

TDK-dolgozat



Készítette: Pintér Judit Mária

Miskolci Egyetem

Gépészmérnöki és Informatikai kar

Automatizálási és Kommunikáció-Technológiai Tanszék



Beszéd felismerés hatékonyságának vizsgálata különböző nyelvtanokkal

Assessment of speech recognition with different grammars

Készítette: Pintér Judit Mária

mérnök informatikus MSc 2. évf.

Témavezető: Dr. Czap László

egyetemi docens, tanszékvezető

Tartalomjegyzék

Bevezetés	5
1. Beszédfelismerés	7
1.1 Beszédfelismerési feladatok [1].....	7
1.2 A számítógépes beszédfelismerők általános felépítése és működése.....	11
1.3 Fejlődését befolyásoló tényezők.....	12
1.4 Beszédfelismerés alapegysége [3]	14
2. Rejtett Markov- modell [2].....	15
2.1 A Markov –lánc	15
2.2 A Markov- modell	15
2.3 A Rejtett Markov- modell (HMM)	15
3. A HTK toolkit [7]	17
4. Az akusztikus beszédfelismerő hatékonyságának vizsgálata	19
4.1 A betanított beszédfelismerő [2].....	19
4.1.1 A Nyelvtan fájl [2]	19
4.1.2 A nyelvtan [2]	20
4.1.3 A HMM modell [2]	21
4.2 Hosszú és rövid felpattanó zárhangok [2]	23
4.3 Különböző állapotszámú HMM modellek tesztelése [2]	24
4.4 Kettős Gauss eloszlású rejtett Markov- modellek tesztelése [2]	26
4.5 Változó HVite paraméterezés tesztelése [2]	27
4.6 Különböző nyelvtanok tesztelése	30
4.6.1 A hosszú felpattanó zárhangok átírása	31
4.6.2 Zöngés- zöngétlen részleges hasonulás (zöngétlenedés).....	32
4.6.3 Zöngétlen-zöngés részleges hasonulás (zöngésedés)	33
4.6.4 Teljes hasonulás.....	34
4.6.5 Egyedi besorolást igénylő esetek.....	36

4.6.6 Összes eset együttes alkalmazása	37
Összegzés.....	39
Irodalomjegyzék	42
Mellékletek	43
1. Átalakítások.....	43
2. Nyelvtan fájl tartalma	43
3. csere6.m matlab fájl tartalma a vizsgálatokhoz.....	46

Bevezetés

A beszéd az emberek közötti legtermészetesebb információátviteli forma. Ezen állítást szem előtt tartva az ember régi vágya, hogy az általa konstruált gépekkel, berendezésekkel emberi nyelven tudjon 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.) [1].

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

A 20. század első felében főként a távközlés motiválta a beszédvizsgálatát, a feldolgozási eljárások legtöbbször a beszédkódolásból származó módszereken alapult.

Az 50-es 60-as években rövid beszédsszeleteket próbáltak beszédhangként besorolni, vagy hosszabb egységek esetén a beszédet hangokra szegmentálták, majd pedig a szegmenseket próbálták felismerni. Nagyobb egységek (pl. szavak) esetén ígéretesebbnek bizonyult az időtengely menti lehetséges torzulások ún. dinamikus idővetemítéssel való kezelése. A *Dynamic Time Warping* feladata a lehető legjobb illeszkedés megkeresése úgy, hogy az illesztésnél a szükséges gyorsítások-lassítások helyét és mértékét kell megtalálni. Az általános dinamikus programozás egy speciális esetéről van szó. Ez a megoldás viszont csak szuboptimálisnak nevezhető.

A 70-es és még részben a 80-as éveket is az előbb említett módszer letisztulása és elterjedése jellemzi, amit kapcsoltszavas felismeréssé egészítettek ki a folyamatos beszédfelismerés előkészítésére. Folyamatos beszédfelismerő megvalósítása nehézséggel jár, főként a határok megállapítása miatt. A felismerés során megpróbálják felhasználni a magasabb szintű (lexikális, szintaktikai, szemantikai) információkat. Beszédfelismerés helyett beszédmegértés a cél, hogy az ember és a gép kommunikációjában a rendszer helyesen reagálhasson az elhangzott utasításokra.

A 80-as évektől egy új módszer, a statisztikai alapú felismerés hódít teret, aminél minden felismerendő egységhez tartozik egy valószínűségi modell. Lenny Baum volt az, aki feltalálta a rejtett Markov-modell alkalmazhatóságát a felismerők területén. A **Hidden Markov-Model** máig használatos módszer az akusztikus beszédfelismerésben, melynek bemenetét a jellegvektorok vagy lényegvektorok (*Feature vector*) képezik. Nehézségeket jelent, ha a bemenet bizonytalan, vagy zajjal terhelt, ha sok a hasonló szótárelem, továbbá a mindig problémát jelentő kiejtésbeli eltérések. A magyar, mint toldalékoló nyelv igen sok problémát vet fel egy megfelelő beszédfelismerő megalkotása esetén, az agglutináló (ragozó) jelleg toldalékok halmozását is lehetővé teszi [2].

Munkám során különböző nyelvtani eseteket alkalmazok egy Rejtett Markov-modelleket alkalmazó HTK toolkit segítségével betanított folyamatos beszélőfüggő beszédfelismerő esetén. Ugyanazon tesztelő anyag felhasználásával kapott hatékonyságra vonatkozó adatokat, elemzem és összehasonlítom az egyes eseteket. Nem csak a nyelvtanra vonatkozó, hanem más különböző módszereket is alkalmazva próbálom növelni a hatékonyságot. Mindezen módszerek, illetve a beszédfelismerő működésének megértéséhez ismertetem a megfelelő elméleti hátteret.

Célom, hogy a vizsgálatok segítségével elérjem a lehető legjobb hatékonyságot, illetve általános szabályokat, összefüggéseket fogalmazzak meg a beszédfelismerő beállításaira vonatkozóan.

1. Beszédfelismerés

A beszédfelismerés az a folyamat, melynek során a beszédfelismerő gép azonosítja a kiejtett beszédjeleket és átalakítja ezeket szöveggé, vagy más, számítógép által feldolgozható adattá. Beszédjelek alatt természetesen érthetünk akusztikus vagy akár vizuális jeleket is (gesztikulációk, arcminika, szájmozgás). A vizuális jelek feltérképezésével, megértésével, és rekonstruálásával az audiovizuális beszédfelismerés foglalkozik. Az általam betanított beszédfelismerő viszont az akusztikus jeleket fogja figyelembe venni, azaz magát a beszédhangot, amit szöveggé fog átalakítani. A beszéd nem más, mint akusztikus hullámok keltése, azaz beszédhangok, fonémák (hangok olyan elemi, elvont egysége, amely szavakat különböztet meg egymástól, önálló jelentéssel nem rendelkezik) kibocsátása. A beszéd nem csupán fonémák sorozata, hanem fontos a hangsúlyozás a hanglejtés és számos más szupraszegmentális jellemző is. Ezek alapján egyértelmű, hogy a beszéd az emberek legfőbb kommunikációs eszköze, amiért az akusztikus beszédfelismerést igen sok területen és különböző céloknak megfelelően alkalmazzák [2].

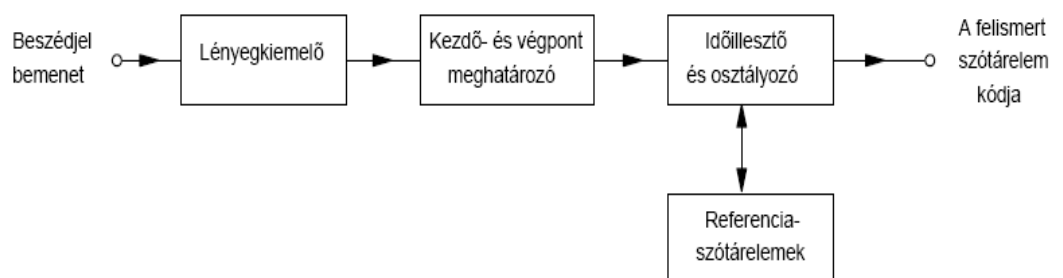
A beszédfelismerés meglehetősen tág témakör. Szűkebb értelemben a tartalom felismerését értjük alatta, tágabb értelemben azonban alkalmazások egész sora használ egészében vagy komponenseként beszédfelismerőt.

1.1 Beszédfelismerési feladatok [1]

- Beszéd-szöveg átalakítás (speech recognition)
- Beszélőfelismerés
 - Beszélő személy azonosítása (speaker verification)
 - Felhasználó felismerése, kiválasztása egy előre definiált halmazból (*speaker identification*)
- Beszéddetekció
- Kulcsszókeresés
- Érzelmi töltet felismerése

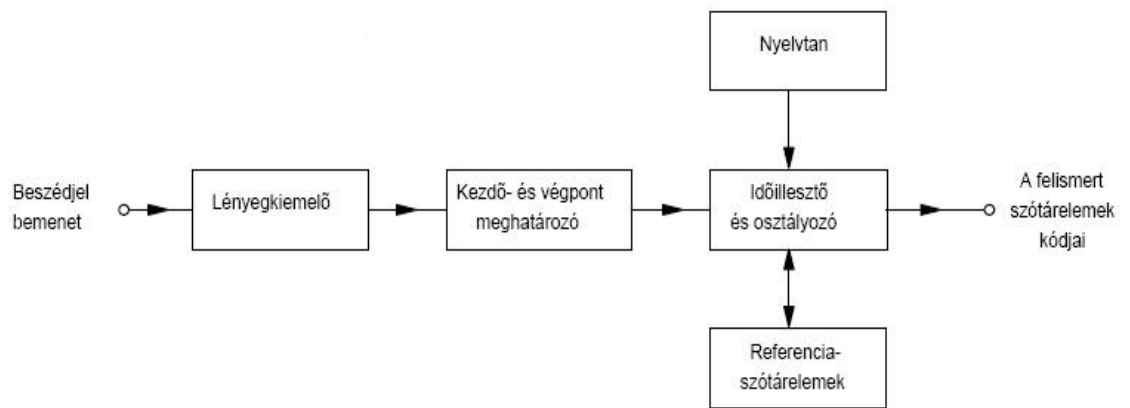
A különböző szempontokat figyelembe véve az alábbi osztályokba sorolhatók a beszédfelismerő szoftverek:

- **A szótár nagysága szerint:**
 - kis (kötött, zárt) szótáras: néhány száz szó
 - középszótáras: néhány ezer szó
 - nagy szótáras (kötetlen, nyílt szótáras), több tízezer szó
- **A megvalósítás módja szerint:** milyen módon kezeli a bemeneti hangadatot, illetve, hogy milyen algoritmusokat használ
 - Mintafelismerő: napjainkban főleg ezt használják, ilyen pl. rejtett Markov-Modell
 - Akusztikus-fonetikus: pl. neurális hálós és az ismeretalapú (tudásalapú) módszer
- **Beszédmód szerinti osztályozás:**
 - Izolált szavak, kijelentések felismerése: egyszerre csak egy kijelentést vár, a beszélőnek a szavak között szünetet kell tartania. Egy szótárelemet határoz meg a szótárból. (1.ábra)



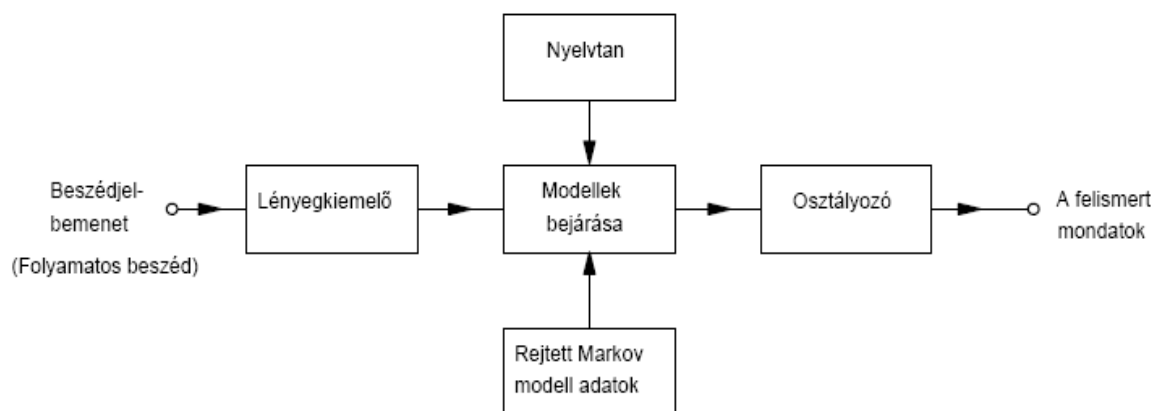
1. ábra – Izolált szavak, kijelentések felismerése

- Kapcsolt szavak felismerése hasonlít az előzőhöz, de a rendszer már képes bizonyos szókapcsolatokat is felismerni, viszont nem ismeri a szavak határát, ezért minden kombinációt figyelembe kell vennie. (2.ábra)



2. ábra – Kapcsolt szavak felismerése

- Folyamatos beszéd felismerés esetén sok nehézséggel jár a rendszer megalkotása, főként a határok megállapítása miatt, viszont jó megoldásnak ígérkezik a valószínűség számítási algoritmusok (pl. HMM) használata. A nagy szótáras diktáló rendszernek előkövetelménye a folyamatos beszéd felismerés. A 3. ábra is azt szemlélteti, hogy a végeredmény nem szavak felismerése, hanem összefüggő mondatoké. Ehhez szükséges szintén a lényegkiemelés, ami ebből is láthatóan minden felismerési típus esetén a kezdeti pont, majd a létrejövő modellek körbejárása, ebben az esetben adatokkal feltöltött rejtett Markov-modellekre kell gondolnunk, amihez szintén szükséges egy nyelvtani alap, a helyes szókapcsolatok kialakításához. Majd az osztályozást követően megkapjuk az eredményt. (3. ábra)



