

Az ABSONIC beszédrögzítő, beszédkezelő és beszédfeldolgozó rendszer és a Microsoft Azure beszédfelismerő és fordító szolgáltatás kapcsolódása

Az ABSONIC v2 digitális hangrögzítő rendszer nagy része a NICOpro és GALLMED Kft. saját fejlesztésű szoftver és hardver elemeiből épül fel. Az ABSONIC v2 hangrögzítő szoftver a mai kor követelményeinek megfelelő technológiákat alkalmazva, több évi fejlesztő munkával készült. Követve a változó piaci igényeket és a technikai fejlődést, a rendszerünket a Microsoft Azure beszédfelismerő és fordító szolgáltatás használatával bővítettük ki.

1. Beszédfelismerés

1.1. A beszédinformáció feldolgozást segíti a beszédfelismerés, amelynek során az élő beszéd vagy a már rögzített és tárolt beszédanyag közvetlenül karakteres szöveggé alakul át, tehát elektronikusan, automatikusan leíródik. Mikrofonokról, illetve mikrofon hálózatokról történő beszédfelismerés esetén, jó műszaki előkészítés után, a leirat pontossága 90-95% fölé is vihető. A távközlési eszközökön folytatott beszélgetések beszédfelismerése rosszabb hatásfokú (a távközlési csatornáktól és zajoktól függően például 40-90%), de ez is javítható, megfelelő műszaki előkészítés után (zajszűrés, szintkiegyenlítés, visszhang mentesítés stb.). Azon alkalmazási helyeken, ahol egyébként nem szükséges a hangfelvétel manuális leírása, de mégis jó, ha megvan a leirat vagy nem rendelkeznek vakon, illetve kétkézesen gépírni tudó munkatárssal, ott a beszédfelismerő használata gyorsabban megtérül. Amennyiben emlékeztetőt, jegyzőkönyvet írnak a hanganyagból, úgy a beszédfelismerő segítségével a leíró manuális munkája átlagosan kb. 66%-kal csökkenthető.

1.2. A Microsoft Azure beszédfelismerő szolgáltatás jellemzői

- 1.2.1. A beszédfelismerő szolgáltatás jelenleg 122 nyelv felismerését biztosítja.
- 1.2.2. Központosítás biztosítása (mondatokra tagolás, kis- és nagybetűk, írásjelek használata) a beszédfelismert szövegben.
- 1.2.3. Trágár szavak kiszűrése a beszédfelismert szövegből, be- vagy kikapcsolhatóan.
- 1.2.4. Automatikus nyelv kiválasztás. Négy tetszőlegesen beállított nyelvből a beszédfelismerő szolgáltatás automatikusan ki tudja választani a beszédfelismerés nyelvét. A beszédfelismerés folyamán, akár mondatonként is változhat a beszédfelismerés nyelve.
- 1.2.5. Tanítható, saját beszédfelismerési modell használata a nagyobb beszédfelismerési pontosság eléréséhez (speciális terminológiák, szaknyelv, stb.).

1.3. Valós idejű beszédfelismerés

- 1.3.1. **Diktálás valós idejű leírása**, a beszédfelvétel és a diktált szöveg letárolása (DICSTATION)
 - 1.3.1.1. A diktálás közben a felismert szöveget figyelni, ellenőrizni és javítani lehet. Az írásjelek beillesztése, kis- és nagybetűváltások, javítások stb. beszédvezérléssel is működhetnek. Jó műszaki előkészítés esetén a felismert szavak pontossága az akusztikus

környezettől és a diktálótól függően 90-95% fölé is vihető. A diktálás hanganyaga a leirattal együtt, azzal összerendelve tárolható, továbbítható, feldolgozható.

1.3.2. **Előadások, oktatások, beszédek az előadóterem kivetítőjén az** előadással egy időben folyamatosan és karakteresen megjeleníthetők, akár a fordításukkal is (LECTURESTATION)

1.3.2.1. A közösségi, egyetemi és más oktatás, képzés, előadás, tájékoztatás, konferenciák az emberi együttműködés, információ átadás napi gyakorlata, így ennek tovább fejlesztése, hatékonyság növelése széleskörű igény és lehetőség minden tématerületen, a legkorszerűbb digitális infokommunikációs megoldásokkal.

