

תקציר

כיום, מערכות לאימות דובר (speaker verification systems), משתמשות במרחב מאפיינים משותף (common feature space) לכל הדוברים. מרחב מאפיינים משותף זה מכיל בד"כ את המאפיינים הקפסטרליים (cepstral features) אשר מנוצלים במערכות של זיהוי דיבור. מחקר זה מבוסס על ההנחה שלכל דובר יש את מרחב המאפיינים ה"אופטימלי" שלו, אשר מבדיל אותו משאר הדוברים בצורה אופטימלית. מטרת עבודה זו היא להדגים את חשיבות השימוש במרחב מאפיינים אינדיבידואלי (individual feature space) במערכות מודרניות לאימות דובר המבוססות HMM ו-GMM.

במחקר זה, פותח קריטריון חדש (בשם RRC - recognition related criterion) עבור בחירת מאפיינים (feature selection), אשר מתאים למשימות אימות דובר ואשר קורלטיבי עם שיעור הזיהוי (recognition rate). מספר פרוצדורות לבחירת מאפיינים מומשו ונבחנו יחד עם הקריטריון החדש, כגון: k-best, forward, Sequential Floating Forward Sequence (SFFS), Dynamic Programming (DP). מימשנו גם שתי מערכות לאימות דובר אשר מכילות את האלגוריתם לבחירת מאפיינים אינדיבידואליים. הראשונה היא מערכת אימות דובר תלוית-טקסט (text-dependent) המבוססת על מסווג מסוג מודל מרקובי חבוי (Hidden Markov Model – HMM) והשנייה היא מערכת אימות דובר בלתי תלוית-טקסט (text-independent) המבוססת על מסווג מסוג מודל בלילת גאוסיינים (Gaussian Mixture Model – GMM).

המערכת המבוססת HMM נבחנה על בסיס נתונים text-dependent. שיפור משמעותי בתוצאות זיהוי הושג על ידי שימוש במרחבי מאפיינים אינדיבידואליים לעומת שימוש במרחב (MFCC) Mel Frequency Cepstrum Coefficients ה"סטנדרטי" ($12 \text{ MFCC} + 12 \Delta \text{MFCC}$). Equal Error Rate (EER) בערך של 0.7% התקבל עבור מרחב המאפיינים MFCC. עבור אותם תנאים, המערכת המבוססת על בחירת מאפיינים אינדיבידואליים (במימד 24) הניבה EER של 0.48% בלבד. נמצא ששתי הפרוצדורות הטובות ביותר לבחירת מאפיינים הם ה-DP וה-SFFS. בכל אופן, ה-SFFS יעיל יותר מן ה-DP במובן של עומס חישוב. מערכת זו נבחנה גם על בסיס נתונים text-dependent מורעש. נמצא

שמערכת זו המשתמשת במרחב מאפיינים אינדיבידואליים מאד רגישה לשינויים בתנאי סביבה, כלומר יחסי אות לרעש שונים בין בסיס הנתונים לאימון ובין בסיס הנתונים לזיהוי. המערכת המבוססת GMM נבחנה על בסיס נתונים text-independent. שיפור משמעותי בתוצאות זיהוי התקבל גם כאן על ידי שימוש במרחבי מאפיינים אינדיבידואליים לעומת שימוש במרחב MFCC. EER בערך של 6.14% התקבל עבור מרחב המאפיינים MFCC. עבור אותם תנאים, המערכת המבוססת על בחירת מאפיינים אינדיבידואליים הניבה EER של 4.15% בלבד.

מילות מפתח

זיהוי דובר, אימות דובר, בחירת מאפיינים, מרחב מאפיינים אינדיבידואלי, מודל מרקובי חבוי, מודל בלילת גאוסיינים.