

Beszédfeldolgozó algoritmusok tesztelése siketek szájról olvasását segítő eszközök fejlesztéséhez

Szakdolgozat

Fehér Anett Adrienn

Konzulens:

Tihanyi Attila



Információs Technológiai Kar
Pázmány Péter Katolikus Egyetem
Budapest, 2012

Nyilatkozat

“Alulírott Fehér Anett Adrienn, a Pázmány Péter Katolikus Egyetem Információs Technológiai Karának hallgatója kijelentem, hogy ezt a szakdolgozatot meg nem engedett segítség nélkül, saját magam készítettem, és a szakdolgozatban csak a megadott forrásokat használtam fel. Minden olyan részt, melyet szó szerint, vagy azonos értelemben, de átfogalmazva más forrásból átvettem, egyértelműen a forrás megadásával megjelöltem. Ezt a Szakdolgozatot más szakon még nem nyújtottam be.”

.....

Tartalomjegyzék

Nyilatkozat	3
Tartalomjegyzék	4
1 Összefoglalás	6
2 Abstract.....	7
3 Bevezetés	8
4 A segítő infokommunikációs technológia fejlődése, hallássérültek kommunikációs csatornái.....	9
4.1 A hallássérültek kommunikációs csatornái és lehetséges segítő technikái napjainkban.....	9
5 Vibrotaktilis projekt. és a beszédfeldolgozásban való szerepe.....	11
5.1 A vibrotaktilis felépítése.....	11
5.2 A vibrotaktilis vezérlőegysége	12
5.3 A vibrotaktilis szoftver követelményei.....	12
6 A beszéd, mint az emberi kommunikáció egyik alapvető eszköze.....	13
6.1 Biológiai szempontból a beszéd	13
7 Beszélt nyelvek sajátosságai, a magyar beszédhangok osztályozása	15
7.1 A magyar beszédhangok osztályozása általánosan.....	15
7.2 Beszédhangok osztályozása a projektet tekintve.....	17
8 Beszédfeldolgozó algoritmusok csoportosítása	18
8.1 Beszéddetekciós módszerek	18
8.2 Mintaillesztésen alapuló beszéddetektálás	18
8.3 Küszöb-alapú algoritmusok beszéddetektálásra	19
8.3.1 Energiaküszöb alapú megközelítés.....	19
8.3.2 Gördülő amplitúdó spektrum alapján	21
8.3.3 Spektrális entrópiaküszöb-módszer	21
8.3.4 Hosszúidejű spektrális divergencián alapuló módszer	22
8.3.5 Lineáris predikció.....	23
8.3.6 Kepsztrum.....	23

8.3.7	Zero-Crossing	24
8.3.8	Mel-frekvenciás kepsztrum	24
9	Spektrális entrópiaküszöb-módszer kiválasztása.....	25
9.1	Algoritmus implementálásának menete.....	25
9.2	Sepktrális-entrópia módszerrel való mérések analízise	27
10	Konklúzió, továbblépés	31
11	Összefoglalás	32
	Köszönetnyilvánítás	33
	Ábrajegyzék.....	34
	Irodalomjegyzék	35
	Mellékletek	37

1 Összefoglalás

A szakdolgozatomban a siketek szájról olvasását segítő vibrotaktilis projekt fejlesztésével foglalkozok. A feladatom, hogy a beszédmegértését segítő eszközhöz keressek olyan algoritmust, amely megfelel a zaj és beszéd elkülönítésére, figyelembe véve az eszköz követelményeit.

A dolgozat első részében bemutatom a hallássérültek beszédmegértési nehézségeit, kommunikációs lehetőségeit, rendelkezésükre álló infokommunikációs eszközöket. Áttekintem a Pázmány Péter Katolikus Egyetemen folytatott siketek beszédértést segítő vibrotaktilis projekt céljait, felépítését, és a beszédfeldolgozásban való szerepét.

A feladatom során áttekintettem a beszéd képzésének és felfogásának biológiai tulajdonságait. Kitérek röviden a beszélt nyelvek sajátosságaira, majd a projekt szempontjából lényeges, magyar beszédhangokra, probléma felvetést követően osztályozást hozok létre.

Szakirodalomban végzett kutatómunkám során tanulmányoztam a beszédfeldolgozó megoldásokat. A projekt szempontjából alkalmasnak tekintett algoritmust kipróbáltam. A tesztelést MATLAB környezetben végeztem el. Az algoritmus megvalósításának lépéseire kitérek dolgozatomban. A tesztelés során kapott eredmények alapján értékeltem a projektben való felhasználhatóságát és meghatároztam továbbfejlesztést követendő lehetőségeket, irányokat.

2 Abstract

In my thesis I deal with the development of a vibrotactile project that helps the hearing impaired in lip reading. My objective is to find an algorithm that is suitable for separating speech from speech noise, concerning the requirements of the device.

In the first part of the dissertation I present the speech-understanding difficulties and the communicational potentials of the hearing impaired and the available infocommunicational devices. I give an overview on the construction of the vibrotactile project in Pázmány Péter Catholic University and I also touch upon its role in speech recognition and its objectives.

During my work I sum up the biological characteristics of speech formation and understanding. I briefly touch upon the peculiarities of the spoken languages and the sounds of speech that are significant regarding the project. I create a classification after posing the problems.

I analyzed the speech processing solutions during my research work in scientific literature. In MATLAB environment I tried out the algorithm that seemed suitable relating to the project. I present the steps of the implementation of the algorithm. I evaluated its applicability in the project based on the results and determined the possibilities of further development.

3 Bevezetés

Beszédfelismerés, beszédszintézis kutatása, fejlesztése több évtizedes múltra tekint vissza nemzetközi és hazai viszonylatban is. Az igény a beszéd alapú szolgáltatásokra egyre csak növekszik. Gyakran olvashatjuk, hogy tervezik a „felhangosított” jövőt is, ma már beszéddel vezérelhető televízió is kapható.

A minél pontosabb beszédfeldolgozás célja nem egyszerűen csak a kényelem megteremtését szolgálná. Olyan eszközök megvalósításához, és eléréséhez is vezetne megfelelő beszédalgoritmusok megtalálása, optimalizálása, amelyek a siket illetve halláskárosult emberek kommunikációját segítenék. A jelnyelv csak akkor működik, ha látjuk a velünk kommunikáló félt, és ismerjük a jelnyelvet. A szájról olvasás, akkor, ha szintén látjuk a hang forrását, és ezzel együtt, az azonosan képzett hangokat meg tudjuk különböztetni. Szükség lenne egy olyan eszközre, ami észleli, a beszédet és jelez a készülék használojának.

Veszélyhelyzetben figyelmeztetni valakit nagyon körülményes, amikor nem ugyanaz a kommunikációs eszköz áll rendelkezésre (beszéd, vagy vizuális nyelv), nem is beszélve arról, hogy ilyen helyzetekben leginkább ösztönösen a hangunkat használjuk. Az egészséges és siket ember közti kommunikáció megkönnyítésére is hasznos lenne egy megfelelő eszköz. A későbbiekben olvasható vibrotaktilis projekt célja az ilyen és ehhez hasonló esetekhez nyújtani segítséget, mikor csak a beszédre támaszkodhatunk, mint kommunikációs lehetőségre. Ezzel a munkával ehhez a projekthez szeretnék hozzájárulni.

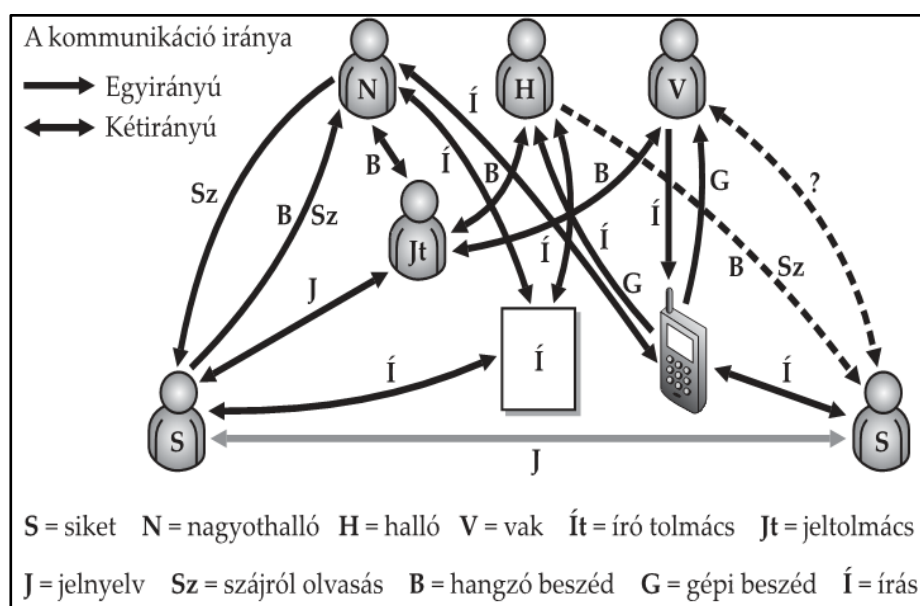
Szakedolgozatom során bemutatom a siket embereket segítő kommunikációs és infokommunikációs lehetőségeket. A beszédet és a beszéd felfogását biológiai szempontból vázolom, és mint az emberi kommunikáció egyik alapvető eszközét. Ismertetem az egyetemünkön folytatott vibrotaktilis projektet, a motivációt, az eszköz tulajdonságait és felépítését valamint a szoftver követelményeit. Beszédfelismerés szempontjából rövid áttekintést teszek a beszélt nyelvek sajátosságaira. A magyar beszédhangok csoportosítását követően, vizsgálom a beszédetektálásra alkalmas algoritmusokat az eszköz funkcionális és nem-funkcionális követelményeit figyelembe véve. A dolgozatom során ismertetem a megfelelőnek talált algoritmust, és implementációját, amit MATLAB környezetben valósítottam meg. A tesztelés során kapott eredményekből levont következtetéseimet összefoglalom, valamint definiálom a projekt továbblépési lehetőségeit.

