

MAGYAR NYELVŰ FOLYAMATOS BESZÉD GÉPI FELISMERÉSE AKUSZTIKAI-FONETIKAI-FONOLÓGIAI SZINTEN*

Vicsi Klára - Vig Attila - Berényi Péter

MTA Akusztikai Kutatólaboratórium

Bemutatjuk az MTA Akusztikai Kutatólaboratóriumában OTKA támogatással létrehozott szövegtől független folyamatos beszédet felismerő számítógépes rendszert, valamint a létrehozásához szükséges előzetes statisztikai vizsgálatokat. A statisztikai vizsgálatok eredményeként a felismerés alapegységéül a fészótagokat választottuk. A rendszer több szintből épül fel. Szintenként a feldolgozási módszer más és más: egészében véve különböző szerkezetű neurális hálózatoknak és szabálybázisú feldolgozásnak mindig az adott szinten elvégzendő feladathoz illeszkedő kombinációja. Az egyes szintek egymással együttműködnek, a felsőbb szintek visszacsatolásával az alsóbb szintek hibája korrigálható. A rendszer kimenete egy fészótaglánc sorozat, ahol a felismerési valószínűségek függvényében első-, másod- és harmadrendű jölötteket tüntetünk fel.

1. Bevezetés

Széles körben ismert, hogy izolált szavas gépi beszéd-felismerésben a kutatók igen szép sikereket értek el. Léteznek már 10000-20000 szavas diktafon berendezések, amelyek beszélőadaptívan képesek szavak között szünettel megszakított beszéd felismerésére (Baker 1993). A folyamatos gépi beszéd felismerés valódi megoldásához azonban még számos komoly problémát kell megoldani.

A folyamatos beszédnél a mérhető jellemzők variáltsága még fokozottabb, mint az izoláltan kimondott szavaknál.

*A tanulmány második részének megjelenését a sorozat következő számában tervezzük.

A kiejtés kevésbé precíz, a beszélők közötti különbségek még fokozottabbak, a beszédtempó még jobban változik. A hangkapcsolatok hatása szóhatároknál is jelentékeny. Végül, új probléma is jelentkezik. Egy szó jelentését egy kifejezésben a hangsúly, nyomaték befolyásolja, ezzel együtt változik annak akusztikai realizációja is. Habár a beszéd különálló szavak soraként érzékelhető, a szóhatárok megkeresése nagyon nehéz. Ilyen körülmények között olyan rendszer, amely csak akusztikus elemzést és mintaosztályozást végez, a folyamatos beszéd felismerésére alkalmatlan.

Folyamatos beszéd felismerésnél figyelembe kell venni az adott nyelv ismert szabályait. A természetes beszéd felismerésénél az emberek alkalmazzák a megtanult nyelvészeti összefüggéseket, szabályokat. Kísérletileg bizonyított tény (Pollak - Pickett 1963), hogy szintaktikai, szemantikai szabályok nélkül az ember sem képes 85%-nál nagyobb felismerési biztonságot produkálni. A hatékony beszéd felismerő berendezések mindenképpen nyelvi elemzésen alapuló tudást használnak fel, és ezt minél több szinten és minél hatékonyabban használják, annál jobb a rendszer.

A nyelvészeti tudás számos ún. szintje létezik. Ezek a következők: akusztikai-fonetikai szint, fonológiai szint, lexikai szint, szintaktikai szint, prozódiai szint, szemantikai szint, pragmatikai szint stb. Általában nem minden szintet kell és szokás használni, de megfelelő kombinációban alkalmazva a különböző szintű szabályokat vagy azok egy részét, a beszéd felismerők képesek lehetnek arra, hogy az adott kiejtés nyelvtanilag helyes, jelentéssel bíró lejegyzését megadják.

2. A felismerés alapegységének kiválasztása, beszéd-elemtár meghatározása

A gépi beszéd felismerés ismert útja az, hogy megkíséreljük a beszéd folyamatot azonosítható elemek sorozatára bontani, majd az azonosítást elvégezve következtetni a

beszéd folyamatban kódolt üzenetre.

