



בי"ס להנדסת חשמל

פרויקט מס' 3481

תכנית עבודה

שם הפרויקט: Audio processing platform

מבצעים:

ת.ז. 323829853

שם: אדר מאורי

ת.ז. 318255072

שם: בן פוזננסקי

מקום ביצוע הפרויקט: אוניברסיטה

לשימוש המנחה:

הנני מאשר את תכנית העבודה המצורפת

שם: ליאור ארבל

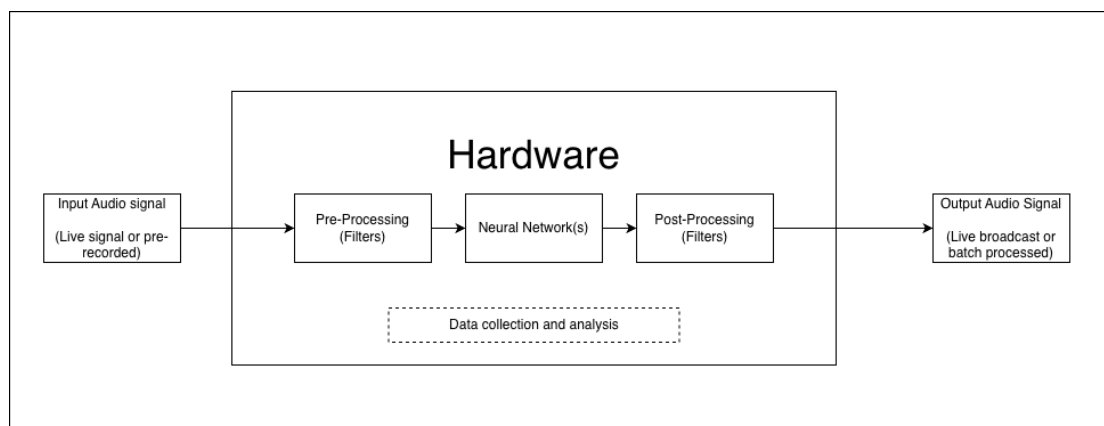
חתימה: ליאור ארבל

1 תקציר

פרויקט זה בוחן את האתגרים וההזדמנויות בשימוש ברשתות נוירונים (NN) עבור יישומי עיבוד אודיו הדורשים תגובה בזמן אמת (real-time). בעוד שרשתות נוירונים מציגות יכולות מתקדמות במשימות אודיו כגון סינון רעשים, יצירת אפקטים מוזיקליים וזיהוי דיבור, הן דורשניות מאוד מבחינת כוח חישוב, דבר המהווה חסם משמעותי במערכות זמן-אמת. מטרת הפרויקט היא לבצע מחקר השוואתי של ביצועי רשתות נוירונים שאומנו מראש למטרות אודיו, על גבי פלטפורמות חומרה מגוונות. במסגרת הפרויקט, נממש ונבחן את אותה הרשת על חומרות שונות, ביניהן מעבד מרכזי (PC), חומרה מיתכנתת (FPGA) ופלטפורמת עיבוד Bela.

בנוסף, נחקור את ההשפעה של אלגוריתמים שונים למימוש פעולות הליבה של הרשת (כגון כפל מטריצות) על הביצועים הכוללים, בדגש על זמן ההשהיה (latency) וניצול המשאבים. תוצר הפרויקט יהיה ניתוח מקיף ומתוקף של החלופות השונות, ופיתוח שיטות מימוש אופטימלית/מתודות חדשות להתאמת עיבוד אודיו בזמן אמת באמצעות רשתות נוירונים, פר פלטפורמה.

דיאגרמת בלוקים של התהליך הנבחן:



2 מוטיבציה

המוטיבציה המדעית וההנדסית לפרויקט זה נובעת מהפער הקיים בין היכולות התאורטיות של רשתות נוירונים (NN) בתחום האודיו, לבין הקושי המעשי ביישומן במערכות זמן אמת. כיום, רשתות נוירונים הן כלי מוביל במשימות מורכבות של עיבוד אותות, אך העלות החישובית הגבוהה שלהן מקשה על שילובן במוצרים הדורשים latency נמוך מאוד, כגון מכשירי שמיעה, מערכות תקשורת חיות, או אפקטים מוזיקליים לנגינה חיה. תרומת הפרויקט היא במתן מתודולוגיה וניתוח כמותי לבחירת צורת המימוש (חומרה ואלגוריתמיקה) היעילה ביותר עבור יישום נתון, ובכך לאפשר את הרחבת השימוש בטכנולוגיות NN מתקדמות בתחומים אלו. השוואה לחלופות קיימות:

