Better Sound Hound - project:

דוד ניר 203487293, ארז הכט 311248959, מעיין שטרית 315512715

תיאור הפרויקט:

הפרויקט שלנו הוא אפליקציה שמאפשרת שליטה מרחוק על המחשב מבלי לגעת בו.
המשתמש יבצע מחווה גופנית - ביצוע מחווה פיזית עם יד אחת (סוג של תנועת "חכה רגע") לכיוון המצלמה, האפליקציה תזהה את המחווה ותעבור למצב דיבור. מצב דיבור מאפשר תקשורת עם המחשב באמצעות פקודות קוליות (כמו סירי ואלקסה). חלק מהפקודות יהיו חיפוש בגוגל, פתיחה ושליטה ביוטיוב, זום אין/אאוט, ועוד. לצורך הנוחות, תמיד תופיע חלונית קטנה המאפשרת למשתמש להבין בבירור מתי המחשב מאזין לו ומתי לא באמצעות אייקון קטן (אייקון של מצלמה בשלב זיהוי המחווה, מיקרופון בשלב הזיהוי הקולי וכו'). בשלב ההאזנה, המשתמש יוכל לומר למחשב פקודת חזרה שתגרום ליציאה ממצב דיבור וחזרה למצב בו האפליקציה מחכה למחווה גופנית, פקודת יציאה שתצא מהתוכנית או פקודת "הצג הוראות" המראה את הפקודות האפשריות ומידע נוסף, בכל אופן, לאחר כמה שניות בלי קלט מצב הדיבור יבוטל והמשתמש יוחזר למצב זיהוי המחווה.

התמונות לחיווי למשתמש:

מאזין לפקודה:

מחכה למחווה:

מבצע פקודה:

פקודה לא ידועה:

קהל יעד:

האפליקציה מותאמת לכל מי שמשתמש במחשב וצריך אפשרות של שליטה מרחוק או ללא מגע, המשתמשים העיקרים יהיו:

- משתמשים שירצו שליטה במחשב בזמן בישול.
- מתאמנים שעושים ספורט עם סרטון הדרכה ביוטיוב ורוצים להתאים את קצב הסרטון אליהם, לעצור כשרוצים לשתות מים, לנוח יותר, לעשות יותר חזרות וכו' יוכלו להשתמש באפליקציה ולעשות זאת בקלות.
- בישול תוך כדי צפייה בסרטון בישול - בשלנים לא מקצועיים יוכלו לעצור את הסרטון ולעשות דברים בקצב שלהם, למשל לחתוך לאט יותר, להחזיר כמה שניות אחורה, לחפש בגוגל, וכל זה בלי מגע עם המחשב בכלל, ובלי צורך לשטוף ולנגב ידיים לפני כל פעולה כזו.

מה חשוב לכם שיחווה קהל היעד מהפרויקט:

חשוב לנו שמשתמשי האפליקציה לא יסתבכו בהפעלה ובשליטה בה. אנחנו רוצים שה flow יהיה חלק וזורם, ברור וקליל, כך שגם משתמשים מבוגרים או אנשים שאין להם זיקה למחשבים יוכלו להשתמש ולהנות מהאפליקציה. בנוסף, חשוב לנו שהמשתמשים לא יצטרכו להתעסק יותר מדי עם GUI ושיהיו כמה שפחות לחיצות וGUI מציק (לכן החלונית הקטנה תפתח בצד ולא תפריע).

אילו שיקולים במימוש הפרויקט נגזרו מהסעיפים הקודמים:

- בחרנו לממש GUI בסיסי ביותר לאפליקציה, עם הדרכה בסיסית לפעולה אחת שצריך לעשות. בצורה זו משתמש שלא מכיר את האפליקציה לא ירתע ולא יצטרך לחקור את האפליקציה על מנת להפעיל אותה.
- כשבחרנו את מחוות הגוף שנעשה התלבטנו בין כל מיני תנועות אפשריות - תנועת עצור עם היד (שנבחרה), נמסטה עם שתי הידיים, תנועת שלום, ועוד. חיפשנו תנועה שלא תביך את המשתמש (קצת מוזר לעשות נמסטה למחשב, במיוחד אם אחת הידיים מלוכלכת), שתיצור כמה שפחות זיהויים שגויים, כלומר שתהיה תנועה בסיסית שהסיכוי שהמשתמש יבצע סתם הוא קטן, ושיהיה קל לאפליקציה לזהות אותה. תנועת עצור בסיסית עם היד ענתה על כל הדרישות ולכן בחרנו אותה.
- רצינו משהו שיבדיל אותנו מאפליקציות רגילות לזיהוי קול כמו סירי ואלקסה, ולכן הרחבנו את השליטה עם האפליקציה שלנו לשליטה על יוטיוב, חיפוש בגוגל, שליטה בספוטיפיי וכו'. כמובן שתמיד ניתן להוסיף עוד ועוד פונקציונאליות למשתמש לפי דרישה.

בדיקות על משתמשים, שינוים ולקחים:

חלק מהאפליקציה נכתב בלינוקס וחלק בווינדוז, אחד מהאתגרים והשינויים הגדולים שהיו לנו בקוד היה האינטגרציה בין החלקים שעובדים בלינוקס לחלקים שעובדים בווינדוז, זה גרם לנו לשנות כמעט את כל הפרויקט ולהתגבר על הרבה קשיים טכניים, לשמחתנו הגענו למצב שהאפליקציה מצליחה לרוץ ולעבוד טוב גם בלינוקס וגם בווינדוז.
לאחר שהצלחנו לחבר את כל חלקי האפליקציה ניסינו להשתמש בה באופן עצמאי כבדיקה ראשונית על מנת למצוא באגים. מצאנו לא מעט בעיות והצלחנו להתגבר כמעט על כולם.

