esp32 ומודול זיהוי קול Elechouse v3

1. מבוא כללי

במאמר זה נחבר אל מיקרו בקר ESP32 מודול זיהוי קול Elechouse V3 .

על מודול זיהוי קול Elechouse V3 בהרחבה כולל עבודה עם ארדואינו ניתן למצוא בקישור :

https://www.arikporat.com/wp-content/uploads/2022/12/voiceRecognition-v3.pdf

מודול ה **Elechouse Voice Recognition Module V3** - הוא חיישן זיהוי קול המאפשר למערכת לזהות פקודות קוליות. המודול תומך בעד 80 פקודות קוליות מוגדרות מראש, כאשר עד 7 מהן יכולות להיות פעילויות במקביל. החיבור עם ESP32 מתבצע לרוב באמצעות UART מה שמקל על האינטגרציה.

1.1 עקרונות זיהוי הקול במודול elechouse v3

במודול **Elechouse Voice Recognition V3** העקרונות והאלגוריתם לזיהוי קול מתבססים על ניתוח אותות קוליים והשוואתם לדגימות קול שנשמרו במערכת. המודול אינו מבצע עיבוד שפה טבעית, אלא מזהה דפוסים קבועים של גלי קול. הנה כמה עקרונות בסיסיים:

א. דגימת האות הקולי:

- כאשר המודול מקבל כניסת קול, הוא מבצע דיגיטציה של האות הקולי, הופך אותו לנתונים דיגיטליים.
- דגימה זו מבוצעת בפרקי זמן קבועים, כך שניתן לקבל ייצוג דיגיטלי רציף של האות הקולי (דגימת הקול).

ב. חילוץ מאפיינים(Feature Extraction):

- המודול מחלץ מאפיינים ייחודיים מתוך האות הקולי, כגון תדרים, עוצמות ושינויים בזמן. תהליך זה מאפשר לבנות דגם של האות הקולי שניתן להשוות אותו לדגמים קודמים.
- בשלב זה נבדקים גם רכיבי התדר של הקול (Fourier Transform) כדי להפריד בין תכונות עיקריות בתדרים שונים.

ג. השוואה לדגימות שמורות:

- המודול משווה את האותות הקוליים שהוא קיבל לדגימות קוליות שנשמרו מראש. ההשוואה נעשית על ידי חישוב דמיון בין מאפייני האות החדש לבין אלו של הפקודות שנשמרו.

ד. תהליך ההכרה(Recognition):

- אם נמצא דמיון מספק בין הדגימה הקולית החדשה לבין דגימה שמורה, המודול מזהה את הפקודה כפקודה מוגדרת.
- במקרים בהם הדמיון אינו מספיק או שיש רעשים ברקע, המודול לא יזהה את הפקודה בצורה נכונה.

ה. אימון המודול(Training):

- כדי לרשום פקודות חדשות, המודול צריך לעבור תהליך של "אימון" שבו הוא לומד את הדפוס של הפקודה הקולית מהמשתמש. תהליך זה מבוצע בממשק תוכנה שמדריך את המודול ללמוד את הפקודות.
- לאחר האימון, המודול שומר את הדגימות בזיכרון שלו לצורך זיהוי בעתיד.

www.arikporat.com

1.2 אלגוריתם הזיהוי:

המודול **Elechouse V3** משתמש באלגוריתם זיהוי קול בסיסי המבוסס על **DTW (Dynamic Time Warping)** עיוות/עיקום זמן דינמי , או וריאציה כלשהי שלו.

מהו DTW ?

Dynamic Time Warping הוא אלגוריתם שמשמש להשוואת שתי סדרות של נתונים (במקרה זה, סדרות של מאפייני אותות קוליים) שאולי יש ביניהן הבדלים קטנים בזמן, אך הן עשויות להיחשב דומות או תואמות.

עקרונות של DTW :

א. זיהוי דפוסים בזמן משתנה:

- DTW מאפשר להשוות בין שני דפוסי קול, גם אם אחד מהם נאמר מהר יותר או לאט יותר מהשני. הוא עושה זאת על ידי עיוות/עיקום (warping) של ציר הזמן כך שהשוואת הדפוסים תהיה מדויקת.

ב. מציאת המסלול המינימלי:

- האלגוריתם מחשב את המסלול המינימלי של עיוות הזמן הדרוש כדי להתאים את הדפוס החדש לדפוס שמור, תוך חישוב ההפרשים בין נקודות שונות בשני הדפוסים.

ג. חישוב עלות התאמה:

- בסוף התהליך ה - DTW מחשב את "עלות" ההתאמה, שהיא סך ההפרשים בין הדפוסים לאורך ציר הזמן. אם עלות ההתאמה נמוכה דיה, הדפוסים נחשבים תואמים.

1.3 יתרונות DTW במודולי זיהוי קול:

- **יכולת לזהות פקודות גם בשינויים קלים** : מאפשר למודול לזהות פקודות גם אם נאמרות בקצבים שונים (מהירות דיבור, דגימות ברעש וכו').
- **דיוק גבוה בפקודות מוגדרות מראש** : האלגוריתם מותאם במיוחד לזיהוי דפוסים חוזרים, ולכן מתאים לזיהוי פקודות קבועות שהוגדרו מראש.