1. **אלגוריתמי עיבוד אותות קלאסיים (DSP):** שיטות מסורתיות כמו מסנני FIR/IIR, ואלגוריתמים אדפטיביים (LMS, RLS) הן יעילות מאוד מבחינה חישובית ובעלות latency נמוך. עם זאת, הן מתקשות להתמודד עם בעיות מורכבות ולא-לינאריות כמו הפרדת מקורות או סינון רעשים לא-סטציונריים, משימות שבהן רשתות נוירונים מצטיינות.
2. **עיבוד מבוסס ענן (Cloud-based Processing):** חלופה זו מאפשרת שימוש במשאבי חישוב כמעט בלתי מוגבלים להרצת רשתות נוירונים מורכבות. החיסרון המרכזי והמכריע הוא זמן ההשהיה (latency) הנגרם מהצורך לשלוח את האודיו לשרת מרוחק, לעבד אותו ולקבל את התוצאה בחזרה. זמן השהיה זה אינו קביל ברוב יישומי זמן האמת.

3 תכולת העבודה

תכולת העבודה בפרויקט מחולקת למספר שלבים עיקריים, החל מרכישת הרקע התאורטי ועד לניתוח התוצאות והסקת מסקנות.

רקע תאורטי נדרש:

- **רשתות נוירונים:** הכרת ארכיטקטורות רלוונטיות לעיבוד אודיו, בעיקר רשתות קונבולוציה (CNN) ורשתות רקורנטיות (RNN/LSTM).
- **עיבוד אותות ספרתי (DSP):** עקרונות דגימה, קוונטיזציה, ניתוח ספקטרום (STFT) וחלון.
- **ארכיטקטורת חומרה:** הבנת ההבדלים בין ארכיטקטורות SoC על PC, ומערכות משובצות (Embedded) כמו FPGA, Bela, בדגש על ביצוע פעולות אריתמטיות מקבילות.
- **אלגוריתמים לחישובים נומריים:** לימוד אלגוריתמים לכפל מטריצות (GEMM) וקונבולוציה, והכרת ספריות ממוטבות כמו cuDNN, BLAS ו-CMSIS-NN.

דרישות ושלבי עבודה:

1. שלב א' - מחקר והכנת סביבה:
 - סקירת ספרות ובחירת 2-3 רשתות נוירונים מאומנות וזמינות בקוד פתוח למשימות כמו סינון רעשים או אפקט גיטרה (למשל, מודל מבוסס WaveNet או RNNNoise).
 - הקמת סביבות הפיתוח והבדיקה על כל אחת מהחומרות: PC (עם Python), FPGA (באמצעות Verilog), ופלטפורמת Bela (עם C++).
2. שלב ב' - מימוש ובחינת ביצועים:
 - מימוש גרסת בסיס של הרצת המודל (inference) על כל פלטפורמת חומרה.
 - מימוש ואינטגרציה של לפחות שתי שיטות שונות לביצוע כפל המטריצות/קונבולוציות ברשת (למשל, מימוש נאיבי מול מימוש המסתמך על ספרייה ממוטבת).
 - פיתוח מערך בדיקות (benchmark) סטנדרטי למדידת ביצועים: זמן השהיה (end-to-end latency), קצב עיבוד (throughput) וניצול משאבים (CPU/RAM).
 - הרצת הבדיקות ואיסוף נתונים כמותיים.
3. שלב ג' - ניתוח ותיעוד:
 - ניתוח השוואתי של התוצאות והצגתן בצורה גרפית.
 - זיהוי צווארי הבקבוק בביצועים עבור כל תצורה (חומרה + אלגוריתם).
 - גיבוש המלצה מנומקת לשיטת מימוש אופטימלית בהתאם למגבלות המערכת (למשל, "עבור latency מינימלי, מומלץ להשתמש ב-FPGA במימוש X", או "עבור מערכת דלת משאבים, Bela עם אלגוריתם Y נותן את התוצאה הטובה ביותר").