4 A segítő infokommunikációs technológia fejlődése, hallássérültek kommunikációs csatornái

A segítőtechnológiák az informatika egyik dinamikusan fejlődő területe. Ennek oka az életszínvonal és a társadalmi felelősségvállalás növekedése a világ számos országában. Ezzel együtt egyre nagyobb a társadalmi igény az emberek életminőségének javítására. Számottevő a piaci kereslet az ilyen technológiai alkalmazásokra. Az alábbiakban a hallássérültek kommunikációs csatornáit, lehetőségeit veszem sorra. Ez a fejezet rávezet, hogy milyen újabb segítséget is nyújtaná a projekt végkimenete az emberek kommunikációjához.

4.1 A hallássérültek kommunikációs csatornái és lehetséges segítő technikai napjainkban

Hallássérültek kommunikációs csatornáit az alábbi ábra mutatja be.



1. ábra: A kommunikáció csatornái [1]

Az 1. ábra elmagyarázza, hogy az emberek között milyen kommunikációs lehetőségek jöttek létre, és milyen kommunikációs csatornák állnak rendelkezésükre, milyen eszközök vannak ma a segítségükre. Az ábra jelöli a siket- (S), halló- (H), nagyothalló- (N), nem látó (V) emberek közötti segédeszközöket. Az eszközök lehetnek jelnyelv (J), szájrólolvasás (Sz), hangzó beszéd (B), gépi beszéd (G), írás, az író tolmács (Ít), jeltolmács (Jt).

A siket emberek egymás közötti kommunikációjának eszköze leginkább a jelnyelv. A jelnyelv teljes értékű vizuális nyelv. Siket emberek között a leghatékonyabb kommunikációs csatorna, ehhez azonban látni kell egymást, még akár ismerni sem árt, ugyanis nem csak országoként, de országban belül is rengeteg válfaja létezik, és még ha ezek a feltételek teljesülnek, akkor is több mint valószínű, hogy csak siketek között akadálymentes a jelnyelven történő kommunikáció, mivel sem a nagyothalló sem a halló többség nem ismeri ezt a már Magyarországon is elfogadott hivatalos nyelvet. Így más közvetítő csatornára is feltétlen szükség van. Ilyen például a jelnyelvi tolmács igénybe vétele, mint ahogy azt az ábra is mutatja [1]. Sajnos ez a mankó sem áll minden esetben rendelkezésünkre, pl.: kisebb településeken erre nincs lehetőség. Magyarországon az is közbe szól, hogy az olyan jelnyelvi tolmács melynek szolgáltatásai igénybe vehetők nem sokkal haladja meg a 60-at. Ezzel szemben hazánk lakosságának 10%-a küszködik valamilyen hallásproblémával, WHO¹ adatok szerint 60 ezer a siketek száma, s mintegy 300 ezer a súlyos fokban nagyothallók száma. Jeltolmács hiányában, a siket embereknek más lehetőséget kell találni, hogy megszólítsák a velük élő embertársaikat. Szájról olvasás, máshogy mondva orális nyelv, vagy az írásban való közlés lehetőség áll fent. A hallássérült emberek többsége nagyon jó szájról olvasó, mégis több feltételnek kell teljesülni. Jó fényviszonyok kellenek, mindig látni kell a beszélő arcát szemből, a beszélőnek artikuláltan kell beszélni, de sokszor ezek teljesülése sem elég és félreértések történhetnek. Csak egy gyakori példát mondjak a zöngés-zöngétlen mássalhangzó párok (b/p, d/t, g/k, v/f, z/sz, zs/s). Ugyan a zöngés mássalhangzókat lazább izomműködéssel képezzük, míg a zöngétleneket feszesebben, mégsem észrevehető a különbség, mikor csak vizuálisan érzékelhetjük ezeknek a hangoknak a képzését. Kipróbálhatjuk mit ért meg a következő szavakból valaki, ha eltátogjuk neki:

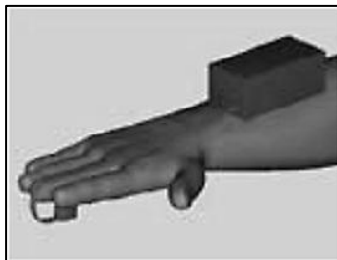
bor	por
babzsák	papság
dél van	tél van
habos	babos
veszélyez	feszélyez

Az 1. ábrán még látható az úgynevezett írótolmács. Ez QWERTY billentyűzettel ellátott mobil készülékekkel valósítható meg. Ennek hátránya, hogy elég lassú kommunikációt tesz lehetővé. Például veszélyhelyzetben erre nincs idő.

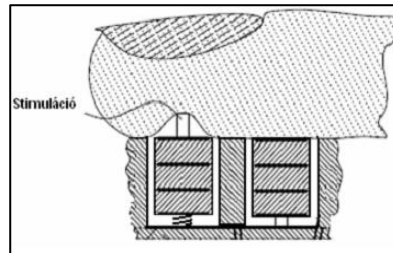
¹ WHO: World Health Organization, Egészségügyi Világszervezet

5 Vibrotaktilis projekt. és a beszédfeldolgozásban való szerepe

Rengeteg információt csak beszéd illetve a hang útján „postázunk” egymás között. Ezen az úton nem mindig ér célba az információ a címzetthez. Siket és halláskárosult emberek beszédmegértéséhez biztosíthat a projekt terméke egy segédeszközt. Ebben a fejezetben az egyetemünkön folytatott Vibrotaktilis projektet mutatom be röviden.



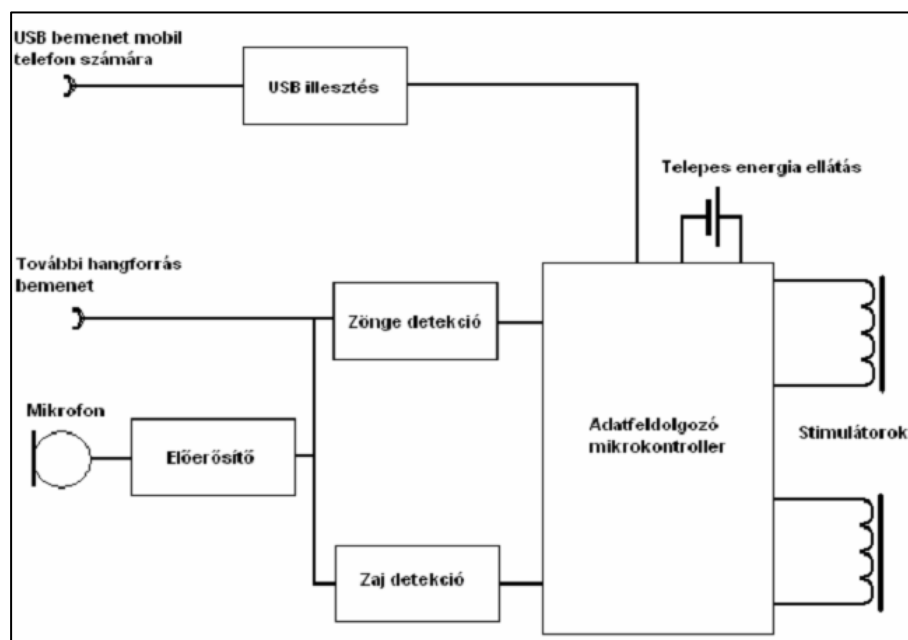
3. ábra Vibrotaktilis eszköz



2. ábra Stimuláció

A 3. ábrán egy vibrotaktilis eszköz látható. Az eszköz lényege, ha beszéd hallható, azt a zajtól megkülönbözteti, majd rezgésekkel jelez a felhasználónak, így nyerve el az eszközt használó figyelmét. Az eszköz ezt követően megkülönböztető rezgésekkel jelzi a hasonlóan képzett hangokat. A rezgéseket tapintás útján érzékeli a taktilis használója, ez látható a 2. ábrán.

5.1 A vibrotaktilis felépítése



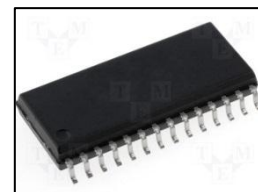
4. ábra Tervezett taktilis kijelező logikai elrendezése [14]

A vibrotaktilis alkatrészei egy nyáklapon két mágnes stimulátor, egy USB csatlakozó, egy mikrofon, és egy vezérlő chip. A 4. ábra bemutatja, a vibrotaktilis eszköz részeinek logikai elrendezését. A mikrofon minimális feszültséget (mV) ad ki magából működés közben, emiatt szükséges az előerősítő, ami 0,775-1,55 V közöttire erősíti fel a jelet. A jelvétel során fut a zaj-beszéd megkülönböztetésére szolgáló algoritmus. Amikor beszédet detektál, elkezd futni a zöngé-zöngétlen hangokat megkülönböztető algoritmus. A feladatokhoz szükséges egy megfelelő mikrokontroller kiválasztása. A stimulátorok a hangok megkülönböztetését szolgáló mechanikus rezgéseket továbbítják, amit a felhasználó tapintás útján érzékel. Az eszköz mobiltelefonra is csatlakoztatható lesz, a telefonos kommunikációt is megkönnyítve ezzel.

5.2 A vibrotaktilis vezérlőegysége

A taktilis kijelzőt vezérlő processzornak egy 16 bites mikrovezérlőt javaslok. Az alábbiakból kiderül, hogy a jelfeldolgozás fontos része a feladatnak. Az erre használatos Microchip által gyártott mikroprocesszorok, és jelfeldolgozó processzorok:

- PIC24FJ
- PIC24HJ
- dsPIC30F
- dsPIC33F



5. ábra PIC24FJ16MC102

A kategória főbb jellemzői: 16 bit szélességű RAM, 24 bit szélességű programmemória.