1.4 מגבלות האלגוריתם:

- **רגישות לרעשי רקע** : במקרים של רעשי רקע חזקים, האלגוריתם עלול להתקשות בזיהוי נכון.
- **מגבלת מספר פקודות בו זמנית** : ניתן להפעיל רק מספר מוגבל של פקודות לזיהוי בכל זמן נתון (עד 7).

1.5 סיכום:

מודול **Elechouse V3** משתמש באלגוריתם זיהוי קול מבוסס DTW או שיטה דומה, שמספקת זיהוי אמין של פקודות קוליות מוגדרות מראש. האלגוריתם מזהה דפוסים חוזרים בגלי קול גם כאשר יש הבדלים בקצב או בעוצמה, מה שהופך אותו למתאים לזיהוי פקודות קוליות מוגדרות בקבצי אימון, כל עוד הסביבה נקייה מרעשי רקע חזקים.

www.arikporat.com

1.6 עקרון העבודה של המודול:

א. **הגדרת פקודות קוליות** : קודם כל, יש צורך לרשום את הפקודות הקוליות שהמערכת תזהה. כל פקודה קולית מקבלת מזהה ייחודי (Group ID) .

ב. **זיהוי פקודות קוליות** :לאחר רישום הפקודות, המודול מקשיב לכניסות קוליות ומשווה אותן לפקודות הרשומות. כשמתבצעת התאמה, הוא שולח למיקרו-בקר את מזהה הפקודה שזוהתה.

1.7 מודולים של זיהוי קול נוספים

ישנם מספר מודולים מתקדמים יותר ל**זיהוי קול** אשר מציעים פונקציונליות משופרת בהשוואה ל **Elechouse V3** -במיוחד עבור פרויקטים שדורשים זיהוי קולי מתקדם, עיבוד שפה טבעית (NLP) או עבודה עם מספר רב של פקודות בסביבות רועשות.

א. Google AIY Voice Kit

- פותח על ידי Google ומאפשר שימוש בזיהוי קול ברמת AI מתקדם.
- מאפשר זיהוי שפה טבעית באמצעות ממשקים ל **Google Assistant** .
- יכול לבצע פעולות מורכבות בהתאם לשפה טבעית ולהתחבר לענן של Google לעיבוד מתקדם.
- מיועד למערכות חכמות שמחוברות לרשת.

יתרונות:

- יכולת ניתוח שפה טבעית (NLP) .
- תמיכה בענן ובמודלים למידת מכונה .
- אינטגרציה קלה עם Google Assistant .

חסרונות:

- דורש חיבור לאינטרנט לזיהוי מתקדם.
- עלול להיות מסובך יותר להטמעה בפרויקטים פשוטים או במערכות לא מקוונות.

ב. Respeaker 2-Mic Array / 4-Mic Array

- סדרת מודולים שפותחה על ידי **Seeed Studio** הכוללים מיקרופונים מרובים לזיהוי קול מתקדם.
- תומך בזיהוי מרחוק (far-field voice recognition) ומסוגל לעבוד טוב גם בסביבות רועשות.
- כולל תמיכה ב Amazon Alexa, Google Assistant,-ובפתרונות עיבוד קול מתקדמים נוספים.

יתרונות:

- עיבוד קול חזק בסביבות רועשות.
- מספר מיקרופונים מאפשרים "כיווניות" של השמע וזיהוי גם ממרחק.
- תמיכה ב AI - ובמודולי זיהוי קול מתקדמים.

חסרונות:

- מורכב יותר לשימוש, במיוחד עבור מתחילים.
- דורש חיבור למיקרו-בקר חזק יותר או חיבור לענן עבור עיבוד מתקדם.

www.arikporat.com

ג. Snips Voice Platform

- פלטפורמה מבוססת AI שמיועדת למערכות חכמות ומאפשרת זיהוי קול ללא חיבור לענן.
- התמקדות בפרטיות – כל העיבוד נעשה מקומית על המכשיר.
- תומך ביצירת ממשקים קוליים עם זיהוי שפה טבעית (NLP) בתוך המכשיר עצמו.

יתרונות:

- עיבוד מקומי ללא צורך בענן – יתרון לפרויקטים שדורשים פרטיות.
- אפשרות לזיהוי שפה טבעית (NLP) .

חסרונות:

- דורש יותר כוח עיבוד בהשוואה למודולים פשוטים.
- יקר ומורכב יותר להתקנה ולהגדרה.

ד. TensorFlow Lite עם ESP32-S3

- ה ESP32-S3-הוא גרסה משופרת של ה ESP32-שכוללת תמיכה ביכולות AI ו ML -מובנות, כולל אפשרות להריץ **TensorFlow Lite** לזיהוי קול מתקדם.
- מאפשר לבצע זיהוי קול באמצעות מודלים של למידת מכונה (כמו זיהוי מילות מפתח) ישירות על המכשיר.

יתרונות:

- מאפשר עיבוד מתקדם על המכשיר עצמו (Edge AI) ללא צורך בחיבור לענן.
- גמישות רבה עם TensorFlow Lite לזיהוי דפוסים קוליים ומילות מפתח.

