BESZÉDVÁLASZÚ RENDSZEREK
(beszédszintetizátor)

Előadja:

Lukács Levente

Tartalom:

- Történelem
- Mi a beszéd?
- Alapfogalmak
- A magyar beszéd
- Gépi beszédkeltés
- Kötött szókészletű rendszerek
- Számfelolvasó
- Szövegfelolvasó
- Szövegfelolvasók minősítése

Célok:

- -automatikus szöveg-felolvasás
- -gépi beszéd-előállítás
- -beszédfelismerés
- -beszédtisztítás
- -beszélő személy felismerése, azonosítása
- -digitális beszédfeldolgozás

Egy kis történelem:

- A beszéd tudománya (a fonetika)
- Számítástechnika lehetőségeit melyek további kibontakozáshoz vezettek a fonetika mindhárom területén:
  - az artikuláció,
  - a beszéd képzése,
  - a beszédészlelés

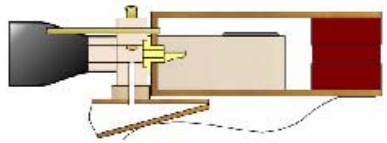
- A **beszédtechnológia** a beszédkutatás új irányzata
  - beszéd alapú (verbális) **gyakorlati alkalmazások** létrehozásával foglalkozik.

- **In memoriam Kempelen Farkas:** ő alkotta meg az első olyan mechanikus szerkezetet, amely az emberi beszédhez nagyon hasonló hangokat tudott kiadni.
  http://fonetika.nytud.hu
- **(1790-ben megjelent könyvében leírt változat rekonstruálása)**

A beszélőgép kívülről

Mit rejt a takaródoboz?

Működés közben



Működési elv

Alapfogalmak, meghatározások

- **A beszéd** hangzó részek és szünetrészek sorozata.

  Ha valamelyik hiányzik, vagy nem megfelelő szerkezetű, akkor sérül a kommunikációs hatékonyság.

- **A beszéd végeredménye, sok paramétertől függ:**
  - akusztikai, (milyen hangsor hangzik el)
  - nyelvi, (milyen információt hordoz a hangsor)
  - egyéni, beszélőtől függő (egyéni megformálás, hangszínezet)

- **A sikeres beszédtechnológiai** fejlesztésekhez komplex munkára van szükség

- **A beszéd elméleti szerkezeti szintjei:**

Szegmentális szerkezet	**Szupraszegmentális szint**
artikuláció	prozódia
akaratunktól ált. nem függő	akarattól függő

- **Hangképzés:** beszédképző szervek segitségével valósul meg
- **Akusztika**: a hangrezgésnek vannak jellemzői: időtartam, frekvencia, intenzitás, spektrum;
  - **Frekvencia:** Fo és a formánsok
  - **Intenzitás:** mondatra, szóra, hangra vonatkozó, ill. hangon belüli

- **Magyar beszéd**: nyelvspecifikus szegmentális és szupraszegmentális szerkezetek;

  -A magyar nyelvre jellemző 14 magánhangzó **típus szerint osztályozva** van;

  -részletesen vizsgálni kell a **formánsszerkezetüket**, az **időtartamukat** és a **hangzósági szintjüket**;

  -A mássalhangzóknál tárgyalja a különböző típusok (zárhangok, zár-rés hangok, réshangok, nazálisok) **akusztikai-fonetikai jellemzőit**;

  -A prozódia keretén belül részletesen foglalkozik a hanglejtéssel, a hangsúllyal, a tempóval és a beszéd ritmusával.

Fo = alapfrekvencia= 50-500 Hz

Formáns = felerősített felhangnyaláb,

Zönge= a gége szintjén lévő kvázi periodikus rezgés (forrásjel), amely a hangszalagoktól ered,

Fojtott zönge= zöngés zár- és zárrés hangok zárszakaszi építőeleme b,d,g,gy,dz,dzs,

Néma fázis= a zöngétlen zár- és zárrés hangok zárszakaszi építőeleme p,t,k,ty,c,cs,

Zárfelpattanás= zár hangok zárszakaszi eleme utáni hangrész, amely befejezi a hangot,

Specifikus időtartam= a hangra jellemző alapidőtartam a hangkörnyezet függvényében,

Specifikus intenzitás= a hangra jellemző alapvető intenzitásszint a többi hanghoz viszonyítva.

- **FONÉMA**

  Például: bár – pár; már- vár.

  A hosszú-rövid hangok is külön fonémaként kezelendők a magyarban, noha inkább csak az időtartamukban különböznek egymástól.

  Például: hal - hall; sok – sók; tör – tőr.

- A **beszédhangok különböző megvalósulásai**
  - hangsorkezdő,
  - Hangsorbelseji,
  - és hangsorzáró helyzetben.
- **Artikulációs sebesség (hang/s),**
- **Beszédsebesség (hang/s),**
- **VOT=** zönge kezdési idő a zöngétlen zárhang zárfelpattanásától a zönge megindulásáig mért idő,

  Értéke nyelvenként változik,

  A magyarra: p=10-20 ms; t=20-30 ms; k=30-100 ms.

Spectrogram:

- **Spektrum:** egy jel meghatározott idointervallumban mért színképi teljesítmény eloszlásfüggvény.

- **Spektogram:** a színképi teljesítmény eloszlásfüggvény idobeli változását adja. A spektogram a beszédhang hangsúlyozását is ábrázolja

- **Szoftverek - Spectrogram**

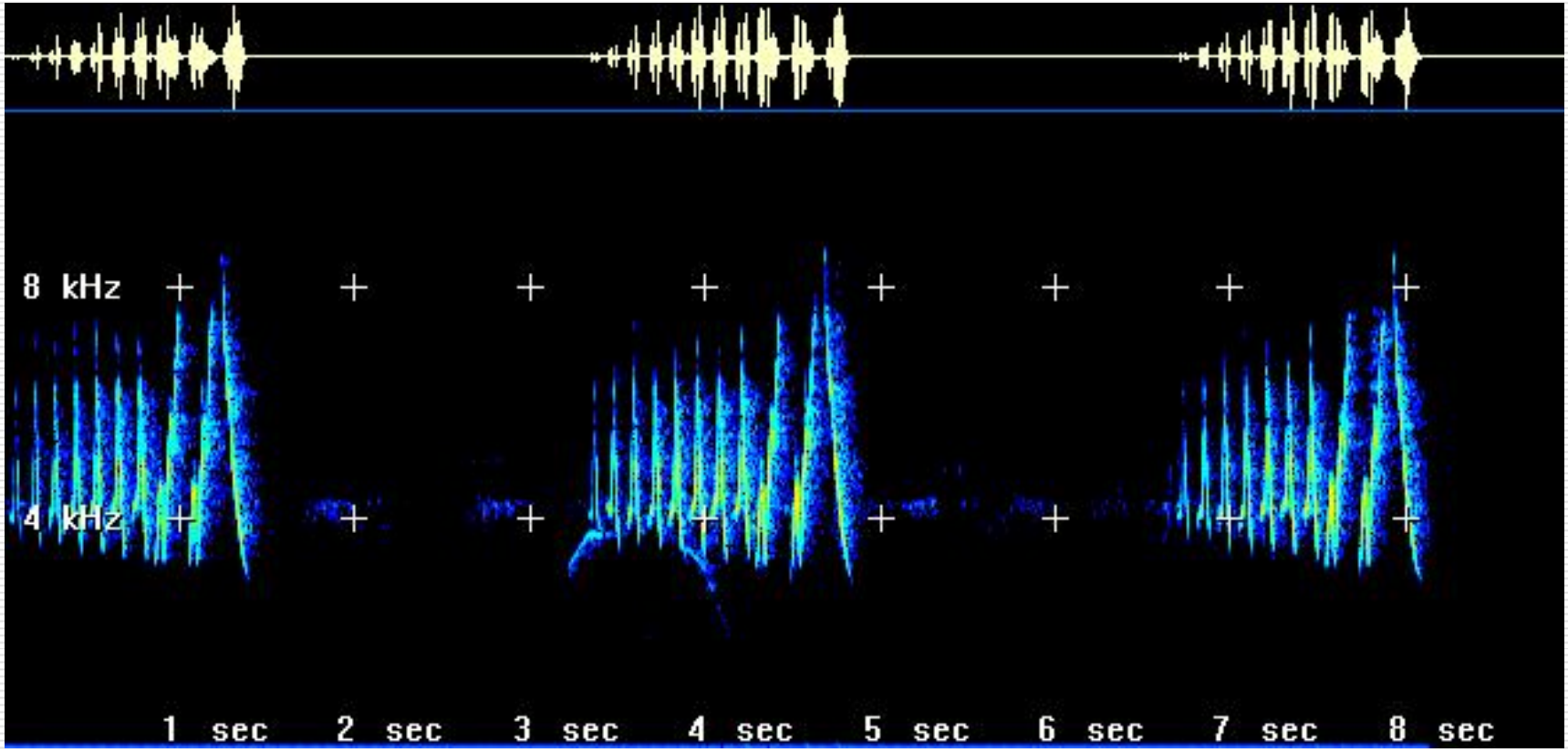
  A Spectrogrammal a számítógép egy kétcsatornás spektrum-analizátorrá alakítható. A hangkártya analóg/digitális konvertere felhasználásával bármilyen bemenetről digitalizálható a jel, s az egyből láthatóvá is tehető a képernyőn. A Spectrogram ideális eszköz a következő minták analíziséhez: emberi hangok és zene analízise, zeneszerszámok hangolása, audio rendszerek hitelesítése.

Honlap: http://www.visualizationsoftware.com/gram.html

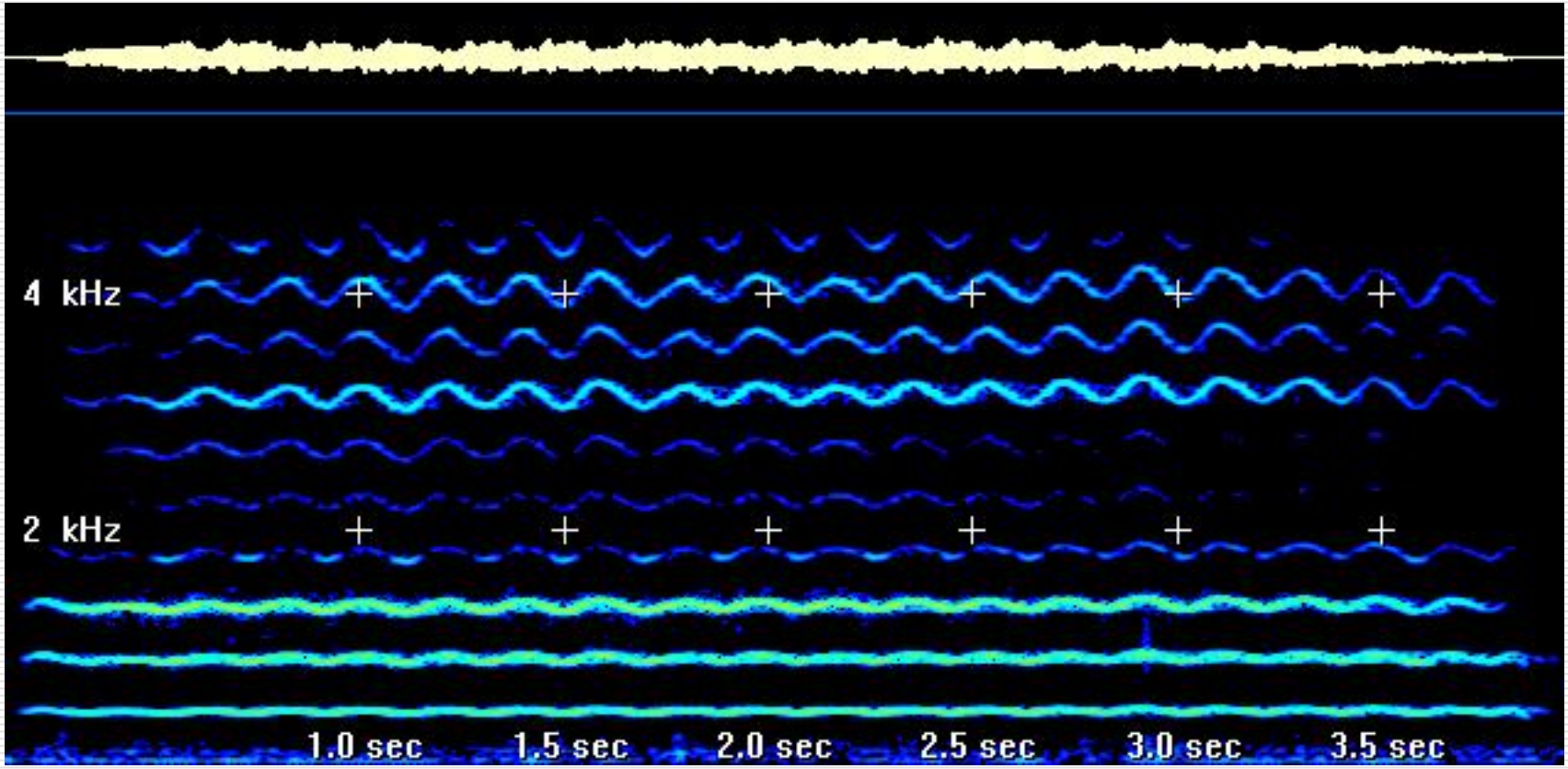
Közvetlen letöltés:
http://www.visualizationsoftware.com/gram/programs/setup.exe (465 kB)

