Multidiszciplináris tudományok, 10. kötet. (2020) 4 sz. pp. 80-91 **https://doi.org/10.35925/j.multi.2020.4.11**

AKUSZTIKUS BESZÉDFELDOLGOZÓ EGYSÉG HATÉKONYSÁGI VIZSGÁLATA KÜLÖNBÖZŐ ZAJOK ESETÉN

Pintér Judit Mária
tudományos főmunkatárs, Automatizálási és Infokommunikációs Intézet
3515 Miskolc, Miskolc-Egyetemváros, e-mail: pinterjm@uni-miskolc.hu

Absztrakt

Az ember-gép kommunikáció során az egyik legfontosabb kritérium a megbízhatóság. Ez az elvárás magába foglalja azt, hogy az észlelt információ megegyezzen a küldő által kibocsátott jel információtartalmával. Tudjuk, hogy ez emberek közötti kommunikáció egyik legalapvetőbb eszköze, és egyben legegyszerűbb és legtermészetesebb módja a beszéd. Ezen állítást szem előtt tartva az ember régi vágya, hogy az általa konstruált gépekkel, berendezésekkel emberi nyelven, a beszéd eszközével tudjon hatékonyan és megbízhatóan kommunikálni. Munkám során teszteléseket fogok végezni egy rejtett Markov-modelleken alapuló beszédfelismerőn, ahol különböző zajokkal terhelt tanító és tesztelő anyagok hatékonyságra gyakorolt hatásait fogom vizsgálni.

Kulcsszavak: *beszédfelismerés, zaj, rejtett Markov-modell, hatékonyság vizsgálat*

Abstract

One of the most important criteria in human-machine communication is reliability. This expectation includes that the detected information be the same as the information content of the signal emitted by the sender. We know that one of the most basic means of communication between people and at the same time the simplest and most natural way of doing so is speaking. With this statement in mind, it is man's old desire to be able to communicate effectively and reliably with the machines and devices he constructs in human language, with speech. During the course of my work, I will perform tests on a speech recognizer based on hidden Markov models. I will strive to select the most optimal number of states for the hidden Markov models, and then I will investigate the effects of different noise-laden teaching and testing materials on efficiency.

Keywords: *speech recognition, noise, hidden Markov model, efficiency study*

1. Bevezetés

Az ember-gép kapcsolatot szűkebb és tágabb értelemben is szokás definiálni. A kapcsolat szűkebb értelmezése az ember és gép közvetlen kapcsolódási felülete, tágabb kört vizsgálva pedig nem más, mint az ember-gép együttélés. Ebből adódóan a szűkebb értelmezés a közvetlen kommunikációs eszközökről, a fizikai kapcsolatról és annak feldolgozásáról, irányításról szól. Amíg a tágabb értelmezés a kommunikációs felületen túlmenően tartalmazza az emberek és gépek alkalmazkodását egymáshoz, valamint az egymásra hatásokat is. [1] Az ember-gép kommunikáció során az egyik legfontosabb kritérium a megbízhatóság. Ez az elvárás magába foglalja azt, hogy az észlelt információ megegyezzen a küldő által kibocsátott jel információtartalmával. Annak érdekében, hogy ennek a feltételnek eleget tegyünk, csak olyan eszközökkel tudjuk az ember-gép közötti kommunikációt megvalósítani, amelyek lekötik valamely testrészünket. Ezekre evidens példaként az egér, a billentyűzet, vagy az érintőképernyő szolgálhat, mely eszközök a látó szervünkkel és a tapintó képességünkkel együtt segíti a kommu-

nikációt. Tudjuk, hogy ez emberek közötti kommunikáció egyik legalapvetőbb eszköze, és egyben legegyszerűbb és legtermészetesebb módja a beszéd. Ezen állítást szem előtt tartva az ember régi vágya, hogy az általa konstruált gépekkel, berendezésekkel emberi nyelven, a beszéd eszközével tudjon hatékonyan és megbízhatóan kommunikálni. A természetes nyelvű ember-gép dialógusnak a beszédmegértésre irányuló elemét nevezzük gépi beszédészlelésnek. A beszédészlelés terminus gyakorlatilag mindent magába foglal abból, amit az ember a beszédpercepció során megállapíthat a másik ember beszédéből (tartalom, beszélő attitűdje, érzelmei, fizikai állapota stb.).

A gépi beszédészlelést megvalósító alkalmazások tipikusan a gépi beszédfelismerők, amelyek puszta beszéd-szöveg átalakítást végeznek anélkül, hogy a beszédben hordozott jelentést megérteni képesek lennének. A beszédfelismerés fejlődése során több módszert is kifejlesztettek ennek megvalósítására.

A rejtett Markov-modell (HMM) máig használatos módszer az akusztikus beszédfelismerésben, melynek bemenetét a jellegvektorok vagy lényegvektorok (Feature vector) képezik. Viszont ez a modell sem tökéletes, hiszen nehézségeket jelent, ha a bemenet bizonytalan, vagy zajjal terhelt, ha sok a hasonló szótárelem, továbbá a mindig problémát jelentenek a kiejtésbeli eltérések.

