

A diplomamunkát ellenőriztem:

beadható

–

nem adható be

dátum

belső konzulens aláírása

Bíráló adatai:
.....
.....

A diplomamunka minősítése

A bíráló javaslata:

érdemjegy

dátum

bíráló aláírása

A Tanszék javaslata:

érdemjegy

dátum

tanszékvezető aláírása

Az ÁVB határozata:

érdemjegy

dátum

ÁVB elnök aláírása

SEVINGER CSABA
BESZÉDJELEK ELŐÁLLÍTÁSA
ÉRTHETŐSÉGI VIZSGÁLATOK SZÁMÁRA
ÖSSZEFOGLALÁS

A szakdolgozat fő célja annak vizsgálata, milyen módszerekkel lehet javítani a beszéd érthetőségét beszédátviteli rendszerekben. Ennek érdekében egy beszédérthetőségi teszt lett kidolgozva és elvégezve.

Az első fejezet a beszéd és a fonetika elméleti alapjainak áttekintéséről szól. Bemutásra kerülnek a beszédképzés és beszédértés élettani alapjai, a beszéd szegmentális, akusztikai, fizikai szerkezete, a koartikulációs hatások, a beszédértés és beszédészlelés folyamatának működése és a beszédadatbázisok létrehozásához, feldolgozásához szükséges ismeretek. A második részben szerepel a feladat kidolgozásának illetve a beszédértés vizsgálatának részletes leírása. Az utolsó fejezet témája a kapott eredmények kiértékelése.

CSABA SEVINGER

**SYNTHESIS OF SPEECH SIGNALS FOR
SPEECH RECOGNITION TESTS**

SUMMARY

The base objective of the thesis is to examine what methods can help the recognition of speech in speech-transmitting systems. In order to this, a speech-recognition test had been developed and executed.

The first chapter details the theoretical basics of speech. It introduces the biological basics of the production and recognition of speech, the segmental, the acoustic and the physical construction of speech, the coarticulation effects, the working of the production and recognition of speech and the necessary knowledge to make and prepare a speech-database. The second chapter shows the development of the exercise, and the examination of the speech-recognition. The third chapter introduces the evaluation of the achievements.

TARTALOMJEGYZÉK

1. BEVEZETÉS	3
2. A BESZÉD ELMÉLETI ALAPJAI.....	4
2.1. A beszéd definíciója	4
2.2. A beszéd élettani alapjai.....	5
2.2.1. A beszédképzés szervei és működésük	5
2.2.2. A hallás szervei és működésük	7
2.3. A beszéd szegmentális szerkezete	8
2.3.1. A fonetikus átírás.....	8
2.3.2. A magyar magánhangzók osztályozása	10
2.3.3. A magyar mássalhangzók osztályozása	10
2.4. A beszédhangok akusztikai szerkezete.....	12
2.4.1. A beszéd mint akusztikai hullám.....	12
2.4.2. A beszédhangok időszerkezete	13
2.4.3. A magánhangzók időtartama	13
2.4.4. A mássalhangzók időtartama	14
2.4.5. Szótagok és szavak időtartama	15
2.4.6. A beszédhangok frekvenciaszerkezete	16
2.4.7. Az átlagos beszédszínekép	16
2.4.8. A magánhangzók frekvenciaszerkezete.....	17
2.4.9. A mássalhangzók frekvenciaszerkezete	19
2.4.10. A beszédhangok specifikus intenzitása	20
2.5. A koartikuláció	21
2.5.1. Koartikulációs hangátmenetek.....	23
2.5.2. Módosult artikulációs konfiguráció.....	24
2.5.3. A beszédhang megváltozása koartikuláció hatására	25
2.6. A szegmentális hangszerkezet észlelése.....	26
2.6.1. A beszédmegértés folyamata	26
2.6.2. A beszédészlelés folyamata	27
2.6.3. A szegmentálás	29
2.6.4. Invariancia, redundancia, irrelevancia	29
2.6.5. Normalizációs folyamatok.....	29
2.6.6. A beszédhangok észlelése.....	30
2.6.7. A magánhangzók észlelése	30
2.6.8. A mássalhangzók észlelése.....	31
2.6.9. Hangsorok észlelése.....	33
2.7. Adatbázisok.....	34
2.7.1. Adatbázisok felosztása	34

TARTALOMJEGYZÉK

2.7.2. Adatbázisok tervezése	35
2.7.3. Adatbázisok nyelvi feldolgozása	36
2.7.4. Szegmentálás	37
2.7.5. Adatbázisok a gyakorlatban.....	37
2.7.6. Magyar nyelvű adatbázisok	40
3. A BESZÉDÉRTÉS VIZSGÁLATA.....	41
3.1. A spondeuszok	41
3.2. A hamis-spondeuszok	42
3.3. A hangfelvételek elkészítése és előfeldolgozása	44
3.4. A szegmentálás.....	45
3.5. A kivágás.....	46
3.6. A feltöltés	51
3.7. A szubjektív lehallgatás	56
4. KIÉRTÉKELÉS	57
4.1. A spondeuszok észlelése	57
4.2. A hamis-spondeuszok észlelése	67
5. ÖSSZEFOGLALÁS	77
6. IRODALOMJEGYZÉK	78
7. MELLÉKLET	79

1. BEVEZETÉS

A szakdolgozat célja annak vizsgálata, milyen módszerekkel lehet javítani a beszéd érthetőségét beszédátviteli rendszerekben. Az egyesült államokbeli EBIRE (East Bay Institute for Research and Education) kutatóintézetben végeztek hasonló vizsgálatokat angol nyelven, ennek mintájára lett kidolgozva a feladat magyar nyelvre. A lehetőségek szerint ugyanazokkal a paraméterekkel készült a magyar vizsgálat, mint az angol, hogy a későbbiekben össze lehessen hasonlítani a két kísérlet eredményeit.

Az emberi beszéd észlelése és megértése egy nagyon bonyolult és komplexen építkező folyamat. Jelentős részének működéséről gyakorlatilag nincsenek vagy alig vannak pontos ismereteink, sejtjük csak, milyen működések történnek az idegpályákon és az agyban. Nagyon sokféle kísérletet dolgoztak már ki, hogy feltérképezzék és megértsék ennek működését. A beszédpercepció vizsgálatára nincsenek külön műszerek, a folyamatok célzott kísérleti eljárásokkal ismerhetők meg.

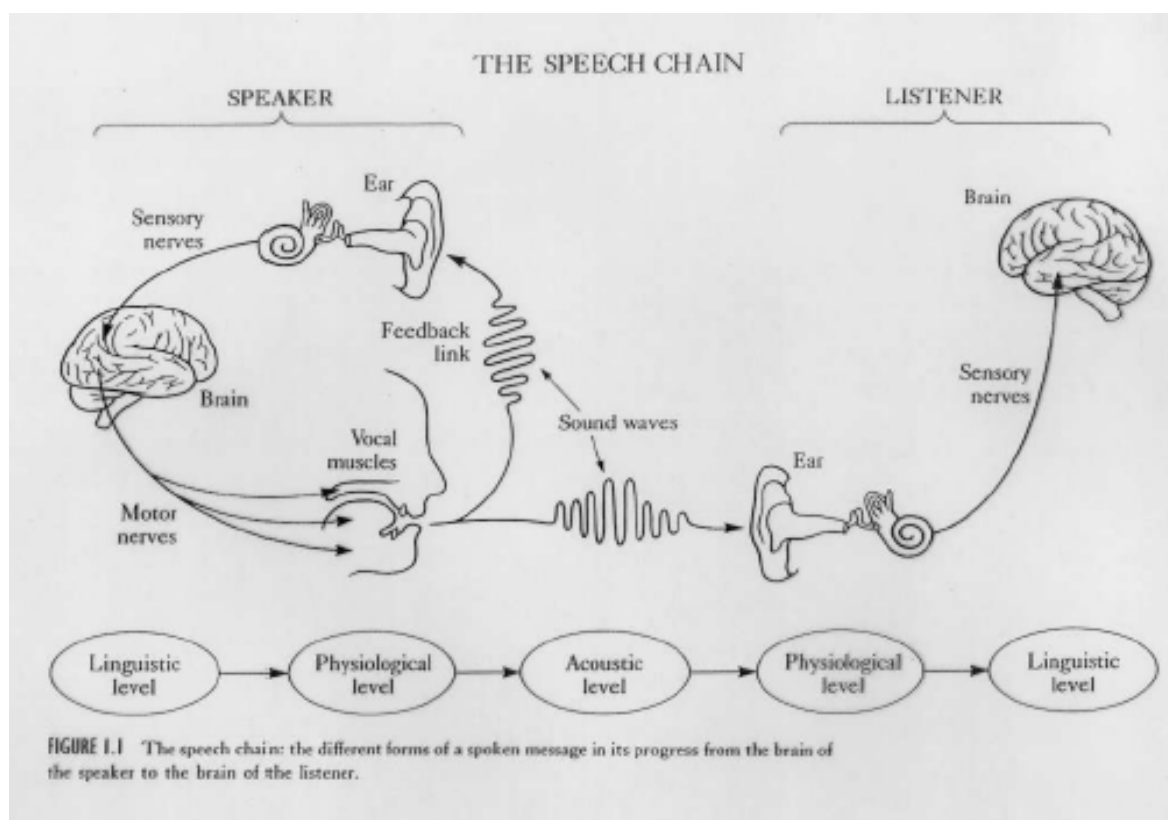
A beszéd érthetőségének vizsgálatához egy kisméretű, szavakból álló beszédadatbázist kellett létrehozni. A 40 értelmes és 40 értelemmel nem rendelkező szó kiválasztása után egy beszélő alkalmazásával elkészültek a hangfelvételek. Ezekből aztán bizonyos részek ki lettek vágva, ezzel modellezve az átviteli kieséseket. A keletkezett lyukak vagy üresen maradtak, vagy fel lettek töltve beszédzajjal vagy fűrészzellel. Ezeket kellett emberekkel meghallgattatni, majd a kapott eredményeket kiértékelni. A lehallgatási tesztekben 20 ember vett részt. Azt vártuk, hogy a beszédzajjal feltöltött lyukak elősegítik az érthetőséget és az információ helyreállítását.

A szakdolgozat első fejezete a beszéd és a fonetika elméleti alapjainak áttekintéséről szól. Bemutatásra kerülnek az élettani alapok, a beszéd szegmentális, akusztikai, fizikai szerkezete, a koartikulációs hatások, a beszédértés és beszédészlelés folyamatának működése és a beszédadatbázisok létrehozásához, feldolgozásához szükséges ismeretek. A második részben szerepel a feladat kidolgozásának illetve a beszédértés vizsgálatának részletes leírása. Az utolsó fejezet témája a kapott eredmények kiértékelése.

2. A BESZÉD ELMÉLETI ALAPJAI

2.1. A beszéd definíciója

Legtömörebb megfogalmazásban a beszéd a nyelvhasználat hangzó formája. Akusztikus gondolatátvitel, ahol az üzenet továbbítása nyelvi formában történik. Gondolataink megformálásának, továbbításának, valamint a gondolatok feldolgozásának eszköze. Két nagy része van: az egyik a beszédprodukció, a másik a beszédfeldolgozás. Ezek váltakozása a beszéd körfolyamata. A folyamat működését az **2.1. ábra** szemlélteti.



2.1. ábra A beszéd körfolyamata

Az agy a vezérlés központja mindkét esetben. Feladata többek között a beszédszervek működésének szabályozása, a folyamat végeredménye az artikuláció, az artikuláció következménye pedig egy akusztikai jel. Működik egy akusztikai visszacsatolás is, a beszélő saját beszédének a hallgatása, ami lehetővé teszi az állandó ellenőrzést, és az azonnali hi-

bajavítást, korrigálást például nyelvbottlások, vagy nem egyértelmű fogalmazás esetén. Az akusztikai jelet a levegő részecskéinek rezgése továbbítja, és a fül fogja fel, a benne lévő hallószervek ingerületté alakítják, és továbbítják az agy felé. Itt zajlik beszédészlelés és a beszédmegértés folyamata. Élő beszéd esetén a vizuális észlelésnek is szerepe van, a hallgató látja a beszélő artikulációjának látható mozgásait, és az így nyert információt is felhasználja a beszédértéshez. [1] [3]

2.2. A beszéd élettani alapjai

A beszédképző szervek elsősorban nem a kommunikációt szolgálják, hanem az élet, életben maradás célját; rágást, táplálkozást, légzést.

A beszédképző és feldolgozó szervek minden embernél azonosak. Méretben és anatómiai felépítésben eltérések vannak, de elsődleges funkciójukban azonosan működnek minden embernél. Hangképzéskor viszont elég nagy különbségek lehetnek az egyes népcsoportoknál, ezért van, hogy a különféle nyelvek hangzása nagyon különböző lehet. [1]

2.2.1. A beszédképzés szervei és működésük

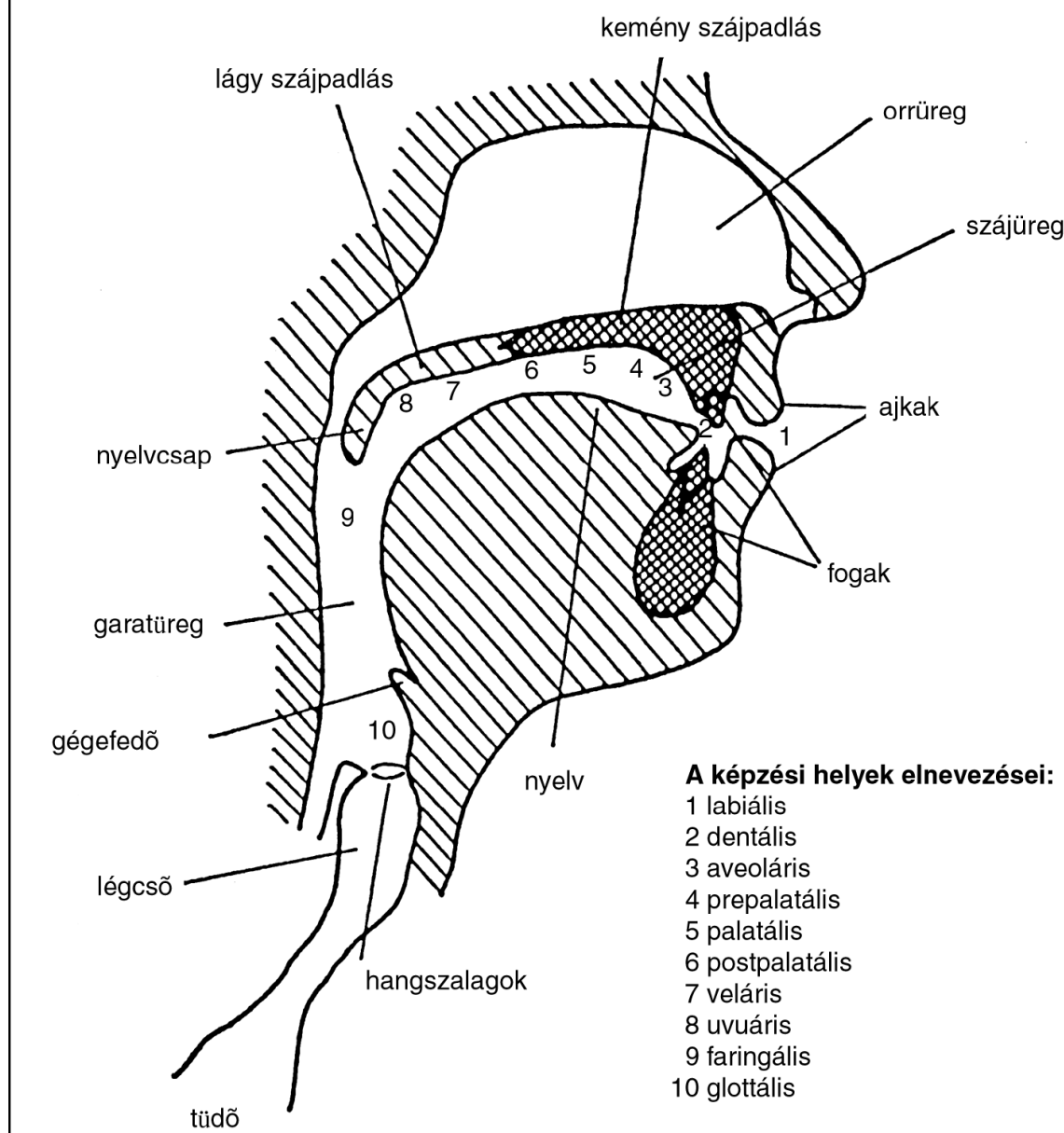
A tüdő elsődleges feladata a légzés, azaz az oxigén felvétele és a széndioxid eltávolítása. Beszéd közben a hangképzéshez szükséges levegő áramlását biztosítja.

A tüdő kivezető csöve a légcső, ennek felső részén helyezkedik el a porcok felépítésű gége. Itt találhatóak a hangszalagok és az ezeket beállító izmok. A hangszalagok a zöngékpzésért felelősek. Rendkívül mozgékonyak, hatféle állást lehet megkülönböztetni, amiket itt nem részletezünk.

A hangszalagoktól a szájnyílásig terjedő terület a toldalékcső. Három üreg alkotja: a garat, a szájüreg és az orrüreg. A hangszalagrezgés következtében kialakult hanghullám ezekbe az üregekbe kerül, amik rezonátorláncként működnek, mintegy rákényszerítik a saját rezgéseiket a gégehangra.

A szájüreg legfontosabb részei: nyelvcsap, nyelv, lágy szájpadlás, kemény szájpadlás, nyelv, fogak, ajkak. Ezek mozgatásával, helyzetük változtatásával lehet alakítani, „hangolni” a szájüreget, akadályt képezni a kiáramló levegőnek.

A hangképző szervek vázlatos képe



2.2. ábra. A beszédképzés szervei

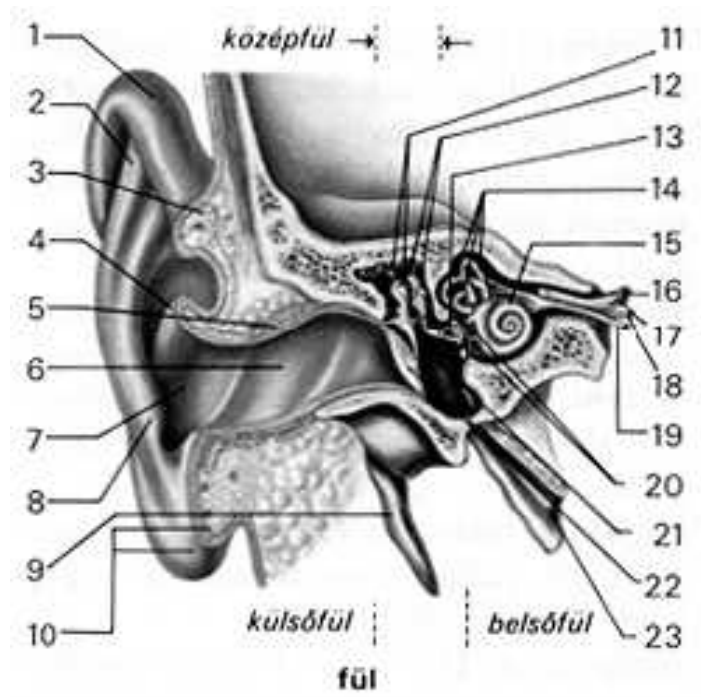
A beszédképző szervek működését egyrészt a központi idegrendszeri, másrészt perifériális idegrendszeri és reflextevékenységek koordinálják. A beszéd vezérlését az agy végzi, nagyobb részben a bal, kisebb részben a jobb féltéke. A beszédalkotás folyamán használt elemi folyamatokat artikulációs bázisnak nevezzük.

2.2.2. A hallás szervei és működésük

A perifériális hallószerv feladata az akusztikai jel felfogása és ingerületté alakítása. Felépítését **2.3. ábra** mutatja.

A külsőfület a fülkagyló és a hallójárat alkotja. A középfültől a dobhártya választja el, ami átalakítja a levegő részecskéinek rezgését a hallócsontok rezgésévé. A középfül részei a dobüreg, az Eustach-kürt.. A dobüregben találhatóak a hallócsontok: a kalapács, az üllő és a kengyel. Ezek továbbítják a rezgést a belső fülbe. Itt van a hallásért és az egyensúlyozásért felelős szerv is. Az előbbi nevezik csigának (Corti-szerv). Itt válik a mechanikai inger idegi ingerületté a folyadékban úszó szőrsejtekben.

Az ingerületeket idegrostok továbbítják a központi idegrendszerhez az agykéregbe. Ez a centrális hallás területe.



2.3. ábra. A hallószervek

1-fülszegély; 2-belsőszegély; 3-a fülkagyló rostos porca 4-fülcsap; 5-mirigyréteg; 6-külső hallójárat; 7-fülkagylóárok; 8-antitragus; 9-csontnyúlvány; 10-fülcimpa és metszete; 11-a kalapács és felső szalagja; 12-üllő és felső szalagja; 13-kengyel; 14-félkörös ívjáratok; 15-csiga (cochlea); 16-belső hallójárat; 17-arcideg; 18-vestibularis ideg; 19-cochlearis ideg; 20-kerek és ovális ablak; 21-dobüreg; 22-dobhártya; 23-fülkürt (Eustach-kürt);

2.3. A beszéd szegmentális szerkezete

A beszéd képzése két részre osztható, mindkettő létrehozása ugyanazokkal a beszéd-szervekkel történik. A szegmentális szerkezeten a beszédhangokat, hangkapcsolatokat és hangsorokat értjük, a szupraszegmentális hangszerkezeten az ezekkel egyidőben képzett beszéddallamot, hangsúlyt, tempót, szünetet, ritmust, hangerőt és hangszínezetet.

2.3.1. A fonetikus átírás

Az írás kialakulásakor a betű megegyezett a hanggal, de a beszéd dinamikusan változott, az írással ellentétben. Ezért a kiejtés és az írás egyre jobban eltávolodott. A jelenség erősen nyelvfüggő, a magyarban nem olyan jelentős, nem úgy mint például a franciában.

A beszélt nyelvben a fonéma a hangok olyan elemi, elvont egysége, amely szavakat különböztet meg egymástól. Más szóval a legkisebb nyelvi egység, melynek cseréjével a szavak értelme megváltozik. Önálló jelentéssel nem rendelkezik, hanem jelentésmegkülönböztető szerepe van. Az allofón a fonológiában azon beszédhangok egyike, melyek azonos fonéma alá tartoznak. [1] [3]

A fonetikus átírás célja, hogy a lehető legpontosabban, nyelvfüggetlenül, írásban rögzítse az elhangzó beszéd sajátosságait. A Nemzetközi Fonetikai Ábécét (International Phonetic Alphabet, IPA) a Nemzetközi Fonetikai Szövetség (International Phonetic Association) készítette, eredetileg angol és francia kezdeményezésre. Története során számos nagy módosításon esett át, és folyamatos fejlesztés alatt áll. Jelöléseit jelenleg főként a latin ábécé betűiből veszi, néhány jelet a görög ábécéből valamint néhány olyat is használ, melynek nincs kapcsolata már létező nyelvek írásképével. Erre azért van szükség, mert a világ nyelveinek hangkészlete annyira változatos, hogy a latin betűk száma nem elegendő a megjelenítésükre. [1] [3]

Az IPA jelölésrendszerével az a probléma, hogy nem illeszkedik a számítógép billentyűzetéhez, ezért bevezetésre került egy új jelölésrendszer SAMPA (Speech Assessment Methods Phonetic Alphabet) néven, ami az IPA jelkészletét a számítástechnikában legelterjedtebb ASCII-kódoknak felelteti meg. Így a gépelés és a különböző számítógépes platformok, programok közti átjárhatóság problémája megoldódott. [1] [3]

A SAMPA azonban nyelvfüggő jelölés, mivel eredetileg csak egyes nyelvekhez készült el, ezért nem alkalmas arra, hogy az IPA fonetikai jeleit egyértelműen jelölje. Ennek kiküszöbölésére jött létre az X-SAMPA, ami egy nyelvfüggetlen jelölésrendszer. A magyar nyelv beszédhangjait leíró IPA és X-SAMPA szimbólumokat a **2.3. táblázat** tartalmazza.

A dolgozatban a továbbiakban az X-SAMPA szimbólumai lesznek használva a beszédhangok jelölésére kettő „ / ” jel között feltüntetve. A fonémák lehetnek rövidek és hosszúak, a hosszúakat a szimbólum utáni „ : ” jelöli.