A gyakorlatban működő algoritmusok persze folyamatos beszéd esetén a felbontást (szegmentálást) és azonosítást (címkézést) nem független munkafolyamatokként végzik, hanem a sok lehetséges (szegmentálási és címkézési) változat közül gyakran igen bonyolult keresési és optimalizálási feladatok megoldásán keresztül próbálják kikeresni azt, amelyik a legjobbnak tűnik. Mindebből az következik, hogy egy gépi beszéd felismerő tervezésénél az első döntési pont (megelőzve az algoritmusok, szoftver- és hardver-eszközök stb. kiválasztását) a beszéd felismerés alapegységének - a szegmentálással elkülönítendő, majd azonosítandó elemeknek - a megválasztása, majd a használandó beszédelemtár pontos definiálása. Első pillantásra kézenfekvőnek tűnik, hogy ezek az elemek a beszédhangok (fonémák) legyenek (Takács 1992). A tüzetesebb fonetikai vizsgálat azonban megmutatja, hogy ezek realizációja erősen kontextusfüggő, így a fonémák önmagukban nem igazán jól azonosítható elemek.

A gyakorlatban is használható elemtárral szemben a következő követelményeket támasztjuk:

1. fedje le a nyelv egészét vagy legalább annak túlnyomó részét,
2. számossága ne legyen elviselhetetlenül nagy,
3. az elemek realizációja ne legyen túl erősen kontextusfüggő.

E követelmények közül az első kettő statisztikai természetű, a harmadik fonetikai.

2.1. A statisztikai vizsgálat leírása

E célból mintegy 230000 szónyi modern magyar prózát dolgoztunk fel, két tucatnyi különböző szerzőtől. Azokat az elemeket vizsgáltuk, amelyek az eddigi kutatások során egyáltalán szóba kerültek mint a gépi beszéd felismerés lehetséges egységei. Ezek pedig a következők: szó, szótag,

félszótag, mássalhangzó-csoport, trifon, valamint difon. E célból a szöveget először megfelelő egységes formára hoztuk, majd egy helyesírás-ellenőrző programon átszűrtük. A következő lépés a fonetikai átírat elkészítése volt a hasonulási és egyéb szabályok figyelembevételével. Minden további vizsgálatot ezen az anyagon végeztünk.

Eredmények

1. SZÓ, pontosabban: szóalak - nem azonos a szótári elemmel, hiszen a morfológiai elemzés nem előzheti meg a felismerést. Így egy szó valamennyi ragozott, képzett formáját, valamint összetételét különböző elemnek tekintettük. Azt a némileg meglepő eredményt kaptuk, hogy a magyar szóalakok száma potenciálisan végtelen, pontosabban minden (eddig általunk vizsgált) elég nagy szövegben a szóalakok kb. 10%-a új.

2. SZÓTAG: intuitív fogalma magától értetődőnek tűnik, de ha egzakt operatív definíciót akarunk nyerni, nehézségekbe ütközünk. Ha eltekintünk néhány vitatható esettől (*autó, pszt, brr, hm...*), megszámlálni könnyű a szótagokat: pontosan annyi lesz belőlük, ahány magánhangzót az illető szöveg tartalmaz. Ez persze korántsem egy univerzális, a vizsgált nyelvtől függetlenül érvényes szabály, de a magyar nyelv esetén jó közelítéssel alkalmazható. A szótaghatár megállapítása már sokkal nehezebb probléma. Az Akadémia helyesírási szabályai nyilván nem lehetnek mérvadóak, hiszen azok az ortografikus reprezentációra vonatkoznak, s nem a fonetikus átíratra. Mi viszont a fonetikai átíraton végeztük a szótaghatár megállapítást az alábbiak szerint: V-V, V-CV, VC-CV, VCC-CV.

3. FÉLSZÓTAG: kezdő és záró félszótagot határoztunk meg. Köztük a határ a magánhangzó kvázistacioner szakaszának a közepe.

4. TRIFON: hanghármás

5. DIFON: hangpár

A 2.-5. hangkapcsolatokra végzett statisztikai vizsgálat eredményét az 1. táblázat mutatja.

Látható, hogy a félszótagok adják a legkompaktabb elemkészletet úgy, hogy közben a kontextus jelentős részét is magukban foglalják. Például 176 db kezdő félszótaggal és 150 db zárófélszótaggal, tehát összesen 326 db félszótaggal egy általános szöveg 95%-a, 576 db félszótaggal pedig a 99%-a leírható. Amennyiben az összes fonémát jellemezzük a következő 9 akusztikai jegy alapján (akusztikai osztály), úgy mint

1. csendszakaszból álló fonémarész:	CSEND (-)
2. zöngeszakasz:	ZÖNGE (+)
3. zöngétlen zárfelpattanási zörej:	- BURST
4. zöngés zárfelpattanási zörej:	+ BURST
5. zöngétlen spiráns zörej:	- SPIR
6. zöngés spiráns zörej:	+ SPIR
7. rezonáns jellegű mássalhangzók	NALI
8. mélyszíneképű magánhangzó:	MéM
9. magasszíneképű magánhangzó:	MaM ,

akkor a félszótagok félszótagtípusokba csoportosíthatók aszerint, hogy az azokat felépítő fonémák a 9 akusztikai jegy alapján melyekkel jellemezhetők. A félszótagtípusok gyakorisági sorrendben a 2. táblázatban láthatók.

