

ISRAEL DEFENSE Tech

מלמדים מכונות להבין דיבור

טכנולוגיות זיהוי דיבור הופכות להיות דבר שבשגרה. בדרך לעוד מהפיכה?

כך פרצתי לארגון ביטחוני
פריצה למערכת מסווגת לא תאמינו כמה זה קל

מחשב בתוך מבחנה
בטכניון מפתחים מחשבים מולקולריים

המלחמה על החלל
אירופה וארה"ב בקרב על הגדרת הגבול האחרון

רובוטים בגודל חרקים
איך הצבא ישתמש בהם



אילוסטרציה: Shutterstock

מלמדים מכונות להבין דיבור

התפתחות טכנולוגיית זיהוי דיבור, תנועות גוף ורגשות יחד עם אינטליגנציה מלאכותית, הופכות את המכונות לחכמות - כמעט כמו בני אדם. פרופסור עמי מויאל מסביר מדוע אנחנו בפתחו של עולם חדש

מאת פרופ' עמי מויאל

בקשות חיפוש בדיבור ושירות של עוזר אישי (Google Voice Actions), ואילו מיקרוסופט מובילה כיוון של תקשורת טבעית אדם-מכונה (NUI - Natural User Interface), הכוללת זיהוי דיבור.

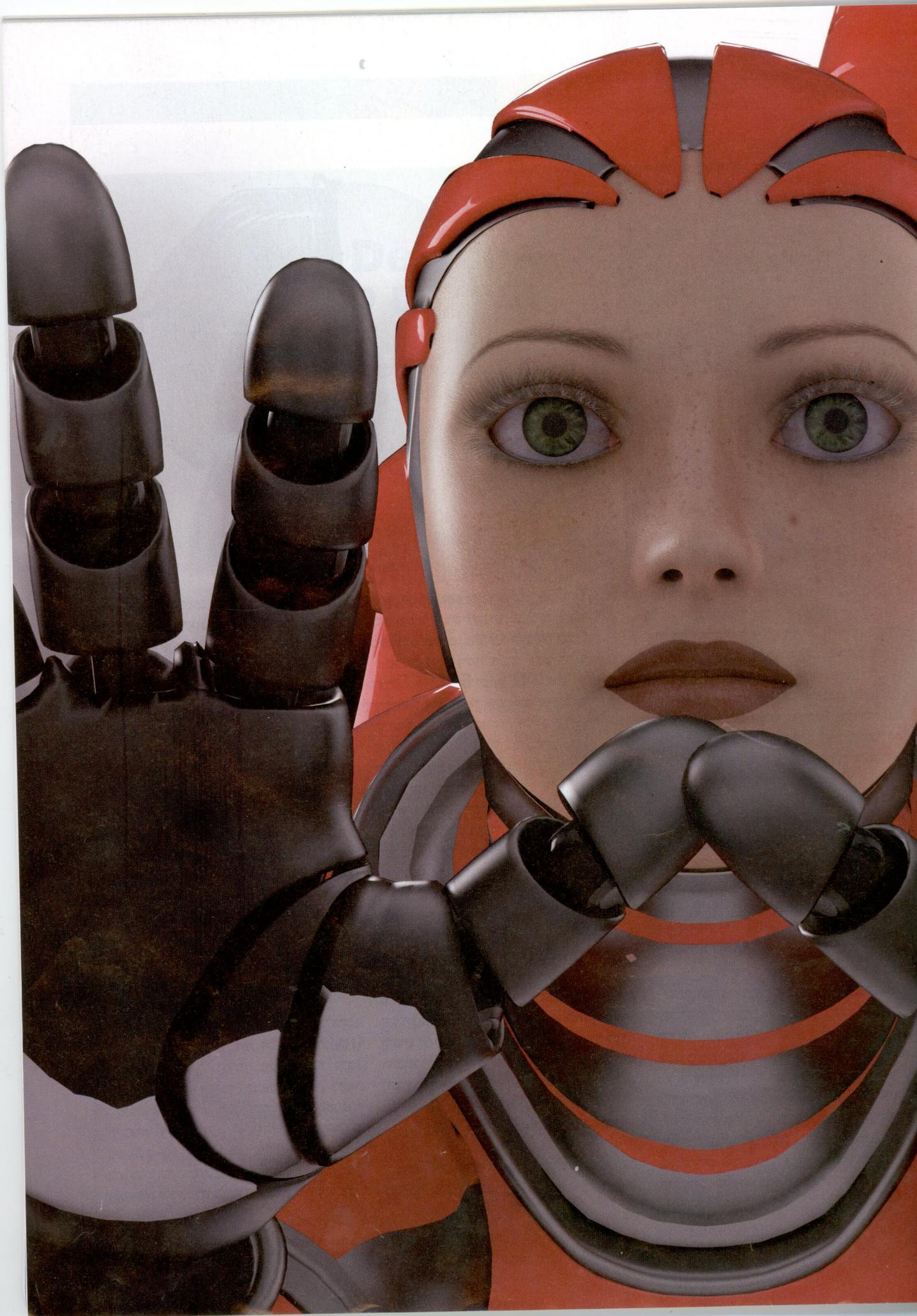
למעשה, שלוש הענקיות עושות שימוש בטכנולוגיה של זיהוי דיבור כחלק מממשק אדם-מכונה טבעי וחכם. סביר להניח, כי התפתחות מגמה זו בעתיד הקרוב תביא לחידה משמעותית יותר של טכנולוגיית זיהוי דיבור לחיינו. נשתמש בטכנולוגיה הזאת בעת הפעלת מכשיר סלולארי, מחשב, קונסולות משחקים או כלי רכב. הממשקים בין האדם למכונה