3. ábra - Folyamatos beszéd felismerés

- **Beszélőre adaptáltság szempontjából:**
 - Beszélőfüggő és adaptív rendszerek (*Speaker Dependence*) esetén a rendszer lehet akár egy adott beszélőre is tervezve, de több felhasználó hangjának az elfogadására is.
 - Beszélőfüggetlen felismerésről akkor beszélünk, amikor a felismerőt használat előtt igen nagyszámú, általában több mint ezer bemondóval előre betanítanak az adott szókészlet, vagy folyamatos szöveg felismerésére. Olyan esetekben használatosak ezek a rendszerek, amikor minden felhasználót azonosan kell kezelni, vagy a betanítás túl hosszadalmas lenne.
- **Akusztikai környezet szempontjából való osztályozás:**
 - Csendes környezet: jó jel-zaj viszony (> 30 dB)
 - Zajos környezet: speciális algoritmusok szükségesek a zajszűréshez
 - Telefonbeszéd: a sávkorlátozott jelleg és az egyre gyakrabban alkalmazott tömörítő kódolás miatt megkülönböztetett és fontos felhasználási terület.
- **Üzem mód szempontjából való osztályozás:**
 - parancsmód
 - diktáló üzemmód
 - másik csoportosítási mód is megadható üzemmód szempontjából, mégpedig az online, illetve az offline működés

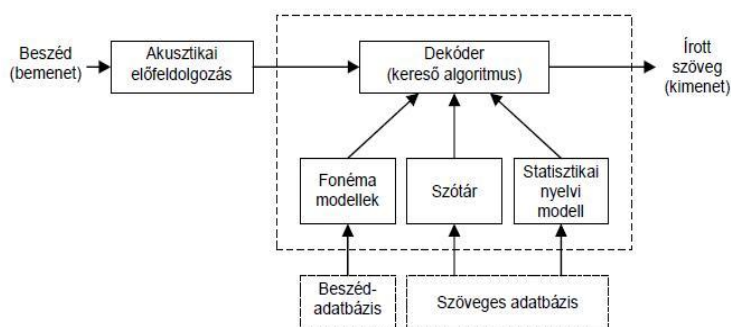
Bár a felsorolás nem teljes körű, jól szemlélteti, a fentebb említett feladatok hétköznapi előfordulásait.

- Diktálás (napjainkban egyik leggyakrabban használt terület)
- Hallássérültek oktatása, beszédsérültek kommunikációjának megkönnyítése
- Telekommunikáció, mobiltelefonok (automata telefonos rendszer, hangalapú tárcsázás)
- Tartalomkeresés audió és videó fájlokban (pl.: *Track ID* zenefelismerő program, híranyag adatbázisokban való kulcsszó alapú keresés)
- Riasztó berendezések vezérlése (pl.: beszédhang, mint egyedi azonosító)
- Épületautomatizálás (pl.: intelligens ház szoftverek)
- Multimédiás eszközök elindítása, vezérlése
- Mesterséges intelligencia

1.2 A számítógépes beszédfelismerők általános felépítése és működése

Egyre inkább világossá vált napjainkban, hogy a puszta beszéd - karaktersorozat átalakítás nem oldható meg a szövegfeldolgozás szintaktikai - szemantikai összefüggéseinek vizsgálata nélkül. E két szint összekapcsolásával működnek a mai felismerők.

A számítógépes beszédfelismerők nem túl részletes működési vázlatát láthatjuk a 4.ábrán, amely fonémák modelljére épülő statisztikai alapú beszédfelismerő rendszert mutat be. A bemenetre kapott beszédjelet előzetesen egy akusztikai feldolgozásnak vetjük alá, amelyet szokás lényegkiemelésnek is nevezni. A cél az, hogy olyan, az egyes beszédhangokat jól elkülönítő jellemzőket tartsunk meg, amelyek a beszédet a lehető legtömörebben reprezentálják lényegesebb információvesztés nélkül. -A tömörítés szükségessége az eredendően hatalmas számítási igény kordában tartása végett is felmerül, a legtöbb alkalmazáshoz kívánatos ugyanis, hogy a beszédfelismerő valós időben is működőképes legyen (online üzemmód). Ezután következhet a dekódolás folyamata. A művelet során a felismerés alapegységeinek modelljeit, a szótárát és a nyelv szintaktikai viszonyait statisztikai alapokon leíró nyelvi modellt használhatjuk fel. A Markov-modellek, a szótár és a nyelvi modell együttesen egyfajta tudást visz a rendszerbe, melyet számítógépes betanítás (training) során alapozhatunk meg. Ehhez beszéd, illetve szöveges adatbázis egyaránt szükséges. A beszédatadatbázisnak a nyelvben előforduló valamennyi építőelemet és kapcsolatot is többszörösen tartalmaznia kell, hogy statisztikailag megfelelő lefedettséget adjon. Hasonlóképpen, a nyelvi modellnek is meg kell felelnie a beszédfelismerő használati területére jellemző szóhasználati szokásoknak [2].



4. ábra - Statisztikai számítógépes beszédfelismerő működése

1.3 Fejlődését befolyásoló tényezők

A beszédfelismerést igen sok tényező befolyásolja. Beszédünk igen változatos, mivel nemcsak az általunk kiejteni próbált mondandót kódolja, hanem számtalan más információt is hordoz magában.

Nagy kihívást jelent a fonetikai változatosság. A magyar nyelv rendszere 65 fonémát használ. Ezen fonémák hangzó formáit tekintjük beszédhangoknak. Beszédkeltés során nem elkülönült fonémák fűződnek egymás után, hanem folytonosan változik hangképző szerveink alakja, amiből következően a szomszédos hangoktól függően a beszédhang megváltozhat (igazodás (koartikuláció), részeleges hasonulás, teljes hasonulás stb.). A beszéd tehát nagyobb változatosságot mutat, mint a fonémák [1]. Az akusztikai körülményektől függően a beszédjelbe bekerülhet a környezetből beszűrődő zaj, vagy a jel torzulhat a mikrofon vagy az átviteli csatorna paramétereitől függően. A beszélők közti változatosság is elég nagy jelentőséggel bír. Különböző beszélők hangmagassága, szájüreg mérete, beszédsebessége, dialektusa stb. meglehetősen különbözhet.

Egy adott beszélő esetén fennálló változatosságot is figyelembe kell venni. Ha változatlan a beszélő, akkor is meglehetősen eltéréseket mutat a beszéd, hisz a beszélő fizikai- (fáradtság, megbetegedés) és lelkiállapota is belekódolódik a beszédsebességébe, a hang minőségébe, hanglejtésbe. Bármennyire akarjuk, nem vagyunk képesek kétszer ugyanúgy elmondani egy üzenetet.

Kicsit más szempontból tekintve a különböző nyelvek is, eltérő esélyeket kínálnak a felismerésükre. A magyar, mint toldalékoló nyelv igen sok problémát vet fel egy beszédfelismerő megalkotása esetén. Az agglutináló (ragozó) jelleg szavanként akár több száz féle toldalék (képző, jel rag) hozzáadását is lehetővé teszi, így megsokszorozza a szóalakokat, míg az angol főleg izoláló nyelv, vagyis a mondatokban a fő nyelvtani funkciókat a szavak sorrendje határozza meg. Alig használ ragozott szavakat, emiatt a szóalakok száma nagyságrenddel kisebb. Ebből adódóan egy angol nyelvre kifejlesztett szóalapú felismerő nem adaptálható a magyar nyelvre.

Felmerül a kérdés, hogy magyar nyelv esetén, - ami ahogy fentebb is megjegyeztük agglutináló jellegű - mi lehet a felismerés alapegysége (ami általánosabban minden nyelvre alkalmazható). Ha nagyobb egységeket választunk, akkor könnyebb megkülönböztetni

egymástól a szótárelemeket, viszont (a már említett toldalékolások miatt is) több tanító elemmel kell számolnunk. Ha alapegységként a jóval kisebb elemet a fonémát választjuk, akkor azok száma jóval kevesebb lesz, viszont a különböző koartikulációs és egyéb hatások miatt azonosíthatatlan lesz. Valahol a kettő között kell megtalálnunk a számunkra legkisebb kompromisszummal járó megoldást. Vicsi Klára (1995) vizsgálta, hogy egy hosszú szöveg részleges lefedéséhez hány kettőshang, félszótag illetve szótag szükséges (5.ábra) az alábbi követelmények szerint:

- fedje le a nyelv egészét, vagy legalább annak túlnyomó részét
- számossága ne legyen túl nagy
- az elemek realizációja ne legyen túl erősen kontextusfüggő

Kumulatív gyakoriság	hangpár	félszótag		szótag
		kezdő	záró	
50%	71	31	12	116
75%	183	75	41	402
90%	399	134	95	1011
95%	586	176	150	1604
99%	956	259	317	3380

5. ábra - Hosszú szöveg részleges lefedéséhez szükséges elemszám (Vicsi, 1996)

A szótag és a táblázatban nem szereplő hanghármas alkalmazása a nagyszámú betanítandó elem miatt nehézkes. A félszótag és a hangpár versengésében a szöveg részleges lefedéséhez a félszótagból kell kevesebb, a teljes lefedéshez a hangpárból. A félszótag mellett szól, hogy általában hosszabb az időtartama, ellene, hogy illeszteni csak a kezdő és záró félszótag határán tudjuk, az így képzett szótagok határán a hangok egymásra hatását nem tudja figyelembe venni. A hangpárok mindkét végükön illeszkednek a szomszédos hangokhoz, viszont rövidebbek. Felmerül a kérdés, hogy a félszótagok és hangpárok felismerését az előbbi tulajdonságok mennyire segítik, vagy gátolják, melyik hatás érvényesül erősebben [3].

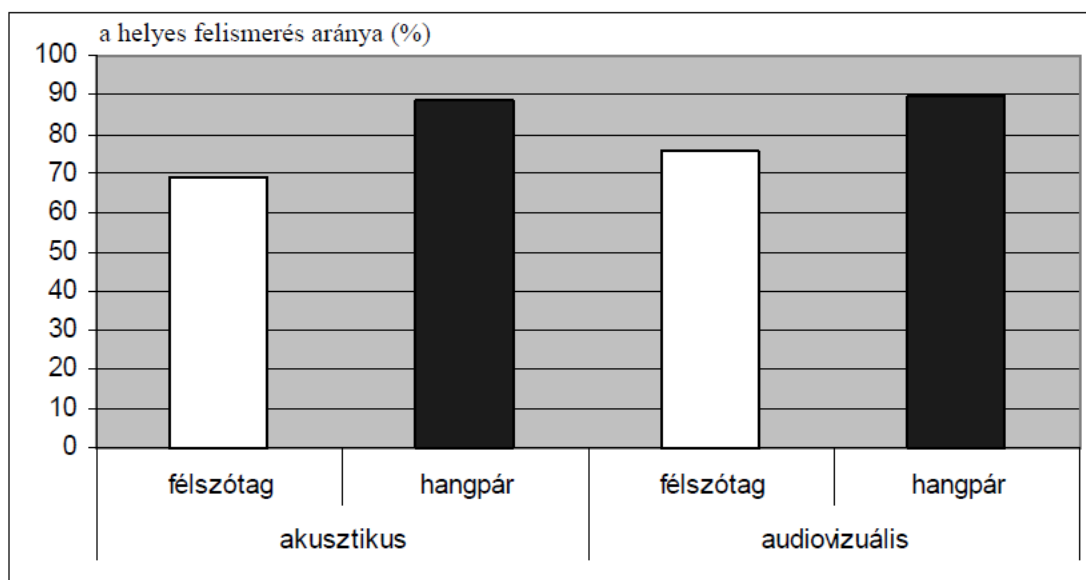
1.4 Beszédfelismerés alapegysége [3]

Elvégezve a vizsgálatot egy 8000 szóból álló tanító és 1400 szóból álló tesztelő akusztikus adatbázisra, az alábbi táblázatban feltüntetett eredmények születtek.

Az 1 996 589 szóból álló klasszikus és modern próza alapján - amely 4 238 066 szótagot tartalmazott – készült statisztikai elemzés szerint az adatbázis szavaiban szereplő 121 kezdő félszótag kumulatív gyakorisága 70,2 %. A 83 záró félszótag kumulatív gyakorisága 80,7 %, a kiválasztott félszótagok 60,1%-át, a szavaknak a 32,6%- át fedték le.

Ugyanezen az adatbázison történt a hangpár alapú felismerés is. Az akusztikus adatbázis 440 hangpárt tartalmaz. Ez a teljes hangpár készletnek mintegy fele.

A felismerő betanítása és tesztelése a HTK toolkit szoftvercsomag alkalmazásával valósult meg, amelyről a későbbiekben esik részletesebben szó.



6. ábra - A félszótag és a hangpár alapú felismerés eredményei

A félszótag és a hangpár alapú felismerési eredmények (6.ábra) alapján megállapítható, hogy a félszótagokra válogatott adatbázison is a hangpár alapú felismerési eredmények kedvezőbbek. Az eredmények azt mutatják, hogy a felismerés alapegysége a tekintetében többet nyom a latba a kontextusfüggés figyelembe vétele, mint a beszédszakasz időtartama.

2. Rejtett Markov- modell [2]

2.1 A Markov –lánc

A Markov-lánc megalkotása a 20. század elejére tehető, amely Andrej Markov orosz matematikus nevéhez fűződik. Az első eredményeket 1906-ban a folyamatok tekintetében kizárólag elméleti szinten fektette le. A Markov-lánc 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 [5].

2.2 A Markov- modell

A Markov-modellek is a fentebb említett tulajdonságokon alapszanak. 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.

2.3 A Rejtett Markov- modell (HMM)

A "*rejtett Markov-modell*" [6] 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).

