוא מדגיש את הפוטנציאל הטמון בשילוב טכניקות מודרניות בעיבוד קול ומציע תשתית להמשך מחקר בתחום. הפתרון המוצע יכול להשתלב ביישומים מסחריים כמו שירותי סטרימינג, עריכת מוזיקה וניתוח סאונד, ומהווה תרומה משמעותית לקידום טכנולוגיות הפרדת אודיו.

רשימת מקורות

[1] Fabian-Robert Stöter, Stefan Uhlich, Antoine Liutkus, and Yuki Mitsufuji. “Open-Unmix - A Reference Implementation for Music Source Separation.” Journal of Open Source Software (2019). Available: https://www.theoj.org/joss-papers/joss.01667/10.21105.joss.01667.pdf

[2] Simon Rouard, Francisco Massa, Alexandre Défossez. “Hybrid Transformers for Music Source Separation”. EESS Audio and Speech Processing (2022). Available: https://arxiv.org/pdf/2211.08553

[3] Yi Luo, Jianwei Yu. “Music Source Separation with Band-split RNN.” EESS Audio and Speech Processing (2022). Avaliable: https://arxiv.org/pdf/2209.15174.pdf

[4] Zafar Rafii and Antoine Liutkus et Al. “The MUSDB18 corpus for music separation.” (2017). Available: https://doi.org/10.5281/zenodo.1117372

[5] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. Avaliable: https://doi.org/10.1162/neco.1997.9.8.1735

[6] Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing, 45*(11), 2673–2681. Avaliable: https://doi.org/10.1109/78.650093