חום וזהו דיבור של אדם על ידי מחשב וזכה לאחרונה להתעניינות רבה, עם השקת שירותים מבוססי ממשק קולי על ידי ענקיות הטכנולוגיה.

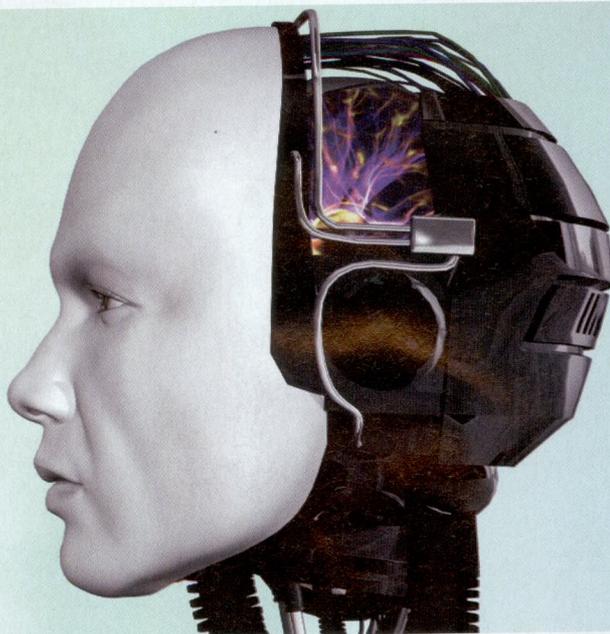
כך, אפל השיקה אפליקציה של עוזר אישי (Siri), המאפשרת תקשורת באמצעות דיבור עם מכשיר האיפון והפעלת אפליקציות שונות כמו קביעת פגישות, עדכון תזכורות, רשימת משימות, קבלת מידע על מזג האוויר ותחזית עניין נוספים.

חברת גוגל השיקה גם היא שירות של חיפוש קולי (Google Voice), והוא מאפשר הזנת

ת



**"המשמעות היא,
שבסיס הנתונים
לאימון המערכת
צריך להיות מייצג
בכל האספקטים של
פעולת הזיהוי, וצריך
לייצג את הדוברים
שיעשו שימוש
בשירות מבחינת
הדיאלקטים, הגילאים
והמין (מגדר)"**



איילון דרור: Shutterstock

האלגוריתם מבצע חיפוש מבוסס תכנות דינמי אחר סדרת המילים בעלת ההסתברות המירבית, תוך שימוש במודלים אקוסטיים, לקסיקון מילים ומודל לשוני. המודלים האקוסטיים הינם מודלי מרקוב המאפשרים חישוב ההסתברות לקבלת מאפיינים שמוצו מאות הדיבור. לקסיקון המילים כולל מילון המתואר כסדרות של פונמות, והמודל הלשוני מתאר את ההסתברויות לקבלת סדרות של מילים.