חסרונות:

- דורש ידע מתקדם ב AI-ו ML-להגדרת המודלים.

סיכום:

- **Elechouse V3**הוא מודול פשוט לשימוש עם יכולות בסיסיות של זיהוי קול, מתאים לפרויקטים פשוטים.
- אם אתה מחפש מודול עם **זיהוי קול מתקדם יותר** ויכולת עיבוד שפה טבעית או עבודה בסביבות רועשות, מודולים כמו **Google AIY**, **ReSpeaker** או פלטפורמת **Snips** יהיו פתרונות טובים יותר.
- ליישומים מתקדמים עם **עיבוד מקומי (Edge AI)**, שימוש ב **ESP32-S3**-עם **TensorFlow Lite** יאפשר לך גמישות רבה בזיהוי דפוסים קוליים ואינטגרציה עם אלגוריתמים למידת מכונה.

2 חומרה

כמה שיקולים בעת חיווט מודול זיהוי הקול אל ESP32 :

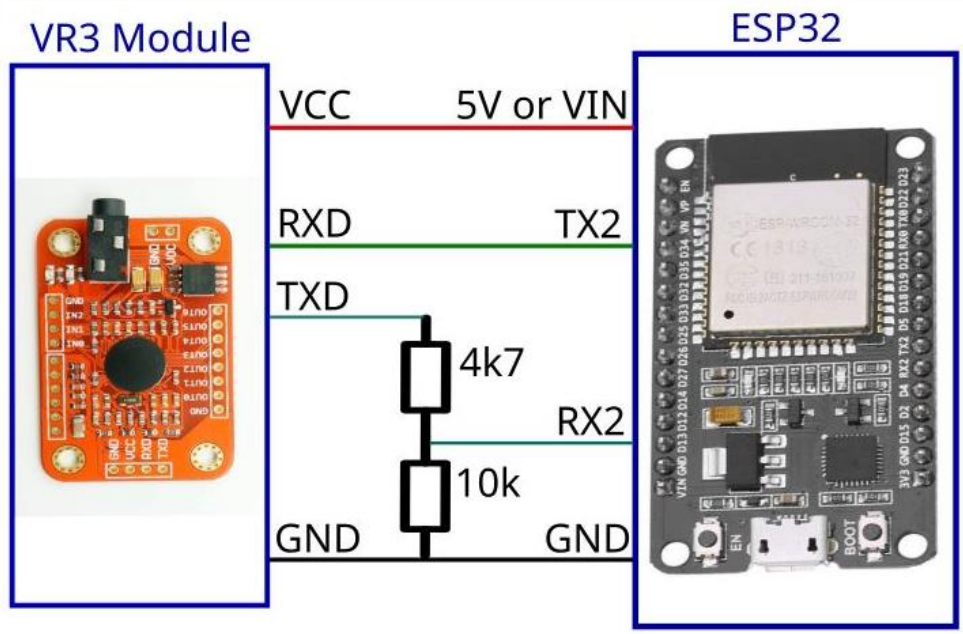
- מודול זיהוי הקול Elechouse VR3 הוא מודול העובד עם 5 וולט, וההדקים של ה - ESP32 אינם עמידים במתח של 5 וולט ויכול לקרות להם נזק. . לפיכך, יש להוריד את המתח היוצא מהדק השידור TXD של מודול זיהוי הקול למתח של

www.arikporat.com

3.3 וולט שנכנס להדק RXD של ה ESP32 . ניתן להשיג זאת בעזרת שני נגדים. נבחר 2 נגדים של 4.7 KΩ ו 10KΩ . כאשר בהדק TXD יש 5 וולט על נגד ה 4.7 KΩ ייפול 1/3 ממתח ה 5 וולט ועל נגד 10KΩ ייפול כ 2/3 ממתח ה 5 וולט (כ 3.3 וולט). מדוע לקחנו נגד של 4.7 KΩ ולא לקחנו נגד של 5 KΩ ? כי הוא קשה להשגה בחנויות רכיבים. 4.7 KΩ הוא נגד נפוץ וניתן להשגה בקלות.

- ESP32 כולל שלושה פורטים טוריים (UART). הראשון הוא U0UXD והוא משמש עבור יציאת ה USB הטורית כולל לתכנות (צריבת התוכנית בהורדה מהמחשב) . השני, U1UXD, משמש בדרך כלל לחיבור לשבב פלאש SPI. מכאן שרק ה-UART השלישי, U2UXD, באמת בטוח לשימוש. מסיבה זו, הן דיאגרמת החיווט המוצגת להלן, והן התוכנה, מניחים שה-Elechouse VR3 מחובר ל- UART השלישי הנקרא U2UXD.

החיבור נראה באיור הבא :



איור 1 : חיבור ESP32 ומודול זיהוי קול Elechouse VR3 .

מודול ה ESP32 מקבל מתח מיציאת ה-USB שלו, ומספק זאת ל-Elechouse VR3. בהתאם ללוח ESP32 שבו נשתמש הדק ה 5 וולט בלוח ESP32 עשוי להיקרא 5V או שהוא עשוי להיקרא גם VIN.

