

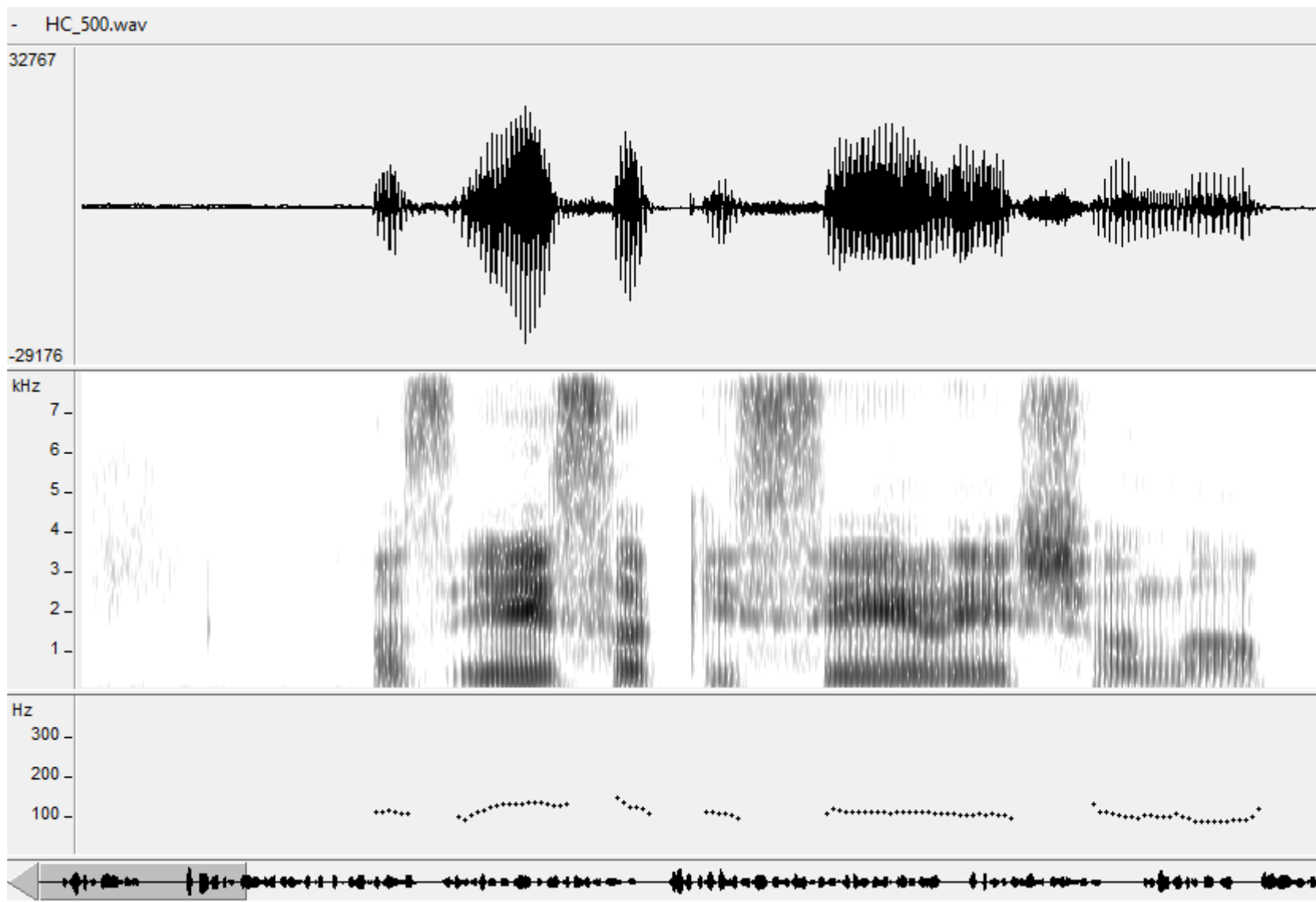
Beszédinformációs rendszerek

3. gyakorlat - Elemi jelfeldolgozás

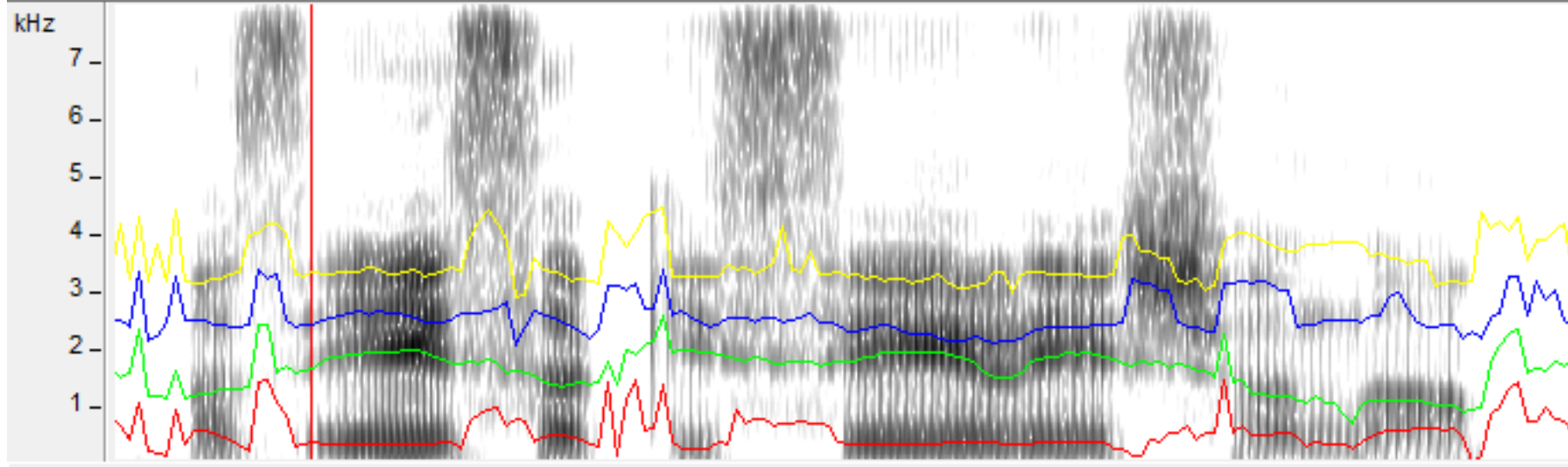
(a beszédjel feldolgozásának lépései)

Beszédinformációs rendszerek

- Kiss Gábor, Tulics Miklós Gábor, Tündik Máté Ákos
- {kiss.gabor,tulics,tundik}@tmit.bme.hu



1. feladat: Az alábbi ábrán egy hangrészletet talál.
- Férfi vagy nő lehet a beszélő?
 - Mennyi az F0 az első magánhangzón?
 - Láthatóak-e valahol formánsok? Ha igen, hol, ha nem, miért nem?