Egy általános szöveg 95%-a 30 különböző félszótag típusba, a 99%-a pedig 36 különböző félszótag típusba csoportosítható.

2.2. Optimális méretű szöveg összeállítása a beszéd-felismerők betanítására

Általánosan elfogadott megoldás, hogy a beszéd-felismerők betanítása nagyméretű adatbázis alapján történik (Cocosda 1992), nagyméretű a szöveganyag, hogy minden lényeges elem elegendő számban forduljon elő benne. Holott statisztikai

1. táblázat

EGYSÉG Kumulatív gyakoriság %	KETTŐSHANG		KEZDŐ FÉLSZÓTAG		ZÁRÓ FÉLSZÓTAG		SZÓTAG		KETTŐSHANG SZÓTAGHATÁROKNÁL	
	FREQUENCY %	ORDER NUMBER OF UNIT'S TYPE	FREQUENCY %	ORDER NUMBER OF UNIT'S TYPE	FREQUENCY %	ORDER NUMBER OF UNIT'S TYPE	FREQUENCY %	ORDER NUMBER OF UNIT'S TYPE	FREQUENCY %	ORDER NUMBER OF UNIT'S TYPE
50	0,36	71	0,91	31	1,47	12	0,16	116	0,23	91
75	0,13	183	0,39	75	0,5	41	0,05	402	0,13	218
90	0,04	399	0,16	134	0,14	95	0,01	1011	0,05	408
95	0,02	586	0,08	176	0,06	150	0,005	1604	0,02	563
99	0,005	956	0,02	259	0,008	317	0,0008	3380	0,006	884
TOTAL NUMBER OF UNIT'S TYPE		1796		629		1052		6419		1677

STATISZTIKAI ELEMZÉS A
FELISMERÉS ALAPEGYSÉGEKNEK
MEGALLAPÍTÁSÁRA

(A TELJES SZÖVEG 527856 SZÓT TARTALMAZ)

Záró félézótágtípusok statisztikája

	db	kumulatív gyakoriság %	típus			
1.	126180	24.1912	MaM			
2.	119312	47.0657	MéM			
3.	58748	58.3288	MaM	Nali		
4.	53711	68.6263	MéM	Nali		
5.	33126	74.9772	MaM	Csend	-Burst	
6.	28808	80.5003	MéM	Csend	-Burst	
7.	20458	84.4225	MaM	-Spir		
8.	12430	86.8056	MaM	Zöngé	+Burst	
9.	10604	88.8386	MéM	-Spir		
10.	9504	90.6607	MéM	+Spir		
11.	7648	92.0925	MéM	Zöngé	+Burst	
12.	6816	93.3993	MaM	Nali	Csend	-Burst
13.	6622	94.6689	MéM	Zöngé	+Spir	
14.	4964	95.6206	MéM	Nali	Csend	-Burst
15.	4355	96.4555	MaM	+Spir		
16.	3515	97.1294	MaM	Zöngé	+Spir	
17.	2045	97.5215	MéM	Csend	-Spir	
18.	1989	97.9028	MéM	-Spir	Csend	-Burst
19.	1831	98.2538	MaM	Csend	-Spir	
20.	1134	98.4712	MaM	-Spir	Csend	-Burst