3 תוכנה

3.א נשתמש בסביבת העבודה של הארדואינו . קישור המסביר כיצד להוריד את תוכנת ה ESP32 לסביבת העבודה של הארדואינו ניתן למצוא בקישור לאתר בשפה העברית:

https://www.arikporat.com/wp-content/uploads/2023/01/introduction-to-esp32.pdf

או קישור לאתר בשפה האנגלית :

https://randomnerdtutorials.com/installing-the-esp32-board-in-arduino-ide-windows-instructions/

ב.3 כעת יש להתקין את ספריית זיהוי הקול V3. בצע את ההוראות מדף Github של Elechouse בקישור :

GitHub - elechouse/VoiceRecognitionV3: Arduino library for elechouse Voice Recognition V3 module

ג.3 יצרנית המודול חברת Elechouse ספריה עבור הארדואינו שאיננה עובדת באופן מלא עם ESP32 . יש להחליף 4 קבצים (לבצע overwrite) . הקבצים נמצאים בקישור :

ESP32-ElechouseVR3/Arduino_Files at main · frankvanhooft/ESP32-ElechouseVR3 · GitHub

ד.3 ארבעת הקבצים הם :

- VoiceRecognitionV3.cpp
- VoiceRecognitionV3.h
- vr_sample_train.ino
- vr_sample_bridge.ino

ה.3 שני הקבצים הראשונים VoiceRecognitionV3.cpp ו VoiceRecognitionV3.h מצורפים על ידי חברת Elechouse וגם 2 הקבצים האחרונים עם הסיומת ino ניתנו גם הם על ידי Elechouse והם שימושיים מאוד.

ו.3 השינוי שבוצע ב 2 הקבצים הראשונים הם מחיקת השורה : #include <softwareSerial.h> כי הספרייה הזו איננה עובדת עם ESP32 . כמו כן עודכן הפורט הטורי לעבודה עם U2UXD (פורט תקשורת טורית UART2).
גם השידורים מה-ESP32 ל-VR3 הואטו מעט. השהיה קטנה הוכנסה בין בתים עוקבים בעת שליחת פקודות ל-VR3 – נראה שזה משפר את האמינות.

ז.3 ברירת המחדל של VR3 היא 9600 באוד, כך שכאשר קוראים לפונקציה ( )setup תמיד עוברים ב-9600 אלא אם כן קצב השידור של VR3 שונה במפורש.

ח.3 לקבלת הוראות ומדוע להשתמש ב- vr_sample_train וב- vr_sample_bridge יש לעיין בתיעוד של Elechouse בקישור

VR3_manual.pdf (elechouse.com)

4 שימוש בתוכנת ההקלטות vr_sample_train.ino :

א.4 העלה/טען את התוכנית לערכת ה ESP32 . תוכנית זו מבצעת "אימון" , כלומר מקליטה הודעות למודול . יש לפתוח את המוניטור הטורי . לשים לב שקצב התקשורת 115200 ו "תחילת שורה ומעבר לשורה חדשה" ,
עבור כל אימון/הקלטת דיבור התוכנית תודיע לנו בעזרת המוניטור הטורי "speak now" – דבר עכשיו ואז נשמיע את ההודעה שנרצה (זמן ההודעה לא יותר מ 1.5 שניות). זמן קצר לאחר מכן נקבל הודעה "speak again" – דבר שוב ואז יש