Elsősorban az izolált szavas, kötött szótáras beszédfelismerésben lehet alkalmazni, amit már a beszédfelismerők osztályozásánál is megállapítottunk. 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 nagyszótáros, folyamatos beszéd felismerésére, viszont ebben az esetben nem használható maga a szó felismerési alapegységként, hanem célszerűbb kisebb egységekből építkezni, azaz fonémákból. Ezek összekapcsolásából kaphatjuk meg a szavak modelljeit, majd végül ezeket körbekapcsolva kaphatunk egyetlen nagy modellt.

Egy beszédfelismerési feladat rejtett Markov- modellekkel matematikailag az alábbiak szerint fogalmazható meg:

$$Szó_{felismert} = \operatorname{argmax}_{minden\ szóra} \{P(szó|X)\},$$

Vagyis azt a szót (vagy más beszédelemet) keressük, amelyre az X adott akusztikai megfigyelés-sorozat valószínűsége a legnagyobb. Számunkra azonban az X megfigyeléssorozat ismert, ezért *Bayes tétele* alapján átalakítva a fenti összefüggést az alábbiak szerint írhatjuk:

$$\text{Bayes- tétel: } P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

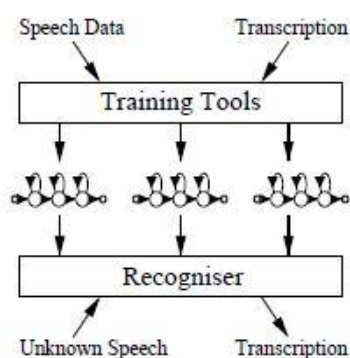
$$Szó_{felismert} = \operatorname{argmax}_{minden\ szóra} \{P(X|szó) P(szó)\}$$

Ebben az alakban a $P(X)$ tagot a nevezőből elhagytuk. A $P(X|szó)$ valószínűséget az akusztikai, a $P(szó)$ valószínűséget pedig a nyelvi modell adja meg. Az akusztikai modellnek tehát arról kell informálnia, hogy adott akusztikai megfigyelés az egyes szavakra milyen valószínűségű, a nyelvi modellnek pedig arról, hogy az egyes szavak előfordulásának mekkora a becsült valószínűsége.

3. A HTK toolkit [7]

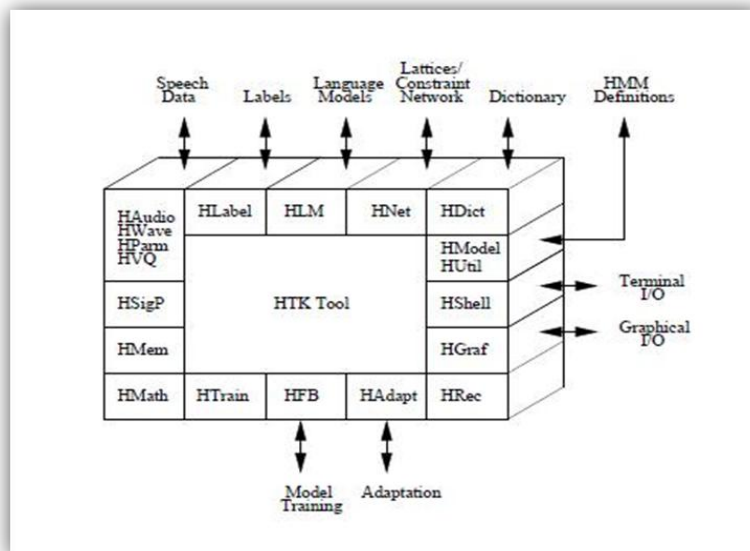
A *HTK* egy eszközugyűjteménynek is tekinthető, amely rejtett Markov-modellek építésére szolgál. A HMM alkalmas bármilyen sorozat modellezésére, és maga a *HTK* is alapjában véve hasonlóan általános célú. Mindazonáltal a *HTK* elsősorban mégis Rejtett Markov-modell alapú beszédfelismerők fejlesztésére szolgál, ami megfelelő nagyságú infrastruktúra háttérrel biztosít ezen feladat számára.

Az 7.ábrán is látszik, hogy két nagy feldolgozó szakaszból áll. Először a *HTK* a betanító eszközeit használja a paraméterek becslésére számos Rejtett Markov-modell, betanító kifejezések és a hozzájuk társított leírások segítségével. Második szakaszban pedig az ismeretlen beszédjelet írja át a *HTK* a felismerő eszközök használatával.



7. ábra - Beszédfelismerés általános szemléltetése HMM-el

A szoftver számos funkciója könyvtári modulként van beépítve a szoftver architektúrába. Ezek a modulok biztosítják, hogy minden egyes eszköz interfészei azonos módon csatlakozzanak a külvilághoz. Ezért az általánosan használt funkciók számára biztosít egy központi készletet is.



8. ábra - Szoftver architektúra

A 8.ábra egy tipikus *HTK* eszköz felépítését, és annak input/output interfészeit szemlélteti. A felhasználói input/output és az operációs rendszer kölcsönhatását a *HShell* könyvtári modulok ellenőrzik, a memóriamenedzselés felügyelete pedig a *Hmem* feladata. A fejlesztéshez szükséges matematikai támogatását a *HMath* látja el, a beszéd analízishez szükséges jelfeldolgozó műveleteket pedig a *HSigP* könyvtári modul tartalmazza.

Egyes fájl típusok esetén a *HTK* külön dedikált interfészeket ír elő. Például a *Label* fájlokhoz a *HLabel* modult rendeli, nyelvi modellek tartalmazó fájlokhoz a *HLM*-t, *HNet* a hálózathoz és nyelvtani hálókhoz előírt, *HDict* a szótárakhoz, *HModel* pedig HMM definíciókhoz tartozó dedikált interfész.

Minden kimeneti és bemeneti hullámforma a *HWave* interfészen keresztül, a paraméterezett változata pedig a *HParm* interfészen keresztül kapcsolódik. Amellett, hogy egységés felületet biztosít a fejlesztők számára, a *HWave* és a *HLabel* több fájlformátumot is képes támogatni, amely lehetővé teszi különböző adatok importálását más rendszerekből.

A közvetlen audió inputot a *HAudio*, az egyszerű interaktív grafikát pedig a *HGraph* modul támogatja. A *HUtil* számos segéd rutint biztosít a HMM-k manipulálására, míg a *HTrain* és a *HFB* különféle támogatást nyújtanak a *HTK* betanító eszközökhöz. *HAdapt* a *HTK* adaptációs eszközökhöz nyújt különféle segítséget. Végül *HRec* tartalmazza a legfontosabb a felismerési folyamathoz szükséges funkciókat.

4. Az akusztikus beszédfelismerő hatékonyságának vizsgálata

4.1 A betanított beszédfelismerő [2]

A fentebb már bemutatott HTK szoftvercsomag alkalmazásával egy 747 hangfájllal, azaz 7470 szóval betanított folyamatos beszélőfüggő (a hanganyagok egy személytől származnak) beszédfelismerőnek a hatékonyságát vizsgáltam egy 225 hangfájlból, azaz 2250 szóból álló tesztanyaggal. Akusztikus lényegkiemelésre az MFCC (Mel-Frequency Cepstral Coefficients) jellemzőket használtam. A felismerő alapegységének – a kísérleti eredményeknek megfelelően - a diádot választottam. A diád olyan fonémakapcsolat, ami két hangból tevődik össze, és az első hang felétől a második hang feléig tart. Így mindkét végükön illeszthetők a szomszédos hangpárokhoz, a kontextusfüggés jól meghatározható. Megjegyzendő, hogy a beszélő az adatbázis létrehozásakor törekedett a gondos artikulációra kerülve a modoros kiejtést.

4.1.1 A Nyelvtan fájl [2]

A betanításhoz elengedhetetlen nyelvtan fájl ezen felismerő esetén az előforduló diádokat tartalmazza. A HTK egy nyelvtani definíciókat létrehozó “nyelvet” nyújt számunkra, aminek a segítségével egyszerű, vagy akár összetett nyelvtani szabályokat is alkothatunk a feladatunknak megfelelően. Ezen szabályok egyfajta reguláris kifejezéseket adnak meg, amelyek állhatnak fonémakapcsolatok sorozatából és metakarakterekből. A definíciók megadhatók egyszerű txt fájlokban is. Ebben az esetben a nyelvtan fájlban nincs semmilyen nyelvtani megkötés arra vonatkozóan, hogy egyes diád összetételek csak egy adott másikat követhetnek, illetve azok csak bizonyos esetekben állhatnak elő. Egyszerűen megfogalmazva a felismerőhöz alkalmazott nyelvtanfájl, csak annyit fogalmaz meg, hogy bármelyik diádot, csak olyan diád előzhet meg, aminek a második hangja megegyezik az őt követő első hangjával, illetve olyan követheti, aminek első hangja megegyezik az előtte levő második hangjával, valamint definiálja a szünetet, amit a HTK *sil* néven képes értelmezni. Tehát a nyelvtan nem tartalmaz szótárt, nem definiál szavakat, morfológiai szabályokat. A diádok sorozatából nem csak értelmes szavak adódhatnak. A szótár és a szabályrendszer figyelembe vétele egy magasabb szintű nyelvtani elemző feladata lehetne.

4.1.2 A nyelvtan [2]

Lényeges kiemelnünk vizsgálati szempontból magát a nyelvtant is, ami szintén elengedhetetlen a betanításhoz. Az elemtár a tanítóanyagban előforduló összes diádot tartalmazza. Beszéd közben a kiejtett hangok nem mindig felelnek meg a helyesírás szabályai szerint lejegyzett betűkhöz tartozó hangoknak. A szomszédos hangok egymásra hatása az írásképnek megfelelő fonéma egy másik fonémával helyettesítését eredményezi. Az adatbázisban szereplő szavak fonotipikus átírása előzte meg a tanítást.

Magánhangzók esetén, mint "a á" és "e é" különbséget kell tennünk mivel más a képzésük ellentétben a többi magánhangzóval. Ezért az á és é betűket A-ra és E-re cseréltem. A többi magánhangzók esetén pedig mivel a hang képzése hasonló csak a kiejtés hossza változik (a magánhangzókat körülvevő hangoktól függően a rövid és a hosszú magánhangzók esetén egyaránt változatos időtartamok fordulhatnak elő, ezért nem lehet azokat megfelelően megkülönböztetni), amit a beszédfelismerő nem tudna helyesen értelmezni, ezért azokat nem vesszük figyelembe, hanem rövid magánhangzóknak tekintjük. Továbbá az "ö ü" betűket O-nak és U-nak feleltetjük meg. Mássalhangzók esetén, mint "sz zs gy ty cs ny" a kezdő betűk nagy változatát használtam jelölésnek. Ilyen esetekben a kettősséget a betűk duplázódásával jelöltem. Az átírás célja, hogy minden fonémát egy karakter jelöljön. Az alkalmazott egykarakteres átírási táblázatot a Melléklet tartalmazza.

Felpattanó zárhangok duplázódása esetén, mint *bb, pp, dd, tt, ggy, tty, gg, kk* két lehetőség közül kell választanunk. Vagy változatlanul hagyjuk ezen elemeket, és megkülönböztetjük a rövid és hosszú párokat, vagy eltekintünk a rövid és hosszú zárhangok megkülönböztetésétől. Hogy melyik eset lesz a hatékonyabb a tesztelés során derül ki.

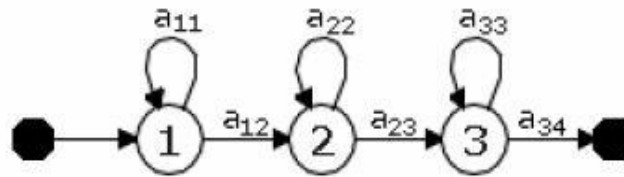
Fontos továbbá figyelembe vennünk néhány nyelvtani szabályt, ami a mássalhangzók kapcsolódására vonatkozik. Ha szavakat mondunk, vagy összefüggően beszélünk, akkor a hangokat általában úgy ejtjük ki, hogy két egymásután következő hang artikulációja, képzésmódja közeledik egymáshoz. A helyes kiejtést ez a tény még nem érinti. Csak azok a hangkapcsolatok lehetnek kérdésesek, amelyekben két szomszéd hang valamely fontosabb képzési mozzanatban hat egymásra. Különbséget kell tennünk egyfelől

magánhangzók és mássalhangzók találkozása közt, másfelől a tekintetben, hogy a kapcsolat egyetlen szóelemben (morfémában), tö és toldalék határán, összetett szó taghatárán, vagy szókapcsolatban szóhatáron fordul-e elő. Szóhatáron különbség lehet a kiejtésben aszerint, hogy egyetlen hangsúlyozási szakaszba, (ún. szólamba) tartozik-e a két szó, vagy a második szó kissé nagyobb hangsúlyt kap, továbbá, hogy milyen sebesen és milyen stílusban beszélünk. Ha két szó közt szünetet tartunk akkora szünet előtti és a szünet utáni hang már nem hat egymásra [8]. Sajnos nincs lehetőségünk mind ezen szempontokat megvizsgálni, ezért csak olyan igazodásokat, részleges hasonulásokat, teljes hasonulásokat, összeolvadásokat veszek figyelembe, amelyek függetlenek attól, hogy hol helyezkednek el a kifejezésen belül. A betűk cseréjét a *matlab* segítségével oldottam meg.

4.1.3 A HMM modell [2]

A szótárban szereplő minden egyes diádhoz létre kell hozni a hozzá tartozó rejtett Markov-modellt. Beszédhangok esetén általában háromállapotú lineáris struktúrájú modellt (ún. balról – jobbra) szokás választani (9.ábra). Magát a modellezést a diádok esetén három vagy több állapot végzi, valójában azonban két további szélső állapotot is találunk, amelyek az egyes beszédelem-modellek összefűzését biztosítják.