Kezdő félézótágtípusok statisztikája

	db	kumulatív gyakoriság %	típus		
1.	84727	16.2433		Nali	MaM
2.	63448	28.4076		Nali	MéM
3.	51869	38.3516	Csend	-Burst	MaM
4.	49686	47.8774			MéM
5.	43878	56.2897	Csend	-Burst	MéM
6.	43097	64.5523			MaM
7.	34864	71.2364		-Spir	MéM
8.	33162	77.5942		-Spir	MaM
9.	29343	83.2198	Zöngé	+Burst	MaM
10.	26036	88.2114	Zöngé	+Burst	MéM
11.	19542	91.9580		+Spir	MaM
12.	18150	95.4377		+Spir	MéM
13.	6380	96.6609	Csend	-Spir	MéM
14.	5124	97.6433	Csend	-Spir	MaM
15.	4006	98.4113	Zöngé	+Spir	MéM
16.	3854	99.1502	Zöngé	+Spir	MaM

vizsgálatra támaszkodva, a nyelv sajátosságait megtartva optimális hosszúságú szöveg állítható elő, amelyben a betanításhoz szükséges alapegységek, jelen esetben a félszótagok elengedő számban, jelen esetben legalább 10-szer fordulnak elő. Az így előállított anyagunk 150 mondatból, összesen 9889 karakterből áll, amelyben a félszótagok 95%-os kumulatív gyakoriságig szerepelnek. Ez az anyag egy egységes adatbázis alapja lehet, amely bármilyen beszédfelismerő magyar nyelvű betanítására alkalmas.

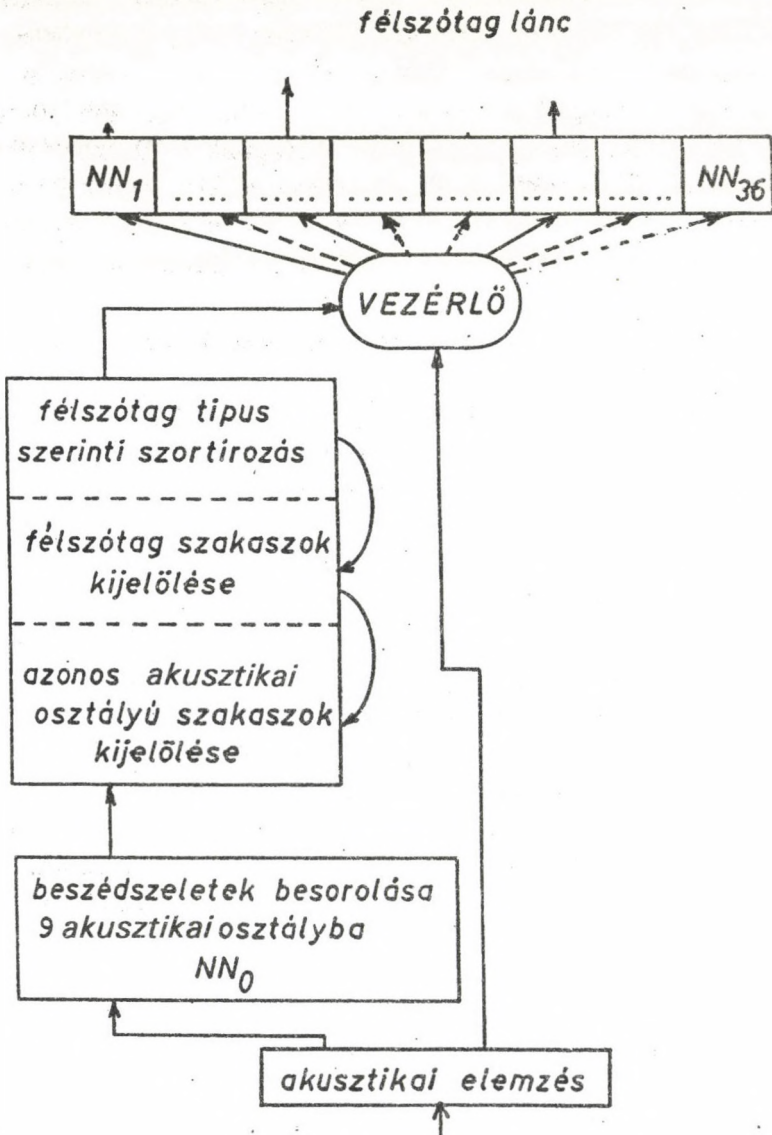
2.3. Optimális méretű szöveg összeállítása a beszéd-felismerők tesztelésére

Tesztelésre egy másik szöveget szerkesztettünk, amelyben az alapegységek eloszlása a nyelvre jellemző. E szövegek beszédátviteli és beszédfeldolgozó rendszerek tesztelésére új szabvány alapját is képezhetik. A szöveg 140 mondatból, és 8880 karakterből áll.

