1. ZH-k

# 2002.01.23

**1.a) Osztályozza a beszédhangokat a létrehozásukhoz használt gerjesztés**

**szempontjából! Jellemezze a beszédhangokat (spektrális-, intenzitás-, idő-)**

**szerkezetük szempontjából. (10 pont)**

Gerjesztés szerint lehet:

* • zöngés: az összes magánhangzó, b, d, g, gy, v, j, m, n, ny, l, r
* • zörejes: p, t, ty, k, c, cs, f, sz, s, j\*, h
* • kevert: dz, dzs, z, zs

Akusztikai szerkezet szerint:

* • Egyszerű: az összes magánhangzó, v, f, z, sz, zs, s, j, h, m, n, l
* • Összetett: b, p, d, t, g, k, gy, ty, c, cs, dz, dzs, ny, r

[Forrás: CD, 138. oldal]

Specifikus időtartamok:

• Magánhangzók: i,u,ü,o,a,e,ö,é, á (70 és 160ms között rendre)

• Mássalhangzók

40ms: r

50ms: n,l

60ms: z, zs, réshangok

70ms: p, t, k, ty

80ms: f, sz, s

90ms: c, cs

Intenzitás:

• Imin: h

• Imax: á,e

• Magánhangzók csökkenő sorrendben: á,e, a, é, ö,o, i, ü, u

**b) Mi az időablak szerepe a beszéd színképe elemzésében? Mi az előnye és mi a**

**hátránya a rövid és a hosszú ablaknak? (8 pont)**

TODO

Tehát amikor a változást követni szeretnénk egy teljes

beszédszakaszon, akkor az elemzést mindig egy meghatározott idıintervallumban, idı

ablakban(∆t ) kell elvégeznünk, és ezt az idıablakot végig kell görgetni a vizsgált

beszédszakaszon. Így kapjuk meg a gördülı teljesítményspektrumot, vagyis a

teljesítményspektrogramot

Előnye a rövidnek, hogy jól láthatóak az egyes komponenstek, de könnyen kihagyhatunk magasab ferekvenciákat: hosszú ablakon elkenődhetnek a lényegi információk passz

**2. Pontokba szedve írja le a tervezés és megvalósítás menetét egy HI-FI**

**számfelolvasó elkészítéséhez. (20 pont)**

Kötöttszótáras rendszer ésszerű.

1. tematika meghatározása -> számfelolvasás
2. felhasználók osztályozása: felhasználó, laikus
3. üzenetek meghatározása: mivel HIFI minőség kell, így igényesebb rendszereknél ajánlott az elemet megelőző és követő hangelemeknek megfelelő változatait is letárolni pl gyezer meg nyezer, nem csak ezer (így is max pár100 elem lesz), állandó meg tartalom nincs a specifikáció szerint
4. felolvasandó szöveg megtervezése, vivőmondatok: utóbbiak nincsenek, tehát elég az összes lehetséges hangelemkapcsolódásnak megfelelő számosságú üzenetet kiválasztani
5. bemondó kiválasztása: mindegy, csak ne a palik
6. hangfelvétel, HI-FI minőségben!
7. digitalizálás, ügyelni kell arra hogy a hangminőség ne romoljon, megfelelő bitráta stb. megválasztása
8. adatbázis elkészítése, elemek kivágása
9. próbaüzem, akusztikai csiszolás
10. rendszerintegrálás

(lépések röviden: : Tematika felderítése - bemondanadó szöveg tervezése - szótár kialakítása - bemondó választása - akusztikai bázis elkészítése - rendszerbeillesztés)

**3. Egy 8 kHz-es mintavételi frekvenciával és az alábbi H(f) karakterisztikájú**

**visszaállítóval működő mintavételező rendszer bemenetére a sas, majd a (gondolom faf) hangsor kerül egymás után férfi ejtésben állandó frekvenciával (F0: 125Hz).**

**1, ha 1 < abs(f) <= 3.5**

**H(f) = (4 – abs(f)) / 0.5, ha 3.5 < abs(f) < 4 a frekv. Mértékegysége [kHz]**

**0, egyébként**

F: zöngétlen réshang, nincsenek zörejgócok, egyenletes eloszlás a 1000-10000Hz

frekvenciatartományban. A környezetében levő magánhangzó formánsaira csak kis

mértékben van hatással.

S zöngétlen réshang: zörejelemek 1800-6500Hz között, intenzív zörejgóc ált. 2500-

3500Hz között. Az s hangot követő magánhangzó formánsaiban kismértékű mozgás van

jelen az átmeneti fázisban

**a) Megkülönböztethető-e a két visszaállított hangsor hangzása? Miért? (8 pont)**

Metalogika alapján: a /c miatt nem a válasz! Indoklás: mivel mindkettő hang

zöngétlen, azaz gerjesztése zörejes (fehérzaj szerű), ezért spektrumukban mindenfele frekvenciakomponens előfordul, és egész magas frekvenciákon is

vannak fontos komponensek, ezeket ez a mintavételezés (telefon) nem viszi át,

ezért az "f" és "s" nehezen megkülönböztethető, a kis mintavételezési frekvencia

miatt fellép az átlapolódás jelensége is.

**b) Mi változik, ha a rendszer bemenetére is egy H(f) karakterisztikájú szűrő**

**kerül? (5 pont)**

Megszűnik az átlapolódás jelensége, az s zörejgóca így könnyebben kivehető és

megkülönböztethető az f hang egyenletes frekvenciaeloszlásától. (ebben nem

vagyok biztos)

**c) Javasoljon egy olyan mintavételi frekvenciát és összetett simító**

**karakterisztikát, amely a fenti hangsorokat helyesen és elfogadható**

**komplexitással megvalósítva átviszi! (7 pont)**

22khz mintavételezéssel, és egy darab hasonló szűrővel 1 és 11khz között

(egyenletes meredekségű) a probléma megoldható.

**4. A vezetékes telefon frekvencia átviteli tartománya 300-3400Hz. Mely**

**beszédhangok torzulnak el leginkább az átvitel során? Miért nem zavar ez minket a**

**gyakorlatban? (10 pont)**

Leginkább a zár- és zárréshangok torzulnak az átvitel során, mivel a spektrális szerkezetükben ezek a hangok rendelkeznek nagyon magas frekvenciaösszetevőkkel

(4khz fölött is), amit a telefonvonal szűrője levág. A gyakorlatban azért nem zavaró,

mivel a hangok 4khz alatti komponensei is adnak némi támpontot, valamint az ember a

magasabb értelmezési szinteken a hang- és szövegkörnyezetből is következtetni tud arra,

hogy milyen hang lehetett ott.

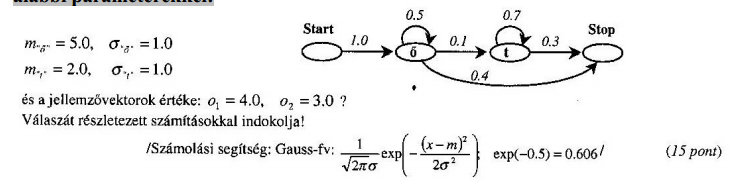
**5. Beszédfelismerési kísérleteket végeztünk lényegkiemelésként csupán a**

**keretenkénti logaritmikus energiát számolva. Az alábbi HMM hálózatra 2 db**

**egymás utáni jellemzővektor került. (Ezek nyilván nem származhattak valós**

**szóbemondásból, inkább hibás szó-detekcióból.). Mit ír ki a felismerő, ha az**

**állapotfüggő megfigyelési valószínűségsűrűség-függvények Gauss-fv.-ből állnak az**

**alábbi paraméterekkel:** 

kicsit később van a megoldás

**6. Témavezető informatikusként a következő feladattal bízzák meg: Web, WAP és**

**telefonos lekérdezési felületet kell megvalósítani a budapesti és a londoni**

**részvénytőzsde értékeihez. Milyen főbb beszédtechnológiai elemeket kellene ill.**

**lehetne alkalmazni a rendszerben? Milyen tervezési lehetőségeket kellene**

**figyelembe venni? Gondolkozzon kreatívan és széles látókörűen! A kérdésekre több**

**jó válaszegyüttes is adható! (17 pon**

WAP:

mivel az adatátvitel szűkös beszédátvitelhez (adatforgalom+beszéd?), ezért lehetne az, h

sms-ben elküldik a kért tőzsdeinformációkat, aztán az sms-t felolvassa az SMSmondó

vagy eleve egy telefonon futó kliensalkalmazás leszedi a szükséges adatokat a szerverről

WAPon keresztül és olvasná fel így integrált módon egymaga oldaná meg a problémát

mindkét esetben egy egyszerű beszédszintetizátor jöhet csak szóba a mobiltelefonok

(mondjuk azt hogy) egyelőre még szűkös kapacitása miatt, tehát formáns vagy diádalapú

megoldás az elsődleges jelölt.

mindkét esetben a kliensprogram nyelvét a felhasználó telepítésnél választhatná ki

Telefon:

Itt már a vezérlés is lehet, sőt ajánlottan beszédalapú. Gondosan megtervezett,

dialógusszerű, 2-3 hierarchiaszintes szerkezetben közölhetné igényét a felhasználó, pl:

Mire kíváncsi? Értékpapír. Melyik cég papírja? Machester United. Milyen mutatójára?

Értékére. 4.52 penny.

vagy kötetlen kérdésfeltevés után kérdezne rá a program a bizonytalan vagy hiányos

részletekre: Hogy állnak a Machester papírjai? 4.52 penny.

a válaszok vegyes (TTS+kötött) felolvasó rendszer segítségével generálódnának,

dinamikus rész lenne a számok, dátumok, esetleg cégnevek felolvasása, a többi statikus,

vivőmondatok

ügyelni kell hogy többnyelvű legyen a rendszer, hisz ez mégis a világ legnagyobb

tőzsdéje, alapból lehetne angol, és ha nem reagál a felhasználó, akkor mondaná adott

nyelveken hogy melyik gombot nyomja meg ha héberül/kínaiul/xu! vagy egyéb nyelven

szeretné hallani a frankót

a beszédfelismerő beszélőfüggetlen legyen, a beszédgenerátor triádosként lenne optimális,

esetleg diádos a hang minősége 3,7khz, ezt mind beszédgenerálásnál (minták minősége), mind

beszédfelismerésnél figyelembe kell venni

PC:

számítógéppel a billentyűzet, de főleg az egér legalább olyan gyors kommunikációt

biztosít mint a beszéd, kivéve kereséseknél

tehát adott, kevés alternatíva közül egér segítségével, sok (cégek neve, dátum keresése)

lehetőségnél beszéddel választanánk

a felismerő futhatna a felhasználó gépén, csökkentve a szerver terheltségét, ugyanakkor

nem szabad túl nagynak lennie, valamint az sem jó ha be kell tanítani minden szóra

(DTW kiesett), tehát HMM jön szóba, moderált méretű paraméteradatbázissal, esetleg

beszélőadaptív lehetne!

a beszédgenerátort mindeképp a felhasználó gépén lenne érdemes megoldani, ahol

alapértelmezetten/adott parancsra felolvasná a lekérdezés eredményét

a program nyelvét a felhasználó választaná ki telepítésnél, de a kliens természetesen

nyelvfüggetlen módon, csak adatok formájában kommunikálna a szerverrel

# 2002.05.23. vigzsa

**1. Magyarázza meg a következő fogalmakat.**

**a) Formáns. Adjon példákat**

* Röviden: A zöngejelbol az artikulációs csatorna üregrendszere által felerosített felhangnyaláb.
* A zöngés beszédhangok létrehozásához két független építőelemre van szükség: a gerjesztő jelre (zönge, alaphang, alapfrekvencia: Fo) és az artikulációs csatornára, amelyik a zönge jelét átformálja. Az átformálás során a zönge adott felharmonikusait az üregrendszer rezonanciái felerősítik. Ezeket a felerősített felhangnyalábokat formánsoknak nevezik. Például az "a" hang formánsai: F1 500-600Hz, F2 900-1100Hz, F3: 2200-2400Hz

**b) Gerjesztés és fajtái. Adjon példákat.**

A beszédhangok létrehozásának egyik alaptényezője a gerjesztés, vagyis a

hangforrás, amiből az artikuláció hatására a tényleges beszédhang kialakul. A gerjesztési

hang alapvetően háromféle lehet: zöngés, zörejes és kevert (tüdő és hangszálak)

* zöngés: az összes magánhangzó, b, d, g, gy, v, j, m, n, ny, l, r
* zörejes: p, t, ty, k, c, cs, f, sz, s, j\*, h
* kevert: dz, dzs, z, zs

**c) Egyszerű és összetett szerkezetű beszédhang. Adjon példákat.**

* Egyszerű szerkezetű a beszédhang, ha időben periodikus vagy állandó.
* Összetett szerkezetű, ha belső időszerkezettel is rendelkezik, ez írja le a hangon belüli akusztikai jelenségek időtartam-értékekeit, azok egymáshoz viszonyított időarányait.
  + Egyszerű: az összes magánhangzó, v, f, z, sz, zs, s, j, h, m, n, l
  + Összetett: b, p, d, t, g, k, gy, ty, c, cs, dz, dzs, ny, r

**d) Erősen, gyengén ls kölcsönösen illeszkedő beszédhangok. Adjon példákat.**

Artikulációjuk a beszédhangok szerint lehetnek stabilak (nem illeszkednek a környező magánhangzókhoz), erősen illeszkedők (nagyban befolyásolja akusztikai jellemzőket, formánsaikat a szomszédos beszédhang) és kölcsönösen illeszkedők (a szomszédos hanggal kölcsönösen befolyásolják egymás jellemzőit, paramétereik közelednek).