2.3. táblázat. IPA és X-SAMPA szimbólumok magyar nyelvhez

Betű	SAMPA	IPA	Példák
A	A	a	hat
Á	a:	a:	hát
B	b	b	bab
C	ts	ʈs	cél
CS	tS	tʃ	cső
D	d	d	dél
DZ	dz	ɖz	bodza
DZS	dZ	ɖʒ	dzsem
E	E	ɛ	vet
É	e:	e:	vét
F	f	f	fél
G	g	g	gép
GY	J\	ɟ	gyár
H	h	h	hét
I	i	i	hit
Í	i:	i:	szít
J	j	j	jön
K	k	k	kép
L	l	l	lép
M	m	m	méz
N	n	n	néz
NY	J	ɲ	nyom
O	o	o	sok
Ó	o:	o:	sók
Ö	2	ø	köt
Ő	2:	ø:	sőt

Betű	SAMPA	IPA	Példák
P	p	p	pad
R	r	r	réz
S	S	ʃ	só
SZ	s	s	szép
T	t	t	tél
TY	c	c	tyúk
U	u	u	fut
Ú	u:	u:	kút
Ü	y	y	süt
Ű	y:	y:	fűt
V	v	v	vér
Z	z	z	zaj
ZS	Z	ʒ	zsír

Allofónok			
Betű	SAMPA	IPA	Példák
j	(X	ç	kapj
m	F	ɱ	kámfor hamvas honvágy honfitárs
n	N	ɳ	ing tönk
h	X	x	doh
h	X	x	ihlet
h	h\	ɦ	lehet

2.3.2. A magyar magánhangzók osztályozása

A nyelv vízszintes mozgása szerint:

- Elöl képzettek (palatálisok): [E, e:, i, i:, 2, 2:, y, y:]
- Hátul képzettek (velárisok): [A, a:, o, o:, u, u:]

A nyelv függőleges mozgása szerint:

- Felső nyelvvállásúak: [i, i:, u, u:, y, y:]
- Középső nyelvvállásúak: [e:, o, o:, 2, 2:,]
- Alsó nyelvvállásúak: [E, A]
- Legalsó nyelvvállásúak: [a:]

Ajaktűköködés szerint:

- Ajakkerekítésesek (labiálisok): [A, o, o:, 2, 2:, u, u:, y, y:]
- Ajakrésesek (illabiálisok): [a:, i, i:, E, e:]

A nyelvi időtartam szerint:

- Rövidek: [A, E, i, o, 2, u, y]
- Hosszúak: [a:, e:, i:, o:, 2:, u:, y:]

2.4. táblázat. Magyar köznyelvi magánhangzók

A nyelv vízszintes→ függőleges↓ mozgása	Elöl képzettek		Hátul képzettek	
	ajak- résesek	ajakkerek- ítésesek	ajak- résesek	ajakkerek- ítésesek
felső nyelvvállásúak	i, i:	y, y:		u, u:
középső nyelvvállásúak	e:	2, 2:		o, o:
alsó nyelvvállásúak	E			
legalsó nyelvvállásúak			A, a:	

2.3.3. A magyar mássalhangzók osztályozása

Képzési mód szerint:

- Zárhangok:- Orális, felpattanó zárhangok: [b, p, d, t, J, c, g, k]
- Nazális zárhangok: [m, n, J, F, N]
- Réshangok (frikatívák): [v, f, z, s, Z, S, h, h\, (X, X)]
- Zár-rés hangok (affrikáták): [ts, tS, dz, dZ]

- Pergőhangok: [r]
- Közelítőhangok (likvidák): [j, l]

Képzési hely szerint:

- Két ajakkal képzettek (bilabiálisok): [b, p, m]
- Ajak-fog hangok (labiodentálisok): [v, f, F]
- Fogmederhangok (alveolárisok): [d, t, n, z, s, dz, ts, r, l]
- Hátsófogmederhangok (posztalveolárisok): [Z, S, dZ, tS]
- Elülsőszájpadlás-hangok (palatálisok): [J\, c, J, j, (X)]
- Hátsúlsőszájpadlás-hangok (velárisok): [g, k, N, X]
- Gégehangok (laringálisok): [h, h\]

Hangszalagműködés szerint:

- Zöngés: [b, d, J\, g, m, F, n, J, N, v, z, Z, dz, dZ, r, j, l, h\]
- Zöngétlen: [p, t, c, k, f, s, S, h, X, (X, ts, tS,]

Nyelvi időtartam szerint:

- Rövidek
- Hosszúak

2.5. táblázat. Magyar rövid mássalhangzók

Képzési mód→ hely↓	Zárhangok			Részhangok	
	felpattanó zárhangok		nazálisok		
	zöngés	zöngétlen	zöngés	zöngés	zöngétlen
bilabiális	b	p	m		
labiodentális			F	v	f
alveoláris	d	t	n	z	s
posztalveoláris				Z	S
palatális	J\	c	J		(X
veláris	g	k	N		X
laringális				h\	h
Képzési mód→ hely↓	Zár-rés hangok		Pergő- hangok	Közelítőhangok	
				mediális	laterális
	zöngés	zöngétlen	zöngés	zöngés	zöngés
alveoláris	dz	ts	r		l
posztalveoláris	dZ	tS			
palatális				j	

2.4. A beszédhangok akusztikai szerkezete

2.4.1. A beszéd mint akusztikai hullám

Az artikuláció következményeként jön létre a rezgés, amint a levegő az ajkakon át kikerül a levegőbe. A beszéd mint rezgés (vagy mint hullám) ugyanazokkal az akusztikai paraméterekkel jellemezhető, mint bármilyen más akusztikai jel. A beszéd tehát ekkor fizikai jelenség, amely meghatározott időtartamban zajlik, meghatározott frekvenciaszerkezete és intenzitás szerkezete van. A rezgéseknek két csoportjuk van, az egyszerű és az összetett hangok. A hangok továbbá lehetnek periodikusak és aperiodikusak. Az összetett hangok felbonthatók tiszta hangokra. Egy teljes rezgésminta idejét nevezzük periódusnak. A rezgések frekvenciája az időegységre eső periódusok száma, ez adja a hang magasságát, amit Hertzben fejezünk ki. Alaphangnak a komplex hang legkisebb frekvenciájú összetevőjét nevezzük, az ennél magasabb frekvenciájúak, amelyek az alaphang legkisebb egész számú többszörösei, a felhangok vagy felharmonikusok. [1]

A beszédhangok fizikai sajátosságai az artikulációjuk következménye. Ez azt jelenti, hogy ha az artikuláció változik, akkor annak akusztikai következménye is változik. A beszédhullám forrása a hangszalagok rezgése, valamint a szájüreg különböző helyein keletkező zörejek. A beszéd frekvencia-, intenzitás- és időszerkezeti sajátosságai a hangszínképpel (spektrum) jól szemléltethetők. A hangszalagrezgés kváziperiodikus, vagyis csaknem teljesen periodikus, a zöngé tehát periodikus összetett hang. A magánhangzók létrejötte akusztikailag úgy történik, hogy a zöngé áthalad az üregrendszeren, amely rezonátorként viselkedik, és a zöngé frekvenciájából az üreg a sajátfrekvenciájának megfelelő vagy ahhoz közel eső felhangokat (vagy felharmonikusokat) felerősíti. Az így létrejött energiakonzentrációt nevezzük formánsnak. A magánhangzók és a zöngés mássalhangzók formánsokkal jellemezhetők. A formánsok sorszáma arra utal, hogy a zöngéhez (F_0) képest milyen távolságra helyezkednek el (F_1 , F_2 stb.) A zöngémássalhangzók szerkezetét az artikulációnak megfelelő akusztikai következmény jellemzi (zárfelpattanás, turbulens zörej). [1]

A beszédet mint akusztikai hullámformát ugyanazon három paraméterrel jellemezzük, mint a többi akusztikai rezgéssorozatot: az időtartammal, a frekvenciával és az intenzitással. [1]

2.4.2. A beszédhangok időszerkezete

A beszéd egy meghatározott időtartamban valósul meg, amelyen belül bármely részletet vagy szakaszt saját időviszonyok jellemeznek. Minden artikulációs mozgás meghatározott időtartamban zajlik, amit univerzális tényezők és nyelvspecifikus tényezők egyaránt meghatároznak. Az időszerkezet azt a struktúrát jelenti, amelyben egy adott beszédesemény megvalósul. Ez létrejöhet szegmentális és szuprasegmentális szinten egyaránt. Szegmentális beszédszerkezetben leggyakoribb a beszédhangok időtartamának vizsgálata, szuprasegmentális szinten pedig a beszédtempó elemzése. Az összetett beszédhangok egyes részeinek időtartama külön-külön is meghatározható. Ilyen például a felpattanó zárhangok zárszakasza és a felpattanás szakasza. Ekkor belső időtartamról beszélünk. [1]

A magánhangzók és mássalhangzók időtartamát számos tényező befolyásolja: bizonyos nyelvspecifikus fonetikai szabályok, a beszédhang minősége, a hangsor terjedelme, a hangkörnyezet, a beszédhang helye a hangsorban, illetőleg bizonyos szuprasegmentális tényezők, mindenekelőtt az aktuális beszédtempó, a hangsúly, a hangerő, a hanglejtés. [1]

A beszédhangok időtartamát nem lehet pontosan meghatározni, mert közöttük hangátmenetek vannak az artikulációs szervek tehetetlensége miatt. Magánhangzók esetében beszélhetünk tiszta időtartamról és átmeneti részről. Kísérleti eredmények szerint a hangátmenetek magyar nyelvben CV- (magánhangzó-mássalhangzó) kapcsolatoknál 20 ms-tól 100 ms-ig szórnak, ezek szintén nyelvfüggőek. A beszédhangok időtartamát általában milliszekundumban adják meg. [1]

2.4.3. A magánhangzók időtartama

A magyarban a magánhangzók időtartama nyelvileg kétféle lehet: rövid és hosszú. A rövidnek időtartama 30-150 ms-ig terjed, a hosszúaké 80- 280 ms-ig. Ez átfedést mutat, amelyet a már említett tényezők, a magánhangzó típusa, a kontextus és egyebek is meghatároznak. A magánhangzók határértékeit hangsúlyos és hangsúlytalan esetben a **2.6. táblázat** mutatja. [1]

A magánhangzó időtartamára hat a képzés konfigurációja. A nyelvemelkedés fokával fordítottan arányos az időtartam. A labiális magánhangzók hosszabbak, mint az illabiálisok, a veláris magánhangzók hosszabbak, mint a palatálisok. [1]

2.6. táblázat Magyar magánhangzók időtartamai

Magán- hangzó	Magánhangzók időtartamának szórása (ms)	
	hangsúlyos	hangsúlytalan
E	78-220	78-227
e:	115-323	74-298
i	59-110	66-196
i:	90-227	105-365
2	77-174	84-255
2:	141-322	133-306
y	55-251	60-180
y:	110-341	140-369
A	82-149	62-231
a:	149-290	105-282
o	68-145	61-180
o:	128-249	117-306
u	65-188	62-243
u:	98-369	110-314

2.4.4. A mássalhangzók időtartama

A mássalhangzók időtartama nyelvileg ugyancsak kétféle lehet, rövid és hosszú. Fizikai megvalósításukat (hasonlóan a magánhangzókéhoz) számos tényező befolyásolja: a mássalhangzó akusztikai alkata, a hangsorban elfoglalt helye, a hangkörnyezet, a hangsor terjedelme, a szupraszegmentumok. A rövidek és hosszúak között itt is átfedések tapasztalhatók. Egy adatközlő ejtését tekintve például a rövidek időtartama 30-270 ms, a hosszúaké 100-400 ms. [1]

A rövid mássalhangzók időtartama a képzési sajátosságaik függvényében a következő (legrövidebbektől a leghosszabbakig): approximánsok, nazálisok, zöngés zárhangok, zöngés réshangok, zöngétlen zárhangok, zöngés affrikáták, zöngétlen réshangok, zöngétlen affrikáták. A palatálisok hosszabbak, mint a velárisok. Minél hosszabb egy magánhangzó időtartama, annál rövidebb a szomszédos rövid mássalhangzó időtartama. Valamennyi mássalhangzó hosszabb a hangsor elején, mint a belsejében. [1]

2.7. táblázat Magyar mássalhangzók időtartamai

Mássalhangzó	Rövidek szórása (ms)		Hosszúak szórása (ms)	
	zöngés	zöngétlen	zöngés	zöngétlen
b, p	60-196	94-196	128-275	181-400
d, t	72-149	90-220	157-330	172-322
c, ʃ	56-282	83-298	227-365	227-381
g, k	27-196	83-220	119-298	171-318
v, f	47-188	78-282	176-192	172-369
z, s	68-227	103-326	114-282	211-357
Z, S	60-267	74-275	180-214	187-400
X		98-165		160-198
h		50-157		178-200
dz, ts	168-337		275	
dZ, tS	133-416		247	
m	44-145		157-251	
F	40-128			
n	36-130		130-279	
J	60-196		121-224	
N	42-142			
l	37-149		94-204	
j	55-215		92-198	
r	31-148		99-196	

2.4.5. Szótagok és szavak időtartama

A szótagok időtartama a szerkezetüktől, illetőleg az őket alkotó beszédhangoktól függ. Szavak szótagjainak az időtartamait szerkezetüktől függően a **2.8. táblázat** szemlélteti.

A szavak időtartama csakúgy számos tényező függvénye, mint a beszédhangoké. Izolált ejtésben ugyanazon szavak időtartama nem mutat jelentős eltérést több beszélő esetén sem, spontán beszédben azonban nagy változatosságot mutat. [1]

2.8. táblázat Magyar szótagok időtartamai (C: mássalhangzó, V: magánhangzó)

Szótagtípus	átlag (ms)	szórás (ms)
CV	194,2	150-320
VC	165,2	130-250
CVC	218,5	180-280

2.4.6. A beszédhangok frekvenciaszerkezete

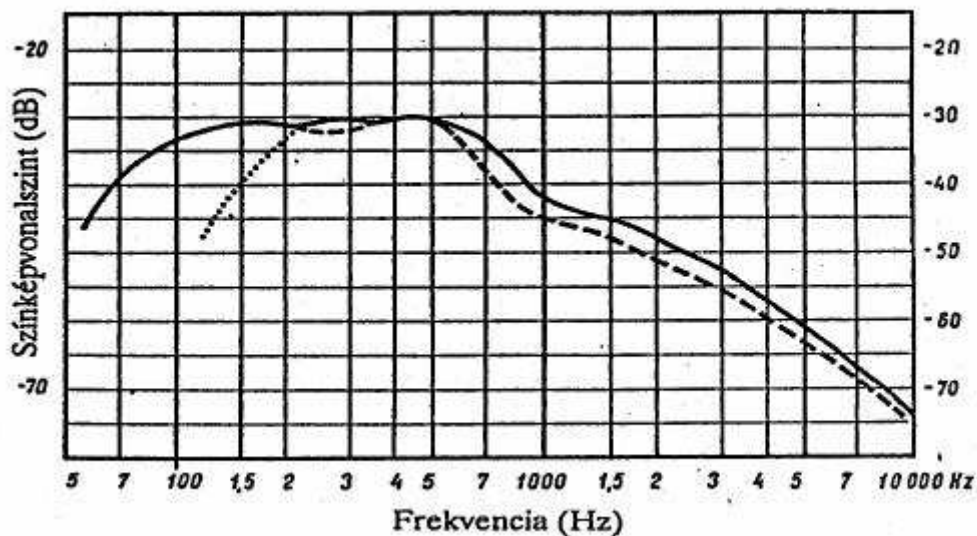
A beszédhangok összetett hangok, amelyek az időben változnak és széles frekvenciatartománnyal jellemezhetők. Ebben a frekvenciatartományban található az alaphang, a felhangok, a formánsok, a zörejösszetevők, tehát mindazon komponensek, amelyek a beszédhangok frekvenciaszerkezetét létrehozzák. Az emberi hallás számára a beszédhangok 70-10000 Hz-es tartománya hasznos, ami azt jelenti, hogy az ebben a frekvenciatartományban megjelenő összetevők akusztikailag egyértelműen jellemzik a beszédhangokat. A 10 kHz feletti részeket általában nem hasznosítjuk a verbális kommunikáció során, ezek az egyénre jellemző részek. A beszéd energiájának 90 százaléka 4000 Hz alatt van, a legnagyobb teljesítményt az 500 Hz körüli összetevők hordozzák. Ha a beszélő megemeli a hangját, hangsabban beszél, általában a frekvenciaösszetétel is megváltozik. A nagyobb frekvenciájú hangok nagyobb mértékben erősödnek fel, mint a kis frekvenciájúak. [1] [3]

Az alaphang, ami a folyamatos beszéd átlagos alaphangfrekvenciáját jelenti, férfiaknál 100-200 Hz, nőknél 150-300 Hz.

2.4.7. Az átlagos beszédszínekép

A beszéd pillanatról pillanatra változó alaphangfrekvenciájú, intenzitású és színképi eloszlású hangjelek egymásutánja. A hangközvetítés, a beszédérthetőség, a beszéd akusztikai jellegének megállapítása céljából szükséges megismernünk a beszéd egészére jellemző színképi eloszlást. A beszédszínekép a beszédben előforduló hangok átlagos színképi eloszlást tünteti fel az amplitúdó és a frekvencia függvényében, más szavakkal a frekvenciaösszetevők intenzitásintjeinek összessége. [3]

A beszédszínekép megállapításához „beszédzörejt” teremtünk, amely nem más, mint sok beszélő beszédéből kapott, homogén eloszlású, időben állandó színképű zörej. A legtöbb esetben 10 személy olvas fel folyamatos szöveget 10 mikrofonba. A mikrofonok áramát felerősítve, egyetlen szalagra veszik fel a hangkeveréket. A szalagot visszajátszva egyetlen beszédzajt lehet észlelni, amelyet elemzési módszerekkel vizsgálnak. A magyar hangképzésből eredő átlagos beszédszíneképet Tarnóczy vizsgálatai alapján ismertetjük. A 62 személy beszédéről készített átlagos színképet **2.4. ábra** mutatja be. [3]



2.4. ábra. 62 magyar beszélő átlagos beszédshínképe

A folytonos vonal a férfiak és nők burkológörbáját jelzi. Ahol a női beszédshínkép ez alatt marad, pontozott vonal jelzi az eltérést, ahol pedig a férfi beszédshínkép marad alatt, szaggatott vonal.

A férfiak beszédében három amplitúdócsúcs emelkedik ki az első 100 Hz-en, a második 260-270 Hz-en, a harmadik 510 Hz-en jelentkezik. A nők beszédében két energia csúcsot vehetünk észre: az első 200 Hz-en, a második 510 Hz-en jelentkezik. [3]

2.4.8. A magánhangzók frekvenciaszerkezete

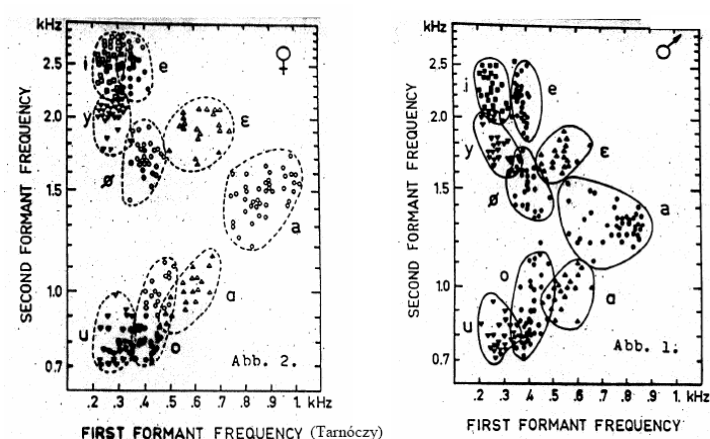
A magánhangzók formánsszerkezetét általában az első három formáns frekvenciaértékével adják meg. A köznyelvi magánhangzók formánsértékei igen különbözőek a nagyfokú variáltság miatt. A 2.9. táblázatban szereplő értékek csak a szemléltetést szolgálják, nem szerepelnek határértékeként. A formánsok az egyéni artikuláció, a fonetikai kontextus és pozíció, avagy a hangsúlyviszonyok következtében is nagyon változatosak. Az értékek egy személy ejtésén belül is jelentős szóródást mutatnak. A formánsok kimutatására többféle elemzési eljárás is rendelkezésre áll, általában FFT –vel (gyors Fourier-analízis) vagy LPC (lineáris predikciós elemzés) –vel vizsgálják. [1]

A magánhangzók kváziperiodikus hangok, a periódusoknak a hangshínképen függőleges vonalak felelnek meg.

2.9. táblázat A köznyelvi magyar magánhangzók formánsainak tájékoztató adatai

Magánhangzó	Első formáns (F1) (Hz)	Második formáns (F2) (Hz)	Harmadik formáns (F3) (Hz)
A	600	1100	2400
a:	800	1400	2500
o	450	900	2300
o:	400	950	2300
u	300	650	2300
u:	280	600	2100
2	380	1600	2400
2:00	400	1550	2500
y	280	1750	2200
y:	260	1800	2150
E	550	1900	2500
e:	420	2100	2800
i	280	2400	3000
i:	260	2500	2850

Az első és a második formáns a magyarban rendszerint egyértelműen meghatározza a magánhangzó minőségét. A harmadik a beszélő személy egyéni ejtésére vonatkozóan is adhat valamelyes információt, általában csak másodlagosan jellemzi a magánhangzót. Rendszerint az első formáns a legintenzívebb, a továbbiak értéke csökken. Az első és a második formáns viszonyáról szemléletesebb képet kapunk, ha egymáshoz képest ábrázoljuk őket. Egy ilyen grafikon látható a **2.5. ábrán**. A magánhangzók akusztikai szerkezetét a formánsok sávszélességével is jellemezhetjük. [1]



2.5. ábra. Magyar női és férfi magánhangzók F1-F2 grafikonja

2.4.9. A mássalhangzók frekvenciaszerkezete

A mássalhangzókat akusztikailag a gerjesztés és a keletkezett akusztikai szerkezet szerint kategorizáljuk. Egy részük kváziperiodikus, zöngés hang, más részük zöngés-zörejes, azaz kevert típusú mássalhangzó. A zöngés zörejmássalhangzók zöngésségének erőssége és tartama változó lehet. [1]

Vannak olyan mássalhangzók is, amelyek teljesen zörejesek, semmiféle periodikus rezgés nem jellemzi őket, ezek a zöngétlen mássalhangzók, obsztruensek. Képzésükkor a vokális traktuson áthaladó levegő útja vagy tökéletesen el van zárva (zárhangok), vagy egy szűkület képezi az akadályt (réshangok), avagy az előbbi kettő kombinációja következik be (zár-rés hangok). A zöngétlen mássalhangzók hangszínképén nincsenek a periodikus rezgést jelző függőleges vonalak. [1]

A hangképzés tekintetében igen különféle felpattanó zárhangok léteznek a világ nyelveiben. A nyelvi sajátosságoktól függően különféle és többszörös akusztikai kulcsok definiálják őket. A zárfelpattanási zörej utáni szakasz a magánhangzóhoz vezető átmenet (ezt nevezik zöngelkedési időnek, rövidítése VOT). [1]

A résmássalhangzókra jellemző turbulens zörej a képzési hely függvényében akusztikailag más és más frekvenciatartományban jelentkezik. A zöngétlenek zörejtartalma nagy, s mintegy 1000-8000 Hz-ig tartalmaznak komponenseket. A laringális (vagy glottális) réshangra általában az alsóbb frekvenciatartományban megjelenő, nem túl nagy intenzitású zörejkomponensek a jellemzők. [1]

A zár-rés hangoknál a teljes zár akusztikai lenyomata az esetek túlnyomó többségében megfigyelhető, a zár- és a résmozzanat az akusztikai szerkezetükben jól elkülöníthető. [1]

A nazális mássalhangzók esetében nincs átmenet a szomszédos magánhangzókhoz, ami artikulációs sajátosságaik következménye. Zöngés gerjesztésűek, egyszerű a szerkezetük, vagyis összetevőik lényegében nem változnak a képzés teljes tartamában. [1]

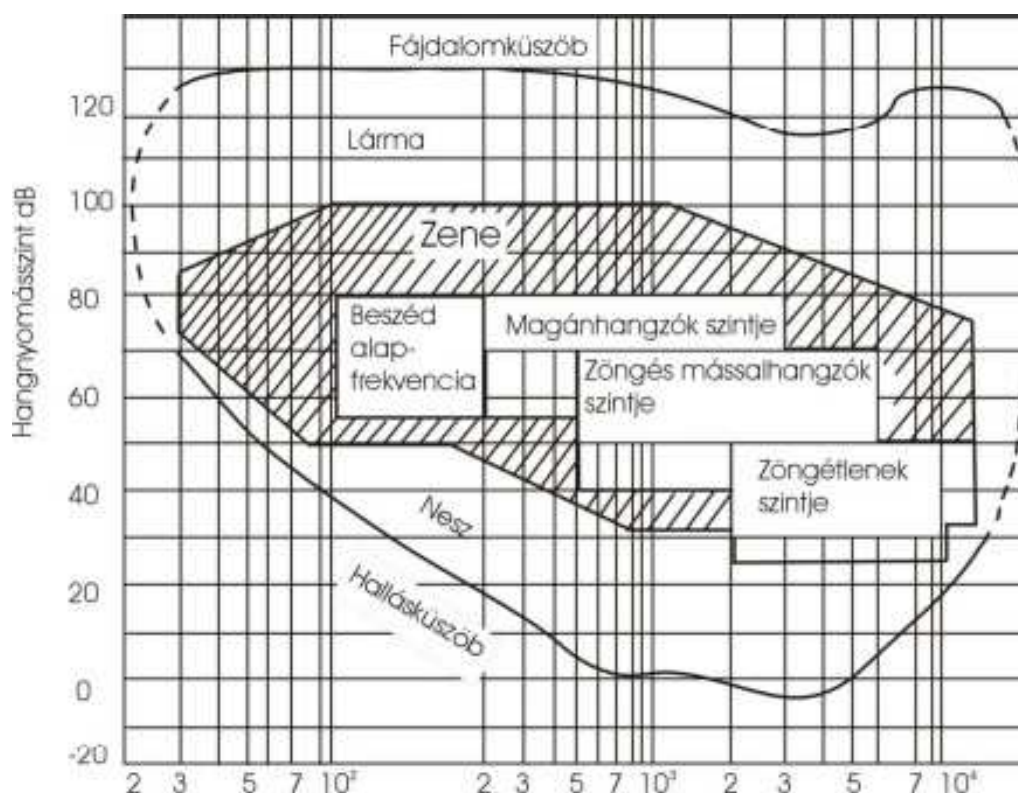
A pergőhang jellegzetessége, hogy pergetéskor a mechanikus oszcilláció mintegy 20-35 Hz, valamint hogy a harmadik formáns többnyire alacsony frekvenciájú. A perdületek száma jól látható a hangszínképen. [1]

A közelítőhangok akusztikai szerkezete igen hasonló a magánhangzókéhoz, néhány esetben például nezenen különíthető el az /i/ magánhangzó és a /j/ mássalhangzó. [1]

2.4.10. A beszédhangok specifikus intenzitása

A hang fizikai erősségét meghatározott időablakra mért intenzitásszint effektív értékével [dB] jellemezzük. A suttogott beszéd intenzitásszintje: 30-40 dB, a kiabálásé 80-90 dB. A különböző beszédhangok specifikus intenzitásának a meghatározása meglehetősen nehéz, mivel a különböző kontextusokban eltérő értékekkel fordulhatnak elő. A szakirodalomban a legintenzívebbnek tekintett hang az /a:/. [1]

A magánhangzókéhoz viszonyítva rendszerint kisebb energiájúak a közelítőhangok és a nazálisok, valamint a pergőhang. Az intenzitás szórásértékeinek alapján megállapítható, hogy az alveoláris és a posztalveoláris zöngétlen réshangok azonban akár a magánhangzókéhoz hasonló intenzitásúak is lehetnek. A legkisebb energiájúak általában a zöngés alveoláris és a posztalveoláris réshangok, a /v/, az affrikáták és a zárhangok, végül a /f/ és a /h/. Mindenképpen megjegyzendő azonban, hogy ezek a megállapítások a mindenkori kontextus és fonetikai pozíció, valamint a szuprasegmentális szerkezet függvényében változhatnak. [1]



2.6. ábra. A beszéd intenzitása és elhelyezkedése a hallástartományban

2.5. A koartikuláció

A koartikuláció a folyamatos beszéd artikulációs jelensége, a beszédprodukciónak kontrollja, a nyelv fonológiai rendszere és annak realizációja közötti folyamatot jelenti. Más szavakkal azt, mondhatjuk, hogy ugyanazon fonéma nem pontosan ugyanolyan artikulációs gesztusokkal vagy ugyanabban a beszédhangban realizálódik minden hangkörnyezetben, hanem különféleképpen változik azért, hogy a – lehetőségek szerint – minél hasonlóbbá váljon a szomszédos vagy közeli beszédhangokhoz. Ez részben az artikuláció gazdaságosságának, részben a nyelvspecifikus (illetve nyelvjárási) jellemzőknek a következménye, sőt nemegyszer egyéni sajátosságok érvényesülése. [1]

Nincs minden egyes anyanyelvi beszédhangunkra külön toldalékcső, amely mindig az adott hang képzésének konfigurációját veszi fel. Egyetlen toldalékcsövünk van, amely rugalmas annyira, hogy valamennyi beszédhangunkat képezni tudjuk vele, de rugalmatlan is annyira, hogy időre van szüksége, amíg az egyik pozícióból a másikba jut. Ez idő alatt pedig megtörténik a toldalékcső fiziológiai átalakulása, s ezt a folyamatot nevezzük koartikulációnak. A koartikuláció mértéke a módosuló hangot tekintve sokféle lehet, sőt a koartikulációs hatás több hangot is érinthet, nemcsak a közvetlen szomszédosat. [1]

A szomszédos beszédhangok egymásra hatása lehet előre ható (progresszív) és hátra ható (regresszív). A progresszív egymásra hatáskor a másodikként kiejtett hang illeszkedik a megelőzőhöz, míg a regresszív egymásra hatáskor az elsőként kiejtett hang közeledik a másodikként képzett hang artikulációs sajátosságaihoz. [1]

A koartikulációs folyamatok még ma sem pontosan ismertek, rendszerezésük bonyolult feladat. Nem egyszerű olyan szempontrendszer kialakítása, amelynek alkalmazásával kompromisszumok nélkül megoldható. A rendszerezés igénye mindenekelőtt a fonetikai és fonológiai koartikuláció meghatározását, elkülönítésüket, valamint az esetleges átfedések sajátosságainak megismerését teszi szükségessé. [1]

A fonetikai koartikuláció időben és térben fokozatosan jelenik meg, nemegyszer különböző szegmenseken keresztül is hat. Az idő itt az artikuláció időviszonyait, a beszédhangok képzésének sorrendiségét jelenti. A tér pedig arra utal, hogy az artikuláció például a képzési helyet tekintve módosulhat, változhat. A fonetikai koartikuláció tehát a definíciónak megfelelően azt jelenti, hogy a szomszédos beszédhangok valamilyen mértékben hatással

vannak egymás artikulációjára, függetlenül attól, hogy a koartikuláció következtében létrejövő módosult artikulációs konfiguráció definiál-e egy beszédhangot vagy nem. Fontos ugyanakkor, hogy a létrejött artikulációs konfiguráció nem lehet az adott nyelv valamely fonémájának az alaprealizációja. [1]

Fonológia folyamatnak tekintjük a beszédhangok egymásra hatásából adódó azon változásokat, amelyek egy másik fonéma alaprealizációját eredményezik, és a változások időben és térben nem tekinthetők fokozatosnak, illetőleg amelyek alaktani kötöttségűek. [1]

A koartikulációs folyamatok rendszerezése a következő:

Fonetikai koartikuláció:

- Módosult artikuláció:
 - CV, VC (kis, kút, munka)
 - CC:
 - Képzési mód (raktál, repce)
 - Képzési hely (hídnál, átnyúl)
 - Képzési mód és hely (vénség, hadból)
 - Időtartam (jobbra – jo[b']ra)
- Beszédhang megváltozása
- Hangátmenetek (gya, li sze, gyb)

Fonológiai koartikuláció:

- Hiátustöltés (fi[j]ú)
- Rövid-hosszú oppozíció (á[l]tam)
- Kiesés (dombtető)
- Hasonulások:
 - Egy képzési jegy (székből):
 - Zöngésség szerint (vágta, képzés)
 - Képzési hely szerint (rongy)
 - Képzési mód szerint (anyja)
 - Több képzési jegy (vitézség)
 - Sorozatos (kínja)
 - Alaktani (kutyussal)

2.5.1. Koartikulációs hangátmenetek

A szomszédos beszédhangok artikulációja hat egymásra, ez az egymásra hatás fiziológiailag magyarázható. Eredménye a hangátmenet. A hangátmenetek nem érintik a szomszédos beszédhangok artikulációs konfigurációját. A hangátmenetek akkor jönnek létre, amikor a beszéd folyamatos artikulációja során a beszédszervek az egyik beszédhang artikulációs konfigurációjáról áttérnek a rá következő, szomszédos beszédhang artikulációs konfigurációjára. A beszédszervek mozgást végeznek, de relatív tehetetlenségük következtében az egyik helyzetből a másik helyzetbe kerülésük folyamatos, azaz az egyik artikulációs konfigurációból folyamatosan rövidebb-hosszabb idő alatt jutnak a másik artikulációs helyzetbe. A hangátmenet akusztikai jellemzői az artikulációs sajátosságok következményei, attól függenek, hogy a szomszédos beszédhangok képzése milyen mértékben tér el egymástól. [1]

Az átmeneti részek nehezítik az egyes beszédhangok szegmentálását, vagyis azoknak a határoknak a kijelölését az akusztikumban, amelyek egy és csak egyetlen egy beszédhangot határoznak meg. Kivételt jelentenek a nazális mássalhangzók, amelyek általában nem idéznek elő szegmentálási problémát a nazális/orális képzés eltérése miatt. A magánhangzók első, de főként a második formánsa hajlik a szomszédos beszédhanghoz, vagyis a folyamatos artikuláció következtében egyfajta igazodás történik a megelőző és a követő beszédhanghoz. A formánsok frekvenciájának értéke ezért csak rövid ideig állandó, a változások mértékét a mindenkor fonetikai kontextus határozza meg. Két szomszédos mássalhangzó között az átmenet sajátosságait a találkozó hangok minősége dönti el. [1]

A hangátmeneteknek különös jelentőségük van a felpattanó zöngés (kisebb mértékben a zöngétlen) zárhangok és a magánhangzók kapcsolatában. Meghatározóak a magánhangzó-minőségek, továbbá az átmenetek mint akusztikai kulcsok biztosítják az adott zármássalhangzó pontos észlelését. [1]

Két magánhangzó találkozásakor a két hang közötti átmenet oly mértékben folyamatos, hogy az artikuláció szempontjából akár diftongus (egy szótagot alkotó két magánhangzó) is lehetne, az időviszonyok azonban jelentős eltérést mutatnak. Két mássalhangzó találkozásakor is megfigyelhetők a hangátmenetek, ezek azonban az adott mássalhangzó-minőségektől függően jobban vagy kevésbé kifejezettek. [1]

2.5.2. Módosult artikulációs konfiguráció

Ebben az esetben a szomszédos beszédhangok hatnak egymás artikulációjára, s ennek következtében az egyik beszédhang artikulációs konfigurációja módosul ugyan, de csak oly mértékben, hogy ne jöjjön létre egy, az eredetitől képzési jegyben eltérő másik beszédhang. Ezt más szóval igazodásnak is nevezik. Többféle altípusát különböztetjük meg. [1]

Mássalhangzó (CC) kapcsolatokban:

- A képzési mód változása:
 - Elmarad a zárhang felpattanása (raktál, repce)
 - Elmarad a nazális hang zár-felpattanása (színház, színlel)Ezek a jelenségek visszahatóak, regresszívek.
- A képzési hely változása (például alveorális zárhangoknál: hídnál, átnyúl)
Ezek is regresszív hatásúak.
- A képzési hely és a képzési mód együttes módosulása (vénség, hadból)
- A beszédhangok időtartamának változása
 - Rövidülés (jobbra /jobrA/, elszálltak /Elsa:ltAk/)
 - Hosszabbodás (katasztrófa /kAtAs:tro:fA/)

Mássalhangzó-magánhangzó (CV, VC) kapcsolatokban:

- A képzési hely változása a követő magánhangzó hatására:
 - Veláris képzésű zárhangoké (k, g: kutya, kicsi, kenyér)
 - Alveorális zárhangoké (d, t: tilt, támad, tulajdon)
 - Alveorális, posztalveoláris réshangoké (szita, szánkó, Zita, Zánka, sík, sál, zsír, zsák)

Ezekben az esetekben a követő magánhangzó ajakműködése visszahat a megelőző mássalhangzóra, tehát ezek a hatások regresszívek.

- Mássalhangzó formánsainak simulása a magánhangzóhoz
Például ha összehasonlítjuk a lúd szó /l/ mássalhangzóját a liba szókezdő mássalhangzójával, jelentős eltéréseket tapasztalhatunk a második és magasabb indexszámú formánsok területén, de kisebb mértékben eltérő az első formáns frekvenciája is.
- Magánhangzók nazalizálódása (marad, olyan, nem, munka)
Ez a jelenség regresszív és progresszív is lehet, sőt egyidejűleg mindkettő.

A fonetikai helyzet hatása is módosíthatja a beszédhang képzését. Ekkor nem a koartikuláció következtében módosul az artikuláció, az ejtésbeli változás mégis másmilyen artikulációs konfigurációt mutat. A felpattanó zárhangok, különösen a zöngétlenek hehezeztessé válnak az abszolút közlésvégen. Ez azt jelenti, hogy a zárfelpattanást a megszokottnál intenzívebb és hosszabb súrlódás követi, ami jól látszik a hangszíneképen. [1]

2.5.3. A beszédhang megváltozása koartikuláció hatására

Két szomszédos mássalhangzó úgy is hathat egymásra, hogy az egyikőjük képzési konfigurációja annyira megváltozik, hogy már egy másik beszédhang képzési jegyeivel lesz jellemezhető. Ennek több altípusa van: a) a képzési hely módosulása, b) a képzési mód módosulása, c) zöngésség változása. [1]

a) Az artikulációs hely megváltozása tapasztalható az alveoláris nazális és a felpattanó veláris (zöngés és zöngétlen) zárhangok találkozásakor: harang, ing, Lenke, inka, hangos stb. Az alveoláris nazális képzés helye eltolódik a veláris területre (megegyezően a követő zárhang artikulációs helyével). Ez a változás egy másik hang artikulációs konfigurációjával egyezik meg (/n/ helyett /N/). [1]

Hasonlóan az alveoláris és a két ajakkal képzett nazális mássalhangzók képzési helye is eltolódik és a labiodentális területre kerül, ha utána labiodentális mássalhangzó következik, mint a hamvas, szenved, kín volt, kámfor, kém volt stb. szavakban és szókapcsolatokban. Az így létrejött beszédhang az /M/. Ezek a folyamatok regresszívek. [1]

A veláris zöngétlen résmássalhangzó, a /X/ (pech, ihlet, technika szavakban) hátul képzett magánhangzó után posztalveoláris képzéshelyű beszédhanggá módosul, mint a doh, sah, jacht, Allah, eunuch vagy a potroh szavakban. Ezek a folyamatok progresszívek. [1]

b) A képzési mód változását eredményezi a koartikuláció, amikor az approximáns /j/ mássalhangzó az öt megelőző zöngétlen mássalhangzó hatására zöngétlen réshanggá alakul, például a kapj, rakj, döfj szavakban. Ez a beszédhang a /(X/. A spirantizálódás és a zöngétlenedés kizárólag szóvégen fordul elő. A változás progresszív irányú. [1]

c) A zöngétlen gégehang /h/, valamint a veláris, zöngétlen réshang /X/ két magánhangzó között általában zöngésedik /hV/. A koartikulációs folyamat a két magánhangzó közötti (intervokális) helyzetben érvényesül, például az éhes szóban. [1]

2.6. A szegmentális hangszerkezet észlelése

2.6.1. A beszédmegértés folyamata

A beszédmegértés folyamata két nagy szakaszból áll: 1. a nyelvi kódok rendszerének megfelelő hangjelenségek észlelése 2. ennek a kódrendszernek az értelmezése. Mindkettő több fokozatból (szintből) épül fel, amelyek törvényszerű együttműködésben biztosítják a hallott beszédjelenségek megértését. Az első nagy szakasz a beszédhangok, a hangsorok, a hangkapcsolatok felismerését jelenti, ezt nevezzük beszédészlelésnek. A második szakasz a kódrendszer értelmezése, vagyis a szavak, a mondatok és a szöveg megértése. Ez a szakasz a folyamatnak már a jelentés felismerését is tartalmazza, ezért már nem észlelésnek, hanem beszédmegértésnek nevezzük. [2]

A beszédmegértés komplex folyamat, amelyben (normális körülmények között) nem válik szét a beszéd értelmes és értelem nélküli szegmensekre vagy szegmentális és szupraszegmentális részre. A folyamat jelentős részének működéséről gyakorlatilag nincsenek vagy alig vannak pontos ismereteink, sejtjük csak, milyen működések történnek az idegpályákon, vagy az agyban. Többféle modellt állítottak fel a folyamat működésének megértéséhez. A hierarchikus modellt láthatjuk a **2.7. ábrán**. [2]



2.7. ábra. A beszédmegértés hierarchikus modellje

2.6.2. A beszédészlelés folyamata

A beszédészlelés szintén egy bonyolultan építkező és komplexen működő mechanizmus. Az észlelés akusztikai, fonetikai és fonológiai szintjei részben automatikusak, ez azt jelenti, hogy problémamentes feldolgozáskor a hallgató tudatosan nem ellenőrzi az ezeken a szinteken végbemenő működéseket, zavar esetén azonban lehetősége van az észlelési részfolyamatok kontrolljára. Az ép hallás biztosítja a hierarchikusan egymásra épülő szintek működését. [1] [2]

A bemenet maga a beszéd, az az akusztikai jelsorozat, amely a nyelvi információt tartalmazza. Az anyanyelv-elsajátítás folyamán tanuljuk meg azokat a stratégiákat, amelyek alkalmazásával képesek vagyunk észlelni a beszédhangok és hangkapcsolatok azonosságát különböző beszélők ejtésében. [1] [2]

Minden nyelvnek saját percepció bázisa van, amely – soros korrelációban az artikulációs bázissal – az anyanyelv-elsajátítás folyamán alakul ki. Percepció bázisnak nevezzük azt a nyelvspecifikus működésmechanizmust, amelynek során az elhangzott közléssorozatot feldolgozzuk, mégpedig úgy, hogy a nyelvi sajátosságok meghatározók, és hatnak a fiziológiai rendszer működésére. [1] [2]

A beszédészlelés folyamatát a hierarchikus szemléletű modellben tekintjük át. A feldolgozás az elsődleges hallási elemzéssel indul, ahol néhány előzetes döntés történik a frekvencia, az intenzitás és az idő szempontjából. Az itt létrejövő döntések nagyjából behatárolják a felfogott akusztikai jel hullámtulajdonságait. Felismerjük például, hogy zene vagy beszéd hangzott-e el, énekszót vagy hangszert hallottunk, gyors volt-e vagy lassú, magas vagy mély, halk vagy erős stb. [1] [2]

Ezek az eredmények az észlelési információtárban összegződnek, ahonnan a beszédfeldolgozás indul. Ezután egy felismerési terv következik a beszédészlelés szintjein. Ez utóbbi egyszerre három, hierarchikusan építkező síkon folyik: az akusztikai, a fonetikai és a fonológiai szinten. Ezeknek a szinteknek az elemzési eredményei az akusztikai-fonetikai tárból őrződnek. Ebből a tárból elegendő mindig a legjellemzőbb, tehát az éppen szükséges adatokat előhívni. A tárolási rendszer azt biztosítja, hogy az alsóbb szinteken hozott döntések a mind időben, mind absztrakcióban távolabb eső szintek eredményeinek ismeretében korrigálhatóak legyenek. [1] [2]

Az akusztikai elemzéssel választ kapunk arra, hogy a hangsornak mint hullámformának mik az akusztikai tulajdonságai: volt-e alaphfrekvencia, észlelhető-e energiakoncentráció valamilyen frekvencián vagy nem, ha igen, akkor hol, a frekvenciaváltozások miként jelentkeznek az időben, holt volt intenzitáscsúcs (volt-e), milyen volt annak a lefutása stb. Az akusztikai elemzés a feltételezések szerint bináris döntések sorozatával valósul meg. [1] [2]

Az akusztikai elemzés adataira épül a fonetikai osztályozás. Ekkor döntési szabályok alapján nyelvészeti-fonetikai szempontból ítéljük meg a bejövő jelet. Egy folyamatosan változó akusztikai jelsorozatban nehéz eldönteni, melyik és mekkora rész jelenti az adott fonéma realizációját. További nehézséget jelent az a tény, hogy gyakran nem konkrét értékek alapján kell döntéseket hoznunk, hanem viszonyok alapján. Az akusztikai jelsorozatban vannak olyan pontok, amelyek általában könnyebben azonosíthatók paramétereik alapján (pl. a magánhangzók tiszta fázisai), mások nehezebben. Mindehhez szükség van szegmentálásra, amely a fonémák különböző realizációit, illetőleg a jelentéssel bíró jeleket különíti el. A fonetikai elemzés eredménye a fonetikai tárho kerül. [1] [2]

A részben automatikus feldolgozási szakasz utolsó lépcsője a fonémadöntés meghozatala a fonológiai szinten. A beszédhangokat itt azonosítjuk a megfelelő fonémákkal. Például, hogy a sebhintőpor szóban a fonetikai szinten zöngétlen /p/-nek felismert hang, valójában a zöngés /b/ megvalósulása. A fonémáról való döntés megvalósulhat izolált megjelenéskor is, amikor a beszédhangot önmagában halljuk. A beszédhangok más részének felismeréséhez azonban szükség van a környezet bizonyos mértékű részvételére is, mivel a hangkörnyezet hordozza a jellemző jegyeket. [1] [2]

A fonológiai szint elemi percepciók egysége a szótag. Nyilvánvaló ugyanakkor, hogy a szótag, mint a fonémadöntés alapja, nyelvspecifikus, nemcsak felépítését tekintve, hanem a tekintetben is, hogy az egy szótaghoz kapcsolódó jelentés milyen gyakori. A magyar kísérleti adatok szerint a fonémadöntés alapja egy magánhangzóból és egy mássalhangzóból álló hangkapcsolat (sorrendtől függetlenül) vagy egy magánhangzóból és több mássalhangzóból felépülő szekvencia lehet. A /b/, /d/, /g/ azonosítása például az őket követő magánhangzótól függ, a fonémadöntés tehát csak hangkapcsolatban működik. A magánhangzók egy részének a tiszta fázis elegendő a felismeréséhez, más részüknél viszont a hangkörnyezet biztosítja a felismerést. [1] [2]

2.6.3. A szegmentálás

A beszédészlelés egyik alapvető funkciója a szegmentálás. A szegmentálás eredménye lehet a beszédet felépítő hosszabb közlés (például a mondat), lehet annál rövidebb, grammatikailag szorosabban összetartozó közlésegység (például a szókapcsolat), lehet szótári egység, azaz maga a szó, vagy lehet például a szótag, a hangkapcsolat, sőt maga a beszédhang is. [1] [2]

A szegmentálásról eddigi ismereteink a következők: a) a szegmentálás nyelvspecifikus folyamat, b) elsősorban a beszédészlelésnek, de adott esetben a beszédmegértés folyamatainak is funkciója, c) a szegmentálás készsége az anyanyelv-elsajátítás során fejlődik ki, d) szoros összefüggést mutat az elhangzó beszéd akusztikumával és az adott nyelv fonológiai sajátosságaival. [1] [2]

2.6.4. Invariancia, redundancia, irrelevancia

A beszéd akusztikai szerkezetében invariánsnak nevezzük azokat a paramétereket vagy paraméter-együtteseket, amelyek egyértelműen meghatározzák az adott szegmentumot, és biztosítják az észlelését. Az invariancia egy adott hangnak olyan akusztikai és/vagy artikulációs jellemzője, amely minden kontextusban képes felidézni az adott hangot. [1] [2]

A beszéd hullámformája nagy mennyiségű többletinformációt tartalmaz, amelyek jó átviteli körülmények között feleslegesek. Ennek egy része bizonyos körülmények között fontos lehet, ezt nevezzük redundanciának. Csendes körülmények között jóval kevesebb akusztikai jel elegendő a beszéd megértéséhez, de például zajos környezetben a redundáns elemek is információhordozóvá válnak. Az irrelevancia olyan elemek összessége, amik semmilyen körülmények között nem hasznosíthatók az üzenet dekódolásához. [1] [2]

2.6.5. Normalizációs folyamatok

A normalizáció egy hipotetikus észlelési folyamat, amelyben a beszélők közötti artikulációs különbségek akusztikai következményei még azelőtt csökkennek, hogy a nyelvi kategóriák felismerése megtörténne. A normalizációs folyamat hivatott a beszélősemleges percepciót biztosítani. Az első lépésben az akusztikailag különböző egységek valamiféle közös reprezentációs formába alakulnak át, és tárolódnak a memóriában. Ennek következtében az információ egy része elvész ugyan, de a jelek változatossága csökken. [1]

2.6.6. A beszédhangok észlelése

Az invariáns jegyet a beszédhangok azonosításában felismerési kulcsokkal fejezhetjük ki. Amikor a felismerési kulcsokat meghatározzuk, tulajdonképpen az akusztikum és a percepció közötti összefüggéseket tárjuk fel. Vannak elsődleges és másodlagos felismerési kulcsok. Az utóbbiak akkor jutnak szerephez, amikor az elsődleges kulcsok eltűnnek (például mélyfrekvenciás zaj nehezíti a felismerést), vagy valamilyen okból nem hozzáférhetőek (például a személynek a mély frekvenciákon halláskiesése van). A másodlagos kulcsok lényegében redundáns tényezők. Az elsődleges felismerési kulcsaiktól megfosztott, izolált magyar magánhangzókat 80% körüli biztonsággal azonosították a kísérleti személyek, vagyis a másodlagos kulcsok igen jó észlelési eredményt adnak. A felismerési kulcsok nyelvspecifikusak. [1]

2.6.7. A magánhangzók észlelése

A magánhangzók felismerésének kérdése, hogy vajon az őket jellemző paraméterek közül melyek invariánsak az észlelés szempontjából. A különféle módszerű kísérletek eredményei a következő paraméterek jelentőségét igazolták a magánhangzók felismerésében:

- A magánhangzó tiszta fázisának időtartama
- Az aktuális hangkörnyezet
- A magánhangzó spektruma

A magyarban és a legtöbb nyelvben a magánhangzók első két formánsa meghatározza a hangminőséget. Ez azt jelenti, hogy nem a konkrét értékek, hanem a formánsfrekvenciák viszonya működik felismerési kulcsként az észlelésben. Az alaphangmagasság változása a felismerés szempontjából általában nem lényeges tényező. A formánsfrekvenciának van a legnagyobb hatása az észlelésre, az intenzitás és az F0 másodlagosak. Kísérleti eredmények szerint az első formáns valamilyen mértékben tartalmazza a második formáns információját is, valamint az F2 az F1-ét. Az időtartam és a frekvencia szoros összefüggésben biztosítja a magánhangzók pontos észlelését. [1]

A magyarban a rövid és hosszú magánhangzók jelentéseket különítenek el, így valószínűsíthető, hogy az észlelésben is elkülönülnek a nyelviileg rövid és hosszú párok tagjai. Nemcsak az időtartambeli de képzésbeli különbségek is vannak köztük. [1]

2.6.8. A mássalhangzók észlelése

A mássalhangzók felismerését (csakúgy, mint a magánhangzókét) a frekvencia, az intenzitás és az idő, illetőleg ezek meghatározott kombinációja biztosítja. Nem elhanyagolható a kontextus hatása sem, noha az egyes mássalhangzóknál a kontextusfüggőség eltérő mértékű. Vannak olyanok, amelyeknél erősen meghatározó (például a felpattanó zöngés zárhangoknál), másoknál alig van jelentősége (például a réshangoknál). [1]

A zöngés felpattanó zárhangok azonosítása igen bizonytalan, erőteljesen kontextusfüggő, és természetesen meghatározó a mássalhangzó képzési sajátosságai következtében létrejövő akusztikai szerkezete. A kísérleti eredmények azt sugallják, hogy többszörös akusztikai kulcsok szükségesek a pontos azonosításukhoz. Ilyenek a követő magánhangzó első formánsának frekvenciája, az alaphangmagasság változása a zárfelpattanást követően, a jellegzetes zörejfrequencia és a második formáns. A magyarban a zöngés felpattanó zárhangok zöngéssége erős, ami azt jelenti, hogy a megelőző beszédhang végétől már tapasztalható a hangszalagok rezgése, és ez a zárfelpattanás alatt és azt követően is tart. Mesterségesen előállított hangkapcsolatokkal végzett kísérletek eredménye szerint a követő magánhangzó a /b/, /d/ esetében nagymértékben, a /g/ esetében kevésbé, de hatással van a zármássalhangzó képzési helyének felismerésére. Az intenzitás és a fonetikai pozíció meghatározó szerepét mutatja, hogy a /b/ zárhang felismerése például a *bá* hangkapcsolatban 100%-os, *áb* hangkapcsolatban csak 13%-os. Ez utóbbi a az intenzitás csökkenésének következménye. [1]

A felpattanó zöngétlen zárhangok észlelése sok esetben még bizonytalanabb, mint a zöngéseknél, mindkét típusnál jellemző a kontextusfüggőség. Ezek azonosításához is többféle akusztikai kulcs szükséges. Vizsgálatok eredményei szerint a /p/ felismerése 90, /k/ és /t/ felismerése 75 százalék körüli a zárfelpattanást követő átmeneti szakasz hiányában is. Ezek a zárhangok általában egymás között keverednek az észlelésben. Az intenzitás itt is meghatározó jelentőségű, de ez sem független a fonetikai pozíciótól. A /p/ mássalhangzó felismerése például *pá* vagy *áp* hangkapcsolatokban igen hasonló, 36-40%-os, ugyanakkor a /k/ felismerése CV-kapcsolatban 68, VC-kapcsolatban csak 17%-os. [1]

Általánosan elmondható, hogy minél messzebb van a magánhangzó képzési helye a mássalhangzó képzési helyétől, annál nagyobb mértékű a pontos észlelés. [1]

Résmássalhangzóinkra jellemző, hogy az 1000 Hz alatti komponenseik önmagukban nem teszik lehetővé azonosításukat. Ezekben az esetekben e hangokat zárhangokként vagy /h/-ként azonosítják. A felismerés függ a réshang típusától, a kontextustól és a fonetikai helyzettől. A zörejes résmássalhangzók, a /z/, /s/, /Z/, /S/ - artikulációjuk következtében - meglehetősen széles frekvenciatartományban valósulnak meg. Az egyes mássalhangzókra jellemző zörejek intenzitása azonban különböző, s meghatározott frekvenciasávban jelentkezik. [1]

A zöngés labiodentális, alveoláris és posztalveoláris résmássalhangzók azonosításához a zöngé jelenléte és valamivel szűkebb frekvenciasáv szükséges, mint a zöngétlenek esetében. A frekvenciasáv mellett döntő akusztikai kulcs a felismerésben az intenzitás. Ha a zörejjeg intenzitása megváltozik, nem lesz többé jellemző az adott mássalhangzóra, és a felismerés egy másik beszédhangba csap át. [1]

A frekvencia és az intenzitás mellett a résmássalhangzók harmadik jellemző felismerési kulcsa az időtartam, jelentős megváltozása megváltoztatja az észlelt hang minőségét. Az időtartam csökkenésével /s/ hangból először /ts/, majd /t/ lesz. A /v/ mássalhangzó közel 80%-osan azonosítható a vá hangkapcsolatban, míg vá kapcsolatban felismerhetetlen, a ví esetében 54%-os, az ív hangkapcsolatban 29%-os. [1]

A zár-rés hangok, azaz az affrikáták azonosításához mind a zárszakasz, mind a résszakasz frekvencia-, intenzitás- és időzítési sajátosságainak megléte szükséges. Ellenkező esetben e mássalhangzók homorgán zárhangokba, illetőleg réshangokba csapnak át. A percepció kísérletek megerősítik, hogy a zörejelem időtartama, a zár- és résmozzanat relatív aránya a homorgán zárhang – affrikáta – réshang elkülönítés elsődleges felismerési kulcsa (Harrington 1988). A specifikus intenzitásértéken túl meghatározó a kontextus és a fonetikai pozíció is. [1]

A pergőhang azonosítása elsődlegesen a perdületek számától függ, illetőleg az azt tükröző frekvenciaszerkezettől. Az intenzitás is elsődleges akusztikai kulcsként funkcionál, s ehhez még a kontextus és a fonetikai pozíció sajátosságai is hozzájárulnak. A /r/ felismerése erőteljesen csökken például a rí hangkapcsolatban (70%), még tovább csökken az ár kapcsolatban (40%), s az ír esetében már csak véletlen találatról beszélhetünk (29%). [1]

A nazális mássalhangzók felismerési kulcsai egy alacsony, mintegy 300 Hz-es és egy 1000 Hz-en megjelenő formánsban, valamint a zöngében határozhatók meg. Az észlelés alapkritériuma, hogy az úgynevezett nazális rezonancia hallható legyen. A kísérleti adatok azt igazolják, hogy a bilabiális /m/ és a palatális /j/ nazálisokat – a fonetikai helyzettől függően – gyakran alveoláris nazálisként /n/ azonosítják, illetőleg a palatális /j/ bilabiálisként /m/, fordítva azonban sohasem. Mégis ezek az orrhangú mássalhangzók tekinthetők a legstabilabb észlelésűeknek. Az alveoláris nazális /n/ elsősorban /l/-ként, ritkábban /r/-ként ismerik fel. [1]

A specifikus intenzitásnak meghatározó szerepe az észlelésben az alveoláris és a palatális mássalhangzók esetében, a bilabiális nazálisnál azonban másodlagos, bár a fonetikai pozíciónak itt is jelentős a szerepe. A relatív időtartam-változás lényegesen nem befolyásolja a nazálisok pontos azonosítását. A nem fonémaértékű nazálisok, az allofónok (/F/, /N/) azonosítása meglehetősen bizonytalan. [1]

A közelítőhangok észlelési sajátosságai a magánhangzókéra emlékeztető akusztikai szerkezetük miatt azokéval mutatnak hasonlóságot. Elsődleges felismerési kulcsuk az első két formáns, bár a /j/ esetében az észlelés keveredik az /i/ magánhangzóval, a /l/ esetében pedig elsősorban a nazális /m/-el, ritkábban a /r/-vel. Az időtartam és a spektrális jellemzők együttjárása ugyancsak a magánhangzóknál tapasztalt észlelési következményeket mutatja. CV hangkapcsolatokban az intenzitás másodlagos felismerési kulcs, VC-kapcsolatokban azonban elsődlegessé válik kontextustól függetlenül, különösen a /j/ esetében. [1]

2.6.9. Hangsorok észlelése

A szakirodalomban rendelkezésre álló hatalmas adathalmaz ellenére még mindig számos kérdésre nincs egyértelmű válasz a beszédpercepció folyamatát illetően. Nem egyértelmű például, hogy az akusztikai jelsorozatnak milyen és mekkora része tartalmazza a nélkülözhetetlen információt, változik-e egy szegmentum percepciója a szegmentumsor nagyságától függően, hogyan történik az akusztikai kulcsok elsődleges és másodlagos felhasználása. A hosszabb szekvenciák azonosíthatóságát a kontextus és/vagy a percepció valószínűség biztosítja. A hangsorok felismerésében meghatározó az akusztikai szerkezet, a fonotaktikai elrendeződés, az adott fonetikai kontextus és az artikulációs tempó. [1]

2.7. Adatbázisok

A digitális beszédfeldolgozás, a nagyméretű tárhelykapacitás lehetővé teszi a különböző típusú adatbázisok létrehozását. A digitális beszédfeldolgozás különböző területei más-más méretű és összeállítású adatbázist igényelnek. [3]

Az adatbázisok számítógép segítségével létrehozott, tárolt és a szükséges magyarázó jegyzetekkel, címkézésekkel és átírásokkal ellátott beszédfelvételek gyűjteményei. Rádióból, TV-ből felvett beszéd nem adatbázis. Az adatbázis lényeges tartozéka a precízen leírt dokumentáció a rögzítés technikájáról, a beszélők számáról és típusáról, a nyelvi tartalomról, oly módon, hogy az adatbázist felhasználók egyszerűen megkapják a gyűjteményre vonatkozó szükséges információt. [3]

Nagyon sokfajta adatbázis létezik. A beszédtechnológiával foglalkozó szakemberek számára igen fontos ezeknek az adatbázisoknak az ismerete, azért hogy közülük egy meghatározott feladatra a legmegfelelőbbet tudják kiválasztani, vagy ha nincs megfelelő adatbázis, hogyan kell az adott feladathoz az optimálisat létrehozni. Az adatbázisok különböznek egymástól abban, hogy milyen felhasználási területre készültek, mekkora a bennük gyűjtött beszéd mérete, a bemondók száma, stb. [3]

2.7.1. Adatbázisok felosztása

A jelenleg elérhető adatbázisok 3 alap kategóriába sorolhatók felhasználás szerint:

- Analitikus – diagnosztikus adatbázis: nyelvi és fonetikai kutatások segítségét szolgálja. Ilyen, pl. a BABEL (EURM0, EUROM1 adatbázis).
- Általános adatbázis: nem specifikus, általános szótárakat tartalmaz, sokfajta felhasználásra alkalmas, mint például a SPECO (gyermek beszédadatbázis).
- Specifikus adatbázis: olyan beszédgyűjtemény, amely meghatározott felhasználási területen készül. Különböző felismerők betanítására alkalmas adatbázis. Ilyen például a SPEECHDAT adatbázis.

