

פרויקט גמר

מגישות: מיה יעקובי ועמית יחזקאל

1. מבוא

פרויקט זה עוסק בפיתוח מערכת תוכנה שנועדה לשפר את התקשורת של אנשים על הספקטרום האוטיסטי המתמודדים עם ליקויי דיבור או דיבור לא ברור. אוכלוסייה זו מתמודדת לעיתים עם קושי משמעותי בהבעה מילולית, מה שיוצר פערים בתקשורת מול הסביבה הקרובה והמטפלים שלהם, ומוביל לאי הבנות ולתסכול רב. המערכת נועדה לספק פתרון טכנולוגי פשוט ונגיש, שיאפשר לסביבה להבין בצורה טובה יותר את כוונתם של המשתמשים באמצעות תרגום הדיבור שלו לפלט טקסטואלי ברור.

מטרת הפרויקט היא לפתח כלי לזיהוי דיבור לא טיפוסי באמצעות התאמה אישית לכל משתמש. המערכת בנויה משני מצבים מרכזיים: במצב הלמידה המשתמש (לעיתים יחד עם הורה או מטפל) מקליט מילים וביטויים קצרים, מתייג אותם על פי המשמעות (למשל "אבא", "כואב לי", "אני רוצה מים"), ובכך נבנה עבורו מחסן מילים קולי אישי. במצב השימוש, המשתמש מקליט את עצמו מדבר, והמערכת משווה את הקלט הקולי למאגר ההקלטות שנאספו בשלב הלמידה, ומציגה על המסך את המילה או הביטוי המזוהים כמתאימים ביותר.

המערכת מבוססת על עיבוד אותות קוליים וטכניקות בסיסיות של למידת מכונה או השוואת דפוסים, במטרה לזהות דמיון בין דגימות דיבור חדשות לבין הדוגמאות הקיימות במחסן המילים. במסגרת העבודה יתוכנן ויושם מנגנון לאיסוף, שמירה והשוואה של דגימות קול, לצד ממשק משתמש פשוט ונוח המותאם לשימוש במכשירים ניידים או מחשב. כחלק מתהליך הפיתוח יבוצעו בדיקות ראשוניות של דיוק הזיהוי והערכת תפקוד המערכת בתרחישים שונים.

האתגר המרכזי בפרויקט נובע מכך שמערכות זיהוי דיבור כלליות אינן מצליחות לזהות היטב דפוסי דיבור אישיים ולא טיפוסיים, ולכן אינן מספקות מענה מתאים למשתמשים אלה. המערכת המוצעת מתמודדת עם קושי זה באמצעות התאמה אישית: כל משתמש בונה מחסן מילים קולי אישי במסך הלמידה, שבו הוא מקליט מילים וביטויים ומקשר ביניהם לבין משמעותם, ולאחר מכן, במסך השימוש, המערכת משווה קלטים חדשים למחסן זה ומציגה את המילה המתאימה כטקסט. כך מתקבלת גישה פשוטה וממוקדת, המבוססת על עיבוד אותות בסיסי, התאמה פר משתמש וממשק ידידותי, ללא צורך במודלי זיהוי דיבור כלליים בלבד.

המערכת המוצעת יכולה לשמש בסיס להמשך פיתוח בתחום התקשורת התומכת והחליפית, ובפרט לפתרונות המותאמים לדפוסי דיבור אישיים ולא סטנדרטיים. בעתיד ניתן יהיה להרחיב את המחסן המילולי, לשפר את אלגוריתמי הזיהוי, ולשלב את המערכת עם כלים טיפוליים וחינוכיים קיימים. אנו שואפות לקדם טכנולוגיה שתפחית את קשיי התקשורת, תחזק את תחושת העצמאות של המשתמשים, ותסייע להם להביע את עצמם בצורה מובנת יותר בסביבתם היומיומית.

2. סקירת מתחרים

בשנים האחרונות פותחו מגוון פתרונות דיגיטליים שנועדו לתת מענה לקשיי תקשורת ודיבור בקרב אוכלוסיות בעלות צרכים מיוחדים, ובפרט אנשים עם דיבור לא טיפוסי או דיבור שאינו מובן לסביבה. חלק מהפתרונות מתבססים על טכנולוגיות מתקדמות של זיהוי דיבור והתאמת מודלים למשתמש, ואחרים משתמשים בתקשורת תומכת וחלופית באמצעות סמלים, טקסט וממשקים חזותיים. כדי להבין באופן מעמיק יותר מהו המצב הקיים בשוק, מהם היתרונות של הפתרונות הדיגיטליים שכבר פותחו בתחום, ומהם הפערים שנותרו בלתי פתורים עבור משתמשים עם דיבור לא טיפוסי, ערכנו סקירת מתחרים. סקירה זו מציגה מספר אפליקציות רלוונטיות בתחום.