* Stabil: gy, ty, j, n, ny, r
* Kölcsönösen illeszkedő: b,p, d, t, dz, c, dzs, cs, v, f, z, sz, zs, s, h, m, l, az összes magánhangzó (?)
* Erősen illeszkedő: g, k

**e) Prozódia. Milyen komponensei és alkomponensei vannak? (3 – 3 pont)**

A prozódia a beszéddallam, a hangsúly, a ritmus, a hangero, a tempó és a hangszínezet nyelvi használata, a beszédképzés szupraszegmentális szintjének része.

Komponensei:

* Dallam
* Hangsúly (ezen belül alkomponensek: alapfrekvencia, intenzitás, időtartam)
* Ritmus
* Hangszín (?)

**2.**

**a) Milyen szempontok szerint lehet minősíteni a beszédfelismerő rendszereket? (10**

**pont)**

* statisztikai alapú (HMM, ANN) vagy szabálybázis/tudásalapú
* beszélőfüggetlen, beszélőfüggő vagy adaptív (avagy beszélők száma alapján)
* akusztikus környezet alapján: robusztus (zajos környezetben is használható), távbeszélő minőséggel vagy kiváló hangminőséggel működik csak
* szociolingvisztika: dialektusra, korra és nemre érzékeny-e
* artikuláció alapján: izolált szavas, kapcsolt szavas vagy folytonos (diktáló) rendszer
* szótárméret: kis (<100 szó), közepes vagy nagy (>20.000 szó)
* beszédstílus: spontán, parancsmódú vagy dialógus-menüszerű
* nyelvfüggés-nyelvazonosítás
* alkalmazói környezet: szakembereknek vagy laikusoknak, egyfelhasználós vagy sokfelhasználós

**b) Miért rejtett a rejtett Markov modell? (5 pont)**

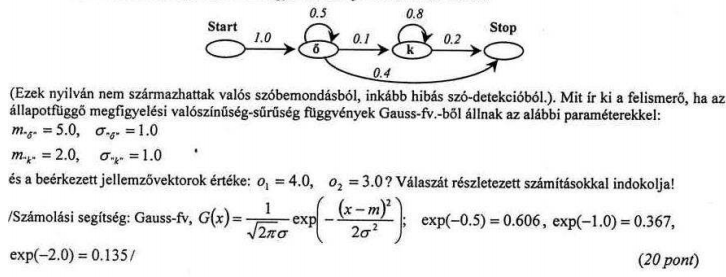
A modell azért rejtett, mert egy megfigyelés esetén nem lehet egyértelmuen meghatározni,

hogy melyik állapot generálta azt

**3. Beszédfelismerési kísérleteket végeztünk, lényegkiemelésként csupán a**

**keretenkénti logaritmikus energiát számolva, Az alábbi HMM hálózatra 2 db**

**egymás utáni jellemzővektor került.**



[itt van bevezető a 7. oldal körül](http://wiki.sch.bme.hu/images/b/ba/Beszed_VizsgaesZH_megoldasok_2010ig.pdf)

m”ő” = 5.0 o”ő” = 1.0

m”k” = 2.0 o”k” = 1.0

jellemzővektorok:

o1 = 4.0 o2 = 3.0

G(x) adott.

A megfigyelést úgy kapod, hogy a vizsgasoron megadott GAuss-függvénybe behelyettesítgetünk. M lesz az állapotok/betűk m-je, X pedig az o megfigyelésvektorok. Tehátt annak a valószínűségét akarod kiszámolni, hogy az “ő” állapotban o1-t figyeled meg, akkor az m = m = 5.0-t és x=o1 = 0.4-et helyettesíted be. exp ( -((4 - 5) ^2 ) / 2\*1 ) = exp (-0.5) = 0.606

ezt kell még szorozni 1/sqrt(2\*pi\*1), azaz ~ 0.4-gyel

Megoldás:

* átmenet STARTból “ő”-be: 1.0
* “ő” állapotban az o1 esemény megfigyelése: 0.4\*0.606 = 0.242
* “ő” áallpotban az o2 esemény megfigyelése 0.4\*0.135 = 0.054
* “ő” állapotból “k” állapotba lépni: 0.1
* “k” állaptoban o1 eseméynt megfigyelni: 0.4\* 0.135 = 0.054
* “k” állapotban o2 eseményt megfigyelni: 0.4\*0.606 = 0.242
* “k” áallpotból a STOP állapotba érni: 0.4

Két út lehetséges, ezek valószínűsége:

elsőnek az o1, majd az o2 eseményt figyeljük meg, a megfelelő élekkel felszorozva.

START-Ő-K-STOP: 1.0 \* 0.242 \* 0.1 \* 0.242 \* 0.2 = 0.00117, azaz 0.117

%

START-Ő-Ő-STOP: 1.0 \* 0.242 \* 0.5 \* 0.054 \* 0.4 = 0.00261, azaz 0.267

%

Tehát "őő"-t ír ki a felismerő.

**4. Számítsa ki a suttogott beszéd (átlagos hangnyomás 1000 mikroP) és a kiabálás**

**(átlagos hangnyomás 1000000 mikroP) közötti dinamikatartományt dB-ben. Mi a**

**jellegzetes különbség a suttogó és a normál beszéd spektrális színképe között? (15**

**pont)**

Suttogó beszéd: 1000 \* 10ˇ-6 P = 10ˇ-3 P

Kiabálás: 1000000 \* 10ˇ-6 P = 1P

Dinamikatartományuk különbsége: 20 \* log (1P / 10ˇ-3 P) = 60dB (mivel 1000x

hangosabb, ez is a jó eredmény -- 20\*3 az 60)

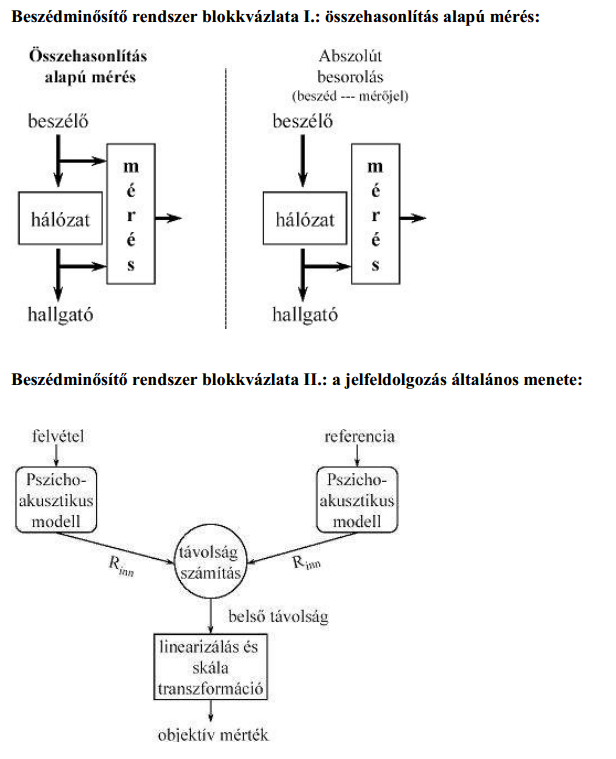
Különbség abban rejlik hogy suttogó beszédben nincs zöngés gerjesztés, így

alapfrekvencia és formánsok sem, tehát a magánhangzók vonalas színképe helyett is

folytonos színképet kapunk spektrális elemzésnél. Suttogásnál beszédtorzítás: zöngék nélkül.

**5. Ismertesse egy gépi beszédminősítő működési elvét:**

**a) adja meg a blokkvázlatot, (8 pont)**



**b) röviden ismertesse minden elem szerepét (7 pont)**

A hálózat az az átviteli technológia vagy kódolás, amelynek beszédminőségét minősíteni

szeretnénk. A mérés menete a második ábrán van kifejtve, ezen belül a pszichoakusztikus

modell az, ami az ember számára lényeges, hallható részek kiemelése és a nem

érzékelhetőek, észlelhetőek elnyomása, majd ezekből ablakozással és Fourier

transzformációval jellemző vektrokat ad ki keretenként. A távolságszámítás ezek alapján

a vektorok alapján megbecsüli a forrásjel és a vett jel eltérését, torzulását, amelyből

linearizálás és skálatranszformáció után kapjuk meg a vizsgált rendszer

beszédminőségének objektív mértékét.

**6. Tervezze meg egy ésszerű BKV menetjegy-bérlet vásárló rendszer párbeszéd**

**folyamatát, ami egyaránt működik beszédfelismerővel és nyomógombos vezérléssel**

**is. A beszédfelismerő egyszerre maximum 10 különböző szót képes felismerni, kb.**

**95%-os biztonsággal. (Elsődleges input a beszédfelismerő!) A rendszert 6-99 éves**

**korig bárki használhatja (külföldieknek is legyen esélye!). A készüléknek egysoros**

**kijelzője van, billentyűzet 10 gombos. (20 pont)**

**Vásárolható:**

**• Jegy (normál, gyűjtő, ..)**

**• Bérlet (30 napos, havi, éves …)**

**• Turista jegy**

**• …**

• első lépéskent a 10 gombbal indítja el a vásárlást az utas, minden gomb egy-egy

nyelvnek felel meg, így levesszük a nyelvazonosítás terhét a rendszerről és tudja

egyből, mikortól van action (esetleg azzal indíthatnák el, hogy bemondják a

kívánt nyelvet: Magyar, English, Deutsch, stb)

• minden, amit a gép mond, sorban kiíródik a kijelzőn is!!

• a nyelv kiválasztása után üdvözöli a rendszer az utast napszaknak megfelelően

majd megkérdezi mit szeretne venni, vagy információt kérni

• ha nincs response, akkor elkezdi mondani a gép hogy miket lehet venni:

(vonal)jegy, szakaszjegy, átszállójegy, gyűjtőjegy, turistajegy, (havi) bérlet, 30

napos bérlet, éves bérlet, stb.

• a választ a következő formában várja a rendszer: "[Kérek/Szeretnék

venni/Adjon/Aggyá] [1-10]

(Vonal)jegyet/átszállójegyet/szakaszjegyet/gyűjtőjegyet/turistajegyet/([diák/nyug

díjas](havi)/30 napos/éves bérletet) [kérek/szeretnék (venni)]", ezek nagyjából

max. 10 hosszú mondatoknak felelnek meg, lehetőleg a kulcsszavakra vadászva,

azokra forszírozva illesztünk

• ha pontatlan a kérés (diák v nyugdíjas bérlet?) akkor rákérdez a rendszer és

megvárja a releváns választ

• mindezen fix választási opciók esetén (tehát nem dátumnál vagy darabszámnál) a

10 gombbal is lehet választani az alternatívák közül

• a felhasználó minden ponton a "vissza" paranccsal egy kérdéssel visszaugorhat

(ha rosszul értette meg a rendszer), "elölről" paranccsal pedig újrakezdheti a

vásárlási procedúrát

• ha a végleges kérést megértette a gép (úgy véli) akkor visszamondja, mit szűrt le

és felszólítaná az utast h dobjon be ennyi meg ennyi pénzt vagy vissza vagy

elölről ha nem tetszik valami. Pl.: "Két darab vonaljegyhez dobjon be 300 forintot.