Adatbázisokra jellemző, hogy milyen nyelvi egységekből épülnek fel, pl. izolált szavakból, mondatokból stb., továbbá a bemondás módja szerint lehet olvasott szöveg, spontán beszéd. [3]

2.7.2. Adatbázisok tervezése

Az adatbázisok, vagy az adott feladathoz legjobban illeszkedő adatbázis kiválasztásánál az alábbi szempontokat kell figyelembe venni: a felvételek és a rögzítés pontos fizikai leírását, a felvett anyag nyelvi jellemzőit, az adatbázis méretét, a beszélők szociolingvisztikai adatait, az adatbázis feldolgozási módját. Adatbázisok jellemző adatai láthatók a **2.10. táblázatban**. [3]

A beszéd adatbázis bemondási körülményei, a rögzítés módja az adatbázis lényeges része. Itt kell figyelembe venni:

- egy, vagy több mikrofon, mikrofon műszaki leírása
- környezet: stúdió, süketszoba, iroda stb.
- felvétel ellenőrzési módszer
- mintavételi paraméterek

2.10. táblázat Adatbázisok jellemző adatai

rögzítés fizikai leírása	Mintavételi paraméterek Felvételi körülmények fizikai leírása Monitor használata
nyelvi jellemzők	Rögzített nyelv, dialektus Nyelvi alapegység: hangkapcsolatok, szavak, mondatok Bemondott anyag leírása Bemondás stílusa: olvasott, spontán beszéd, dialógus
méretbeli jellemzők	Beszélők száma Rögzített anyag időbeli hossza Nagysága CD-k száma
szociolingvisztikai jellemzők	nem, kor, beszéd stílusa
adatbázis feldolgozása	Címkézés Átírás Szegmentálás Spektrális elemzés

Beszélők száma szerint különféle adatbáziscsoportok léteznek:

- **Kevés beszélő adatbázisa**

Például a beszédszintézis fejlesztés céljait szolgálja. Lényeges jellemzője a lehető legnagyobb fonetikai variáltságú anyag összegyűjtése. Az anyagban hangsúlyozottan szerepet kapnak a beszéd mikroszegmentális jellemzői. Rendszerint a bemondást szakértő végzi.

- **Adatbázis közepes számú beszélővel**

A felismerésénél használt modell paraméterek becslésére szolgál. Éppen ezért a nyelvi szöveg variáltsága nagy. Általában csendes helyiségekben történik a felvétel. Beszélők száma kisebb, mint 50.

- **Adatbázis sok beszélővel**

Ezek az adatbázisok a beszélő független felismerők betanítására szolgálnak. A beszédstílus és a rögzítési körülmények nagy variáltsága, több mint 100 beszélő szükséges. [3]

A szociolingvisztikai jellemzők főleg a bemondók leírására szolgálnak. Férfiak, nők, dohányoznak, nem dohányoznak, anyanyelvükön történik-e a bemondás, tájszólások vannak-e rögzítve az adatbázisban. Milyen a bemondók koreloszlása. [3]

2.7.3. Adatbázisok nyelvi feldolgozása

Beszédatadatbázisoknak a beszéd digitális tárolása mellett annak nyelvi információ tartalmát is rögzíteniük kell. Ezért a hullámforma tárolása mellett a hozzátartozó ortografikus karaktereket is rögzítik. Különböző zajok, embertől származóak,- (ilyen a köhögés, nyelés, különböző szájmozgásból adódó zajok), vagy környezeti (járművek, motorok zaja, székcikorgás stb.) bejelölésre kerülnek a legtöbb adatbázisban, vagy a szöveganyagban, vagy magában az időfüggvényben. [3]

Akusztikai jelek fonetikai átírásakor rendszerint a karaktereket hozzárendelik a rögzített hullámformához, vagyis a folyamatos beszédet pl. beszédhang egységekben kvantálják, bejelölik a beszédhangok elejét és végét, valamint beírják a beszédjelhez tartozó írásos szimbólumokat. Ezek a szimbólumok lehetnek egy adott nyelv betűi, de ha az adatbázis nemzetközi célra készül, akkor célszerű nemzetközi jelölésrendszert használni, hogy a külföldi szakemberek is pontosan tudják milyen hangok sorozatáról van szó. Az európai adatbázisok már SAMPA karaktereket használnak. [3]

A fonetikai átírásnak számos szintje létezik:

- **Kanonikus fonetikai átírás**

Az adott szöveg karaktereinek olyan átírása, amelyben az ortografikus karaktereket fonémák sorozatára alakítjuk ki, de az adott szöveggörnyezetet nem vesszük figyelembe. Tehát a hasonulás és a koartikuláció nincs figyelembe véve.

- **Fonotipikus fonetikai átírás**

A karakterek átírását, az adott nyelv fonetikai szabályainak alapján végezzük, a szöveggörnyezet függvényében (pl. A hasonulási szabályok figyelembe vételével).

- **Hallás alapján történő fonetikai átírás**

A figyelmesen lehallgatott szöveg hallás alapján történő lejegyzése. Tehát itt, az írott szöveg figyelembe vétele nélkül, kizárólag a hallott hangok kerülnek lejegyzésre.

- **Audio-vizuális fonetikai átírás**

A fonémáknál kisebb egységek alapján történik az átírás, a közel stabil akusztikai-fonetikai részek bejelölésével. Az átírást a szöveg hallgatása, és az időfüggvény vagy a színekép elemzése alapján hajtják végre. [3]

2.7.4. Szegmentálás

Szegmentálás során a beszéd időfüggvényében bejelölik a beszédhangok, vagy egyéb fonetikai egységek határait, és beírják a megfelelő fonetikai szimbólumokat. Kézi vagy automatikus szegmentálást szokás használni. A kézi szegmentálás pontos, de fárasztó és időigényes. Az automatikus, vagy félautomatikus szegmentálás gépi felismerésekkel hajtható végre. Gyors eljárás, de elég pontatlan, ilyenkor kézi korrekciót kell végezni. [3]

2.7.5. Adatbázisok a gyakorlatban

Beszédatadatbázisokat elsősorban a gépi beszédfelismerésben használunk. Széles felhasználói terület még az automatikus beszéd-szintézis, kódolás, elemzés, beszédazonosítás, nyelvazonosítás. Mindezen területek nagy adatbázist igényelnek. Nem csak betanításra használatosak, hanem tesztelésre is, hiszen segítségükkel rögzített, állandó anyaggal, tehát ismétlési lehetőséggel lehet végrehajtani a tesztelést. [3]

Néhány jellemző adatbázis adatait a **2.11.** és **2.12. táblázat** mutatja be.

2.11. táblázat Adatbázisok minőségi jellemzői

Adatbázis neve	Forrás	Formátum kHz	Rögzítési környezet	Bemondás módja	Feldolgozás	
					alapegység	átírás
TI Digits	mikrofon	20	csendes szoba	felolvasás	szó	nem
TIMIT	mikrofon	16	csendes szoba	felolvasás	beszédhang	igen
NTIMIT	telefon	8	telefonon keresztül telefonfülke, iroda, lakás, utca stb.	felolvasás	beszédhang	igen
ATISO	mikrofon	16	hivatal	spontán felolvasás	mondat	nem
Switchboard (Credit Card)	telefon	8	telefonon keresztül telefonfülke, iroda, lakás, utca stb.	spontán beszéd	szó	igen
Switchboard (Credit Card)	telefon	8	telefonon keresztül telefonfülke, iroda, lakás, utca stb.	spontán beszéd	szó	igen
MARSEC	mikrofon	16	változó	spontán	beszédhang	igen
ATIS2	mikrofon	16	hivatal	spontán	mondat	nem
EUROM-1 BABEL	mikrofon	20	csendes helyiség	olvasott	beszédhang szó	igen
SpeechDat SpeechDat-E	telefon	8	vezetékes telefon, mobil telefonfülke, iroda, lakás, utca stb.	olvasott, spontán	beszédhang szó	nem

2.12. táblázat Adatbázisok mennyiségi jellemzői

Adatbázis neve	CD száma	Felvételi idő	Méret gigabájt	Beszélők száma	Egységek száma
TI Digits	3	~14	2	3260630	>2 500 szám
TIMIT	1	5,3	0,65	630	6 300 mondat
NTIMIT	2	5,3	0,65	144	6 300 mondat
ATISO	6	20,2	2,38	69	10 722 kiejtés
Switchboard (Credit Card)	1	3,8	0,23	16	35 dialógus
Switchboard (Credit Card)	30	250	15	100	2 500 dialógus
MARSEC	1	5,5	0,62	351	53 mondat
ATIS2	6	~37	~5	>124	12 000 kiejtés
EUROM-1 BABEL	3-5	nyelvfüggő	nyelvfüggő	100	számok, fonetikailag kiegyensúlyozott mondatok, hangkapcsolatok, szavak
SpeechDat SpeechDat-E	4-6	nyelvfüggő	nyelvfüggő 1,5-4	nyelvfüggő 500-5000	számok, nevek, intézmények, utasítások, fonetikailag gazdag mondatok

2.7.6. Magyar nyelvű adatbázisok

Ez ideig 3 egymástól igen különböző magyar beszédatbázis készült el, amelyek összefoglaló adatait a **2.13. táblázat** tartalmazza. [3]

2.13. táblázat Magyar beszédatbázisok összefoglaló adatai

	BABEL	SpeechDat-E	SPECO gyermek adatbázis
forrás	mikrofon	telefon	mikrofon
formátum	20 kHz, 16 bit	8 kHz, 16 bit (ISDN)	20050 Hz, 16 bit
rögzítési kör- nyezet	süketszoba (tisztá beszéd)	iroda, lakás, utca, telefon- fülke stb.	süketszoba
bemondás módja	olvasott szöveg	80% olvasott, 20% spon- tán szöveg	olvasott, utánzott szöveg
szövegtípus	hangkapcsolatok számok, szavak folyamatos szöveg	betűzött szavak dátumok, pénzösszegek számok, telefon- és hitel- kártyaszámok szavak, tulajdonnevek, mondatok	kitartott beszédhangok hangkapcsolatok számok, szavak mondatok
bemondók száma	60	1000	76
feldolgozás	fonotipikus átírás fonémaszintű szeg- mentálás	karakteres leírás nincs szegmentálás zajok, hibák jelölése	fonotipikus átírás fonémaszintű szeg- mentálás

3. A BESZÉDÉRTÉS VIZSGÁLATA

A vizsgálat egy az egyesült államokbeli EBIRE (East Bay Institute for Research and Education) kutatóintézetben angol nyelven végzett kísérlet mintájára készült magyar nyelvre. Alapja az a feltevés, hogy a beszéd érthetőségét javítani lehet, ha a például átviteli kiesések miatt keletkezett rövid, pár száz milliszekundumos „lyukakat” feltöltjük különböző jelekkel. Várakozások szerint a beillesztett beszédzaj elősegíti az információ helyreállítását, azaz érthetőbbé válik a beszéd.

Ennek a vizsgálatához úgy lett összeállítva a feladat, hogy meg lehessen határozni mely esetekben tapasztalható javulás, ha egyáltalán van ilyen. A lehetőségek szerint ugyanazokkal a paraméterekkel készült, mint az angol nyelven végzett kísérlet, hogy a későbbiekben össze lehessen hasonlítani a kettő eredményeit. Ehhez először is egy, az amerikai kísérletnél használt beszédadatbázishoz hasonló adatbázist kellett létrehozni.

3.1. A spondeuszok

A spondeusz eredeti jelentése két hosszú szótagból álló versláb az időmértékes verselésben. Jelen feladat értelmezése szerint olyan kétszótagú összetett szó, aminek az első tagja mássalhangzóra végződik, a második pedig azzal kezdődik, tehát a szó közepén legalább két mássalhangzó van, és a tagok külön-külön is értelmesek. Mind az angol, mind a magyar nyelvben elég sok ilyen szó található, és egyszerű kezelni őket, amikor egy 200-600 ms-os, a szótaggal nem megegyező darab kivágása a feladat.

Először is tehát ilyen szavakat kellett keresni, majd ki lett választva közülük negyven darab a beszédérthetőségi teszthez. A válogatáskor az elsődleges szempont az volt, hogy lehetőleg mindenféle mássalhangzó szerepeljen az első tagok végén, és a második tagok elején, és a magánhangzók közül is minél többféle legyen benne. Természetesen ahhoz, hogy minden magánhangzó-mássalhangzó kapcsolatot lehessen vizsgálni sokkal több szóra lenne szükség, de ennyi elegendő ahhoz, hogy lássuk egyáltalán van-e alapja a feltevésnek, miszerint a lyukak feltöltésével javítani lehet az érthetőségen, ha igen, milyen esetekben, és van-e értelme további vizsgálatoknak.

A szavak listája és azok fonotipikus fonetikai átírása X-SAMPA szimbólumokkal a **3.1. táblázatban** látható. A betűk átírása beszédhangokká a magyar nyelv fonetikai szabályai alapján történt, figyelembe véve a szöveggörnyezetet is (pl. koartikuláció, hasonulás). Található köztük kettő kivétel is: szakág, szótó, amik közepén csak egy mássalhangzó van.

3.1. táblázat. A spondeuszok listája

1.	bükk	fa	byk:fA	21.	nap	pá	nAp:a:
2.	csepp	kő	tSEp:k2:	22.	nyak	lánc	JAkla:nts
3.	éj	fél	e:jfe:l	23.	ős	kor	2:Skor
4.	el	fut	Elfut	24.	rak	part	rAkpArt
5.	gép	zsír	ge:pZi:r	25.	rak	tár	rAkta:r
6.	gomb	lyuk	gombjuk	26.	sír	kő	Si:rk2:
7.	gőz	gép	g2:zge:p	27.	szak	ág	sAka:g
8.	hang	szál	hANgsa:l	28.	szak	ma	sAkmA
9.	itt	hon	it:hon	29.	szám	tan	sa:mtAn
10.	jég	hegy	je:ghEJ\	30.	szét	tép	se:t:e:p
11.	jég	krém	je:k:re:m	31.	szín	kép	si:Nke:p
12.	jog	díj	jogdi:j	32.	szó	tő	so:t2:
13.	kap	tár	kApta:r	33.	táv	cső	ta:vtS2:
14.	kék	fény	ke:kfe:J	34.	tér	kép	te:rke:p
15.	kód	szó	ko:dso:	35.	test	vér	tESTve:r
16.	kor	só	korSo:	36.	toll	szár	tol:sa:r
17.	köz	kincs	k2skintS	37.	tűz	fal	ty:sfAl
18.	köz	pont	k2spont	38.	új	kor	u:jkor
19.	kulcs	csont	kultStSont	39.	vér	csepp	ve:rtSEp:
20.	láb	szár	la:bsa:r	40.	víz	szint	vi:s:int

3.2. A hamis-spondeuszok

Ezekből lehet előállítani a hamis spondeuszokat, amik egy létező összetett szót utánoznak, úgy, hogy vagy az első, vagy a második felet rímelik, azaz a helyettesítő egyszótagú szóban a magánhangzó megegyezik az eredeti szótag magánhangzójával, a mássalhangzók tetszőlegesen. Így biztosítjuk azt, hogy megmarad a magánhangzó-harmónia, és csak olyan magánhangzó-mássalhangzó kapcsolatok fordulnak elő a szótagokban, amik amúgy is előfordulnak a magyar nyelvben. Ez a felismerésnél fontos szempont, mert a hangsorépítési (fonotaktikai) sajátosságok meghatározóak a hangsorok felismerésében.

Az EBIRE-ben végzett kísérlethez négyféle hamis spondeusz-listát készítettek:

- az első tag változatlan (arm-hair)
- a második tag változatlan (yard-chair)
- mindkét tag meg lett változtatva (art-fair)
- véletlenszerű szótagokból összeállított spondeuszok (key-stone)

A zárójelekben olyan példák vannak, amiket az „armchair” szóból képeztek. Mindegyik lista 80 elemből áll. Az eredeti spondeuszokkal együtt ez 400 szó. Ezekből készítették a hangfelvételeket. Két beszélő olvasta fel a listákat, így összesen 800 felvétel készült.

A magyar nyelven végzett vizsgálathoz csak az első hamis-spondeusz listát állítottuk elő, úgy hogy az első tagot változatlanul hagytuk, a második tagokat pedig összekevertük a magánhangzó-egyezés figyelembevételével. Így lett például a „bükkfa”-ból „bükkpart”. A hamis-spondeuszokat és a szavak fonotipikus fonetikai átírását mutatja a **3.2. táblázat**.

3.2. táblázat. A hamis-spondeuszok listája

1.	bükk	part	byk:pArt	21.	nap	szál	nApsa:l
2.	csepp	cső	tSEp:tS2:	22.	nyak	tár	JAkta:r
3.	éj	vér	e:jve:r	23.	ős	csont	2:StSont
4.	el	lyuk	Ej:uk	24.	rak	lánc	rAkla:nts
5.	gép	díj	ge:pdi:j	25.	rak	tan	rAktAn
6.	gomb	fut	gombfut	26.	sír	tő	Si:rt2:
7.	gőz	fény	g2:sfe:J	27.	szak	fa	sAkfA
8.	hang	tár	hANGta:r	28.	szak	szár	sAksa:r
9.	itt	pont	it:pont	29.	szám	fal	sa:FfAl
10.	jég	csepp	je:gtSEp:	30.	szét	krém	se:tkre:m
11.	jég	kép	je:k:e:p	31.	szín	gép	si:Nge:p
12.	jog	zsír	jogZi:r	32.	szó	kő	so:k2:
13.	kap	lánc	kApla:nts	33.	táv	kő	ta:vk2:
14.	kék	fél	ke:kfe:l	34.	tér	fény	te:rfe:J
15.	kód	só	ko:dSo:	35.	test	gép	tEZge:p
16.	kor	szó	korso:	36.	toll	ág	tol:a:g
17.	köz	kor	k2skor	37.	tűz	part	ty:zpArt
18.	köz	szint	k2s:int	38.	új	pont	u:jpont
19.	kulcs	kor	kultSkor	39.	vér	hegy	ve:rhEJ\
20.	láb	szál	la:bsa:l	40.	víz	kincs	vi:skintS

3.3. A hangfelvételek elkészítése és előfeldolgozása

A beszédatadabázishoz a felvételek a BME Beszédakusztikai Laborjában lévő süketszobában készültek, vagyis csendes, zajmentes környezetben. A bemondás mikrofonba történt, ami egy számítógép hangkártyájának a bemenetéhez csatlakozott. Egy női beszélő olvasta fel egy listáról először a spondeuszokat, majd a hamis-spondeuszokat, a szavak között rövid szünetet tartva. Így egy összesen 80 szóból álló hangfelvétel készült.

Ezen a felvételen aztán szószintű szegmentálás lett végrehajtva, minden szó külön, 16 bites, 44100 Hz mintavételezésű, PCM formátumú wav fájlokban lett tárolva. Beszédatadabázisoknál nem szokás 20 kHz-nél nagyobb mintavételi frekvenciát használni. Shannon tétele szerint elegendő a jelben előforduló legnagyobb frekvencia kétszerese, és a beszéd szükséges elemeit a 10 kHz alatti frekvencia-összetevők tartalmazzák. Ebben az esetben viszont az adatbázis kis mérete miatt nem probléma a nagyobb mintavétel. Végül a hangfájlok egy automatikus normalizációs eljárással azonos hangerőszintre lettek hozva.

3.3. táblázat Az adatbázis adatai

Forrás	mikrofon
Rögzítési környezet	süketszoba, tiszta beszéd
Bemondás módja	felolvasás
Beszélők száma	1
Szövegtípus	spondeuszok (értelmes, kétszótagú összetett szavak) hamis-spondeuszok (értelemmel nem rendelkező szavak)
Egységek száma	40 spondeusz, 40 hamis-spondeusz
Formátum	44.1 kHz, 16 bit, PCM
Méret	6,5 Mbyte
Felvételi idő	150 s
Feldolgozás	fonotipikus átírás, szószintű szegmentálás, részleges fonémaszintű szegmentálás

3.4. A szegmentálás

Különböző alkalmazásokhoz készült beszédatbázisok feldolgozásakor általában fonémaszintű szegmentálást kell végezni, azaz bejelölni a beszédben a fonémák határait, és beírni a megfelelő fonetikai szimbólumokat. Ez egy meglehetősen fárasztó és időigényes folyamat. A munka megkönnyítésére készültek automatikus szegmentáló programok, azonban ezek nem elég pontosak, kézi korrigálásra is szükség van.

A feladat célkitűzése szerint nem volt szükség teljes szegmentálásra. Elég volt meghatározni a magánhangzók elhelyezkedését, a kezdeti- és végpontjukat, és az időtartamukat. Ennek ellenére a - spondeuszok szerkezetéből adódóan - ez a legtöbb esetben már behatárolta szinte az összes fonémát a szóban. Minden szónál kézi szegmentálás történt a beszédjel hullámformája (oscillogram) és a hangszínképelemzés révén előálló spektrogram, valamint lehallgatás alapján. Ez a Cool Edit 2000 nevű programmal történt, ami képes a hangfájlok hullámformájának és spektrogramjának megjelenítésére, továbbá tetszőleges hangrészlet lejátszására.

Az oscillográfia és a hangspektográfia a beszéd akusztikai szerkezetének vizsgálatához használatos eljárások. Az oscillogramon látható az egyes időpillanatokban megjelenő rezgések amplitúdója az idő függvényében. Az oscillogram voltaképpen a rezgés idődiagramja. A hangspektogramon a beszéd három paraméterének változásai olvashatók le: az idő (a vízszintes tengelyen), a frekvencia (a függőleges tengelyen) és az intenzitás (a feketedés mértékéből). [1]

A **3.1. és 3.2. ábrákon** kettő szó, a *bükkfa* /byk:fA/ és a *számfa* /sa:FfAI/ hullámformája és spektrogramja látható. Függőleges vonalakkal vannak jelölve a fonémahatárok, alul pedig az X-SAMPA szimbólumok sorozata. Az időtartománybeli képen könnyen észrevehető a magánhangzók sokkal nagyobb energiája a mássalhangzókhoz képest, és az ebből az intenzitásbeli különbségből adódó szükséges felfutási és lefutási idő. A spektrogramon ki-vehetőek a formánsok, valamint megfigyelhető a periodikus jelleg zöngés hangoknál a függőleges csíkozásból. Látszik a hangok egymásra hatása, a koartikuláció is: a *számfa* szóban, az /a:/ hang formánsai a /F/ formánsaihoz közelítenek. Az is kitűnik, hogy vannak esetek, amikor a fonémák jól elkülöníthetőek, pontos határvonalat lehet húzni például az /f/ és /A/ beszédhangok között. Ellenben az /A/ és /I/ hangok között folyamatos átmenet van.

Azokban az esetekben, amikor a hangok nem különíthetők el élesen egymástól, például a magánhangzó-rezonáns mássalhangzó kapcsolatokban, a határ az átmeneti rész 50%-ánál lett bejelölve. A feladat jellegéből kifolyólag nem volt szükség olyan pontos szegmentálásra, mint például egy gépi beszédfelismerő betanításához készült adatbázisnál, de a pontosság így is milliszekundumos nagyságrendű. A fonémahatárokat sem kellett eltárolni, ezekre csak a vágás helyének meghatározásához volt szükség, a további feldolgozáshoz nem.

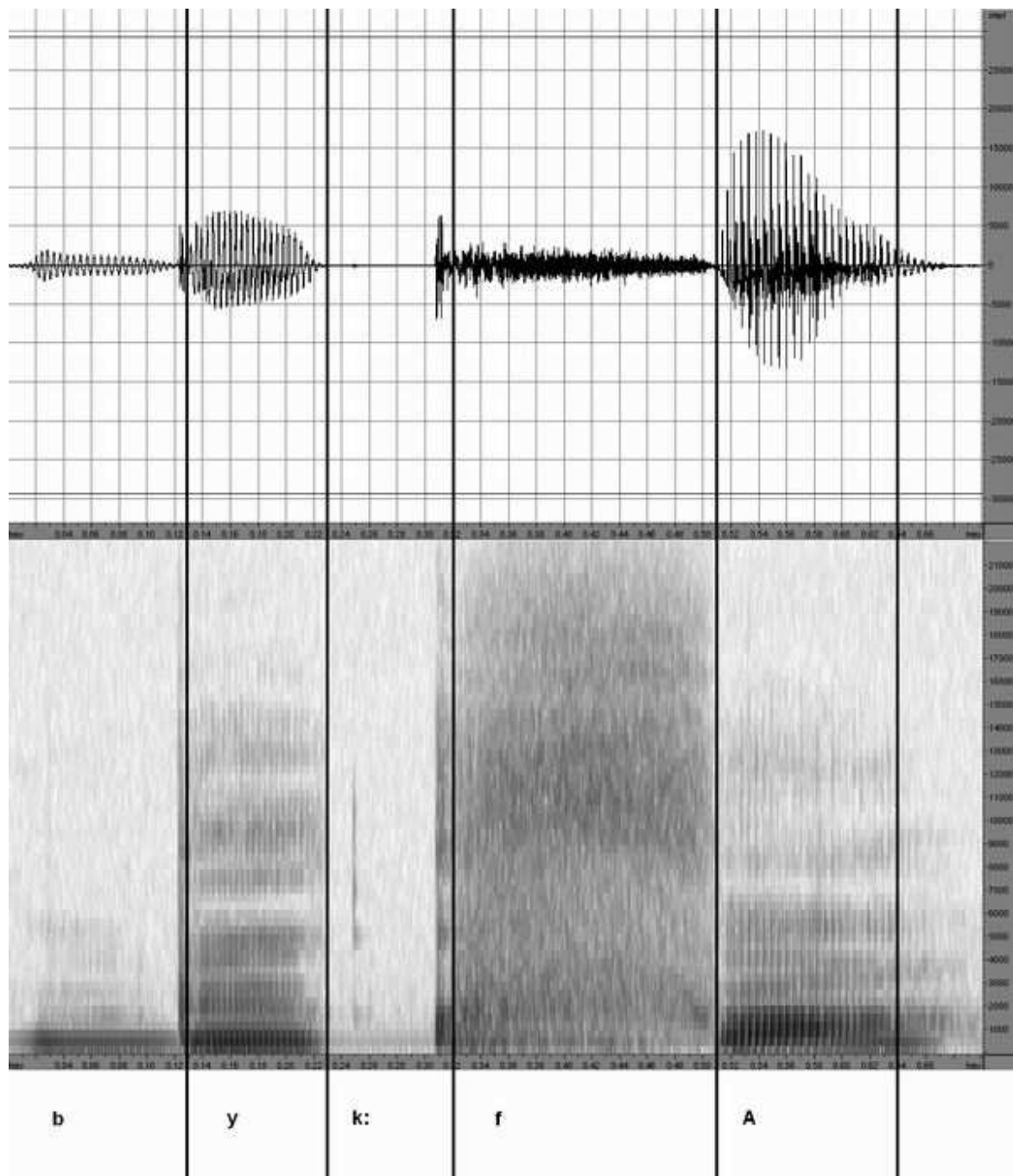
3.5. A kivágás

Minden igazi és hamis-spondeusz közepéből ki lettek vágva a mássalhangzók teljes egészében, illetve a magánhangzókból a teljes időtartamukhoz viszonyítva 25 vagy 50 százalék. A kivágások szolgáltak a beszédátviteli rendszerekben történő rövid (például a csatorna zavartságából adódó) átviteli kiesések modellezésére.