F0-mérés

- 1. feladat: Az alábbi ábrán egy hangrészletet talál.
 - Férfi vagy nő lehet a beszélő?
 - Mennyi az F0 az első magánhangzón?
 - Láthatóak-e valahol formánsok? Ha igen, hol, ha nem, miért nem?
 - → **Wavesurfer**

- 2. feladat: Egy telefonos rendszerben a következő felharmonikusokat sikerült kimérnünk: 400~Hz, 800~Hz, 1000~Hz. Mennyi lehet az alapfrekvencia?
 - → **Lásd 2. gyakorlat**

SAMPA táblázat

Betűk	Példák	IPA	SAMPA	Betűk	Példák	IPA	SAMPA
a	hat	ɔ	o	p	pad	p	p
á	hát	a:	A:	b	bab	b	b
e	vet	ɛ	E	t	tél	t	t
é	vét	e:	e:	d	dél	d	d
i	hit	i	i	k	kép	k	k
í	szít	i:	i:	g	gép	g	ɛ
o	sok	o	o	c	cél	tʃ	tʃ
ó	sók	o:	o:	dʒ	bodza	dʒ	dʒ
ö	köt	ø	2	cs	cső	tʃ	tʃ
ő	söt	ø:	2:	dʒ	dzsem	dʒ	dʒ
u	fut	u	u	ty	tyúk	c	t'
ú	kút	u:	u:	gy	gyár	j	d'
ü	süt	y	y	f	fél	f	f
ű	fűt	y:	y:	v	vér	v	v
				sz	szép	s	s
				z	zaj	z	z
				s	só	ʃ	S
				sz	szár	ʒ	Z
				h	hét	h	h
				r	réz	r	r
				l	lép	l	l
				j	jön, nyuk	j	j
				m	méz	m	m
				n	néz	n	n
				ny	nyom	ɲ	J

A /h/, /j/, /n/ és /m/ fonémák
 megkülönböztetett ejtés-
 változatai (allofónjai)

j	kapj	ç	x
---	------	---	---

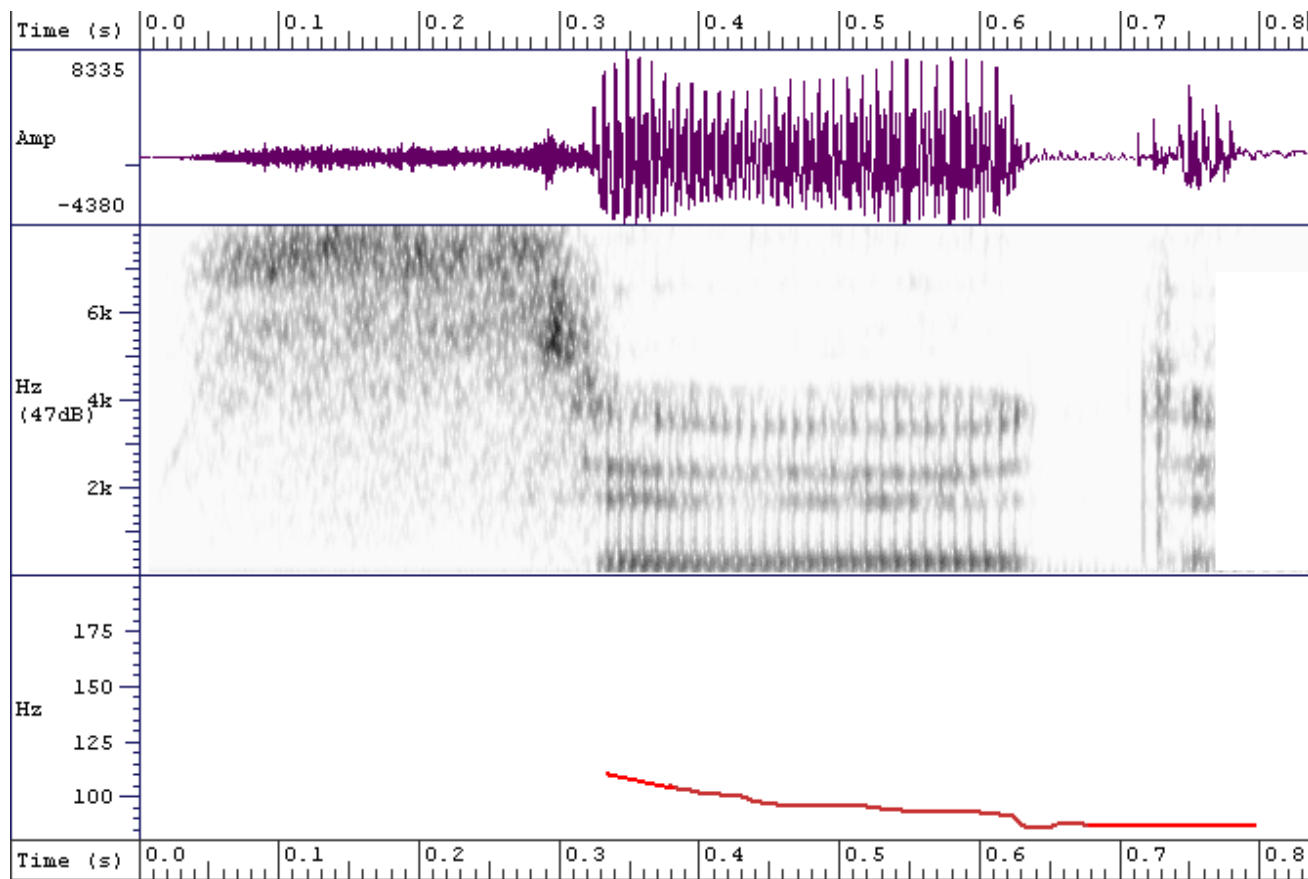
m	kémfőz, hamvas hosszágy, hosszítás	m̥	F
n	ing tönk	n̥	N

h	doh	x	x
h̥	ihlet	x̥	x̥
h	lehet	h̥	h̥

Beszédképzés-akusztika

Feladat: Ejtse ki az édesanya teljes nevét és a kiejtett hangsort jellemezze (egy táblázatot készítsen az összes válaszhoz, szinkronizált ábrázolással)