Pár példa:

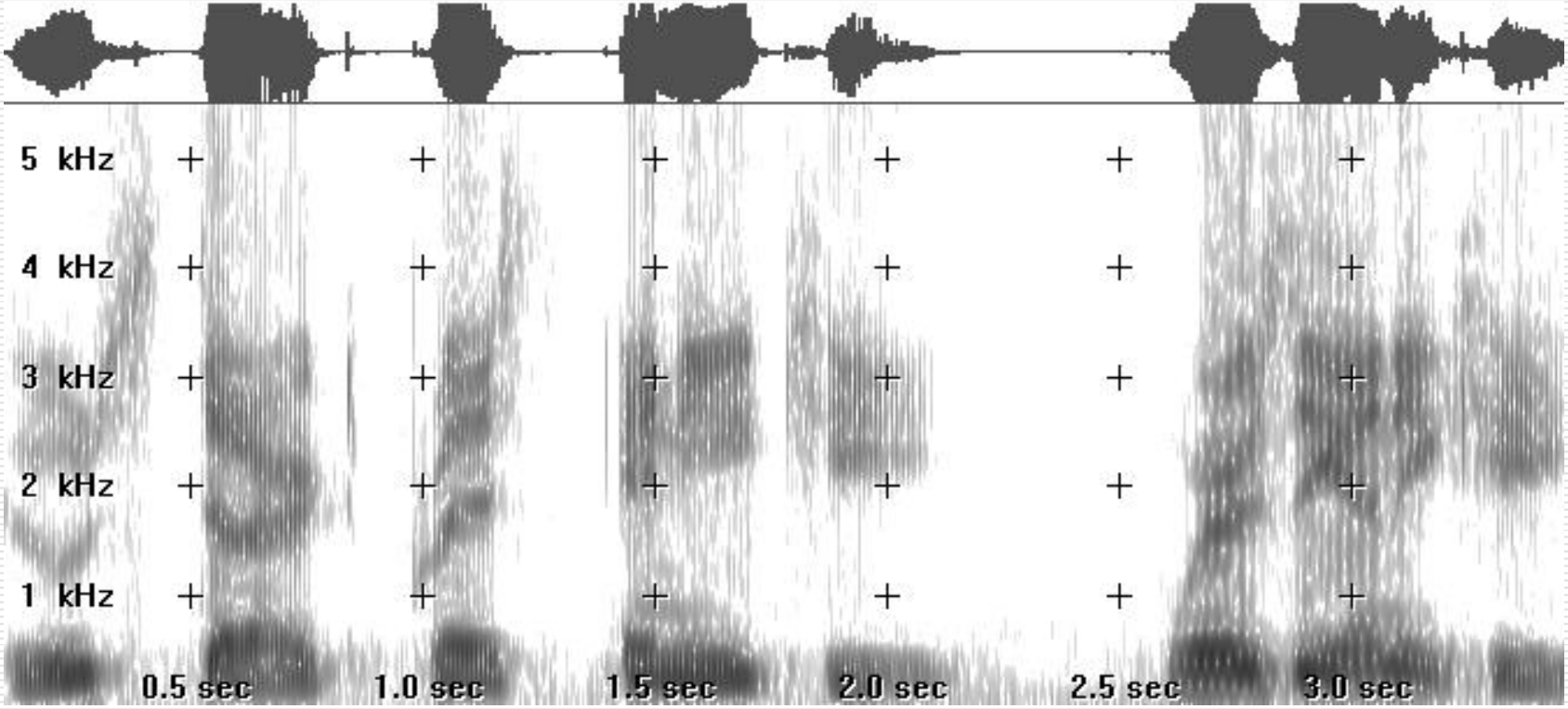
- ***Audio Spectrum of the Song of the Chestnut-sided Warbler***
  Recorded from Birds of North America V2



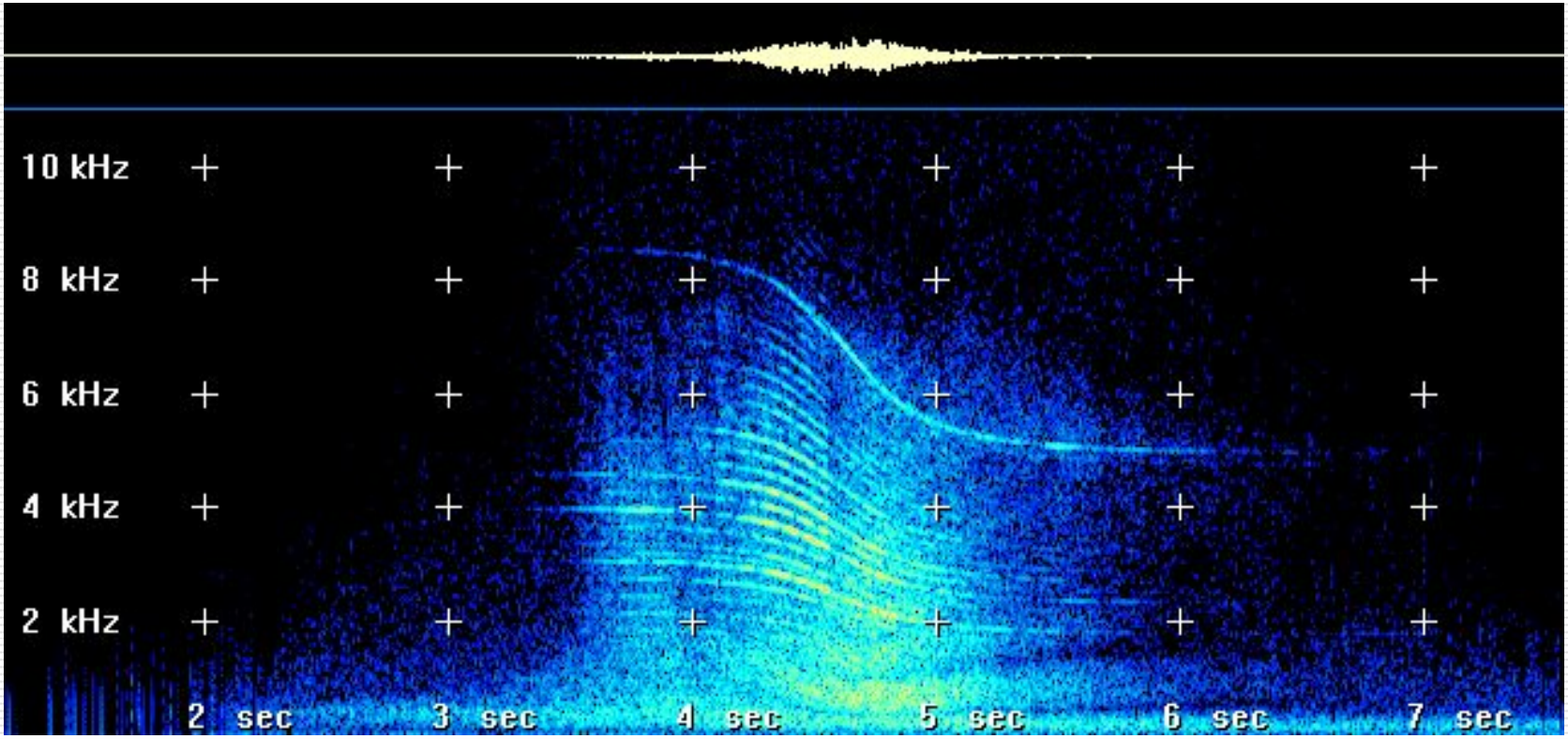
Audio Spectrum of the Singing Voice



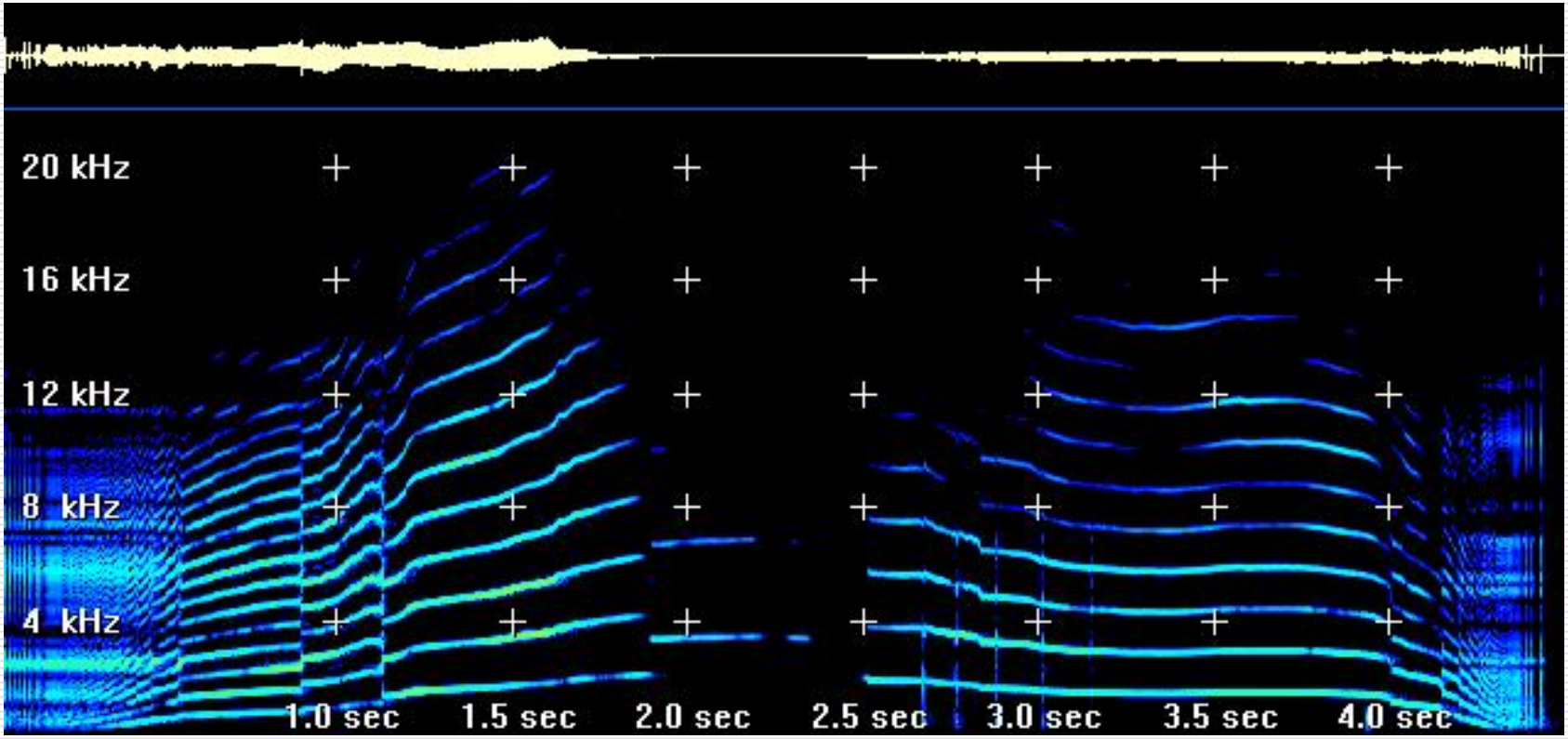
Broadband Audio Spectrum of the Speaking Voice