כלים, סביבה וחומרה:

- תוכנה
 - C/C++
 - Python (PyTorch, TensorFlow, Librosa, PANDAS)
 - CUDA
- חומרה
 - PC (SoC)
 - FPGA ייעודי
 - Bela Starter Kit

4 תוצרי הפרויקט

תוצר לסוף סמסטר א' (הצגת התקדמות):

- דוח מסכם של סקר הספרות ובחירת המודלים והאלגוריתמים הנבחרים לכל פלטפורמה.
- מימוש ראשוני של לפחות מודל אחד על פלטפורמת ה-Bela כולל תוצאות המדיגמות את פעולתו (למשל, קובץ אודיו לפני ואחרי סינון רעשים או הוספת אפקט).
- מימוש ראשוני של לפחות מודל אחד על פלטפורמת ה-PC כולל תוצאות המדיגמות את פעולתו (למשל, קובץ אודיו לפני ואחרי סינון רעשים או הוספת אפקט).

תוצרים סופיים:

- תיאור דרישות המערכת:
 - כלל החומרות:
 - זמן השהיה (Latency): 10% שיפור בזמן הריצה תוך פגיעה של עד 5% באיכות.
 - ניצול משאבים: המימוש לא יעבור 80% מניצול הזיכרון והמעבד בממוצע, כדי לאפשר פעולת מערכות אחרות.
 - PC:
 - שיפור ב-20% בהשהיה לעומת מודל דומה בענן, ללא פגיעה באיכות.
- קוד מקור וסימולציות:
 - מאגר קוד (Git Repository) המכיל את כל המימושים, סקריפטים להרצת הבדיקות, וניתוח הנתונים.
 - דוח ביצועים מקיף המשווה את כלל התצורות שנבדקו.
- שיטות בדיקה וסביבת בדיקה:
 - סביבת בדיקה: המערכת תקבל קלט אודיו מקובץ WAV סטנדרטי. הפלט המעובד יישמר לקובץ WAV נפרד. מדידות latency וניצול משאבים יתבצעו באמצעות כלי פרופילינג תכנותיים (software profiling).
 - בדיקות:
 - בדיקות פונקציונליות: האזנה סובייקטיבית לתוצאות ובדיקה ויזואלית של הספקטרוגרמה כדי לוודא שהעיבוד מתבצע כמצופה.
 - בדיקות ביצועים כמותיות: הרצת כל מימוש 100 פעמים על אותו קובץ קלט ומדידת ממוצע וסטיית תקן של זמן העיבוד וניצול המשאבים.

5 לוח זמנים

אבן דרך	פירוט (2-3 שורות)	תאריך יעד לביצוע	הערות
השלמת סקר ספרות	בחינה של כל רכיב חומרה - IO, ממשק תוכנתי, קצבי עבודה. רשתות נוירונים - סקירת המודלים והתאמתם לחומרות אלגוריתמים - IO, זמני ריצה	21.12.2025	
הקמת סביבות עבודה	הקמת סביבת עבודה בפלטפורמת PC ובחינתה עם מודל/אלגוריתם פשוט	1.2.2026	
	הקמת סביבת עבודה בפלטפורמת Bela ובחינתה עם מודל/אלגוריתם פשוט	8.2.2026	
	הקמת סביבת עבודה בפלטפורמת FPGA ובחינתה עם מודל/אלגוריתם פשוט	15.2.2026	
הגשת מצגת האמצע		פברואר-מרץ 2026	
הצגת מצגת אמצע			
סיכום התקדמות אמצע סמסטר	כל הפלטפורמות תקינות, התקדמות משמעותית בבחינת מודלים ואלגוריתמים רלוונטיים	12.4.2026	
שילוב pre ו-post processing	שילוב פילטרים ואפקטים אחרים לפני/אחרי הרצת רשתות הנוירונים	20.4.2026	
שילוב אלגוריתמים ומודלים מתקדמים	שילוב האלגוריתמיקה המתקדמת וביצוע אופטימיזציה לזמן הריצה במימוש המודלים על PC.	29.4.2026	
	שילוב האלגוריתמיקה המתקדמת וביצוע אופטימיזציה לזמן הריצה במימוש המודלים על Bela.	7.5.2026	
	שילוב האלגוריתמיקה המתקדמת וביצוע אופטימיזציה לזמן הריצה במימוש המודלים על FPGA.	14.5.2026	
ניתוח תוצאות	ניתוח זמני ריצה, עלות חומרתית ואיכות התוצרים כתלות באלגוריתמיקה, פילטרים, חומרה, תוכנה	21.5.2026	
הגשת הפוסטר וסיום העבודה בפרויקט		24/5/2026	
אישור ספר הפרויקט		שבוע לפני הגשת ספר הפרויקט	
הגשת ספר הפרויקט ומצגת הסיום		יולי-אוגוסט 2026	