Felismeréskor a rendszer számára minden keret érkezésekor két lehetőség áll fent, vagy állapotot változtat, vagy helyben marad, bizonyos valószínűséggel. Ezeket nevezzük állapotátmeneti valószínűségeknek, melyek becslése a tanítás során történik. Ez a mechanizmus biztosítja az időbeli illesztést a modell és az aktuális keret között. A rendszer az adott (belső) állapotból két keret érkezése között egy megfigyelést bocsát ki, mely tulajdonképpen egy hasonlósági mérték az adott állapotra jellemző jellemzővektor-eloszlás és az aktuálisan érkezett, a külső megfigyelést reprezentáló jellemzővektor között. Lényegében azt mondhatjuk, hogy e hasonlósági mérték a mérőszáma a megfigyelt jellemzővektor és a modellállapot spektrális illeszkedésének. Egy állapotra jellemző jellemzővektor-eloszlást általában sűrűségfüggvényével adunk meg, amelyről feltételezzük, hogy normális (*Gauss*) eloszlások lineáris kombinációjából áll elő. Ezt szokás kibocsátási valószínűségnek is nevezni [7].



9. ábra - 3 állapotú lineáris modell

Az ezek alapján betanított beszédfelismerő az alábbi eredményeket szolgáltatta, amit tekinthetünk a további vizsgálatok kiindulási pontjának is. Az eredmények arra az esetre vonatkozik, amikor megkülönböztetjük a rövid és hosszú párokat.

```

===== HTK Results Analysis =====
Date: Wed May  5 20:42:50 2010
Ref : efull.mlf
Rec : recout3.mlf
----- Overall Results -----
SENT: %Correct=0.00 [H=0, S=225, N=225]
WORD: %Corr=82.66, Acc=41.66 [H=21418, D=82, S=4410, I=10624, N=25910]
=====

```

10. ábra - A hosszú és rövid felpattanó zárhangok megkülönböztetése esetén a felismerés eredménye

Az eredmények első sora a nagyobb felismerési egység (SENT) eredményeit mutatja. A 225 (N) „mondatból” (10 egymás után ejtett független szó jelentett egy egységet) egy sem akadt, amelynek mind a tíz szavát hiba nélkül felismerte a rendszer (10.ábra). A második sorban a diád (WORD) felismerés eredményei jelennek meg. Széttagoló nyelveknél (pl. angol) a felismerés alapegysége a szó, ezek sorozata a mondat. A táblázatban szereplő kifejezések ennek felelnek meg. Esetünkben az agglutináló jelleg miatt a felismerés alapegysége a szónál kisebb diád, a mondat szerepét az egymás után, szünetekkel kiejtett független szavak vették át. A „mondat” így a szavak számával összehasonlítva sokkal több építőelemet tartalmaz. H mértéke jelenti a helyesen felismert alakzatok számát, D a törlések számát, S azokat az eseteket, amikor az adott egységet tévesen ismerte fel, I a beszúrások számát. A két alábbi képlet mutatja meg, hogyan is számolja a HTK a helyességet és a pontosságot.

$$\%Correct = \frac{H}{N} \times 100\%$$

$$\%Accuracy = \frac{H - I}{N} \times 100\%$$

4.2 Hosszú és rövid felpattanó zárhangok [2]

Felmerül a kérdés, hogy érdemes-e megkülönböztetni a hosszú és rövid felpattanó zárhangokat. Ezek a zárfelpattanás előtti szünet (zöngétlenek), illetve fojtott zöngé (zöngések) időtartamában térnek el, ilyenformán az azonosan képzett rövid és hosszú magánhangzók mintájára is lehetne kezelni őket. A 10. ábra alapján megállapítható, hogy abban az esetben, amikor a hosszú és rövid felpattanó zárhangokat megkülönböztetjük, és öt állapotú HMM modellt alkalmazunk a tanítás utáni tesztelés 82, 66% -os eredményt mutat a szavak felismerési szintjén.

```
===== HTK Results Analysis =====  
Date: Wed May 5 20:19:46 2010  
Ref : efull3.mlf  
Rec : recout3.mlf  
----- Overall Results -----  
SENT: %Correct=0.00 [H=0, S=225, N=225]  
WORD: %Corr=82.68, Acc=44.99 [H=20936, D=28, S=4359, I=9543, N=25323]  
=====
```

11. ábra - A felismerés eredménye a hosszú és rövid felpattanó zárhangok megkülönböztetése nélkül

A hosszú és rövid felpattanó zárhangok megkülönböztetése nélkül a fonotipikus átírás után a *bb*, *dd*, *gg*, *GG*, *kk*, *pp*, *tt* és *TT* diádok nem fordulnak elő, a diádok összes számából (N) megállapítható, hogy 587 olyan eset volt a tesztelő anyagban, ahol ezek előfordultak, ennyivel csökkent a diádok száma. A helyesen felismert szavak száma 482-vel csökkent. Ez magában nem nagymértékű eltérés, mint ahogy a törlések közti különbség sem, viszont a beszúrások száma 1081-el csökkent, ami már jelentős javulásnak tekinthető, ennek köszönhetően a pontosság is több mint 3 százalékkal javult.

A törlések számának csökkenésénél szerepet játszanak olyan esetek is, mikor a helyesírás megköveteli, hogy kettős betűket használjunk, viszont beszédben nem ejtjük ki hosszan őket. Példaként tekinthető a *jobbra* szó, ahol a *b* felpattanó zárhangot röviden ejtjük. Mindezen tényezőket figyelembe véve, a tanítást és tesztelést, valamint a hatékonyság növelést azokkal a tanító nyelvtanokkal fogom folytatni, amelyek nem különböztetik meg a hosszú és rövid felpattanó zárhangokat. (Sajnos visszaállítási folyamatoknál ezekben az esetben nem lesz egyértelmű a megfeleltetés, a szótár és a nyelvi ellenőrzés dönti el, hogy hosszú vagy rövid a zárhang.)

4.3 Különböző állapotszámú HMM modellek tesztelése [2]

A felismerés hatékonysága nagy mértékben függ a választott HMM modell állapotainak számától. Diádok esetén kézenfekvő volt öt állapotú modell használata, viszont elképzelhető, hogy a felismerés javulni fog, ha ötnél több állapotú modellt használunk [7].

A lentebbi eredmények (1.táblázat) az öt, hét, nyolc, kilenc és tizenkét állapotú HMM modellek alkalmazásával végrehajtott tanítás eredményét szemléltetik.

1. táblázat: Tesztelési eredmények különböző állapotszámú hmm modellek esetén (szavak szintjén)

Állapotszám	Helyesség	Pontosság	H	D	S	I	N
5	82,68%	44,99%	20936	28	4359	9543	25323
7	88,37%	75,60%	22379	104	2840	3234	25323
8	88,66%	80,84%	22451	172	2700	1980	25323
9	86,92%	81,41%	22011	339	2973	1395	25323
12	67,74%	65,13%	17154	2097	6072	662	25323

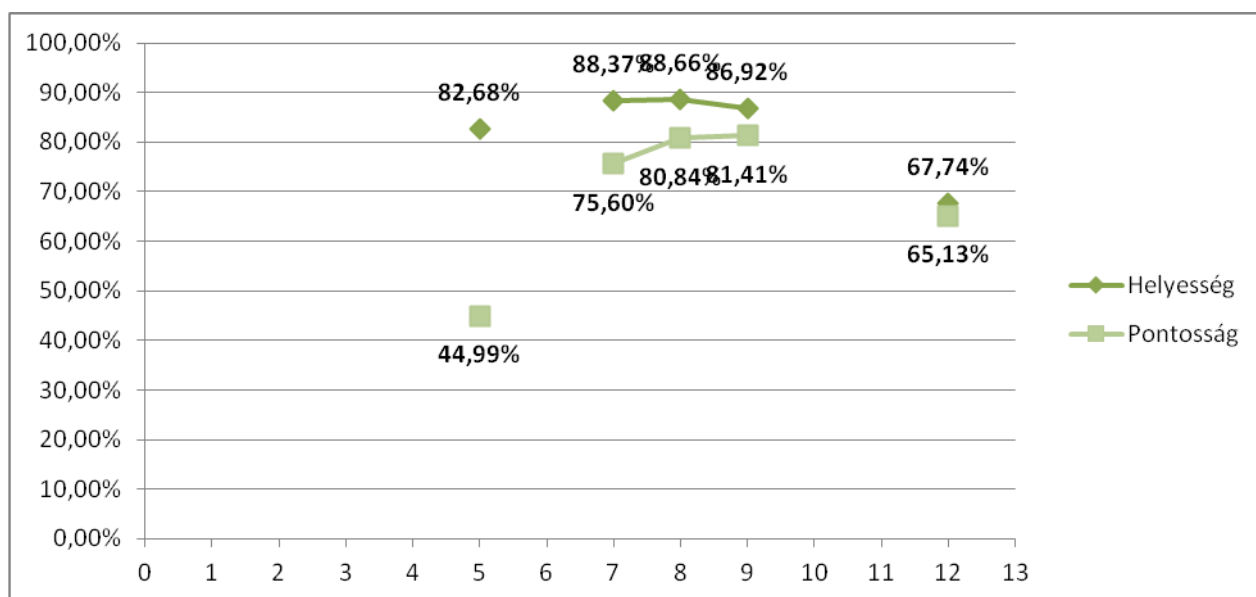
Az öt és a hét állapotú HMM modellek közti pontosság különbség nagymértékű eltérést mutat, hozzávetőlegesen 30%-os növekedést, ami a beszúrások nagymértékű (6309) csökkenésének, és a rosszul felismert diád elemek 1519 mintával való csökkenésének köszönhető. Ezáltal a helyesen felismert szavak száma is nőtt 1443 darabbal, megnövelve a helyesség mértékét is.

Nyolc állapotú modell esetén észlelhető egy kisebb mértékű javulás, de a beszúrások száma 1254-el tovább csökkent a hét állapotú HMM-hez képest, ami még mindig nem elhanyagolható értéknek tekinthető.

Kilenc állapotú HMM tesztelésének eredményét összevetve az eddig legjobb esetnek tekinthető 8 állapotú modellel, már némi romlást mutat. Bár a beszúrások száma továbbra is csökken, a törlések száma majdnem a duplájára növekedett, és a helytelenül felismert diád elemek száma is 273-al növekedett. Ebből következően a helyes

felismerések száma is romlik. Ezek alapján megállapítható, hogy továbbra is a nyolc állapotú modell tekinthető a leghatékonyabbnak.

Kicsivel kiugróbb nagyságrendű állapotszámmal rendelkező modellt is teszteltem, hogy valóban egy tendencia mutatkozik e meg az eredményekben. A kilenc állapotszámú modellhez képest a törlések és a tévesztések száma jelentősen megnövekedett, ennek hatására a százalékos eredmények is nagymértékben romlottak. Az eredmények alátámasztják, azt a következtetést, hogy a felismerés hatékonysága, ha az állapotok számát folyamatosan növeljük nyolcig, akkor szintén növekszik, viszont ha az állapotokat tovább növeljük, folyamatos hatékonyságcsökkenést tapasztalunk. Az alábbi diagram, amit a 12.ábra szemléltet, ezt a tendenciát támasztja alá.



12. ábra - Értékek változása a HMM állapotszámok megválasztása függvényében

A további teszteléseket nyolc állapotú rejtett Markov- modellek felhasználásával végzem el.

4.4 Kettős Gauss eloszlású rejtett Markov- modellek tesztelése [2]

Létrehozhatók olyan rejtett Markov- modellek is, amelyekben minden egyes állapot több Gauss eloszlásból tevődik össze. Mivel a lényegvektorok egy diád elemet képviselnek, melyek értékét befolyásolja az őket körülvevő diádok képzése, ezért azok állapota több féle lehet, hiába egy adott elemet képviselnek. Az által, hogy kettős Gauss eloszlást használunk állapotonként, ezek az egyes állapotok elkülönülhetnek egymástól a tanítási folyamat során, így hatékonyabban lehet az állapotokat egymásnak megfeleltetni. Az elkülönülés mértékét az egyes összetevők súlyainak beállításával lehet változtatni. Tesztelése e súlyok különböző megadásán alapszik [15].

A súlyok összegének mindig eggyel egyenlőnek kell lennie. Megjegyzendő továbbá, hogy a súlyoknak nem kell minden egyes állapot esetén, ugyanolyan arányban lennie. A súlyok beállítása egyes modellek esetén a 2.táblázatban található.

2. táblázat: A súlyok értékei egyes hmm modellek esetén

proto836mix1.hmm	minden egyes állapot esetén: <Mixture> 1 0.4 <Mixture> 2 0.6
proto836mix2.hmm	minden egyes állapot esetén: <Mixture> 1 0.1 <Mixture> 2 0.9
proto836mix3.hmm	2 és 7 állapotok esetén: <Mixture> 1 0.3 <Mixture> 2 0.7 3-6 állapotok esetén: <Mixture> 1 0.4 <Mixture> 2 0.6

proto836mix4.hmm	2 és 7 állapotok esetén: <Mixture> 1 0.1 <Mixture> 2 0.9 3-6 állapotok esetén: <Mixture> 1 0.3 <Mixture> 2 0.7
------------------	--

3. táblázat: Tesztelési eredmények különböző súlyozások esetén (szavak szintjén)

hmm fájl	Helyesség	Pontosság	H	D	S	I	N
proto836mix1.hmm	91,23%	85,78%	23102	174	2047	1381	25323
proto836mix2.hmm	91,58%	86,31%	23190	172	1961	1333	25323
proto836mix3.hmm	91,39%	86,04%	23142	174	2007	1353	25323
proto836mix4.hmm	91,50%	86,32%	23170	177	1976	1311	25323

Az eredmények (3.táblázat) jelentős javulást mutatnak az egy Gauss eloszlást tartalmazó Markov-modellhez képest. Egymással összehasonlítva azonban nem mutatnak ennyire jelentős eltérést. A betanítás utáni Markov-modelleket részletesen megvizsgálva megállapítható, hogy az egyes Gauss eloszlások megfelelő súlyozását a rendszer megtanulja. A kiindulási állapotok a tanulás sebességét határozzák meg, ez az oka a minimális eltérésnek.