Voiceitt

מערכת זיהוי דיבור לאנשים עם דיבור לא טיפוסי, המבוססת על טכנולוגיית זיהוי דיבור ולמידת מכונה כדי ללמוד את דפוסי הדיבור האישיים של המשתמש ולהמירם לטקסט או לדיבור ברור. המערכת מיועדת לאנשים עם לקויות דיבור שונות, ומטרתה לאפשר תקשורת עצמאית ונגישה יותר עם הסביבה.

Google Project Relate

אפליקציית תקשורת ניסיונית לאנדרואיד, המיועדת לאנשים עם דיבור לא טיפוסי. האפליקציה בונה מודל זיהוי דיבור מותאם אישית על בסיס הקלטות של המשתמש, ולאחר מכן מאפשרת לתמלל את דיבורו לטקסט, להשמיעו מחדש בקול ממוחשב ברור.

Proloquo2Go

אפליקציית תקשורת תומכת וחלופית (AAC) מותאמת אישית למכשירי iOS המיועדת להוות "קול" עבור אנשים שאינם מסוגלים לדבר. האפליקציה מבוססת על מערכת סמלים, אוצר מילים מובנה ומנגנון המרת טקסט לדיבור, ומותאמת במיוחד לאנשים עם קשיי תקשורת כגון אוטיזם, שיתוק מוחין ותסמונת דאון, תוך אפשרות להתאמה גמישה לצורכי המשתמש.

הטבלה הבאה מסכמת ומשווה בין שלוש המערכות שהצגנו לפי מספר קריטריונים מרכזיים:

מערכת קיימת	סוג הפתרון	אופן הקלט	אופן הפלט	למידה לדפוס דיבור אישי	אפשרות להוספת אוצר מילים אישי	פלטפורמה	תמיכה בשפה העברית
Voiceitt	מערכת זיהוי דיבור לא טיפוסית המבוססת על בינה מלאכותית ולמידת מכונה	קלט קולי	טקסט קריא ודיבור ממוחשב	✓	✓	יישום ווב הפועל בדפדפן	✓
Google Project Relate	אפליקציית תקשורת מבוססת מודל זיהוי דיבור מותאם אישית	קלט קולי	שלושה מצבים עיקריים: תמלול לטקסט, חזרה בקול ממוחשב ברוח, העברת הפקודות ל-Google Assistant	✓	✓	אפליקציית אנדרואיד (בטא) בלבד	✗
Proloquo2Go	אפליקציית תקשורת תומכת וחלופית מבוססת סמלים וטקסט	בחירת סמלים, מילים וביטויים מתוך לוחות תקשורת בממשק חזותי	הצגת סמלים וטקסט לצד המרת טקסט לדיבור	✗	✓	מכשירי iOS	✗

מן ההשוואה בין המערכות הקיימות בטבלה עולה כי קיימים כיום שני כיוונים עיקריים: מערכות זיהוי דיבור מותאם אישית (כמו Voiceitt ו-Google Project Relate) המבוססות על מודלים מורכבים של בינה מלאכותית, ואפליקציות תקשורת תומכת וחלופית כמו Proloquo2Go העוקפות את הערוץ הקולי ומציעות תקשורת באמצעות סמלים וטקסט. למרות שמערכות אלו מאפשרות התאמה אישית ברמות שונות ואף הרחבת אוצר המילים, רובן אינן מותאמות באופן מובנה לשימוש בעברית, וחלקן אינן עושות שימוש בדיבור הקיים של המשתמש אלא מחליפות אותו בממשק חזותי.

המערכת המוצעת בפרייקט זה מתמקדת בזיהוי דיבור לא טיפוסית בשפה העברית על בסיס מחסן מילים קולי אישי, שנבנה בתהליך למידה עם המשתמש, ובכך משלבת התאמה אישית לדפוס הדיבור של המשתמש עם פיתוח טכנולוגי ישים ומאפשרת להמיר דיבור לא ברור לפלט טקסטואלי מובן בשפת המשתמש.