3. A felismerő rendszer általános leírása

A bevezetőben említett nyelvészeti szintek közül egyelőre 2 szintet kapcsolunk be a rendszerünkbe, az akusztikai-fonetikai, valamint részlegesen a fonológiai szintet. A rendszer bemenetén mérjük az akusztikai hangnyomásváltozást, és a kimeneten a félszótagok láncolatát kapjuk meg.

A feladatot több lépésre bontva oldottuk meg az 1. ábra blokkvázlata szerint. Mivel a félszótagokat választottuk a felismerés alapegységének, az első feladat ezek határait megkeresni. Ha ez megvan, akkor a félszótagszakaszok a megfelelő típusú előzetesen betanított neurális hálózat ($NN_1 - NN_{36}$) segítségével felismerhetők. A félszótaghatárok megkereséséhez először minden 10 ms-nyi beszédszeletet jellemezzük a fentiekben leírt 9 akusztikai jegy valamelyikével, vagyis minden beszédszeletet az előzetesen betanított speciális szerkezetű neurális háló (NN_0) segítségével besoroljuk a 9 akusztikai osztály valamelyikébe. Szabálybázisú algoritmusok



1. ábra

A beszédfelismerő rendszer blokkvázlata

segítségével meghatározzuk az azonos akusztikai osztályú szakaszokat, pl. spiráns szakasz, magas magánhangzó szakasz, nazális szakasz stb. Ezután ugyanilyen alapon végezzük a felszótagok határainak a kijelölését. Itt vesszük figyelembe a különböző fonológiai szabályokat is. A határok kijelölése után lehet a felszótagszakaszokat a megfelelő felszótag típusba besorolni, és az adott típusú felszótagok azonosítására szolgáló neurális hálóba elküldeni.

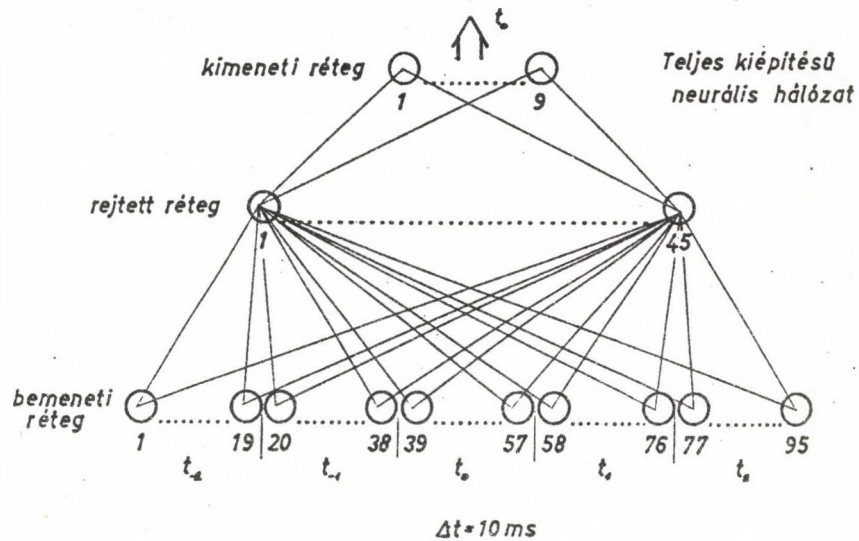
A fent körvonalazott felismerési rendszer megvalósításához host számítógépként egy EVEREX 486 PC-t használtunk. Az A-D átalakítást, a digitális szűrést egy RTD 200 típusú jelprocesszor kártyán, a különböző felépítésű neurális hálószerkezeteket ARIEL DSP32C típusú jelfeldolgozó kártyán valósítottuk meg.

3.1 Akusztikai elemzés

Méréskor a pszichoakusztikában jól ismert kritikus sávokat (Zwicker 1982) megvalósító digitális szűrősort használunk 200 Hz és 7700 Hz közötti frekvenciasávban (19 szűrő). Az akusztikai elemző kimenetén minden 10 ms-ban (beszédszeletben) 19, a szűrő kimenetét jellemző energiaadat jelenik meg (Vicsi - Mattila - Berényi 1990)

3.2 Beszédszeletek besorolása 9 akusztikai osztályba neurális háló segítségével

Teljes kiépítésű neurális hálót (back propagation) (Klimasauskas 1989) használunk a beszédszeletek 9 akusztikai osztályba sorolásához a 2. ábrán bemutatott felépítésben. A bemeneti réteg 5×19 neuronból áll, azaz 5 beszédszelet 19 szűrőadatát adjuk be egyszerre a hálóba, amivel csak a középső t_0 szeletet jellemezzük. A t_0 leírásához tehát figyelembe vesszük a 2 megelőző és a 2 követő szeletet is, hiszen a hangképzés folyamatos cselekvés, időben folyamatos hangeseményt hozva létre. A pillanatnyi hangeseményt a múlt és a cél hangesemény befolyásolja. A háló 9 kimenete a 9 akusztikai osztálynak felel