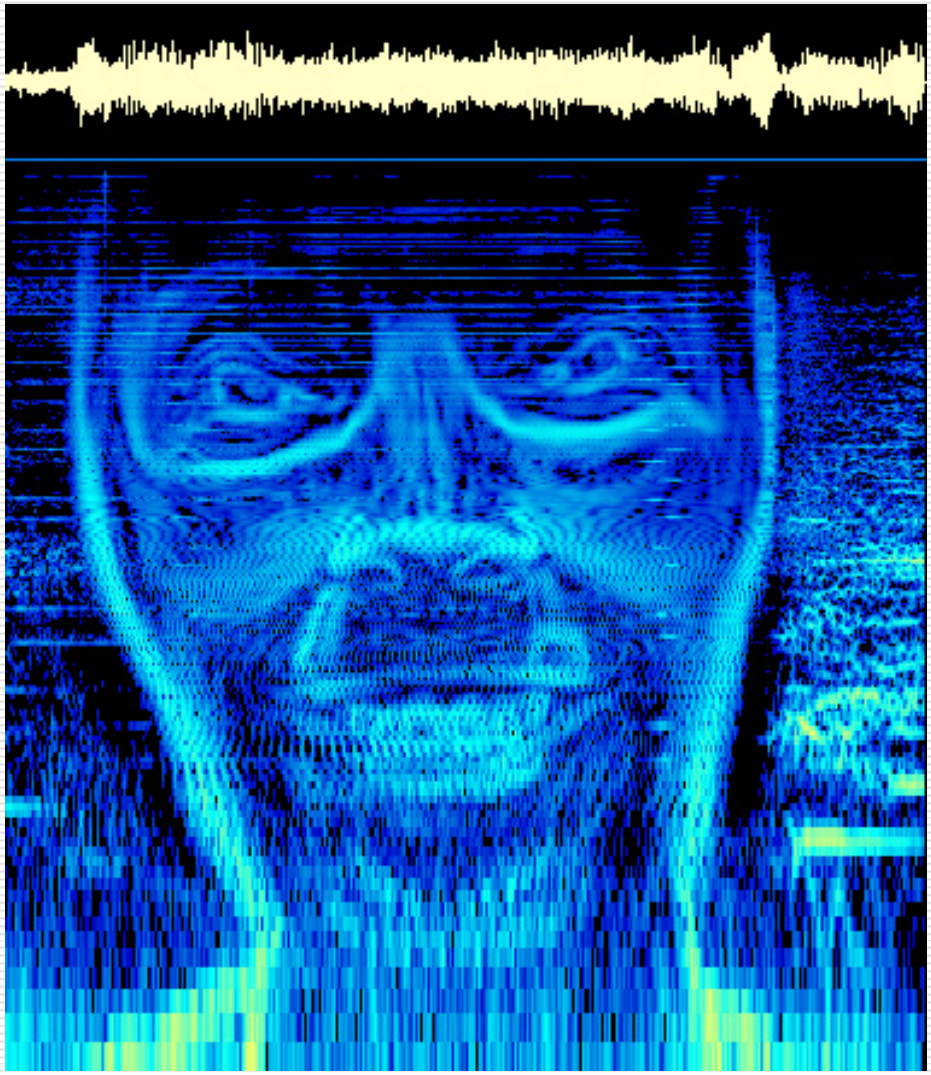
Audio Spectrum of the Sound of a Low Flying Jet



- ***Audio Spectrum of the Sound of a Creaking Door***



- ***"Equation" by Aphex Twin***



Az emberi és gépi beszédkeltés különbsége

- Az emberi beszédkeltés **párhuzamos folyamatok** eredménye: több paraméteres függvény, amely egyazon időben valósul meg (hangképzés, hangerő, időzítés, hangsúlyozás, dallamformálás stb.).

- A gépi beszédkeltésben kénytelenek vagyunk ezt a folyamatot több részre bontani: **soros (esetleg párhuzamos)** megvalósítás.

- **Hol használunk gépi beszédet?**
  - 1. a vizuális visszajelzés mellé kiegészítő információ
  - 2. az ember-gép kommunikáció természetesebbé tétele
  - 3. információ nyújtása ott, ahol máshogyan nincs mód
  - 4. az emberi fárasztó munka kiváltása (telefonközpontoknál, információs rendszereknél, stb.)

Általános szabályok a beszédszintetizátorok tervezésével, fejlesztésével, megítélésével kapcsolatban

- A beszédszintetizátort a társadalom a hangminősége és érthetősége alapján véleményezi (nem szól jól a szintetizátor).
- A gépi beszédelőállítás csak csapatmunka esetén lehet sikeres (nyelvész, akusztikus, mérnök és alkalmazói szakembergárda)
- A gépi beszédelőállítás kompromisszumokat követel (technikai követelmények állnak szemben a nyelvi kívánalmakkal)

AZ ALKALMAZÓ ÉS A FEJLESZTŐ KÖLCSÖNÖS EGYÜTTMŰKÖDÉSE SZÜKSÉGES!

- A beszédszintetizátorok fejlesztési folyamata:

koncepció+tervezés => fejlesztés <=> tesztelés =>integrálás.

A beszédszintézis és a beszéd akusztikai szerkezete közötti kapcsolat 4 paramétere:

- A) A beszédhangok **spektrális szerkezetét** és a hangkapcsolódások spektrális vetületeit definiálni kell.
- B) Az Fo(t) függvényt (**beszéddallam, hangsúlyozás**) külön modulként kell kezelni.
- C) Az I(t) függvényt (a **beszédjel intenzitásváltozása** mondat, frázis, szó, hang szintjén) külön modulként kell kezelni.
- D) A **beszéd időszerkezeti** függvényét T(t) (hangidőtartamok, ritmusváltozás az artikulációs sebesség változtatásával, tagolási pontokon szünetek generálása, ezen szünetek hosszának meghatározása) külön modulként kell kezelni.
- **AKUSZTIKAI FOLYTONOSSÁG: a fenti 4 alkotórész mindegyikére érvényesíteni kell a folyamatosság (interpolálás) elvét.**

Gépi beszédkeltés alapfogalmai: három kategóriát különböztetünk meg

- **Kötött szókészlet:**
  - tudjuk, hogy a rendszernek mit kell majd mondania
  - állandó üzenet („a hívott szám nem elérhető”, kiterjesztett magnetofon)
  - változó elemek
    - primitív: „Önnek üzenete érkezett 2000 május”
    - bonyolultabb: „A hívott szám megváltozott, az új szám: 325-29-48”

- **Kötetlen szókészlet (text to speech, szövegfeldolgozó)**
  - gyakorlatilag ilyen nincs
  - széles szókinccsel kell rendelkeznie kiinduló állapotban, és tetszőlegesen bővíthető
  - ha tudjuk a tematikát, kifejezéseket, akkor meg lehet tanítani

- **Vegyes rendszerek**
  - vannak állandó üzenetek (ezeket nem célszerű TTS-sel megoldani, mert fárasztó)
  - vannak változó üzenetek

szókészlet * minőség = konstans

Kötött szótáras beszédszintetizátor

Adott, előre meghatározott üzenetek kimondására alkalmas. A kimondandó üzeneteket emberi felolvasásból készítik el.

Tervezési lépések:

- Az összes üzenet feltérképezése: **állandó / változó üzenetrészek**. A változókra kell koncentrálni, azokat kell akusztikailag hozzáilleszteni az állandó üzenetekhez, hogy **együttesen** is jól érthető információt kapjunk. *„A tegnapi leveleinek száma: 23."*

- A bemondó által felolvasandó szövegek összeállítása. Fontos: a szintetizátor adatbázisában eltárolt mondatok és az adatbázis elkészítéséhez felolvasott mondatok **nem ugyanazok**. Meg kell vizsgálni: milyen vivőmondatba kell a tényleges (majdan elhangzó) üzenetet beágyazni? Cél: az elhangzó üzenet teljes hosszában (az üzenet állandó és változó részében is) jó legyen az intonáció (dallam, ritmus, intenzitás).

- Összefűzési szoftver elkészítése: egy szabályrendszert valósít meg (mit, mikor, mivel kell összefűzni).

Kötött szókészletű rendszerek tervezési szempontjai

- **Emberi hangot** tárolnak és abból építkeznek, hangelem-összefűzéssel.
  típusai: mondat, szófüzér, szó.
- **Tematika felderítése**
  - az adott rendszeren mik azok az információk, melyeket el kell juttatni a felhasználóhoz
  - mik ezeknek a módjai
  - a felhasználók figyelembevétele (kezdő + profi különböző)
- **Bemondandó szöveg tervezése**
  - Szabály: ami betű szinten összekapcsolható, az nem biztos, hogy hangzás szinten is jól fog sikerülni.
  - Vivőmondatokba kell ágyazni az egyes eltárolandó elemeket. Így biztosítható, hogy az akusztikai folytonosságot meg tudjuk közelíteni az elemösszefűzésnél is.
  - A vivőmondatot kell felolvastatni a bemondóval.
- **Szótárkészlet kialakítása**
  - ☐ az előzővel szinkronban
  - ☐ kompromisszum a minőség és a bonyolultság között
  - ☐ szótárelemek számára algoritmus
- **Bemondó választása**
  - ☐ akusztikai arculat (igényes cégnek saját „hangja” van)

- **Akusztikai adatbázis elkészítése**
  - **felvétel készítése**
  - **elemek kivágása** a vivőmondatokból. Behelyezés az elemtárba.
  - **tesztelés** (az összeillesztett elemekkel az üzenetek meghallgatása), szükség esetén **akusztikai csiszolás** (hangsebészet, intenzitás illesztés, Fo illesztés, ritmikai javítások az elemekben). Az így csiszolt elemtárat kell beépíteni a véglegesen kialakított rendszerben.
- **Rendszerbeillesztés**

- **Tervezési problémák:**

• A rendszer később nehezen bővíthető: új felvétel szükséges, a bemondó hangja szükségképpen változik, ezért el fog ütni a korábban felvettekétől.

- **Kész rendszer üzembevétele esetén fellépő kérdések**

• Külföldi rendszer átvétele, honosítása esetén: hogyan lehet magyarítani (hazai fejlesztésben meg lehet-e oldani)

• Műszaki megbízhatóság + akusztikai arculat (szépen beszéljen, a beszéd és szünet aránya jó legyen)

• Célszerű akusztikai / fonetikai szakértő alkalmazása az elembázis összeállítására

Konkrét példa: Számfelolvasó (példa tipikus kötött szótáras rendszerre)

- **Hagyományos megoldás:** az írás szintjén meghatározható számelemek összefűzése szünetekkel. Eredmény: döcögős, kiegyenlítetlen, nehezen érthető hangzás.

  Ma, ennek ellenére a legtöbb rendszerben ez működik (telefonszámok bemondása, árak, időpontok, dátumok felolvasása, banki információ szolgáltatása, stb.)
- **A korszerű megvalósítás alapelve:** a számelemek fonetikai kapcsolódásának figyelembe vétele. Ehhez olyan hangfelvételt kell készíteni, amelyből minden számelem minden pozíciójára és kapcsolódásaira folyamatos akusztikai illeszkedés hozható létre.
- **A felvételhez össze kell állítani** olyan számsorokat (400-500 többjegyű szám), amelyekben az összes számelem előfordul az összes lehetséges pozícióban (számcsoport eleje, közepe, vége) és az összes lehetséges hangzókörnyezetben, egynél többször, hogy több választási lehetőség legyen.
- **Tagolás:** A számokat a nyelvnek megfelelő szabályok szerint tagoljuk (312 = három száz tizen kettő; drei hundert zwölf). A vezérlő (összekapcsoló) algoritmus is nyelvfüggő.