1.3.2.2. Az előadás hanganyagának és karakteres leiratának rögzítése történhet az előadóteremben lévő számítógépre és hálózati kapcsolaton keresztül feltölthető egy szerverre is, illetve szinkronizálható felhő alapú tárhelyre is (Azure, One drive, Google drive, stb).

1.3.2.3. Az előadás hanganyagának karakteres leirata a kivetítő vizuális érzékelése mellett követhető még valós időben WiFi-s közvetítéssel a teremben lévő hordozható számítógépek, okos telefonok böngészőjében is. A beszéd felismeréssel egy időben a beszéd felismert szöveg fordítása is megtörténhet és megjelenhet.

1.3.3. **A számítógép, billentyűzet, egér vezérlés/kezelés beszéddel**

1.3.3.1. Kézhasználatban korlátozottak életminőségének, tanulásának és munkájának támogatása számítógép-, billentyűzet- és egér beszéddel történő vezérlésével.

1.4. **Letárolt felvételek beszéd felismerése**

1.4.1. **Megbeszélések, értekezletek rögzítés utáni beszéd felismerése,** karakteres leírata (jegyzőkönyv, emlékeztető stb.), mely manuális korrekciókat igényelhet (MEETINGSTATION)

1.4.1.1. A beszéd felismeréssel leíratott szöveg a környezet zajaitól, zavaraitól függően karakterhibákat is tartalmazhat, amelyeket a felismert szöveget megjelenítő szövegszerkesztő felületen ki lehet javítani. Viszont jó műszaki előkészítés esetén a leirat pontossága az akusztikus rendszer és környezet minőségétől, valamint a beszélők beszédmintájától, valamint a (szak)szöveg gyűjteménytől függően, a betanító szolgáltatás (saját beszéd felismerő modell) segítségével 90-95% fölé növelhető. Az értekező hanganyaga a leirattal együtt, azzal összerendelve tárolható, továbbítható, feldolgozható. A beszéd felismeréssel egy időben a beszéd felismert szöveg fordítása is megtörténhet.

1.4.2. **Infokommunikációs beszélgetések rögzítés utáni beszéd felismerése, karakteres leírata** (INFOCOMSTATION)

1.4.2.1. A különböző távközlési és informatikai eszközökkel, vonalakon, csatornákon folyó beszélgetések tartalma, értéke megkívánhatja a rögzítésüket, a beszélők saját számítógépén (is). Ezekhez saját gyártású távközlési adaptereket szállítunk. A letárolt infokommunikációs beszélgetéseket a beszéd felismerő szolgáltatás segítségével le lehet írni, a beszédátvitel minőségétől és a beszélők

beszédjellegétől függő minőségben, valamint fordítás is készülhet róluk. Ezeknek a leiratoknak a pontossága valamivel kisebb, mint a jó minőségű mikrofon felvételeké.

1.4.2.1.1. Rögzíthető vonalak, illetve készülékek típusa, egy időben 64 csatornáig:

- analóg-hagyományos telefon
- ISDN-digitális telefon
- VoIP-internetes telefon
- GSM-mobil telefon
- SKYPE, Viber, Telegram, Messenger
- műholdas telefon (INMARSAT, IRIDIUM, THURAYA stb.)

1.5. Saját beszéd felismerő (betanított) modell használata

1.5.1. Az ABSONIC Microsoft Azure alapú beszéd felismerő szolgáltatás minőségét, pontosságát tovább javíthatjuk saját beszéd felismerő modell használatával, amely lehetővé teszi a beszéd felismerő szolgáltatás bővítését egy adott szakterületen használt résznyelvre (gazdasági, jogi nyelv, orvosi terminológia, építészeti, önkormányzati szaknyelv stb.) jellemző lexikális sajátosságokra, szakzsargonra, a megrendelők által használható beszédre és szövegre vonatkozó betanító szöveges anyagokkal.

1.5.2. Betanító anyag készítése

1.5.2.1. A betanító anyag készítése az ABSONIC beépített szövegszerkesztőjében történhet.

1.5.2.2. Olyan szöveges adatokat kell használni (beírni), amelyek közel állnak a várt kimondott szövegekhez. Erre a célra használhatóak a már elkészült, kijavított szakszövegek is.

1.5.2.3. A használt szövegeknek nem kell teljesnek vagy nyelvtanilag helyesnek lenniük, de pontosan le kell írniük azt a kimondott szöveget, amelyet a saját betanított modelltől várunk.