2. ábra
 SEPT1 neur. hálózat szerkezete

meg. A beszédszeletet abba az osztályba soroljuk be, amelyiknél a háló kimeneti aktivitása a legnagyobb. Egy ismeretlen mondat esetében kapott tévesztési mátrixot a 3. ábrán mutatjuk be. A 4. ábrán a kilenc akusztikai osztálynak megfelelő kimenetekben mért aktivitás változását mutatjuk be az idő függvényében.

3.3 Félszótagszakaszok kijelölése, szortírozás

Ezt a lépést szabálybázisú algoritmusok végrehajtásával oldottuk meg, mert előzetes próbálkozásaink szerint így kapjuk a legjobb eredményt, valamint a különböző fonológiai szabályok alkalmazása így a legegyszerűbb. Ennek a lépésnek minél tökéletesebb végrehajtása a felismerés kulcskérdése, hiszen ha itt a félszótaghatár meghatározásban tévesztés van, vagy egy hangzótípust kihagyunk, a vezérlő az adott szótagszakaszt egy nem hozzá illő neurális hálótípusba küldi el. Ebben az esetben biztos a félszótag tévesztése (pl. egy MaM+Nali félszótag helyett csak MaM félszótagtípust ítél meg, és a MaM-nak megfelelő neurális hálóba küldi el az adott félszótagot.). Ilyen és ehhez hasonló tévesztésekkel pedig mindig számolnunk kell. Ezért a 3. ábrán közölt tévesztési mátrix tanulságai alapján nem egy, hanem több neurális hálóba küldjük el az adott félszótag szakaszt, attól függően, hogy rendszerünkben a 9 akusztikai osztályba soroláskor milyen tipikus tévesztések fordulnak elő.