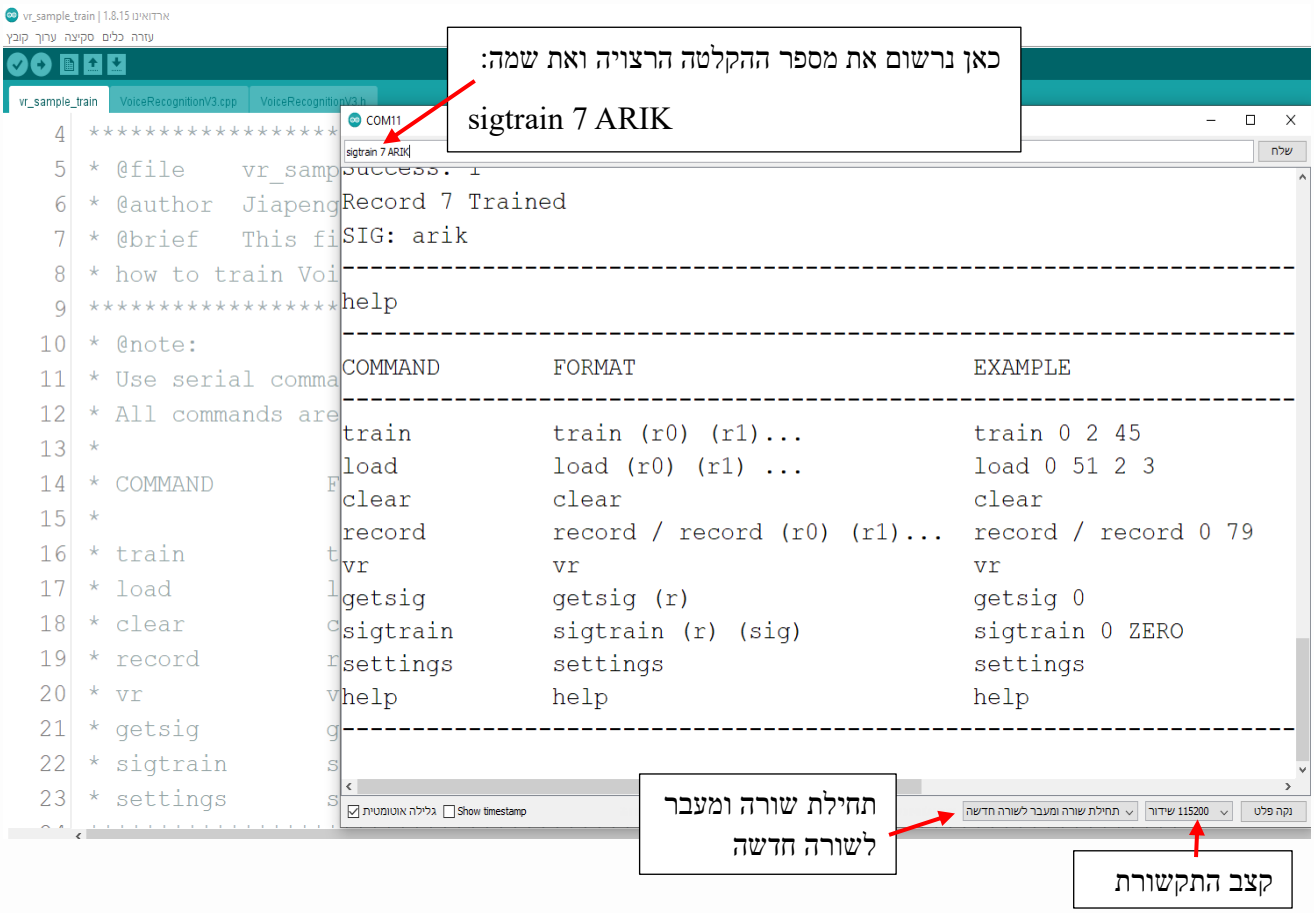
www.arikporat.com

להשמיע אותה הודעה. אם 2 ההודעות זהות (כמעט ..) נקבל success – הצלחה ו matched – בוצע תאום. אם לא נמצא תאום בין 2 ההודעות נקבל הודעה cannwt matched – לא יכול לתאם ומיד לאחר מכן יתבצע ניסיון תאום נוסף של דיבור. דיבור חוזר וניסיון התאמה .

ב.4 נרשום למעלה בצד שמאל את מספר ההקלטה הרצויה לנו עם שם/חתימת ההקלטה לדוגמה : sigtrain 7 ARIK .

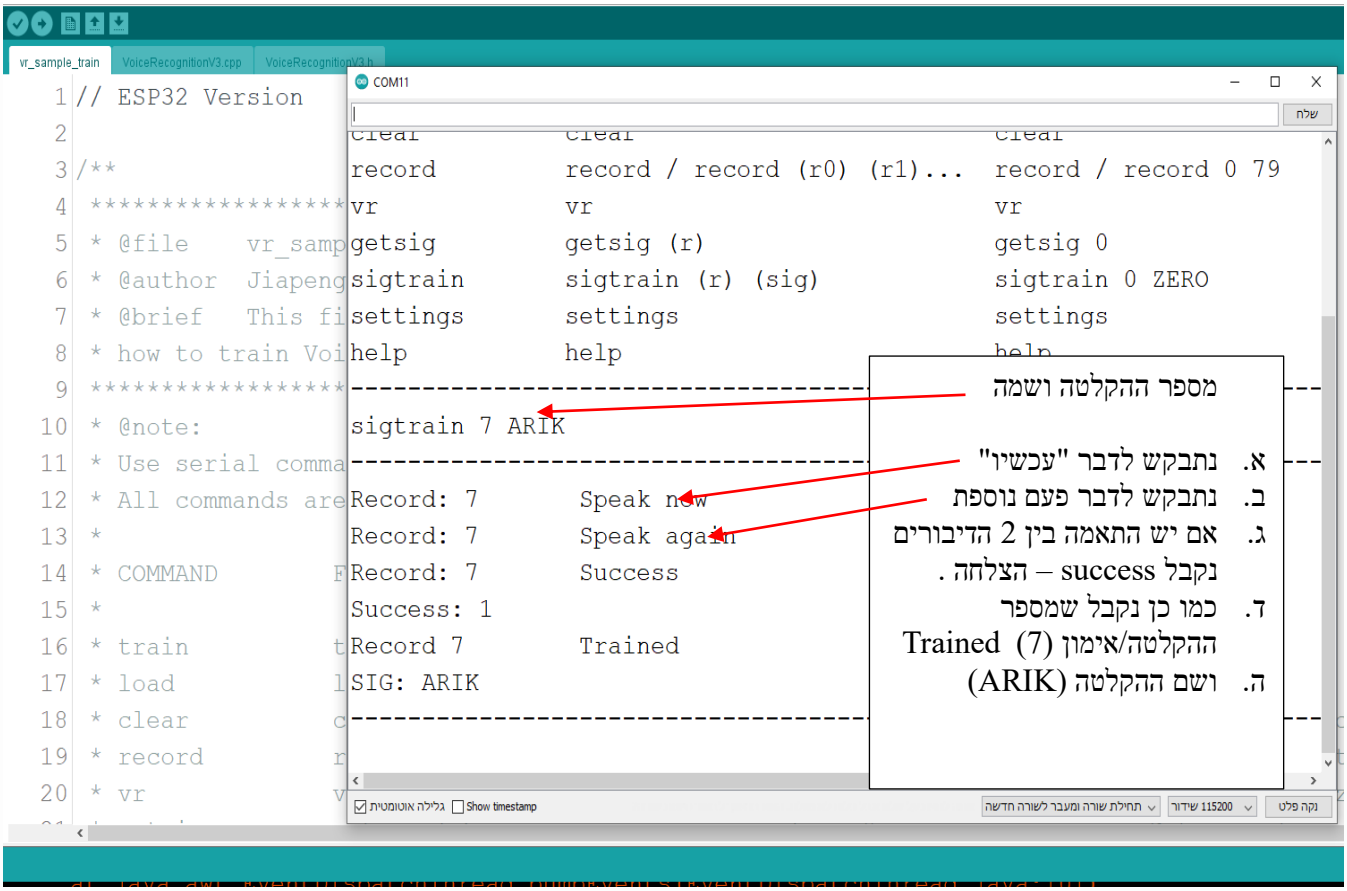
כלומר אימון/הודעה 7 ששמה/חתימתה ARIK . בסיום הכתיבה נלחץ על מקש Enter או "שלח" מצד ימין.