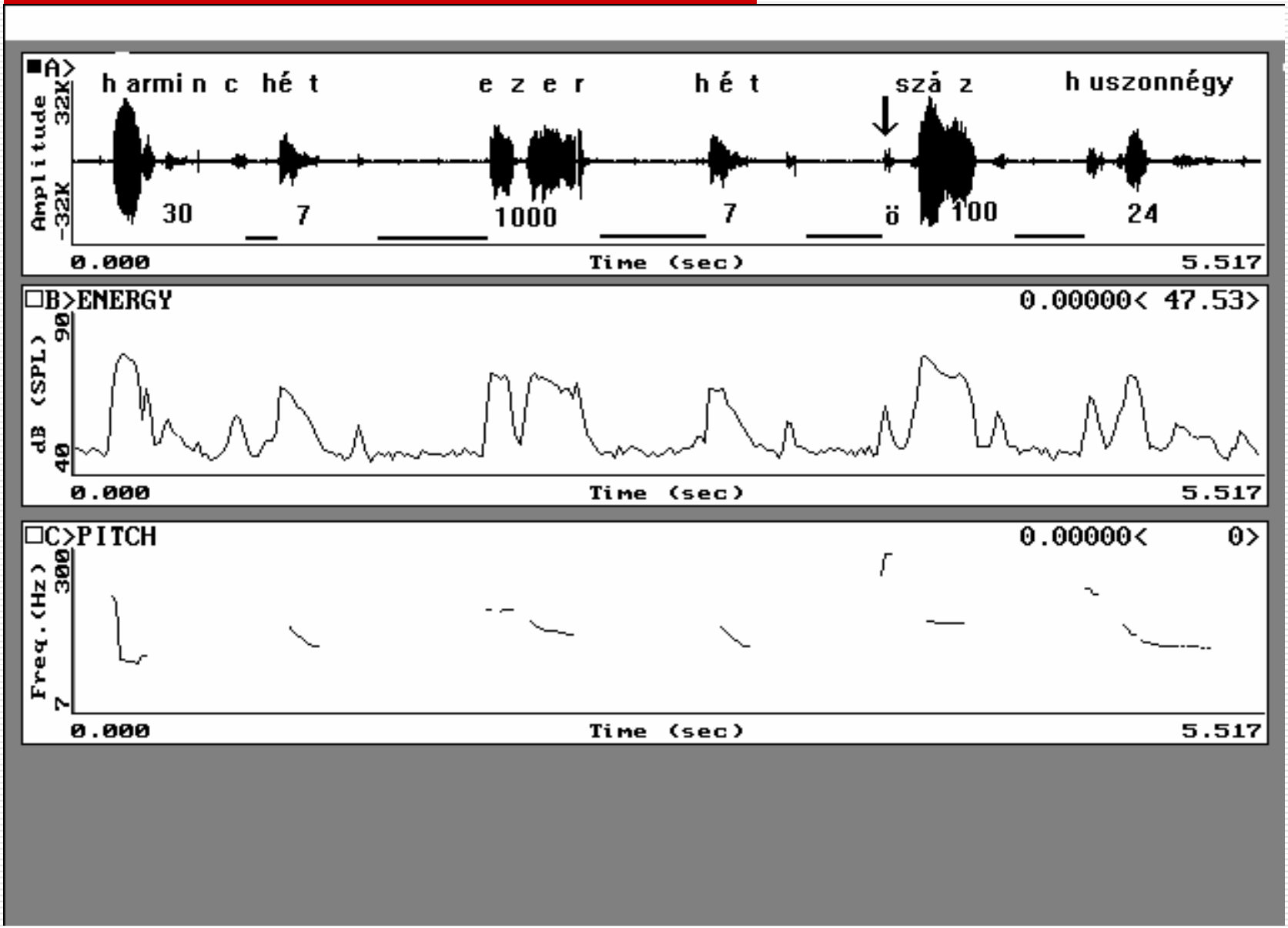
Pl: 125000:

(English)	one	hundred	And	twenty	five	Thousand
(German)	ein	hundert		fünf	und	zwanzig tausend
(Hungarian)	száz	huszon		Öt	ezer	
(Portuguese)	cento	e vinte		cinco	mil	

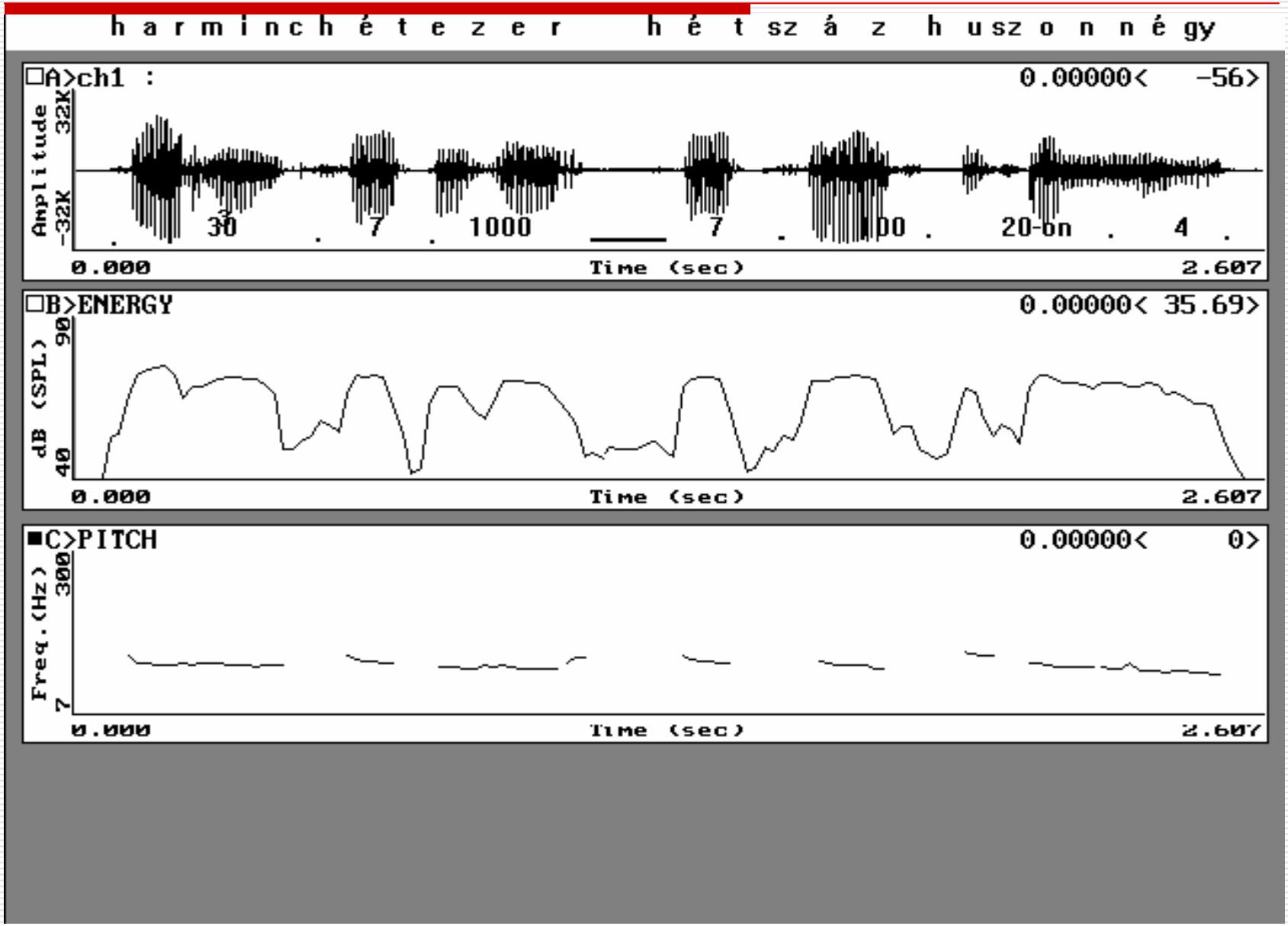
Magyar: 25 db elem, portugál: 53 db elem

	Basic element	English	German	Hungarian	Portuguese
1.	1	one [ω ℘ ν]	ein [αι ν]	egy [εϑ]	um [ũ]
2.	1--	--	eins [αι ν σ]	--	--
3.	1--	--	eine [αι ν ↔]	--	--
4.	2	two [τυ:]	zwei [τσ ω αι]	kettõ [κΕτ:Ο:]	dois [doj∫]
5.	3	three [Τρι:]	drei [δ ρ αι]	három [ηα:ρομ]	três [tre∫]
6.	4	four [ϕ□:]	vier [ϕ ι: □]	négy [νε:ϑ]	quatro [kwatru]
7.	5	five [ϕαιϖ]	fünf [ϕ Φ ν ϕ]	öt [Οτ]	cinco [s ĩ ku]

Hagyományos megoldás



Újrendszerû számfelolvasó



Számelemek kiejtésének fonetikai vizsgálata:

- **A vizsgált paraméterek:**
  - a) Spektrális folytonosság (koartikoláció) az elemek kapcsolódó hangjainál
  - b) ritmikai változás (az elemek hossza a pozíció függvényében)
  - c) intenzitásszerkezet (az elemek intenzitásszintje a pozíció függvényében)
  - d) a hangsúlymegvalósulás (az elemek hangsúlyszintje a pozíció függvényében)

- **Eredmény:**
  - a)-ra: egyedi elemek kialakítása minden számelem kapcsolathoz megfelelő kategorizálás után.
  - b), c)-re három kategória: kezdő, belső, utolsó (befejező, felsoroló)
  - A d)-re: a számérték-elemek hangsúlyosak, a száz, ezer, millió belső helyzetben nem hangsúlyos

- **A természetes hangzású számfelolvasási szintézishez biztosítani kell:**
  - a folyamatos kiejtést, helyes pozíciójú és hosszúságú szünetekkel,
  - a számelemek kiejtési helytől függő időszerkezetének megvalósítását,
  - a spektrális- és intenzitás-folytonosságot (koartikuláció figyelembe vétele) az elemhatárokon, szóhangsúlyok és alapfrekvencia változások helyességét.

Folyt.

- **A számelemek kiejtési helytől függő időszerkezete**

  Kezdő (B, beginning, pl. **1**234567), középső (M, middle, 1**23156**7), záró (L, last, 123456**1**) elem szükséges a többi szempont szerint kiválasztott minden elemből (elvileg).nagyszámú (közel ezer) kimondott szám vizsgálata alapján

- **Spektrális- és intenzitás-folytonosság (koartikuláció figyelembe vétele) az elemhatárokon**
  - Minden elemre hat az előző és a következő elem. A kédés, hogy mely esetekben kell ezért külön elemet kialakítani és eltárolni.
  - Lehetséges pozíciók:
    - ☐ Egyedül áll a szám (6)
    - ☐ Felsorolás (12, 2, 56.)
    - ☐ Első (elemXXX)
    - ☐ Belső (XXXelemXXX)
    - ☐ Záró (XXXelem)

Az 1 számelem példája

- Magyarul *egy*:
  - (1) szabály: *egy* egyedül áll (1, 2, 3 stb.), *egy*XXX
  - (2) szabály: *egy millió* és *egy milliárd*)
  - (3) szabály: *egy ezer*, pl. 31000), zöngés alveolo-palatális zárhang és magánhangzó találkozása,
  - (4) szabály: *egy száz* pl. 3125000, zöngés alveolo-palatális zárhangot zöngétleníti a *száz sz* hangja,

  XXX*egy*
  - (5) szabály: ....*n egy*, pl. 51, 61, 71, etc.) a nazális hang módosítja az *e*-t,
  - (6) szabály: *millió egy*, pl. 5000001) magánhangzó-magánhangzó kapcsolat.

  XXX*egy*XXX
  - (5) + (2), (5) + (3), (6) + (3), (6) + (4)
  - Összesen: 10 (1+3+2 +4) elméleti lehetőség.