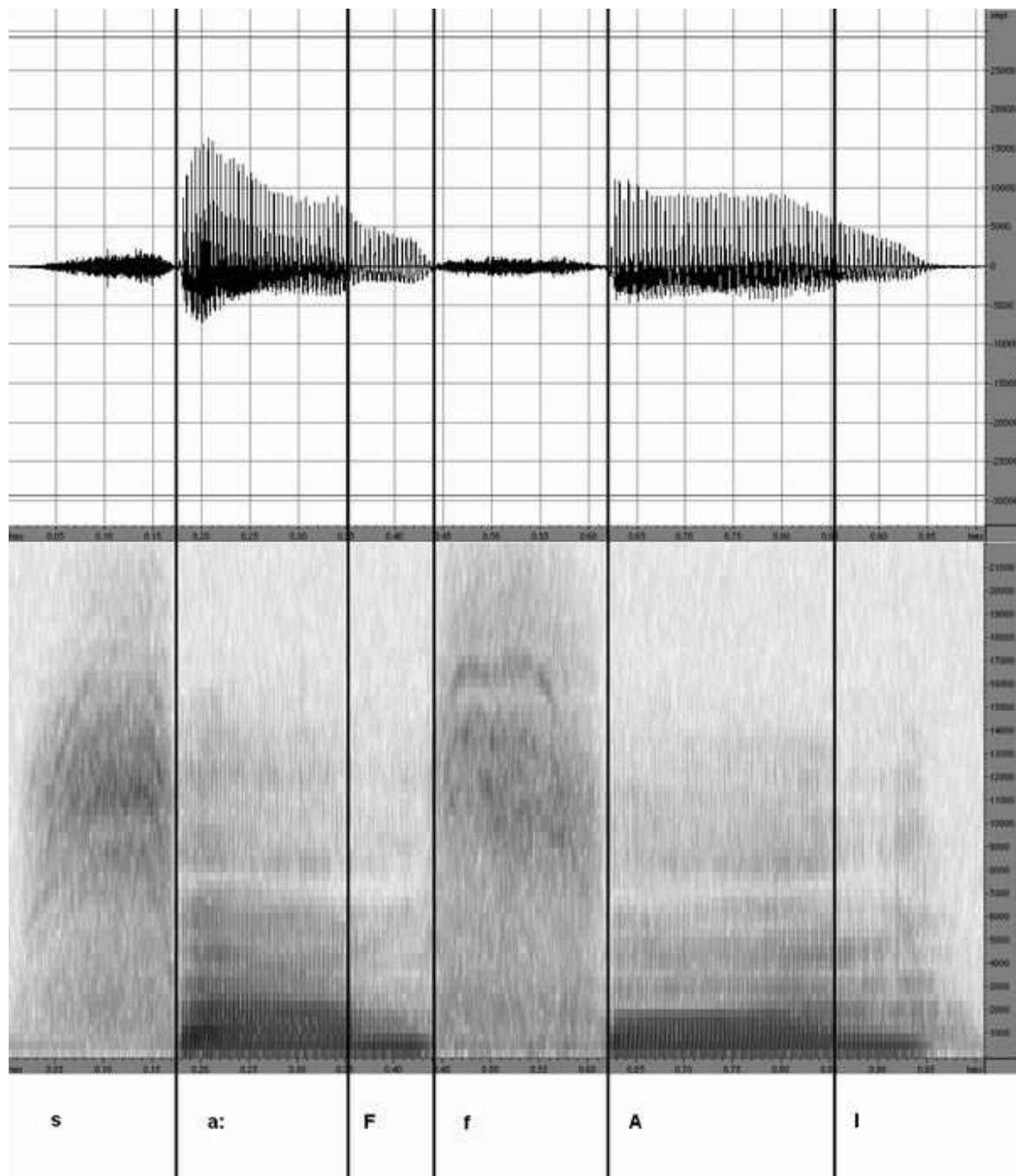
A vágás a 25 és 50 százalékos részhez legközelebb eső nullátmeneteken történt a hirtelen ugrások elkerülése végett. Tehát az ezen pontok között lévő rész ki lett törölve, és a felvétel időtartama nem változott. A kivágás hossza a fonémák számától, időtartamától, elhelyezkedésétől és típusától függően változott. 25 százalékos esetben 250-450 milliszekundumig terjed, 50 százalékosnál 340-590 milliszekundum.

A **3.2. és 3.4. ábrán** a *bükkfa* és a *számfal* szavak vágott verziói láthatók. Az egyik a 25, a másik az 50 százalékos kivágást mutatja. A 25 százalékos esetben általában még marad a megelőző vagy a követő hang koartikulációs hatásából, az 50 százalékos vágás azonban sokszor már az állandósult résznél van.

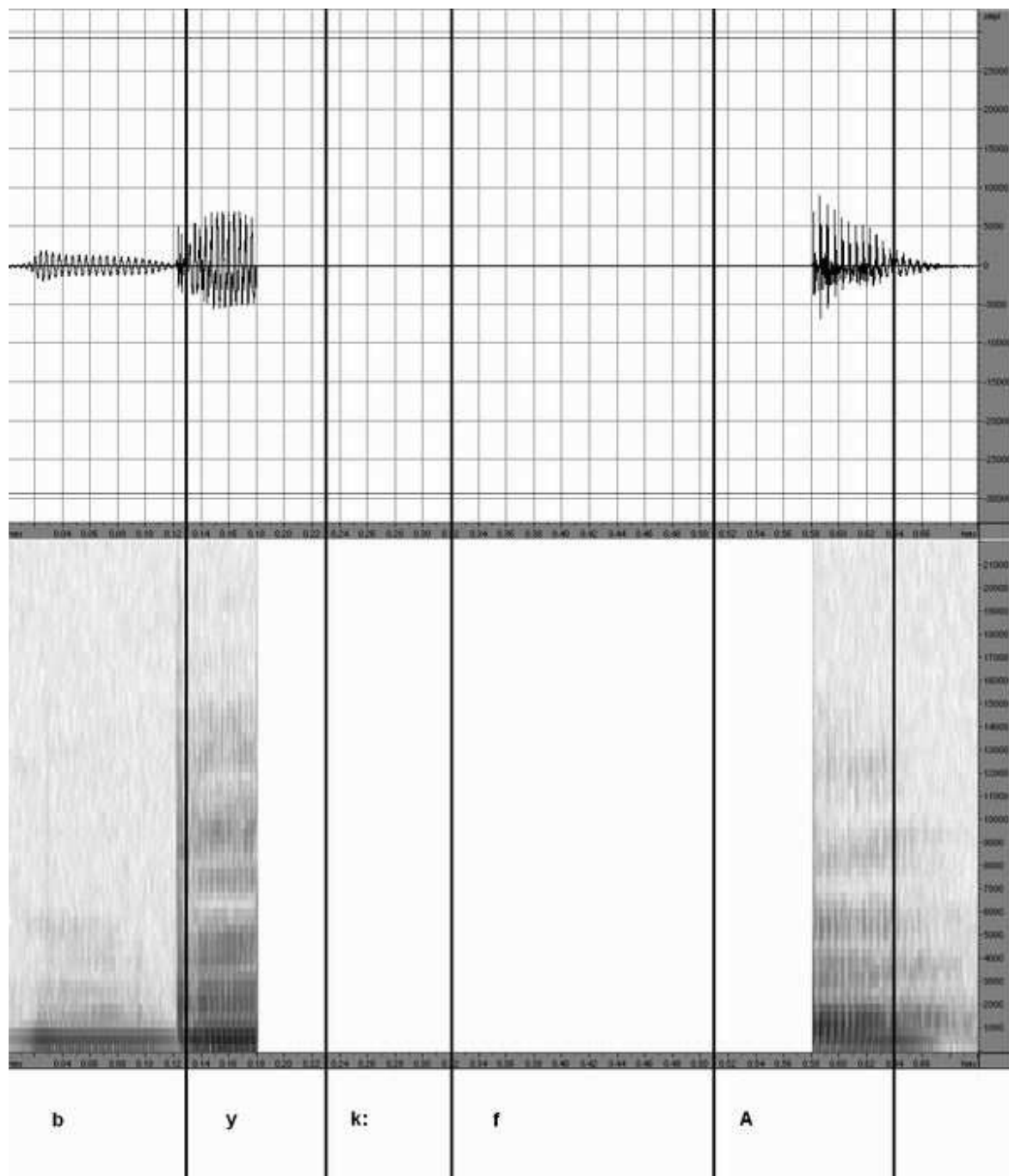
A „lyukas” közepű spondeuszok önmagukban is a hallgatás tárgyát képezik. Ezek lesznek összehasonlítva a „feltöltött” szavakkal, így lehet kideríteni, hogy van-e javulás az érthetőségben. A következő lépés tehát a lyukak helyére különféle jelek beillesztése.



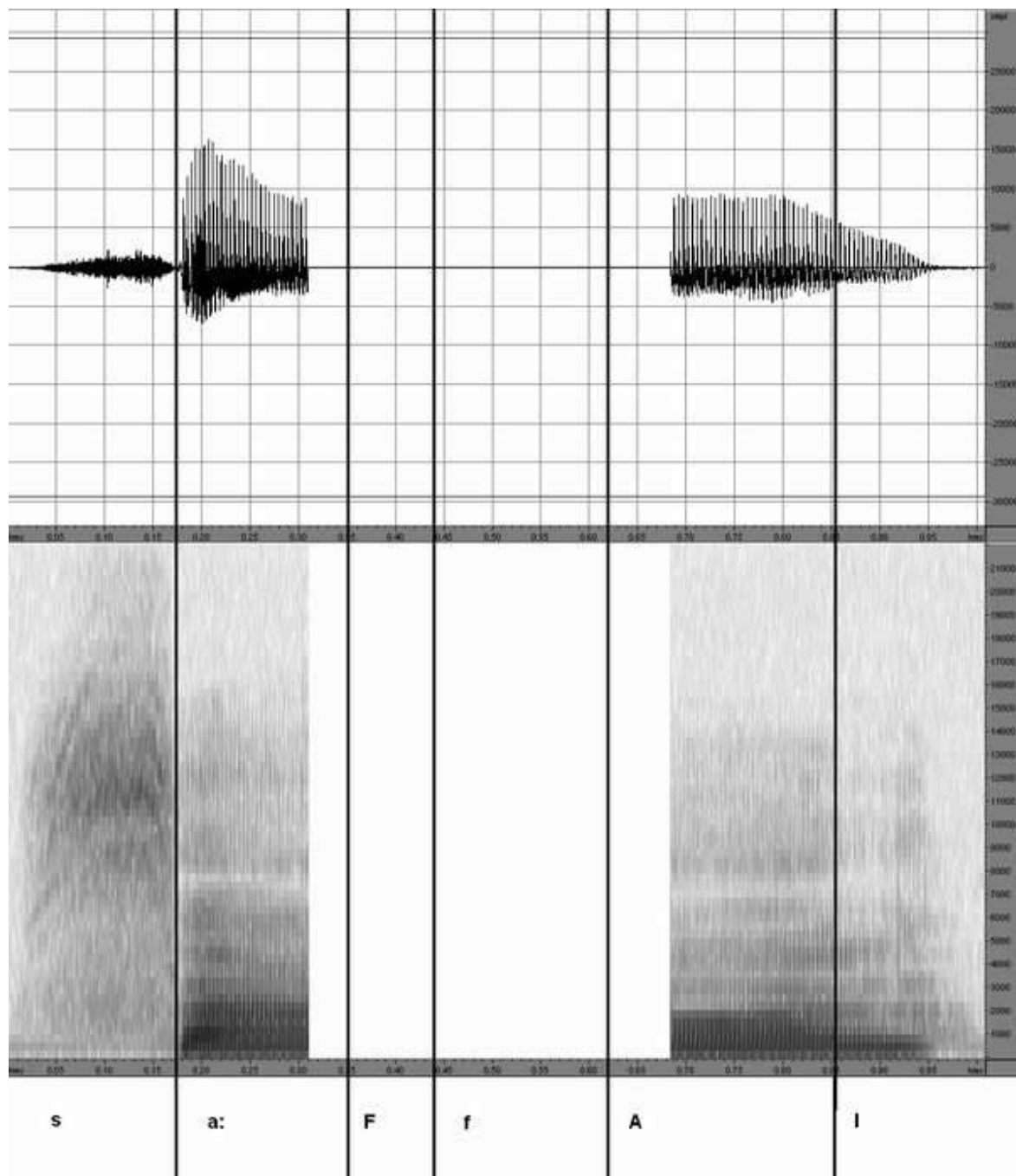
3.1. ábra. A bűkfa szó hullámformája és hangszínképe



3.2. ábra. A számfal szó hullámformája és hangszíneképe



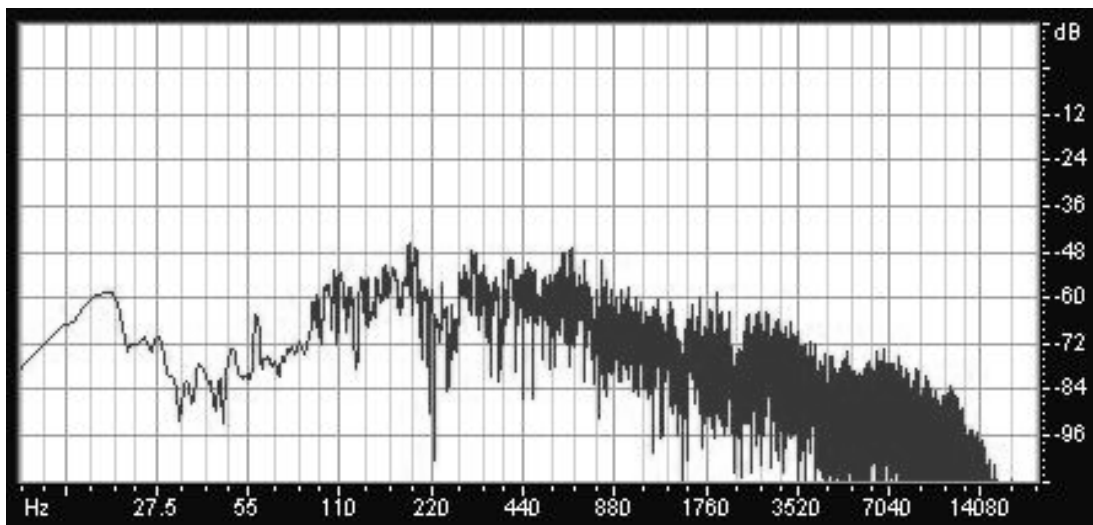
3.3. ábra. A bükkfa szó hullámformája és hangszíneképe 50 százalékos kivágás esetén



3.4. ábra. A számfal szó hullámformája és hangszínképe 25 százalékos kivágás esetén

3.6. A feltöltés

A feltöltés kétfajta jellel történt: beszédzajjal illetve fűrészzel. A beszédzaj olyan zaj, aminek a spektrális eloszlása megegyezik a beszéd átlagos beszédszínképével. A magasabb frekvenciák felé csökkenő intenzitást mutat. A beilleszteni kívánt beszédzaj kivágásánál is figyelembe lettek véve a nullátmenetek, és az időtartama megegyezik a kivágott rész időtartamával, tehát a szó hossza nem változott. Átlagos intenzitása megegyezik a szó átlagos intenzitásával.



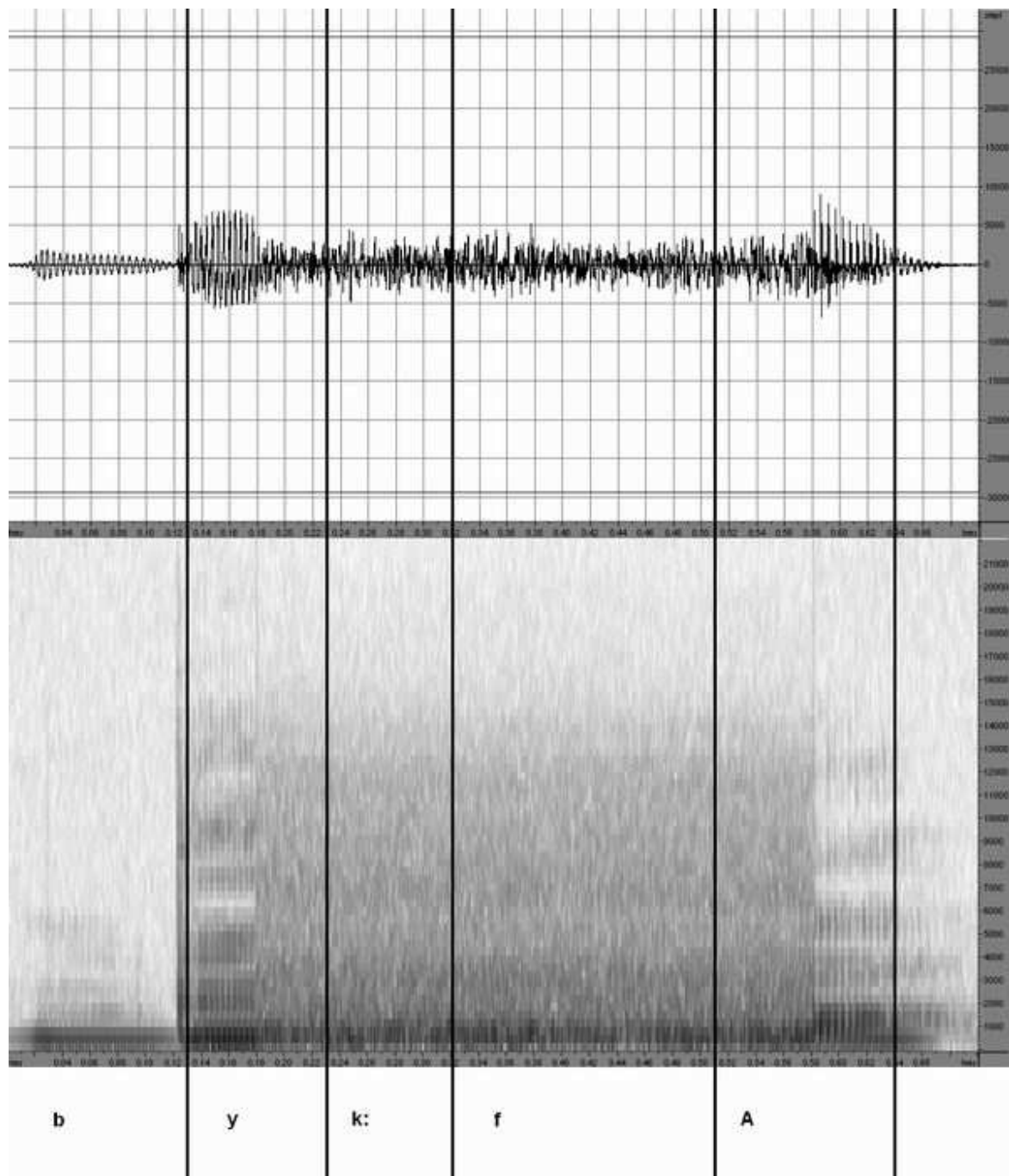
3.5. ábra. A beszédzaj spektrális eloszlása

A fűrészel alapharmonikusa a beszélő alaphangja, ami körülbelül 240 Hz. Átlagos energiája ennek is a szó átlagos energiájával megegyező.

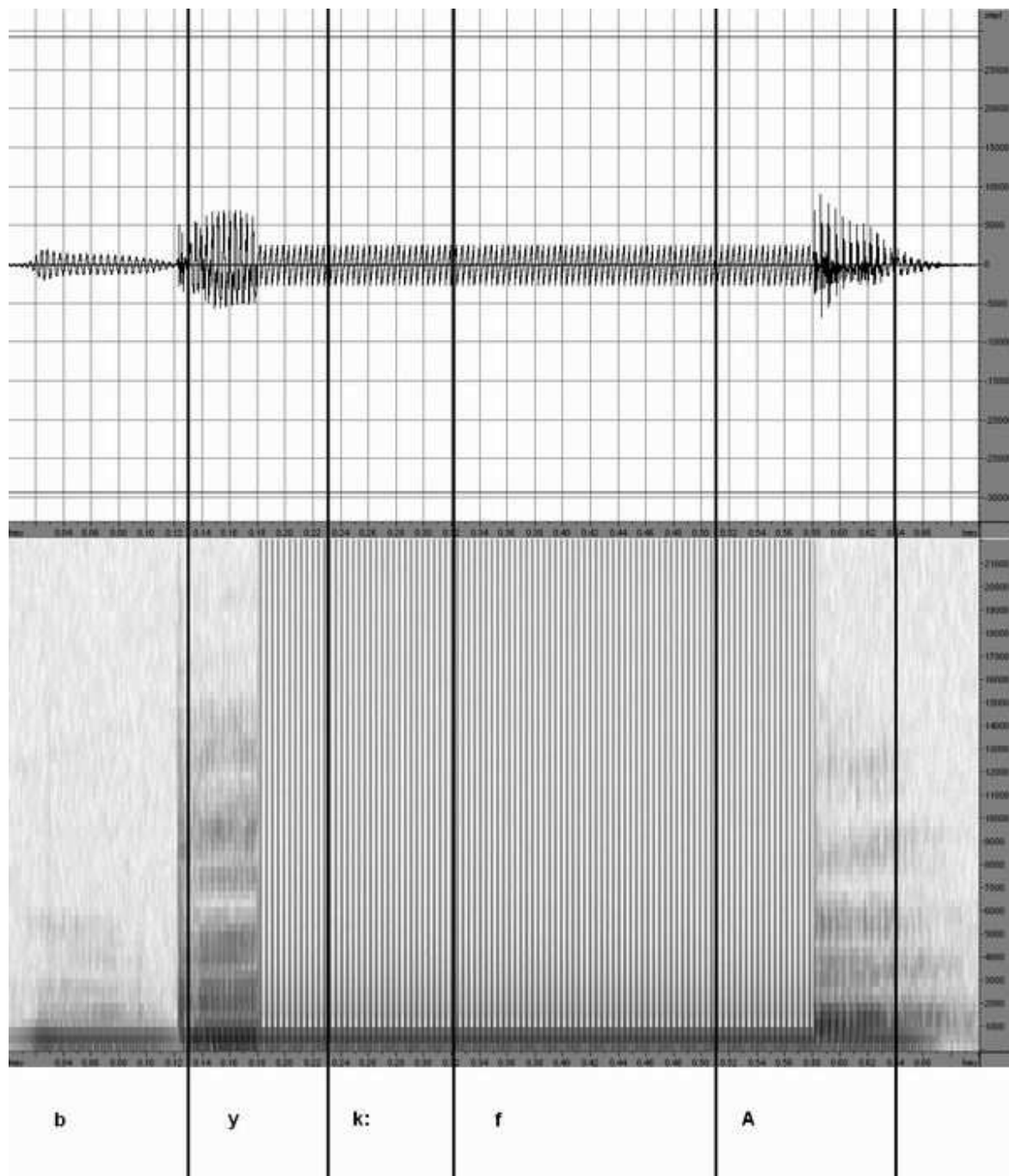
Így tehát minden szóból hat verzió készült, összesen 480 hangfelvétel a vizsgálatokhoz:

- 25%-os kivágás
- 25%-os kivágás, beszédzajjal feltöltve
- 25%-os kivágás, fűrészzel feltöltve
- 50%-os kivágás
- 50%-os kivágás, beszédzajjal feltöltve
- 50%-os kivágás, fűrészzel feltöltve

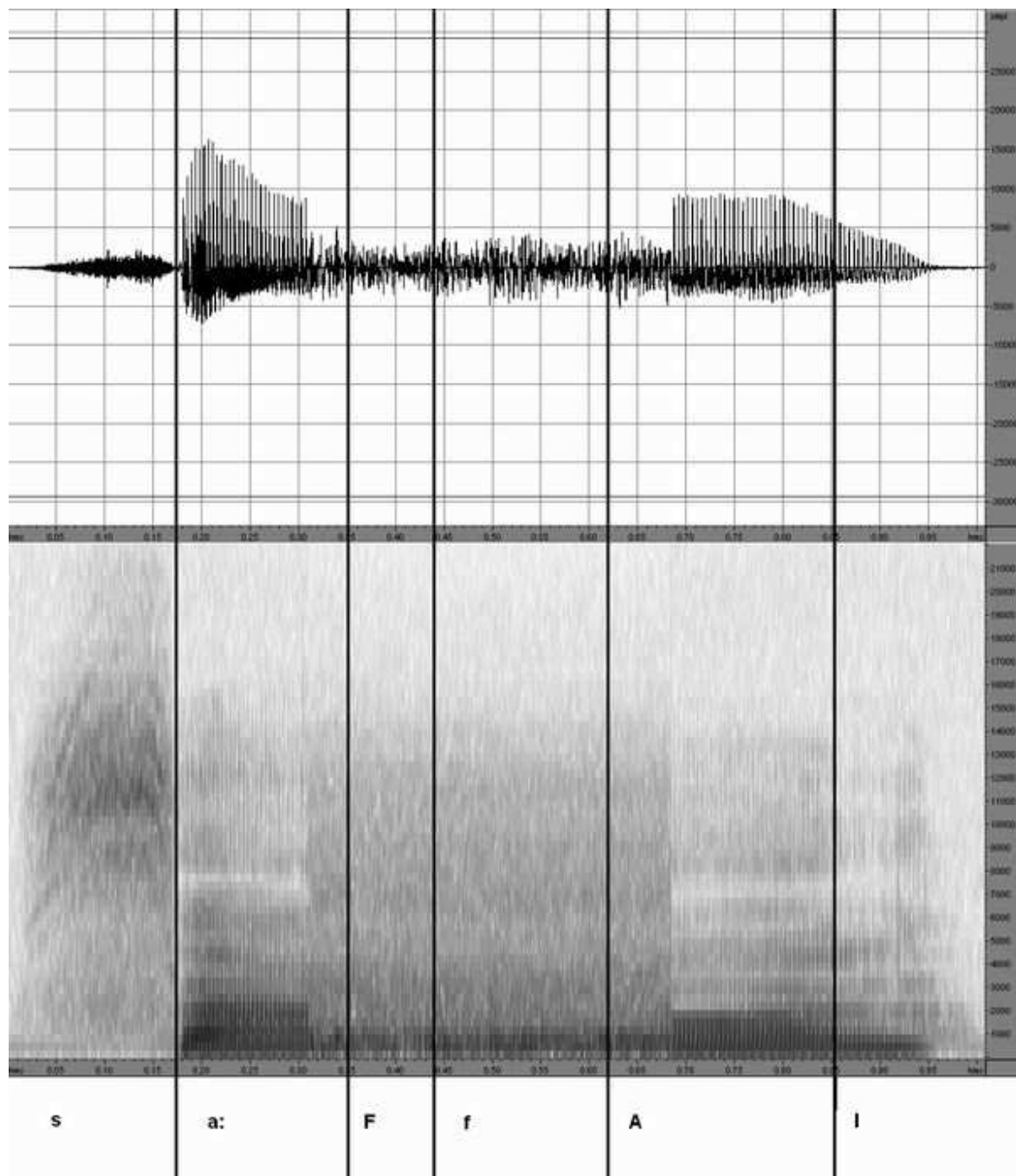
A vágás és a lyukak feltöltése a Cool Edit 2000 nevű programmal készült.