5.3 A vibrotaktilis szoftver követelményei

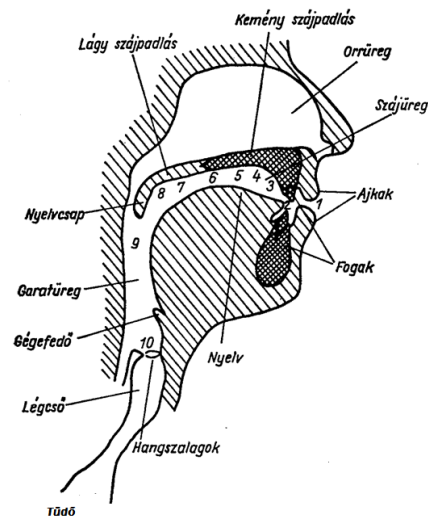
Az elképzelés az, hogy az eszköz egy készenléti állapotban várakozzon, és ha beszédet észlel, bekapcsoljon, majd ha egy bizonyos ideig nem észlel beszédaktivitást, ismét készenléti állapotra váltson. Bekapcsolt állapotban a beszédet kiemeli a zajból. A további algoritmus során kapott értékeket mechanikus rezgésekké alakítja át, amit a felhasználó a bőr felületén, stimuláció útján fog érzékelni (2. ábra). Ezek a jelek fognak segíteni a felhasználónak megkülönböztetni a hasonlóan képzett hangokat. Fontos itt megemlíteni, hogy az eszköz a beszédmegértését támogató segítőrendszer. A vibrotaktilis projekthez olyan megfelelő algoritmusok megtalálása a cél, melynek funkcionális követelménye, hogy a beszédet a zajtól el tudja különíteni, felismerje, hogy beszédaktivitás történik. A másik funkció, hogy a hasonlóan képzett hangokat meg tudja különböztetni. További követelményei a készítendő szoftvernek a gyorsaság, hogy valós idejű legyen, a pontosság elérése a lehető legtakarékosabb erőforrás igény mellett. A felhasználónak, annyi a feladata, hogy meg kell tanulni, hogy a mechanikus rezgések pontosan mit jelölnek.

6 A beszéd, mint az emberi kommunikáció egyik alapvető eszköze

Az ember számára a beszéd az egyik legfontosabb kommunikációs eszköze. A beszéd legkisebb egysége a hang. A hang a fülünk által érzékelt inger. Az alábbi fejezetben a beszédet biológiai szempontból vázolom röviden a hangkeltéstől a hangfelfogásáig.

6.1 Biológiai szempontból a beszéd

A beszédhangok képzésében részt vesz: a tüdő, a légcső, a gégefő a hangsálakkal, a garatüreg a nyelvcsappal, az orrüreg és a szájüreg. Ezeket közös néven beszélőszerveknek hívjuk, 6. ábra. A beszédhangokat kilégzés közben képezzük. Kilégzéskor a tüdőből kiáramló levegő a légcsövön keresztül a gégébe jut. Itt helyezkedik el két hangszalag, melyeknek különféle helyzetei befolyásolják a hangképzést. Hangszalagok alapállásai a következők: zárállás, résállás, nyitott állás, suttogóállás, és belégzéskor lélegző állás. A levegő a nyelvcsap irányításával a szájüregbe vagy az



6. ábra Beszédképzésben résztvevő szervek

orrüregbe kerül. Onnan a szájon vagy orron keresztül távozik. A hangok többsége szájhang, de vannak orrhangok is. A magyar nyelvben például csak három ilyen van, amikor is az orron keresztül távozik a levegő, és ekkor képződnek az 'm', 'n', 'ny' hangok. A szájüregben levő beszélőszervek a nyelv, szájpadlás, a fogak, és az ajkak.

A beszéd és az egyéb zaj által keltett levegőrezgések a hangvezetés, melyhez a hallás képességére van szükségünk. A hallás az egyik legalapvetőbb érzékelésünk, kapcsolatunk a külvilággal. Egészséges hallással rendelkező ember 20 és 20000 Hz között hallja a hangrezgéseket. A beszéd frekvencia spektruma leginkább 100 és 4000 Hz közé esnek.

A hallás érzékszerve a fül. Ez egy összetett érzékszerv, amely részben a hanghullámok formájában érkező információk felvételére, részben a testhelyzet érzékelésére szolgál (egyensúlyozó érzék). A fül felépítését három részre különítjük el: külső fül, mely lényegében egy tölcsérszerű vezeték, ahhoz, hogy a hangrezgések eljussanak a hallószervhez. Aztán van a középfül, ez egy feszes dobhártyától kezdődik, mely magát a hallójáratot zárja le, ezt a dobhártyát rezgésbe hozzák a hanghullámok. Mögötte egy üregben helyezkednek el a hallócsontocskák, melyek a dobhártya rezgéseit továbbítják. A hallócsontocskák közül a

kalapács egyik vége hozzánőtt a dobhártyához, míg a másik vége ízesül az üllőhöz, az pedig a kengyelhez. Végül belső fülnek nevezzük a kengyel talpától kezdődő ovális ablaknak nevezett részt, azt követő hallószervet tartalmazó csigát. Ebben a mechanikai inger ingerületet vált ki, melyet a hallóideg vezet tovább az agytörzsön és a talamuszon keresztül a halántéklebeny agykérgi hallóközpontjába. Ennek felfedezéséért Békésy György biofizikus, 1961-ben Nobel-díjat kapott. Munkásságával tisztázta, hogy a hangrezgések hatására a csigában hullámok keletkeznek, és ezeknek a maximális kitérése meghatározott helyhez kötött. Az adott hely membránjának rezgése pedig ingerületet kelt a hallóreceptorokban. Az egyensúlyozó szervvel kapcsolatos félkörös ívjáratok is a belsőfülben találhatóak. A fül belső fül részét folyadék veszi körül [14].

A hallásveszteség mértékét decibelben szokás megadni. A decibel két mennyiség logaritmikus mértéke.

$$I_{dB} = 10 \cdot \lg \frac{I_{mért}}{I_0} (dB)$$

Hangerősség esetében I_0 referenciának általában a 10^{-12} W/m^2 értéket választják, mivel ez közelítőleg az emberi hallásküszöb értéke.

Hallást érzékelő rendszerünk betegsége, sérültsége miatt alakulhat ki nagyothallás, ez 30-70dB közötti hallásveszteség. Kialakulhat siketség, ami pedig 70dB-nél nagyobb hallásveszteség.

A receptorszerv a fül (a nervus cochlearis), illetve a hallóideg (nervus acusticus), a cortex (a temporalis lebenyben lévő Heschl-féle tekervény), károsodása folytán létrejött hallási fogyatékokosságok léteznek aszerint, hogy melyik a sérült rész.

Egy ilyen sérülés „nemcsak a specifikus, ezúttal a hallási funkció kiesését vagy gyengülését okozza, hanem a személyiségfejlődés egészét megváltoztatja. Ily módon a hallássérülés megbontja a pszichés jelenségek funkcionális egységét, a kommunikációs akadályok következtében a gondolkodási műveletek sajátosan szerveződnek, lecsökken és átalakul a külvilágról való információ felvétel, lelassul az ismeretszerzés, nehezítetté válik a szocializálódás”- olvasható Dr. Szabó Terézia Gyógypedagógiai Alapismeretek című könyvében. Többek között ezért is fontos, hogy a kommunikációs lehetőségeit javítsuk a hallássérültek számára, ezzel is könnyebbé téve a további korlátok leküzdését.

7 Beszélt nyelvek sajátosságai, a magyar beszédhangok osztályozása

A nyelv egy eszközkészlet, a beszéd ennek a működtetése, elsődleges megnyilvánulása. A beszédhangok a nyelvre jellemzőek. Egyes nyelvekben a hangmagasság hajlítása is megkülönböztet beszédhangokat.

A beszédhangokat osztályozni a következő szempontok szerint szokták: gerjesztés szerint, hogy zöngés, zörejes vagy kevert. Szerkezet szerint, illeszkedés szerint. Osztályozni lehet a mássalhangzókat képzésük helye és módja szerint, nazális vagy orális; rés vagy zár a hangszalagok állása szerint; pergő, folyékony vagy felpattanó. Ebben a fejezetben a magyar beszédhangok osztályozását veszem végig, majd a projekt szempontjából előnyösnek tekintett osztályozást hozok létre.

7.1 A magyar beszédhangok osztályozása általánosan

Gerjesztés szerint:

- Zöngés: b, d, g, gy, v, j, m, n, ny, l, r, és a magánhangzók
- Zöngétlen: p, t, ty, k, c, cs, f, sz, s, j*, h
- Kevert: dz, dzs, z, zs

A gerjesztési hang megvalósulásához az energiát a tüdőből kipréselt levegőáram adja. A zöngés gerjesztés, más nevén zöngé, a szabályosan ismétlődő, kvázi-periodikus hanghullám. A hang keletkezése, hogy a tüdőből kiáramló levegő megrezegteti a hangszalagokat. A zörejes gerjesztés aperiodikus rezgés, és itt a tüdőből kiáramló levegő nem rezgeti meg a hangszalagokat, majd csak a szájüregben ütközik akadályba. A kettő keveréke adja a kevert hangok képzésének alapját. Gerjesztés speciális fajtája a néma fázis, amikor is nincs hangzó gerjesztési jel. A néma fázis az egyes beszédhangok például a zöngétlen zárhangok fontos része.