4.5 Változó HVite paraméterezés tesztelése [2]

A tanítást végrehajtva és a tesztelési eredményeket kiértékelve a leghatékonyabb módszernek a nyolc állapotú kettős Gauss eloszlású rejtett Markov-modellek alkalmazása bizonyul, az előzőekben megválasztott súlyok alkalmazásával. A p paraméter alapértelmezett értéke a HVite használatakor 0. Az érték változtatásával hatást gyakorolhatunk a felismerés folyamatára. Különböző mértékű büntetést szabhatunk ki a

szavak (estünkben diádok) beillesztéséért. Ez azt jelenti, hogy befolyásolni tudjuk a beszúrások és törlések egyensúlyát, egyiket a másik rovására csökkenthetjük. A kiindulási eredményeket tekintve, azaz amikor a $p=0$, esetemben a törlések száma 172, a beszúrásoké pedig 1333 (3.táblázat – proto836mix2.hmm). Az értékekből egyértelműen arra tudok következtetni, hogy a p paramétert csökkenteni kell. Ezeket a kísérleteket a következő fejezetben részletezett optimális nyelvtani szabályok alkalmazásával végeztem el. Elsőként p értékét -2-re csökkentettem (4.táblázat).

4. táblázat: Tesztelési eredmények különböző p paraméterek esetén (szavak szintjén)

p értékei	Helyesség	Pontosság	H	D	S	I	N
-2	91,60%	87,13%	23195	181	1947	1132	25323
-4	91,56%	87,60%	23187	195	1941	1005	25323
-6	91,54%	88,09%	23180	219	1924	873	25323
-10	91,56%	88,65%	23186	241	1896	737	25323
-15	91,47%	89,07%	23164	265	1894	608	25323
-20	91,34%	89,39%	23130	312	1881	493	25323
-25	91,18%	89,45%	23089	358	1876	438	25323
-30	91%	89,54%	23045	411	1867	372	25323
-35	90,71%	89,44%	22970	477	1876	322	25323
-40	90,43%	89,29%	22900	539	1884	288	25323
-50	89,63%	88,69%	22696	694	1933	237	25323

Az eredményjavulást mutat, főként a beszúrások számát tekintve, ami 201 esettel csökkent, míg a törlések száma csak 9 esettel növekedett. Figyelembe véve ezeket az eredményeket, újabb tesztelést hajtottam végre, ahol a p paramétert -4 értékűnek választottam (4.táblázat).

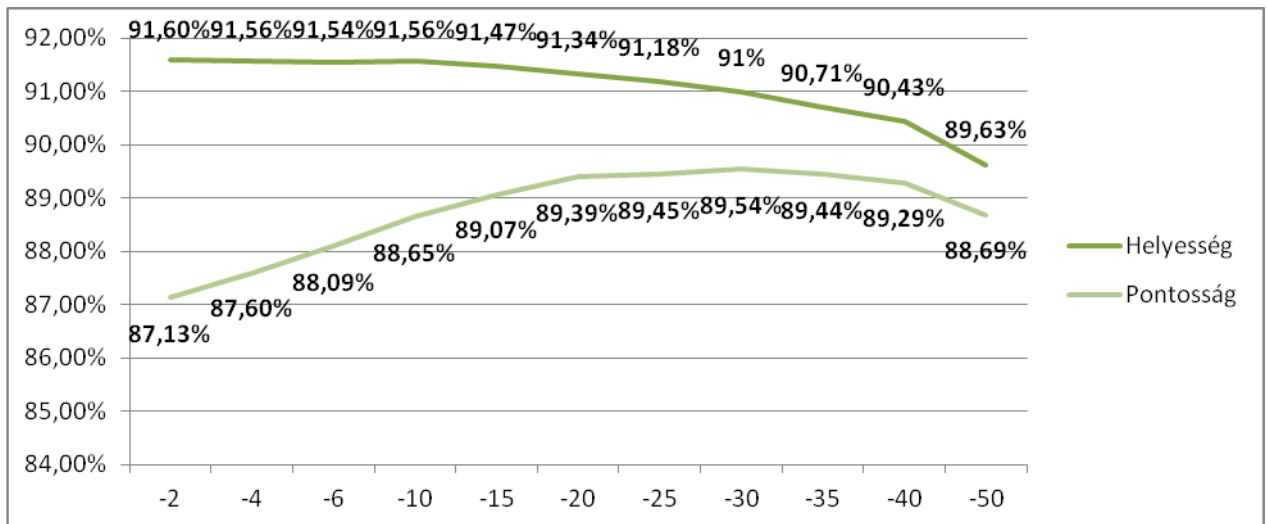
A további csökkentés eredménye azt mutatja, hogy a helyesség elhanyagolhatóan kicsi mértékben romlott, a pontosságé pedig 0,5%- al javult. A beszúrások száma 127 esettel csökkent a törléseké pedig 14-el nőtt. A kettő közti különbségből arra következtetek, hogy a felismerés tovább javítható a p érték csökkentésével, amit -6-nak választottam meg.

A 4.táblázatban látható, hogy a helyességi és pontosság ugyancsak kis mértékben változtak, de ha a beszúrák és törlések számát figyeljük meg, akkor abból szintén az következik, hogy tovább lehet csökkenteni a p paramétert.

A következő mérésnél viszont egy jóval alacsonyabb értéket adtam meg, hogy megvizsgálhassam a tendencia irányát, amit az eddigi adatokból lehetett kikövetkeztetni. A -30- nál (4.táblázat) elért értékeket összevetve a $p = -6$ esetén kapott eredményekkel, már szembetűnőbb különbségeket mutatnak. A helyesség megközelítőleg 0,5 százalékkal csökkent, de a pontosság 1,5 százalékkal növekedett. A beszúrák száma 501-gyel csökkent, a törléseké 192-vel nőtt. Ez az elmozdulás még mindig hatékonyság növekedésnek tekinthető. Ha összevetjük $p = 0$ esetén kapott eredményekkel (3.táblázat - proto836mix2.hmm), akkor még szembetűnőbb az a különbség, ami bár a helyesség százalékát 0,58 százalékkal lerontotta, viszont a pontosságot 3,23 százalékkal javította.

A következő mérés eredménye viszont egyértelműen hatékonyság visszaesésnek minősül a $p = -30$ -nál kapott eredményekhez képest, mivel 1,37%-al csökkent a helyesség, és a pontosság is 0,85 százalékkal. A törlések száma is 283-al növekedett, míg a beszúrák száma csupán 135-el csökkent. Egyértelműen megállapítható, hogy $p = -50$ esetén a hatékonyság csökkent. Mivel a két utolsó p érték között igen nagy a különbség, köztes értéket választva elvégeztem két mérést.

A 4.táblázatban az értékek $p = -40$ esetén még mindig azt mutatják, hogy a hatékonyság ebben az esetben is csökkent, ezért $p = -35$ értéknél további tesztelést végeztem. Továbbra is csökkenés figyelhető meg, tehát a javulási tendencia a -30-as értékig figyelhető meg, a további tesztelések csökkenő hatékonyságú folyamatot támasztanak alá, ha a p paraméter értékét -30-nál kisebbnek választjuk. (13.ábra)



13. ábra - Értékek változó p paraméter esetén

A *Hvite* modul p paraméterének megválasztása a diagram vizuális kiértékelése alapján -20 és -30 között tűnik optimálisnak. A helyesség folyamatosan csökken, a pontosság -30-ig nő. A helyesség és pontosság közeledése azt jelenti, hogy a beszúrák száma csökken, ezért döntöttem a $p=-30$ mellett. Ezek alapján 91 százalékos helyességű és 89,54 százalékos pontosságú akusztikus beszédfelismerőt sikerült létrehoznom.

4.6 Különböző nyelvtanok tesztelése

A továbbiakban a különböző nyelvtani eseteket fogom megvizsgálni, amelyek:

- hosszú és rövid felpattanó zárhangok megkülönböztetése
- részleges hasonulásra vonatkoznak, a zöngés mássalhangzó helyett a megfelelő zöngétlent ejtjük a rá következő zöngétlen hatására
- részleges hasonulásra vonatkoznak, zöngétlen mássalhangzó helyett a megfelelő zöngés ejtjük a rá következő zöngés hatására
- teljes hasonulásokra vonatkoznak
- egyedi besorolást igénylő esetekre vonatkoznak
- az összes nyelvtani esetet figyelembe vettem.

A különböző nyelvtani esetek tesztelésénél a többi beállítás, az eddigi ismereteknek megfelelően történt, azaz:

- 8 állapotú HMM modellt alkalmaztam
- kettős Gauss eloszlásúak a HMM modellek, a súlyok minden egyes állapot esetén 0,1 és 0,9 eloszlásúak
- *HVite* modul p paraméterének értéke: -30.

Ezeket a javításokat átvezetve a hosszú felpattanó zárhangok átírása utáni állapotra, a beszédfelismerés kiindulási eredményeihez jutunk.

4.6.1 A hosszú felpattanó zárhangok átírása

Bár nem a hasonulások témakörét érinti, logikailag ide is tartozik a 4.1.2 fejezetben a fonémák kódolásánál leírt hosszú és rövid felpattanó zárhangok kezelése. A felpattanó zárhangok módosítását a 4.2 fejezet indokolja, így az első vizsgálat eredménye az azóta elvégzett javítások (Markov-modell állapotainak megválasztása, a Gauss eloszlások számának növelése és a beszúrások büntetése) végrehajtása után kapott helyesség és pontosság leginkább viszonyítási alapként fog szolgálni a többi eset elemzéséhez.

Az így kapott eredmény:

	Corr	Acc	H	S	N	D	I
SENT	0.00	-	0	225	225	-	-
WORD	85,82	84,71	22052	2901	25696	743	286

Tehát az alapként szolgáló helyesség 85,82%-os, a pontosság pedig 84,71%-os 25696 diád esetén.

A 4.6.2 – től a 4.6.5-ig terjedő fejezetekben csakis mássalhangzó – mássalhangzó koartikulációs hatások szerepelnek. Ezekben az esetekben a vizsgálatokkal megállapítható, hogy a kapcsolódásban résztvevő hangok valóban hatnak-e egymásra, és történnek-e változások azok akusztikai szerkezetében. A mássalhangzó - kapcsolódásoknál az egyik mássalhangzó képzési helyéből, módjából és gerjesztéséből kell a másik kapcsolódó

mássalhangzó képzésére jellemző artikulációs helyzetbe vezérelni a beszéd szerveket. Ez bizonyos esetekben a mássalhangzó- kapcsolatokra jellemző akusztikai átalakulásokkal is jár. A fejezetek a magyar nyelvtani szabályok által is megfogalmazott változások vizsgálatát írják le, és hogy milyen eredmények születtek az egyes esetekben [1].

4.6.2 Zöngés- zöngétlen részleges hasonulás (zöngétlenedés)

A részleges hasonulás két szempont alapján csoportosítható:

- zöngésség szerint
- képzés helye szerint

A 4.6.2 fejezetben és a 4.6.3 fejezetben az elemzések a zöngésség szerinti csoportosításukat vizsgálja. A zöngésség szerinti hasonulás azt jelenti, hogy ha egy zöngés és egy zöngétlen mássalhangzó egymás mellé kerül, akkor a kiejtésünkben az egyik megváltozik, és az utána következő mássalhangzóhoz fog hasonulni. A fejezet pedig azt az esetet boncolgatja, amikor egy zöngés mássalhangzó helyett a megfelelő zöngétlent ejtjük a rá következő zöngétlen hatására (zöngés-zöngétlen párok: *b-p, d-t, g-k, v-f, z-sz, zs-s, dz-c, dzs-cs, gy-ty*, a többi esetben páratlanok) [10].

A zöngétlenedés vizsgálatkor az alábbi eseteket vettem figyelembe:

bc - pc	bC - pC	bh - ph	bk - pk	bs - ps	bS - pS	bt - pt	bT - pT	dc - tc	dC - tC
df - tf	dh - th	dk - tk	dp - tp	gc - kc	gC - kC	gf - kf	gh - kh	gp - kp	gs - ks
gS - kS	gt - kt	gT - kT	Gf - Tf	Gh - Th	Gk - Tk	gG - kG	Gp - Tp	SdZ - zd	SZ - zZ
vc - fc	vC - fC	vh - fh	vk - fk	vp - fp	vs - fs	vS - fS	vt - ft	vT - fT	zc - Sc
zC - SC	zf - Sf	zh - Sh	zk - Sk	zG - SG	zp - Sp	zt - St	zT - ST	Zc - sc	ZC - sC
Zf - sf	Zh - sh	Zk - sk	Zp - sp	Zt - st	ZT - sT	bf - pf	Gc - Tc	GC - TC	Gt - Tt

Az így kapott eredmény:

	Corr	Acc	H	S	N	D	I
SENT	0.00	-	0	225	225	-	-
WORD	87,30	86,14	22437	2548	25700	715	300

A helyesség és a pontosság közelítőleg 1,5 százalékkal javult. A törlések és a beszúrások számai elhanyagolhatóan kis mértékben változtak, ami azt mutatja, hogy a hangkapcsolatok egyértelműen egy másik hangkapcsolatnak lettek megfeleltetve a módosítás után is. Mivel a helyesen felismert diádok száma 385 esettel nőtt, a nyelvtani változtatások hatékonyságnövelést eredményeztek.

4.6.3 Zöngétlen-zöngés részleges hasonulás (zöngésedés)

Ahogy a fejezet címe is elárulja az előző vizsgálat fordított hatását elemzi, hogy milyen hatékonyság változás figyelhető meg, ha figyelembe vesszük, hogy zöngétlen mássalhangzó helyett a megfelelő zöngés ejtjük a rákövetkező zöngés hatására.