Vizsgálataim során teszteléseket fogok végezni egy rejtett Markov-modelleken alapuló beszédfelismerőn, aminek célja ipari környezetben való vezérlés megvalósítása. Napjainkban a beszédjelek kiemelése a zajos környezetből, vagyis a zajjal terhelt beszédjelek javítása kiemelt kutatási téma. Ennek oka a számos felhasználási lehetőség, amellyel egy hatékony beszédfelismerő rendszer rendelkezik. Ma már számos technológiai eljárás létezik ennek megvalósítására. Általánosságban elmondható, hogy a mai kifejlesztett beszédfelismerők megcélzott felhasználási környezete alacsony zajszintű. Éppen ezért egy ilyen felismerő alkalmazása zajos környezetben akkor lehetséges, ha a felismerő bemenetére már zajszűrt beszédjel kerül, vagy a felismerő akusztikai elő-feldolgozó eljárását zajtűrő eljárásra cseréljük. A zajszűrés külön problémát jelent különösen olyan esetekben, amikor változó, hol állandó, hol impulzusszerű, valamint változó hangszintű és –színezetű zaj váltakozva van jelen [2].

A már fentebb említett alacsony zajszintű beszédfelismerőknél a tanító anyagok jó minőségűek, azaz zajmentesek. Viszont ha a beszédfelismerőt zajos mintákkal tanítjuk be, akkor valószínűleg nagyobb hatékonysággal fogja felismerni a zajjal terhelt bemenetet anélkül, hogy zajszűrést végeznénk.

A vizsgálataim során a tanító és a tesztelő mintákat hat különböző szintű fehér zajjal terheltem, és minden tanító anyag esetén (beleértve a zajmentes mintákat is) elvégeztem a betanítást. A tanítás során nem használtam semmilyen zajszűrő eljárást, a lényegkiemelés konfigurációja minden esetben megegyezett azzal, amit a zajmentes hangmintáknál is használtam, majd megvizsgáltam, hogy a hat zajos plusz a zajmentes teszt anyagokat mennyire hatékonyan ismeri fel. A hangminták zajjal terhelését a Matlab szoftverrel végeztem el, a betanítási és tesztelési vizsgálatokat pedig a HTK toolkit szoftvercsomaggal.

2. Az operátori kezelőfelület vezérléshez szükséges beszédfelismerő

A tervezett felismerő elkészítésének elsődleges célja a navigáció megkönnyítése és hatékonyabbá tétele volt. Bizonyos funkciók előhívása, vagy több ugyanabból a lépésekből álló utasítások, kiértékelési folyamatok végrehajtása kiváltható lenne egészen rövid kifejezéseket tartalmazó szóbeli paranccsal. Ezeknek a kifejezéseknek természetesen egyértelműnek kell lenniük, azaz tartalmazniuk kell például egy rögzített kifejezést amivel ”jelezzük” a rendszer felé, hogy egy utasítást fogunk közölni, nem pedig éppen csak beszélgetünk valaki mással. A ”nyitó” kifejezést követően pedig következhet a fő utasítás, például adott funkció megnyitása, vagy értékek módosítása, lekérdezése stb., vagyis adottak lesznek kulcsszavak, amikkel egyértelműsíthető lesz a parancs.

Figyelembe kell venni azt is, hogy a felületet nem csak egy adott személy fogja nagy valószínűséggel használni, ezért a szóbeli navigációnak ugyanolyan jó hatékonysággal kell majd működnie, minden egyes ember esetén. Ezt két féle módon valósítható meg. Létrehozhatunk egy beszélőfüggetlen felismerőt, aminek minden személy esetén megfelelően kell működnie, vagy minden felhasználó esetén külön tanítási folyamatot hajtunk végre és profilszerűen az adott személyhez kapcsolódó beszéd modelleket (rejtett Markov-modellek) alkalmazzuk a beszédfelismerőben, és attól függően, hogy ki az aktuális felhasználó úgy töltődik be az megfelelő konfiguráció. Elsődlegesen egy beszélőfüggetlen felismerő létrehozása mellett döntöttem. Triviálisnak tűnő elvárás, de mégis meg kell jegyeznünk, hogy a szóbeli navigáció lehetőségének folyamatosan adottnak kell lennie, és alkalmazkodnia kell a folyamatos emberi beszédhez. Az igények, és a kritériumok áttekintése után összegezhetjük, hogy a navigáció megvalósításához pontosan milyen beszédfelismerőre is van szükség:

- kötött szótáras (kulcsszó alapú felismerés);
- mintafelismerő (rejtett Markov-modellek alkalmazása);
- kapcsolt szavak felismerésére alkalmas;
- beszélőfüggetlen;
- zajos környezetben alkalmazható;
- parancs üzemmódú;
- online üzemmódú.