1.5.2.4. Ha lehetséges, akkor egy mondatot vagy kulcsszót egy külön sorban kell leírni.

1.5.2.5. Ha növelni szeretnénk egy kifejezés, például a terméknevek tanulási súlyozását, akkor több mondatot kell a szöveghez hozzáadni, amelyek tartalmazzák a kifejezést, lehetőleg többször. Javasolt a kifejezés 10-15 megjelenése, különböző leíratokból.

1.5.2.6. A betanító anyagot text (szöveg) formátumban kell letárolni.

1.5.3. Betanító anyag feltöltése az Azure szerverre

1.5.3.1. A betanító anyagot közvetlenül az ABSONIC szövegszerkesztő felületéről kell feltölteni feldolgozásra az Azure szerverre.

1.5.4. Saját modell készítése a betanított anyagokból.

1.5.4.1. Az ABSONIC beszéd felismerést beállító felületén kell kiválasztani az adott terminológiához (szaknyelvhez) tartozó, feltöltött, szöveges betanító anyagokat, amelyekből a saját beszéd felismerő modell készül.

1.5.4.2. Több saját modell is készíthető a különféle szaknyelveknek megfelelően és kiválasztható, hogy melyiket használja a rendszer a beszéd felismeréshez.

1.5.4.3. Beszédfelismerési nyelvenként is külön készíthetők a saját beszédfelismerő modellek.

1.6. **A beszédfelismerés során elkészült szövegben kereshetünk** tetszőleges szóra, kifejezésre és a keresési eredmény alapján könnyen megtalálhatjuk a kapcsolódó hangfelvételben is a keresett szakaszt, amit az ABSONIC le is játszik számunkra.

1.7. **Beszédfelismerés költsége**

1.7.1. Alap beszédfelismerés esetén: 1 perc = 112 szó = 0,27 oldal A4: 6,56Ft (1óra~400 Ft, 1000 óra ~400 000 Ft)

1.7.2. Saját beszédfelismerő (betanított) modell esetén: 1 perc = 112 szó = 0,27 oldal A4: 9,68Ft (1óra~600 Ft, 1000 óra ~600 000 Ft)

2. **Fordítás**

2.1. **A Microsoft Azure fordító szolgáltatás jellemzői**

2.1.1. A fordító szolgáltatás jelenleg 43 nyelv fordítását biztosítja.

2.2. **Valós idejű beszédfelismeréssel egyidejűleg készülő fordítás**

2.2.1. Az ABSONIC beállító felületén kiválasztott nyelven jeleníthető meg a valós idejű beszédfelismeréssel felismert szöveggel egyidőben a fordítás szövege is, amelyek utólag is megtekinthetők és szerkeszthetők az ABSONIC szövegszerkesztő felületén.

2.2.2. A lefordított szöveg vagy a szövegszerkesztő felületen jelenik meg a beszédfelismert szöveggel egyidőben, vagy egy böngésző ablakban (HTML megjelenítéssel) vagy egy átfedő (overlay) felületen egy videó/prezentáció felett.

2.2.3. Az ABSONIC a rögzített hangfelvételhez csatolja a beszédfelismert és a lefordított szöveget is.

2.3. **Letárolt és beszédfelismert szöveg utólagos fordítása**

2.3.1. Egy már elkészült, beszédfelismert szöveg utólagos fordítására is van lehetőség.

2.3.2. A fordítás nyelvének kiválasztása után megtörténik a szöveg (dokumentum) fordítása és az eredmény a beépített szövegszerkesztő felületén jelenik meg.

2.4. Fordítás költsége: 1 perc = 855 karakter: 4,16Ft (1 óra~250 Ft, 1000 óra~250 000 Ft)

3. **Saját, offline, független szerver használata az internet kapcsolat nélküli beszédfelismerés és fordítás igénybevételéhez a nagyobb biztonság érdekében.**

3.1. **A Microsoft Azure alapú saját, offline, független beszédfelismerő virtuális szerver jellemzői**

3.1.1. A megrendelő saját céges szerverén futtatható, LAN-WLAN eléréssel.

3.1.2. A beszéd/hanganyag és annak leírata maximálisan levédhető illetéktelenekkel szemben, mivel a beszéd/hanganyag és annak leírata a személyes adatok védelmében nagyobb biztonságban van.

3.1.3. Saját betanított beszédfelismerési modell használata esetén a már betanított modell használatára is van lehetőség a szerveren.