Ha meggondolta magát vagy félreértettem, mondja vagy nyomja hogy vissza vagy

kezdjük elölről" (A biztos hatás érdekében ez a két parancs gombbal is legyen

kiadható) A gép szövegelését bármikor meg lehet szakítani pénzbedobással vagy

gombnyomással. • a pénz bedobását hanggal is nyugtázná a gép (bár ez lehet zavaró, tehát ha 5 mpig

nem dob be új pénzt, és nem elég akkor szólalna meg hogy hello, ez így sovány)

• ha elég pénzt bedobott az utas, akkor megköszöni a gép, kiadja a jegyeket és a

visszajárót, megköszönné hogy igénybevették a szolgáltatását, stb.

# 2006.05.22. vizsga

**1. Mondjon 3 – 3 példát arra, hogy milyen tényezők okozzák az akusztikai**

**paraméterek variáltságát, egy személyen belül és a személyek között. (12 pont)**

Személyen belül:

• érzelmi állapot (nem sikerült beszédvizsga, lediplomáztam, lemerült a telóm)

• egészségügyi állapot (rekedt, megfázott, csuklik)

• szituáció (családi ebéd, szónoklat, történetmesélés)

Személyek közötti:

• nem (női hang magasabb, ffi mélyebb)

• ritmus (hadar, dadog, megfontolt)

• beszédhibák (selypít, raccsol)

esetleg közeg?

**a) A hangkapcsolódások osztályozásánál az egymással kapcsolódó hangokra milyen**

**kategóriákat lehet megkülönböztetni? Adjon példákat. (5 pont)**

ez már volt, ugye, de

Erősen illeszkedő: pl a g,k hang erősen illeszkedik a szomszédos magánhangzóhoz

Kölcsönösen illeszkedő: sok hang, pl b,p,d,t kölcsönösen hatással van a szomszédos magánhangzókkal egymás formánsmozgásaira

Stabil: gy, ty hangra kevés hatással van a környezet

**b) Melyik magyar beszédhangok tartalmazzák a legmagasabb**

**frekvenciakomponenseket és ezeknek mi a jellemző frekvencia tartománya? Miért**

**fontos tudni távközlési alkalmazásokban? (5 pont)**

• dz, c: 4000-5000 5500-7000Hz

• dzs, cs: 3700-8000Hz

• v, f : 1000-10000Hz

• z, sz: 4000-4500, 5000-8000Hz

• zs, s: 3700-8000Hz

a mintavételezés és a helyes antialiasing-szűrő megválasztásánál szükéges tudni a

zörejgócokat és frekvenciatartományokat, hogy érthető maradjon a beszéd az átvitel

során. Egyes hangoknál ha nem jut át megfelelő számú komponens, felismerhetetlen a hang. Ezen kívül a skála kialakítása miatt is fontos

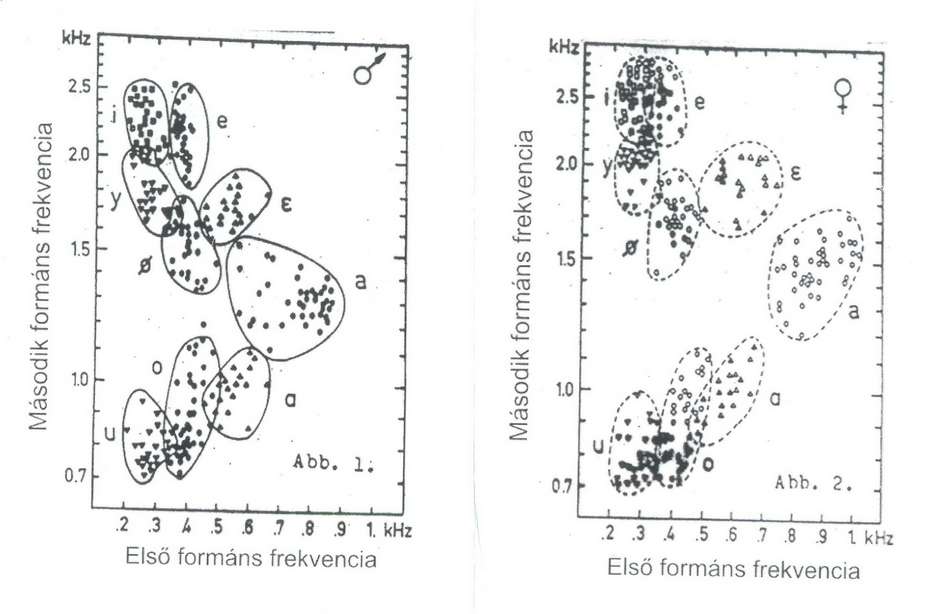
**c) Az y=F2, x=F1 koordináta rendszerben helyezze el az i, u, á hangokat. Rajzolja**

**fel a rájuk jellemző spektrumot, ha feltételezzük, hogy F0= 200 Hz. (5 pont)**

• iF1: 250-350Hz, iF2: 2300-2500Hz

• uF1: 250-350Hz, uF2: 500- 600Hz

• áF1: 700-800Hz, áF2: 1300-1400Hz



**d) Milyen lenne a spektruma ezeknek a hangoknak, ha suttogva ejti ki a beszélő. (5**

**pont)**

a formánsszerkezet megmaradna, de a suttogás miatt a teljes frekvenciatartományban

megjelennek kisebb komponensek, a spektrumképen a teljes frekvenciatartomány kicsit

"beszürkülne".

Különbség abban rejlik hogy suttogó beszédben nincs zöngés gerjesztés, így

alapfrekvencia és formánsok sem, tehát a magánhangzók vonalas színképe helyett is

folytonos színképet kapunk spektrális elemzésnél.

**a) Mi az a DTMF? Van-e szerepe a beszéd értehetőségében? Azonosítható-e és**

**hogyan a jel spektrumában (5 pont)**

Dual Tone MultiFrequency, DTMF jelek esetén nincs beszédjel, így zavarja az

érthetőséget, mert 2 szinusz hang szólal csak meg, így a jel spektrumában könnyen

felismerhető lesz a 2 kiugró amplitudó

**b) Mi az F2 ill. B2? Hogyan határozhatóak meg? (5 pont)**

F2 a beszédjel második formánsa, avagy az akusztikum második legkisebb felerősített

felhangnyalába, a B2 pedig ennek a formánsnak a sávszélessége. F2 meghatározható a jel

spektrumából, ez a második legkisebb erősítési hely (lokális maximum), a B2-t pedig

ezen a maximum alatt 3 dB-lel meghúzott vonal és a burkológörbe metszéspontja jelöli ki.

**c) Mi az ITU p.800? A beszéd mely jellemzőire vonatkozik? (5 pont)**

ITU P.800: az ETSI egyik szubjektív beszédminősítő szabványa. Minősíthetünk

* abszolút módon, előre definiált skála alapján (ACR)
* 'jelenség' észlelési tesztek
* romlás megfigyelése eredetihez képest (DCR)
* referencia rendszerrel összehasonlítás (MNRU)

**d) Mi a VXML, a SAPI és SUI kapcsolata? (5 pont)**

Mindegyik a beszédinformációs rendszerek felépítését segíti, illetve annak egy eleme.

* VXML: Voice eXtensible Markup Language dialógusok tervezését segítő leírónyelv
* SUI: Speech User Interface, avagy beszédalapú felhasználói felület
* SAPI: Speech Application Programming Interface - a Microsoft beszédalapú felhasználói felület API-ja. Ezzel még

**5. Egy mai teherautóba épített beszédfelismerővel működő navigációs rendszert**

**többen is szeretnénk használni. Milyen specifikációs feltételekkel lehetséges ez?**

* Nem kiváló a hangminőség, robosztus rendszer kell
* Nem lehet emiatt diktáló rendszer, maximum kapcsolt szavas felismerő
* Beszélőfüggetlen kell legyen
* Előzőek miatt kis-közepes szótárnagyság a reális
* A rossz körülmények miatt fel kell készíteni spontán beszéd felismerésére
* Egyértelműen statisztikai alapú felismerő jön szóba (ilyenek működnek is, rossz a hangminőség és sok a beszélő)
* Mivel a GPS-nek ez nem a fő funkciója, fontos szempont hogy olcsó legyen a megvalósítása
* Ne kelljen a túlzott társzükséglet miatt növelni a készülék fizikai méreteit (LOL)

**6. a) Tervezze meg egy telefonos, magyar nyelvű, magyarországi egyetemi felvételi**

**információs rendszer párbeszédfolyamatát és sorolja fel specifikált**

**beszédtechnológiai elemeit. A rendszer egyaránt működik beszédfelismerővel és**

**nyomógombos vezérléssel is. Az izolált szavas beszédfelismerő egyszerre maximum**

**500 különböző szót képes felismerni, kb. 90 %-os biztonsággal. A rendszert legalább**

**10 éves időtartamra lényegi módosítás nélküli megoldással tervezze meg.**

**Szükséges információk:**

**• Milyen szak(párok)ra kíváncsi**

**Kimenet:**

**• Az adott szó(párok) indító egyeteme(i) és kara(i) neve, címe, felvételi**

**feltételek, pontszámítás és korábbi évek adatai**

**Gondolkozzon kreatívan és széles látókörűen! A kérdésekre több jó válaszegyüttes is**

**adható! (13 pont)**

Egy gombbal lehet indítani a rendszert, ezzel együtt esetleg nyelvet is ki lehet

választani, így nem kell nyelvadaptációt és beszéddetekciót végeznünk

• A felhasználót megfelelően üdvözöli a rendszer, majd megkérdezi hogy milyen

szakra, szakpárra kíváncsi • A következő bemenetet várja: [A] {szak}/{szakpár} [ra/re vagyok kíváncsi]/[után

érdeklődök] (a kérdéssel jól behatároltuk az adható válasz formáját!)

• Amennyiben a rendszer nem biztos a szakban, felsorolná a 10 legvalószínűbb

szakot, amit mondhatott a felhasználó, és felszólítaná h válasszon közülül vagy

mondja be újra

• A felismerő HMM alapú, robosztus, közepes nagyságú, beszélőfüggetlen.

• A válasz következő formátumban generálódna: A {szak}/{szakpár}t a következő

egyetemek indítják: ( [egyetem], [kar] )\*

• Ha a felhasználót nem érdekli az adott egyetem, "tovább" vagy "vissza" szavakkal

léptethet (gyorsabban) közöttük (gombbal is)

• Ha felkelti érdeklődését valamelyik, a "címe", "felvételi (feltételek)",

"pontszámítás", "korábbi évek" paranccsal kérheti le az őt érdeklő adatokat a

karról (gombbal is választhat)

• Cím esetén: "A(z) {egyetem} {kar} címe [város] {közterület neve} [közterület]

{házszám}, {irányítószám}

• Felvételi feltételek: [adott kar feltételei], a paraméterszerű adatokat dinamikusan

generálja

• Pontszámítás: [pontszámítás menete], paraméterszerű számokat, adatokat

dinamikusan

• Korábbi évek: [évben] [a ponthatár] {szám} [pont volt, a felvettek száma] {szám}

[fő, a jelentkezők száma:] {szám} [fő] stb.

• Vegyes felolvasó rendszert használunk: TTS+kötött

• a [] elemek kötöttszótáras módon, előre felvéve vannak letárolva, a {} részek

generálása pedig triádos szövegfelolvasó rendszer feladat\*a

• A felhasználó a "lista" paranccsal tér vissza a megfelelő egyetemek listáájához

(gombbal is)

• "Köszönöm" esetén vagy 1 perc tétlenség után a rendszer alaphelyzetbe áll

**b) Milyen főbb szempontokat kell figyelembe venni, ha a feladat a spanyol vagy a**

**szlovén felvételire vonatkozna az adott ország nyelvén? (4 pont)**

Szórendre, dátumra, számok felolvasására kell figyelni

• Teljesen más lehet a felsőoktatás menete, pontszámítási módszerek, ezeket is

megfelelően át kell alakítani

• A felismerés során más sorrendben adja meg az adatokat a felhasználó

• Más temperamentumú beszéd, más beszédstílus: újra kell paraméterezni a

felismerőt, nem csak a felismerendő szavakat kell kicserélni

• A katalán és a spanyol nyelv eltér, érdemes a nyelvek közé mindkettőt felvenni

(spanyol rendszer esetén)

számok felolvasásának különbözősége

Amik kellenek szerintem mindig ilyen feladathoz:

1. Kezdésnél nylevl kiválasztása, gombok feladata stb.
2. kérdéseknél jól körbeírni, hogy mit várunk “A <> szakra vagyok kíváncsi”
3. felismerésnél több lehetőség is jelenjen meg, a pontatlanság miatt
4. legyen megnevezve a modell, méret, stb.
5. Több menüpont, lehetséges zótár
6. mi történjen kilépéskro, idle mód stb.
7. számok?
8. stb.