3. Rejtett Markov-modell

A felismerés alapját képező rejtett Markov-modell diszkrét sztochasztikus folyamatot ír le. Az a fogalom, hogy valami Markov-tulajdonságú azt jelenti röviden, hogy adott jelenbeli állapot mellett, a rendszer jövőbeni állapota nem függ a múltbeliektől. Másképpen megfogalmazva, ez azt is jelenti, hogy a jelen leírása teljesen magába foglalja az összes olyan információt, ami befolyásolhatja a jövőbeli helyzetét a folyamatnak. [3]

E modelleket a tudomány számos más területén – fizika, statisztikai folyamatok, internet, matematika, biológiai modellezés, gazdasági elemzések, szerencsejátékok - is alkalmazzák. A Markov-modellek bonyolultabbak a döntési fa modelleknél, de lényegesen kevesebb programozói ismeretet és kisebb adatmennyiséget igényelnek, mint a szimulációs modellek. A Markov-modell magában foglalja a döntési fa lényeges tulajdonságait, és ezen felül már az események bekövetkezésének idejét is figyelembe tudja venni. A "rejtett Markov-modell" [4] kifejezésben a "rejtett" jelző arra utal, hogy mi csak a modell működésének az eredményét, a kimenetet (azaz a generált szekvenciát) ismerhetjük, a modell maga és a paraméterei számunkra ismeretlenek. Így mi csak a kimenetből következtethetünk a modell felépítésére és a működését leíró paraméterekre (az átmeneti és a kibocsátási valószínűségekre).

A szótár minden egyes eleméhez tanulással - approximációs eljárással - el kell készíteni egy-egy Markov-modellt, majd a felismerés során a kiejtett elemhez ki kell számítani minden modell esetén azt a valószínűséget, amely valószínűséggel a modell ezt az elemet ilyen kiejtéssel generálhatta. Ha ezek között a valószínűségek között van pontosan egy kiemelkedő, akkor a felismerés sikeres, és a kiemelkedő valószínűséghez tartozó szótári elem lesz az eredmény. (A rejtett Markov-modell érzékeny a túltanulásra.) Tehát az ilyen modellekre épülő beszédfelismerés tisztán statisztikai alapú. A HMM előnye, hogy elég egyszerűen kiterjeszthető nagyszótáras, folyamatos beszéd felismerésére.

4. A lényegkiemelés módja

A lényegkiemelés a hallás bemutatása során megismert biológiai jellemzők ismeretében valósítható meg. A HTK toolkit keretrendszer tartalmaz egy olyan feldolgozó egységet programot, amivel elvégezhető az átalakítás, vagyis a hangfájlok fájlok lényegkiemelt fájlokká konvertálhatók. Munkám során ezzel valósítottam meg a lényegkiemelést. A folyamat a jelből megkísérli meghatározni a beszéd tartalmát hordozó mennyiségeket, azaz a fontos információkat, és kiküszöbölni a felismerés szempontjából érdektelen információkat (zaj, fázis, torzítások). A digitalizált (beszéd) jelből egy diszkrét idejű, adott dimenziójú lényegvektor-sorozatot alkot. Fontos információ alatt itt a mel-frekvenciás kepsztrális komponenseket (Mel-Frequency Cepstral Coefficient / MFCC) értjük [5]. Az MFCC jellemzők a beszéd lényegi tartalmának kinyerésére szolgálnak és napjainkban a beszédfelismerő rendszerek jelentős többsége e jellemzők (vagy ennek kicsit módosított változatai) segítségével próbálják reprezentálni a beszéd lényegi információ-tartalmát.

5. A zaj

A zaj több eltérő frekvenciájú és intenzitású jel zavaró összessége. A jelek forrása és frekvenciaspektruma attól függ, milyen zajról van szó. Az információelméletben a zaj csökkenti a kommunikációs csatornán átvihető információmennyiséget, azaz csökkenti a csatorna kapacitását. Különféle kódolási eljárásokkal csökkentik ennek a zajnak a hatását. A zaj leírására nem alkalmas egyetlen szám (például a hang intenzitása), ezért azt többnyire egy színképpel írjuk le. A színkép egy adott zaj vagy hang hangnyomásszint értékeinek a frekvencia függvényében történő ábrázolása. Megkülönböztetünk folytonos színképet, amikor nagyjából minden frekvencián van valamekkora hangnyomásszint érték, és vonalas színképet, amikor csak bizonyos frekvenciákon van hangnyomásszint.

Néhány színképnek gyakorlati jelentőségük miatt külön nevet adtak. Ilyen a fehérzaj, szürke zaj és a rózsaszín zaj. [6]

- **Fehérzaj:** a fehérzaj olyan, hangtechnikában használatos véletlenszerű zaj, amire igaz az, hogy a teljes vizsgált frekvenciatartományban (emberi érzékelő esetén 20 Hz – 20 kHz) a hangnyomásszintje állandó.
- **Szürke zaj:** a szürke zaj esetén egy jól meghatározott, szűk frekvenciatartományban folytonos hangnyomásszint van, míg az összes többi frekvencián nem mérhető hangnyomásszint.
- **Rózsaszín zaj:** olyan zaj, melynek hangnyomásszintje a frekvenciával fordítva arányosan esik, és az olyat is, melynek a hangnyomásszintje a frekvencia négyzetével fordítottan arányosan esik. A rózsaszín zaj a hangtechnikában egyértelmű jelentéssel bír: véletlenszerű zaj, amelynek a teljes vizsgál frekvenciatartományban (jellemzően 20 Hz – 20 kHz) a hangnyomásszintje oktávonként 3dB-lel csökken.
- **Színes zajok:** az olyan zajokat, melyek frekvenciája határozottan nem állandó értékű, de gyakorlatilag jól meghatározható frekvenciasávba esik, színes zajoknak nevezik. A fehérzajtól eltérően nincs a különféle színes zajspektrumoknak általánosan elfogadott meghatározása.