- a) Adja meg nyomtatott betűkkel a kiejtett nevet
- b) Végezze el a betű-hang átalakítást, és adja meg a kiejtett hangsort a tanult hangkód készletek valamelyikével
- c) Alkalmazzon egy bináris kategorizálási formát a tanultak közül a hangsor beszédhangjaira és adja meg azt hangonként.
- d) Jelölje, hogy melyik hangban van fojtott zöngé (FZ).
- e) F-el jelölje be, hogy melyik hangban van F1, F2, F3 formáns
- f) Osztályozza nevének hangjait külön-külön legalább 3 beszédhang-osztályozási forma szerint.
- g) Rajzolja fel a kiejtett név hangnyomásváltozására jellemző közelítő intenzitás-idő függvényt. Jelölje benne a hanghatárokat is. Időosztás: 100 ms = kb. 2 cm.
- h) Mennyi lehet az átlagos alaphangmagasság frekvencia értéke, ha Ön ejti ki a saját nevét? Indokoljon!



Az alábbi ábrán egy három beszédhangból álló szó időfüggvénye (felső sáv), spektrogramja (középső sáv) és alapfrekvencia görbéje (alsó sáv) látható.

- Húzza be a beszédhangok határait. Indokoljon! Jellemezze a beszédhangok sajátosságait a tanultak alapján.
- Próbálja meghatározni a hangsort. Indokoljon!
- Minden hangra adja meg a gerjesztés típusát!
- Női, férfi, vagy gyermek beszéde van a spektrogramon?

Átírás - szegmentálás

1. feladat: Mit jelent egy beszédatbázis szöveganyagának annotálása, és mit jelent a szegmentálása?

- *Annotálás:* a hangfelvétel tartalmával kapcsolatos általános információkat adunk meg, melyeket egy címkefájl tartalmaz. Pl. szófaji címkék, ortografikus jelölés, kanonikus alak
- *Szegmentálás:* kérdéses részek elkülönítése jelzésekkel/eseményhatárokkal a beszéd időfüggvényében, ahol a gyűjtési szempont szerinti jellemzés a hanganyagban megváltozik.
 - ❑ Leggyakrabban fonéma-, szó- és mondatszintű
- **2. feladat: Végezze el az alábbi szövegrészlet fonotipikus átírását a SAMPA szimbólumrendszer felhasználásával!**
- Azt mondta, hogy kipihen magát. Róla azonban mindjárt megfeledkezett.

Annotálás

- **Milyen paraméterek rögzítése történik az annotálásakor? Mi ennek a jelentősége?**
 - ❑ Meghatározott feladathoz más és más információk felcímkézése szükséges: pl. érzelmek, prozódiai jellemzők (pl. hangsúly)
 - ❑ Fontos pl. az akusztikai-fonetikai modellek betanításához is.
- **Milyen zajok jelölése szükséges annotálásakor? Adjon meg egy tipikus zajcsoportosítást annotálásához!**
 - ❑ Embertől származóak: pl. köhögés, nyelés, szájmozgásból adódó zajok
 - ❑ Környezeti zajok: pl. járművek, ajtónyikorgás, madárcsicsergés
- **Mi a beszélőn belüli és beszélők közötti variáltság? Mit kezdünk vele szintéziskor? És mit felismeréskor?**

Annotálás

- **A beszéd variáltsága**
- \sim =A beszéd jellemző fizikai paramétereinek variáltsága
- Beszédhang fizikai jellemzőit befolyásolja
 - a hangképző szervek mérete
 - a vezérlő izmok feszítettsége
 - az üregek falának rugalmassága.
 - környezeti, akusztikai körülmények(pl. a zajos környezet, visszhangok)

Annotálás

- **Beszélőn belüli variáltság**
- A hangképzés során folyamatosan változó mozgások összessége.
- **Koartikulációs hatás:** A folyamatos hangképzőszervi mozgások miatt az egyik hang fizikai tulajdonságai befolyásolják az azt megelőző és követő hangok fizikai tulajdonságait
- Különböző hangkörnyezetben ejtett ugyanazon beszédhangok fizikai jellemzői, pl. **formánsfrekvenciái** eltérnek egymástól
- **Ritmus, hangerő, hangmagasság, hanglejtés, nyomatékbeli különbségek**
- A **megfázás** igen nagymértékben megváltoztatja a hangok akusztikai paramétereit, hiszen a folyadékkal terhelt üregek rezonanciafrekvenciái eltolódnak
- A **környezeti hatások**, izgalom, meglepetés stb. szintén hatással vannak a létrehozott beszéd akusztikai tulajdonságaira.

Annotálás

- **Beszélők közötti variáltság**
 - **Biológiai tényezők** pl. a beszédképző szervek méretkülönbsége (férfi-nő-gyerek között, de egy csoporton belül is)
 - **Nyelvi különbözőségek**, egy nyelvközösséghez tartozó embercsoportok **ejtésbeli különbözőségei**
- Beszélőn belüli <-> beszédszintézis
- ❑ A felhasznált adatbázis lényeges jellemzője a lehető legnagyobb fonetikai variáltságú anyag összegyűjtése
 - ❑ Az anyagban hangsúlyozottan szerepet kapnak a beszéd mikroszegmentális jellemzői. Kevés, pontosan modellezett beszélő.
- Beszélők között <-> beszéd felismerés
- ❑ A beszélőfüggetlen felismerők betanítására
 - ❑ A beszédstílus, és a rögzítési körülmények nagy variáltsága szükséges

Adatbázisok - általános

1. Mire való a beszédatbázis?

- Nagyméretű, nyelviileg és akusztikailag feldolgozott, tárolt hangadathalmaz, magyarázó jegyzetekkel, címkézésekkel és átírásokkal ellátva
- Lehetőséget teremt a beszéd változékonyságának és változatlanságának tanulmányozására

2. Mi lehet az oka annak, hogy egy német nyelvű, 50 hangot tartalmazó beszédszintézis hangadatházis egy felvétel alapján 4 különböző változatban, 7 Mbyte, 3.5 Mbyte, 2.54 Mbyte és 1.27 Mbyte méretben is elkészült?