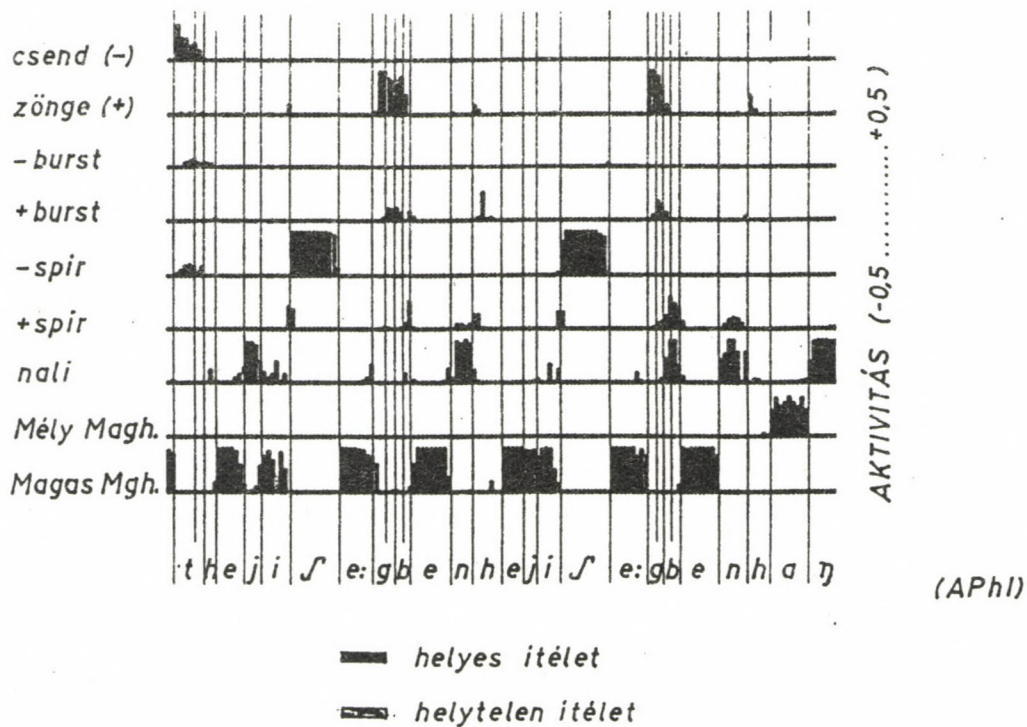
3.4 Félszótagfelismerés

A rendszerben annyi félszótag felismerésére szolgáló neurális háló van, ahány félszótag típust akarunk felismerni. A neurális háló szerkezetének felépítése a félszótag típusától függ. Jelen megoldásban 36 db különböző felépítésű hálót építettünk be. A hálók szerkezetének leírása a 3. és a 4. táblázatban olvasható. Egy adott háló felépítése zöngés zárhangból és magas magánhangzóból álló félszótagra az 5. ábrán látható. Az 1. ábrán látható vezérlő szabja meg, hogy a félszótag típusától függően a félszótagszakaszok mely részeiből és hány darab beszédszeletet küld az adott háló bemenetére az akusztikai

	csend	zöngé	-burst	+burst	-spir	+spir	nali	MeMagh	MaMagh
csend	34	0	9	0	0	0	2	0	0
zöngé	0	16	0	6	2	0	0	0	0
-burst	0	0	17	0	2	1	0	0	1
+burst	0	2	0	2	1	0	0	0	0
-spir	0	0	0	0	21	0	0	0	3
+spir	0	0	1	1	1	2	3	0	0
nali	1	3	0	4	1	0	49	1	23
MeMagh	0	0	0	0	0	0	1	21	0
MaMagh	0	4	1	0	5	0	24	0	170

Összes szegmens száma: 435
A helyes ítéletek száma: 332
Százalékban: 76.321839 %

3. ábra
Egy ismeretlen mondat esetében a STEP1 neurális háló
kimenetén kapott tévesztési mátrix



4. ábra

A STEP1 neurális háló 9 kimenetén kapott aktivitás az idő függvényében

elemzőből.

Mivel az alsóbb szinteken a tévesztés természetes, ezt javítandó, egy adott félszótagszakaszt egyszerre több neurális hálóba küldünk el (a szakasz határai is változhatnak). Ez annyit jelent, hogy egyszerre több háló kimenetén kaphatunk eredményt. A kimeneti aktivitás függvényében ezek sorrendbe rakhatók, kiválasztható az első-, másod- és harmadrendű jelölt. Tehát a kimeneten a jelöltekből álló félszótagláncolatot kapunk, amelyre a nyers felismerési biztonság egyelőre 72%. Itt még a félszótagok sincsenek egymáshoz illesztve.

4. Eredmények, kitekintés

Célunk most a teljes felismerési rendszer, mint működő komplex egész bemutatása volt. Nyomdai kapacitási korlátok miatt a részletekről itt beszámolni nem tudunk, de a részeredmények a laboratóriumban rendelkezésre állnak. Ilyenek pl. a statisztikai vizsgálatok eredményei, az összeállított szövegminták, a felismerési biztonság változása a neurális háló szerkezetének változása függvényében stb.

A második részben számolunk be a betanítási eljárásról, az egyes szintek működésének részletes leírásáról, az egyes szinteken kapott részfelismerési eredmények folyamatos javításáról, valamint arról, hogy miként optimalizáltuk a neurális hálók szerkezetét és a betanítási módszert.

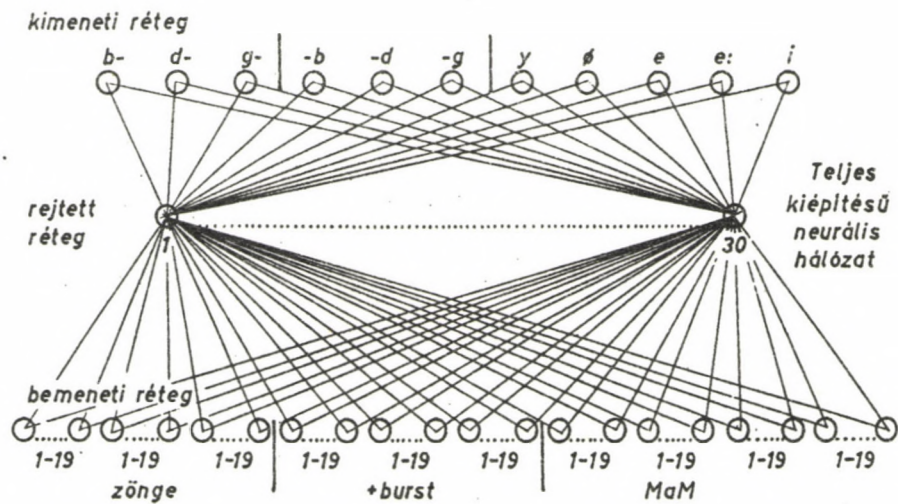
Az egyes szintek tökéletesítésére, a végleges felismerési biztonság növelésére a vizsgálatok most is folynak. Tervezzük a további nyelvészeti szintek bevonását a rendszerbe, hiszen csak így érhetünk el a felhasználó számára is elfogadható felismerési biztonságot. Tervezzük továbbá a felismerő beszélőadaptív változatának a kidolgozását is.