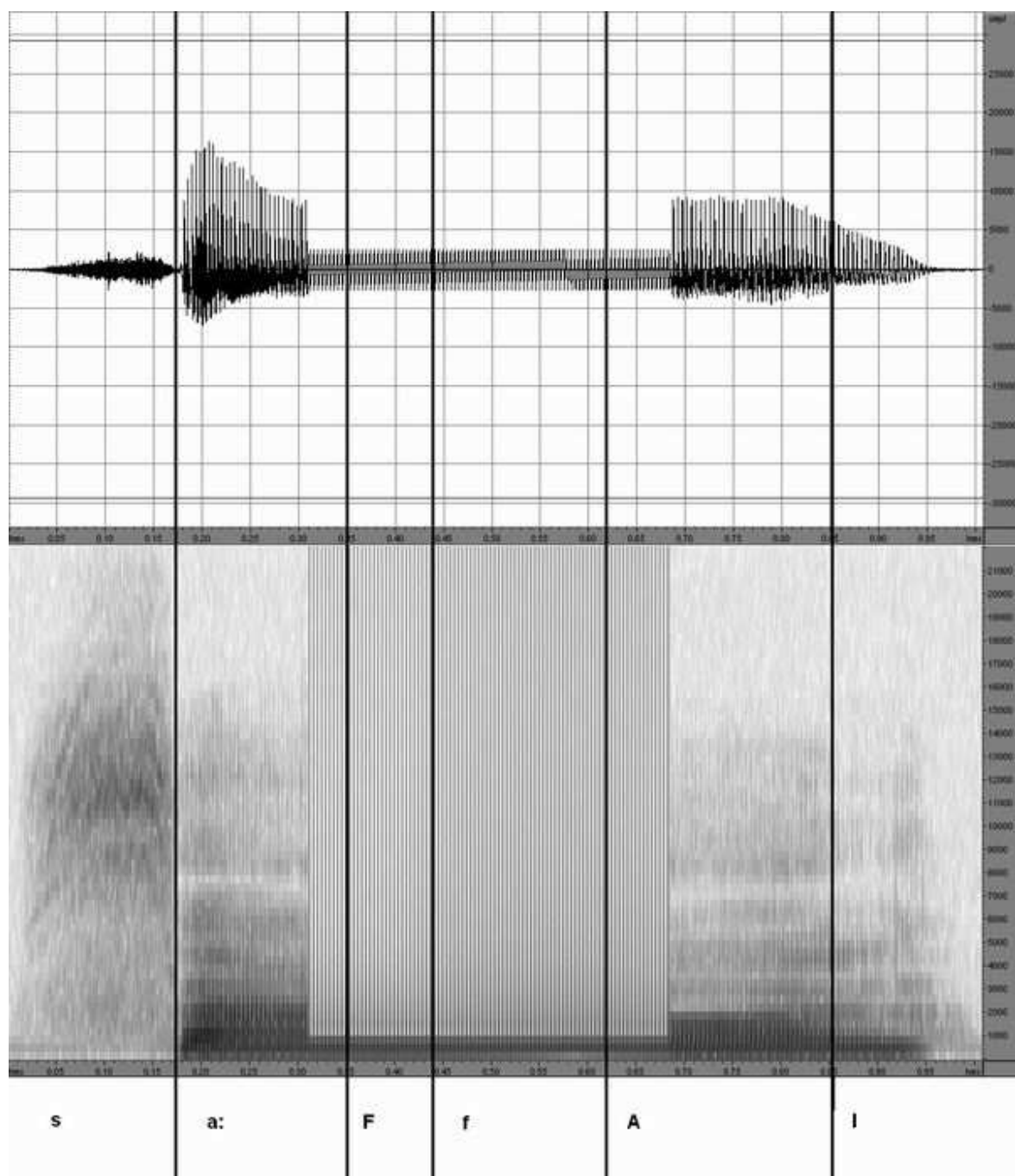
3.6. ábra. A bükkfa szó hullámformája és hangszínképe 50 százalékos kivágás esetén, beszédzajjal feltöltve



3.7. ábra. A bükkfa szó hullámformája és hangszíne 50 százalékos kivágás esetén, fűrészjellel feltöltve



3.8. ábra. A számfal szó hullámformája és hangszínképe 25 százalékos kivágás esetén, beszédzajjal feltöltve



3.9. ábra. A számfal szó hullámformája és hangszíneképe 25 százalékos kivágás esetén, fűrészjellel feltöltve

3.7. A szubjektív lehallgatás

A beszédértés vizsgálatára rengeteg féle kísérletet végeztek már, ezekben a résztvevő személyeknek a rögzített hanganyag (beszédhangok sorozata, dallamminták, változó intenzitású szótagok, zajos mondatok stb.) meghallgatása és valamilyen reakció, válaszadás a feladatuk. Ennek a kísérletnek a típusa izolált szavas felismerés. [1]

A teszthez nem volt másra szükség mint a hangfájlokra, egy számítógépre, egy lejátszóprogramra és egy fejhallgatóra. Így a vizsgálat helyfüggetlenül szinte bárhol elvégezhető, ahol rendelkezésre állnak ezek az eszközök. A kísérleti személyek ismerőseim és családtagjaim köréből kerültek ki, összesen 15 ember vett részt benne. A legtöbben két alkalommal hallgatták meg az anyagot, először a 25, majd az 50, vagy először az 50, és utána a 25 százalékos vágású hangfájlokat. Minden esetben egyszerre csak egy hallgatóval folyt a vizsgálat, csendes, nyugodt környezetben, hogy semmi se vonja el a figyelmüket. Először rövid tájékoztatásban részesültek a mérés céljáról és folyamatáról, majd kaptak egy formanyomtatványt, amin kettő üres táblázat volt, mindegyik negyven rubrikával, ezekbe kellett a hallottakat leírni. Előtte még meghallgattak két-három szót, gyakoroltatás céljából, ezek nem kerültek bele az eredményekbe.

Két részből állt a hallgatás. Az elsőben az igazi spondeuszok felismerése volt a feladat, a másodikban a hamis spondeuszoké. A hallgatók tájékoztatva lettek arról, hogy az első részben értelmes, a másodikban értelemmel nem rendelkező szavakat fognak hallani, de az első részben sem kell értelmes szavakat írniuk, csak azt amit hallottak, vagy hallani véltek. Egyesével történt a hangfájlok lejátszása, csak akkor lett elindítva a következő, ha a hallgató végzett az írással, kérésre pedig egyszer újra le lett játszva az aktuális szó. Abban az esetben, ha egyáltalán nem tudták értelmezni a hallottakat, kihagyták azt a rubrikát, és áttértünk a következő szóra.

Egy vizsgálat alkalmával vagy csak 25, vagy csak 50 százalékos vágású szavakat hallgattak. Mindegyikből három különböző anyag lett összeállítva, úgy hogy harmadrészt tartalmazott kivágott, beszédzajjal feltöltött és fűrészzellel feltöltött felvételeket, de minden szó csak egyszer szerepelt benne. Ezek véletlenszerű sorrendben lettek lejátszva. A 25 százalékos vágású szavakat 15 ember hallgatta meg, így minden szóra 5 eredmény született, az 50 százalékos kilencen hallgatták meg, ezeknél minden esetben 3 eredményt kaptunk.

4. KIÉRTÉKELÉS

Kiértékeléskor a szavakat két részre bontjuk, hogy külön lehessen megvizsgálni a visszafele és az előrefele irányuló folyamatokat. Valójában viszont a két tagot nem lehet teljesen különválasztani, főleg értelmes szavak esetében.

Egy hallgató miután meghallgatta a vágott anyagot, meghallgatta az eredeti, vágás nélküli hangfelvételeket is. Kivétel nélkül helyesen azonosította nemcsak az értelmes, de az értelem nélküli szavakat is. Ez azt bizonyítja, hogy a felvett szavak tisztán ejtettek és felismerhetők.

4.1. A spondeuszok észlelése

A magánhangzók és a szókezdő, illetve szóvégi mássalhangzók felismerése közel 100 százalékos pontosságú. Ez annak köszönhető, hogy a magyar kiejtésben a magánhangzók képzése teljes értékű, vagyis függetlenül a szóban, mondatban elfoglalt helyétől, a magánhangzó nem redukálódik. Napjaink spontán beszédének vizsgálata azonban egyértelműen igazolta, hogy a szükséges hangminőségek helyett mintegy 70 %-ban semleges magánhangzók találhatók. [1] Ebben az esetben viszont nem spontán beszédéről van szó, hanem tisztán kiejtett szavakról. Magánhangzók esetében, ha téves volt az észlelés, a hasonló képzéseket cserélték fel: e:→i:, i:→e:, o:→u, o→a.

A spondeuszok felismerésénél erősen meghatározó volt, hogy mennyire ismert, gyakori az adott szó a nyelvben. Például a *központ* szó első tagját közel 100%-ban azonosították helyesen, a *közkincs* szóban ugyanazt a szótagot viszont csak körülbelül 40%-ban. A *színkép* szót akik ismerik, használják, jó eséllyel azonosították helyesen, akik viszont nem ismerték, egyszer sem találták el, helyette más ismert szót írtak (pl.: széttép). Előfordult, hogy hozzátettek egy mássalhangzót a szó végéhez, így lett például a *nappá*-ból *naptár*.

A spondeuszok százalékos felismerésének összesítését mutatja a **4.1. táblázat**. Látszik, hogy nagy különbségek nincsenek a kivágott, és a beszédzajjal illetve a fűrészzel feltöltött szavak észlelésében. 25%-os vágás esetén a beszédzajjal való feltöltés kis mértékben nagyobb, 50%-osnál kisebb felismerést mutat, de ez az eredmény egyáltalán nem szignifi-

káns. Gyakorlatilag nincs különbség a különböző méretű vágások között sem. Az első tagok felismerése valamivel biztosabb, mint a másodiké, de ez is csak 8-10 % körüli különbség.

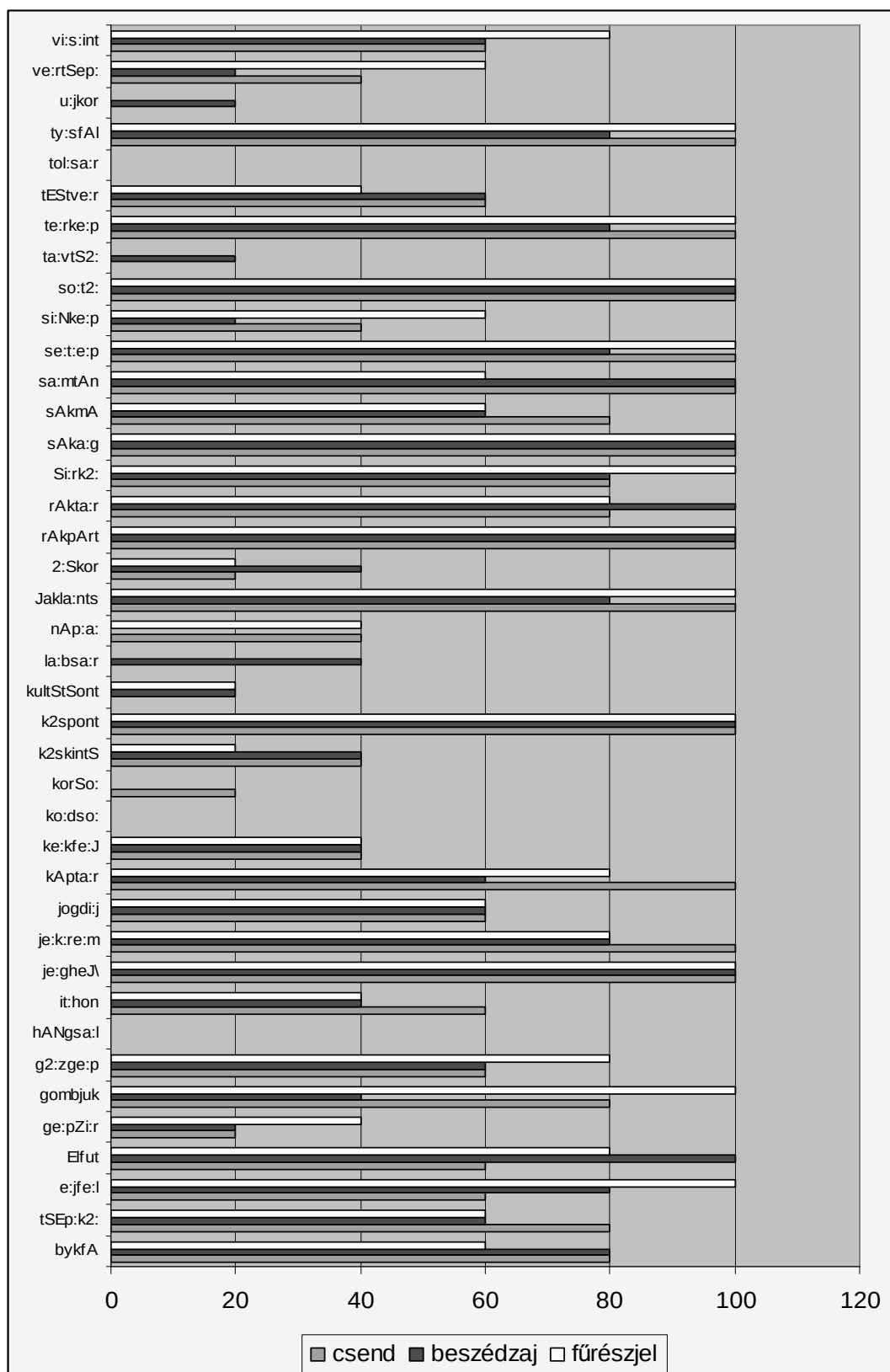
4.1. táblázat Spondeuszok felismerésének összesítése százalékban kifejezve

	25%-os vágás			
	csend	beszédzaj	fűrészel	összes
teljes szó	59	55,5	59	57,8
első tag	70,5	66,5	69	68,7
második tag	60,5	58	61	59,8

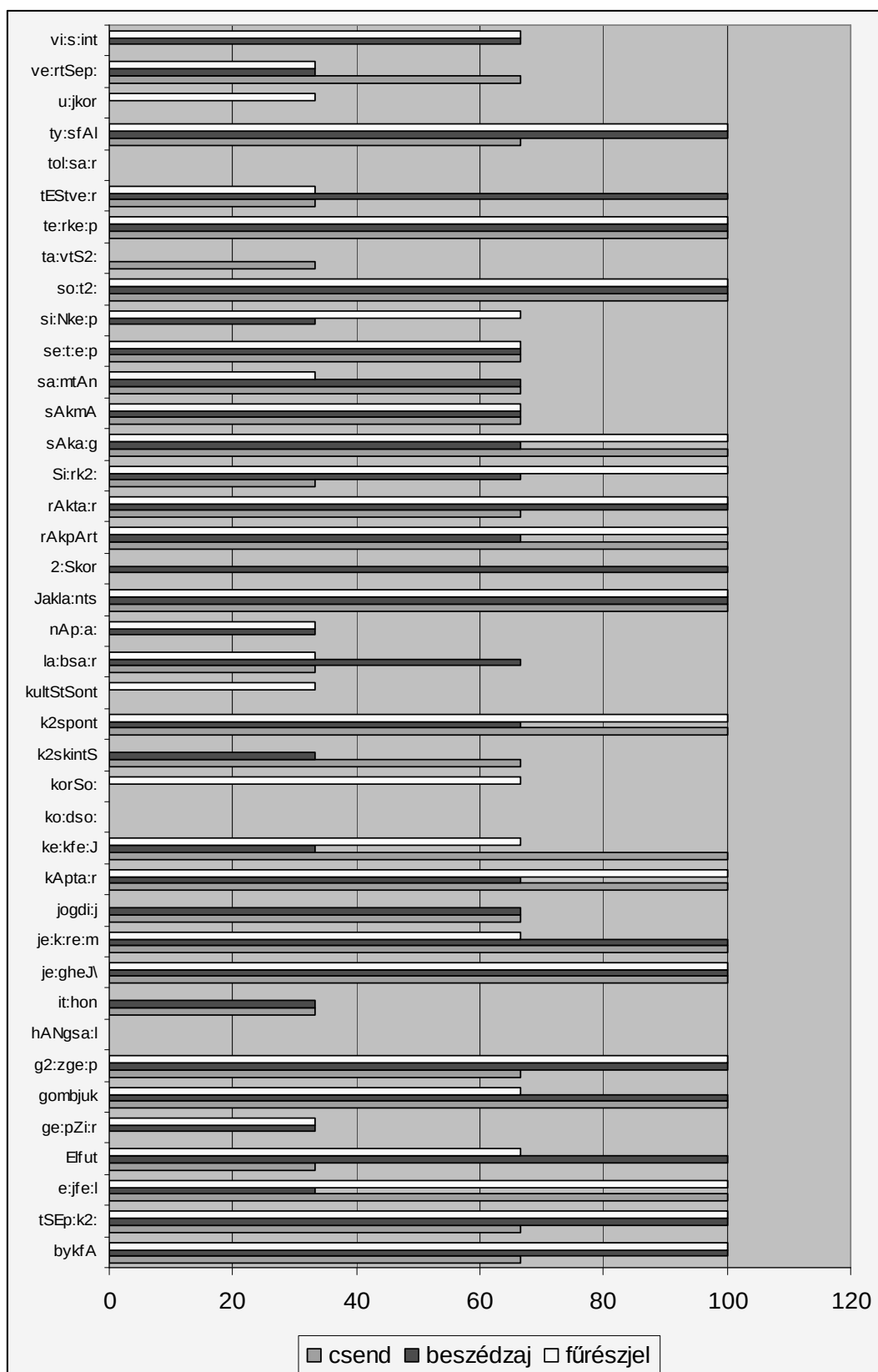
	50%-os vágás			
	csend	beszédzaj	fűrészel	összes
teljes szó	53,3	60	59,2	57,5
első tag	61,7	70,9	65,8	66,1
második tag	54,2	65	64,2	61,1

A **4.1.-4.2. ábrákon** a teljes szó helyes azonosításának eredményei láthatóak. Azokban az esetekben, amikor a szó közepén csak egy rövid, vagy egy hosszú mássalhangzó van (szótő, szakág), a felismerés 100%-os. A *nappá* helyett viszont sokszor *naptárt* azonosítottak. A legrosszabb azoknak a felismerése, amik nem túl gyakori szavak a köznyelvben: *hangszál, kódszó, tollszár, újkor*.

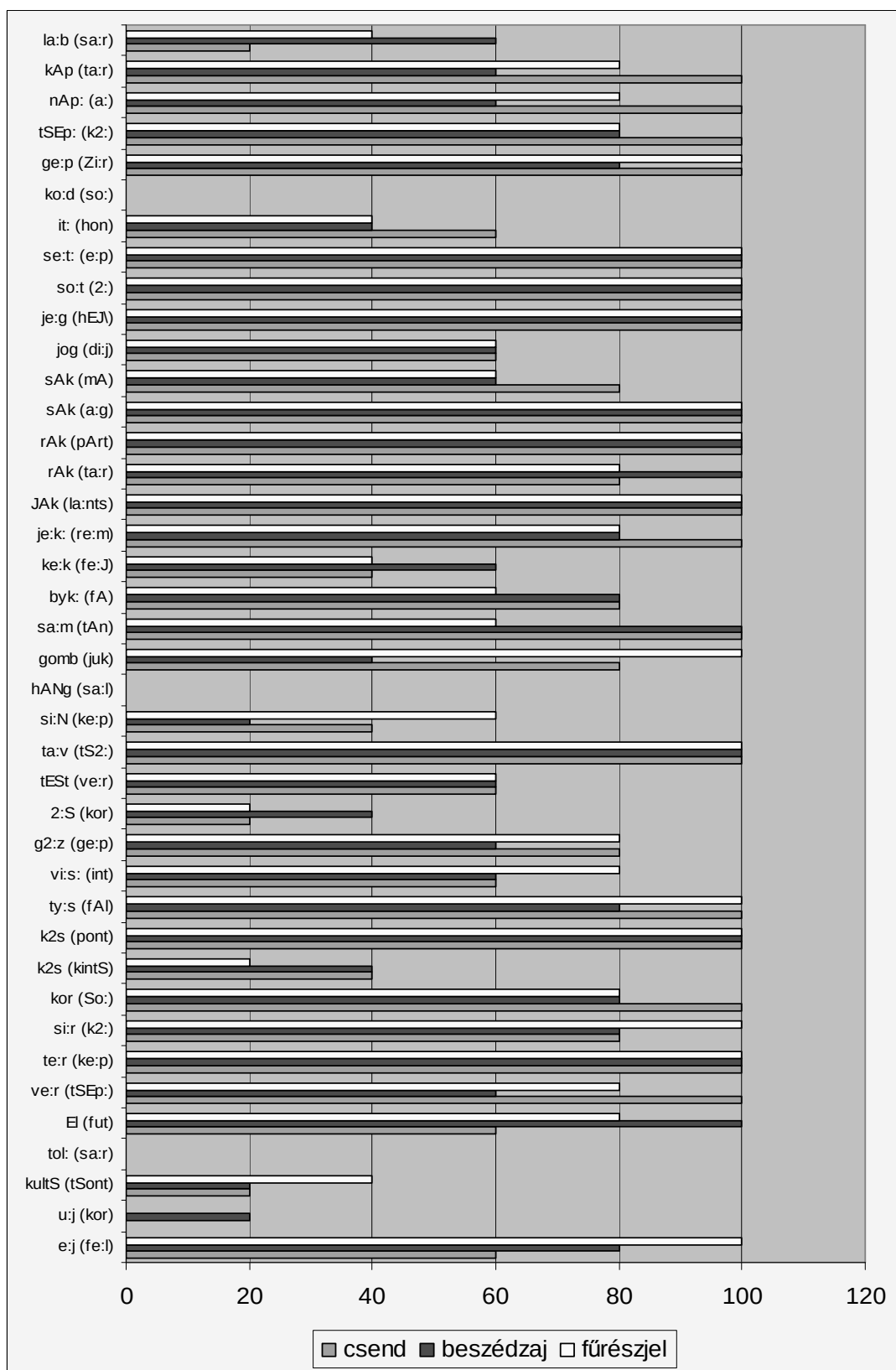
A **4.3.-4.6. ábrákon** a külön láthatóak a szavak első, illetve második tagjának helyes felismerési eredményei százalékban kifejezve, a **4.7.-4.10. ábrákon** pedig a szó közepén lévő, azaz a kivágott mássalhangzók azonosításai. Az értékekből messzemenő következtetést nem lehet levonni, mert egyes mássalhangzók csak egy szóban szerepeltek, de ezeken a grafikonokon is látszik, hogy a feltöltéssel általánosságban nem javult az érthetőség. Az első tagokban a legbizonytalanabb a zöngés zárhangok (/b/, /d/), a réshangok (/z/, /s/, /S/), a közelítőhangok (/l/, /j/) és az /N/ hang azonosítása, a második tagokban a /z/, /s/, /S/, /tS/ hangoké. A zajszerű, turbulens zörejt tartalmazó réshangok (/v/, /f/, /z/, /s/, /Z/, /S/) felismerhetősége valamelyest javult beszédzaj hatására.



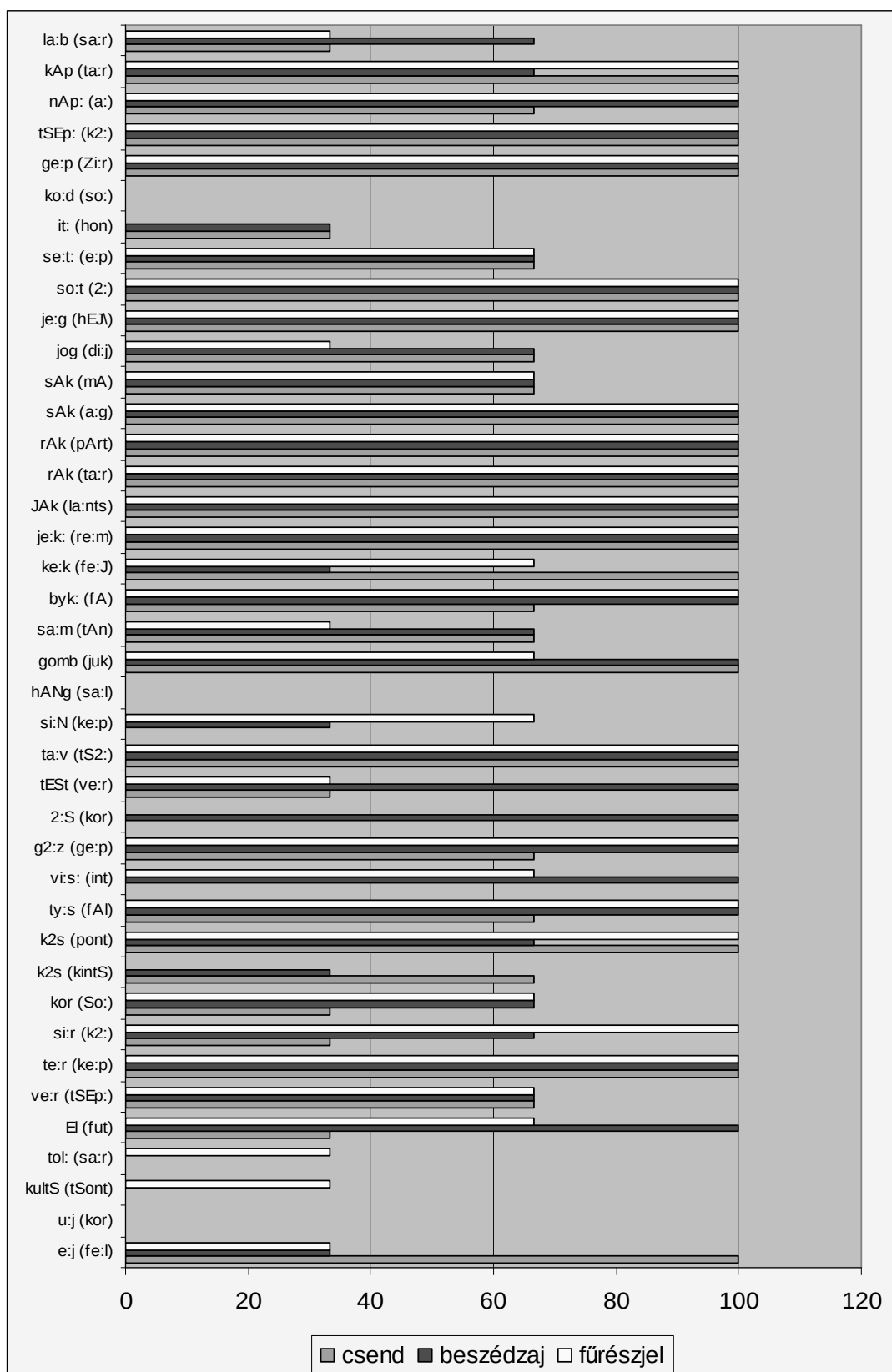
4.1. ábra. 25%-os vágású spondeuszok felismerése



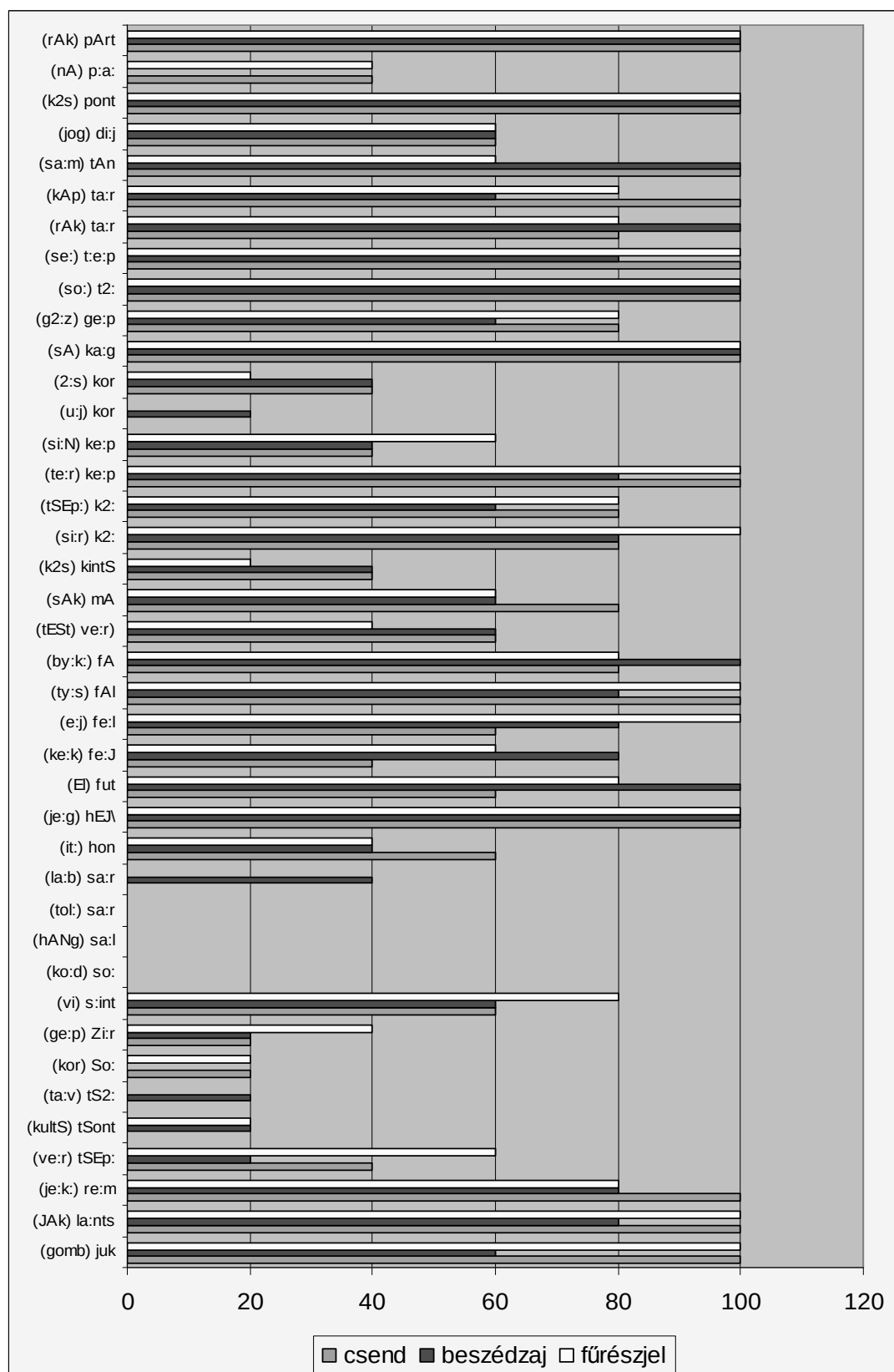
4.2. ábra. 50%-os vágású spondeuszok felismerése