# 2006.06.02. vizsga

1. feladat

**Milyen jellel mérjük a beszédátviteli rendszerek minőségét?**

Természetes emberi beszéddel, de érdektelen felvételeket kell felolvastatni az alanyokkal! (nem vagyok teljesen biztos h ezt kérdezik..)

**Az objektív minősítő rendszer hatékonyságát mihez képest mérjük?**

Az objektív minősítés célja a szubjektív minősítés közelítése, tehát azt nézzük, hogy mennyire egyezik az eredménye az egyéni véleményekkel.

**Ha a gépi minősítés a szubjektív minősítéshez képest egyes méréseknél lényegesen jobb, más méréseknél lényegesen rosszabb eredményt ad, akkor a minősítő mely komponensét kell módosítani?**

A pszichoakusztikus modellt, esetleg a belső távolság számításának a módszerét (amivel a referenciafelvételtől való eltérést mérjük, számítjuk)

**A csomagkapcsolt beszédátviteli rendszerek (pl. VoIP) mely tulajdonsága okozza a legnagyobb nehézséget a beszédminőség mérése során?**

(A hálózat paramétereinek nem stabil volta. Teljesen más minőséget kapunk ha kis illetve szélessávon mérünk, illetve változatos kapcsolat (műholdas, kábel, adsl) esetén is jelentős eltéréseket tapasztalhatunk a beszéd minőségében, a hálózatforgalmi szituációkat nem is említve (pl. ha közben töltünk is).) Nem a rizsára voltak kíáncsiak. A válasz: Jitter (késleltetés-ingadozás)

**Mikor és ki készítette az első beszédkeltő gépet a világon? Hol látható?**

Kempelen Farkas

1791-ben.

Az egyetlen megmaradt példány ma a müncheni Deutsches Museumban van.

**Mikor és ki adta be a világ első szabadalmát tetszőleges szöveg felolvasására alkalmas beszélőgépre?**

Bánó Miklós

1916-ban.

**Mi az artikulációs sebesség? Milyen érték jellemző a magyarra? Mi a beszédsebesség?**

Az artikulációs sebesség az időegység alatt ejtett hasznos beszédhangok száma folyamatos ejtésnél, szünetek nélkül.

A magyar beszédnél tipikus értéke 13 hang/s.

A beszédsebesség a beszéd hangzásának teljes idejében, szünetekkel, időegység alatt elhangzott beszédhangok száma, a nem hasznos beszédjeleket is beleértve. (Magyar beszédnél 14 hang/s)

artikulációs sebesség <= beszédsebesség

**Mi a VOT? A beszédjel mely részén mérhető? Adjon 5 konkrét példát indoklással!**

\*VOT\*: Voice Onset Time avagy zöngekezdési idő

felpattanó zárhangok esetén a zár felpattanása és az azt követő magánhangzó megszólalása között eltelt idő

Tipikusan a beszéd azon helyen mérhető, ahol gerjesztésváltás történik, és zöngétlen hangot zöngés hang követ.

A fentiek fényében a VOT pl. p után 8ms, t után 15ms, k után 26ms. (Ide lényegesen többet nem tudok írni, főleg az indoklás részét nem értem)

**Mi a spektrális átlapolódás oka mintavételezéskor? Hogyan előzhető meg? Adjon példát.**

\*Spektrális átlapolódás\*: ha a hang mintavételezésénél a mintavételezési frekvencia kisebb, mint a legnagyobb frekvenciakomponens kétszerese, a visszaállításnál nemkívánatos jelek kerülnek visszaállításra, a jel nem állíthatő elő egyértelműen/hűségesen.

Megelőzhető megfelelő karakterisztikájú aluláteresztő szűrővel a bemeneten. (Sávkorlátozás)

Példát mindenki remélem tud adni ezek alapján :]

**Mi a néma fázis? Sorolja fel az összes beszédelemet, amelyre vonatkozhat!**

\*Néma fázis\*: A zárhangok azon része, amelyben nincs hangképzés. A tüdőből kiáramló levegő a toldalékcsőben képzett akadály miatt feltorlódik és a zárfelpattanásig levegőáram nem hagyja el az artikulációs csatornát.

A fentiek alapján néma fázis található a zöngétlen zár- és zárréshangoknál így: p, t, k, ty, c, cs.

**Igaz-hamis**

**a.) A lényegkiemelő feladata, hogy digitalizált beszédjelből előállítson egy disszkrét idejű vektoriális fonémasorozatot -** HAMIS -Nem fonémasorozatot kell előállítania, hanem egy olyan 10-40 dimenziós vektort, melyeknek kicsi az intraindividuális és az interindividuális jellemzője.

**b.)A lényegkiemelő olyan akusztikus információt emel ki a bemenő beszédjelből, amely alapján következtethetünk arra, hogy egy adott kimenő vektor melyik beszédhanghoz tartozik. -** IGAZ

**c.) A lényegkiemelő eljárásoknál a beszéd kepsztrális elemzése elsősorban a prozódiai jegyek kiemelését célozza** - HAMIS . a prozódia teljesen más.

**d.) A lényegkiemelő a beszédfelismerőkbe ágyazott beszédértő azon része, amely kiemeli a közlés tárgyát.** - HAMIS, kiszűrni a n00bokat van ez a kérdés

**a.) A mintaillesztsés feladata, hogy a bemenő bexzédhangsorozatot a felismerési hálózathoz illesztve megpróbálja a kimenetén előállítani a felismert szósorozatot -** IGAZ

**b.) Létezik olyan mintaillesztési módszer, amelyet ML Maximum likehood lértelemben mindig optimális illesztést valósít meg a bement és felimerési hálózat között. -** IGAZ

**c.) A mintaillesztés csak osztályozást jelent )(vagyis az egyes felsimerési lehetőségekhez hasonlósági mértékek rendelését) az időillesztés egy másik lépésben történik meg. -** HAMIS A mintaillesztés ergyik lényege a különböző ritmusú ejtések között is tudjon mintailleszteni.

**d.) A dinamikus idővetemítés (DTW) nem mintavételezés.** - HAMIS, De, az.

**a.) A rejtett Markov-modellek abban hasoníltanak a Markov-láncokhoz, hogy állpotomk és állapot-átmenet valószínsűgégek is értelmezettek mindkét esetben. -** IGAZ

**b.) A rejtett Markov-modellek olya nmódon jellemzik a beszédhangokat, hogy kizárólag egy adott áallapot megfigyelési sűrűségfüggvénye alapján el tudjuk dönteni, hogy egy bemenő vektor az adott állapot által modellezett beszédhangokhoz tzartozik-e, vagy sem. -** HAMIS: valószínűségekkel dolgozik a HMM, így teljes biztonsággal sosem tudja megmondani, hogy egy emgfiygyelés adott állapothoz tartozik vagy éppen nem tartozik.

**c.) A mintaillesztés rejtett Markov-modellek esetén nem más, mint a felismerési hálótaz kezdő és végpontja közti legkisebb valószínsűségű útvonal megtalálása. -** HAMIS. nyilván a legnagyobb valószínűséget keresi, nem a legkisebbet.

**d.) Az órán bemutatott (Viterbi) algoritmusnál a mintaillesztés számításigénye megközelítőleg exponenciálaisan függ a felismerési hálózat állapotainak számától.-** HAMIS, lineárisan,pont ez a lényege.

**4.**

**a) Mi az LPC? Van-e szerepe a beszédértésben? Kapcsolatba hozható-e és hogyan a**

**jel spektrumával?**

Linear Prediction Coding / Coefficients. Lineáris elõrejelzés. Olyan matematikai

eljárás, amellyel a megelõzõ mintákból jósolni lehet a következõ mintát. LPC

segítségével az akusztikus jelbõl meghatározható például az artikulációs

üregrendszer átviteli karakterisztikája is.

• ??? (Ha jól meghatározhatók az LPC együtthatók, jobban érthetők a hangok?) A

formánsokat jól lehet vele követni.

• Igen, a LPC analízis is egyfajta spektrumát adja meg a jelnek. (ide még lehetne

írni)

**b) Mi az F0 ill. F1? Hogyan határozhatók meg?**

**•** F0 az alapfrekvencia, azaz a hangforrás gerjesztésének frekvenciája. F1 pedig a

legkisebb (első) formáns azaz felerősített felhangnyaláb.

• F0 meghatározható a zöngés hangok periódusidejéből (megegyezik azokkal). F1

pedig a jel spektrumára illesztett burkológörbe első (lokális) maximumhelye.

**c) Mi a Hamming-ablak és mi a szerepe a beszédfeldolgozásban?**

• A Hamming-ablakot a jelre illesztve egy véges időtartományban kell csak

elvégezni a Fourier-integrálást. A szerepe az, hogy adott időpillanatban releváns

frekvenciákat felerősítse, a távoliakat gyengítse hogy adott időpillanatra jó

spektrumot kapjunk a Fourier-integrálás után. 0.540-46\*cos (2\*pi\*(t/T0) )

**d) Mi a screen reader és a TTS kapcsolata?**

• A screen reader csak egy illesztő alkalmazás a képernyő és a TTS(TextToSpeech) között, a

képernyőn található információt adja át felolvasásra a TTS számára.

**5. Adjon meg min. 5 specifikációs szempontot egy távközlési szolgáltató számára**

**tervezett e-levél felolvasó rendszerhez! Adjon meg min. 5 felhasználási lehetőséget is!**

5 specifikációs szempont:

• Nyelv

• Operációs rendszer

• Beszéd minősége : érthetőség, természetesség

• Milyen hangokon szólaljon meg (ffi/női)

• Mennyire legyen paraméterezhető: hangmagasság, sebesség, szünetek hossza, stb.

• Vezérlési felület, API

• Bővítési, továbbfejleszthetőségi lehetőségek

5 felhasználási lehetőség:

• Emailek felolvasása telefonon keresztül

• Vakok és gyengénlátók számára

• Rendszerüzenetek, ajánlatok természetesebb közlése

• Előfizetési információk közlése emailen keresztül

• Gyerekek számára

• Call Center IVR (telefonos menürendszer) elemeinek dinamikus létrehozása, esetleg nagy kiterjedésű hiba esetén az 'üdvözlőszöveg' amiben bemondják hogy tudnak a hibáról és javítás alatt van, felolvasó nélkül beállítható

**6. Sorolja fel a gépi beszédfelismerők jellegzetes fajtáit működési elv szerinti,**

**használati módja szerinti, és méret szerinti osztályozásban.**

Működési elv:

• Szabálybázisú

• Statisztikai alapú: HMM, ANN

• Sablon alapú: DTW (Dynamic Time Warping)

Használat módja:

• Spontán beszéd (folyamatos beszéd, pl diktáló rendszerek)

• Parancsmódú vezérlés (izolált szavas)

• Dialógusvezérlés (kapcsolt szavas, a szavak közötti szünetek minimálisak)

Méret:

• Kicsi: párszáz szó

• Közepes

• Nagy: 20-80 ezer szó

másikcsoport

**3. Adjon meg min 5 specifikációs szempontot egy távközlési szolgáltató számára**

**tervezett SMS felolvasó rendszerhez! Adjon meg min. 5 felhasználási lehetőséget is!**

A szempontok kb ugyanazok, a felhasználási lehetőségek:

• Előfizetési információk természetesebb közlése

• Vakok és gyengénlátók segítése

• Idős felhasználók segítése, akik nem tudnak/akarnak kis képernyőn olvasni

• Autóval való közlekedés során is elolvashatjuk SMS-einket

• Email-eket SMSben továbbítva, azokat elolvashatjuk

• Minden olyan helyzetben előnyt jelenthet, amikor a nyomógombok használata

vagy a kijelzőn megjelenő szöveg olvasása nem megoldható.

**4.**

**a) Mi a SAMPA? Van-e szerepe a beszédértésben? Kapcsolatba hozható-e a jel**

**spektrumával?**

• SAMPA: Speech Assessment Methods Phonetic Alphabet. Beszédhangok jelölése

7 bites ASCII karakterekkel.

• A SAMPA-val a beszédhangok egyértelműen leírhatók, segíthet a beszédértésben.

• Szerintem nem hozható kapcsolatba a jel spektrumával. Vagy csak nagyon

összetett, indirekt módon.

**c) Mi a négyszögletes ablak és mi a szerepe a beszédfeldolgozásban?**

A Fourier-integrálás során egy kis időkeret analízise úgy történhet meg, hogy az időben

folyamatos jelet egymással átlapolódó négyszögletes ablakokkal kiablakozzuk. Így kis

időszakaszokra megkaphatjuk a jel spektrumát, ami a magasabbrendű beszédfeldolgozás

fontos alapeleme.