Szerkezet szerint:

- Egyszerű: v, f, z, sz, zs, s, j, h, m, n, l, és a magánhangzók
- Összetett: b, p, d, t, g, k, gy, ty, c, cs, dz, dzs, ny, r

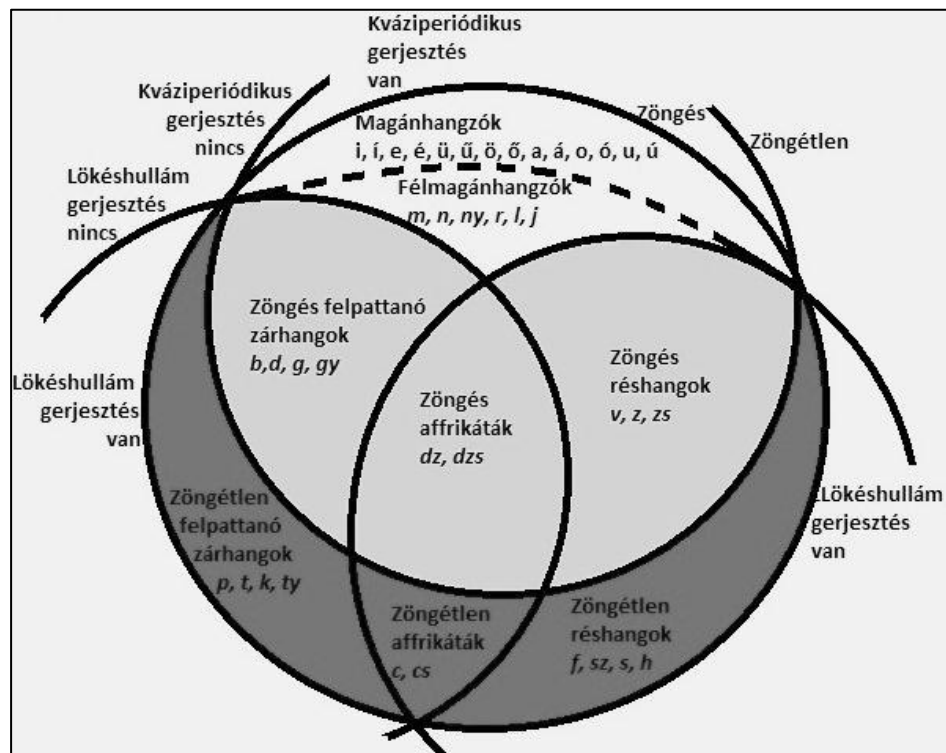
Illeszkedés szerint:

- Stabil: gy, ty, j, n, ny, r
- Kölcsönösen illeszkedő: b, p, d, t, dz, c, dzs, cs, v, f, z, sz, zs, s, h, m, l
- Erősen illeszkedő: g, k

Képzés szerint a mássalhangzók:

- Réshang (frikatíva): f, v, s, zs, sz, z, h
- Likvida (hosszan, folyékonyan ejtett hangok): l, j
- Nazális (orrhangok) zárhang: m, n, ny, ng (beszéd közben az egymáshoz idomulás miatt)
- Pergő (termuláns) zárhang: r
- Felpattanó (explozíva) zárhang: t, d, p, b, k, g, ty, gy
- Zár-réshang (affrikáta): c, cs, dz, dzs

Az alábbi ábrán összefoglalva láthatjuk a fonetikai és gépi feldolgozás alapján létrehozott beszédhangok osztályozását.



7. ábra Fonetikai és gépi feldolgozás alapján beszédhangjaink

7.2 Beszédhangok osztályozása a projektet tekintve

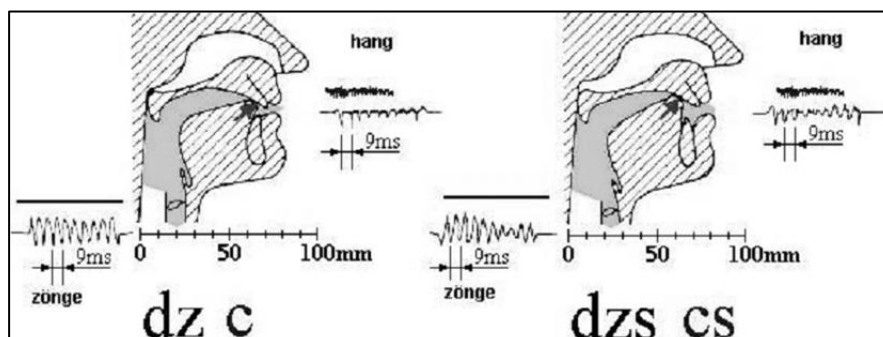
A projektet tekintve a hasonlóan képzett hangokat kell megkülönböztetni. A magyar nyelvben hasonlóan képzett hangok a mássalhangzóink között vannak.

	zárhangok								zár-réshangok				réshangok						nazálisok						
	b	p	d	t	gy	ty	g	k	c	dz	cs	dzs	v	f	z	sz	zs	s	h	m	n	ny	j	l	r
két ajak	☒	☒																		☒					
ajak-fog													☒	☒											
fog-fogmeder			☒	☒					☒	☒					☒	☒					☒				
fogmeder											☒	☒					☒	☒						☒	☒
elhúzó szájpadlás					☒	☒																☒	☒		
hátsó szájpadlás							☒	☒																	
gége																			☒						

8. ábra Zár-, zár-rés-, réshangok, nazálisok [6]

A 8. ábra a hangok képzés szerinti csoportosítását ábrázolja. Jól leolvasható a pusztán artikulációból nem megkülönböztethető hangok. Két ajakkal zártan képezzük a „b, p” hangokat. Fog és fogmeder zárásával képezzük a „d, t” hangokat. Zár-rés párok a „c, dz”. Rés hangpárok „z, sz”. Aztán hasonló képzésük miatt párok „gy, ty”, a „g, k”, a „cs, dzs”, a „zs, s”, „v, f”. Ezeket a hangokat nem tudjuk megkülönböztetni szájról olvasás folyamán, mert a képzés során az artikuláció majdhogynem egyforma. A magyar nyelvben rengeteg olyan szó van, ami ebből kifolyólag félreérthető. Erre példa: papság, babzsák.

Gerjesztésük szerint különböznek. A párok egyike zöngé a másik zöngétlen, esetleg kevert. A 9. ábra mássalhangzóink közül a dz, c és dzs, cs képzését, gerjesztését ábrázolja.



9. ábra Zár-rés hangok

8 Beszédfeldolgozó algoritmusok csoportosítása

A projekt szempontjából két algoritmikus feladatot kell megoldani. El kell dönteni, hogy beszédet detektál majd meg kell különböztetni a hasonlóan képzett hangokat. Van egy készenléti állapot. Ebben az állapotban, a beszéd jelenlétét kell vizsgálni, ha csak hangos háttér zaj van, akkor maradjon készenléti állapotban a készülék, ha beszédet érzékel, akkor bekapcsol a készülék. A taktilis további funkciója akkor aktiválódik, ha bizonyos százalékban biztosak vagyunk benne, hogy az beszéd. Ezen fejezet beszéd-zaj elkülönítésével foglalkozik, a beszéddetektálásra alkalmas algoritmusokat, lehetőségeket veszem sorra, és vizsgálom, hogy melyik megoldás lenne a legmegfelelőbb a projekt munkához.

8.1 Beszéddetekciós módszerek

A beszédfelismerés lehet izolált, más nevén kötött szótáras, vagy folyamatos. A folyamatos beszédfelismerésre neuronhálós, HMM² megoldások, küszöb-alapú algoritmusok, mintaillesztésen alapuló módszerek is léteznek.

A feladatom az, hogy olyan algoritmust keressek, amely a folyamatosan bejövő hangjelből észleli, hogy beszéd szakasz van, legminimálisabb késéssel, nagy zajban is.

Beszéd zaj elkülönítésére alkalmas megoldásokat két nagyobb csoportba sorolhatóak, ezek a küszöb-alapú algoritmusok és a mintaillesztésen alapuló módszerek.

8.2 Mintaillesztésen alapuló beszéddetektálás

A beszéddetektálás hasonlóan történik, mint a beszédfelismerés folyamata. A mintaillesztésen alapuló beszéddetektálás módszere, hogy a beszédről és a zajról is modellt kell alkotni, és a modell paramétereit meg kell becsülni.

A projekt szempontjából nem javasolt az alábbiak miatt:

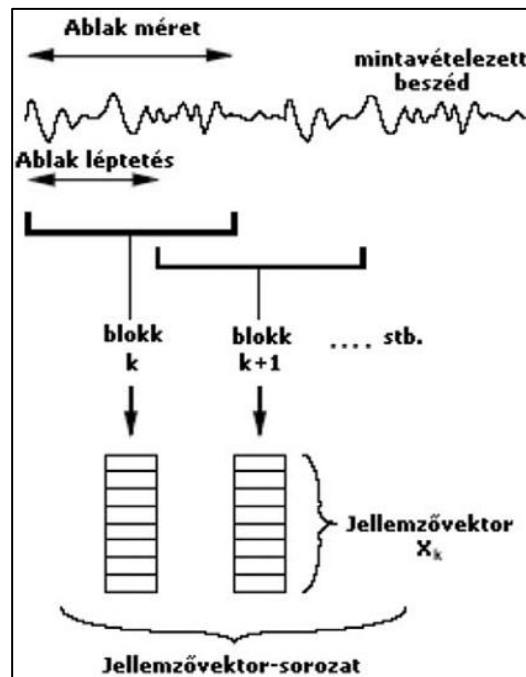
- Szükségesek tanító adatok, ezeket nehéz meghatározni, rengeteg féle zajról kellene modellt alkotnunk. Olyan eszközt szeretnénk, amit a felhasználó bárhol megfelelően tud működtetni.
- Nagyok az erőforrás igényei. A projekt szempontjából, ez az egyik legfontosabb nem funkcionális követelmény, hogy az erőforrás igénye a lehető legkisebb legyen. Egy minél kisebb fizikai méretű eszköz a cél, melynek tárhelykapacitása kevés.
- A végrehajtási sebességnek gyorsnak kell lenni, valós idejű beszéddetektálásra van szükség.