הערה : יש לשים לב שבזמן ה "אימון" אין רעשי רקע כמו רעש מכוניות בכביש, מטוסים, דיבורי של אנשים או טלוויזיה עובדת וכו'. האיור הבא מתאר את הנאמר.



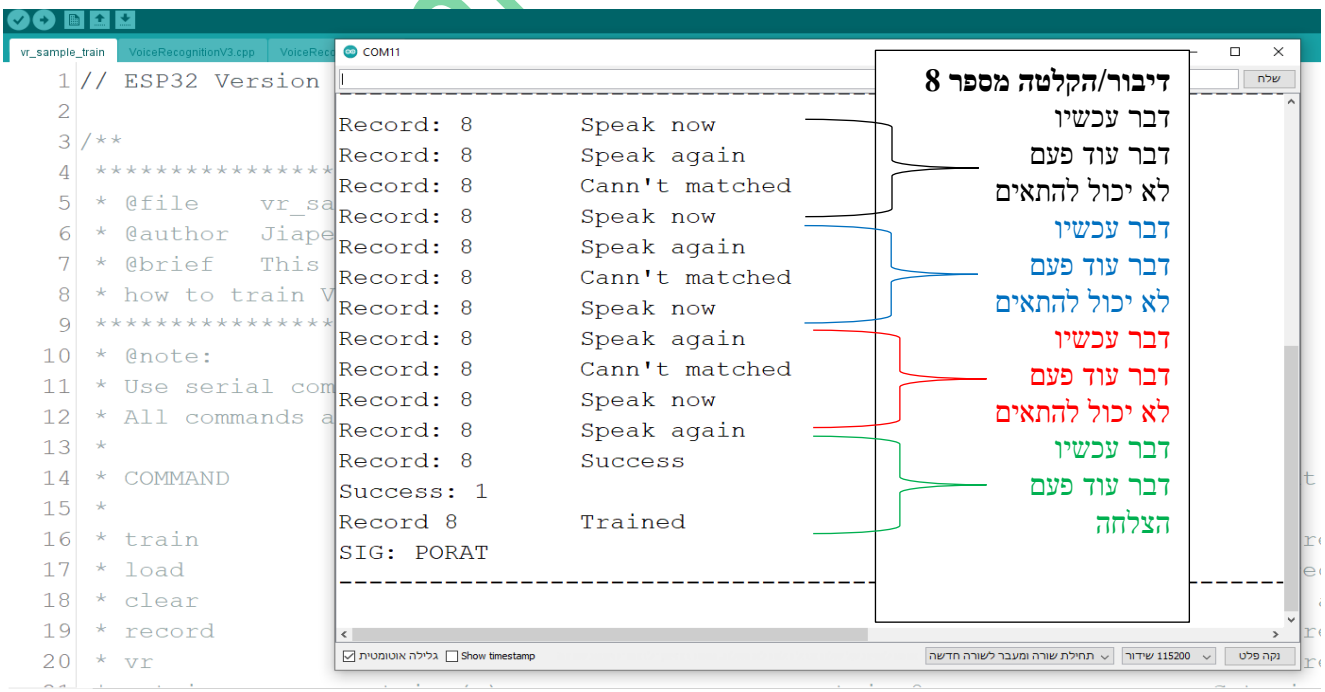
איור 2 : כתיבת מספר ההקלטה הרצויה ואת שם ההקלטה.

ג.4 כאשר נלחץ על "שלח" או מקש Enter נקבל את המסך המתואר באיור הבא :



איור 3 : מסך הקלטות הודעה .

