

SignSight – Zajtalan párbeszéd: Egy intelligens szemüveg forradalma a siket közösség számára

SilentSpellers

Rekecki Alex

Felkészítő: Szabó Beatrix

*Szegedi SzC Tóth János Mórahalmi Szakképző Iskola és Szilágyi Mihály Kollégium,
6782 Mórahalom, Dosztig köz 3.*

1. Bevezetés

A halláskárosult személyek társadalmi és kommunikációs nehézségei jelentős akadályokat állítanak a mindennapi életük és társadalmi integrációjuk elé. A meglévő támogató technológiák, például a beszédfelismerő alkalmazások, okostelefonos feliratkészítők és különféle kiterjesztett valóság (AR) szemüvegek, ugyan segítséget nyújtanak a kommunikációban, de számos hátránnyal rendelkeznek.

A projekt célja egy olyan eszköz megalkotása, amely a siket emberek számára lehetővé teszi a környezetükben lévő beszéd hatékony felismerését és szöveges megjelenítését. A rendszer tervezésekor kiemelt szempont volt a hordozhatóság, a használati kényelem és a megbízható működés.

2. Probléma megoldásának menete

A prototípus megvalósítása Raspberry Pi 4 mikrovezérlőre épült, amelyhez OLED kijelző, mini USB csatlakozik. A szemüveg ergonomikus keretét 3D nyomtatással állítottuk elő. A szoftver Python alapú, és a SpeechRecognition könyvtár, valamint a Google Speech API felhasználásával biztosítja a valós idejű hangfelismerést. A rendszer működése során a környezeti hangokat szöveggé alakítja, amely azonnal megjelenik a beépített OLED kijelzőn. A szemüvegről készült kép az 1. ábrán látható. Az első fázis célja egy okos szemüveg kifejlesztése volt, amely képes a környezeti beszédhangok feldolgozására és azok megjelenítésére.

Majd a projekt tovább lett bővítve egy mobil alkalmazással. Az OLED kijelzőt mobil alkalmazás váltotta fel. A szoftverről készült kép a 2. ábrán tekinthető meg.

Felül egy legördülő menü található, amelyben a mikrofon bemenetet lehet kiválasztani. Az alapértelmezett opció a "Microsoft Sound Mapper", amely általában a rendszer alapértelmezett mikrofonját jelenti.

A következő legördülő menüben kiválasztható a beszédfelismerés nyelve. A szoftver magyar és angol nyelven működik.

Szöveglablak: Egy nagy, üres szövegmező található középen, ahol a szoftver valós időben jeleníti meg a felismert szöveget.



1. ábra: Az első prototípus

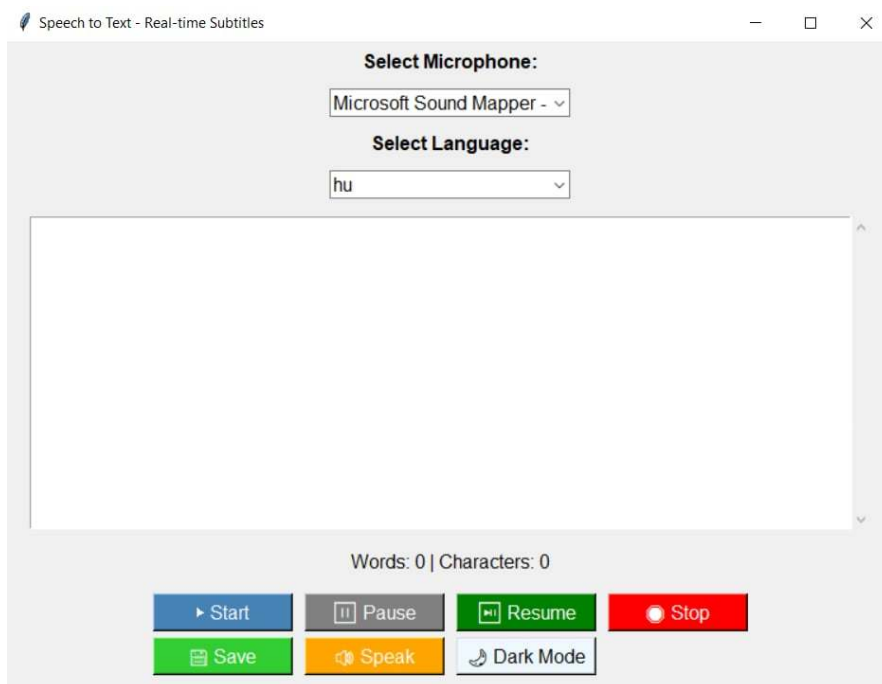
Statistika: A szövegmező alatt a felismert szavak és karakterek számát követhetjük nyomon.