Emiatt a többértelműség miatt a tudományos cikkek az 1/f zaj fogalmat olyan folyamatokra alkalmazzák, melyek zaj-teljesítménysűrűsége fordítottan arányos a frekvenciával.

6. A jel-zaj viszony

A jel-zaj viszony egy műszaki kifejezés, ami két teljesítmény hányadosát jelenti. A jel (információ) és a háttér zaj teljesítményének hányadosa [6]:

$$\text{jel/zaj} = \frac{P_{\text{jel}}}{P_{\text{zaj}}} = \left(\frac{A_{\text{jel}}}{A_{\text{zaj}}}\right)^2 \tag{1}$$

A jel/zaj meghatározásánál a logaritmikus decibelskálát használják. Decibelekben mérve, a jel-zaj viszony az amplitúdók hányadosának 10-es alapú logaritmusának 20-szorosa vagy a teljesítményarány logaritmusának 10-szerese:

$$\text{jel/zaj(dB)} = 10\log_{10}\left(\frac{P_{\text{jel}}}{P_{\text{zaj}}}\right) = 20\log_{10}\left(\frac{A_{\text{jel}}}{A_{\text{zaj}}}\right) \tag{2}$$

ahol *P* az átlagos teljesítmény, *A* az amplitúdók négyzetes átlaga. A jeleket és a zajokat azonos sávszélességi rendszerben mérik. A decibel definíciójának megfelelően a jel-zaj viszony azonos eredményt ad függetlenül attól, hogy a jel milyen jellemzői alapján számítjuk ki (teljesítmény, áram, feszültség). A jel-zaj viszonyt gyakran használják egy átlagos jel-zaj viszony indikátorának, ezért lehetséges, hogy egy pillanatnyi jel-zaj viszonyt teljesen eltérően értékelnek. Általában a magasabb jelszint azonos zaj mellett kedvezőbb; ilyenkor a jel „tisztább” [6]. A vizsgálatokhoz a tanító és a tesztelő hanganyagokat 6 különböző zajszinttel terheltem: -6 dB, 0 dB, 6 dB, 12 dB, 18 dB, és 24 dB.

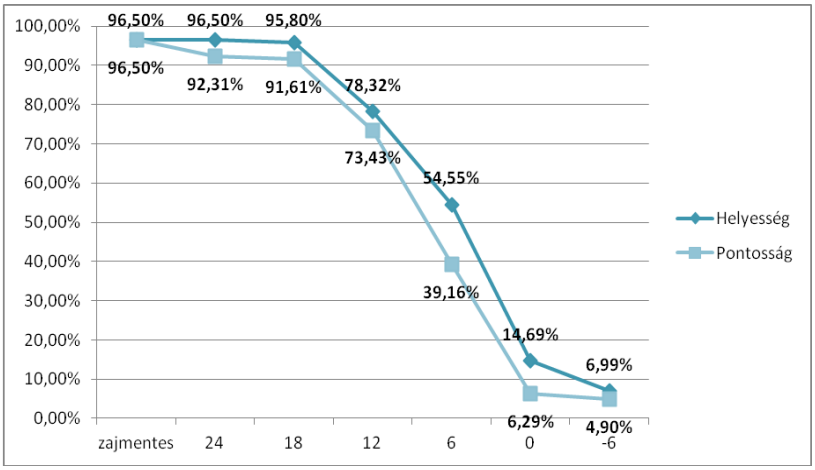
7. Elvégzett hatékonysági vizsgálatok

7.1 Zajmentes tanító minták vizsgálata

1. táblázat. *A zajmentes tanítóanyagokra elvégzett tesztek eredményei*

Teszt anyagok jel/zaj viszonya (dB)	Helyesség	Pontosság
zajmentes	96,50%	96,50%
24	96,50%	92,31%
18	95,80%	91,61%
12	78,32%	73,43%
6	54,55%	39,16%
0	14,69%	6,29%
-6	6,99%	4,90%

Az 1. ábrán megfigyelve a helyességi és pontossági görbét láthatjuk, hogy az értékek a zaj növekedésével csökkennek. A 18dB és 0 dB közötti értékek mutatnak leginkább nagyfokú romlást. -6dB-nél a helyesség már csak 6,99% a pontosság pedig 4,90%, vagyis a felismerő ebben az esetben teljesen használhatatlan. Összességében megállapítható a görbe és az 1. táblázat értékei alapján, hogy ha egy felismerőt zajmentes hangmintákkal tanítunk be, akkor annak a zajérzékenysége igen nagy, csak a 24dB-es és 18dB-es jel-zaj viszonyoknál mutat a beszédfelismerő minimális zajtűrő képességet, itt az értékek csak néhány százalékban változnak.