- 2-2 fájl méret egymásnak kétszerese → lehet egy 8 és 16 bites vagy kHz-es változat (különböző mintavételezés és kvantálás)
- Beszédszintetizátor bemenetek igény szerint változtathatók; a kisebb kevesebb helyet foglal és gyorsabb, de minőségben ennek ára van

Adatbázisok - általános

3. Milyen követelményeket támasztana egy beszédszintézishez, és milyeneket egy beszédfelismeréshez használatos adatbázissal szemben?

- **Beszédszintetizátor:** profi bemondó, minden hangot ill. hangkapcsolatot számos változatban bemondja, jó hangminőség
- **ASR:** sok bemondó, zajoktól, hibáktól nem kell tisztítani, ha az a normál felhasználási körülményekbe belefér, nyelvfüggetlenség

4. Mit jelent az, hogy az adatbázis fonetikailag gazdag? Mi a trifón? Mi közük van egymáshoz?

- **Fonetikailag gazdag mondatok** → magyar nyelvre is legyen elegendő (statisztikailag megfelelő lefedettséget biztosító), a fonéma-alapú felismerők tanítási kritériumainak megfelelő hanganyag;
 - ❑ a magyarban előforduló fonémák, fonémakapcsolatok(pl. bi- és trifónok: kettős, hármashangzatok – leggyakrabban CVC -) megfelelő számban fordulnak elő
 - ❑ Pl. alma szó az _al és a ma_ triádokból valamint az lm diádból

Adatbázisok - általános

5. Egy 15 beszélő anyagából készült, beszélőnként 1,5 órás, változatos szövegű, fonetikailag átírt, telefonon keresztül felvett beszédatbázist kap kézhez. Mire lehet alkalmas ez és mire nem? Indoklás!

- Elemösszefűzéses beszéd szintetizálás
- Beszédfelismerő irodai diktálásra
- Beszédfelismerő call centerben
- Beszédfelismerés adaptáció

Megoldás:

- Beszédszintetizátorhoz a telefonos minőség nem megfelelő
- Beszédfelismerő tanításához (legyen az diktálás vagy call center) a 15 ember kevés
- Leginkább így csak adaptációt tudunk végrehajtani -> beszédfelismerés adaptációra alkalmas a kapott adatbázis

Appendix

- A magyar beszéd (19.o. – 26.o, 3. fejezet: Fiziológiai, fizikai alapok)
- **3.1.1 Gégeszintű hangképzés – Zöngés hang**
- Zöngéállásnál a **hangszalagok kváziperiodikusan szétnyílnak-összezáródnak**, a levegő nyomásviszony-változásának következtében (tüdőből préseljük) → létrejön a **zöngéshang**
- Ez a **zöngés hang ejtésének ideje alatt folyamatosan ismétlődik**, melyet egy **periódusidővel** jellemzünk
 - ❖ Férfiaknál : 8-12 ms
 - ❖ Nőknél : 4-6 ms
 - ❖ Gyermekéknél még rövidebb
- **Átlagos alaphangfrekvencia** a beszélő nemére enged következtetni:
 - ❖ Férfi: 100 Hz, Nők : 180 Hz, gyermekek: több száz
- Speciális zöngéképzés: **Fojtott zöngé**
 - Az orális és nazális csatorna le van zárva, nem áramlik ki levegő az artikulációs csatornából → pl. baba szóban a „b”-nél
 - Rövid időtartamú, amíg kisebb a szájüregben lévő nyomás a légcsőből érkezőnél.
 - Az arcüreg megrezegteti a külső levegőt

Alaphang(F0): beszéd alaphangfrekvenciája
Reciprok kapcsolat a periódusidővel(1/s,Hz)

Appendix

- **Zöngétlen hangok esetén**

- ❖ Nincs fonáció

- ❖ Fúvó állásban a gége nem jelent akadályt a tüdőből kipréselt levegőnek -> turbulens áramlás → réshangok

- ❖ Az akadályok az artikulációs csatornában hozhatók létre:

- 1. Zárhangok: Lökéshullámok eredményeképpen, néma szakasz

- 2. Zár-rés hangok

- ❖ Kivétel: H-állás

Appendix

• 3.1.2 Az artikulációs csatorna, toldalékcső (szupraglottális)

- A hangszalagok és az ajkak közötti hangképzési terület
- Három üregrész: garat, szájüreg, orrüreg
- Itt formálódnak ki a beszédhangok, **artikulációs szervek** segítségével

❖ **Aktív:**

- **Nyelv:** előre-hátra ill. fel-le mozgás : magánhangzó képzés \leftrightarrow ha akadályt is képez: mássalhangzó (zárhang)
- **Ajak:** kerekítés, széthúzás: magánhangzó \leftrightarrow összezárás + fogakkal interakció : mássalhangzó

❖ **Passzív:** orr (orrüreg)

• Így adódik a mássalhangzók csoportosítása:

- **Képzés helye:** bilabiális | labiodentális | dentialveorális | alveorális | palatális | veláris | faringális | nazális
- **Képzés módja:** zárhang | réshang | zár –rés hang | pergőhang | közelítőhang