Vezérlőgombok:

1. Start: Elindítja a beszédfelismerést.
2. Pause: Szünetelteti a felismerést.
3. Resume: Folytatja a szüneteltetett felismerést.
4. Stop: Leállítja a folyamatot.
5. Save: Lehetőség van a felismert szöveg elmentésére.
6. Speak: Felolvasási funkció, amely visszamondja a felismert szöveget.
7. Dark Mode: A sötét mód bekapcsolása a kényelmesebb olvashatóság érdekében.

3. Elért eredmények

Az eddigi tesztek során a rendszer képes volt a környezeti beszéd felismerésére és megjelenítésére magyar és angol nyelven. A 3D nyomtatott keret tartósnak és kényelmesnek bizonyult. Bár a rendszer működött, a szemüveg használata nem bizonyult kényelmesnek, mivel az OLED kijelző túl közel volt a szemhez, ami nehézkessé tette a mindennapi használatot. Az első fázis problémáira reagálva, a második fázisban a projekt átalakult. Az alkalmazás használata kényelmesebb, mivel nem szükséges közvetlenül a szemhez közelíteni a kijelzőt, így a felhasználók számára kényelmesebbé vált a mindennapi használat, miközben megőrzi a célzott funkciók elérését.



2. ábra: A szoftver felépítése

A beszédfelismerés során számos kihívással szembesülhet a rendszer, amelyek befolyásolhatják a felismerés pontosságát és hatékonyságát. A rendszer érzékenysége korlátozott lehet, és ha a beszéd túl halk, a mikrofon nem érzékeli megfelelően a hangokat, így hiányos vagy pontatlan szöveg jöhet létre. Zajos környezetben, például forgalmas utcán vagy irodában, a háttérzajok megzavarhatják a beszédfelismerést, mivel a rendszer nehezen különbözteti meg a beszédhangokat a zajtól. Az eltérő akcentusok és dialektusok jelentős eltéréseket okozhatnak a kiejtésben, amelyekkel a beszédfelismerő algoritmusok nem mindig tudnak megbirkózni, különösen, ha nem kaptak megfelelő mennyiségű tanítási adatot ezekhez a nyelvváltozatokhoz. A szavak pontatlan kiejtése félreértelmezéshez vezethet, különösen olyan nyelveknél, ahol a kiejtés és a jelentés között szoros kapcsolat van. A gyors tempójú beszéd esetén a rendszer nem tudja megfelelően elkülöníteni a szavakat, ami téves szövegeredményekhez vezethet. Ha a beszélő nem tart szünetet a szavak vagy mondatok között, a rendszer gyakran összefüggő szöveggként kezeli az elhangzottakat, ami a szöveg jelentésének torzulásához vezethet. A nagyon hosszú mondatok és

bonyolult szerkezetek nehezíthetik a beszéd felismerést, mivel a rendszernek egyszerre nagy mennyiségű információt kell feldolgoznia.

A következő lépés, hogy egy olyan eszközt dolgozzunk ki, ahol a szemüveg és a szoftver is kompatibilisen és praktikusán működhet együtt. Valamint látható, hogy számos környezeti hatás befolyásolhatja az észlelés pontosságát. Céлом, hogy ezeket a hatásokat kiküszöböljem és a szoftver megbízhatóságát növeljem.