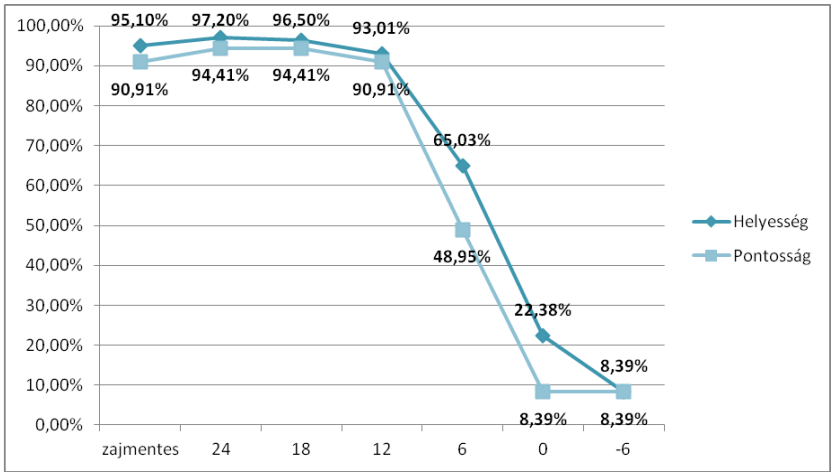


1. ábra. *A tesztelési eredmények ábrázolása*

7.2 24dB zajjal terhelt tanító minták vizsgálata

2. táblázat. *A 24dB zajjal terhelt tanítóanyagokra elvégzett tesztek eredményei*

Teszt anyagok jel/zaj viszonya (dB)	Helyesség	Pontosság
zajmentes	95,10%	90,91%
24	97,20%	94,41%
18	96,50%	94,41%
12	93,01%	90,91%
6	65,03%	48,95%
0	22,38%	8,39%
-6	8,39%	8,39%



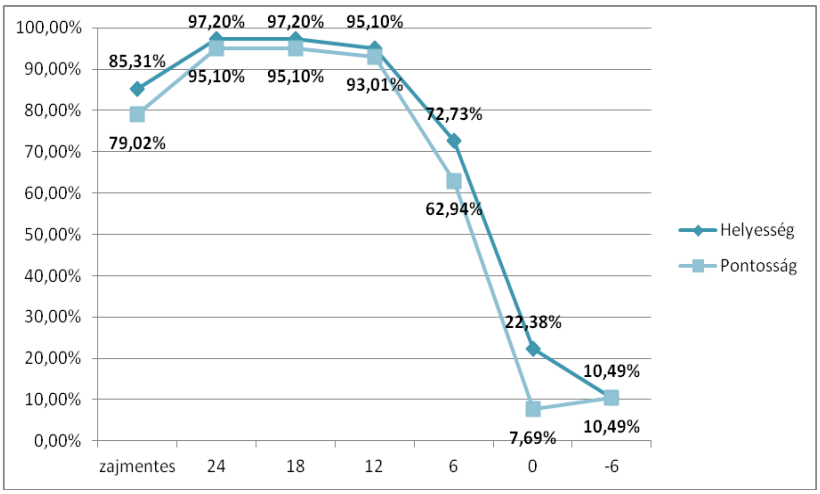
2. ábra. *A tesztelési eredmények ábrázolása*

A 2. ábrán a 24 dB zajjal terhelt tanító anyaggal betanított felismerő hatékonysági eredményeit láthatjuk a különböző teszt anyagok felismerésénél. Az értékek hasonlóan az előző esetben mért eredményekhez, a zaj növekedésével csökkennek. Azonban megállapítható az is, hogy míg a zajmentes tanítóknál a zajtűrő képesség a 24dB-es és 18dB-es értékeknél érzékelhető, ebben az esetben viszont azzal, hogy a tanítót 24dB-es zajjal terheltük a zajtűrési tartománya megnövekedett. 12dB-es zajjal terhelt tesztelő anyagok vizsgálatánál is a helyességi és pontossági értékek 90% felettiek. Az is megállapítható, hogy a 24dB zajjal terhelt tanító anyag használata a legzajosabb tesztelő anyag (-6dB) felismerésénél is növelt a hatékonyságon, egyedül a zajmentes tesztelő anyagoknál mutatkozik enyhe romlás, de összességében hatékonyabb lett a felismerésünk.