בעוד שניתן לראות כי הטכנולוגיה הגיעה לבשלות המאפשרת שימוש ביישומים מסחריים שונים, הרי שישנם עדיין מספר נושאים מאתגרים, הקשורים בעיקר לחסינות הטכנולוגיה והפיכתה למוצר שימושי באפליקציות שונות.

אחד האתגרים עימם מתמודדת הטכנולוגיה קשור לירידת הביצועים בתנאי רעש: רמת זיהוי הדיבור בסביבות רועשות, כמו רכב נוסע או מקום ציבורי, היא נמוכה יותר, ולפעמים אף אינה מאפשרת הפעלה של שירות מבוסס זיהוי דיבור. כל שוני בין תנאי האימון לתנאי הזיהוי מוביל לירידה בביצועים. המשמעות היא, שבסיס הנתונים לאימון המערכת צריך להיות מייצג בכל האספקטים של פעולת הזיהוי, וצריך לייצג את הדוברים שיעשו שימוש בשירות מבחינת

השימוש במחשוב מתקדם, זמין וזול יותר, השקה של מוצרים ושירותים מבוססי זיהוי דיבור בשווקים שונים, בתהליך שנמשך עד היום.

טכנולוגיית זיהוי דיבור היא רב-תחומית, ועושה שימוש בגישות ובתחומים רבים כמו עיבוד אותות לצורך מיצוי מאפיינים מתוך אות הדיבור, מידול סטטיסטי (לרוב על ידי

מודל מרקוב חבו) של הפונמות, שהן יחידות בסיס לשוניות, מהן ניתן להרכיב את כל מילות השפה, תהליכי אימון אוטומטיים (מונחים ולא מונחים) של המודלים הסטטיסטיים מתוך בסיס נתונים בדיבור בגישה מונחית מידע, מידול לשוני של השפה, המתאר את ההקשרים הסמוכים והרחוקים יותר בין מילות השפה

(ההסתברות שמילים יופיעו בסמיכות למילים אחרות) ואלגוריתמים של חיפוש לצורך ביצוע חיפוש יעיל ואופטימלי של סדרת המילים בעלת ההסתברות הגבוהה ביותר, בהינתן אות דיבור והמידול האקוסטי והלשוני.

אתגר הטכנולוגיה

אלגוריתם החיפוש מקבל בכניסתו אות דיבור, ומייצר ביציאתו סדרת מילים טקסטואלית.

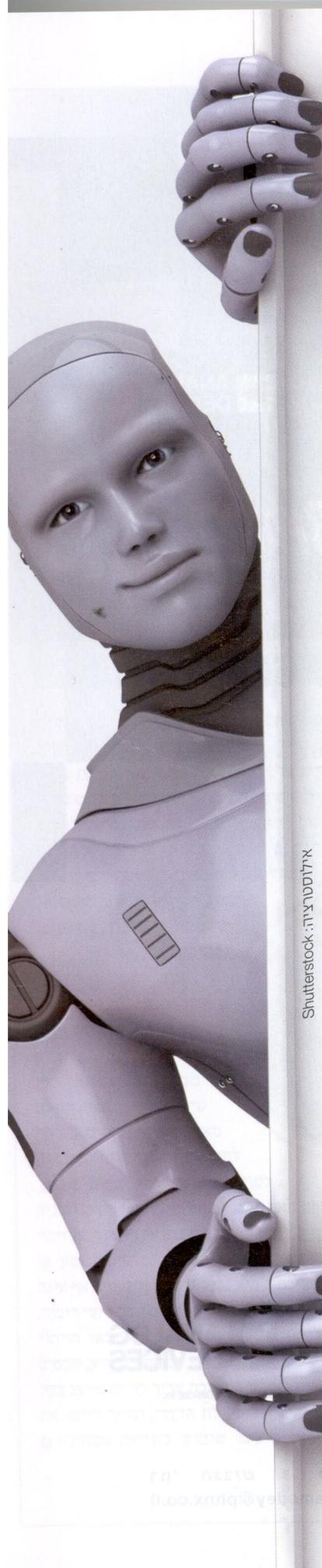
נעשים מתוחכמים יותר, וכוללים זיהוי דיבור כמו גם זיהוי תנועות גוף ומרכיבים נוספים, בהם נעשה שימוש בתקשורת בין אדם לאדם.