Az első esetet kiegészítve az alábbi módosításokat hajtottam végre:

cb - dZb cd - dZd cg - dzg cG - dzG Cb - dZb Cd - dZd CG - dZG cZ - dzZ

Cz - dZz fb - vb fd - vd fdZ - vdZ fg - vg fG - vG fz - vz fZ - vZ

kb - gb kd - gd kdZ - gdZ kG - GG kz - gz kZ - gZ pd - bd pdZ - pdZ

pg - bg pG - bG pz - bz pZ - bZ sb - Zb sd - Zd sdZ - ZdZ sg - Zg

sG - ZG Sb - zb Sd - zd Sg - zg SG - zG Sz - zz tb - db tg - dg

tG - GG tz - dz tZ - dZ Tb - Gb SdZ - zd zC - SC SZ - zZ dN - GN

tN - TN nb - mb np - mp nG - NG nT - NT

Az így kapott eredmény:

	Corr	Acc	H	S	N	D	I
SENT	0.00	-	0	225	225	-	-
WORD	86,71	85,54	22292	2684	25708	732	288

A fejezetben fentebb összefoglalt változtatások 0,89%-os helyesség és 0,88%-os pontosság javulást eredményeztek. A törlések és beszúrások száma hasonlóan az előző fejezet értékeihez minimális eltérést mutatnak, melynek oka szintén az egyértelmű megfeleltetés. 217-el nőtt a helyesen felismert diádok száma. Összességében csaknem ugyanolyan arányban mutat hatékonyságnövekedést ez a vizsgálat, mint az előző fejezetben taglalt.

4.6.4 Teljes hasonulás

A teljes hasonulás a mássalhangzó-hasonulás másik fajtája. Azt jelenti, hogy ha két mássalhangzó találkozik egymással, akkor kiejtésünkben gyakran azonossá válnak. A teljes hasonulásnak két fajtája van. Az egyik fajtáját írásban is jelöljük, a másik fajtáját viszont nem. Tehát a nevük az írásban jelölt és az írásban jelöletlen teljes hasonulás.

A negyedik vizsgálati szempont, amivel kiegészítettem az első esetet a teljes hasonulás volt.

A módosítások:

kg - gg pb - bb sZ - ZZ Sz - zz td - dd bp - pp dt - tt gk - kk GT - TT vf - ff

zS - SS Zs - ss lj - jj nj - NN dG - GG dj - GG tT - TT Tj - TT Nj - NN Gj - GG

tj - TT ddZ - dZZ dS - cc ds - C tc - cc tS - cc ts - C tC - CC tdZ - ddZ

Az így kapott eredmény:

	Corr	Acc	H	S	N	D	I
SENT	0.00	-	0	225	225	-	-
WORD	87,22	86,06	22318	2620	25589	651	296

A helyesség 1,4% -os a pontosság viszont 1,89% -os javulást mutat. Ez arányaiban eltér az előző kettő vizsgálat eredményeitől. Az eltérés a hibás törlések és beszúrások számában mutatkozik meg, ahol a törlések száma 92-vel csökkent, míg a beszúrások száma csak 10-el. Teljes hasonulás esetén a mássalhangzók kiejtésükben (bizonyos esetekben írásukban is) azonossá válnak, azaz megduplázódnak. Ebben az esetben a HMM modellek közül azok, amelyek adott mássalhangzót egyszeresen tartalmazzák nagyobb valószínűséggel fognak illeszkedni az adott szóba, mint ami valóban helyes lenne. Ebből adódóan téves törlések léphetnek fel. Miután a teljes hasonulások a tanító és a tesztelő anyagban is megmutatkoznak a változtatások elvégeztével, a törlések száma nagymértékű javulást mutat.

4.6.5 Egyedi besorolást igénylő esetek

A magyar nyelv tartalmaz olyan hangkapcsolatokat is, amik egyedi besorolást igényelnek, viszont figyelembe vettem zöngességük szerint a második és a harmadik vizsgálatban is. Ezért ezeket külön is megvizsgáltam.

A *ch* hangkapcsolat esetén általában *h* hangot ejtünk: *ar-chívum*, *ma-chináció*, *or-chidea*, *pszi-chológia*, *Ri-chárd*.

A *h* réshangnak többféle variánsa is kialakul, mind a gerjesztést, mind pedig a képzési helyet illetően. Ha második hangként ejtjük, és egy *c* hang előzi meg azt (vizsgált eset), akkor zöngés változata jelenik meg. Ez a kettős viselkedés megnehezíti a felismerését, viszont a kettős *Gauss eloszlás* alkalmazásával korrigálható is (bizonyos mértékig), ha felismerési alapegységnek magát a hangot választjuk.

Figyelembe kell venni, hogy helyesírásunk bizonyos meghatározott esetekben a hagyományt követi a mai hangjelölési rendszer, a mai kiejtés, illetőleg a szóelemzés rovására.

A hagyomány érvényesülése ennél természetesen jóval szélesebb körű helyesírásunkban. A hagyomány alakította ki és őrzi betűink alakját, külön- és egybeírási szokásainkat, sőt a kiejtés és a szóelemzés elvének érvényesülését is. Hagyományos írásmódról, azaz a hagyomány elvének érvényesüléséről – a szónak szűkebb értelmében – mégis csak akkor beszélünk, ha a máig megőrzött írásmód sem a kiejtés, sem a szóelemzés elvével nem magyarázható, ilyenek a régies családnevek, pl.: *Zichy* [e.: zicsi], *Madách*, *Moholy* [e.: moholi], [e.: madács], *Czuczor* [e.: cucor], *Hajnóczy* [e.: hajnóci]. A példák a *cz* hangkapcsolatot is szemléltetik, amely esetben *cz* helyett némileg módosult *c* hangot ejtünk [16].

Az *m* hang szerkezete nem változik lényegesen hangpár kapcsolatokban, folyamatos zöngés hangként van jelen. Egyetlen kapcsolatnál fordul elő, hogy a képzési helye megváltozik, bilabiálisból labiodentálissá válik, ha labiodentális mássalhangzó követi (*hamvas*, *kámfor*). Ennek a változásnak az eredménye az *m* hang egy variánsa, aminek az akusztikai szerkezete nem teljesen egyértelmű, mivel nem tér el lényegesen a bilabiális *m* hangétól. Ezért módosítása ezekben az esetekben nem feltétlen indokolt [1].

A figyelembe vett hangkapcsolatoknál a képzés helye szerinti részleges hasonulás figyelhető meg, ha a hangokat nem azonos helyen képezzük.

Az átalakítások:

ch - h cz - c mf - nf mv - nv nb - mb np - mp dN - GN tN - TN nG - NG nT - NT

Az így kapott eredmények:

	Corr	Acc	H	S	N	D	I
SENT	0.00	-	0	225	225	-	-
WORD	86,28	85,10	22151	2804	25672	717	304

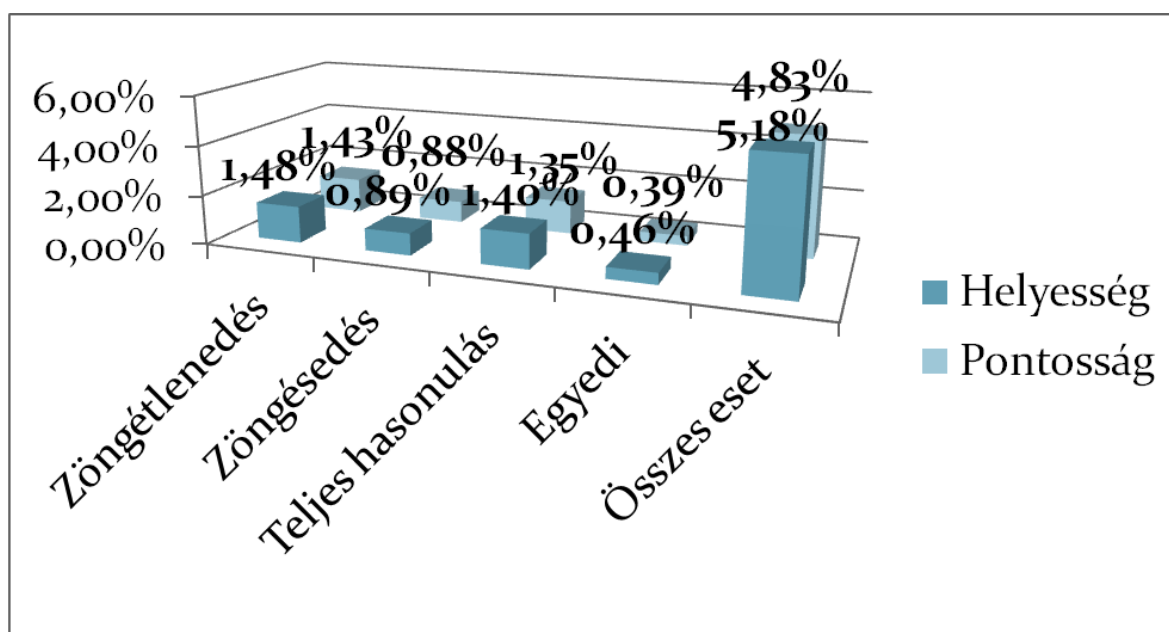
Összevetve az alapként szolgáló értékekkel <0,5 százalékos javulás figyelhető meg mind a helyesség mind a pontosság eredményénél. Egyedül a beszúrák számánál mutatkozik romlás, ami a *ch* és *cz* hangpárok egy hanggal való nem egyértelmű megfeleltetéséből adódhatnak.

4.6.6 Összes eset együttes alkalmazása

A nyelvtani eseteket megelőző, a HMM modellekre és a beállítási paraméterekre vonatkozó teszteléseket az összes nyelvtani eset együttes alkalmazásával végeztem el, így az eredmény:

	Corr	Acc	H	S	N	D	I
SENT	0.00	-	0	225	225	-	-
WORD	91,00	89,54	23045	1867	25323	411	372

A 14.ábra foglalja össze, hogy a különböző nyelvtani esetekben az átalakítások a viszonyítási alapként szolgáló hosszú felpattanó zárhangok átírása után kapott helyességi és pontossági értékekhez képest milyen mértékű javulást eredményeztek.



14. ábra - Helyességi és pontossági javulások

Összegzés

A HTK szoftvercsomag alkalmazásával egy 747 hangfájllal, azaz 7470 szóval betanított folyamatos beszélőfüggő beszédfelismerőnek a hatékonyságát vizsgáltam egy 225 hangfájlból, azaz 2250 szóból álló tesztanyaggal. A felismerő alapegységének – a kísérleti eredményeknek megfelelően – a diádot választottam. A rejtett Markov modell paramétereinek optimalizálása – ezen belül is az állapotok számának megválasztása – okozta a legjelentősebb javulást. A helyesség 6 %-kal, a pontosság csaknem 40 %-kal nőtt. A tesztelésekkel főként a különböző nyelvtanokat vizsgáltam, azok alkalmazása milyen hatékonyságbeli változást idéznek elő.

Az esetek:

- hosszú felpattanó zárhangok helyettesítése rövid párjukkal
- zöngétlenedés
- zöngésedés
- teljes hasonulás
- egyedi besorolást igénylő nyelvtani esetek (képzés helye szerinti részleges hasonulás)

Minden esetben hatékonyság javulás volt megfigyelhető. Az adott esetek vizsgálatára speciálisan összeállított tanító illetve tesztelő adatbázisok esetén az eredmények szemléletesebbek lettek volna, de sajnos ilyen hanganyagok nem álltak rendelkezésemre.

5. táblázat Adott jelenség egy előfordulása mekkora javulást okoz a pontosságban

	Pontossági javulások	Előfordulások	Javulás/100 előfordulás
Kettős felpattanó zárhangok	3,33%	253	1,316%
Zöngétlenedés	1,43%	269	0,532%
Zöngésedés	0,83%	191	0,435%
Teljes hasonulás	1,35%	195	0,692%
Egyedi esetek	0,39%	77	0,506%

Az 5.táblázat szemlélteti, hogy az adott jelenség egy előfordulása hány század százalékos javulást eredményez a pontosságban. Így az értékek már összehasonlíthatók. Az értékek alapján látható, hogy az egyedi esetek is közel hasonlóan hatékonyak, és kimagaslóan a kettős felpattanó zárhangok módosítása adja a legjobb eredményt, de meg kell jegyeznünk, hogy a vizsgálata nem azonos körülmények között zajlott a többi nyelvtani esetekhez képest.

Megjegyzendő, hogy bár a változtatások javulást eredményeztek, azok visszaalakítása már nem egyértelmű, például *lábtól* -> *láptól*.

A teljes hasonulások eseténél érdemes lenne további teszteléseket elvégezni egy speciális adatbázison, ahol nem fordulnak elő mássalhangzó duplázódások, csak szóbeli teljes hasonulások. Ebben az esetben támasztható alá vagy cáfolható meg a feltételezés miszerint ezekben az esetekben a HMM modellek közül azok, amelyek az adott mássalhangzót egyszeresen tartalmazzák nagyobb valószínűséggel fognak illeszkedni az adott szóba, mint ami valóban helyes lenne, így növelve a téves törlések számát.

Az egyedi besorolást igénylő esetekre vonatkozó teszteléseknél érdemes lenne további vizsgálatokat végezni csak a *ch* illetve *cz* hangkapcsolatokra, hogy valóban a nem egyértelmű megfeleltetés miatt mutatkozott romlás a beszúrások számánál.

Az összes felsorolt hasonulási szabály együttes alkalmazásával a felpattanó zárhangok megkülönböztetése nélküli kiinduló eredményekhez képest a helyességben több mint 5%, a pontosságban csaknem 5% javulást lehet elérni a számítási igény növelése nélkül.

A nyelvtanok illetve a többi hatékonyságnövelő módszerek, mint a HMM modellek 8 állapotúnak való megválasztása, a *kettős Gauss eloszlás* alkalmazása 0,1 és 0,9-es súlyokkal, a HVite *p* paraméter értékének -30 megadása, együttes alkalmazásával egy 91 százalékos helyességű és 89,54 százalékos pontosságú beszédfelismerőt hoztam létre. Összevetve a kiindulási adatokkal, ahol a helyesség 82,66 %, a pontosság 41,66 %, a fentebb felsorolt módszerek együttes alkalmazásával a helyesség 8,34 %-ot a pontosság 47,88 %-ot javult.