7.3 18dB zajjal terhelt tanító minták vizsgálata

3. táblázat. *A 18dB zajjal terhelt tanítóanyagokra elvégzett tesztek eredményei*

Teszt anyagok jel/zaj viszonya (dB)	Helyesség	Pontosság
zajmentes	85,31%	79,02%
24	97,20%	95,10%
18	97,20%	95,10%
12	95,10%	93,01%
6	72,73%	62,94%
0	22,38%	7,69%
-6	10,49%	10,49%



3. ábra. *A tesztelési eredmények ábrázolása*

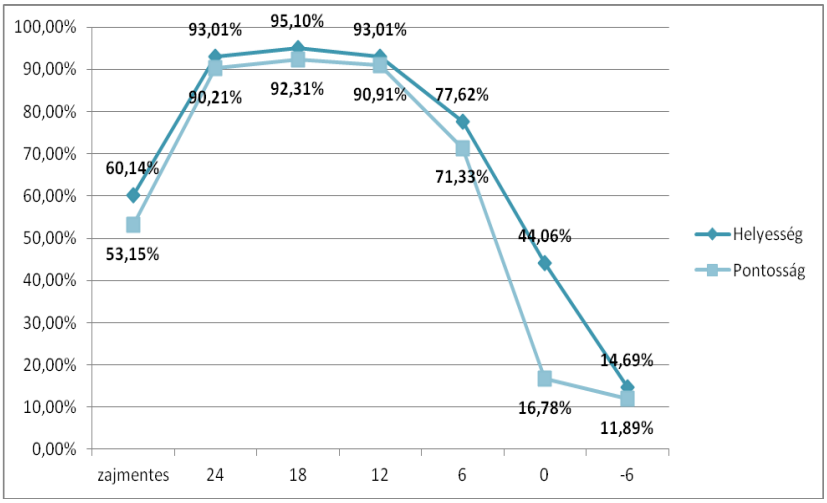
A tesztelésekhez 18dB zajjal terhelt tanító mintákat használtam fel. A 3. ábra diagramját megtekintve láthatjuk, hogy a 24dB-es és 18dB-es zajjal terhelt tesztelő anyagoknál a leghatékonyabb a felismerés (helyesség: 97,20%, pontosság: 95,10%). Ez az eredmény meglepő, mivel a helyesség értéke 0,7%-kal jobb, mint a zajmentes tanító – zajmentes tesztelő párosítás vizsgálatánál, ami eddig a legjobb elérhető eredmény volt a hatékonyság növelő vizsgálatok során. Tovább vizsgálva az ábrát, azt is

tapasztaljuk, hogy a zajtűrő képesség intervalluma a 24dB-es zajjal terhelt tanító anyag tesztelésénél látható intervallumhoz képest eltolódott a zajosabb tesztelő anyagok irányában. A 24dB-es 18dB-es, és 12 dB-es tesztelő anyagok esetén születtek 90% feletti eredmények. A zajmentes tesztelő mintánál az eredmények tovább romlottak.

7.4 12dB zajjal terhelt tanító minták vizsgálata

4. táblázat. *A 12dB zajjal terhelt tanítóanyagokra elvégzett tesztek eredményei*

Teszt anyagok jel/zaj viszonya (dB)	Helyesség	Pontosság
zajmentes	60,14%	53,15%
24	93,01%	90,21%
18	95,10%	92,31%
12	93,01%	90,91%
6	77,62%	71,33%
0	44,06%	16,78%
-6	14,69%	11,89%



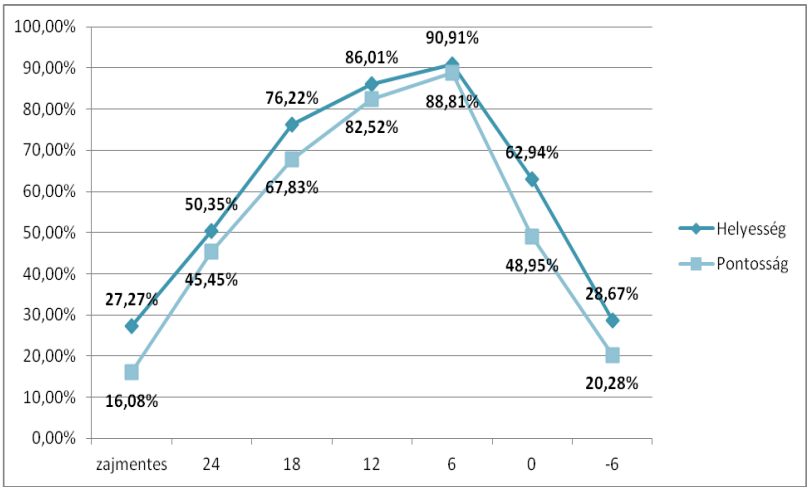
4. ábra. *A tesztelési eredmények ábrázolása*

12dB zajjal terhelt tanító minták alkalmazásánál a 4. táblázatban látható eredmények keletkeztek. Azzal, hogy még zajosabb tanító anyagot alkalmaztunk a felismeréshez, a zajmentes tesztelő minták vizsgálatánál az eredmények újból csökkentek. A zajtűrő képesség tartománya továbbra is a 24dB-es, 18dB-es és 12dB-es tesztelő minták vizsgálatánál érzékelhető (az eredmények szintén 90% felettiek). 12dB és -6dB között a helyességi és pontossági értékek itt is meredeken csökkennek, de összehasonlítva az előző két mérés eredményeivel folyamatos javulást vehetünk észre.