² Hidden Markov Model, Rejtett Markov Model

8.3 Küszöb-alapú algoritmusok beszéd-detektálásra

Ezeknek az algoritmusoknak a működési elve, hogy a bejövő jelből beszéd eldöntésére alkalmas paraméterek kinyerése után, előre beállított küszöbérték szerint történik a beszéd-felismerés. A küszöbérték lehet adaptív, idővel változó, globális, vagy környezethez alkalmazkodó [2].

Az alábbi ábra, ennek a műveletét mutatja be.



10. ábra Paramétervektor előállítás a jelből

Az 10. ábra értelmezése, hogy beérkező jel k . szakaszából veszi a szükséges paramétert, amiből előállítja az X_k paraméter-oszlopvektort. A küszöb-alapú algoritmusok úgy működnek, hogy a detektálás alatt kinyert paramétervektor alapján létrejön egy döntés, hogy beszéd vagy nem-beszéd a beérkező jel. Beszéddetektációs paramétereket úgy kapja meg, hogy azok különböző eloszlást mutatnak a beszédre és a zajra. A paraméterek számára szükséges egy megfelelő adatbázis.

8.3.1 Energiaküszöb alapú megközelítés

Energiaküszöb módszer úgy működik, hogy először a küszöböt állítja be, tehát felveszi a zaj energiaszintjét. Ennél a megközelítésnél a küszöbszintet többféle módon is beállíthatjuk, például csúszó ablakos energiaátlagolással, vagy a mintavett jel kezdő keretét megelőző intervallumból minimális energiaszintet választva. Mindkét lehetőség esetén ez időkéseletetést jelent.

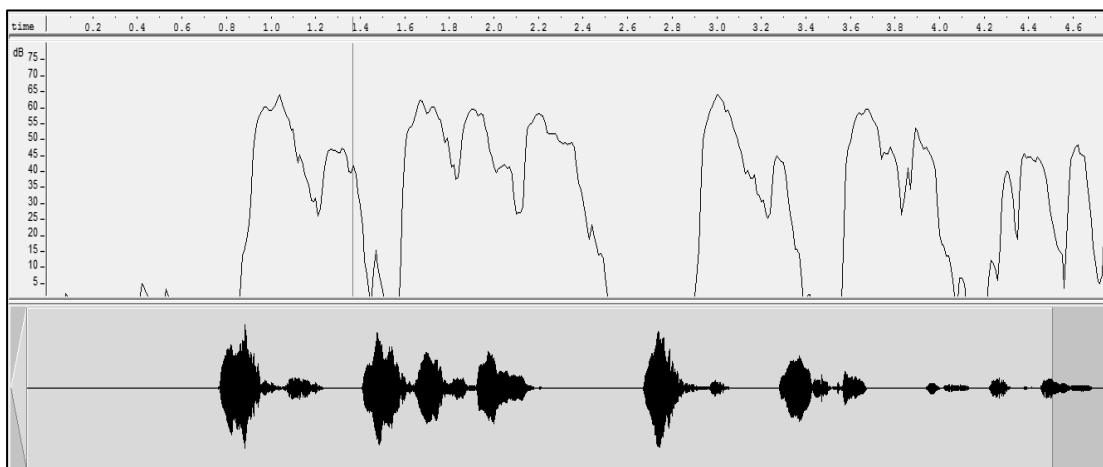
A mintavett jel energiáját (E_{jel}):

$$E_{jel}(t_0) = \sum_{t=t_0}^{t_0+T-1} y^2(t)$$

képlet szerint számolja, ahol T a minta hossza, t_0 a keret kezdete.

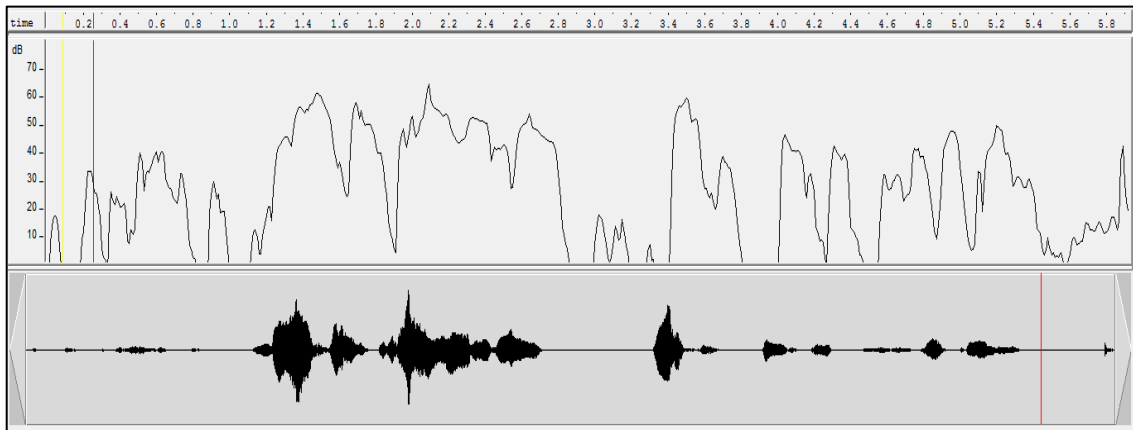
Beszédet az algoritmus úgy detektál, hogy beszéd nem beszéd döntésénél minimum 6dB-lel a küszöb fölé emelkedő szakaszokat tekinti. A beszédet pici szóközi szünetek jellemzik, ezért a túl hosszan tartó nagy energiájú részek esetén úgy veszi, hogy az eddigi jel is zaj volt. Tehát a küszöbszint újra megállapítása következik. Van úgy, hogy a beszéd nem ott kezdődik, ahol az algoritmus azt megállapítja, mert a mássalhangzók halkak, és nem mindig ütik meg a küszöbszintet.

Előnye, hogy a zajt nem kell ismernünk. Hátránya, hogy amikor nagy energiájú zaj van, akkor azt is beszédnek detektálhatja. Alacsony SNR (Jel-zaj viszony) esetén pedig a halk beszédet zajnak veheti. Tehát zajos környezetben ez a megközelítés nem hatékony. Lehetséges egy időeltolódás is.



11. ábra "Hol volt hol nem volt, volt egyszer egy kis vakond"

11. ábrán két diagram látható. Az alsón 8kHzen 8biten mintavételezett jel képét látjuk, a felsőn ennek az energia szintje szerinti ábrázolását. Viszonylag zajmentes a környezet.



12. ábra "Hol volt hol nem volt, volt egyszer egy kis vakond" _zajban

12. ábra a 11-től abban különbözik, hogy zajban vettem fel. A projekt szempontjából az energiaküszöb megközelítés nem elég pontos, és az időkéseletetés sem kecsegtető.

8.3.2 Gördülő amplitúdó spektrum alapján

Spektrális analízisben jól bevált a gördülő amplitúdó spektrum, angol nevén running-spectra. A módszer, hogy a beérkező jelet átlapolódó blokkokra bontja (10. ábra). A blokkokon FFT (Fast Fourier Transformation) hajtodik végre, aminek eredménye a jel gördülő spektruma:

$$Y_{jel}(f, t_0) = \sum_{t=0}^{T-1} y(t_0 + t) \cdot h(t) \cdot e^{-\frac{j2\pi ft}{T}},$$

ahol t diszkrét idő, $y(t)$ a vizsgált jel, f frekvencia, t_0 az aktuális keret kezdete, $h(t)$ a súlyozó ablak. Jól alkalmazható beszéddetektálásra, a zaj becslése mellett. A projekt szempontjából ez nekünk nem biztos, hogy elég. A következő módszer ennek továbbgondolása.

8.3.3 Spektrális entrópiaküszöb-módszer

Az információ elméletből vett entrópia egy véletlenváltozó bizonytalanságát írja le. A módszer a következő lépésekből áll [4] [2]:

Az előző fejezetben való elgondolás alapján megkapjuk a jel gördülő spektrumát. Majd a spektrum rendezettségét vizsgáljuk. Ennek mérésére entrópia számolást használunk. Az amplitúdó spektrum entrópiáját az alábbi képlet alapján kapjuk meg:

$$H(S) = -\sum_{i=1}^N P(s_i) \cdot \log\{P(s_i)\},$$

ahol s_i a forrásból származó i .szimbólum, $P(s_i)$ az i . szimbólum adási valószínűsége.

Ezek alapján a t . keret f frekvencián kiszámolt spektrumának entrópiája:

$$H(|Y_{jel}(f, t)|^2) = -\sum_{f=1}^F P(|Y_{jel}(f, t)|^2) \cdot \lg\{P(|Y_{jel}(f, t)|^2)\},$$

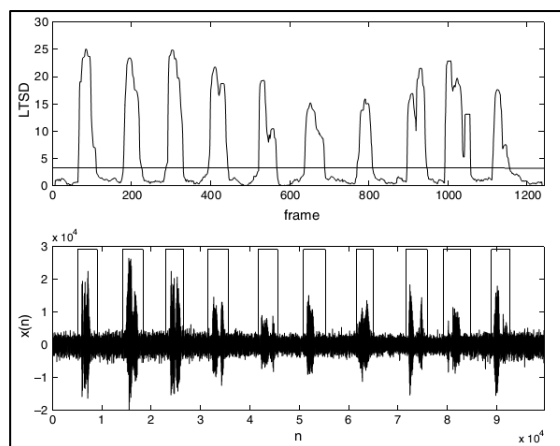
ahol:

$$P(|Y_{jel}(f, t)|^2) = \frac{|Y_{jel}(f, t)|^2}{\sum_{f=1}^F |Y_{jel}(f, t)|^2}.$$

A zaj és a beszéd más spektrális karakterisztikával rendelkezik, az entrópia számolása alkalmas lehet beszéddetektálásra. Ha a vizsgált jel fehérzaj az entrópia maximális, ha pedig tiszta szinusz, akkor minimális. Az entrópia értéke a jelszinttől független. Így változó szintű, állandó spektrális karakterisztikájú zaj esetén azokat a szakaszokat, ahol beszédet tartalmaz, jól detektálható. Adaptív és statisztikus módszerrel is megválasztható a küszöb. Nem állandó zaj bezavarhat a karakterisztikába, erre a szakirodalom zajbecslési eljárásokat javasol.