Koartikulációs szabályok

- **A legfontosabb regresszív koartikulációs zabályok**

az előző elem utolsó hangja	az alábbira változik, ha	a következő elem első hangja
b, d, g, v, z, 3	p, t, k, f, s, S	zöngétlen
Ts	ts felpattanás (burst) nélkül	S
T	t felpattanás (burst) nélkül	N
N	n(k)	K
n	n(h)	H
N	Nn	N
N	):	)
N	M	m, b, p
Magánhangzó	átmeneti szakasz	magánhangzó
Magánhangzó	palatalizált átmeneti szakasz	Palatális

- **A legfontosabb progresszív koartikulációs szabályok**

ha az előző elem utolsó hangja	és a következő elem első hangja	akkor a következő elem leső hangja az alábbira változik
Nazális	magánhangzó	nazalizált átmeneti szakasz
Palatális	magánhangzó	palatalizált átmeneti szakasz
Magánhangzó	magánhangzó	Átmeneti szakasz

Szóhangsúlyok és alapfrekvencia változások helyessége számfelolvasásnál:

sample number	pronounced style	comment
121	onehundredandtwentyone. AB N N AM AL	.=full stop AL= accent and falling intonation in the last item
2151	twothousandonehundredandfiftyone. AB N AM N N AM AL	AB, AM= accents in the number

A számok kimondásakor több hangsúly is megjelenik:

AB: kezdő hangsúly
AM: közbenső hangsúly
AL: záró hangsúly, eső intonáció
N: semleges, hangsúlytalan elemek

- Ha a számelem a mondat végén áll, (pl. Az ön számlájának egyenlege: **53424** forint) **eső jellegű intonációja lesz**.
- Ha a mondat közepén helyezkedik el, (pl. Az ön számláján **53424 forint** összegű tranzakció valósul meg.) a számelem **intonációja laposabb**, **lebegőbb.**

A számkimondó megvalósítása

- **Előzmény**: az elemi (hagyományos) **építőkockák**, számelemek meghatározása a **kimondási szabályrendszerek** (időtartam, koartikulációs,

  hangsúly és intonáció) meghatározása
- ***A felolvasandó szöveglista meghatározása.***

  Vivőszöveg kialakítása az építőkockák és a szabályrendszer alapján.
- ***A felolvasandó szöveg felvétele***

  A vivőszöveget célszerű redundánsra tervezni (minden elem legalább kétszer forduljon elő).

  Az egyes elemek között kb. 1 sec szünetet célszerű tartani.

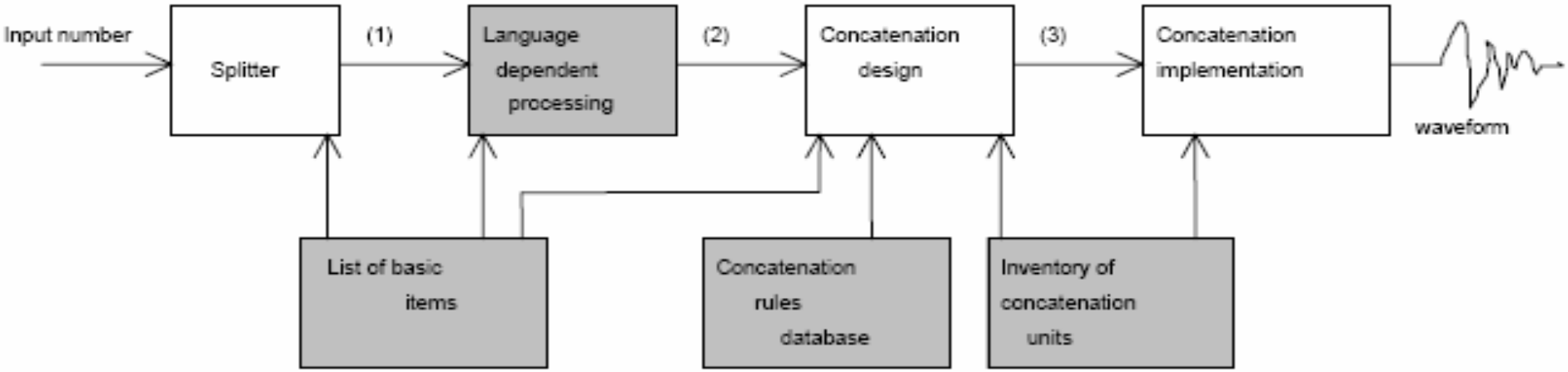
  Nagyobb egységenként (pl. oldalanként) érdemes hosszabb szünetet tartani a felolvasásban.

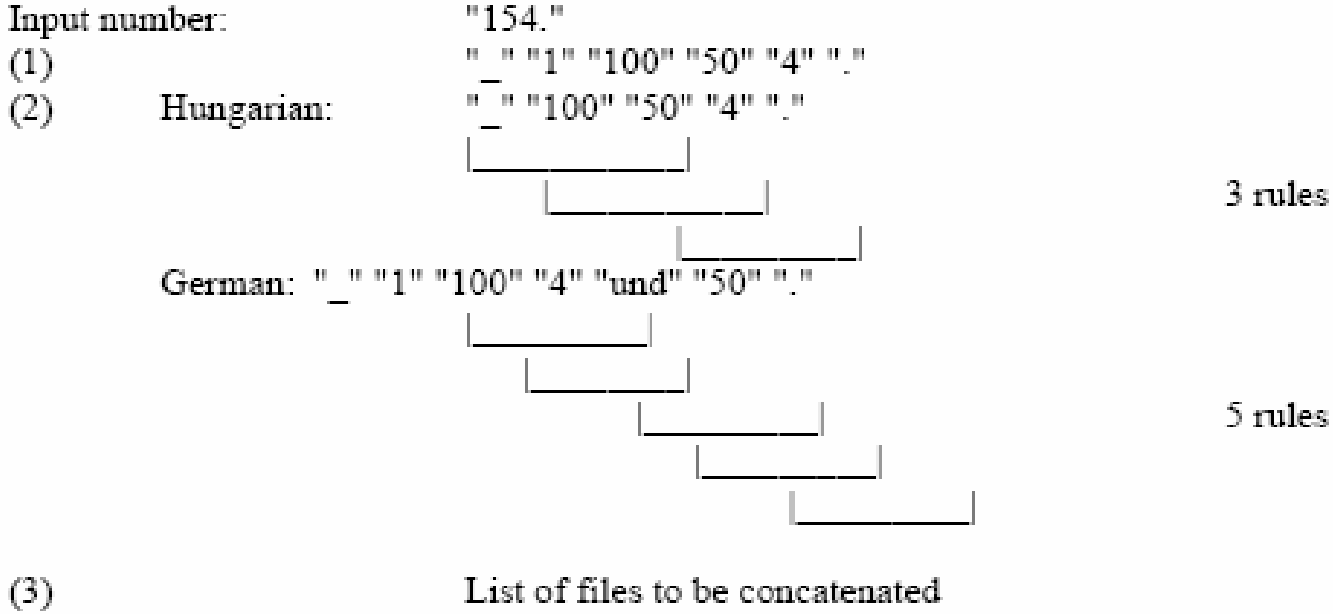
  Összpontosítás a szám természetes ritmusú, dallamú, hangerejű kimondásához.

A hangelemek kivágása a felolvasott vivőszövegből

- Kivágás előtt a felolvasás helyességét ellenőrizni, hiba esetén a redundáns elemet kell elővenni.
- Időbeli (esetleg spektrális) vizsgálat alapján a számelemek határait meg kell állapítani.
- Az elemeket ki kell vágni a vivőszámból, el kell menteni az építőelem lista és a szabályrendszernek megfelelő logikus rendben (adatbázis, könyvtárstruktúra, stb.)

Egy lehetséges megvalósítási struktúra





Block diagram of NTS algorithm implementation

- **A működő rendszer tesztelése és javítása**

  Analízis szintézissel módszer (analysis by synthesis).

  (**Olaszy törvény (1984):** Ahhoz, hogy egy beszédszintetizátor fejlesztéséhez hozzá tudjunk kezdeni, első lépésben létre kell hozni egy beszédszintetizátort.)

- A durvább hibák megszüntetése után vesszük észre a finomabb hibákat.
- Szint-, időtartam-/sebesség-, hangzáskiegyenlítés, hangsebészet alkalmazásával.

Szövegfelolvasó rendszerek

- **CSAPAT MUNKA!**
- Két komponensből épül fel: **agy** (szabályrendszer, tudásbázis, adatszintű előkészítés), majd a **megvalósítás** (hangelemek, adatbázis és jelfeldolozás)
- **Elvárások:**
  - --Úgy szóljon, mint egy rádióbemondó produkciója
  - --Úgy szóljon, mintha egy adott személy beszélne
- **Szövegfelolvasó (text to speech):** adott nyelv köznapi szókincsében előforduló szövegek felolvasása (kb. egy 8-10 éves gyerek szókincsének megfelelő)
- **Üzenet felolvasó (concept text to speech):** a kifejezni kívánt üzenetre vonatkozó jelekkel ellátott szöveg felolvasása
  - pl. [Conf_Req] A gépkocsi típusa [Car_Type] Volgswagen Golf
- **Többnyelvű TTS (miltilingual):**
  - azonos építőelemek minél nagyobb halmazának egységes keretben történő felhasználása
  - Ideális esetben azonos program kód (ami cél és nem pedig a valóság)
  - egységes szerkezetű, külső adatbázisban

- **Poliglott TTS:** azonos hangon szóló TTS
  - zseniális paraméteres leírás, mely nincs emberi hanghoz kötve
  - egy bemondó sok nyelven mondja el a szöveget

- **Kötött tematikájkú (domain specific) TTS:** csak egy adott témakörű (pl. menetrend, időjárás, szállodafoglalás) szöveg felolvasására alkalmas rendszer. Átmenet egy hagyományos kötött szókészletű és egy TTS rendszer között.

Osztályozási szempontok

- **milyen nyelveken** szeretnénk felolvastatni
- **milyen szövegeket**
  - **szövegtípus:** általános, szakszöveg, e-mail, SMS, stb.
  - milyen **mondattípust** tudjon megszólaltatni: kijelentő, kérdő, felkiáltó, egyéb érzelem kifejezése, CTS
- **milyen minőséget célozunk meg**
  - **érthetőség** : intelligibility
  - **természetesség**: naturalness és ezek nem is feltétlenül korrelálnak egymással
- **milyen hangokon** – egy illetve több hangon, illetve például amit kiemelünk, más hangon szóljon