7.5 6dB zajjal terhelt tanító minták vizsgálata

5. táblázat. *A 6dB zajjal terhelt tanítóanyagokra elvégzett tesztek eredményei*

Tesztanyagok jel/zaj viszonya (dB)	Helyesség	Pontosság
zajmentes	27,27%	16,08%
24	50,35%	45,45%
18	76,22%	67,83%
12	86,01%	82,52%
6	90,91%	88,81%
0	62,94%	48,95%
-6	28,67%	20,28%



5. ábra. *A tesztelési eredmények ábrázolása*

Az 5. ábrán látható görbe képe, ami a 6dB zajjal terhelt tanító anyagok vizsgálatainak eredményeit szemlélteti, az előzőekben jellemzett esetektől rendkívül eltérő. 90% feletti értéket csak egy pontban találhatunk, mégpedig abban az esetben, amikor a tesztelő minták szintén 6dB zajjal terheltek, de ebben az esetben is már csak a helyességi érték 90% feletti (helyesség: 90,91%, pontosság: 88,81%). Ez a jel-zaj viszony, egy választó vonalnak is tekinthető, mivel a 6dB-nél kisebb zajjal terhelt tesztelő mintáknál az eredmények jelentősen csökkentek, míg a zajosabb mintáknál nagymértékben növekedtek, az eddigi eredményekhez viszonyítva.

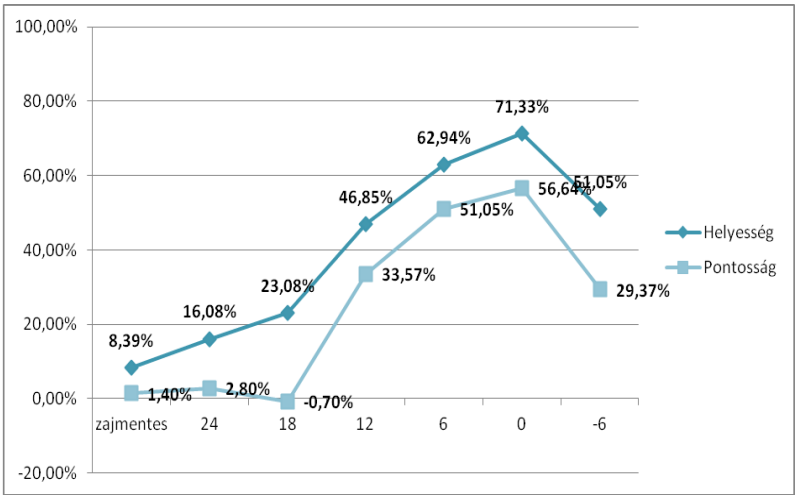
7.6 0dB zajjal terhelt tanító minták vizsgálata

A 0dB zajjal terhelt tanító anyag tesztelési eredményei alátámasztják az előző fejezetben megfogalmazott állítást, miszerint a 6dB-es jel-zaj viszony egy választó vonal. A 6. ábra értékei egy ”súlypont” átbillenést szemléltetnek. Itt már a kevésbé zajos tesztelő anyagok felismerésénél (zajmentes, 24dB, 18dB) tekinthető működés képtelennek a beszédfelismerő, ellentétben a 9.3 fejezetben tapasztaltakkal. A görbe itt is abban az esetben éri el a maximumát, amikor a tanító és tesztelő minták ugyanazzal a zajjal lettek terhelve (helyesség: 71,33%, pontosság: 56,64%). Ez alapján az is megállapítható, hogy a

beszédfelismerő, ebben az esetben elveszítette a zajtűrő képességét, és egyik esetben sem tekinthető megfelelően működőképesnek. (A 18dB-es teszt anyagnál a pontosság értéke már 0% alatti.)

6. táblázat. *A 0dB zajjal terhelt tanítóanyagokra elvégzett tesztek eredményei*

Teszt anyagok jel/zaj viszonya (dB)	Helyesség	Pontosság
zajmentes	8,39%	1,40%
24	16,08%	2,80%
18	23,08%	-0,70%
12	46,85%	33,57%
6	62,94%	51,05%
0	71,33%	56,64%
-6	51,05%	29,37%