8.3.4 Hosszúidejű spektrális divergencián alapuló módszer

A jel gördülő amplitúdó spektrumából indul ki ez a módszer is. Az alapgondolat minden frekvencia komponensen SNR-t mérünk, és ezt átlagoljuk [11]. A szakirodalom szerint egyik legpontosabb módszer arra, hogy egy hangjelben elkülönítsük a beszédet a zajtól, ezt az alábbi ábrán is láthatjuk.



13. ábra Long-term spectral envelope, hosszú idejű spektrális divergencia

A módszer nagy előrettekintő ablak igénye miatt valósidejű beszéd zaj elkülönítésére nem felel meg.

8.3.5 Lineáris predikció

Beszédsspektrum burkolójának kinyerésére, a beszéd átvitele során lényegkiemelésre és tömörítésre is használják. A módszer: vesszük a beszédminták sorozatát, amit s_i -vel jelölök, ahol $0 < i < n-1$, n pedig a minták száma.

$$s_p(n) = -\alpha_1 \cdot s(n-1) - \alpha_2 \cdot s(n-2) - \dots - \alpha_n \cdot s(1)$$

A beszédjel n -edik mintája becsülhető a megelőző p beszédminta lineáris kombinációjával. α együtthatók meghatározása a becslés négyzetes predikciós hibájának minimalizálásával történik. Beszédfelismerőkhöz és beszéd szintetizátorokhoz használják.

8.3.6 Kepsztrum

A jel periodicitásának megfigyelésén alapszik a kepsztrum módszer. A beszédnek is jellemzője a szabályosság így ebből az irányból is érdemes megközelíteni a detektálást. Ez a módszer a következő [8][9]:

A kepsztrum értelmezhető a spektrum spektrumaként, mert a spektrum további Fourier-transzformáltját nevezzük kepsztrumnak. A spektrumot sok kisebb szinusz hullámkomponens összegének tekintjük. A spektrum minden hullámkomponensének amplitúdóját kepsztrum (spektrum Fourier transzformáltjaként kapott hullám) frekvenciájának függvényében ábrázoljuk. Egy magas periodicitású jelnek a spektrumban lesz egy nagy kiemelkedésű csúcs az alapprofrekvenciánál, és több kiemelkedése ennek többszöröseinél. Ezek a csúcsok szabályos intervallumokként helyezkednek el. Ez az intervallum felel meg a kepsztrum alapperiódusának. Így egy nagy kiemelkedésű csúcs lesz látható az alapperiódusnál. Egy alig periodikus jel spektrumában sok kiemelkedő csúcsú lesz sok frekvenciánál, amelyek nem követik egymást meghatározott szabályossággal. Így egy gyengén periodikus jel alacsony amplitúdójú sok kepsztrális csúcsot eredményez, különböző frekvenciákon. A kepsztrumba berajzolt lineáris regressziós görbével, amely reprezentálja az átlagos hangenergiát, kiküszöbölhetjük a hangkeltés amplitúdójának eltérését különböző szituációk és személyek között. A legnagyobb amplitúdóval rendelkező csúcs a kepsztrumban és a lineáris regressziós görbe közti különbséget kepsztrális csúcs kiemelkedésnek nevezzük. Gyengén periodikus jelnél ennek értéke kicsi, erősen periodikusnál pedig nagy. Alkalmazható beszéddetektálásra, de elég számításigényes.

8.3.7 Zero-Crossing

A zero-crossing módszer a null-átmenetek megfigyelése. Ehhez először a jel kepsztrumát kell számolni. Majd meg lehet figyelni, hogy periódusonként hányszor van null-átmenet. Ezekkel az átmenetekkel a zöngés zöngétlen mássalhangzókat jól meg lehet különböztetni. A beszéd detektálására is alkalmazható módszer, mert ezek a pozitívból negatívba váltások a beszédet jellemzik. Azonban a módszer számításigényes a beszéd aktivitás jelzéséhez, hiszen kepsztrumot kell számolni hozzá. Az eszköz további funkciójához jó módszer lehet a zero-crossing.

8.3.8 Mel-frekvenciás kepsztrum

A Mel-frekvenciás kepsztrum (továbbiakban MFCC³) abban különbözik a kepsztrumtól, hogy az MFCC a frekvenciák Mel-skáláját veszi, ami jobban illeszkedik a beszéd jellemzőihez. A módszer menete, hogy vesszük a jel amplitúdó spektrumát, ablak alkalmazásával, ami egyenletes frekvenciaosztásokkal tartalmazza az adott keret energiájának eloszlását. Az MFCC módszer az emberi hallás érzékenységet figyelembe veszi, modellezi, és ennek alapján használja a következő képletet az f frekvenciák átváltásához.

$$M(f) = 2595 \log_{10} \left(1 + \frac{f}{700} \right).$$

Így kapjuk meg a Mel-skálát. Majd a Mel-skálás eredményt DCT (Discrete Cosinus Transform) használatával átváltjuk. Az így kapott kepsztrum amplitúdói az MFCC együtthatói [2]. A módszert elterjedten alkalmazzák beszédfelismeréshez. Nekünk a zajt kell megkülönböztetni a beszédetől. Azt gondolom, hogy ezt kisebb számítási igénnyel is el lehet érni.

³ Mel-frequency cepstral coefficients

9 Spektrális entrópiaküszöb-módszer kiválasztása

Az irodalomkutatás és a beszédfelismerő módszerek vizsgálata során arra jutottam, hogy a spektrális entrópiaküszöb-módszer beszédaktivitás jelzésére alkalmas lehet számunkra, és érdemes egy tesztkörnyezetet kialakítani rá.

A spektrális entrópia alapú algoritmus nem használ nagy adatbázist, csak a küszöb paramétereire kell.

A [2] irodalom vizsgálata és ROC⁴ görbékkel való ábrázolása során arra jutott, hogy a hosszú idejű spektrális-divergencián alapuló és a spektrális entrópia módszer a két legjobban teljesítő beszéd-detekciós módszer. A vizsgálat során a nem-detektált beszédszakaszok arányát és a helyesen detektált szünetek arányát vizsgálta. A vibrotaktilis projekt szempontjából a hosszú idejű spektrális-divergencia alapú módszer nem jó választás, mert valósidejű rendszert szeretnénk.

9.1 Algoritmus implementálásának menete

A megvalósítást MATLAB környezetben végeztem.

Vegyük az inputnak 8biten és 8kHz mintavételi frekvencián mintavételezett hangjelet. A bejövő jelet 30ms-os keretekre osztottam fel. A keret méretének megválasztása fontos, mert túl nagy keret esetén a gyors válaszok kiátlagolódnak, túl kicsi méret esetén a beszédjellemzők lesznek meghatározhatatlanok. 10ms-os lépésekkel mentem a jelen végig, ez 8000Hz mintavételi frekvencia esetén (8000 minta másodpercenként):

$$n_3 = 0.03 \cdot f_s = 0.03s \cdot 8000Hz = 240 \text{ mintát jelent keretenként, és 66%-ban átlapoltak.}$$

Az átlapolódást úgy tudjuk kiszámolni, hogy $n_1 = 0.01 \cdot f_s = 0.01s \cdot 8000Hz = 80$, $n_3 - n_1 = 160$ minta, 8kHz esetén 30ms-os keretenként. Átlapolódásra azért van szükség, hogy a hangfeldolgozás során a beszédjellemzők változását nyomon tudjuk követni.

A tesztelést más paraméterekkel mintavételezett hangjellel is elvégeztem, az eljárás ugyanaz.

Majd jött egy előkiemelés, aminél elsőfokú FIR szűrőt használtam, ennek során az alacsonyabb frekvenciájú jelek kisimultak. Átviteli függvénye: $w(z) = 1 - 0.95z^{-1}$

⁴Receiver Operating Characteristic

Ezt követően 0.03s-os kereteknél Hamming ablakot használtam, hiszen nem periodikus a jel, és FFT-vel (Fast Fourier Transformation) előállítottam a spektrumot.

Hamming függvény:

$$h[n] = \begin{cases} 0.54 - 0.46 \cos\left(\frac{2\pi n}{L-1}\right), & 0 \leq n \leq L-1 \\ 0 & \text{egyébként} \end{cases}.$$

Az így kapott jel képlete keretenként:

$$Y_{jel}(f, t_0) = \sum_{t=0}^{T-1} y(t_0 + t) \cdot h(t) \cdot e^{-\frac{j2\pi t f}{T}},$$

Ahol t_0 a keret kezdete, t a diszkrét idő, $h(t)$ a súlyozó függvény, ami Hamming ablak.

A keletkezett komplex spektrum abszolút érték négyzetét veszem. Az elhagyott információ beszédjel szempontjából nem kell.

Következő lépés az információelméletből vett Shannon-féle entrópia kiszámítása. Végig mentem a kereteken és a komponensek adási valószínűségét számoltam.

$$P(|Y_{jel}(f, t)|^2) = \frac{|Y_{jel}(f, t)|^2}{\sum_{f=1}^F |Y_{jel}(f, t)|^2}$$

A keret spektrális rendezettségét

$$H(|Y_{jel}(f, t)|^2) = - \sum_{f=1}^F P(|Y_{jel}(f, t)|^2) \cdot \lg\{P(|Y_{jel}(f, t)|^2)\}$$

képlet alapján kapom meg.