ד.4 במקרה ודיברנו 2 הודעות והמודל לא מצא ביניהן זהות הוא מודיע שאין התאמה.



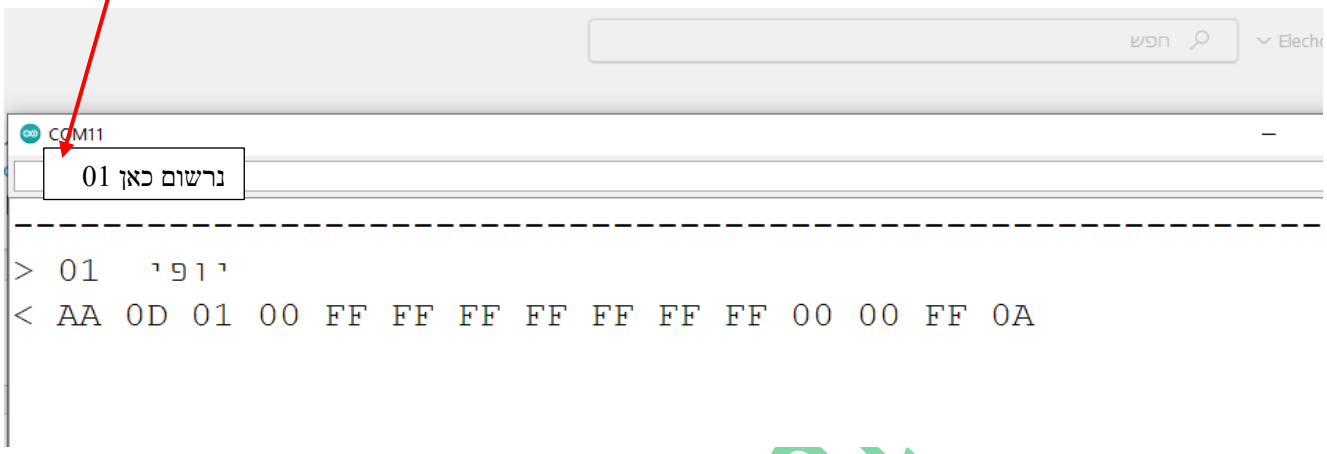
איור 4 : דיבור של הודעות ללא התאמה עד להתאמה.

ניתן לבצע הקלטה אחת בו זמנית . כך נבצע הקלטה אחת אחרי השנייה עד לסיום כל ההודעות הרצויות.

5 שימוש בתוכנית זיהוי ההודעות vr_sample_bridge.ino :

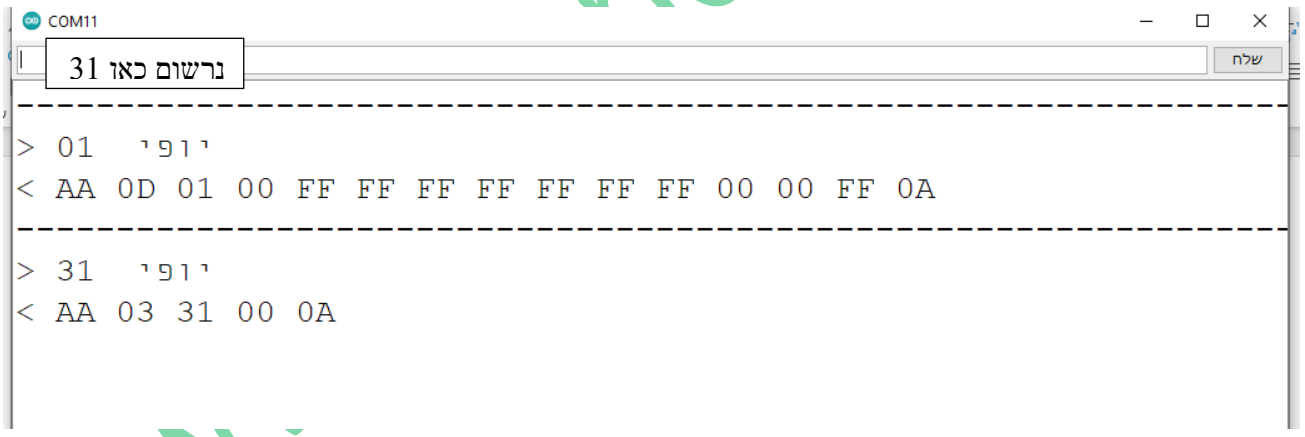
זוהי תוכנית שבה נשמיע את ההודעות ונקבל זיהוי של ההודעות שהשמענו .

5.א נטען את התוכנית לכרטיס ה ESP32 . בתחילה נבדוק שיש קשר עם המודול . לשם כך נרשום בשורת הפקודה 01 שאומרת בדיקת רקוגנייזר ונלחץ על Enter או שלח ונקבל את המסך שבאיור הבא :



איור 5 : תגובת המודול לפקודה 01 .