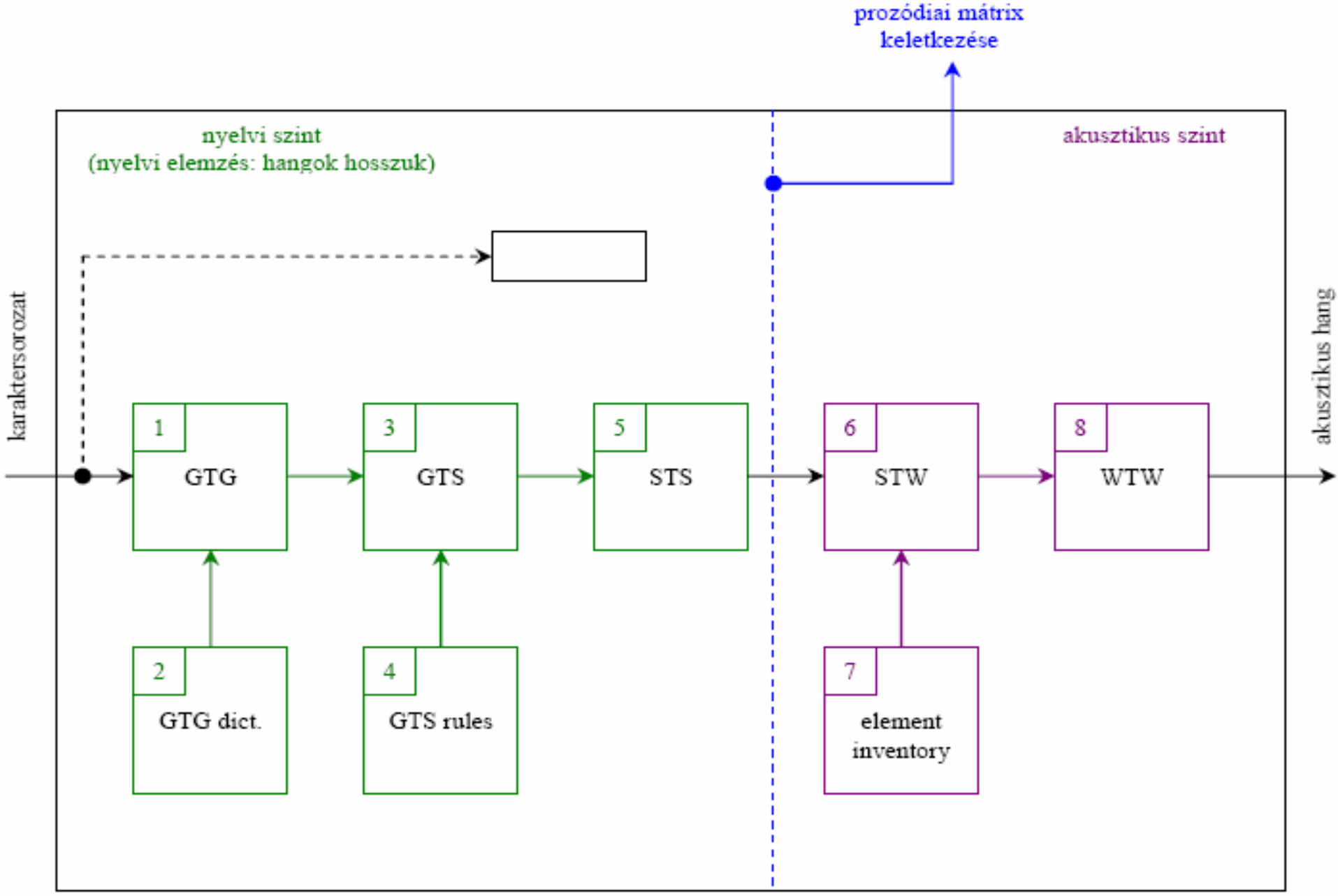
- milyen paraméterek állíthatók:
  - sebesség
  - hangmagasság
  - suttogás
  - rekedtség
  - szünetek hossza
  - betűzés
- milyen platformokon fusson:
  - hardware
  - operációs rendszer (Windows, Unix, OS/2)
  - erőforrásigény, csatorna – nem mindegy, hogy mobiltelefonban vagy távközlési központban
- milyen vezérlési felületeket kell biztosítani, API-k
- bővítési, továbbfejlesztési lehetőségek – mit ad hozzá a felhasználó és mit a fejlesztő, pl. rövidítésfeloldó

Felépítés: néhány alapprobléma

- **Az írás diszkrét**, a szavakat szünetek választják el. A beszédben a szavak **folyamatosan következnek egymás** után, csak nagyobb egységeket (prozódiai egység) választ el szünet. A beszédben a folyamatosság megértése teszi nehézé a megértést.

- **Az írott hibákat másképp kezeljük**: az akusztikus formára sokkal érzékenyebbek vagyunk.

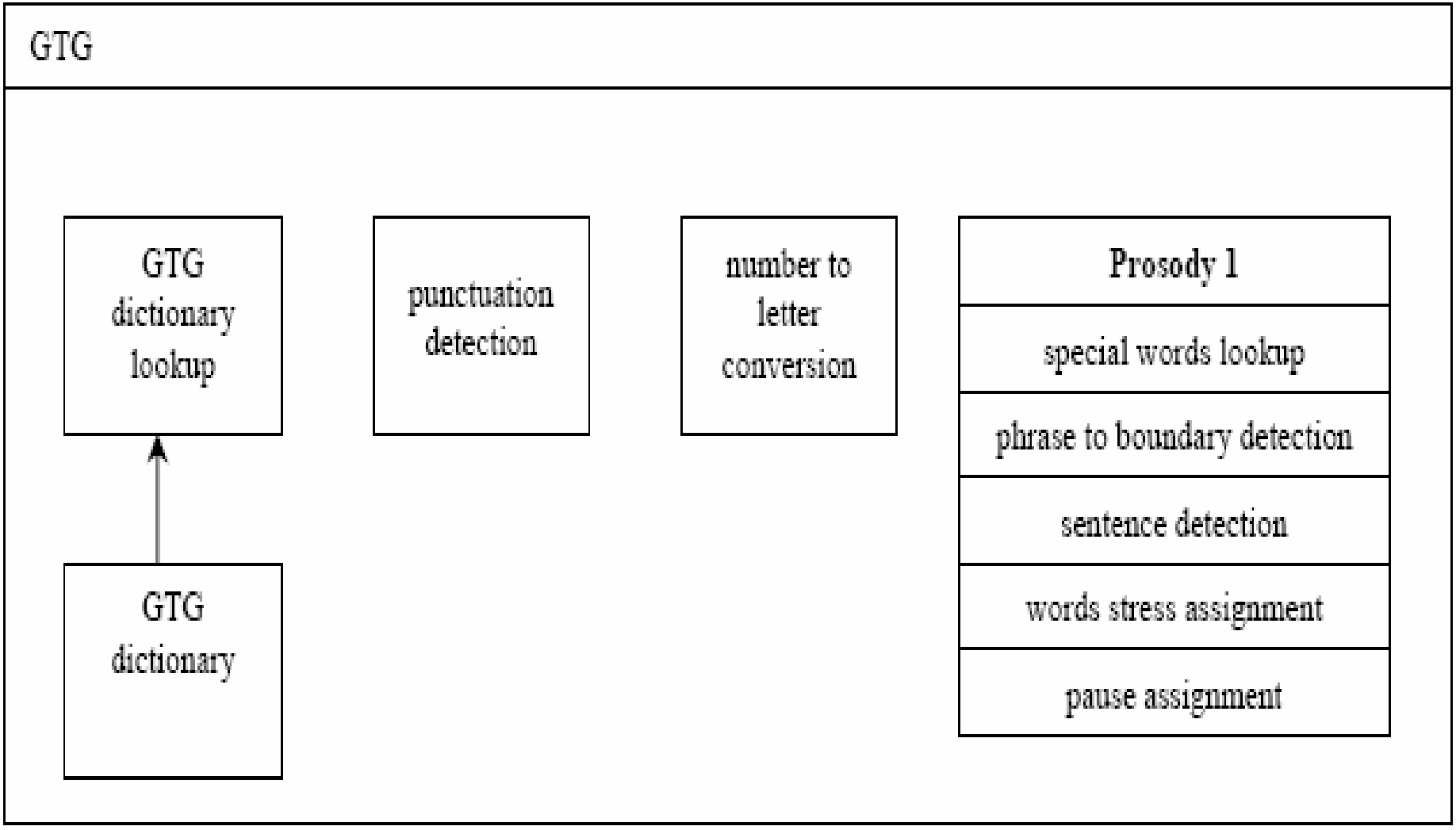
- **Fontos**: a TTS bemenetére minél helyesebb és minél részletesebb szókimondást segítő információt tartalmazó jelsorozat érkezzen.


prozódiai mátrix
keletkezése
nyelvi szint
(nyelvi elemzés: hangok hosszuk)
akusztikus szint
karaktersorozat
akusztikus hang
1
GTG
3
GTS
5
STS
6
STW
8
WTW
2
GTG dict.
4
GTS rules
7
element
inventory

Magyzarázat:

- 1. GTG : Grapheme to grapheme (írásjel→betű)
- 2. GTG dict. : GTG dictionary (szótár)
- 3. GTS : Grapheme to Sound (betű→hang)
- 4. GTS rules : szabály és szótár
- 5. STS : Sound to Sound (hang→hang)
- 6. STW : Sound to Wave (hang→hanghullám)
- 7. element inventory : hangelem-tár, akusztikai adatbázis
- 8. WTW : Wave to Wave (hanghullám-feldolgozás)
- ☐ 1-5 elvileg lehet nyelvfüggetlen, viszont 6-8 mindenképpen nyelvfüggő
- ☐ ha igazán általánossá akarjuk tenni, akkor nagyon bonyolult és nagy leíró nyelvre van szükség

GTG

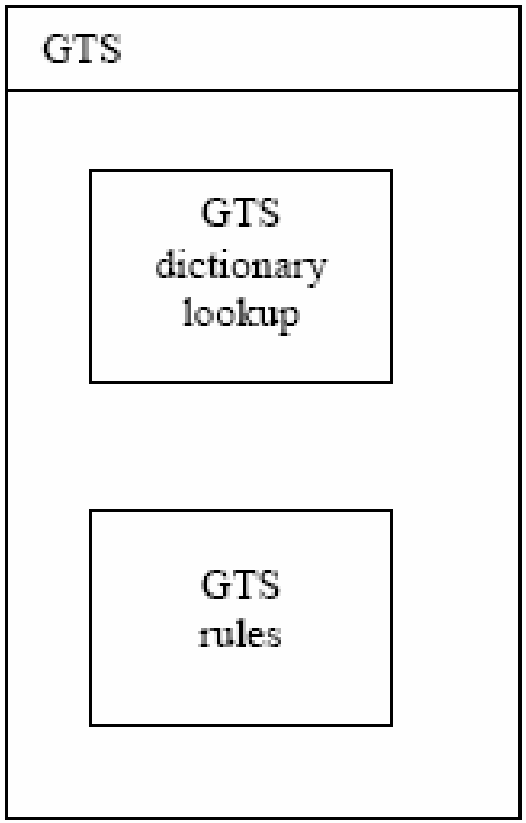

GTG
GTG dictionary lookup
GTG dictionary
punctuation detection
number to letter conversion
Prosody 1
special words lookup
phrase to boundary detection
sentence detection
words stress assignment
pause assignment

Tehát többszintű nyelvi elemzés szükséges

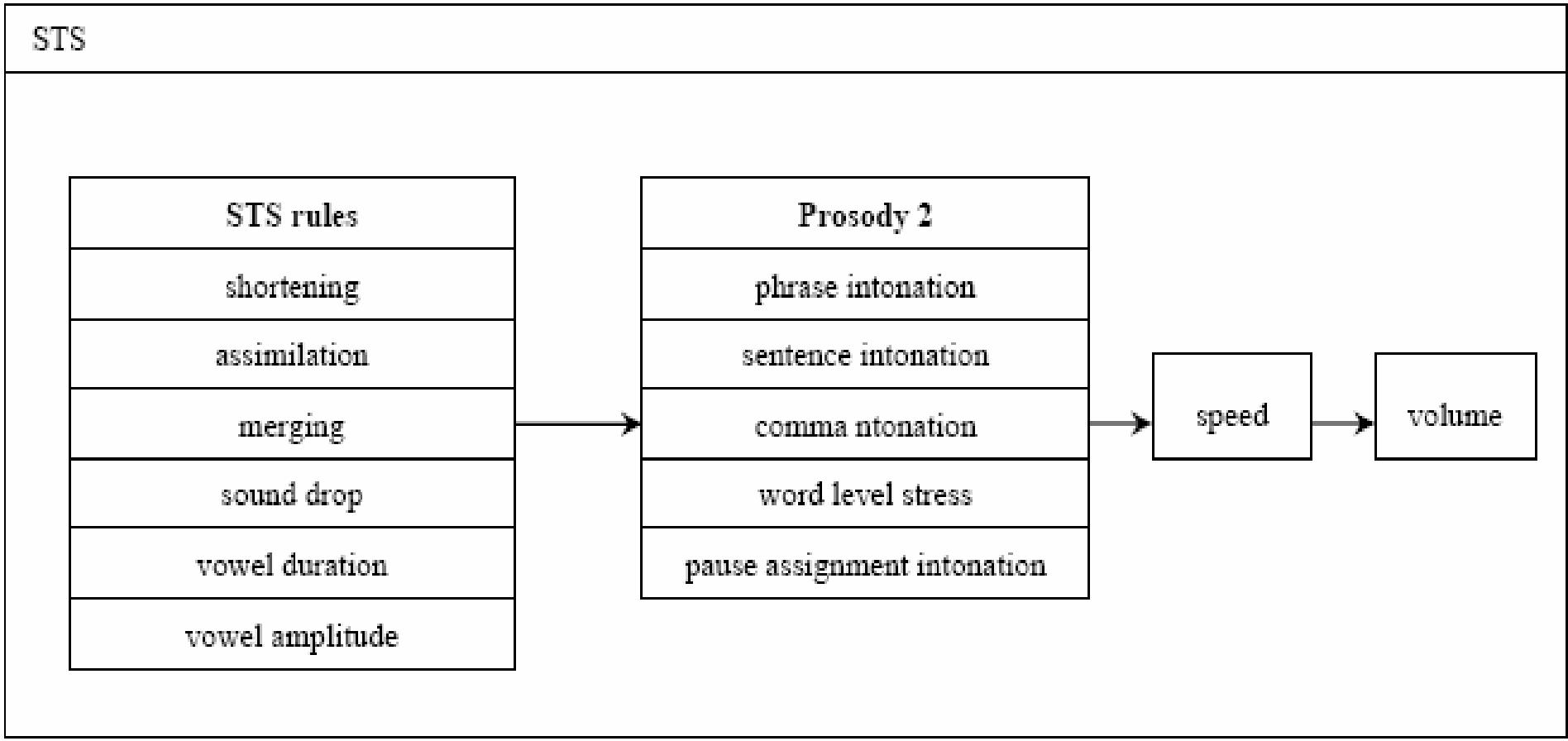
- **punctuation detection**: azért fontos, mert pl. egy mondatban pont sok helyen lehet (rövidítések, …, mondat vége, stb.)
- **special words lookup:** pontosvessző, pont, vessző, zárójel, csillag, bizonyos dolgokat nem mindig akarunk hallani
- **phrase boundary detection:** egységként kimondható szavak, frázisok között mikor tartunk szünetet
- **sentence detection:** intonáció egy mondatra
- szövegfelolvasó mit dolgozzon fel egy egységként (az egész egy mondat)