Appendix

- **3.1.2.2 A gégeműködés és az artikulációs csatorna viszonya**
- A **gégében**, levegőráamlás által létrejött **hang** a beszéd alapkomponense → **Az artikulációs csatorna erre a hangra épít, a beszéd itt nyeri el végleges rezgésformáját.**
- A kisugárzott beszédhangok jellemző frekvenciakomponenseiket
 - A toldalékcsőben kialakult rezonanciafrekvenciáknak (~ magánhangzók)
 - Létrehozott szűkületek helyének és alakjának (~ mássalhangzók)
 - köszönhetik.
- A toldalékcső rezonátorüregként vesz részt a hangképzésben -> bonyolult rezgésformák jönnek létre, többféle rezonanciafrekvencia van a rezonátorüregben

Appendix

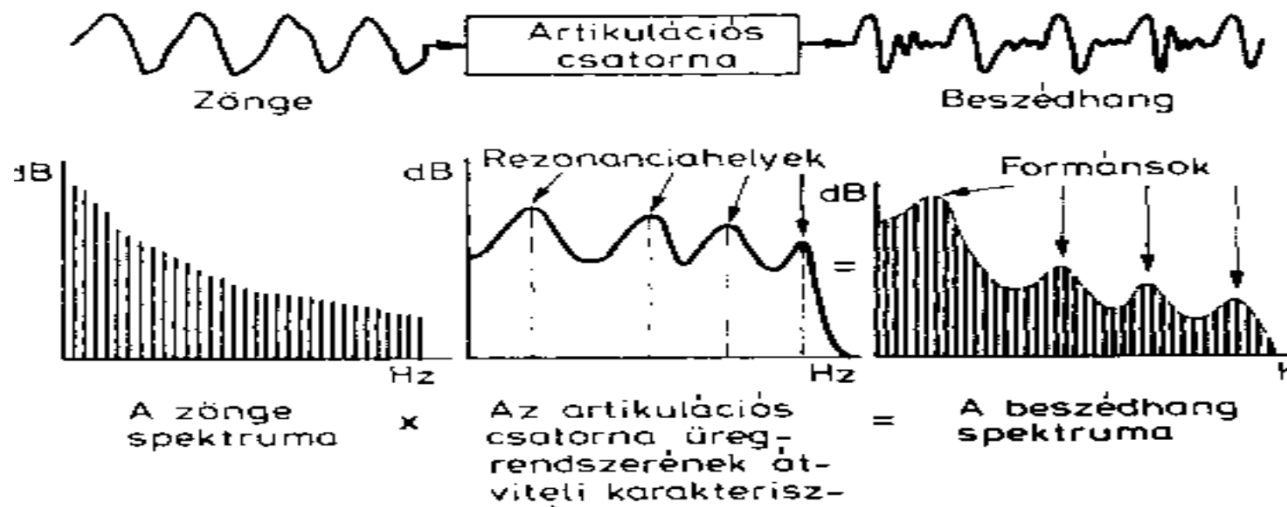
• 3.3.5.1. Formáns, zörejgóc

- A zöngés elemek frekvenciaszerkezete a gégeszintű hangforrás és a toldalékcső (mint rezonátorüreg) együttes tulajdonságaiból adódik, pillanatnyi spektrummal jellemezhető -> akusztikus vetület
- Fonáció esetén a zöngé rezgésformája gerjeszti az artikulációs csatornát
- **A zöngé: az alaphang + felharmonikusok**, kb. 5000 Hz-ig
- (gyakorlatban 4000 Hz-ig tudjuk leolvasni)
- **A vonalas spektrumban** a felhangok amplitúdó 12 db/oktáv csökkenéssel vannak jelen, legnagyobb amplitúdójú az alaphang
- **Felharmonikusok** egyes csoportjai **a pillanatnyi rezonanciafrekvenciákon** és környékükön **felerősödnek** → **formánsok**
- A vonalakra helyezett burkológörbe csúcsainál vannak
- A nagyobb sorszámú távolabb van az alaphangtól, nagyobb a sávszélessége, az amplitúdó esése: 6 db/oktáv

Appendix

• 3.3.5.1. Formáns, zörejgóc

- A magánhangzókra jellemző az F1-F2 + azok sáv szélessége
- Zárhangok zárszakaszában nem mérhető
- Zöngétlen hangoknál turbulens áramlások, zárfelpattanási zörejek (lökéshullámok) gerjesztik a csatornát, képzés helyének megfelelő **zörejgócok** alakulnak ki



Forrás: Dr. Olaszy Gábor , Dr. Németh Géza: Beszédinformációs rendszerek – Beszédkeltés gyakorlat (2012)

3.3.4. Összetett rezgések frekvenciaelemzése

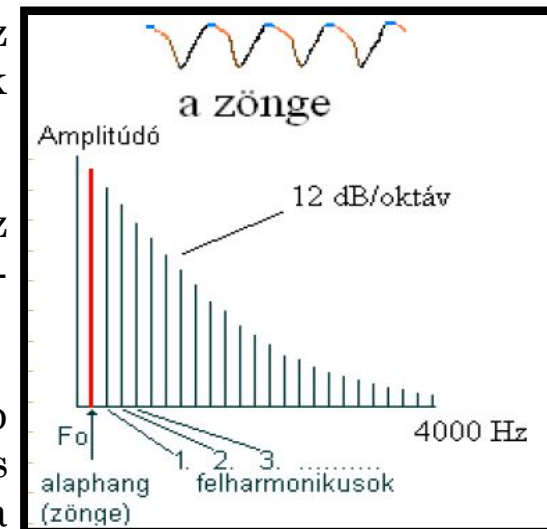
• Frekvenciakomponensekre bontás → összetett rezgés hangnyomás-idő függvényét felbontjuk frekvenciatartományban értelmezhető függvényre

• Többféle spektrum: nyomásamplitúdó, teljesítmény, energia: az adott időpontban a frekvencia-összetevők melyik eloszlást adják meg.