- מבחינת זיהוי המחווה הוספנו כל מיני קריטריונים וזיהויים שונים שימנעו שגיאות גם מסוג False Positive וגם מסוג False Negative, שיפרנו ממש את איכות הזיהוי. בהתחלה נתקלנו בבעיה שכאשר המרפק לא נמצא בתמונה האפליקציה לא מצליחה לעשות את הזיהוי, וכמו כן קורה לפעמים שהמצלמה מזהה את המחוה סתם. הפתרונות שעשינו הם שיפור האלגוריתם של זיהוי היד, כעת האפליקציה מזהה את היד גם אם רק כף היד נמצאת בתמונה והמרפק מוסתר, באופן זה הקטנו משמעותית את השגיאות של זיהוי סתמי. בנוסף הוספנו בדיקות של המיקום היחסי של האצבעות כדי לגלות את האוריינטציה של היד (למשל לזהות מתי כף היד פונה למצלמה ומתי גב כף היד פונה למצלמה). בהתחלה רצינו שהאפליקציה תחפש גם מבט של המשתמש, לכן כתבנו אלגוריתם (די בסיסי) שיודע לזהות מתי המשתמש מסתכל על המצלמה, אך בהמשך לא היינו צריכים את זה בכל מקרה.
מעבר לשיפור הזיהוי, הוספנו מנגנונים חיצוניים לשיפור ביצועים, כגון על מנת לזהות מחוה נבצע חישובים על הפריים ה-n, (במימוש שלנו, כל 4 פריימים היה פרמטר טוב מספיק). הוספנו ספירת פריימים מהזיהוי האחרון כדי להמנע מזיהויים רצופים בעקבות עדכון לא מספיק מהיר של המצלמה או דרישה למספר מסויים של פריימים בהם זוהתה מחווה כדי לקרוא לה מחווה אמיתית.
- מבחינת הזיהוי הקולי לא הצלחנו להתגבר לגמרי על בעיית חוסר הזיהוי שיש לפעמים הנובעת בגלל מבטא או בעיות במיקרופון, להבנתנו זו בעיה שאין לה עדיין פתרון מוצלח בשוק ולכן לא הצלחנו לשפר בהרבה את הזיהוי מלבד לנסות לומר את הפקודות עם מבטא. בנוסף נתקלנו בקשיים שנבעו מתקשורת (השתמשנו במנוע של google בשביל הזיהוי), למשל בהדגמה שלנו היה latency גבוה משמעותית ממה שהיה בבדיקות שלנו וזה אפילו פגע בהדגמה שניסינו לעשות בפני כל הקורס. פתרון מאוחר יותר לבעיה הזו היה לפתוח hotspot דרך טלפון כדי לקבל קצב העלאה טוב יותר (חבילות אינטרנט של טלפונים כוללות קצב העלאה משמעותית גבוה יותר מחבילות אינטרנט של בתים) ולהגדיל את רוחב הפס שיש למחשב שמריץ את האפליקציה.

כשסיימנו לבנות את האפליקציה, מאוד רצינו לתת לאנשים שונים (הרבה אנשים) שלא הכירו את האפליקציה בכלל לנסות להשתמש בה על מנת לראות אם ה flow זורם, לצערנו בעקבות הקורונה זה לא התאפשר, אבל כן נתנו למספר מצומצם של אנשים קרובים מהמשפחה להתנסות, גילינו שמילות החיפוש שאנחנו הגדרנו לא כיסו את מילות החיפוש של חלק מהאנשים ולכן הרחבנו את מאגר המילים שהאפליקציה מזהה עבור כל פקודה.

<u>מה היינו רוצים לעשות אחרת – לו היה מתאפשר:</u>

אחד הדברים שהיו חסרים לנו הוא בדיקת האפליקציה על ידי משתמשים זרים - מה שלא התאפשר בעקבות הקורונה. זה היה מאפשר לנו לראות זוויות נוספות ודרישות אמיתיות של אנשים אמיתיים ובכך לקבל פידבק מאוד מועיל שלא היה לנו.

כמו כן היינו רוצים לממש יותר פקודות שמזוהות באמצעות מחוות, אבל גילינו שמאוד קשה לזהות מחווה בצורה טובה כשכל מה שלרשותנו הוא מחשבים די מוגבלים מבחינת חומרה ומצלמות די בסיסיות (webcam רגילות מובנות בלפטופ), ולכן המחווה היא רק למעבר לשליטה קולית כרגע.

בנוסף, היינו רוצים לממש מודל המבוסס על למידה או להשתמש באחד קיים או כפי שינון הציע לבדוק אם אפשר לפקס את המנוע שהשתמשנו בו של גוגל למשפטים ספציפיים על מנת לשפר את איכות וכמות הזיהויים הקוליים שלנו. אנו בטוחים שזה היה משפר בהרבה את חווית המשתמש.

כמו כן היינו רוצים להשקיע יותר זמן בUI ובחוויית המשתמש (חלונות יותר יפים ומושקעים, דיבור וצלילים יותר מושקעים ומחושבים) ולהוסיף עוד אפשרויות (כרגע יש מספר די מוגבל של דברים שהגדרנו, אבל בגלל שאנחנו מתרגמים פקודה קולית לפעולה שהמחשב מבצע, השמיים הם הגבול).

לבסוף, אם היינו יכולים להמשיך לעבוד על האפליקציה בלי מחסומי חומרה (הן מבחינת GPU/מחשבים חזקים והן מבחינת תקשורת latencyו מול APIים של גוגל ושות'), היינו מקבלים תמונה הרבה יותר טובה של מה באמת אפשר לעשות עם האפליקציה הזו וזה היה יכול להיות ממש נחמד.