9.2 Sepktrális-entrópia módszerrel való mérések analízise

Több szempontot figyelembe véve teszteltem a MATLAB környezetben megvalósított algoritmust. Az egyik ilyen volt a megbízhatóság. Igyekeztem többféle mérési körülményt teremteni a tesztelésre. Különböző háttérzajok mellett rögzített beszédre végeztem teszteket és többféle minőségben vettem fel a hangjeleket. 8, 16 biten 8kHz és 16 kHz mintavételi frekvencián mintavételezett hangadatokon is futattam a tesztet. Másik ilyen szempont volt a gyorsaság, a tesztek során mértem a lefutási időt. A mérésekhez a hangjeleket számítógép beépített mikrofonjával vettem fel.

Az entrópia módszernél úgy állapíthatjuk meg a küszöbértékeket, hogy 4 alatti értékeket beszédnek tekintjük, a másik küszöbérték pedig a nagyon rövid 4 alatti részekre fog határt adni. Így nem tekintjük beszédnek 0.2 másodpercnyi 4 küszöbérték alatti részeket. Ez azt jelenti a projekt szempontjából, hogy a folyamatos beszédetektálás során, 0.2s késleltetésünk már biztos, hogy lesz, ha az algoritmus megfelel.

Mérés 1:

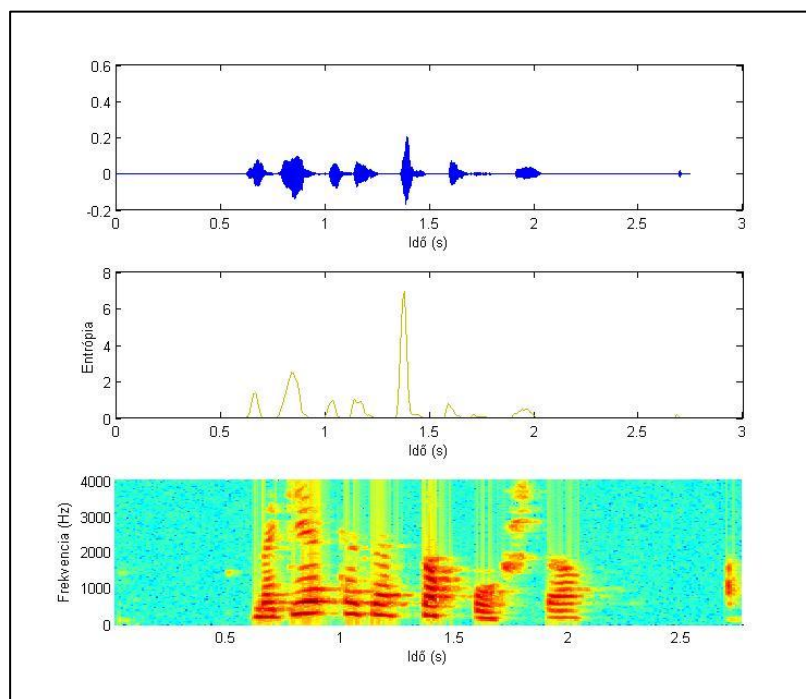
Az alábbi négy mondatot rögzítettem. Az első négy mondatot 8biten 8kHz mintavételi frekvencián vettem fel. A mérést megismételtem 8bit, 16kHz és 16bit, 16kHz mintavételezési minőségben. A projekt szempontjából nagyobb mintavételezési számokat nem érdemes venni, az eszköz használatánál erre nincs mód.

Mondatok:

88_1	Megérkezett a posta.
88_2	Kész van az ebéd.
88_3	A vonat 5 perc múlva indul.
88_4	Hat órákor kezdődik a műsor.

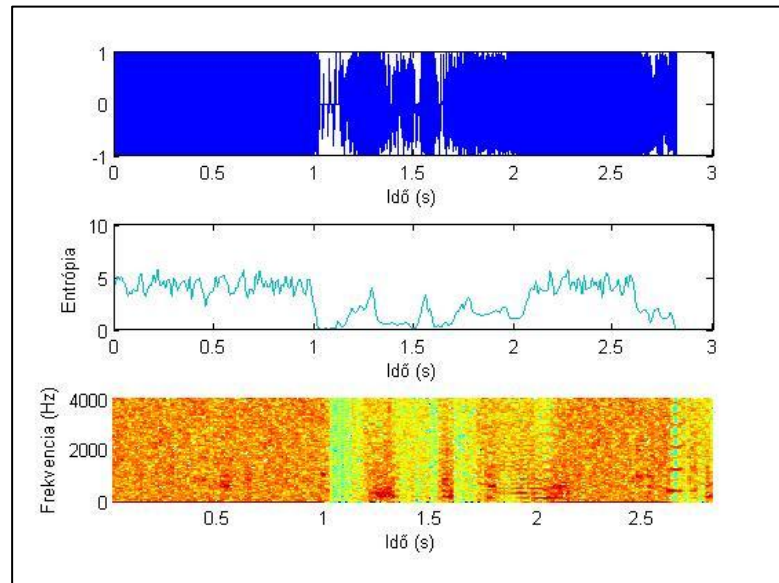
4 féle mondat				
	88_1	88_2	88_3	88_4
Spektrum kiszámítása (s)	0,0924	0,0736	0,0989	0,0919
Spektrum-entrópiáig számítás (s)	0,2657	0,198	0,2705	0,239
Teljes (s)	0,3581	0,2716	0,3694	0,3309
n (bit)	8	8	8	8
Mintavételi frekvencia (kHz)	8	8	8	8

Ebből az eredményből azt a következtetést vonhatjuk le, hogy a spektrum kiszámításhoz kevesebb, mint feleannyi idő kell, mint utána a spektrum entrópiájának kiszámításához. A számítások a gépen gyorsan lefutottak.



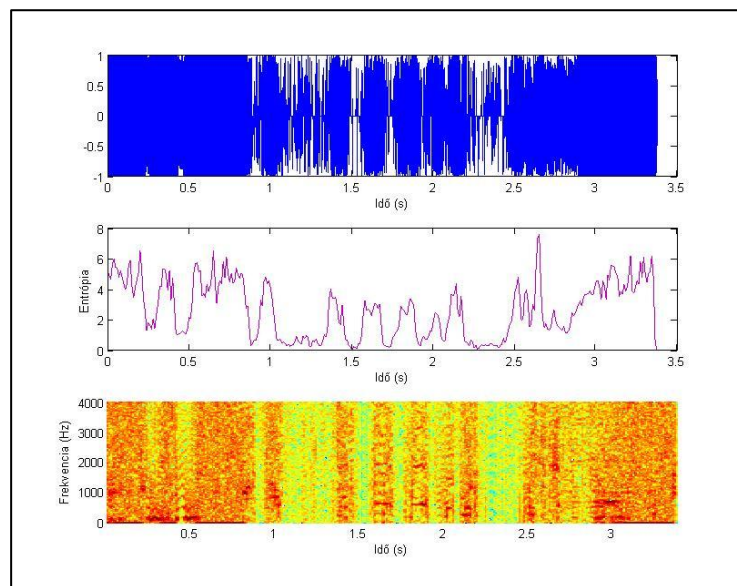
14. ábra 16kHz, 8 bit "Megérkezett a posta"

A 14. ábra zajtalan környezetben. Az algoritmus jól alkalmazható, a spektrum entrópia módszer ott jelzett ahol kell. Az alábbi kép nagyobb kihívás a beszéd-detekciós megoldásoknak.



15. ábra (88_2) "Kész van az ebéd"

A 15. és a 16. ábrán az első függvény a mintavételezett jelet mutatja. Látható, hogy rossz minőségű és zajos jel van mind a két ábrán. 8biten 8kHz-en mintavételezett. A második függvény pedig a spektrum-entrópia ábrázolása. Mint ahogy fentebb is említettem ez a módszer a spektrum rendezettségét adja ki. A 15. ábrán látványosan elkülönül a középső rész az időtengelyen nézve 1 másodperctől.

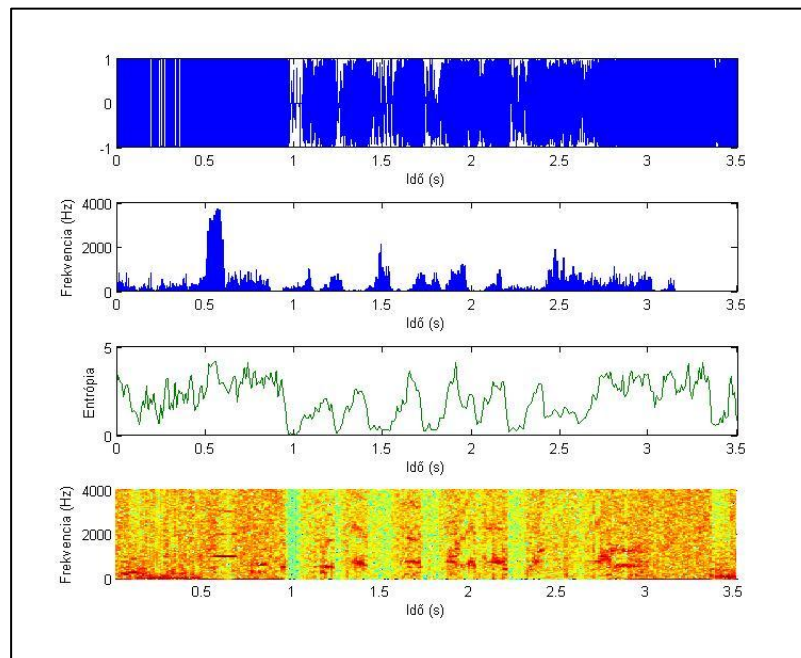


16. ábra (88_4) "Hat órákor kezdődik műsor"

Mérés 2:

Egy mondatot néztem meg négy minőségben:

<i>"Összetörték a poharakat."</i>				
	besz1	besz2	besz3	besz4
Spektrum kiszámítása (s)	0,119312	0,284	0,1806	0,1429
Spektrum-entrópiáig számítás (s)	0,246611	0,572007	1,390536	1,168449
Teljes (s)	0,365923	0,856007	1,571136	1,311349
n (bit)	8	16	8	16
Mintavételi frekvencia (kHz)	8	8	16	16



17. ábra "Összetörték a poharakat"

17. ábra a táblázatban levő „besz1” nevű oszlophoz tartozik. Változó zajú rossz minőségű jel az input. A spektrumát látjuk a második koordináta-rendszerben. A negyedikben a spektrogramot, és a harmadikban az entrópiát. Az időtengelyen azt látjuk, hogy egy magas energiájú zaj kicsit elnyomta a beszédszakaszt. Az algoritmustal ellátott vibrotaktilis eszközön azt tapasztalnánk, hogy be majd ki kapcsolna a tényleges beszédszakasz megkezdése előtt.