• Periodikus, összetett rezgések jellemző alapperiódusa az **alaphang**: az összetett hangot felépítő összes frekvencia-összetevő közül a legalacsonyabb frekvenciájú (F_0)

• Többi összetevő - felharmonikusok: a legalacsonyabb frekvenciájú alaphang egész számú többszöröse → vonalas spektrum jellemző a periodikus rezgésekre, hiszen csak a felhangok frekvenciáin vannak spektrumösszetevők

• Hallási tulajdonságunk révén akkor is meg tudjuk ítélni az alaphangot, ha nincsen benne az F_0 a beszédjelben (pl. telefon)



Forrás: Dr. Olasz Gábor , Dr. Németh Géza: Beszédinformációs rendszerek – Beszédkeltés gyakorlat (2012)

Összetett rezgések frekvenciaelemzése

- A nemperiodikus jelekre folytonos spektrum jellemző
 - sok frekvenciájú szinuszos összetevőből állnak, a frekvenciatartomány bármely pontján lehetnek, az összetevők nem meghatározott frekvenciáknál koncentrálnak
- Ábrázolások a frekvenciatartományban
 - ❖ Spektrum: 2D – vízszintes tengely a frekvencia, függ.: intenzitás
 - ❖ Spektrogram: 3D – x: idő, y: frekvencia, z: amplitúdó(szürkeárny.)

A beszédjel elemzése

- Fontos, mekkora időintervallumot választunk
- Rövid ablakszélesség: Időben gyors változásokhoz: (pl. zárfelpattanás), zöngés hangperiódusokból az alapfrekvencia kiszámítható
- Hosszú időablak: Finom frekvenciafelbontás, zöngés hangok felhangjai látszanak
- Minél feketébb a kép, annál nagyobb a hangrész frekvenciaösszetevőjének az intenzitása
- Zöngés-zöngétlen szakaszok megállapítására is alkalmas
- Kisebb-nagyobb energiájú szakaszok váltogatják egymást
- A TF (mintavételi ablak) $\cdot \Delta\omega$ (frekvenciafelbontás, frekvenciaösszetevők távolsága DFT-ben) = állandó érték.

Az írott szöveg és a hangalak kapcsolata

- A hangalak, fonetikai átírat meghatározása két irányból közelíthető: ortografikus szöveg -> beszéd, elhangzó beszéd -> szöveg
- Folyamatos beszéd lejegyzése esetén fontos figyelni a szavak összekapcsolásakor létrejövő koartikulációra, szünettartásra, további hangjelenségekre (pl. megakadásjelenségek)
- A lejegyzés nyelvfüggő is, magyarban az ortografikus lejegyzés közel áll a hangalakhoz
- Kanonikus írás: a helyesírás szerinti ortografikus formát, egységesebb, írott szavakat tartalmazó formára hozzuk. -> feloldások ott, ahol nem felel meg a kiejtés alakjának
- Pl. gépi beszédfelismerésben fontos, hogy meglegyen az ortografikus alak, a fonemikus alak, és a rövidítések feloldása is (lásd kényszerített illesztés, fonemikus alakból dolgozik)

A beszéd szegmentális szerkezete

- Szegmentális: alapvető elemek a beszéd létrehozásához: beszédhangok, hangkapcsolódások, nyelvi időtartamok
- Szupraszegmentális: beszéddallam, hangsúlyozás, ritmika (prozódia) – az alapvető szegmentumokra épül, nyelvfüggő, nagyobb nyelvi egységen értelmezett elemek
- **Akusztikai vetület:** F1-F2-F3 formánsok által meghatározott, a rezonátorüregként működő toldalékcső ha zöngehanggal gerjesztett, akkor a keletkezett hang formánsstruktúrája jellemzi a pillanatnyi artikulációs állapotot
- Fontos a szerepe beszéd-szintézisben (formánsszintetizátor), és útmutatást ad a beszéd felismerésben is