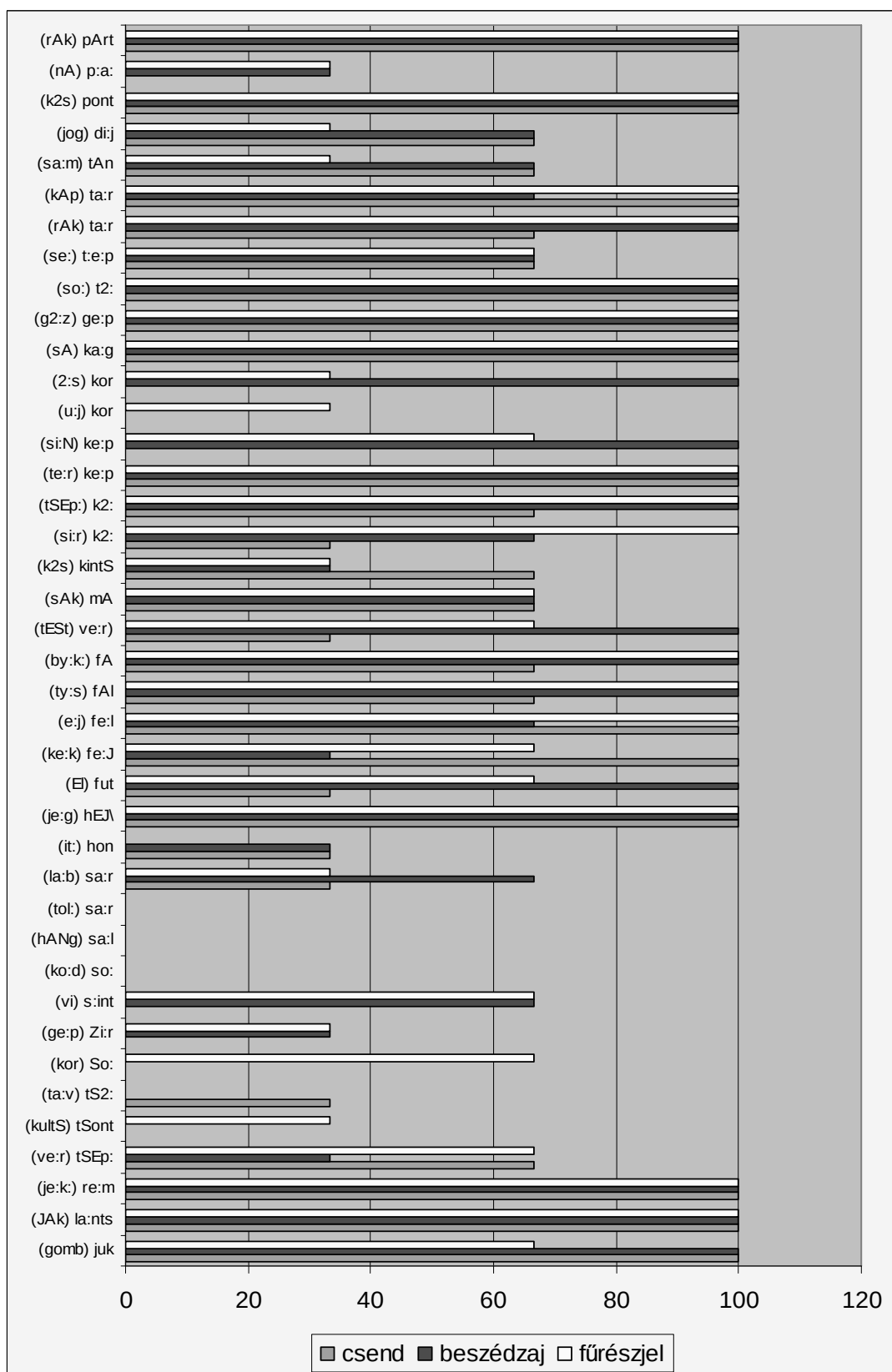
4.3. ábra. 25%-os vágású spondeuszok első tagjának észlelési eredményei



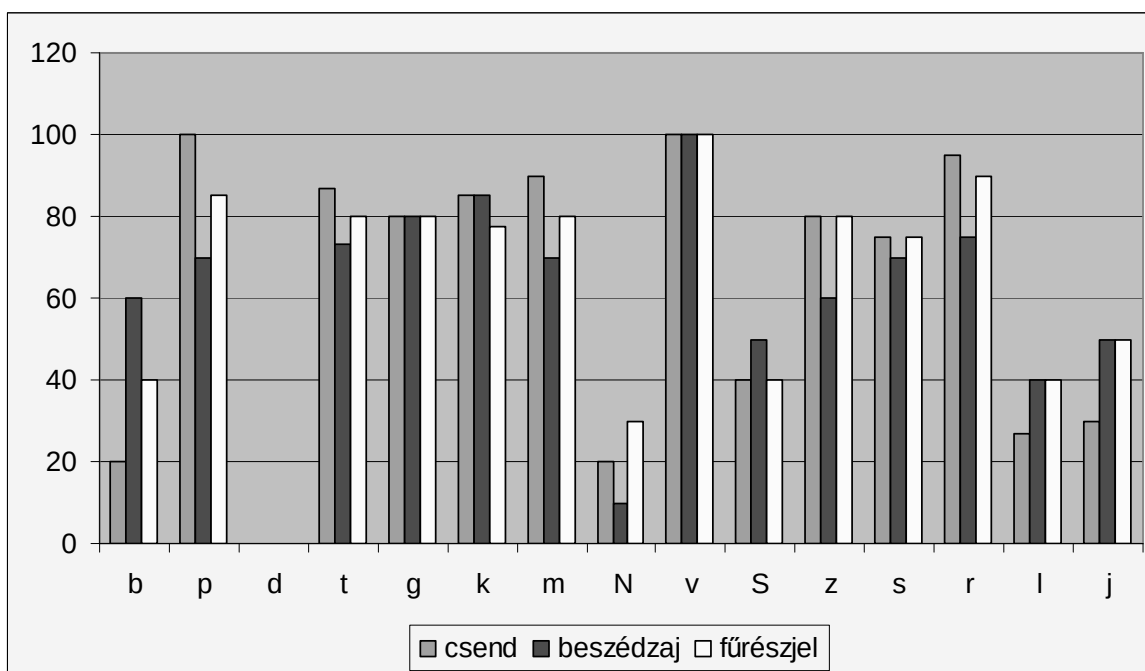
4.4. ábra. 50%-os vágású spondeuszok első tagjának észlelési eredményei



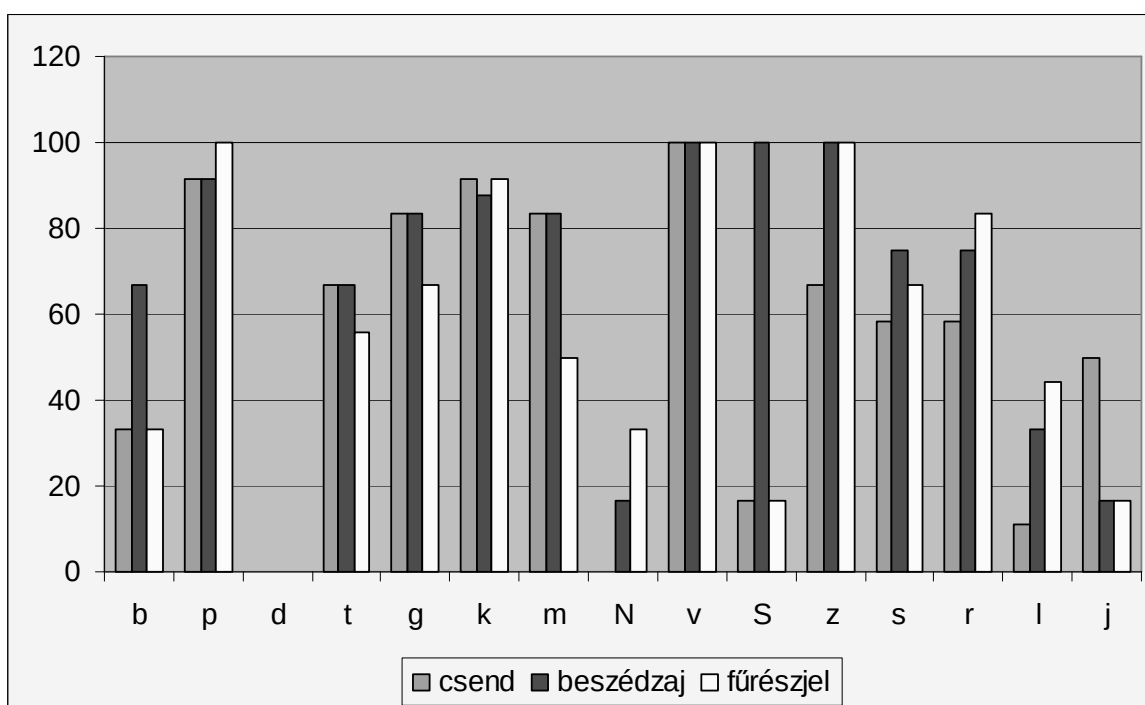
4.5. ábra. 25%-os vágású spondeuszok második tagjának észlelési eredményei



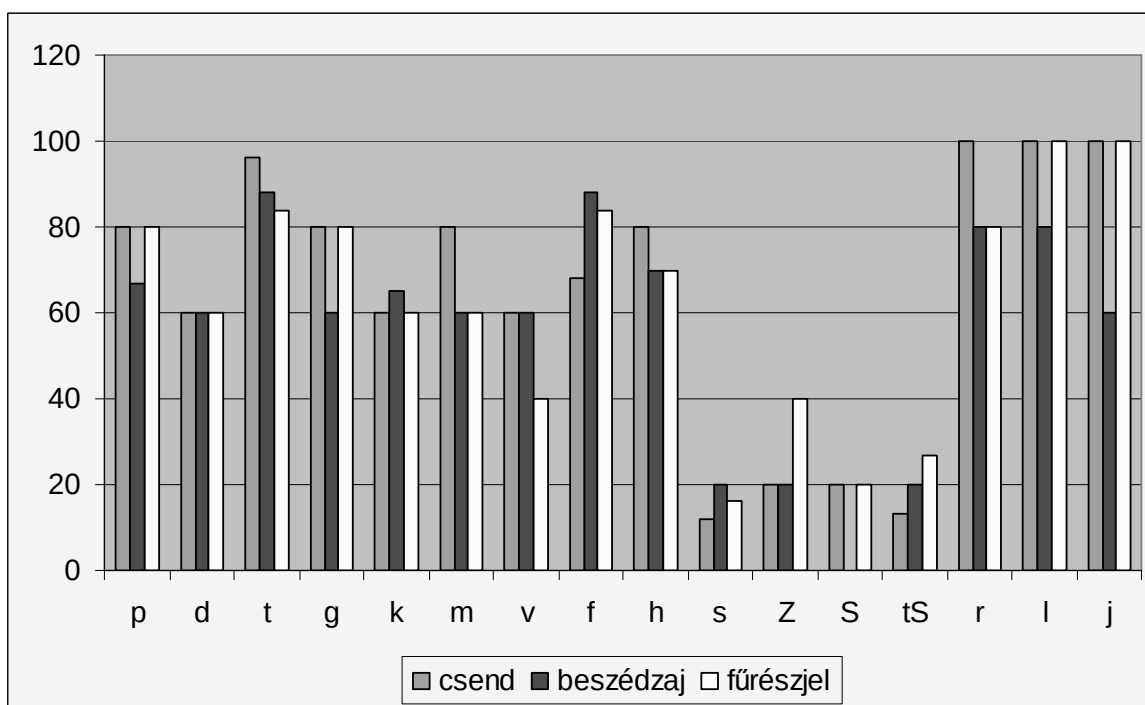
4.6. ábra. 50%-os vágású spondeuszok második tagjának észlelési eredményei



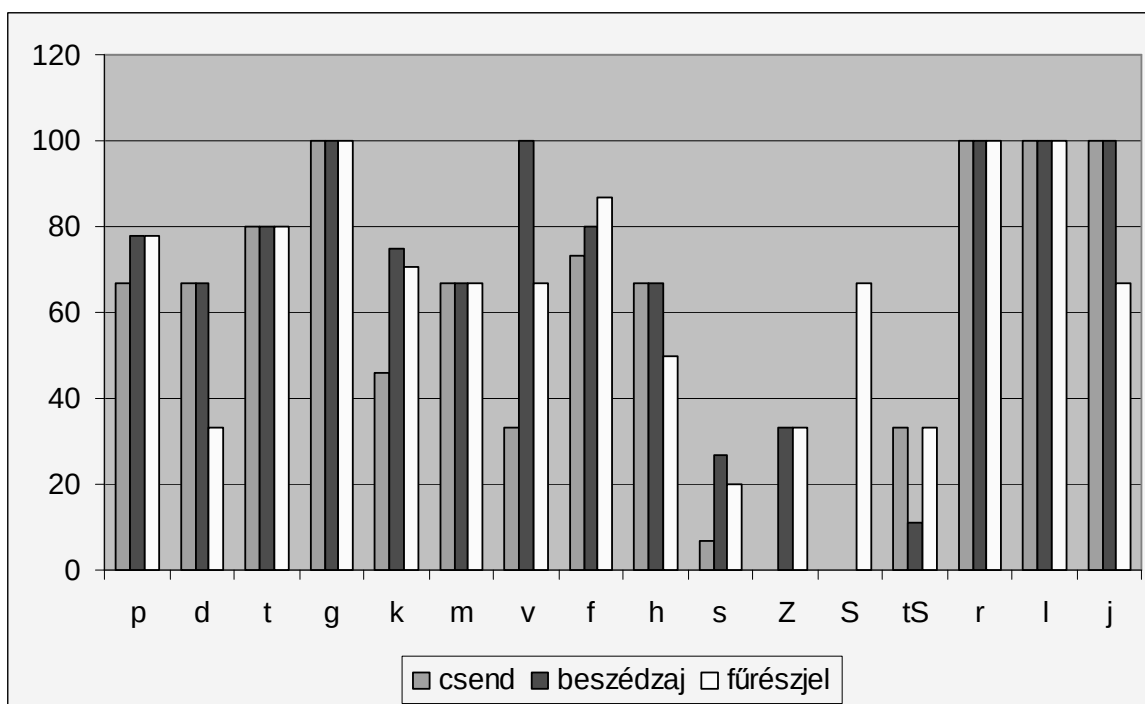
4.7. ábra. Mássalhangzók észlelési eredményei a 25%-os vágású spondeuszok első tagjában



4.8. ábra. Mássalhangzók észlelési eredményei az 50%-os vágású spondeuszok első tagjában



4.9. ábra. Mássalhangzók észlelési eredményei a 25%-os vágású spondeuszok második tagjában



4.10. ábra. Mássalhangzók észlelési eredményei az 50%-os vágású spondeuszok második tagjában

4.2. A hamis-spondeuszok észlelése

A mentális lexikon jelentőségét a beszédfelismerésben mutatja az az eredmény is, hogy az értelemmel nem rendelkező szavak felismerése csupán 10% körüli, míg az értelmeseké majdnem 60%. Bár a hallgatók tudták, hogy nem értelmes szavakat fognak hallani, mégis sokszor azt vélték felfedezni benne.

A **4.2. táblázat** mutatja a hamis-spondeuszok felismerésének összesített eredményeit. A teljes szó és a külön a tagok azonosításában itt nagyobb különbség van, mint a spondeuszoknál. Ez is azt bizonyítja, hogy itt nagyobb a szerepe az akusztikai szerkezet elemzésének. Az első tagok azonosítása itt is valamivel biztosabb, de ez a különbség is 10% alatti minden esetben. A 25 és 50%-os vágású szavak eredményeiben - a várakozásokkal ellentétben - itt is elhanyagolható a különbség. A beszédzaj itt is 25%-os esetben kicsit inkább rontott az érthetőségen, 50%-osban kismértékben javított, de ez sem jelentős.

A magánhangzók és a szókezdő, szóvégi mássalhangzók azonosítása ezeknél is közel 100%-os, de itt már kicsit gyakrabban fordult elő, hogy mást értettek, általában amikor értelmes szóra egészítették ki a hallottakat. Előfordult, hogy más mássalhangzót értettek ennek hatására a szó végén vagy elején (pl.: kaplanc helyett nyaklanc, gőzfény helyett kötvény, térfény helyett kötvény), hozzátettek egy mássalhangzót a szó végéhez vagy az elejéhez (öscsont→központ), vagy pedig elhagytak egyet (kékfél→éjfél). Ritkán ezeknél is megváltozott a magánhangzó (színgép→széttép).

4.2. táblázat Hamis-spondeuszok felismerésének összesítése százalékban kifejezve

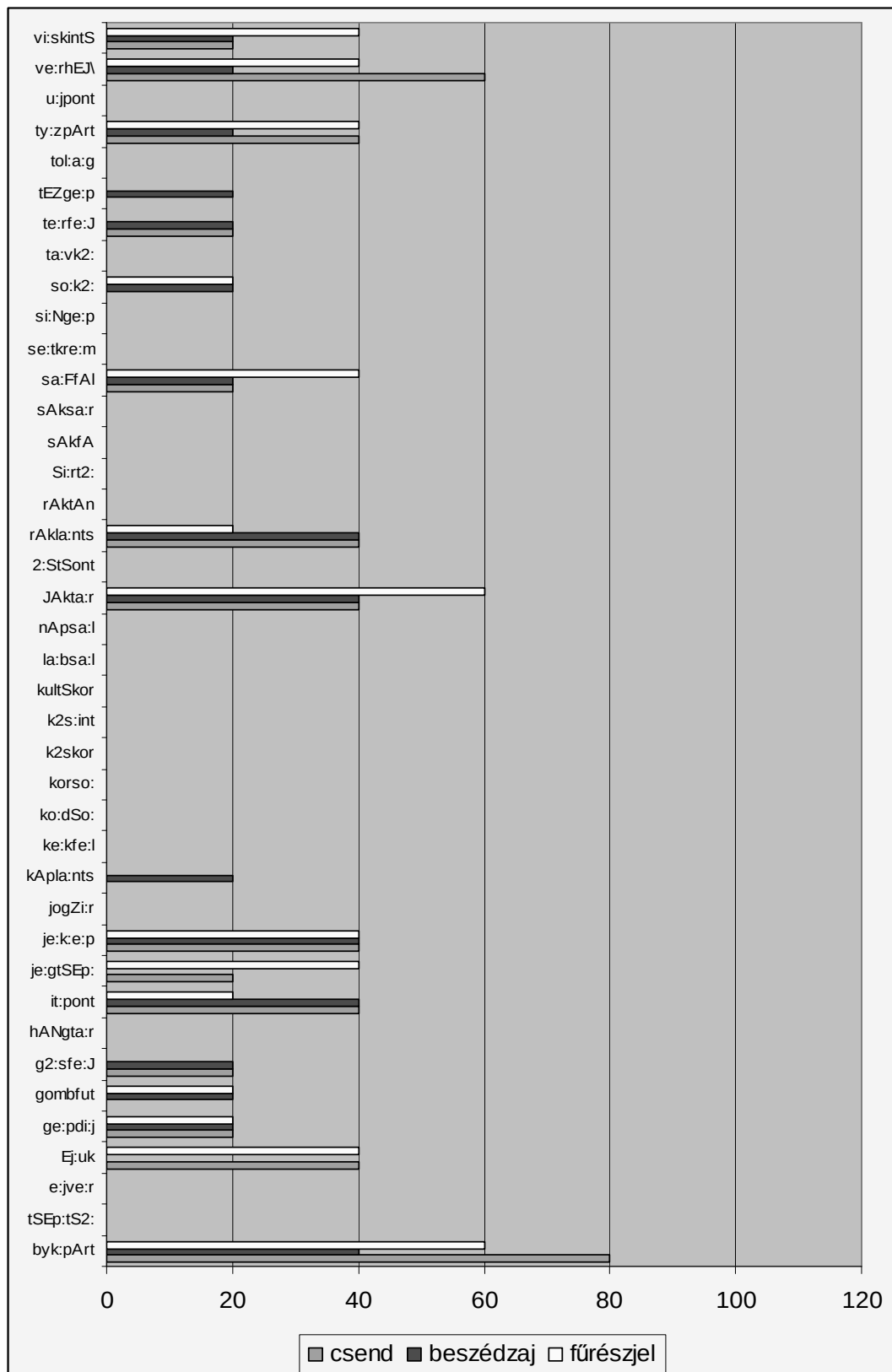
	25			
	csend	beszédzaj	fűrészel	összes
teljes szó	12,8	10,7	12,8	12,1
első tag	37,4	31,3	37,4	35,4
második tag	27,2	26,7	32,8	28,9

	50			
	csend	beszédzaj	fűrészel	összes
teljes szó	7,7	10,3	13,7	10,5
első tag	33,3	35	42,8	37
második tag	24,8	31,6	33,3	29,9

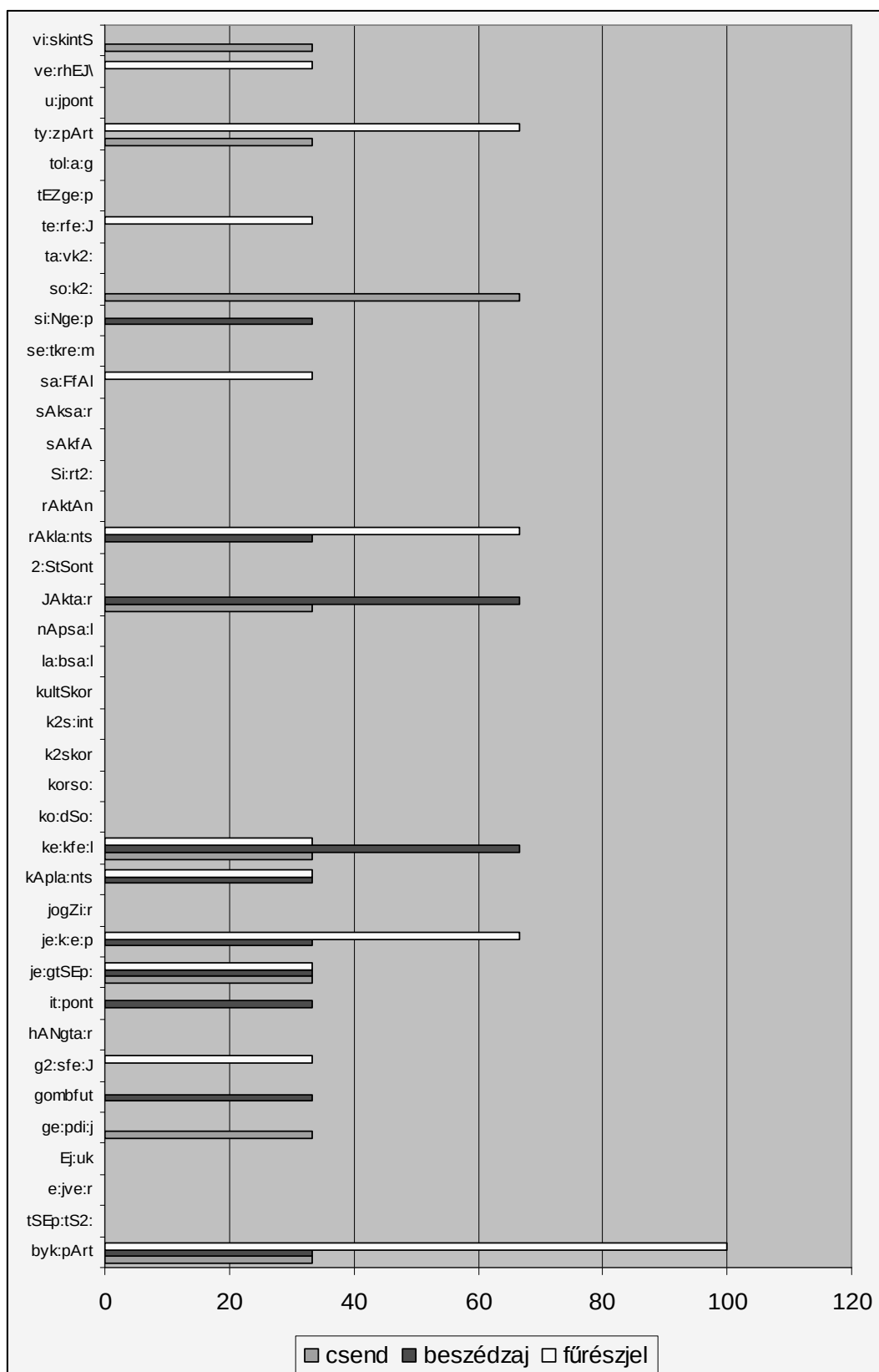
A **4.11.-4.12. ábrákon** a teljes szó helyes azonosításának eredményei láthatóak. Azokban az esetekben, amikor a szó közepén csak egy rövid, vagy egy hosszú mássalhangzó van (*szókő*, *tollág*, *közzsint* /k2s:int/, *ejjuk*) itt már nem olyan jó a felismerés. A *szókő* helyett a legtöbbször itt is *szótő*-t értettek.

A **4.13.-4.16. ábrákon** a külön láthatóak a szavak első, illetve második tagjának helyes felismerési eredményei százalékban kifejezve, a **4.17.-4.20. ábrákon** pedig a szó közepén lévő, azaz a kivágott mássalhangzók azonosításai. A hamis-spondeuszok esetén is a legbizonytalanabb az első tagokban a zöngés zárhangok (/b/, /d/), a réshangok (/z/, /s/, /S/), a közelítőhangok (/l/, /j/) és az /N/ hang azonosítása, a második tagokban a /z/, /s/, /S/, /tS/ hangoké.