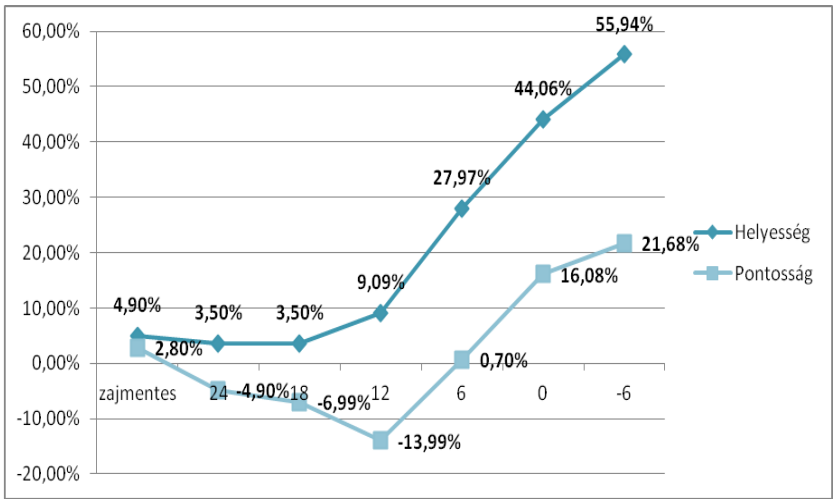


6. ábra. *A tesztelési eredmények ábrázolása*

7.7 -6dB zajjal terhelt tanító minták vizsgálata

7. táblázat. *A -6dB zajjal terhelt tanítóanyagokra elvégzett tesztek eredményei*

Teszt anyagok jel/zaj viszonya (dB)	Helyesség	Pontosság
zajmentes	4,90%	2,80%
24	3,50%	-4,90%
18	3,50%	-6,99%
12	9,09%	-13,99%
6	27,97%	0,70%
0	44,06%	16,08%
-6	55,94%	21,68%



7. ábra. *A tesztelési eredmények ábrázolása*

A -6dB zajjal terhelt tanító minták a legzajosabbak mindegyik eset figyelembe véve. Az eredmények (7. táblázat és 7. ábra) közül a legjobb hatékonyság itt is abban az esetben volt mérhető, amikor a tanító és tesztelő mintákat ugyanúgy -6dB zajjal terheltem (helyesség: 55,94%, pontosság 21,68%), de ebben az esetben is az értékek rendkívül rossznak minősülnek. Több esetben, ahol a teszt anyagok kevésbé zajjal terheltek, a pontossági értékek 0% alá csökkentek. Az előző esethez viszonyítva, azzal hogy még zajosabb tanító anyagot használtunk, a beszédfelismerő működőképessége tovább romlott. A 6dB-es zajjal terheltségig teljesen működésképtelennek tekinthető, és csak ettől a ponttól kezdenek intenzíven növekedni a helyességi és pontossági értékek.

8. Összefoglalás

Munkám során a HTK toolkit szoftvercsomag alkalmazásával betanítottam egy kis szótáras beszédfelismerőt, amin vizsgálatokat végeztem különböző jel-zaj viszonyokkal (-6dB, 0dB, 6dB, 12dB, 18dB, 24dB). A minták zajjal való terhelését a Matlab szoftver segítségével valósítottam meg. Az eredmények alapján általánosságban elmondható, hogy a helyességi és pontossági értékek abban az esetben a legjobbak – vagy legalább azon érték körül maximalizálódnak-, amikor a tanító és a tesztelő anyagok jel-zaj viszonya megegyezik, azok zajjal való terheltségük különbsége minél kisebb). A vizsgálati eredmények és a következtetések alapján összességében megállapíthatjuk, hogy ha egy felismerőt zajos környezetben akarunk alkalmazni, akkor érdemes a környezetben fellelhető zajokkal specifikusan foglalkozni vagy a zajszűrő eljárás alkalmazásának szempontjából vagy jelen esetben a hangmintákat is ugyanazon zajjal terhelt környezetben rögzíteni és azokat felhasználva betanítani a beszédfelismerőt, mert így a felismerő jobb hatékonysággal fog működni, mint zajmentes tanító és tesztelő anyagok esetén, hiszen azok nem a valós felhasználói környezetet reprezentálják.

9. Köszönetnyilvánítás

A cikkben ismertetett kutató munka az EFOP-3.6.1-16-2016-00011 jelű „Fiatalodó és Megújuló Egyetem – Innovatív Tudásváros – a Miskolci Egyetem intelligens szakosodást szolgáló intézményi fejlesztése” projekt részeként – a Széchenyi 2020 keretében – az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Irodalom

[1] Németh, G., Olaszy, G. *A magyar beszéd*, Akadémiai Kiadó, Budapest, 2010

[2] Sztahó, D., Szaszák, Gy., Vicsi, K. *Zajszűrő eljárások alkalmazása, teljesítményük vizsgálata zajos beszéd automatikus felismerésénél*, VI. Magyar Számítógépes Nyelvészeti Konferencia, Szeged, 2009., pp. 195-205.

[3] Brooks, S., et al., eds. *Handbook of Markov Chain Monte Carlo*, CRC press, 2011. **https://doi.org/10.1201/b10905**

[4] Deng, L. *A generalized hidden Markov model with state-conditioned trend functions of time for the speech signal*, Signal Processing 1992, 27(1):65-78. **https://doi.org/10.1016/0165-1684(92)90112-A**

[5] Muda, L., Begam, M., Elamvazuthi, I. *Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques*, Journal of Computing 2010, 2(3):138-143. ISSN 2151-9617

[6] Dal Degan, N., Prati, C. *Acoustic noise analysis and speech enhancement techniques for mobile radio applications*, Signal Processing 1988, 15(1):43-56. **https://doi.org/10.1016/0165-1684(88)90027-8**