3. táblázat: Kezdő félszótagok neurális hálózatai

Sorszám	Félszótag típus		Bemeneti réteg elemszáma	Rejtett réteg elemszáma	Kimeneti réteg elemszáma	
K0.		MaM	4 x 19	30	5	
K1.		MéM	4 x 19	30	4	
K2.	Nali	MaM	7 x 19	30	11	
K3.	Nali	MéM	7 x 19	30	10	
K4.	-Spir	MaM	7 x 19	30	9	
K5.	-Spir	MéM	7 x 19	30	8	
K6.	+Spir	MaM	7 x 19	30	8	
K7.	+Spir	MéM	7 x 19	30	7	
K8.	Csend	-Spir	MaM	10 x 19	30	11
K9.	Csend	-Spir	MéM	10 x 19	30	10
K10.	Zöngé	+Spir	MaM	10 x 19	30	11
K11.	Zöngé	+Spir	MéM	10 x 19	30	10
K12.	Zöngé	+Burst	MaM	10 x 19	30	11
K13.	Zöngé	+Burst	MéM	10 x 19	30	10
K14.	Csend	-Burst	MaM	10 x 19	30	11
K15.	Csend	-Burst	MéM	10 x 19	30	10

4. táblázat: Záró félszótagok neurális hálózatai

Sorszám	Félszótag típus			Bemeneti réteg elemszáma	Rejtett réteg elemszáma	Kimeneti réteg elemszáma	
V0.	MaM			4 x 19	30	5	
V1.	MéM			4 x 19	30	4	
V2.	MaM	Nali		7 x 19	30	11	
V3.	MéM	Nali		7 x 19	30	10	
V4.	MaM	-Spir		7 x 19	30	9	
V5.	MéM	-Spir		7 x 19	30	8	
V6.	MaM	+Spir		7 x 19	30	8	
V7.	MéM	+Spir		7 x 19	30	7	
V8.	MaM	Csend	-Burst	10 x 19	30	11	
V9.	MéM	Csend	-Burst	10 x 19	30	10	
V10.	MaM	Zöngé	+Burst	10 x 19	30	11	
V11.	MéM	Zöngé	+Burst	10 x 19	30	10	
V12.	MaM	Zöngé	+Spir	10 x 19	30	11	
V13.	MéM	Zöngé	+Spir	10 x 19	30	10	
V14.	MaM	Csend	-Spir	10 x 19	30	11	
V15.	MéM	Csend	-Spir	10 x 19	30	10	
V16.	MaM	Nali	Csend	-Burst	13 x 19	30	17
V17.	MéM	Nali	Csend	-Burst	13 x 19	30	16
V18.	MaM	Spir	Csend	-Burst	13 x 19	30	15
V19.	MéM	-Spir	Csend	-Burst	13 x 19	30	14



5. ábra
Az egyik (K12) neur. hálózat szerkezete a STEP 2-ben

Irodalom

Baker, J.M.: Dictation, directories, and data bases; Emerging PC Applications for Large Vocabulary Speech Recognition. Eurospeech '93 Berlin, 1993, 3-10.

COCOSDA Proceedings of the International Coordinating Committee on Speech Databases and Speech I/O Systems Assesment Session IV. Banf, 1992.

Klimasauskas, C. : Neural Computing, NeuralWare, Inc. 1989. Nc-111.

Pollak, I.- Pickett, J.M.: The intelligibility of excerpts from conversation. Language and Speech 6. 1963, 165 - 171.

Takács, Gy.: Acoustic-phonetic recognition of continuous speech by artificial neural networks. Journal of Communications XLIII, 1992, 9-19.

Vicsi, K. - Mattila, M. - Berényi, P.: Continuous speech recognition using different methods. Acustica 71. 1990, 152.

Zwicker, E. : Psychoakustik. Springer Verlag, Berlin, 1982.