3. סקירה ספרותית

בפרק זה מוצגת סקירת ספרות העוסקת במאפייני הדיבור אצל אנשים על הרצף האוטיסטי, ובאתגרים והפתרונות המחקריים הקשורים בזיהוי דיבור לא טיפוסי. מטרת הסקירה היא לבסס רקע תיאורטי רלוונטי לפרויקט, המבוסס על שבעה מקורות עדכניים, תוך הדגשת הצורך בפתרון מותאם משתמש.

3.1 מאפייני שפה, דיבור ופרוזודיה באוטיזם

המחקר על דיבור אצל אנשים עם אוטיזם מצביע על מורכבות רבה ושונות גבוהה בין דוברים שונים. Vogindroukas ועמיתיו [1] מציינים כי קשיים בשפה (כמו אוצר מילים מצומצם וקשיים בהבעה ובהבנה) מלווים לעיתים קרובות בקשיים פרוזודיים (אינטונציה, קצב, והטעמה- האופן שבו אומרים את המילה), המשפיעים על המובנות. לפי ממצאיהם, דוברים על הרצף עשויים להציג דיבור מונוטוני (ללא שינויי גובה קול), שינויי קצב קיצוניים, והפסקות שאינן מותאמות למשפט, דבר המקשה על הבנת הכוונה.

מחקרם של Godel ועמיתיה [2], המבוסס על ניתוח אקוסטי (מדידה מדויקת של מאפייני קול), מצא קשר ישיר בין מדדים כמו גובה קול, שונות בגובה ומשך הברות לבין רמת חומרת האבחון. הם הראו כי ניתן לזהות "חתימות קול" אופייניות לאנשים על הרצף, מה שמצביע על כך שההבדלים בדיבור אינם רק סובייקטיביים אלא ניתנים למדידה אובייקטיבית.

המטא אנליזה של Ma ועמיתיה [3] מאשרת כי שילוב של כמה מדדי פרוזודיה מאפשר להבחין בין דוברים על הרצף לבין דוברים טיפוסיים בדיוק גבוה יותר. עם זאת, הממצאים גם מדגישים את האתגר המרכזי: שונות גבוהה בין דוברים, עד כדי כך שקשה מאוד לבנות מודל אחיד שמתאים לכולם. ממצא זה תואם ממצאים מעולם הדיסארטריה ולקויות הדיבור הנוספות, ומבסס את הצורך בגישה מותאמת משתמש.

בכך, שלושת המחקרים [1-3] מספקים תמונה רחבה. יש מאפיינים אקוסטיים ושפתיים ברורים שמבדילים דיבור של אנשים עם ASD אך השונות הרבה בין דוברים מונעת שימוש במודל אחד כללי. נקודה זו מהווה בסיס ישיר לצורך בפתרון אישי כמו זה שמוצע בפרויקט.

3.2 זיהוי דיבור לא טיפוסי ומערכות לזיהוי דיבור

מחקרים בתחום זיהוי הדיבור מצביעים על כך שמערכות זיהוי דיבור אוטומטיות ASR (מערכות שממירות דיבור לטקסט) מתקשות מאוד להתמודד עם דיבור לא טיפוסי. Gale ועמיתיו [4] מצאו כי מערכות אלו מציגות שיעור שגיאה גבוה במיוחד אצל ילדים עם ASD או לקות שפה, גם כאשר משתמשים במודלים מתקדמים. הם מסבירים זאת בכך שמודלים אלו מאומנים בעיקר על דיבור "תקין", ולכן אינם מתאימים לאוכלוסייה עם דפוסי דיבור ייחודיים. הם מציעים שימוש ב- Transfer Learning (התאמת מודל קיים באמצעות דוגמאות אישיות), אך גם גישה זו מוגבלת כאשר מספר הדוגמאות קטן.

תחום הדיסארטריה מציג תובנות דומות. Yadav ועמיתיו [6] מציינים כי השונות הגבוהה בדיבור והמובנות הנמוכה מקשים על המערכת לזהות מילים בצורה מדויקת. הם מדגישים שמערכות ASR רגילות אינן מסוגלות להתמודד עם דיבור לקוי ללא התאמות נרחבות לדובר הספציפי.

Almadhor ועמיתיו [7] פיתחו מודל End-to-End (מודל המבצע ישירות מעבר מקול לטקסט ללא שלבים נוספים) שהצליח לשפר את הדיוק בזיהוי דיבור דיסארטרי. עם זאת, מודל זה דורש כמויות עצומות של דאטה ומשאבי חישוב, מה שהופך אותו ללא מעשי עבור פרויקטים קטנים ושפות בעלות מעט נתוני אימון, כמו עברית.