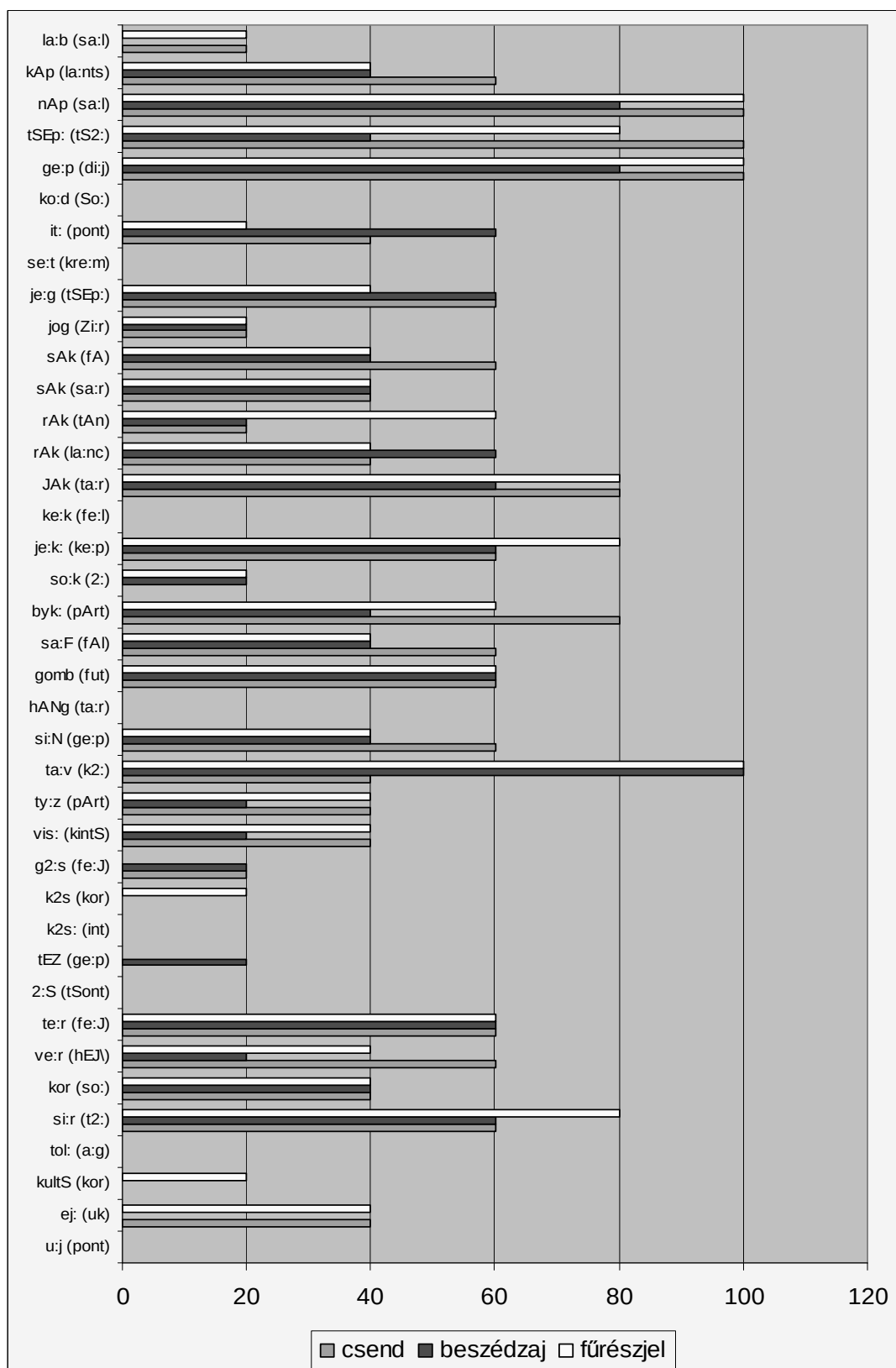
A hamis-spondeuszok észlelési eredményei alapján is egyértelműen kijelenthető, hogy a beszédzajjal, vagy fűrészzel való feltöltéssel nem sikerült javítani az érthetőséget. Ezek sokszor inkább zavaró hatásúak, bár összességében az érthetőséget nem is rontották. A beszédzajt elég sokszor /f/ hangként azonosították, ennek a hangnak a frekvenciaszerkezete hasonlít legjobban a beszédzajára. A fűrészelet sok esetben /p/, ritkábban /b/ hangnak észlelték. A kivágott szavaknál is sokszor felpattanó zárhangokat értettek, legtöbbször /p/, ritkábban /b/, /t/ hangot. Ez logikus, hiszen a csend megfeleltethető a felpattanó zárhangok zárfelpattanási szakaszának.



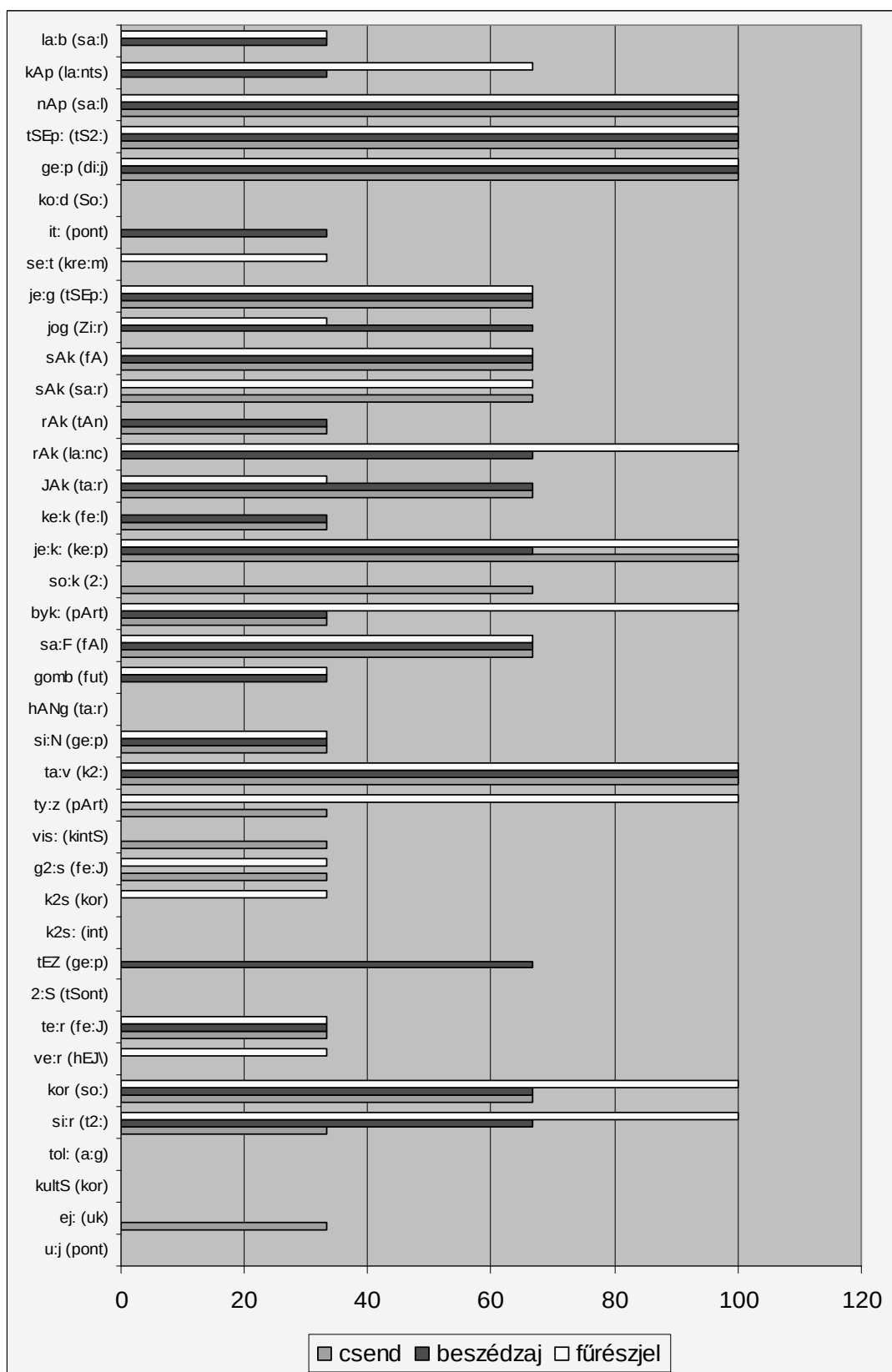
4.11. ábra. 25%-os vágású hamis-spondeuszok felismerése



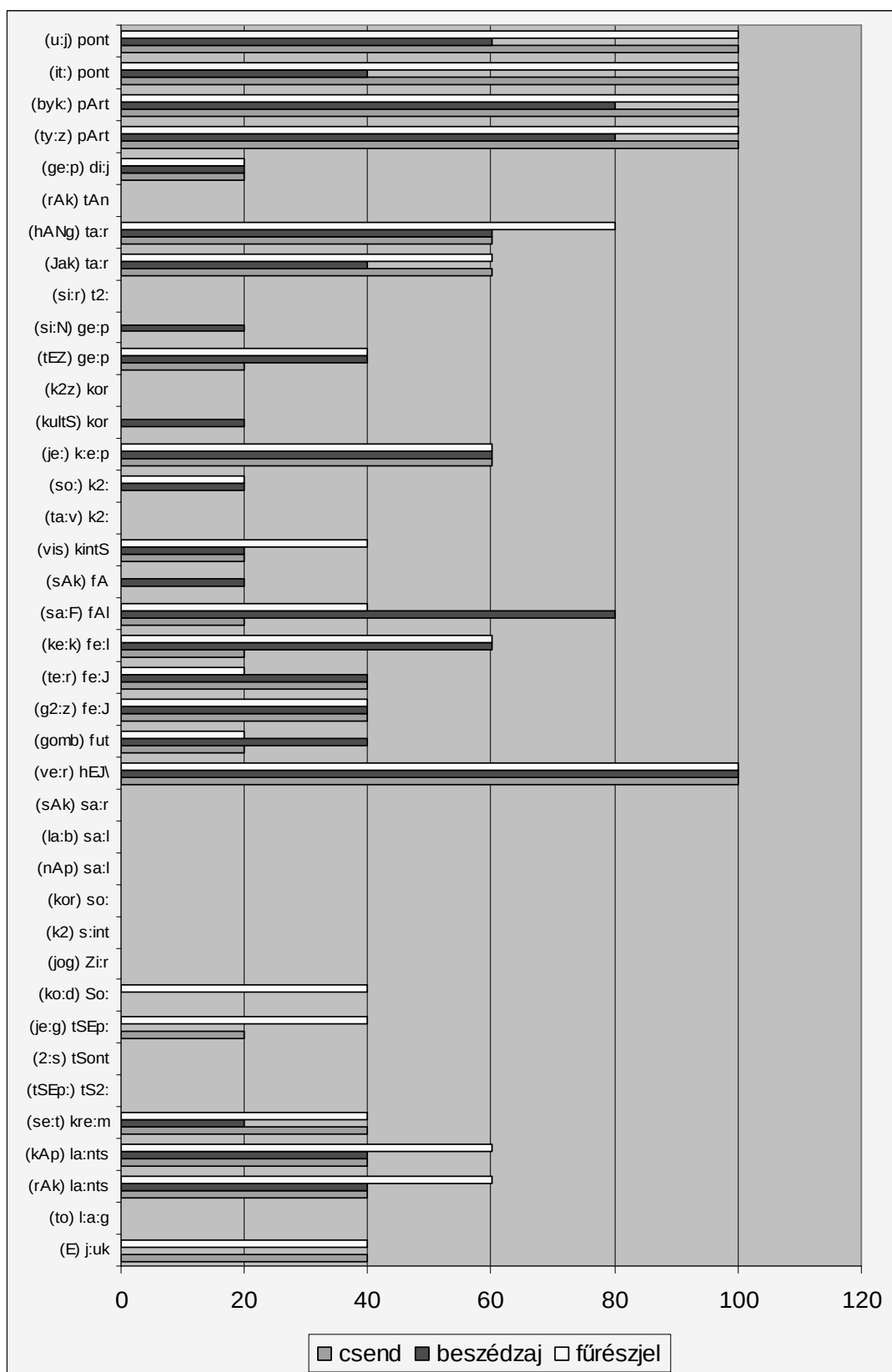
4.12. ábra. 50%-os vágású hamis-spondeuszok felismerése



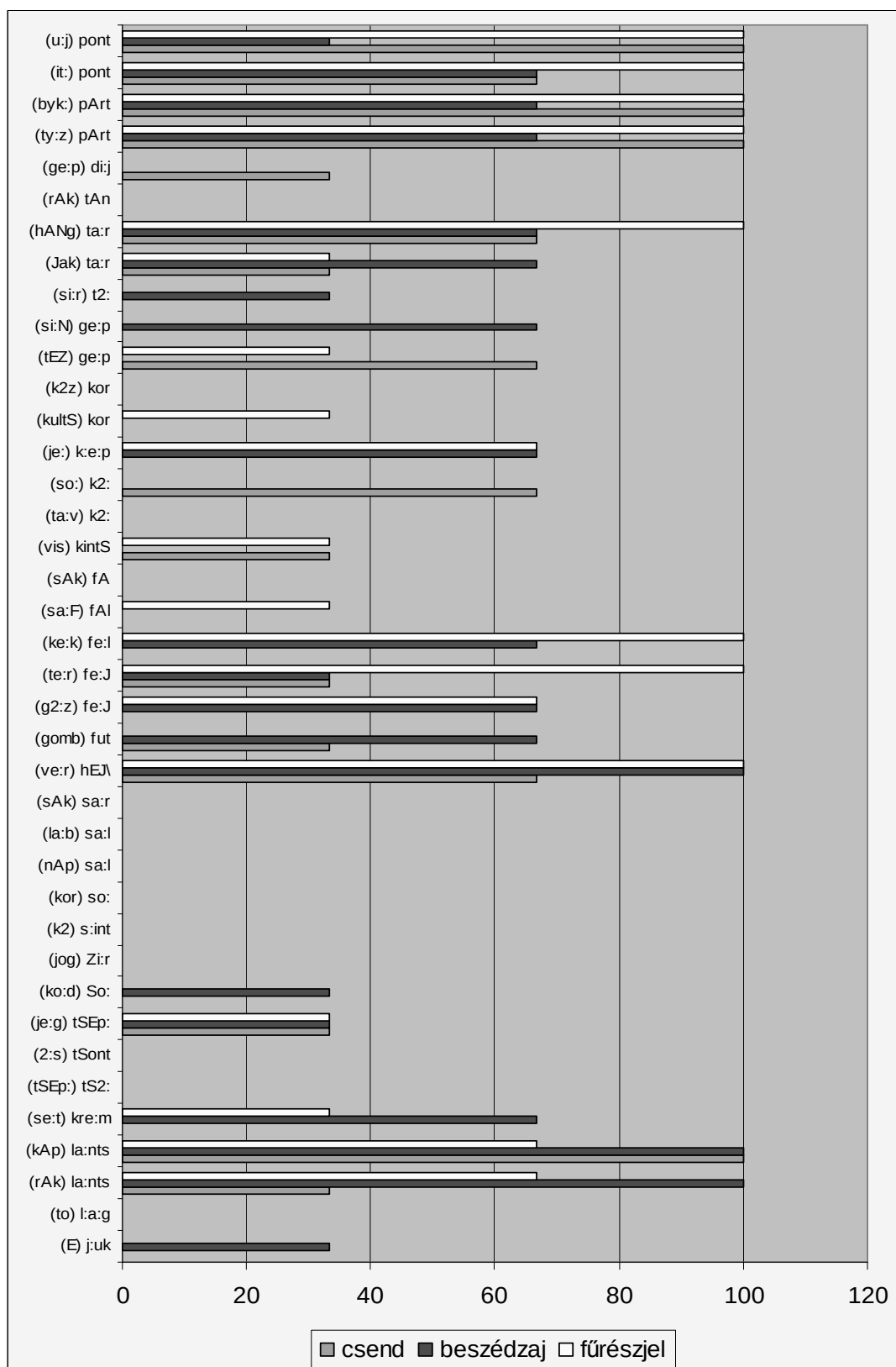
4.13. ábra. 25%-os vágású hamis-spondeuszok első tagjának észlelési eredményei



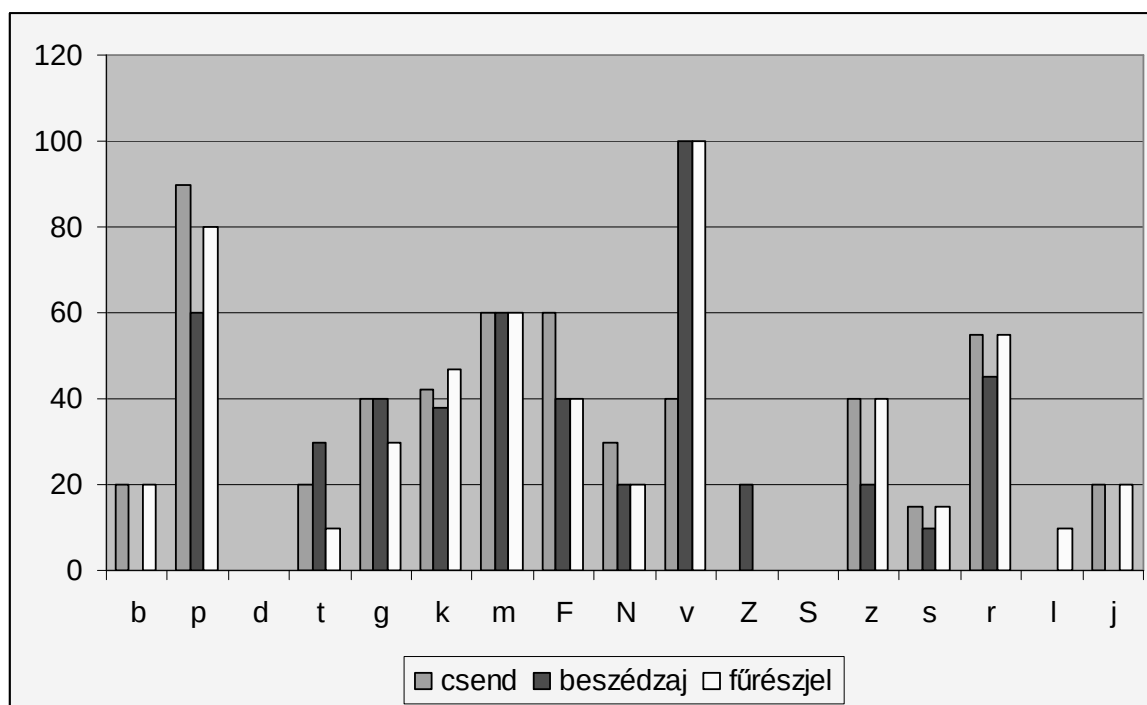
4.14. ábra. 50%-os vágású hamis-spondeuszok első tagjának észlelési eredményei



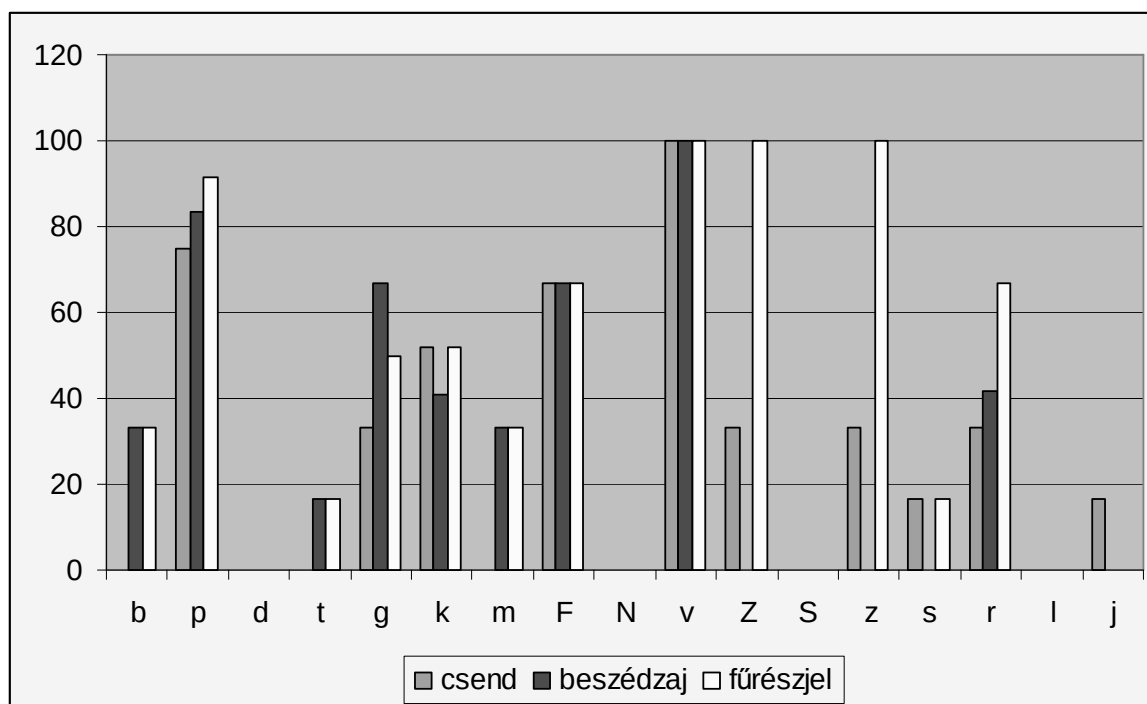
4.15. ábra. 25%-os vágású hamis-spondeuszok második tagjának észlelési eredményei



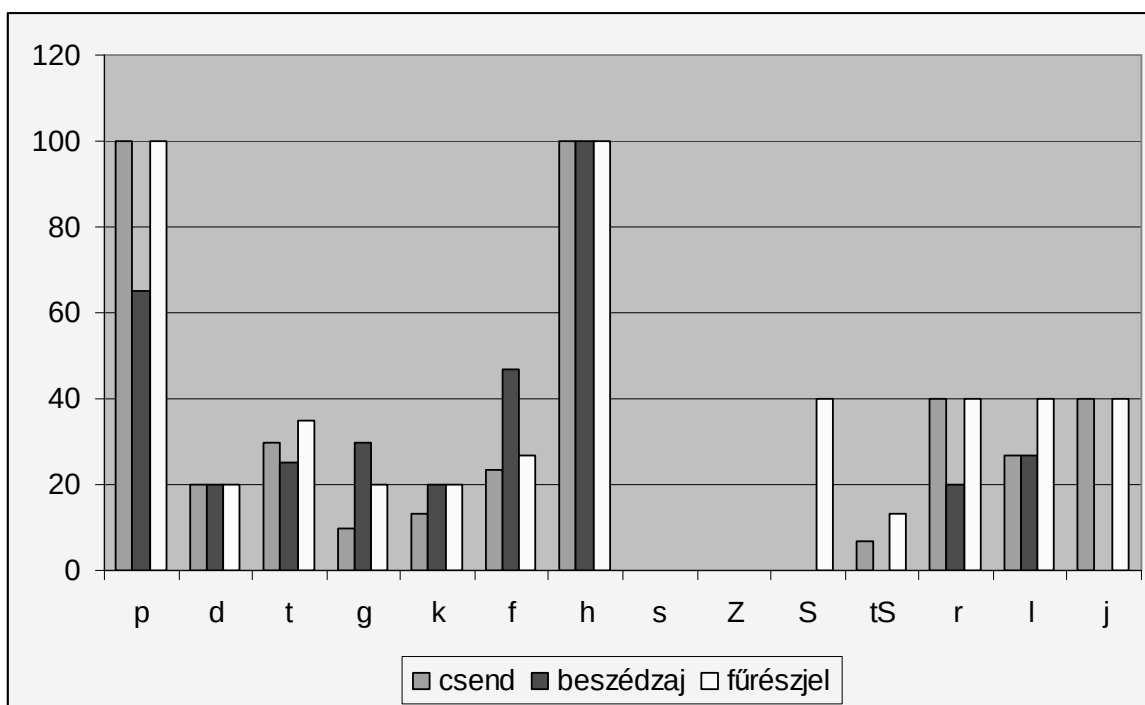
4.16. ábra. 50%-os vágású hamis-spondeuszok második tagjának észlelési eredményei



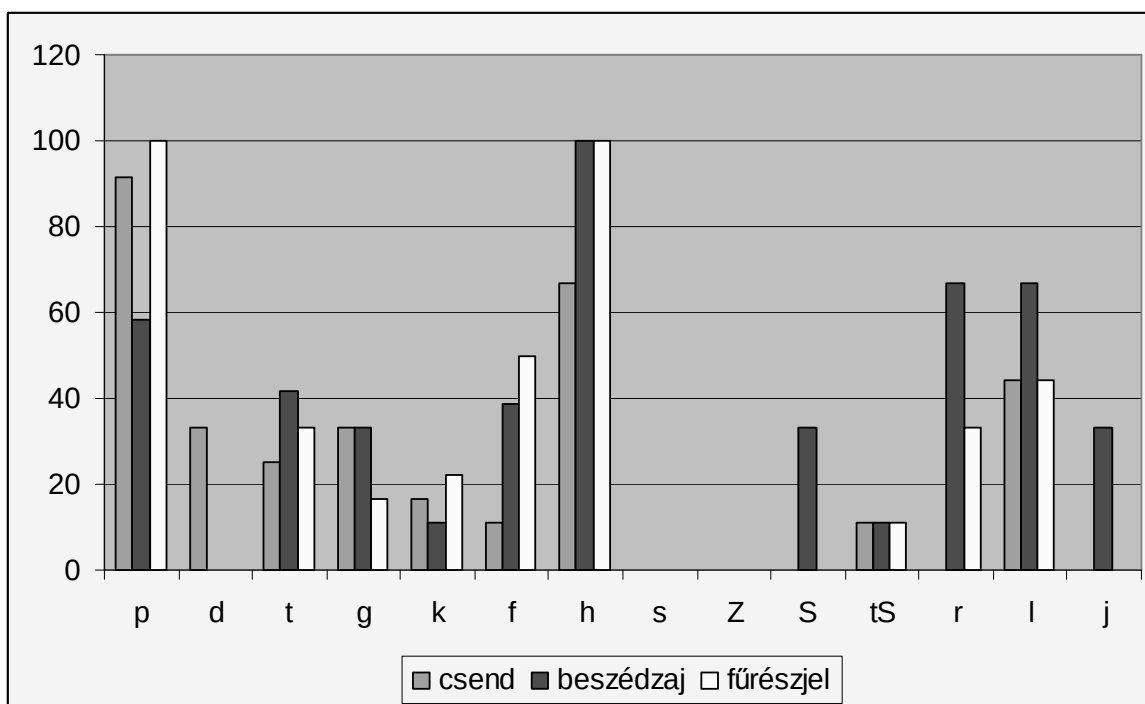
4.17. ábra. Mássalhangzók észlelési eredményei a 25%-os vágású hamis-spondeuszok első tagjában



4.18. ábra. Mássalhangzók észlelési eredményei az 50%-os vágású hamis-spondeuszok első tagjában



4.19. ábra. Mássalhangzók észlelési eredményei a 25%-os vágású hamis-spondeuszok második tagjában



4.20. ábra. Mássalhangzók észlelési eredményei az 50%-os vágású hamis-spondeuszok második tagjában

5. ÖSSZEFOGLALÁS

Bár nem áll rendelkezésre nagy számú vizsgálati eredmény, de ezek is egyértelműen bebizonyították, hogy ezekkel a módszerekkel nem lehet javítani a beszéd érthetőségét. Viszont az elkészült beszédatbázis lehetőséget ad arra, hogy további eljárásokat lehessen kipróbálni. A kivágott közepű szavaknál lehetne például valamilyen fajta interpolációt alkalmazni. Többféle interpolációs eljárást is ki lehetne próbálni (pl.: lineáris, logaritmikus, beszéd spektruma alapján történő stb.) Abban az esetben pedig, ha valamelyik módszer eredményesnek tűnik, meg lehetne vizsgálni spontán beszédnél is a hatását.

6. IRODALOMJEGYZÉK

- [1] Gósy Mária: *Fonetika, a beszéd tudománya*
Osiris Kiadó, 2004.
- [2] Gósy Mária: *Pszicholingvisztika*
Corvina, 1999.
- [3] Dr. Vicsi Klára: *Beszédkommunikáció* (Jegyzettermű)
- [4] Ray D. Kent – Charles Reed: *The Acoustic Analysis of Speech*
Singular Publishing Group, INC. San Diego, California, 1992.

7. MELLÉKLET

A mellékelt CD-n találhatóak az eredeti és a vágott, feltöltött hangfájlok, valamint a feltöltéshez használt beszédzaj és fűrészel. Az [1] könyvtárban az igazi spondeuszok, a [2] könyvtárban a hamisak. Az ezeken belül lévő könyvtárak tartalma a következő:

- [0]: eredeti hangfájlok
- [25]: 25%-os kivágás
- [25_sn]: 25%-os kivágás, beszédzajjal feltöltve
- [25_st]: 25%-os kivágás, fűrészeljellel feltöltve
- [50]: 50%-os kivágás
- [50_sn]: 50%-os kivágás, beszédzajjal feltöltve
- [50_st]: 50%-os kivágás, fűrészeljellel feltöltve

A CD-n rajta van még a meghallgatásokhoz összeállított anyag, a [hallgat] könyvtáron belül. Itt is az [1] könyvtárban az igazi spondeuszok, a [2] könyvtárban a hamis-spondeuszok találhatóak, külön a 25%-os és az 50%-os vágásúak.