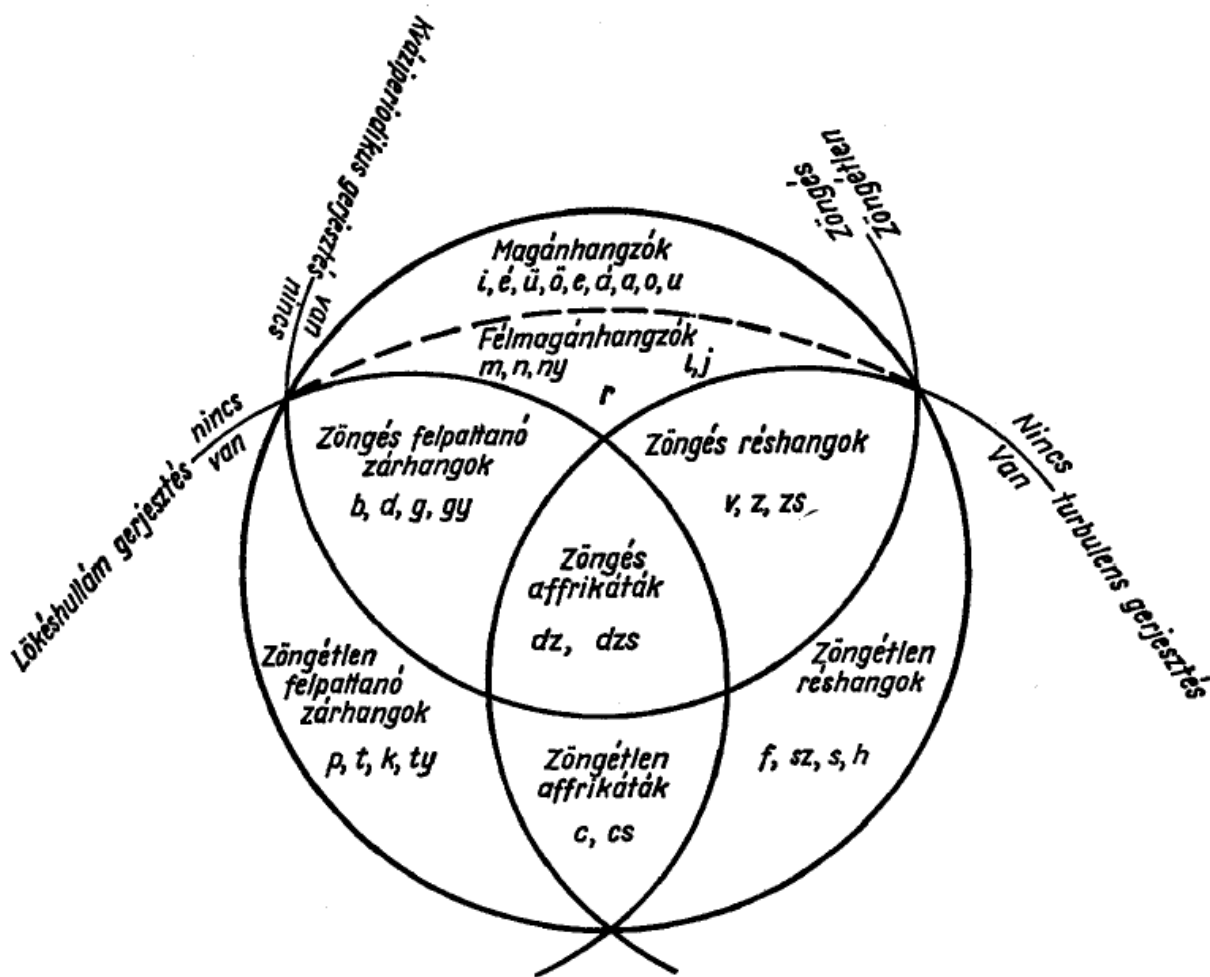
Magyar beszédhangok

- 65 fonéma -> megszólaltatott formájuk : beszédhang
- Akusztikai szempontból jól megkülönböztethetők, de ejtésük a beszélőfüggő, és környezetfüggő is (koartikuláció) -> Így több beszédhang-realizáció van, mint fonéma

Beszédhangok osztályozása (Bináris)

- Magánhangzó-mássalhangzó: a magánhangzó szótagképző, valamint nincs akadály az artikulációs csatornában
- Rövid-hosszú hang: Jelfeldolgozásban nehezebb a modellezés
- Orális-nazális hang: akusztikai szerkezeti különbségek szerint, hol van hangkiáramlás
- Gerjesztési hely: gégeben állandó a zöngéképzés helye, viszont ha zörejeket tartalmaz, az artikulációs csatorna különböző pontjain keletkezik
- Gerjesztési forma: zöngés, zörejes, kevert gerjesztés (szuperponáció)
- Egyszerű-összetett: folyamatosan ejthető / nem ejthető

Gordos-körök a beszédhangok osztályzására



Forrás: Gordos-Takács: Digitális beszédfeldolgozás (1983)

A magyar mássalhangzók

- 25 rövid-hosszú pár (megkülönböztető szerepűek) + 5 darab variáns
- Megkülönböztetjük őket a képzési hely, mód és gerjesztés formája szerint
 1. Képzési mód: zárhangok, réshangok, zár-rés, közelítő, pergő, nazális
 2. Képzési hely(szűkület helye az art. csatornában): bilabiális, labiodentális, dentalveorális, alveorális, palatális, veláris, faringális
 3. Gerjesztés szerint: zöngés, zöngétlen, vegyes

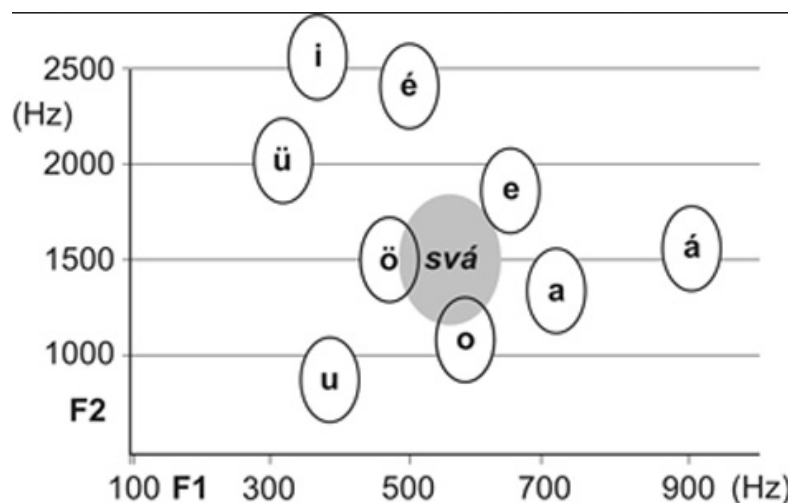
Forrás: Gósy Mária: Fonetika, a beszéd tudománya c. könyv – A beszédhangok akusztikai szerkezete c. fejezet

A magyar mássalhangzók akusztikai jellemzése (átlagos, az adott csoportra jellemző tájékoztató értékekkel)

Mássalhangzók	Szerkezet	Gerjesztés	F1 (Hz)	F2 (Hz)	Z1 (Hz)	Z2 (Hz)
felpattanó zöngés zárhangok	összetett	zöngés	300–500		500–2200	
felpattanó zöngétlen zárhangok	összetett	zörejes			500–2200	
nazálisok	egyszerű	zöngés	300	800–2200		
zöngés réshangok (kivéve [v])	egyszerű	kevert	300	1400–1800	1600–4600	3200–8000
[v]	egyszerű	kevert	300	1200	1500–7000	
zöngétlen réshangok	egyszerű	zörejes			1600–4600	3200–8000
zöngés affrikáták	összetett	kevert	300	1400–1800	1600–4800	3500–7000
zöngétlen affrikáták	összetett	zörejes			1600–4800	3500–7000
pergőhangok	egyszerű	zöngés	300–500	1000–1600		
közelítőhangok	egyszerű	zöngés	300–400	1600–2500		

A magyar magánhangzók

- A magyar beszéd (106.o, 5. fejezet: A beszéd szegmentális szerkezete)
- 9-féle + vannak hosszú párjaik + svá
- Három fő artikulációs alapparaméter: nyelv, ajkak, állkapocs helyzete
- Egyszerű szerkezet, kitartva ejthető, zöngés gerjesztésűek → a zöngékváziperiodikus jelként gerjeszti a gége feletti artikulációs csatornát
- A magánhangzók formánsai adott sávon belül mozognak → az artikulációs csatorna hosszabb: lefelé tolja a formánsokat + a koartikulációval is mozoghatnak