**d) Mi a triád? Előnyei? Hátrányai? Mennyi egy nyelv lefedéséhez szükséges**

**elemszám?**

• Triád: Olyan hangkapcsolat, amelyben a középső hang egészben, a két szélső pedig részben van jelen. Beszédszintézisnél használják, elsősorban a magánhangzók szerepelnek középső helyzetben.

Előnyei:

• A magánhangzóknál nem lép fel torzítás a formánsok megtörése miatt.

• Természetesebb hangzás

• Könnyebb szövegtervezés

Hátrányai:

• Sok munkát jelent a felvétel

• Sok memóriát foglal

• Sok szöveget kell felolvastatni

• Diádokat és egyéb elemeket is igényel az adatbázis

Szükséges elemszám: , ennél némileg kevesebb mivel nem fordul elő minden hármas + a

szükséges diádok: (szerintem a tisztán triádos adatbázis egyszerűen a fonémák köbével

arányos. Az már a kevert adatbázis ahol diádok is vannak. Vagy? )

# 2007. 05. 25 vizsga

**1. 11,025 kHz mintavételi frekvenciával, 16 bites lineárisan kvantált digitalizált**

**beszéd felvételeink vannak. Spektrális elemzésre 256 pontos FFT-t számolunk (egy**

**spektrum kiszámításához ennyi mintát használunk fel).**

**a) Mekkora lesz a spektrális elemzés legjobb idő-felbontása és jel/zaj viszony értéke?**

Idő-felbontás: 256 pontos, és 11,025 KHz --> 90,7 mikroSec , innen az időfelbontás: 256

\* 90,7 mikroSec = 23,2 msec.

SNR=1,74+n\*6,02=1,74+16\*6,02=98,06 dB

**b) Mely beszédhang-csoportok spektrális vizsgálatát tudjuk és melyikét nem tudjuk**

**ezekkel a felvételekkel lényegileg pontosan elvégezni?**

Azokat nem tudjuk, melyeknek lényeges frekvenciakomponenseik vannak 5,5 kHz fölött,

így például a zár és zárréshangok jó részét nem tudjuk így spektrálisan vizsgálni. Azért

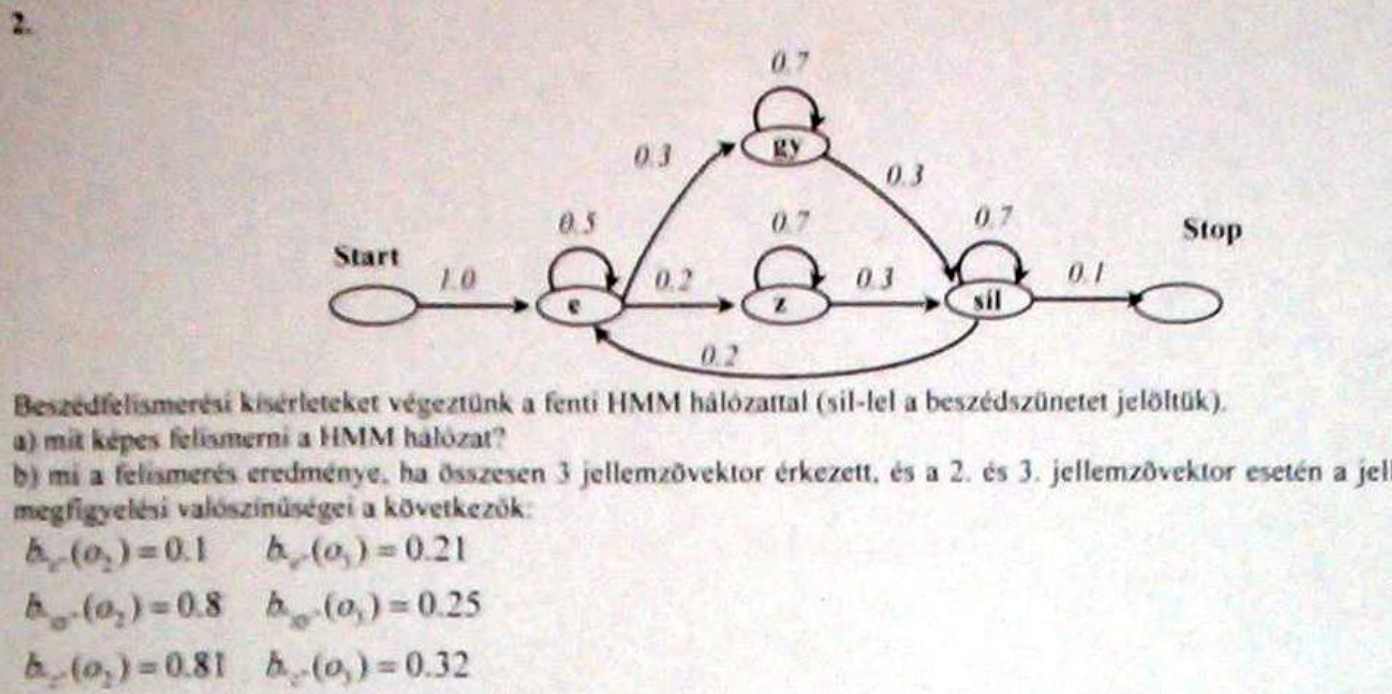
nem, mivel a mintavételezési frekvencia túl kicsi. Mint tudjuk, a mintavételezési

frekvenciának 2x nagyobbnak kell lennie a legnagyobb frekvenciaösszetevőnél, így

11kHz esetén az 11/2=5,5kHz a legmagasabb frekvencia, amiket még jól tudunk

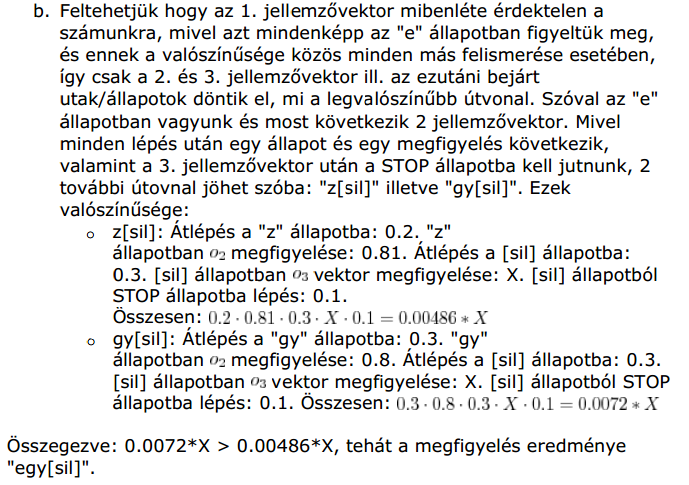
mintavételezni, az ennél magasabbak átlapolódnak.

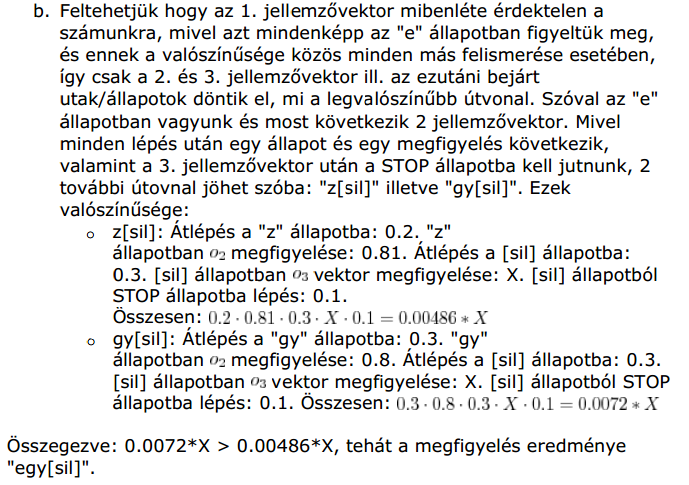
**2.**



a.) ([egy|ez]sil)+ a sil: szünetet jelez.

**b.)**





**3. 800 Mbyte kapacitású CD lemezen (44,1 kHz mintavételi frekvencia, sztereó**

**felvétel, 16 bites lineáris kvantálás) állnak rendelkezésre egyenként átlagosan 3 perc**

**hosszú zeneszámok. Szeretnénk belőlük csengőhangot készíteni egy olyan**

**mobiltelefonra, ami 11,025 kHz-s mintavételi frekvenciával tud mono, 8 bites, A-törvényű logaritmikus kvantálású mintákat lejátszani és 16Mbyte szabad**

**memóriája van**

a) Ábrákkal illusztrálja az átalakítás folyamatát! (8 pont)

Ábrák helyett az egyes lépések (kis dobozkákat rajzolnék egymás után, bennük az egyes

lépések neveit írnám):

• Visszaállítom a kvantált, mintavételezett jeleket (sztereó!) analóggá.

• Átlagolom a két jelet időtartományban, amplitudó szerint 1 mono jellé.

• Aluláteresztő szűrő, mely 5 kHz-ig engedi át a jelet, persze 5 kHz körül lineáris

gyengítéssel.

• Mintavételezés 11,025 kHzen.

• Kvantálás 8 biten.

**b) Hány zeneszám van a lemezen? Valamennyi zeneszám átalakítható-e? Ha nem,**

**mi lehet a megoldás? (6 pont)**

1 sec hanganyag tárigénye: 44,1kHz mintavételezés, 16 bit, sztereó hangsávok:

44100\*16\*2= 1,411,200 bit = 172kbyte. 800Mbyte/172kbyte= 4763, azaz 4763 sec

hanganyag tárolható, ami kb 79 perc. Ez 3 perces zeneszámokkal számolva 26 zeneszám.

Nem alakíthatók át azok a számok, mely 5kHznél magasabb frekvenciakomponenseket

tartalmaznak. Megoldás erre a fentebb már említett aluláteresztő szűrő.

**c) Vissza lehet-e állítani az eredeti felvételt a telefonos formából? Ha igen, hogyan?**

**Ha nem, miért nem? (6 pont)**

Nyilván nem lehet visszaállítani a telefonos formából, ennek több oka is van. Egyrészt a

monó hang átlagolással készült a sztereó hangsávokból, ezt lehetetlen visszaszűrni. (2 és

6 átlaga 4. 4 melyik két szám átlaga?). Másrészt az alacsony mintavételezés miatt

elvesztjük az 5kHz feletti komponenseket, ezeket sem tudjuk visszanyerni. Harmadrészt

pedig a 8 bites logaritmikus kódolás nem arányos a lineáris 16 bitessel, ezért főleg a

magasabb tartományokban nagyobb lesz a kvantálásból eredő zaj nagysága

**4. a) Mi a teljesítmény sűrűség spektrum, az akusztikai dB és a Phon érték kapcsolata?**

* Az akusztikai dB-ből visszakövetkeztethetünk a hangjel amplitudójára (10-es hatványraemelés), az így kapott időjel négyzete a teljesítmény sűrűség spektrum. (ha jól mondom :] )
* A Phon görbe pedig az azonos hangosságérzetű görbék serege, ahol a frekvencia-frekvanceia az 1kHz. Azaz 1kHz-es hangok esetén a phon érték megegyezik az akusztikai dB-vel.

**b) Mi a Hanning-ablak és a szonogram kapcsolata?**

• Ha gördülő spektrumot avagy szonogramot szeretnénk készíteni, akkor az időben

folytonos jelünket bizonyos kis szeletekben mintavételeznünk kell. A kis kivágott

időintervallumokból akkor kapunk jó spektrumot, ha azt megfelelően

kiablakozzuk és nem csak simán kivágjuk egy négyzetes ablakkal. Egy ilyen jól

bevált ablakozó függvény a Hanning ablak, melynek képlete:

0.5 – 0.5 \* cos (2π \* t/T)

**c) Mi a VXML, a SUI és a DTMF kapcsolata a beszédinformációs rendszerekkel?**

• Mindegyik a beszédinformációs rendszerek felépítését segíti, illetve annak egy

eleme.

• A VXML avagy Voice eXtensible Markup Language interaktív dialógusok

leírását és tervezését könnyíti meg ember és számítógép között.

• A SUI avagy Speech User Interface az ember-gép kapcsolatot beszéd és hangok

által teremti meg.

• A DTMF avagy Dual Tone Multi Frequency egy jeltovábbítási megoldás avagy

mechanizmus a normál telefonvonalon keresztül, ahol 2 frekvencia együttes

megszólaltatásával összesen 16 különböző jelet generálhatunk (4\*4=16).