5.ב נרשום את הערך 31 בשורת הפקודה ונקיש Enter או שלח ונקבל את המסך שבאיור הבא:



איור 6 : תגובת המודול לפקודה 31

5.ג כדי לבצע זיהוי של הודעות יש לטעון לרקונייזר את ההודעות שהקלטנו. בדוגמה שנביא כאן נטען לרקוגנייזר את ההקלטות הרצויות לנו עם הפקודה 30 . לדוגמה אם נרצה לטעון הודעה 7 נרשום 07 30 . ניתן לטעון מספר הודעות בפקודה אחת . לדוגמה: 30 02 03 04 05 06 07 08 אומרת לטעון הודעות 02 03 04 05 06 07 08 (מקסימום 7 הודעות !) . בהקלטות שביצעתי יש את ההודעות הבאות:

02 SHALOM

03 KADIMA

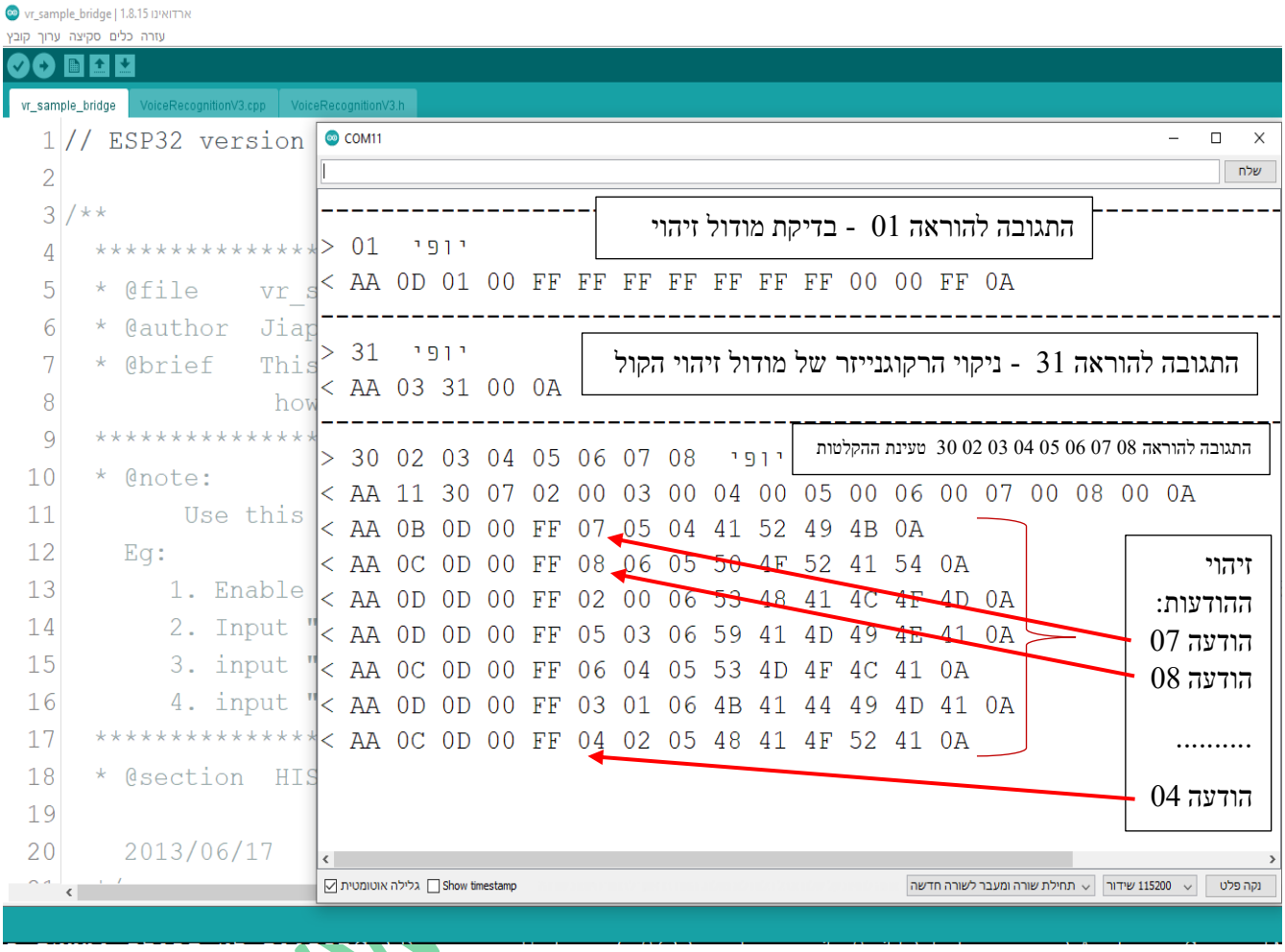
04 HAORA05 YAMINA

06 SMOLA

07 ARIK

08 PORAT

המסך שנקבל עם השמעת ההודעות הרצויות נראה באיור הבא :



איור 7 : מסך זיהוי מספר ההודעה .

עבור כל הודעה המתקבלת במחרוזת תשובה מהמודול זיהוי קול רואים באיבר השישי את מספר ההודעה.

6 ביבליוגרפיה (קישור לאתרים) :

1. https://blog.frankvh.com/2022/02/21/esp32-support-for-elechouse-voice-recognition-module-v3/
2. GitHub - elechouse/VoiceRecognitionV3: Arduino library for elechouse Voice Recognition V3 module
3. libref.pdf
4. Chat GPT