- **Írott szövegnek nem egy az egyben felelnek meg a kimondott hangok**
  - ☐ hasonulások
  - ☐ hangkiesések (mondtam=> MONTAM)
  - ☐ hangbetoldások (fiú => FIJÚ)
  - ☐ röviden írjuk, hosszan ejtjük és viszont (újítás => UJJÍTÁS)
  - ☐ mássalhangzó torlódások (hármas, négyes)
  - ☐ betűkép helyes értelmezése szó illetve morféma határon (malacság, egészség, táncstúdió)

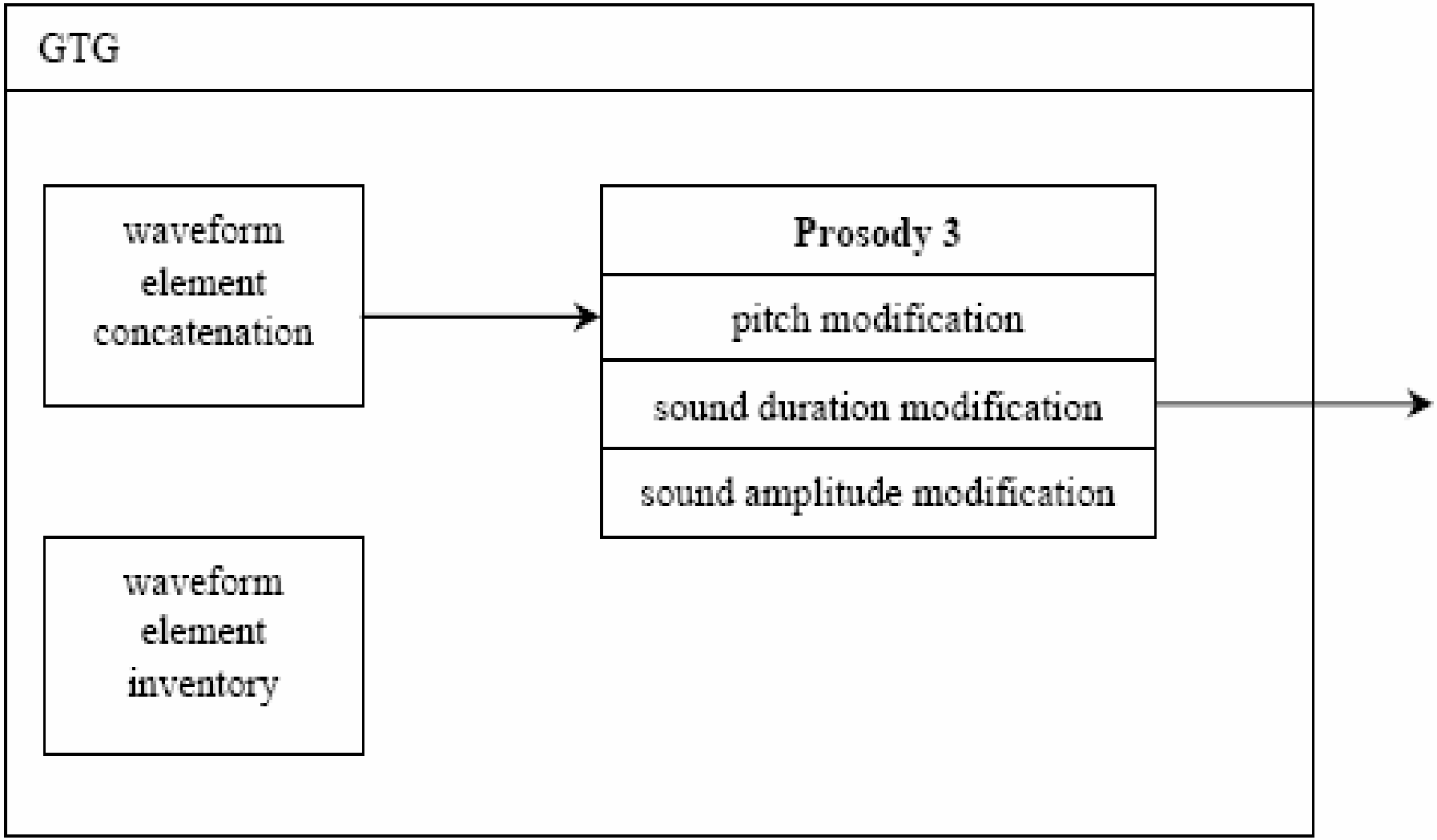
GTS



STS



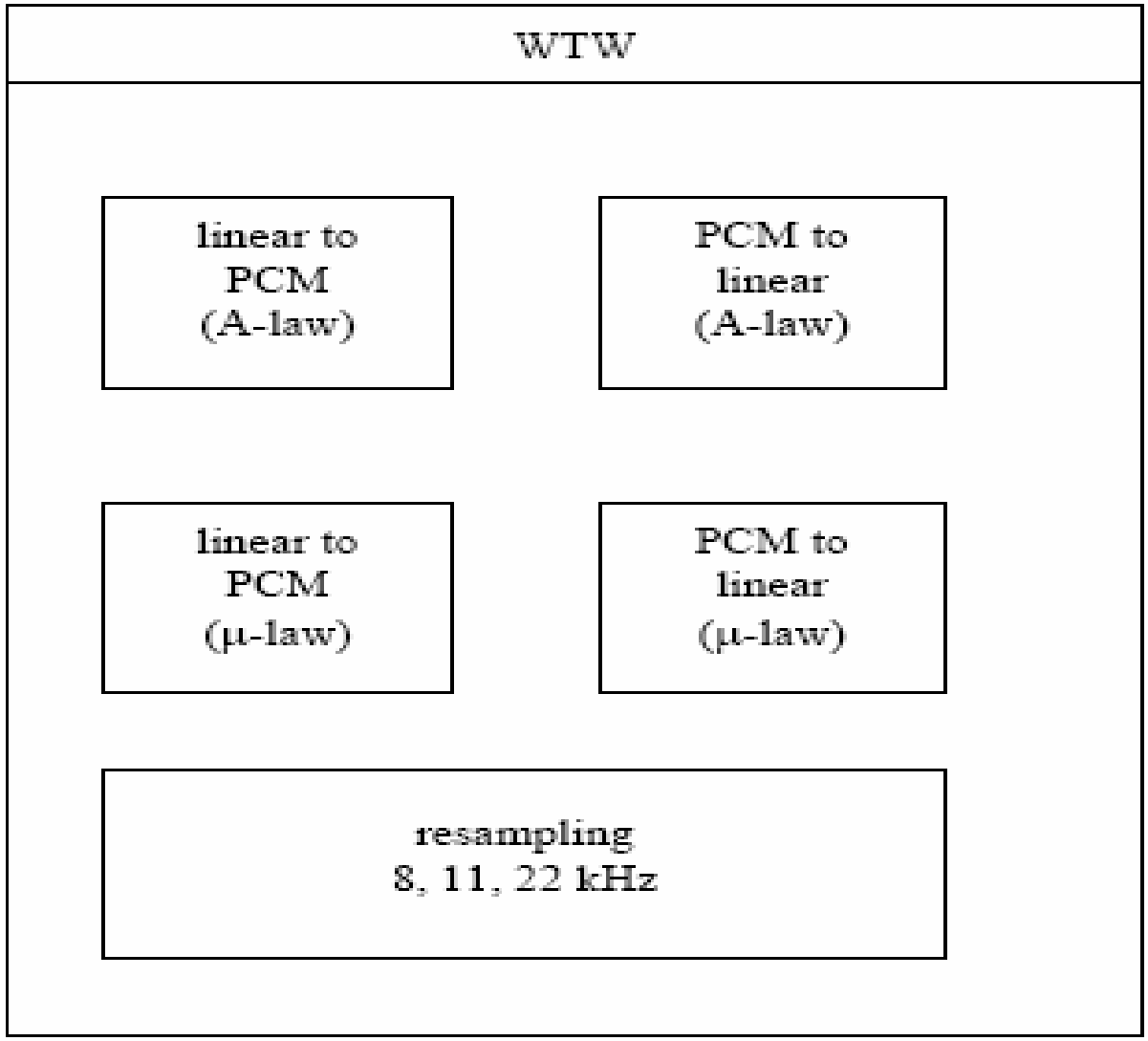
- **speed:** szünetek kivágása (vigyazni kell vele, mert ha figyelmetlenül vagdossuk ki a szüneteket, nem ugyanazt a hangot kapjuk)
- **prosody 2:** magasszintű leírás


GTG
waveform element concatenation
Prosody 3
pitch modification
sound duration modification
sound amplitude modification
waveform element inventory

- **waveform element inventory:**

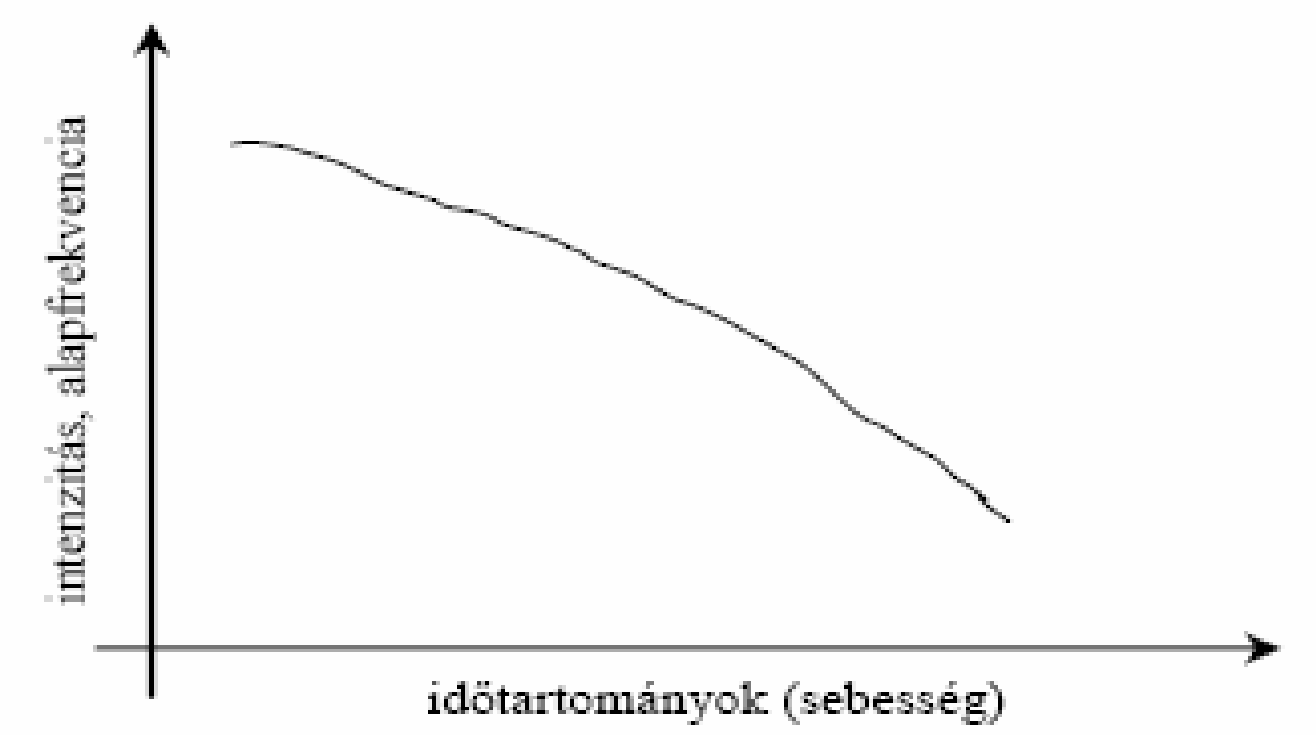
  valamilyen akusztikai adatbázis

- **prosody 3:**
  - a jó minőségű szövegfelolvasók esetén kulcsfontosságú
  - a prozódiai mátrixban előírtakat el kell végezni (módosítások és folytonosság biztosítása)