בנוסף Vaezipour ועמיתיו [5], שנבחנו בפרק הטכנולוגיות, מציינים שגם אפליקציות מתקדמות אינן מסוגלות להתמודד היטב עם זיהוי דיבור לא טיפוסי משום שרובן אינן משתמשות בניתוח דיבור אישי אלא במבנים מוכנים מראש או תרגול קבוע.

לכן, מכל ארבעת המחקרים [4-7] עולה כי כדי לזהות דיבור לא טיפוסי, יש צורך במערכת הלומדת את הדיבור של המשתמש עצמו ולא במודלים כלליים.

3.3 אפליקציות ומערכות לתמיכה בדיבור ותקשורת

Vaezipour ועמיתיו [5] בחנו יישומים דיגיטליים המיועדים לשיקום תקשורת ומצאו כי למרות מגוון הכלים, רובם אינם מבוססי ראיות מחקריות ואינם מותאמים באמת לדיבור אישי. ממצא זה תואם את מסקנות מחקרי ה-ASR [4-7], המדגישים את הקושי לזהות דיבור לא טיפוסי ללא התאמה אישית.

המחקרים בתחום הפרוזודיה [1-3] משלימים תמונה זו. הם אומרים כי אם דיבור אצל אנשים על הרצף מאופיין בשונות גבוהה, אז אפליקציות שאינן מתייחסות לדפוס הדיבור האישיים של המשתמש לא יצליחו לתת מענה אמיתי.

לכן, מכלל הספרות עולה שהאתגר אינו רק טכנולוגי אלא גם תפיסתי. נדרש פתרון שמבוסס על דיבור אישי, בשפה הרלוונטית למשתמש.

3.4 סיכום הפרק והצגת הפער

המחקרים מציגים תמונה עקבית-אנשים על הרצף מציגים שונות רבה במאפייני הדיבור שלהם [1-3], ומערכות זיהוי דיבור אינן מצליחות להתמודד עם שונות זו [4-7]. גם כלים טכנולוגיים קיימים אינם מבוססים על התאמה לדיבור אישי [5].

מכאן עולה שישנו פער מרכזי. אין כיום פתרון פשוט, מותאם משתמש, המבוסס על דיבור אישי ובוודאי שלא בשפה העברית, שבה קיימים פחות נתוני אימון.

הפרויקט נותן מענה ישיר לפער זה באמצעות מחסן מילים קולי אישי המבוסס על הקלטות של המשתמש עצמו, ומתאים עצמו לדפוס הדיבור שלו.

- [1] Vogindroukas, I., Stankova, M., Chelas, E. N., & Proedrou, A. (2022). Language and speech characteristics in autism. *Neuropsychiatric Disease and Treatment*, 18, 2367-2377.
- [2] Godel, M., Robain, F., Journal, F., Kojovic, N., Latrèche, K., Dehaene-Lambertz, G., & Schaer, M. (2023). Prosodic signatures of ASD severity and developmental delay in preschoolers. *npj Digital Medicine*, 6, 99.
- [3] Ma, W., Xu, L., Zhang, H., & Zhang, S. (2024). Can natural speech prosody distinguish autism spectrum disorders? A meta-analysis. *Behavioral Sciences*, 14(2), 90.
- [4] Gale, R., Chen, L., Dolata, J., van Santen, J., & Asgari, M. (2019). Improving ASR systems for children with autism and language impairment using domain-focused DNN transfer techniques. In *Interspeech 2019* (pp. 11–15).
- [5] Vaezipour, A., Campbell, J., Theodoros, D., & Russell, T. (2020). Mobile apps for speech-language therapy in adults with communication disorders: Review of content and quality. *JMIR mHealth and uHealth*, 8(10), e18858.
- [6] Yadav, S. S., Yadav, D. M., & Desai, K. R. (2023). A comprehensive survey of automatic dysarthric speech recognition. *International Journal of Informatics and Communication Technology*, 12(3), 242–250.
- [7] Almadhor, A., Irfan, R., Gao, J., Saleem, N., & Rauf, H. T. (2023). E2E-DASR: End-to-end deep learning-based dysarthric automatic speech recognition. *Expert Systems with Applications*, 222, 119797.