**d) Mi a locus, az F2 és F0 kapcsolata?**

• A CV átmenet jellegzetessége a locus: megfigyelték, hogy pl. a d után ejtett

magánhangzók felfutó szakaszait, ha visszafelé meghosszabbítjuk, ezek egy

pontban metszik egymást – a legtöbb mássalhangzó az őt követő magánhangzó

vagy őt megelőző magánhangzó második formánsát (F2) a szóban forgó

mássalhangzót jellemző frekvenciára kényszeríti, ezek a locusok.

• Az F2 pedig nem más, mint a hangszalagoknál képzett gerjesztő jel

alapfrekvenciájából (F0) a vokális traktusban felerősített, második legkisebb

felhang-nyaláb (Fn).

**5. Egy kötött szótáras telefonos információs rendszert kell terveznie egy áruház**

**üzleti nyitva tartásának automatikus bemondására hetes időszakra. Csütörtökön az**

**üzlet 20 óráig van nyitva, egyébként 18 óráig. Szombaton 11 óráig. Specifikálja a**

**beszédtechnológiai alrendszereket és tervezze meg az információs rendszer**

**dialógusát. Állítsa össze a felolvasó alrendszerben az építőelemek tárát úgy, hogy a**

**koartikulációs hatásokat is figyelembe veszi a hullámforma összefűzésnél. Sorolja fel,**

**hogy milyen elemeket fog tartalmazni az elemtár. Rajzolja fel az információs**

**rendszer blokkvázlatát.**

Először meg kell tervezni, hogy mit kell pontosan felolvasni a rendszernek. A leírás

annyira kötött hogy a legegyszerűbb lenne egy egyszeri felvétel, mely szépen egy

hanganyagban tartalmazná az összes információt. Ez nyilván elég merev lenne, másrészt nem tennénk eleget abbéli kívánalmakban, miszerint kötött szótáras, telefonos rendszert

kell készítenünk. Ekkor érdemes úgy megtervezni a rendszert, hogy információt fogadni

is tudjon avagy egy beszédfelismerő modul is szükségeltetik mindehhez. Az

információkérés avagy dialógus nagyjából így tervezhető meg:

• Üdvözlő szöveg, a végén kérdéssel, hogy melyik nap nyitvatartására kíváncsi a

telefonáló. Ez egy fix szöveg.

• Ügyfél válasza, melyben a hét napjait (hétfő..), relatív utalásokat (ma, holnap)

illetve konkrét dátumot (május 29) keresünk.

• A válasz értelmezése után esetleg visszakérdezés, ha nem értettünk semmit,

esetleg DTMF-es megoldáshoz való folyamodás

• Válasz generálása egy mondatba ágyazva, a következő opciókkal: Az üzlet

(ma/holnap ... hétfőn/kedden ... január 29-én) (szám) órától (szám) óráig tart

nyitva.

A beszédfelismerő lehetne egy HMM-s rendszer pár szóra (kis szótár) minél

robusztusabban (zajra érzéketlen, beszélőfüggetlen) betanítva. A következő szavakat

kéne felismernia: hétfő-vasárnap, ma-holnap-holnapután-tegnap-tegnapelőtt, hónapok, 1-

én ... 31-én. Ezt most nem is részletezem mert sztem nem erre kíváncsiak.

Beszédszintetizátor tervezése: A fix vivőmondat adott, a változtatandó részek: időpontok

(ma/holnap, hétfőn-vasárnap, január-december, 1-én-31én) illetve számok (0-24-ig). Az

időpontokat elég egyszer felvenni hiszen a mondatban csak egy helyen szerepelnek,

viszont a hónap-nap kapcsolatokban előfordulhatnak bizonyos kivételek, amelyekre

figyelni kell, bár most nem találtam ilyet (vki?). A számokat viszont kétszer kéne

felvenni, mivel két pozícióban is szerepelnek (hangsúly, prozódia!), viszont nincs belőlük

olyan sok (25 szám) ezért nem kell vacakolni a még kisebb egységekre bontással.

Innentől meg a szokásos szövegek elkészítése - bemondó kiválasztása - felvétel - tárolás -

csiszolás - rendszerintegrálás blabla, meg valami ábra a fenti elemeket összefűző ábrával.

Ne felejtsük itt el az értelmezőt és a szabályok alapján való elemkiválasztást!

**6.**

**a) Mi a lényeges különbség a felhasználás szempontjából a beszélő-függő és a beszélő**

**független beszédfelismerők között?**

A beszélőfüggetlen rendszereket bárki, bármikor használhatja előzetes betanítás nélkül,

viszont általában kisebb szótárral és megbízhatósággal rendelkeznek. A beszélőfüggő

rendszerek általában beszélőadaptívak is egyben, azaz használatukhoz szükséges egy

előzetes betanítási fázis, ezután azonban több szót és jobb megbízhatósággal képesek

felismerni, izolált szavak helyett akár kapcsoltszavas vagy akár diktáló üzemmódban is.

**b) Betanításnál milyen típusú adatbázis kell az egyik és a másik rendszerhez?**

Beszélőfüggetlen rendszer esetén több beszélőtől szükséges hanganyag, hogy ebből

közös jellemző vonásokat tudjunk kivonni a betanítás során a minél robusztusabb

működéshez. Beszélőfüggő rendszer esetében pedig a hangok paraméterbecslésére nincs

szükség (vagy jóval kisebb adatbázis is elegendő), hiszen a betanítási fázis során pont

ezeket a paramétereket hangoljuk az adott beszélő alapján. Minden más vonatkozásban (szótár felépítése, nyelvi modellek stb) a két megoldás nem különbözik, illetve max. a

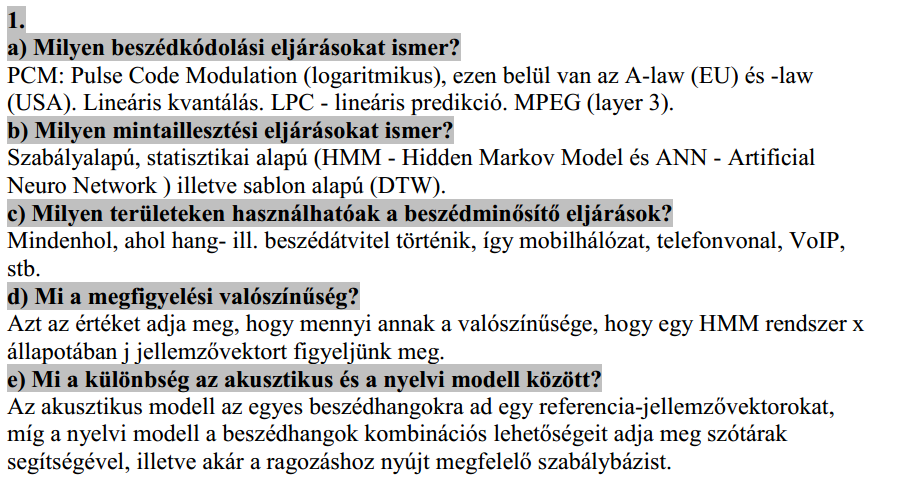
szavak számában.

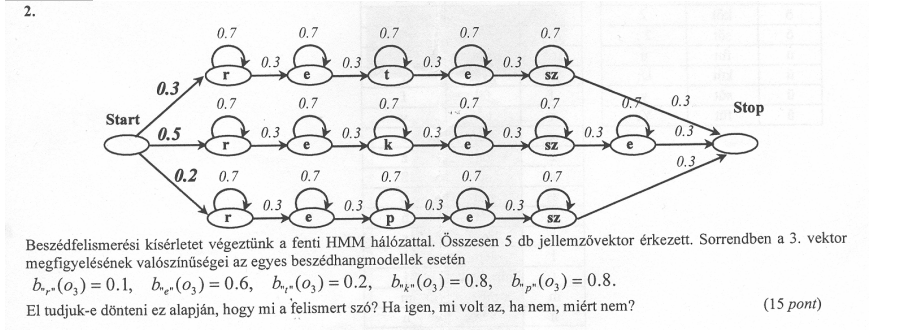
**c) Milyen egyéb szempontokat kell figyelembe venni?**

Szótárméret, tematika, a hangkörnyezet (zajos utca v csendes iroda), beszédmodor

(spontán vagy dialógusszerű), stbstb.

# 2007.06.15 vizsga





El tudjuk dönteni. Mivel HMM-ről van szó, és a mintaillesztéshez feltétel hogy a START

állapotból a STOP állapotba jussunk el úgy, hogy eközben lépések (állapotváltások) és

megfigyelések váltogassák egymást, könnyen látható hogy a középső szó (rekesze) kiesik,

hiszen 6 állapotot tartalmaz, míg nekünk 5 megfigyelési vektorunk van, így ezen az úton

nem juthatunk el a STOPig. Másrészt megfigyelhető hogy a rekesz ill. repesz szónál is

minden állapotváltás valószínűsége rendre megegyezik, sőt egyetlen állapotban, a

középsőben különböznek (k vs p). Ebből triviálisan adódik hogy az egyetlen különbséget

a két út valószínűsége között az adja, hogy mekkora a kérdéses középső állapotban a 3. jellemzővektor megfigyelése, minden más valószínűségi szorzótényezőben

(állapotváltások és megfigyelések: mindig rendre ugyanazt kell megfigyelni ugyanabban

az állapotban) megegyeznek.

Mivel p állapotban O3 megfigyelése 0.8, és t állapotban csak 0.2, a "repesz" szó lesz a

felismert szó.

És mivel az elején sem egyformák a valószínűségek, azt is bele kéne venni... 0.3\*0.2 vs.

0.2\*0.8 de így is repesz. -- Csádám - 2010.12.14.

**3.**

**a) Mit jelent egy beszédadatbázis szöveganyagának annotálása, és mit jelent a**

**szegmentálása?**

Annotálás: címkézés, azaz a megfelelően szegmentált időintervallumokat ellátjuk a

megfelelő magyarázatokkal: milyen hangról van szó, hangsúlyos-e, zöngés-e, stb. A

szegmentálás pedig a hanganyag időfüggvényén a hanghatárok bejelölését jelenti.

**b) Készítse el az alábbi mondat SAMPA fonotipikus átiratát: „Elmondtam Havadtői**

**Csillának. Odahívta közben azt a csöppséget, aki megfogta a kilyukadt zacskót”**

**(segédlet a hátlapon)**

(segédlet a hátlapon)

Nincs táblázatom, ezért a lényeg: ennél a feladatnál a különböző hangváltásokra kell

odafigyelni (hasonulások, összeolvadások, rövidülések és kivetések). Ennek a szövegnek

esetében konkrétan:

• elmondtam --> elmontam, a d hang kiesik!

• havadtői --> havattői, részleges hasonulás, zöngétlenesedés.

• közben --> köszben, részleges hasonulás, zöngétlenesedés. (kétségeim vannak,

"hasonulás" helyett éppen hogy különbözővé válna)

• azt --> aszt, részleges hasonulás, zöngétlenesedés.

• csöppség --> csöpség, rövidülés

• megfogta --> megfokta, részleges hasonulás, zöngétlenesedés.

• kilyukadt --> kilyukatt, részleges hasonulás, zöngétlenesedés.

• megfogta->mekfokta

• odahívta --> odahífta: részleges has., zöngétlenedés

(Írásban nem jelölt) teljes hasonulásra példa: anyja --> annya, hagyja --> haggya másik

irányban működő: község --> kösség, tizennyolc

**4. A hangszalagrezgést elektroglottográf segítségével (10KHz-es, 16 bites lineáris**

**mintavételezéssel) rögzítjük, majd visszajátsszuk. A beszélő a következő szöveget**

**mondta: "Eljössz velem? -em megyek. -em? Bárcsak eljönnél, úgy szeretném!"**

**a)Milyen beszédjellemzőket lehet meghallani egy ilyen hangszalagrezgésről készített**

**hangfelvételből**

A következő beszédjellemzőket lehet meghallani: a beszélő neme (F0 frekvenciájából). A

mondatok típusa nagyjából (prozódiából, azaz alaphang-változásokból kifolyólag).

Ugyanebből kitalálhatók a hangsúlyok helyei is. Beszéddallam. Emellett a zöngés / zöngétlen hangok határait is nagyjából el lehet találni. Gond a CC és VV kapcsolatoknál

van.

**b) Hallható-e a beszéd szegmentális elemei közül valamelyik? Ha igen, akkor**

**melyik(ek). Ha nem, akkor miért nem?**

Szegmentális szint: a hangok specifikus időtartamai nagyjából kiolvashatók (?), de nem

konkrét hang(kapcsolatok)ra, hanem csak általánosan

**c) Hallható-e a beszéd szupraszegmentális elemei közül valamelyik? Ha igen, akkor**

**melyik(ek). Ha nem, akkor miért nem?**

Szupraszegmentális szinten: beszéddallam, hangsúlyok, esetleg ritmus, tempó. tehát ez igen.