Mivel a rendszer beszélőfüggő, fontos, hogy más személyek általi tesztelés során az eredményesség nagymértékben csökkenhet, viszont több személytől származó tanító hanganyagok segítségével robosztusabbá és hatékonyabbá tehető a rendszer.

Végezetül szeretnék köszönetet mondani **Dr. Czap Lászlónak** konzulensemnek mindenkori szíves segítségéért.

"A bemutatott kutató munka a TÁMOP-4.2.1.B-10/2/KONV-2010-0001 jelű projekt részeként az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósult meg."

Irodalomjegyzék

- [1] Németh Géza- Olasz Gábor: A Magyar Beszéd, Akadémiai Kiadó, Budapest, 2010
- [2] Pintér Judit Mária: Akusztikus beszédfelismerés rejtett Markov- modellek alkalmazásával, Szakdolgozat
- [3] Dr. Czap László: Audiovizuális beszédfelismerés és beszéd-szintézis PhD értekezés
- [4] Vicsi, K., Vigh, A. (1995) Text independent neural network/rule based hybrid, continuous speech recognition. EUROSPEECH'95. Madrid: pp. 2201-2204
- [5] <http://hu.wikipedia.org/wiki/Markov-l%C3%A1nc>
- [6] http://en.wikipedia.org/wiki/Hidden_Markov_model
- [7] <http://htk.eng.cam.ac.uk/ftp/software/htkbook.pdf.zip>
- [8] Nyelvművelő kézikönyv I (Akadémiai kiadó, Budapest)
- [9] Rabiner, L. R.: A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2):257-286, 1989.
- [10] <http://www.ngkszki.hu/seged/csej-lyesj/hangkepzes.htm>
- [11] <http://www.retorika.hu/hallasfejlesztes-artikulacio-kiejtes-javitas-beszedhibak>
- [12] http://cs.ubbcluj.ro/~csatol/mestint/diak_pdf/2005/benkerika.pdf
- [13] <http://www.mathworks.com/>
- [14] <http://voxforge.org/home/dev/acousticmodels/linux/create/htkjulius/how-to>
- [15] http://en.wikipedia.org/wiki/Mixture_model
- [16] <http://mek.niif.hu/01500/01547/01547.pdf>

Mellékletek

1. Átalakítások

Egykarakteres kódokká alakítás:

ü - U ö - O é - E í - i ű - U ú - u ő - O ó - o ccs - CC cs - C

ssz - SS sz - S tty - TT ty - T á - A ggy - GG gy - G zzs - ZZ zs - Z nny - NN

ny - N lly - jj ly - j

A hosszú felpattanó zárhangok helyettesítése rövid párjukkal:

bb - b tt - t dd - d pp - p kk - k GG - G gg - g TT - T

2. Nyelvtan fájl tartalma

```
$TLOOP_BEGIN_FLLWRS=#a|A|#b|B|#c|C|#d|D|#e|E|#f|F|#g|G|#h|H|#i|I|#j|J|#k|K|#l|L|#m|M|#n|N|#o|O|  
#p|P|#r|R|#s|S|#t|T|#u|U|#v|V|#z|Z|#t|T|#u|U;
```

```
$TLOOP_END_PREDS=sil|a#|A#|b#|B#|c#|C#|d#|D#|e#|E#|f#|F#|g#|G#|h#|H#|i#|I#|j#|J#|k#|K#|l#|L#|m#|M#|n#|N#|o#|O#|p  
#|P#|s#|S#|t#|T#|u#|U#|v#|V#|z#|Z#|t#|T#|u#|U#;
```

```
$k=ks|kv|kc|kj|kS|kl|km|kf|kh|kn|kr|k#|ki|ke|kE|kt|ka|ko;
```

```
$a=a#|aa|ab|ac|aC|ad|ae|aE|af|ag|aG|ah|ai|aj|ak|al|am|an|aN|ao|ap|ar|as|aS|at|aT|au|aU|av|az|a  
Z;
```

```
$A=A#|AA|Aa|Ab|Ac|AC|Az|AZ|Av|Ad|Af|AE|Ae|Ag|AG|Ah|Ai|Aj|Ak|Al|Am|An|AN|Ap|  
Ar|As|AS|At|AT;
```

```
$b=b#|ba|bA|bd|be|bg|bE|bi|bj|bl|bm|bn|bN|bo|bO|br|bu|bU|bv|bz;
```

```
$c=c#|ca|cA|cc|ce|cE|ch|ci|ck|co|cO|cr|ct|cu|cU|cv|cl;
```

```
$C=C#|Ca|CA|CC|Ce|CE|Ci|Ck|Cm|Cn|Co|CO|Cr|Cu|CU|Cv;
```

```
$d=d#|da|dA|db|de|dE|dg|di|dl|dm|dn|do|dO|dr|du|dU|dv|dz|dZ;
```

```
$e=e#|ea|eA|eb|ec|eC|ed|ee|eE|ef|eg|eG|eh|ei|ej|ek|el|em|en|eN|eo|ep|er|es|eS|et|eT|eu|eU|ev|e  
z|eZ;
```

```

$E=E#|Ea|Eu|Eb|Ec|EC|Ed|Ez|EZ|EE|Ef|Eg|EG|Eh|Ei|Ej|Ek|El|Em|En|EN|Ep|Er|Es|ES|Et|ET|Ev;
$f=f#|fa|fA|fe|fE|ff|fi|fj|fl|fo|fO|fr|fS|ft|fT|fu|fU|fh;
$g=g#|ga|gA|gb|gd|ge|gE|gi|gj|gl|gm|gn|gN|go|gO|gr|gu|gU|gv|gz;
$G=G#|Ga|GA|Gb|Ge|GE|Gi|Gl|Gm|Gn|GN|Go|Gd|GO|Gr|Gs|GS|Gu|GU|Gv|Gz;
$h=h#|ha|hA|he|hE|hi|hl|hm|hn|ho|hO|hr|ht|hu|hU|hk;
$i=i#|ia|iA|ib|ic|iC|id|ie|iE|if|ig|iG|ih|ii|ij|ik|il|im|in|iN|io|iO|ip|ir|is|iS|it|iT|iU|iV|iZ|iZ;
$j=j#|ja|jA|jb|jd|je|jE|jf|jg|jh|ji|jj|jk|jl|jm|jn|jo|jO|jr|js|jS|jt|jT|ju|jU|jv|jz|jz;
$k=k#|ka|kA|kc|kC|ke|kE|kf|kh|ki|kj|kl|km|kn|kN|ko|kO|kp|kr|ks|kS|kt|ku|kU|kv;
$l=l#|la|lA|lb|lc|lC|ld|le|lE|lf|lg|lG|lh|li|lk|lm|ln|lN|lo|lO|lp|lr|ls|lS|lt|lT|lu|lU|lv|lz;
$m=m#|ma|mA|mb|mC|me|mE|mG|mh|mi|mj|mk|ml|mm|mn|mo|mO|mp|mr|ms|mS|mt|mu|mU|mz|mN;
$n=n#|na|nA|nc|nC|nd|ne|nE|nf|ng|nh|ni|nk|nl|nm|nn|no|nO|nr|ns|nS|nt|nu|nU|nv|nz;
$N=N#|Na|NA|Nb|Ne|NE|Nf|NG|Nh|Ni|Nk|Nl|Nn|NN|No|NO|Np|Nr|Ns|NS|Nt|NT|Nu|NU|Nv|Nz|NZ;
$o=o#|oa|oA|ob|oc|oC|od|oe|of|og|oG|oh|oi|oj|ok|oe|ol|om|on|oN|oo|op|or|os|oS|ot|oT|ou|ov|oz|oZ;
$O=O#|Oa|OA|Ob|Oc|OC|Od|Oe|OE|Of|Og|OG|Oh|Oi|Oj|Ok|Ol|Om|On|ON|Oo|OO|Op|Or|Os|OS|Ot|OT|Ou|Ov|Oz|OZ;
$p=p#|pa|pA|pc|pC|pe|pE|pf|ph|pi|pj|pk|pl|pn|po|pO|pr|ps|pS|pt|pT|pu|pU|pv|pm;
$r=r#|ra|rA|rb|rc|rC|rd|re|rE|rf|rg|rG|rh|ri|rj|rk|rl|rm|rn|rN|ro|rO|rp|rr|rs|rS|rt|rT|ru|rU|rv|rz|rZ;
$s=s#|sa|sA|sc|sC|se|sE|sf|sh|si|sk|sl|sm|sn|sN|so|sO|sp|sr|ss|st|sT|su|sU|sv;
$S=S#|Sa|SA|Se|SC|SE|Sf|Sh|Si|Sk|Sl|Sm|Sn|SN|So|SO|Sp|Sr|Ss|SS|St|ST|Su|SU|Sv;
$t=t#|ta|tA|te|tE|tf|th|ti|tk|tl|tm|tn|to|tO|tp|tr|tu|tU|tv;
$T=T#|Ta|TA|Te|TE|Tf|Th|Ti|Tk|TN|To|TO|Tp|Tt|Tu|TU|Tv;
$u=u#|ua|uA|ub|uc|uC|ud|ue|uf|ug|uG|uh|ui|uj|uk|ul|um|un|uN|up|ur|us|uS|ut|uT|uv|uz|uZ;
$U=U#|Ub|UC|Ud|Ue|UG|Uf|Uh|Uj|Uk|Ul|Um|Un|Up|Ur|Us|US|Ut|UT|Uv|Uz|UZ;
$v=v#|va|vA|vb|vd|ve|vE|vi|vj|vn|vo|vO|vu|vU|vv;
$z=z#|za|zA|zb|zd|ze|zE|zg|zi|zl|zm|zn|zo|zO|zr|zu|zU|zv|zz;
$Z=Za|ZA|Zb|Zd|Ze|ZE|Zg|ZG|Zi|Zl|Zo|ZO|Zu|ZZ|Z#;
$#=sil|TLOOP_END;
$q=#a|#A|#b|#c|#d|#e|#E|#f|#g|#G|#h|#i|#j|#k|#l|#m|#n|#N|#o|#O|#p|#r|#s|#S|#t|#T|#u|#U|#v|#z|#Z|#t|#T|#u|#U;
( sil <<      TLOOP_BEGIN+TLOOP_BEGIN_FLLWRS |
TLOOP_END_PREDS-TLOOP_END |