לומדים להבין דיבור

השאלה המתבקשת היא מדוע דווקא עכשיו? הרי מדברים על חדירה של טכנולוגיית זיהוי דיבור כבר שנות דור, בעוד שלמעשה חווינו בשנים האחרונות רק גידול איטי (אם כי קבוע) בשימוש באפליקציות אלו.

הסתכלות לאחור על התפתחות טכנולוגיית זיהוי דיבור משקפת תהליך הבשלה איטי, המזכיר לימוד דיבור של ילד. ההתחלה כללה זיהוי דיבור, הכולל מילים בודדות במילון קטן יחסית, שתלוי במי שאומר את הדברים, ורק לאחר מכן התבצע מעבר לאופן פעולה בלתי תלוי בדובר, והגדלה של מילוני הזיהוי לאלפי מילים עד כדי זיהוי דיבור ספונטני במילונים בני עשרות אלפי מילים.

הפעילות המחקרית בתחום זיהוי דיבור החלה לפני כ-50 שנה במעבדות המחקר של IBM ו-Bell בארה"ב. דחיפה משמעותית קיבל התחום בשנות ה-70 של המאה הקודמת בזכות תמיכה של משרד ההגנה האמריקאי, במסגרת תוכנית DARPA לעידוד מחקר מתקדם. בשנות ה-80 החל שימוש במודלים סטטיסטיים בתחום זיהוי דיבור, שאיפשר הגדלה משמעותית וגמישות של מילון הזיהוי. בשנות ה-90 איפשר



אילוסטרציה: Shutterstock

יהיה להרחיב את האפליקציה ליתר המכשירים בבית. ומה ברכב? תתאפשר הפעלה של מערכות לא קריטיות באמצעות דיבור.

בסוג השני ניתן לעשות שימוש בזיהוי דיבור בכל מקום שבו יש תוכן קולי או תוכן מולטימדיה, המאפשר לתמלל את פס הקול, ואז ניתן לעשות שימוש בכלי עיבוד טקסטואליים כמו אינדוקס, וחיפוש על התוצר הטקסטואלי של מנוע זיהוי הדיבור תוך התחשבות בשגיאות הזיהוי של המנוע. בין השימושים האפשריים של זיהוי דיבור בתחום המרת תוכן דיבור לטקסט, ניתן למנות גם המרה של הודעות קוליות לטקסט ושליחתן לנמען כהודעת טקסט. בתחום המודיעין העסקי - הקלטת השיחות במרכזי שירות לקוחות ותמלולן תאפשר ניתוח טקסטואלי של תמלול השיחות לצורך קבלת מידע עסקי מתוכן השיחות, כמו גם אימון ומשוב לנציגי שירות הלקוחות.

בשירות המודיעין

טכנולוגיית זיהוי דיבור מעניינת גם את גופי הביטחון. על פי פרסומים גלויים, ניתן למצוא בתחום זה מספר יישומים.

כך, הקלטת שיחות על ידי ארגוני מודיעין ותמלולן על ידי מכונה יכולה לאפשר קבלת מידע מתוכן השיחות המתקיימות. תמלול השיחות לאורך זמן תוך ביצוע אינדוקס של התמלול המתקבל, יאפשר ביצוע חיפושים בבסיס הנתונים של השיחות.