**d) Lejegyezhető-e a beszélő személy által mondott szöveg?**

Nem. Rengeteg információ hiányzik, kb csak annyi állapítható meg hogy magánhangzó

vagy mássalhangzót ejt az illető, de még ezek határa is nehezen meghatározható.

**e) Megállapítható-e a beszélő személy neme egy ilyen hangfelvételből?**

Igen. Az alapfrekvencia megfigyelhető, és ebből következtethetünk a nemére is.

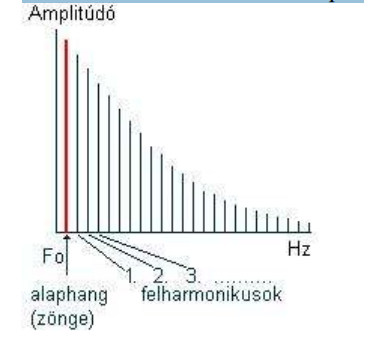
**f) Rajzolja le a periodikus hangszalagrezgés spektrális képét.**

A hangszalagrezgés képe: van egy alapfrekvencia (x Hz, ahol x 100-300 között van), ami

a spektrumban egy vonal. Ennek felharmonikusai, azaz többszörösei (n\*x Hz) is

megjelennek a spektrumban, de egyre kisebb amplitudóval. A csökkenés -12 dB

felharmonikusonként. Lásd a képet:



**5. Tői hangot digitalizálunk8 kHz, 16 bites lineáris mintavételezéssel.**

**Az átlapolásmentesítő szűrő hibás, az átviteli karakterisztikája a 4000 Hz-es felső**

**határ helyett már 2000 Hz-től levág 60 dB/oktáv meredekséggel. A bemondott**

**üzenet a következő: „-Nyolcezerötszáz lesz a végösszeg.”**

**a) Milyen szöveget fogunk észlelni a helyes rekonstruáló szűrővel ellátott visszaállító**

**kimenetén?**

Érthetetlen lesz, hiszen rengeteg fontos frekvencia ill. formáns van a 1000-2000 Hz-es

tartományban, pl a magánhangzók második formánsának jórésze bele esik ebbe a

tartományba. Valami mély mormogást hallunk, gyanítom. (Egyéb, pontosabb ötlet?)

**b) Mennyi lesz a jel/zaj viszonya az így elkészített beszédnek?**

Jel/Zaj viszony: SNR=1,74+n\*6,02=1,74\*16\*6,02=98,06

**c) Mennyire sérül a beszéd dallama a hibás szűrő miatt?**

A beszéd dallama nem sérül, hiszen ezt az alapfrekvencia adja meg (F0), aminek a

mozgását a hangterjedelem adja meg. Ez pedig tipikusan 100-400Hz közötti érték, amit a

szűrő még átvisz.

**6. Egy triádos adatbázisú, hullámforma-összefűzéses szintetizátorral a következő**

**mondatot állítjuk elő: "Miért 40% a határ?". Írja le milyen feldolgozási lépések**

**valósulnak meg a példamondaton, amíg a szövegből a végleges hullámforma előáll!**

Első lépés: begyűjtés! helyett Graféma->Graféma konverziók, avagy a különféle

jelölések feloldása, hogy csak betű legyen az output, mégpedig: "Miért negyven

százalék a határ?"

• Graféma->Fonéma konverziók avagy a g és y nem külön g és y hanem "gy".

Karakterek helyett beszédhangokat írunk. Ezt valami SAMPA átírással lehetne jól

leírni.

• Fonéma->Fonéma konverziók avagy nem negyven-nek ejtjük ezt a szót így,

hanem netyven-nek. Hasonulások, összeolvadások, rövidülések, kivetések.

Eredmény (SAMPA-ban lenne ildomos írni): Mi(j)ért netyven százalék a határ?

• Mindezekkel párhuzamosan fontos a prozódia mondatszintű, szószintű stb

lebontása, relatív megadása. Ugyanígy intenzitással is. Amennyire lehetséges,

hangsúlyhatárokat is bejelöljük (pl vessző előtt felmegy).

• Ha mindez megvan, egy adatmátrixot kapunk, melyben a szöveg minden lényeges

elemét hangokra lebontva megadtuk, ami a kiejtéshez kell. Ezek főbb vonalakban:

frázishatárok, szünetek, hangsúly, időtartam, F0, F0 töréspont, intenzitás. Utóbbi

4-et %-ban célszerű megadni.

• Ezt az adatmátrixot kapja meg a triádos beszédgenerátor.

• A beszédgenerátor veszi a hangkódokat a jelölésnek megfelelően. CVC helyzetbe

triádot keres, egyéb helyzetekben pedig diádot.

• Ezek hangosságát, frekvenciaszerkezetét és periódusidejét megváltoztatja a

megadott százalékoknak stb. megfelelően.

• A szükséges helyekre megfelelő nagyságú szünetet illeszt be.

• Az egyes elemeket simító algoritmusokkal összefűzi.

• Utolsó lépés: a profit!

**2001.04.10 ZH**

**1. Adja meg a megfelelő mértékegységben annak a 80 Hz-es szinuszos hangnak a**

**hangnyomásszintjét, (érzeti) hangosságszintjét és (érzeti) hangosságát, amelynek**

**effektív hangnyomása 0.02T/m2! (15 pont)**

Hangnyomasszint, mas neven akusztikai decibel (ld. jegyzet 2. o.)

L=20\*lg(P\_eff/(20\*10^-6\*Pa)) [dB] erzeti hangossagszint: ehhez phon gorbesereg kell

(zh-n adnak), most ld. 5. o. ha pl. 60 dB szamoltal ki az elobb a 80Hz szinuszos hangra,

akkor megnezed, hogy 80Hz-nel melyik gorbe van a legkozelebb 60 dB-hez. Utana nezd

meg, hogy ennek a gorbenek (fuggvenynek) mi az erteke 1 kHz-nel, mondjuk legyen 80

dB. Akkor a valasz 80 phon. erzeti hangossag: veszed az elobb kiszamolt phon erteket, es

40 phon -> 2^0=1 son, 50 phon -> 2^1=2 son, 60 phon -> 2^2=4 son, es igy tovabb. van,

ahol nem ilyen szepen viselkedik a son-gorbe, olyankor adnak egy abrat.

**2. Vázolja egy 100Hz frekvenciájú szinuszjel és egy ugyanilyen alapfrekvenciájú**

**magánhangzó spektrumának jellemző tulajdonságait és ismertesse az ezzel**

**kapcsolatban tanult fogalmakat! (10 pont)**

100 Hz szinuszos jel spektrumanak egy nemnulla komponense lesz, pontosan 100Hz-nel,

es az erteke a jel amplitudoja (itt nincs megadva az amplitudo). persze a spektrum

szimmetrikus az y tengelyre, tehat -100Hz-nel is oda kell biggyeszteni. maganhangzoknal

ugye van egy kvaziperiodikus "alapjel" gerjesztes, ami a hangszalaktol jon felfele. ez

ffiaknal ~100Hz, noknel ~200Hz, gyerekeknel ~300Hz (ez az f\_0). a maganhangzokat

ezen kivul azert szeretjuk, mert a spektrumuk formans strukturat mutat. a formansok a

spektrumra illesztet burkologorbe maximumai, es f\_0 tobbszorosenel vannak. konkret

peldak pl. jegyzet 12. o. ide lehet meg irni, hogy a jo megerteshez a komm. eszkoznek at

kell vinnie az elso ketto-harom formanst, amit a telefon meg is tesz, ezert a mgh. jol

ertjuk telefonban. massalhangzonal (ez itt nem kerdes asszem) ugye a gerjesztes inkabb

feherzaj szeru, tehat mindenfele frekvenciakomponens elofordul, es egesz magas

frekvenciakon is vannak fontos komponensek, ezeket a tel. nem viszi at, ezert pl. "f" "s"

(asszem) nehez megkulonboztetni.

**3. Egy 8 kHz-es mintavételi frekvenciával és az alábbi, H(f) karakterisztikájú**

**visszaállítóval működő mintavételező rendszer bemenetére két szinuszos jel összege**

**kerül (jellemzőik: 2kHz, 6Vpp és 5kHz, 2Vpp). H(f) = 1, ha abs(f) ≤ 3.5; (4 - abs(f))**

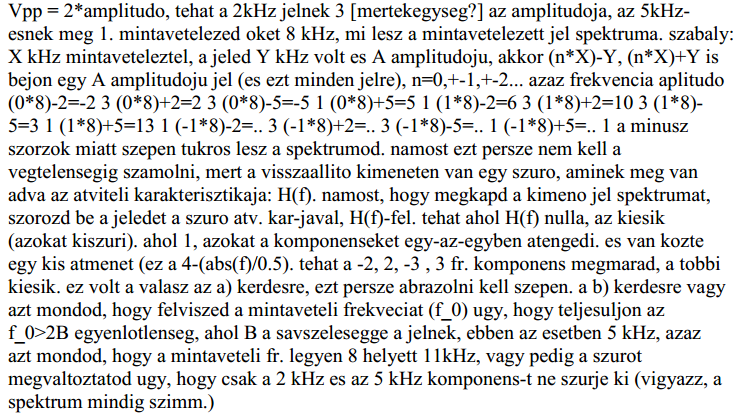
**/ 0.5 , ha 3.5 < abs(f) < 4 ; 0, egyébként (a frekvencia mértékegysége [kHz])**

**a) Milyen jel kerül visszaállításra? (5 pont)**

**b) Javasoljon egy olyan mintavételi frekvenciát és összetett simító karakterisztikát,**

**amely a fenti jelet helyesen és elfogadható komplexitással megvalósítva átviszi! (15**

**pont)**



4**. Egy jelet másodfokú predikciót alkalmazó rendszerrel viszünk át bináris**

**csatornán. a) Határozza meg a prediktort, ha R11 = R22 = 1, R12 = R21 = R01 = 0.8 és R02 =**

**0.6! (10 pont)**

**b) Rajzolja fel a kódoló és a dekódoló részletes felépítését! (5 pont)**

**c) Hány bites kvantálót kell alkalmazni a 60dB jel-zaj viszony eléréséhez, ha a**

**predikciós nyereség 30dB? (5 pont)**

itt egy lin. egyenletrendszert kell megoldani. R = [ R\_11 R\_12 R\_21 R\_22 ], W = [ w\_1

w\_2 ] B = [ R\_01 R\_02 ] es R\*W=B 2 ismeretlen, 2 egyenlet. matlabbal, mert az jo: >

R=[1 0.8

0.8 1]; // az egyutthato matrix > B=[0.8

0.6]; // az eredmeny oszlopvektor > RR=inv(R) // ezzel majd balrolszorozzuk

RR = 2.7778 -2.2222 -2.2222 2.7778

> W=RR\*B // most szorozzuk balrol

W = // ez az eredmeny

0.8889 -0.1111 > R\*W // ellenorzeskeppen

ans =

0.8000 0.6000 // es visszakaptuk B-t. kiraly.

b) rajzolja fel... nem tudom! c) ezt meg azt mondta a bacsi hogy nem tanultuk

# 

# 

# 2002.04.11 ZH

**1. Egyszerre szól egyenként 500, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500 és 5000**

**Hz alapfrekvenciájú és 70dB intenzitásszintû hang. Mekkora az ezen komponensekbõl**

**álló komplex hangnak az össz intenzitásszintje? (10 pont)**

L = 20\*lg(P / 20 \* 10ˇ-6) = 10\*lg( I / 10ˇ-12W)

70dB = 10\*lg( I / 10ˇ-12W)

7dB = lg( I / 10ˇ-12W)

10ˇ7 dB = ( I / 10ˇ-12W)

10 komponensre:

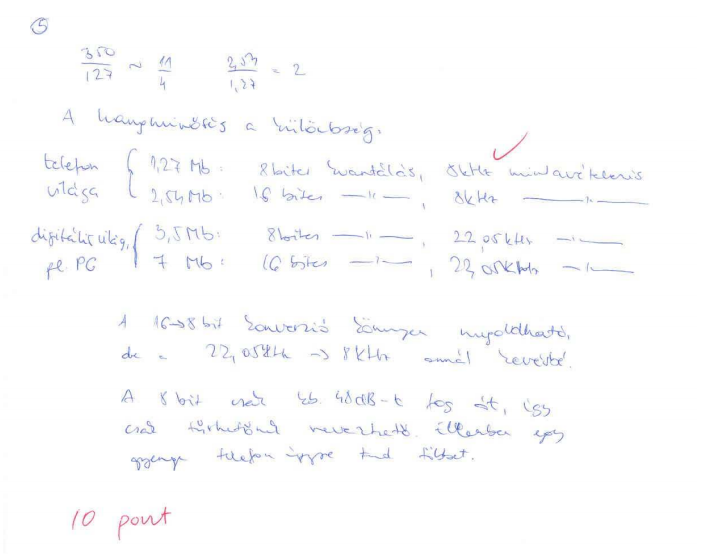
I = 10 \* (10ˇ7 / 10ˇ12) \* (10 dB / W) <- lehet, hogy nem helyesen van lemásolva a sor :D

