

# פרויקט גמר

מגישות: מיה יעקובי ועמית יחזקאל

## 1. מבוא

פרויקט זה עוסק בפיתוח מערכת תוכנה שנועדה לשפר את התקשרות של אנשים על הספקטרום האוטיסטי המתומדים עם ליקוי דיבור או דיבור לא ברוחו. אוכלוסייה זו מתמודדת לעיתים עם קשיים משמעותיים בהבנה מילולית, מה שגורם פערים בתקשורת מול הסביבה הקרויה והמטפלים שלהם, ומוביל לאי הבנות ולתסכול רב. המערכת נועדה לספק פתרון טכנולוגי פשוט ונגיש, שיאפשר לסייע להבין בצורה טובה יותר את בונכם של משתמשים באמצעות תרגום הדיבור שלהם לפטט טקסטואלי ברוחם.

מטרת הפרויקט היא לפתח כלizi דיבור לא טיפוסי באמצעות התאמת אישית לכל משתמש. המערכת בנויה משני מרכיבים מרכזיים: במצב הלמידה המשמש (לעתים יחד עם הורה או מטפל) מקלט מילים וביטויים קצרים, מתייג אותם על פי המשמעות (למשל "אבא", "בוֹאָבָּא", "אני רוצה מים"), ובקר נבנה עבورو מחסן מילים קולי אישי. במצב השימוש, המשתמש מקלט את עצמו מדובר, והמערכת משווה את הקלט הקולי למגר הקלטות שנאספו בשלב הלמידה, ומציגה על המסך את המילה או הביטוי המדוים 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 ויישם מגננון לאיסוף, שמיירה והשווה של דגימות קול, לצד שימוש משתמש פשוט ונוח המותאם לשימוש במכשירים ניידים או מחשב. בחלק מהתהליך הפיתוח יבוצעו בדיקות ראשוניות של דיקן הדיזהוי והערכת תפקוד המערכת בתרחישים שונים.

האתגר המרכזי בפרויקט נובע מכך שמערכות דיזהוי דיבור כליליות אינן מצליחות לזהות היבט דפוסי דיבור אישיים ולא טיפוסיים, ולכן אינן מספקות מענה מתאים למשתמשים אלה. המערכת המוצעת מתמודדת עם קשיים זה באמצעות התאמת אישית: כל משתמש בונה מחסן מילים קולי אישי במסך הלמידה, שבו הוא מקלט מילים וביטויים ומקשר ביניהם בין משמעותם, ולאחר מכן, במסך השימוש, המערכת משווה קלטים חדשים למחסן זה ומציגה את המילה המתאימה בטקסט. בכך מתקבלת גישה פשוטה ומשמעותית, המבוססת על עיבוד אותות בסיסי, התאמת פרט משתמש וממשק ייחודי, ללא צורך במודול דיזהוי דיבור בלבד.

המערכת המוצעת יכולה לשמש בסיס להמשך פיתוח בתחום התקשרות התומכת והחליפית, ובפרט לפתרונות המותאמים לדפוסי דיבור אישיים ולא סטנדרטיים. בעתיד ניתן יהיה להרחיב את המחסן המילולי, לשפר את אלגוריתמי הדיזהוי, ולשלב את המערכת עם כלים טיפולים וחינוכיים קיימים. אנו שואפות לקדם טכנולוגיה שתפחית את קשיי התקשרות, תחזק את תחושת העצמאות של המשתמשים, ותסייע להם להביע את עצםם בצורה מובנת יותר בסביבת חיים יומיומיות.

## 2. סקירות מתחברים

בשנים האחרונות פותחו מגוון פתרונות דיגיטליים שנועדו לחתת מענה לקשיי תקשורת ודיבור בקשר אוכלסיות בעלות צרכים מיוחדים, ובפרט אנשים עם דיבור לא טיפוסי או דיבור שאינו מובן לסביבה. חלק מהפתרונות מtabסים על טכנולוגיות מתקדמות של זהוי דיבור והאמת מודלים למשתמש, ואחרים משתמשים בתקשורת תומכת וחולפית באמצעות סמלים, טקסט וממשקים חזותיים. כדי להבין באופן עמוק יותר מהו המצב הנוכחי בשוק, מהם היתרונות של הפתרונות הדיגיטליים שכבר פותחו בתחום, ומהם הבעיות שנוטרו בתחום פתרונים עבור משתמשים עם דיבור לא טיפוסי, ערכנו סקירות מתחברים. סקירה זו מציגה מספר אפליקציות רלוונטיות בתחום.