```

sil+q|#a+a|a#++#|A+A|A#++#|aa+a|AA+A|Aa+a|Ae+e|ab+b|Ab+b|ac+c|aC+C|Ac+c|AC+C|a
d+d|Ad+d|ae+e|aE+E|AE+E|af+f|Af+f|ag+g|aG+G|Ag+g|AG+G|ah+h|Ah+h|ai+i|Ai+i|aj+j|
Aj+j|ak+k|Ak+k|al+l|Al+l|am+m|Am+m|an+n|aN+N|An+n|AN+N|ao+o|ap+p|Ap+p|ar+r|Ar
+r|as+s|aS+S|As+s|AS+S|at+t|aT+T|At+t|AT+T|au+u|aU+U|av+v|Av+v|az+z|aZ+Z|Az+z|A
Z+Z|#b+b|b#++#|ba+a|bA+A|bd+d|be+e|bE+E|bi+i|bj+j|bg+g|bl+l|bm+m|bn+n|bN+N|bo+o|b
O+O|br+r|bu+u|bU+U|bv+v|bz+z|#c+c|c#++#|C+C|C#++#|ca+a|cA+A|Ca+a|CA+A|cc+c|CC
+C|ce+e|cE+E|Ce+e|CE+E|ch+h|ci+i|cl+l|Ci+i|ck+k|Ck+k|Cm+m|Cv+v|Cn+n|co+o|cO+O|
Co+o|CO+O|cr+r|Cr+r|ct+t|cu+u|cU+U|Cu+u|CU+U|cv+v|#d+d|d#++#|da+a|dA+A|db+b|de+
e|dE+E|dg+g|di+i|dl+l|dm+m|dn+n|do+o|dO+O|dr+r|du+u|dU+U|dv+v|dz+z|dZ+Z|#e+e|e#++
#|E+E|E#++#|ea+a|eA+A|eb+b|Eb+b|ec+c|eC+C|Ec+c|EC+C|ed+d|Ed+d|ee+e|eE+E|EE+E|
ef+f|Ef+f|eg+g|eG+G|Eg+g|EG+G|eh+h|Eh+h|ei+i|Ei+i|ej+j|Ea+a|Eu+u|Ej+j|ek+k|Ek+k|el+
l|El+l|em+m|Em+m|en+n|eN+N|En+n|EN+N|eo+o|ep+p|Ep+p|er+r|Er+r|es+s|eS+S|Es+s|ES
+S|et+t|eT+T|Et+t|ET+T|eu+u|eU+U|ev+v|Ev+v|ez+z|eZ+Z|Ez+z|EZ+Z|#f+f|f#++#|fa+a|fh+
h|fA+A|fe+e|fE+E|ff+f|fi+i|fj+j|fl+l|fo+o|fO+O|fr+r|fS+S|ft+t|fT+T|fu+u|fU+U|#g+g|g#++#|
G+G|G#++#|ga+a|gA+A|Ga+a|GA+A|gb+b|Gb+b|Gd+d|gd+d|ge+e|gE+E|Ge+e|GE+E|gi+i|
Gi+i|gj+j|gl+l|Gl+l|gm+m|Gm+m|gn+n|gN+N|Gn+n|GN+N|go+o|gO+O|Go+o|GO+O|gr+r|
Gr+r|Gs+s|GS+S|gu+u|gU+U|Gu+u|GU+U|gv+v|Gv+v|gz+z|Gz+z|#h+h|h#++#|ha+a|hA+A|hk+k|h
A+A|he+e|hE+E|hi+i|hl+l|hM+m|hn+n|ho+o|hO+O|hr+r|ht+t|hu+u|hU+U|#i+i|i#++#|ia+a|iA
+A|ib+b|ic+c|iC+C|id+d|ie+e|iE+E|if+f|ig+g|iG+G|ih+h|ii+i|ij+j|ik+k|il+l|im+m|in+n|iN+N|i
o+o|iO+O|ip+p|ir+r|is+s|iS+S|it+t|iT+T|iu+u|iU+U|iv+v|iz+z|iZ+Z|#j+j|j#++#|ja+a|jA+A|jb+b|
jd+d|je+e|jE+E|jf+f|jg+g|jh+h|ji+i|jc+c|jj+j|jk+k|jl+l|jm+m|jn+n|jo+o|jO+O|jr+r|js+s|jS+S|
jt+t|jT+T|ju+u|jU+U|jv+v|jz+z|#k+k|k#++#|ka+a|kA+A|kc+c|kC+C|ke+e|kE+E|kf+f|kh+h|ki
+i|kj+j|kl+l|km+m|kn+n|kN+N|ko+o|kO+O|kp+p|kr+r|ks+s|kS+S|kt+t|ku+u|kU+U|kv+v|#l+l|
l#++#|la+a|lA+A|lb+b|lc+c|lC+C|ld+d|le+e|lE+E|lf+f|lg+g|lG+G|lh+h|li+i|lk+k|lm+m|ln+n|l
N+N|lo+o|lO+O|lp+p|lr+r|ls+s|lS+S|lt+t|lT+T|lu+u|lU+U|lv+v|lz+z|#m+m|m#++#|ma+a|mA
+A|mb+b|mC+C|me+e|mE+E|mG+G|mh+h|mi+i|mj+j|mk+k|ml+l|mm+m|mN+N|mn+n|mo
+o|mO+O|mp+p|mr+r|ms+s|mS+S|mt+t|mu+u|mU+U|mz+z|#n+n|n#++#|N+N|N#++#|na+a|n
A+A|Na+a|NA+A|Nb+b|nc+c|nC+C|nd+d|ne+e|nE+E|Ne+e|NE+E|nf+f|Nf+f|ng+g|NG+G|n
h+h|Nh+h|ni+i|Ni+i|nk+k|Nk+k|nl+l|Nl+l|nm+m|nn+n|Nn+n|NN+N|no+o|nO+O|No+o|NO
+O|Np+p|nr+r|Nr+r|ns+s|nS+S|Ns+s|NS+S|nt+t|Nt+t|NT+T|nu+u|nU+U|Nu+u|NU+U|nv+v|
Nv+v|nz+z|Nz+z|NZ+Z|#o+o|o#++#|O+O|O#++#|oa+a|oA+A|Oa+a|OA+A|ob+b|Ob+b|oc+c|
oC+C|Oc+c|OC+C|od+d|Od+d|oe+e|Oe+e|OE+E|of+f|Of+f|og+g|oG+G|oe+e|Og+g|OG+G
|oh+h|Oh+h|oi+i|Oi+i|oj+j|Oj+j|ok+k|Ok+k|ol+l|Ol+l|om+m|Om+m|on+n|oN+N|On+n|ON
+N|oo+o|Oo+o|OO+O|op+p|Op+p|or+r|Or+r|os+s|oS+S|Os+s|OS+S|ot+t|oT+T|Ot+t|OT+T|
ou+u|Ou+u|ov+v|Ov+v|oz+z|oZ+Z|Oz+z|OZ+Z|#p+p|p#++#|pa+a|pA+A|pm+m|pc+c|pC+C|
pe+e|pE+E|pf+f|ph+h|pi+i|pj+j|pk+k|pl+l|pn+n|po+o|pO+O|pr+r|ps+s|pS+S|pt+t|pT+T|pu+u|
pU+U|pv+v|#r+r|r#++#|ra+a|rA+A|rb+b|rc+c|rC+C|rd+d|re+e|rE+E|rf+f|rg+g|rG+G|rh+h|ri+i
|rj+j|rk+k|rl+l|rm+m|rn+n|rN+N|ro+o|rO+O|rp+p|rr+r|rs+s|rS+S|rt+t|rT+T|ru+u|rU+U|rv+v|r
z+z|rZ+Z|#s+s|s#++#|S+S|S#++#|sa+a|sA+A|Sa+a|SA+A|sc+c|sC+C|se+e|sE+E|Se+e|SE+E|s
f+f|Sf+f|sh+h|Sh+h|si+i|Si+i|sk+k|Sk+k|sl+l|Sl+l|sm+m|Sm+m|sn+n|sN+N|Sn+n|SN+N|so+
o|sO+O|SC+C|So+o|SO+O|sp+p|Sp+p|sr+r|Sr+r|ss+s|Ss+s|SS+S|st+t|sT+T|St+t|ST+T|su+u|
sU+U|Su+u|SU+U|sv+v|Sv+v|#t+t|t#++#|T+T|T#++#|ta+a|tA+A|Ta+a|TA+A|te+e|tE+E|Te+
e|TE+E|tf+f|Tf+f|th+h|Th+h|ti+i|Ti+i|tk+k|Tk+k|tl+l|tm+m|tn+n|TN+N|to+o|tO+O|To+o|T

O+O|tp+p|Tp+p|tr+r|Tt+t|tu+u|tU+U|Tu+u|TU+U|tv+v|Tv+v|#u+u|u#++#|U+U|U#++#|ua+a|
uA+A|ub+b|Ub+b|uc+c|uC+C|UC+C|ud+d|Ud+d|ue+e|Ue+e|uf+f|ug+g|uG+G|Ug+g|UG+G
|Uf+f|uh+h|Uh+h|ui+i|uj+j|Uj+j|uk+k|Uk+k|ul+l|Ul+l|um+m|Um+m|un+n|uN+N|Un+n|up+
p|Up+p|ur+r|Ur+r|us+s|uS+S|Us+s|US+S|ut+t|uT+T|Ut+t|UT+T|uv+v|Uv+v|uz+z|uZ+Z|Uz
+z|UZ+Z|#v+v|v#++#|va+a|vA+A|vb+b|vd+d|ve+e|vE+E|vi+i|vj+j|vn+n|vo+o|vO+O|vu+u|v
U+U|vv+v|#z+z|z#++#|Z+Z|Z#++#|za+a|zA+A|Za+a|ZA+A|zb+b|Zb+b|zd+d|Zd+d|ze+e|zE+
E|Ze+e|ZE+E|zg+g|Zg+g|ZG+G|zi+i|Zi+i|zl+l|Zl+l|zm+m|zn+n|zo+o|zO+O|Zo+o|ZO+O|zr
+r|zu+u|zU+U|Zu+u|zv+v|zz+z|ZZ+Z >>sil)

3. csere6.m matlab fájl tartalma a vizsgálatokhoz

A tanító illetve a tesztelő anyagok átírásához alkalmazott matlab fájl tartalma, amiben minden eset szerepel

```
function csere
x = input('Adja meg a txt fájl nevét kiterjesztéssel együtt:','s');
text = fileread(x);
strrep(text,'ü','U');
strrep(ans,'ö','O');
strrep(ans,'é','E');
strrep(ans,'í','i');
strrep(ans,'ű','U');
strrep(ans,'ú','u');
strrep(ans,'ő','O');
strrep(ans,'ó','o');
strrep(ans,'ccs','CC');
strrep(ans,'cs','C');
strrep(ans,'ssz','SS');
strrep(ans,'sz','S');
strrep(ans,'tty','TT');
strrep(ans,'ty','T');
strrep(ans,'á','A');
strrep(ans,'ggy','GG');
strrep(ans,'gy','G');
strrep(ans,'zsz','ZZ');
strrep(ans,'zs','Z');
strrep(ans,'nny','NN');
strrep(ans,'ny','N');
strrep(ans,'lly','jj');
strrep(ans,'ly','j');
%hasonulások
%zöngétlenedés
strrep(ans,'ch','h');
strrep(ans,'cz','c');
strrep(ans,'mf','f');
strrep(ans,'mv','v');
strrep(ans,'nb','b');
strrep(ans,'np','p');
strrep(ans,'bc','c');
strrep(ans,'bC','pC');
strrep(ans,'bh','ph');
```

```

strrep(ans,'bk','pk');
strrep(ans,'bs','ps');
strrep(ans,'bS','pS');
strrep(ans,'bt','pt');
strrep(ans,'bT','pT');
strrep(ans,'df','tf');
strrep(ans,'dh','th');
strrep(ans,'dk','tk');
strrep(ans,'dp','tp');
strrep(ans,'gc','kc');
strrep(ans,'gC','kC');
strrep(ans,'gf','kf');
strrep(ans,'gh','kh');
strrep(ans,'gp','kp');
strrep(ans,'gs','ks');
strrep(ans,'gS','kS');
strrep(ans,'gt','kt');
strrep(ans,'gT','kT');
strrep(ans,'Gf','Tf');
strrep(ans,'Gh','Th');
strrep(ans,'Gk','Tk');
strrep(ans,'gG','kG');
strrep(ans,'Gp','Tp');
strrep(ans,'vc','fc');
strrep(ans,'vC','fC');
strrep(ans,'vh','fh');
strrep(ans,'vk','fk');
strrep(ans,'vp','fp');
strrep(ans,'vs','fs');
strrep(ans,'vS','fS');
strrep(ans,'vt','ft');
strrep(ans,'vT','fT');
strrep(ans,'zc','sc');
strrep(ans,'zf','sf');
strrep(ans,'zh','sh');
strrep(ans,'zk','sk');
strrep(ans,'zG','SG');
strrep(ans,'zp','sp');
strrep(ans,'zt','st');
strrep(ans,'zT','sT');
strrep(ans,'Zc','sc');
strrep(ans,'ZC','sC');
strrep(ans,'Zf','sf');
strrep(ans,'Zh','sh');
strrep(ans,'Zk','sk');
strrep(ans,'Zp','sp');
strrep(ans,'Zt','st');
strrep(ans,'ZT','sT');
% némileg módosult artikulációval
strrep(ans,'bf','pf');
strrep(ans,'Gc','Tc');
strrep(ans,'GC','TC');

```

```

strrep(ans,'Gt','Tt');
% zöngésedés
strrep(ans,'cb','dZb');
strrep(ans,'cd','dZd');
strrep(ans,'cg','dzg');
strrep(ans,'cG','dzG');
strrep(ans,'Cb','dZb');
strrep(ans,'Cd','dZd');
strrep(ans,'CG','dZG');
strrep(ans,'fb','vb');
strrep(ans,'fd','vd');
strrep(ans,'fdZ','vdZ');
strrep(ans,'fg','vg');
strrep(ans,'fG','vG');
strrep(ans,'fz','vz');
strrep(ans,'fZ','vZ');
strrep(ans,'kb','gb');
strrep(ans,'kd','gd');
strrep(ans,'kdZ','gdZ');
strrep(ans,'kG','GG');
strrep(ans,'kz','gz');
strrep(ans,'kZ','gZ');
strrep(ans,'pd','bd');
strrep(ans,'pdZ','pdZ');
strrep(ans,'pg','bg');
strrep(ans,'pG','bG');
strrep(ans,'pz','bz');
strrep(ans,'pZ','bZ');
strrep(ans,'sb','Zb');
strrep(ans,'sd','Zd');
strrep(ans,'sdZ','ZdZ');
strrep(ans,'sg','Zg');
strrep(ans,'sG','ZG');
strrep(ans,'Sb','zb');
strrep(ans,'Sd','zd');
strrep(ans,'Sg','zg');
strrep(ans,'SG','zG');
strrep(ans,'Sz','zz');
strrep(ans,'tb','db');
strrep(ans,'tg','dg');
strrep(ans,'tG','GG');
strrep(ans,'tz','dz');
strrep(ans,'tZ','dZ');
strrep(ans,'Tb','Gb');
strrep(ans,'dN','GN');
strrep(ans,'tN','TN');
strrep(ans,'nb','mb');
strrep(ans,'np','mp');
strrep(ans,'nG','NG');
strrep(ans,'nT','NT');
strrep(ans,'cZ','dZ');
strrep(ans,'Cz','dZ');

```

```

strrep(ans,'SdZ','zd');
strrep(ans,'SZ','zZ');
strrep(ans,'zC','SC');
%teljes hasonulás
strrep(ans,'kg','gg');
strrep(ans,'pb','bb');
strrep(ans,'sZ','ZZ');
strrep(ans,'Sz','zz');
strrep(ans,'td','dd');
strrep(ans,'bp','pp');
strrep(ans,'dt','tt');
strrep(ans,'gk','kk');
strrep(ans,'GT','TT');
strrep(ans,'vf','ff');
strrep(ans,'zS','SS');
strrep(ans,'Zs','ss');
strrep(ans,'lj','jj');
strrep(ans,'nj','NN');
strrep(ans,'dG','GG');
strrep(ans,'dj','GG');
strrep(ans,'tT','TT');
strrep(ans,'Tj','TT');
strrep(ans,'Nj','NN');
strrep(ans,'Gj','GG');
strrep(ans,'tj','TT');
strrep(ans,'ddZ','dZZ');
strrep(ans,'dS','cc');
strrep(ans,'ds','C');
strrep(ans,'tc','cc');
strrep(ans,'tS','cc');
strrep(ans,'ts','C');
strrep(ans,'tC','CC');
strrep(ans,'dc','tc');
strrep(ans,'dC','tC');
strrep(ans,'tdZ','ddZ');
strrep(ans,'auguStus','aguStus');
%hosszú felpattanó zárhang helyett rövid párja
strrep(ans,'bb','b');
strrep(ans,'tt','t');
strrep(ans,'dd','d');
strrep(ans,'pp','p');
strrep(ans,'kk','k');
strrep(ans,'GG','G');
strrep(ans,'gg','g');
strrep(ans,'TT','T');
fid = fopen(x, 'w');
fprintf(fid, '%s', ans);
fclose(fid);
disp('KÉSZ');
open (x);

```