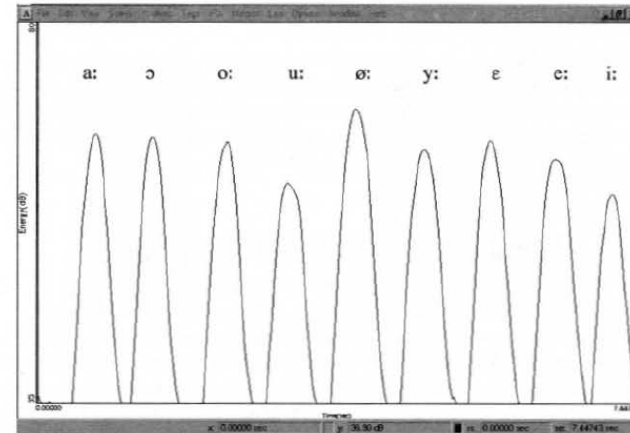


Magánhangzók csoportosítása

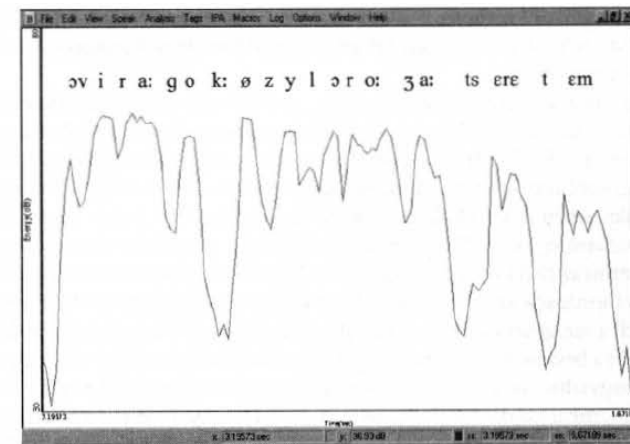
A nyelv vízszintes mozgása szerint	Hátul képzett (veláris)				Elöl képzett (palatális)			
Ajakkműködés szerint	Ajakkerekítéses (labiális)		Ajakréses (illabiális)		Ajakkerekítéses (labiális)		Ajakréses (illabiális)	
Időtartam szerint	rövid	hosszú	rövid	hosszú	rövid	hosszú	rövid	hosszú
A nyelv függőleges mozgása szerint								
felső nyelvvállású (zárt)	u	ú			ü	ű	i	í
középső nyelvvállású (középzárt)	o	ó			ö	ő		é
alsó nyelvvállású (középnylt)	a						e	
legalsó nyelvvállású (nyílt)				á				

Intenzitásértékek

- Magánhangzókhoz viszonyítva rendszerint kevésbé intenzívek a közelítőhangok és a nazálisok, valamint a pergőhang
- Az alveoláris és a posztalveoláris zöngétlen réshangok akár a magánhangzókéhoz hasonló intenzitásúak is lehetnek
- A legkevésbé intenzívek általában a zöngés alveoláris és a posztalveoláris réshangok, a [v], az affrikáták és a zárhangok, végül a [f] és a [h].
- Ezek a megállapítások a mindenkori kontextus és fonetikai pozíció, valamint a szupraszegmentális szerkezet függvényében változhatnak



75. ÁBRA
Izoláltan ejtett magánhangzók egymáshoz viszonyítható intenzitása



76. ÁBRA
Az A virágok közül a rózsát szeretem közlés intenzitás szerkezete

Forrás: Gósy Mária: Fonetika, a beszéd tudománya c. könyv –
A beszédhangok akusztikai szerkezete c. fejezet

Adatbázisok a beszédtechnológia szolgálatában

- Beszédatbázisok csoportosítása:
 1. Analitikus-diagnosztikus: nyelvi, fonetikai kutatáshoz pl. BABEL
 2. Általános: sokfajta területhez: pl. BEA spontán adatbázis
 3. Specifikus: adott felhasználáshoz: MRBA, SPEECHDAT, szintézishez
- Beszédatbázisok forrásai(Spoken Language Resources)
 1. Telefonos
 2. Mikrofonnal felvett
 3. Híryananyag
 4. Rádióanyag
 5. Kiejtési anyagok
- Vannak multimodális adatbázisok is.

Adatbázisok a beszédtechnológia szolgálatában

- **Tanító adatbázis gépi beszédfelismeréshez**
 - ❖ Beszélőfelismerés
 - ❖ Nyelvfelismerés
 - ❖ Érzelemfelismerés
 - ❖ Tartalomfelismerés
 - ❖ Egészségi állapot felismerés
- Manapság a statisztikai modell alapú (HMM) illetve a neurális háló alapú rendszerek elterjedtek
- Egy pontos paraméterbecslési lépés nagyszámú minta vizsgálata alapján lehet sikeres, megfelelő címkézés, átírás kell
- Adatbázis létrehozása: a véletlenszerű folyamat egyes megvalósulásainak összegyűjtése egészen a tervezett használati mód közel teljes lefedéséig -> fontos az adatok jósága!

Adatbázisok a beszédtechnológia szolgálatában

- Többféle szegmentálást kell alkalmazni
 1. Diktálásnál: beszédhang
 2. Prozódia: hosszabb hangsorelem
 3. Modalitás: mondatnyi
 4. Érzelemfelismerés: érzelmeket kifejező beszédegységek
- →Specifikus felismerők
- Két tanító adatbázis kell, akusztikai és nyelvi modellhez
 - ❑ Beszédfelismerők osztályozása:
 - Beszéd: beszélőfüggő, beszélőfüggetlen, beszélőadaptív
 - Beszédegység: hang, hangkapcsolat, szó, szókapcsolat
 - Beszédmód: izolált, kapcsolt szavas, folyamatos
 - Beszédstílus: olvasott, előadás, spontán
 - Szótár: felismerendő elemek száma
 - ❑ Szövegtörzsből: unigramok, bigramok képzése

Beszédből készített elembázisok beszédpszintézishez

- Hullámformarészletek (hangelembázis)
- Folyamatos beszédek, mondatok
- Néha csak az fontos, hogy egy darab jó minőségű hang alakuljon ki, elég egy hangra készíteni
- Minél jobb hangszínezetet akarunk, annál nagyobb adatbázis kell
- Hangjegyek felolvasása pl.
- Elemösszefűzéses, diád/triád alapú – szabályrendszert valósít meg