### Voiceitt

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

### Google Project Relate

אפליקציה תקשורת ניסונית לאנדרואיד, המיועדת לאנשים עם דיבור לא טיפוסי. האפליקציה בונה מודל זהוי דיבור מותאם אישית על בסיס הקלטות של המשתמש, ולאחר מכן מאפשרת לתמוך את דיבורו לטקסט, להשמיעו מחדש ממוחשב ברוח.

### Proloquo2Go

אפליקציה תקשורת תומכת וחולפית (AAC) מותאמת אישית למ襇רי iOS המיועדת להוות "קול" עבור אנשים שאינם מסוגלים לדבר. האפליקציה מבוססת על מערכת סמלים, אוצר מילים מובנה ומנגן המרת טקסט לדיבור, ומותאמת במיוחד לאנשים עם קשיי תקשורת כגון אוטיזם, שיתוק מוחין ותסמונת דאון, תוך אפשרות להתאמאה גמישה לצורכי המשתמש.

הטבלה הבאה מסכמת ומשווה בין שלוש הממערכות שהציגו לפי מספר קритריונים מרכזיים:

מערכת קיימת	סוג הפתרון	אופן הקלט	טקסט קריא ודיבור ממוחשב	למיזה לדפוס דיבור אישי	אפשרות אוצר מילים אישי	פלטפורמה	תמיינה בשפה העברית
Voiceitt	מערכת זיהוי דיבור לא טיפוסי הmbוססת על בינה מלאכותית ולמידת מכונה	קלט קולי	טקסט קריא ודיבור ממוחשב	✓	✓	שימוש ווב הפועל בטלפון	✓
Google Project Relate	אפליקציה זיהוי דיבור מבוססת מודל זיהוי דיבור מותאמת אישית	קלט קולי	שלושה מצבים עיקריים: תמלול לטקסט, חרורה בקול ממוחשב ברורה, הverbת הפקדות ל- Google Assistant	✓	✓	אפליקציה זיהוי דיבור אנדרואיד (בטא) בלבד	✗
Proloquo2Go	אפליקציה זיהוי דיבור תקשורת תומכת וחלופית מבוססת סמלים וtekסט	בחירה סמלים, מילים וביוטיים מתוך לוחות תקשרות במסך חוותי	הצגת סמלים וטקסט לצד המרה טקסט לדיבור	✗	✓	מכשורי iOS	✗

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

ה מערכת המוצעת בפרויקט זה מתמקדת בזיהוי דיבור לא טיפוסי בשפה העברית על בסיס מחסן מילים קולי אישי, שבננה בתהליך למידה עם המשתמש, ובכך משלבת התאמה אישית לדפוס הדיבור של המשתמש עם פיתוח טכנולוגיות יישום ומאפשרת להמיר דיבור לא ברור לפט טקסטואלי מובן בשפת המשתמש.

## 3. סקירה ספרותית

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

### 3.1 מאפייני שפה, דיבור ופרוזודיה באוטיסטים

המחקר על דיבור אצל אנשים עם אוטיזם מצביע על מורכבות רבה ושותות גבוהה בין דברים שונים. Vogindroukas ועמיתו [1] מצינים כי קשיים בשפה (כמו אוצר מילים מצומצם וקשיים בהבנה ובהבנה) מלווים לעיתים קרובות בקשישים פרוזודים (אינטונציה, קצב, והטעה) – האופן שבו אומרים את המילה), המשפיעים על המובנות. לפי ממצאיםם, דברים על הרצף עשויים להציג דיבור מונוטוני (לא شيئا' גובה קול), שינוי קצב קיצוניים, והפסקות שאינן מותאמות למשפט, דבר המקשה על הבנת הכוונה.