WTW
linear to PCM (A-law)
PCM to linear (A-law)
linear to PCM (μ-law)
PCM to linear (μ-law)
resampling 8, 11, 22 kHz

Néhány elvi probléma

- **alapvető feldolgozási egység:**
  - **ember esetében** a mondatnál nagyobb
  - mondat egy **gépi megoldásban**: a prozódiai algoritmusok legmagasabb szintje is a mondat
- Pl. Kijelentő mondat eső jellegű, ez azonban humán megközelítés, mérnöki módon hogyan fejezzük ezt ki



- **A prozódia három megvalósítási paramétere**, amelyek különböző nyelvi szinteken fejtik ki hatásukat(szótag, szó, mondat, szöveg)
  - intenzitás
  - alapfrekvencia
  - időtartománybeli jellemzők

- **absztrakciós szintek:**
  - hang szint (legalacsonyabb)
  - szótag szint
  - szó szint
  - tagmondat szint (prozódiai fázis)
  - mondat
- **ezekre mérhető, fizikai paramétereket kellene találni**
- **beszédfelismerés kulcsterülete:** a sokrétű szinteket megkülönböztetni, elválasztani és a beszédben ezek folytonosak.

Prozódiai címkézés szöveg szinten

- Címkézni kell a **mondatdallam változását**, a **hangsúlyokat**, a **ritmikai változásokat**, a **szüneteket**, az **intenzitás változásokat.**
- Példa az Fo változások címkézésének megtervezésére.
- Az Fo változást két szinten címkézzük:
  - a) mondatdallam leírása, jelölése
  - b) hangsúlyozás leírása, jelölése

    - **Mondatdallam Fo építőelemek:**
      - lineárisan emelkedő fokozatok (13 féle adott kezdő és végponttal, XY jelöléssel)
      - szinttartó (6 értékkel)
      - lineárisan ereszkedő fokozatok (13 féle adott kezdő és végponttal)
  - **Mondatdallam mátrix:** az építőelemek számazonosítóval történő ellátásához mátrix ábrázolást alkalmazunk (sor=X, oszlop=Y). Összesen 32 féle dallamelemet helyezünk el a mátrixban. Ezeket lehet használni a címkézés során. Ezekkel az elemekkel a legtöbb magyar mondat dallama leírható.

- **Dallamelem címkék:** bárhol elhelyezhetők a szövegben, hatásuk a következő zöngés hangtól kezdődik.
  - /XY = folyamatos Fo kapcsolódási pont
  - //XY= szünettel megszakított Fo kapcsolódási pont

- **Hangsúlyozási építőelemek:** Fo csúcsok meghatározott csúcskiemelkedéssel, felfutással és lefutással
  - [Fz]= fókusz hangsúly fokozatok (z=fokozat), + 25%- + 50% Fo csúccsal (3 fokozat)
  - [Wz]= hangsúly, +15%-+25% Fo csúccsal (3 fokozat)
  - [-]= negatív hangsúly (-10%-20% Fo csökkentéssel), 2 fokozat
  - [N]= nincs hangsúly jelölés (hangsúlytalan eset)

Példamondat a címkézésre:

//11[W1]*Sült* [N]*csirkét* /24[N]*rendelt* //36[N]*és* [N]*még* /26[W1]*bort* [N]*is* [N]*hozatott.*

//11[W1]*Tizennyolc* /24[N]*éves* [N]*lehetett*, [-]*mikor* [-]*az* /26[W1]*apja* [W1]*meghalt.*

Megoldási stratégiák

Elektronikus megoldások: Kísérletek mechanikus rendszer készítésére (Japán, 2000-től)

Szabályalapú elektronikus megoldás

- lebontás: címkézés
  - **mondat:** kijelentő, kérdő, felkiáltó
  - **szó:** alany, állítmány, tárgy, határozó, jelző; hangsúlyos/hangsúlytalan
  - **szótag:** hangsúlyos/hangsúlytalan
  - **hang:** szó eleje, szó közepe, szó vége; alacsony/mély hangrendű;
  - magánhangzó/mássalhangzó; hangkörnyezet stb.
- **szabályok megalkotása nagyon lassú** – hiba esetén a szabályok kijavíthatók
- a nyelv nem reguláris szerkezetű, vannak kivételek
- a nyelv változik, nem statikus

Gépi tanulás (machine learning)

- a gyakorlatból, mint nagy adatbázisból **kinyerjük a szabályokat** (nagyon nehéz konkrét szabályokat kinyerni)
- vegyünk sok, egymással összefüggésbe hozott, címkézett adatot – ez elég nagy adatbázis, címkézett szöveggel: **neurális hálóval** megvalósítva a rendszer következtetni tud
- a hosszú, absztrakciós szinten történő munkát kiváltjuk: sok adatban **korrelációk**, **összefüggések** keresése
- **Problémák:**
  - **adatbázis létrehozása:** több millió adatot kézzel kell felcímkézni (☹☐ hangszintig le kell menni)
  - a rendszer **jól működik arra az adatbázisra**, amelyre be lett tanítva, de a többire nincs garancia
  - milyen alapon ítéljük meg **egy rendszer jóságát** (Ezt egyrészről a belső tesztelő, másrészről a felhasználó dönti el!)

A valóságban a két módszer ötvözetét használják. Egy adott, zárt problémakört fed le a gépi tanulás módszere, a kimaradó halmazra valamilyen szabályalapú megoldást alkalmaznak.

Szövegfelolvasók minősítése

Szövegfelolvasók minősítésére **nincsenek kialakult szabványok**

- A szövegfelolvasók „beszédének” minősítésére ma még nincs egységesített, szabványos eljárás. A következőkben **javaslatot láthatunk** egy általános minősítési eljárásra, amelyben különböző szempontok figyelembe vételével lehet a szövegfelolvasók teljesítményét megadni.
- A minősítéshez szövegeket kell a szintetizátor bemenetére adni és meghallgatással kell az elhangzottakat - egy megadott skálán belül - leosztályozni. A vizsgálat lépcsőfokai a következők:

I. Alapteszt

Ezt az akusztikai alapvizsgálatot célszerű minden új szintetizátorra – függetlenül az alkalmazás területétől – elvégezni.

- **a) A beszédhangok akusztikai tartalma** – itt azt vizsgáljuk, hogy a szintetizátor által előállított beszédhangok, hangkapcsolatok megfelelnek-e a nyelv hangjainak, hangkapcsolatainak.
- **b) A beszédhangok hangzósági szintjei** – itt azt vizsgáljuk, hogy a szintetizátor által előállított beszédhangok intenzitása megfelel-e a nyelv hangjaira vonatkozó specifikus szinteknek (kiegyenlítette a beszéd, nincsenek-e benne túl erősen, túl gyengén hangzó hangok, hangkapcsolatok).

- **c) A beszédhangok időtartamai és a szünetek** – itt azt vizsgáljuk, hogy a szintetizátor által előállított **beszédhangok időtartamai kiegyenlítettek-e**, megfelelnek-e a nyelv hangjaira vonatkozó specifikus időtartam arányoknak. (nincsenek-e a beszédben kirívóan hosszan „ejtett”, illetve túl rövid hangok, amikor a rendszer folyamatosan beszél). A szünet a beszédfolyamat szerves alkotórésze, fontos funkciója van a beszéd értelmezése, megértése szempontjából. A szünetek helyes megvalósítása a gépi szövegfelolvasás fontos eleme. A szünetek nagy részét az írásban is jelöljük a mondatvégi írásjelekkel, továbbá a vessző, kettőspont, pontosvessző, gondolatjel stb. jelekkel. A szünetek helyes megvalósítását percepciós teszttel kell ellenőrizni.

II. Nyelvi teszt

Ezt csak igényes, változatos nyelvi alkalmazásra szánt szövegfelolvasóknál kell elvégezni (e-levél- , regény-, fax-, név- és címfelolvasók). **A teszt célja, hogy megállapítsuk a szintetizátor kifejezési képességét** (a tervezők milyen mélységig építettek be nyelvi szabályokat a rendszerbe). **Itt a nyelv prozódiai sajátosságainak megvalósítását vizsgáljuk**. A vizsgálati területek a következők: mondatfajták (kijelentés, kérdés, felszólítás, óhajtás, figyelmeztetés), mondathossz (meg van-e oldva a mondatfajták és a mondathossz összefüggése), Folyamatos dallamvonulat megvalósítása mondatról mondatra szöveg szinten (pl. dialógusok, regényrészletek, elbeszélések felolvasásához).

Ennek a tesztnek az elvégzéséhez össze kell állítani egy olyan szöveget, amelyik tartalmazza a magyar prozódiára vonatkozó legfontosabb mintákat.

III: Funkcionális teszt

- Ennek a tesztnek a célja az, hogy az alkalmazandó felhasználási területhez kapcsolódó **speciális működéseket** vizsgálja.
- Speciális feldolgozásra van szükség például **elektronikus levelek felolvasásánál**, ahol a bejövő karakter sorozatot erősen át kell alakítani, mielőtt a szintetizátor bemenetére küldjük. Ki kell szűrni a nem szöveg karaktereket (csillagok, kötőjelsorozatok per jelek stb.), szét kell választani a szövegrészeket (cím, dátum, az üzenet szövege stb.), meg kell állapítani a szöveg nyelvét (az egész szöveg magyar-e, egyes szavak nem magyarok stb.)
- Hasonló feldolgozásra van szükség az **SMS felolvasók** tesztelésénél is, ahol például a felhasználók által használt (az SMS kommunikációban kialakult) betűsorozatokat kell kimondásra alkalmas szöveggé konvertálni.
- **Hangos könyv** szolgáltatásnál például azt kell vizsgálni. hogy a hosszú mondatokon belül a mondat tagolását milyen sikerrel végzi el a rendszer, továbbá, hogy a mondatok között milyen szüneteket tart.
- **A szövegben elhelyezett XML** jelzések feldolgozását is ellenőrizni kell (lehet, hogy a szintetizátor nincs felkészítve ilyen értelmezésre)
- **Képernyő olvasóknál** (vakok és gyengén látók számára kialakított rendszerekben) azt kell vizsgálni, hogy például a betűzési funkcióban a szintetizátor **hogyan mondja ki a betűket**, továbbá, hogy a beszéd gyorsítása megoldott-e stb.

`

Ajánlott olvasmányok:

• International Journal of Speech Technology 2000/3-4: beszédszintézissel kapcsolatos írások

• Acta Linguistica Hungarica 2002 dec. (MTA könyvtár

Linkek:

- Fonetikai vonatkozások, oktatási anyagok, Kempelen beszélőgépe
  http://fonetika.nytud.hu/
- **Beszédtechnológiai laboratórium**
  http://speechlab.tmit.bme.hu/postnuke/index.php

- **Köszönöm a figyelmet!**