10 Konklúzió, továbblépés

A kiválasztott algoritmus használható beszéd-detektálásra, ha a háttérzaj egyenletes. A projekthez a környezet változatossága miatt, szükség lesz zajbecslésre. A mérések azt mutatták, hogy ha változatos zaj van a környezetben, és nagy energiájú, akkor rendezettebb spektrummal rendelkezhet megközelítve a beszédet, és az entrópia-idő diagramon nehezebb lenne, vagy nem lehet elkülöníteni a beszéd-től. Akkor is bejelezné a készülék, amikor nem kell. Ezeknek a téves jelzéseknek a számát a továbbiakban le kell redukálni. Megoldás lenne, ha a zajt meg tudnánk becsülni, ezt jelöljük Y_{ismert} . Az irodalomban a zajbecslésre több módszer is megtalálható. Létezik olyan, hogy az átlag spektrummal osztja az aktuális keret spektrumát. Egy másik ilyen, hatékonyabb becslésnek bizonyul, hogy a keretet a megközelítőleg ismert zaj spektrumával osztja.

$$Y_{akt}(f, t_0) = \frac{Y(f, t_0)}{Y_{ismert}}.$$

Az Y_{ismert} becslésére is van többféle lehetőség. Található az irodalomban olyan, ami a múltbeli jelek spektrumának minimumát veszi, ami nem lehet a valós zaj frekvenciájánál nagyobb. Az így kapott becslött zajjal fehéríteni próbálja a bejövő jelet. Aztán létezik olyan zajbecslés, ami szerint nem csak múltbeli jelből becsül zajt, hanem a bekövetkező jelből is, csak a múltbelitől kisebb időintervallumban. Pontosabb, de ez egy további időkésleltetést eredményez a detektálás során. Zajbecslésnél mérlegelni kell az idő késleltetést, a gyorsaság fontos, viszont le kell csökkenteni a téves bekapcsolások számát. A projekt szempontjából lehetne egy múltbeli zajbecslést végezni. Az eszköznek minél gyorsabban kell reagálni beszédaktivásra, lehet ez fontosabb szempont, mint annak elkerülése, hogy esetenként bekapcsoljon tévesen. Mert téves bekapcsolást követően is folyamatosan vizsgál a módszer és számításai során észlelni fogja, hogy hirtelen megváltozó zajt detektált, akkor visszkapcsol takarékos üzemmódra. A másik oka ennek, hogy nem probléma, ha az eszközt használó is észreveszi egy jelzés végett, hogy megváltoztak a zaj viszonyok. Ez nem okoz hátrányt, ha belegondolunk egészséges hallással is észleljük, ha hirtelen megváltozik a zaj körülöttünk. Például, megyünk az utcán, beszélgetünk partnerünkkel és hirtelen egy mentő autó szirénáját haljuk. Akarva akaratlanul felfigyelünk rá, készülék is bejelezné. Továbbá a projekt szempontjából, lehetnének olyan funkciók az eszközön, különböző üzemmódok, otthoni (csendesebb), utcai (zajosabb, színesebb zaj), tárgyaló üzemmód például, amivel „rásegítenénk” az algoritmusnak

egy teljesebb működésre. Ez azt jelentené, hogy a felhasználónak egy készüléken való gombnyomással kiválaszthat egy működési üzemmódot. Ami beállítana egy kezdő Y_{ismert} -et.

11 Összefoglalás

Jelen dolgozatomban áttekintettem a vibrotaktilis eszköz elvét, működését, felhasználhatóságát a siket emberek beszédmegértési nehézségeinek leküzdéséhez. A szájról olvasáson alapuló beszédmegértési eljárás szempontjaiból tett problémafelvetést követően, ismertettem az eszköz követelményeit, és azok tudatában mutattam be a beszédfeldolgozásban alkalmazott módszereket, különös tekintettel a beszéd-zaj elkülönítésére alkalmas beszéd-detekciós algoritmusokat. Az áttekintést követően kiválasztottam egy lehetséges megoldást. A kiválasztott módszert megvalósítottam és elemeztem a kapott eredményeket. A tesztelés eredményeinek tudatában meghatároztam az algoritmus továbblépési lehetőségeit az eszközhöz, és az eszköz fejlesztésére további javaslatot tettem.

Köszönetnyilvánítás

Szeretnék köszönetet mondani a családomnak, akik mindig mindenben támogattak és mellettem álltak.

Tihanyi Attila Tanár Úrnak, a konzulensemnek, aki segített megérteni a feladatomat, türelemmel és nagy segítséggel koordinálta a munkám menetét.

Valamint a Pázmány Péter Katolikus Egyetem Információs Technológiai Karának, ahol a munkámat végeztem.

Ábrajegyzék

1. ábra: A kommunikáció csatornái [1]	9
2. ábra Stimuláció.....	11
3. ábra Vibrotaktilis eszköz	11
4. ábra Tervezett taktilis kijelező logikai elrendezése [14]	11
5. ábra PIC24FJ16MC102	12
6. ábra Beszédképzésben résztvevő szervek.....	13
7. ábra Fonetikai és gépi feldolgozás alapján beszédhangjaink.....	16
8. ábra Zár-, zár-rés-, réshangok, nazálisok [6]	17
9. ábra Zár-rés hangok.....	17
10. ábra Paramétervektor előállítása a jelből.....	19
11. ábra "Hol volt hol nem volt, volt egyszer egy kis vakond"	20
12. ábra "Hol volt hol nem volt, volt egyszer egy kis vakond" _zajban.....	21
13. ábra Long-term spectral envelope, hosszú idejű spektrális divergencia.....	22
14. ábra 16kHz, 8 bit "Megérkezett a posta"	28
15. ábra (88_2) "Kész van az ebéd".....	29
16. ábra (88_4) " Hat órákor kezdődik műsor".....	29
17. ábra "Összetörték a poharakat"	30

Irodalomjegyzék

- [1] E. Jókai, K. Koloszar, S. Mogánné Tölgyesy és M. Pataki, Rehabilitációs támogató technológiák, Budapest: Typotex, 2010.
- [2] Z. Tüske, P. Mihajlik, Z. Tobler, T. Fegyó és P. Tatai, „Beszéddetekciós módszerek vizsgálata és optimalizálása gépi beszédfelismerő rendszerekhez,” *Híradástechnika*, p. 59, 2006.
- [3] T. György, Szerző, *Emberi beszédkeltés és beszédérzékelés kurzus*, 2012.02.16. [Performance]. 2012.
- [4] P. Renevey és A. Drygajlo, *Entropy Based Voice Activity Detection in Very Noisy Conditions*, Switzerland, 2001.
- [5] Z. Tüske, P. Mihajlik, Z. Tobler és T. Fegyó, „Robust Voice Activity Detection Based on the Entropy of Noisesuppressed Spectrum,” *Eurospeech*, 2005.
- [6] G. Olaszy, „A korpusz alapú beszédszintézis nyelvi, fonetikai kérdései,” *Híradástechnika*, 2006.
- [7] J. Ramírez, J. M. Górriz és J. C. Segura, „Voice Activity Detection. Fundamentals and Speech Recognition System Robustness,” 2007.
- [8] *Beszéd- és hallásdiagnosztika előadás*. [Performance]. BME, 2012.
- [9] J. A. Haigh és J. S. Mason, *Robust Voice Activity Detection using Cepstral Features*, Swansea, UK.
- [10] J. A. Haigh és J. S. Mason, *A Voice Activity Detector BasedOn Cepstral Analysis*, Swansea, UK.
- [11] J. Ramírez, J. C. Segura és C. Benítez, „Efficient voice activity detection algorithms using long-term Speech information,” *Speech Communication*, 2004.
- [12] MathWorks, *MATLAB R2012b*, 2012.
- [13] I. Shafran és R. Rose, *Robust Speech Detection And Segmentation For Real-Time ASR*

Application, 2003.

[14] G. Dr. Lénárd, „Biológia”.

[15] A. Eszter, „Taktilis kijelző alkalmazások,” 2005.

[16] J.-w. H. L.-s. L. Jia-lin Shen, „Robust Entropy-based Endpoint Detection for Speech Recognition in Noisy Environments,” Taiwan, Republic of China.

[17] G. Gordos és G. Takács, „Digitális beszédfeldolgozás,” Műszaki Könyvkiadó, Budapest, 1983.

[18] A. Kovács, *Siketek beszédértést segítő taktilis kijelző készítéséhez zajtűrő beszédfelismerő algoritmusok áttekintése, fejlesztése*, Budapest, 2012.

[19] G. Németh és Olasz Gábor, *A magyar beszéd*, Budapest: Akadémiai Kiadó, 2010.

Mellékletek

Szakdolgozat teljes anyaga és tartalmi kivonata az elkészített kódokkal megtalálható a szakdolgozathoz tartozó CD-n, és az alábbi linken:

http://users.itk.ppke.hu/~fehan/bsc_szakdolgozat