מחקר של Godel ועמיתה [2], המבוסס על ניתוח אקוסטי (מדידה מדעית של מאפייני קול), מצא קשר ישיר בין מזדים כמו גובה קול, שותות בגובה ומשך הברות לבין רמת חומרת האבחון. הם הראו כי ניתן להזיהות "חתימות קול" אופייניות לאנשים על הרצף, מה שמצוין על כך שהבדלים בדיבור אינם רק סובייקטיביים אלא ניתנים למדידה אובייקטיבית.

המطا אנליזה של Ma ועמיתה [3] מאשרת כי שילוב של כמה מזדי פרוזודיה מאפשר להבחן בין דברים על הרצף לבין דברים טיפוסיים בדיק גובה יותר. עם זאת, הממצאים גם מדגישים את האתגר המרכזי: שותות בגובה בין דברים, עד כדי כך שקשה מאוד לבנות מודל אחיד שמתאים לכלום. ממצא זה תואם ממצאים מעולם הדיסארטريا ולקיים הדיבור הנוסף, ומבסס את הצורך בגישה מותאמת משתמש.

בקיר, שלושת המחוקרים [3-1] מספקים תמונה רחבה. יש מאפיינים אקוסטיים ושפתיים בחרורים שמבדילים דיבור של אנשים עם ASD אך השונות הרבה בין דברים מונעת שימוש במודל אחד בלבד. נקודה זו מהוות בסיס ישיר לפתרון אישי במו זה שמצוע בפרויקט.

### 3.2 זיהוי דיבור לא טיפוסי ומערכות לזייהו דיבור

מחקרים בתחום זיהוי הדיבור מצביעים על כך שמערכות זיהוי דיבור אוטומטיות ASR (מערכות ששמירויות דיבור לטקסט) מתקשות מאוד להתמודד עם דיבור לא טיפוסי. Gale ועמיתה [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 מאמנים בעיקר על דיבור "תקין", וכן אינם מתאימים לאוכלוסייה עם דפוסי דיבור ייחודיים. הם מציעים שימוש ב-Transfer Learning (התאמת מודל קיים באמצעות דוגמאות אישיות), אך גם גישה זו מוגבלת באשר מספר הדוגמאות קטן.

תחום הדיסארטריה מציג תובנות דומות. Yadav ועמיתיו [6] מציינים כי השונות הגדולה בדיור והמודעות הנמוכה מקשים על המعرفת להזות מילים בצורה מדויקת. הם מדגימים שמערכות ASR רגילות אינן מסוגלות להתמודד עם דיור לQUI לא התאמות נרחבות לדיבור הספציפי.

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

בנוסף Zoueziour Vaezipour ועמיתו [5], שנבחנו בפרק הטכנולוגיות, מציינים שגם אפליקציות מתקדמיות אינן מסוגלות להתמודד היטב עם זיהוי דיבור לא טיפוסי משומם שרובן אינן משתמשות בניתוח דיבור אישי אלא במבנהים מוכנים מראש או תרגול קבוע.

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

### 3.3 אפליקציות ומערכות לתמיכה בדיור ותקשות

Vaezipour ועמיתו [5] בוחנים יישומים דיגיטליים המיעדים לשיקום תקשורת ומצאו כי למגוון הכלים, רובם אינם מבוססי ראיות מחקריות ואינם מותאמים באמת לדיבור אישי. ממצא זה תואם את מסקנות מחקרי ה-ASR [7-4], המדגימים את הקושי להזות דיבור לא טיפוסי ללא התאמת אישית.

החוקרים בתחום הפרוזודיה [3-1] משלימים תמונה זו. הם אומרים כי אם דיור אצל אנשים על הרצף מאופיין בשונות גבואה, אז אפליקציות שאינן מתייחסות לדפוסי הדיבור האישיים של המשתמש לא יכולים לחתן מענה אמיתי.

לכן, מכל הספרות עולה שהאתגר אינו רק טכנולוגי אלא גם תפיסתי. נדרש פתרון שمبוסס על דיור אישי, בשפה הרלוונטית למשתמש.

### 3.4 סיכום הפרק והציג הפער

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

מכאן עולה שישנו פער מרכזי. אין ביום פתרון פשוט, מותאם משתמש, המבוסס על דיור אישי ובוודאי שלא בשפה העברית, שבה קיימים פחות נתונים אימון.

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

**רשימת מקורות (APA):**

- [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.