נעשה שימוש בטכנולוגיה גם לטובת מערכות תרגום לחיילים בשטח. על פי הפרסומים בתקשורת, IBM סיפקה לצבא האמריקאי בעיראק מערכת המתרגמת מערבית עיראקית לאנגלית, לצורך קשר עם האוכלוסייה והצבא המקומי. גם סימולטורים, מערכות אימון לטייסי ולפקחי טיסה, יכולים להתבסס על זיהוי דיבור. וכמו בבית החכם, ניתן יהיה להפעיל מכשירים צבאיים שונים כסביבות ההפעלה הצבאיות בשימוש בזיהוי דיבור.

ומה יהיה כאשר המכונות יבינו גם את "שפת הגוף"? גם התחום הזה כבר אינו בגדר מדע בדיוני, זיהוי דיבור בשילוב זיהוי תנועות הגוף וזיהוי רגשות יאפשר תקשורת אדם-מכונה בדומה לתקשורת אדם לאדם. ©

פרופ' עמי מויאל הוא ראש המחלקה להנדסת חשמל וראש המרכז לעיבוד שפה באפקה - המכללה האקדמית להנדסה בתל-אביב. לאקדמיה הגיע פרופי מויאל לאחר 15 שנה בתעשיית ההיי-טק. בעבר היה מנכ"ל חברת nsc, שפיתחה מנוע זיהוי דיבור.

הדיאלקטים, הגילאים והמין (מגדר). כמו כן, עליו לייצג את הסביבות השונות בהן יבוצע הזיהוי כמו המשרד, הבית, רכב חונה, רכב נוסע ומקומות ציבוריים וגם את המיקרופונים ומכשירי הקצה, בהם נעשה שימוש.

בסיס הנתונים צריך לכלול, בנוסף, מילון עשיר, שיאפשר קבלת חזרות בכמות מספקת ומייצגת של הפונמות שיש לאמן. על פי רוב, נעשה שימוש בבסיס נתונים הכולל הקלטות של אלף דוברים לצורך אימון ראשוני של מנוע זיהוי הדיבור, ולאחר התקנת השירות מבוצע אימון נוסף עם הקלטות שנאספו מהשירות. יש לזכור, כי הטכנולוגיה תלויה בשפה, ובעת המעבר לשפה חדשה נדרשים בסיסי נתונים גדולים כדיבור, לצורך תהליך אימון יעורי.

יישומים של תקשורת אדם-מכונה

במעבר מהדיון בטכנולוגיה עצמה לדיון ביישומים האפשריים שלה, ניתן לאפיין שני סוגי יישומים עיקריים לזיהוי דיבור: האחד בא לידי ביטוי בתקשורת הישירה אדם-מכונה באמצעות דיבור, והשני בהמרה של תוכן דיבור לטקסט, לצורך עיבוד טקסטואלי.

בסוג הראשון ניתן לעשות שימוש בזיהוי דיבור בכל מקום בו יש תקשורת אדם-מכונה המתקיימת כיום על ידי הקשה על מקשים. מנוע זיהוי הדיבור יזהה את בקשת המשתמש, מנוע המרת טקסט לדיבור (TTS - Text To Speech) יאפשר למכונה לדבר, ואפליקציה תנהל את הדיאלוג בין האדם למכונה עם יכולת לזיום שאלות במהלך הדיאלוג (Initiative Dialogue).

בין השימושים האפשריים לדיאלוג כזה ניתן למנות יישומים כמו מרכזי שירות לקוחות שבהם לפחות חלק מהשיחה תתבצע מול מכונה. אפשרי גם ממשק כדיבור עם המכשיר הסלולרי לצורך הפעלתו כמו גם הפעלת אפליקציות שונות הפועלות על המכשיר הסלולרי, כגון הכתבת הודעות טקסט קצרות (SMS) או דואר אלקטרוני, קביעת פגישות, קבלת מידע וחיפוש קולי. בתחום הרפואי יתאפשרו הכתבות של דיאגנוזות רפואיות לתיקו של החולה, תוך מתן אפשרות לרופא לברוק את החולה במקביל להכתבת הדיאגנוזה.

ב"בית חכם" מתאפשרת הפעלה קולית של המוצרים בבית: כבר בתקופה הקרובה ניתן יהיה לראות טלוויזיות המופעלות כדיבור, ומשם ניתן