L = 10 \* lg ( (10\* 10ˇ7) / 10ˇ-12) = 10 \* lg(10ˇ8) dB = 80 dB

**5. Mi lehet az oka annak, hogy egy német nyelvû, 50 hangot tartalmazó diádos**

**adatbázis 4 különbözõ változatban, 7 Mbyte, 3.5 Mbyte, 2.54 Mbyte és 1.27 Mbyte**

**méretben is elkészült? (10 pont)**



**a) Magyarázza meg, hogy mit jelent a diád és a triád hangsorépítõ elem a fizikai**

**valóságban. Ismertesse mindkettõ használatának az elõnyeit és hátrányait. Hogyan**

**lehet kiküszöbölni a hátrányokat? (5 pont)**

**b) Milyen hangtulajdonságok határozzák meg egy diádos hangsorépítõ elem fizikai**

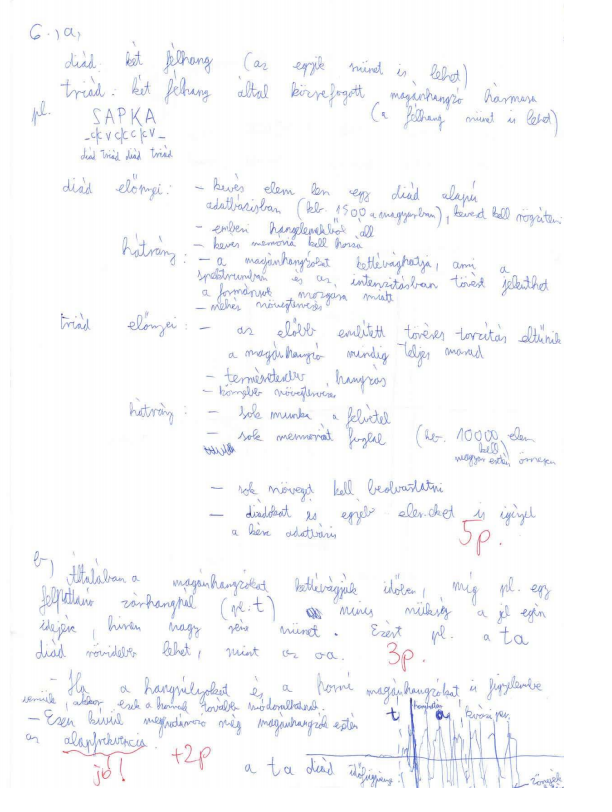
**hosszát? Adjon példákat rajzzal. (5 pont)**

**c) A magyar nyelvre kb. hány diádot kell elkészíteni, hogy szöveget lehessen**

**felolvastatni egy beszédszintetizátorral? (5 pont)**

**d) A triádos koncepciójú elembázisba hány triádos elemet célszerû tervezni (magyar**

**nyelv esetén). A triádos koncepciójú elembázisban milyen a triádok és diádok aránya?**



# 

# 

# 

# 

# 

# 

# **2012 vizsga, teszt**

**PSOLA eljárással módosítjuk a “szín” szó alapfrekvenciáját 250Hz-ről 200Hz-re. A szó 100+100+100 = 300 ms hosszú. Mennyi lesz a módosított szó időtartama, ha az időváltozást nem kompenzáljuk?**

megoldás: (fordítottan arányosak)

250 hz = 0.004 - 300 ms :

200 hz = 0.005 - (szerintem) 3.75

A megoldókulcsban 350 szerepel, nemtom why.

**Szubjektív beszédminősítés szabványa:**

ITU-T P.800

Rejtett markov modell (HMM) beszédfelismerőse mely állítás igaz, mely hamis?

- 1440 elemes diádos adatbázissal működik - Hamis - (nem sok köze van a konrkét megvalósításhoz)

- betanítást igényel - igaz

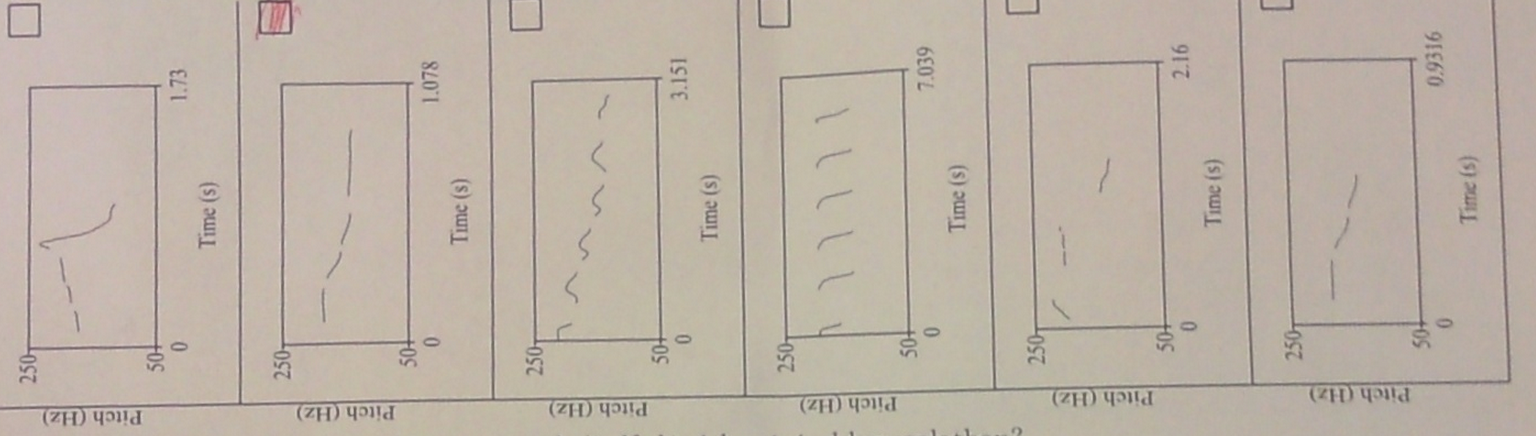
- NEm tartalmaz lényegkiemeltő - hamis

- A felsimerésnél a modell paraméterek rejtettek, azaz ismeretlenek - hamis - nem ezt jelenti a rejtett, hanem hogy az állapotot nem tudjuk, hogy hol tartunk, erre kell a megfigyelésből következtetni

- Csak beszélőfüggő lehet - hamis

- Érzéketlen a zajra - hamis

**A “10.000ft”-os összeg felfolvasását egy magyarországon helyesen lokalizátt rendszerből rögzítjük. Melyik F0 átmenet tartozik hozzá?**



Szerintem azért az, mert egyrészt magas hangsúlybl megyünk mélye. A tíz - ez-zer -forint hangok határánál szokott ilyen lenni (Z kis szünet). pont 4 rész.

**Mely állítások igazk az időtartománybéli elfedés jelenségével kapcsoaltban?**

(jegyzet 1.4.5)

- Az LPC kódoló figyelembe veszi - hamis

- Az MP3 kódoló figyelembe veszi - igaz

- Az A-law kvantálás figyelembe veszi - hamis (itt jegyzem meg, hogy görény kérdés imo.)

- A később érkező hang elfedheti a korábban érkezőt - igaz (ez a lényege kb.)

- Csak korábban érkező hang fedheti el a később érkezőt - hamis

- Csak azonos ferekvenciájú hnagoknál jelentkezik - hamis - mert van egy ablak, amin belül jelentkezik

**Ha ismeri egy magánhangzót tartalmazó bemondás első formánsának frekvenciaértékét, megállapítzható-e egyértelműen a beszélő neme, illetve megállapítható-e egyértelműen, hogy melyi magánhangzó került bemondásra?**

1.9.1.1 a jegyzetben

- A magánhangzót nem tudjuk megmondani, mert nem az első, hanem a második(F2) formánstől függ a pontos hang

- a beszélő neme sem mondható meg, mert bár a nők magasabb F1, F2-ket használnak, de a lefedett freki területek átfedésben vannak: kellene a vonalas spektrum (férfiaknál 100hz, nőknél 200hz-enként van ugrás) vagy az F1-F2 érték(távolság alapján már be lehet lőni ha “férfi nem mondhatja”)

**Az emberi hallás hány kritikus sávval írható le?**

adatbázisos terület: az a terület, ahol a hangosságérzet nem érzékeny a sávszélességre (1.4.3)

válasz: 24, [8. oldal](http://ganymedes.lib.unideb.hu:8080/dea/bitstream/2437/97014/1/szd_btzs.pdf)

**Mely beszélőszervek szükségesek a zönge kézpéséhez?**

Hangszalagok, Tüdő. Ha nem vágod, nézd végig a CD mellékletet: a lényeg, hogy a zöngéből “csinál” a többi érthető hangot, de nem azok keltik a zöngét magát.

**A hullámforma összefűzés szintetizátorban…**

- Nem csak diádok használhatóak - IGAZ - , lehet triádok is.

- CVC triádok esetén nem szükségesek diádok: - HAMIS - . Diád: két féhlnagból álló hullámforma. Triád: CVC elemre a C-ket felénél, V egészben marad, aztán C.

- A diádok 2 félhagnot és eg egészet tartalmaznak - HAMIS - ez a triád

- A PSOLA nem használható - HAMIS -

- A triádok tetszőleges F0-ra jó minőségben módosíthatóak - HAMIS - ezt nem teljesen értem

**-** szükséges lehet a hangidőtartamok módosítása - IGAZ - pl. egy magánhangzó minél hátrébb csúszik egy szóban, annál röidebben ejtjük. Óra - karóra

# **2011. vizsga**

**Teszt**

**Lehet-e beszédet tömöríteni MPEG1 Layer3-as kódolóval úgy , hogy j minőségű maradjon?**

Igen, de csak nagy bitsebesség mellett - mpeg1 32 kb/s, a telefon 64kb/s-kel tömörít, szal tippre ezért. egyébként akinek jó a füle, a 196 és 256-os tömörítéskkete is kihallják

**LPC alapú beszédátvitelnél milyen hosszúságú ablakokat használunk?**

(Linear Predictive Coding):

[Ez alapján](http://support.ircam.fr/docs/AudioSculpt/3.0/co/LPC.html) 10.000Hz - 22.000 Hz passz

**Ha egy 8 bites lienáris kvantálás helyett 16 bites lineáris kvantálást használunk, akkor a kvantálási zajra vonatkozóan hány dB-es jel-zaj viszony javulást érünk el?**

2.3.2.. SNR ratio N kvantálási szint esetén C/delta = N/2.

tippre eggyik sem (4 dB talán

**A “só” szót kiejtve milyen rész nem található a beszédjelben?**

9. oldal a jegyzetben

- Zörejes - VAN - , “s”

- Periódikus - VAN - “ó”

- zárfelpattanás - NINCS

- 5kHz feletti komponens - NINCS (1.9.1.1)

- zönétlen - VAN - “s”

- Fojtott zönge - NINCS - ez a lökéshullámoknál van

**LPC alapú beszédátvitelnél mennyi az általában alkalmazott LP paraméterek száma?**

- 12 db, passz miért. (gondolom a 24 db-os kritikus izé miatt)

**Melyik érték lehet egy férfi beszélő mgaánhangzójának F1 frekvenciája?**

- F1: 200Hz és 1000 Hz között, (F2: 700 és 2500Hz között)

**Melyik hang zöngétlen?**

- G, d, zs, dz, z zöngés

- s pl. zöngétlen

**Egy 7kHz-es szinusz helet szűrés nélkül mintavételezünk 10kHz-en. A mintevett jelben hol jelenik meg a bementi szinusz jel?**

- TODO

**Melyik eljárást használná kvázi-periodikus beszédhangok visgálatzához?**

- Fourier-sorfejtét. A kvázi-periodikus a zenei hangok, sok komponens élesen elkülönül, stb.

**Feladatok:**

**2. Rendezlezésre áll a következő mondat telefonon felvett hullámforma állománya A-törvényű kvantálással, 8kHu mintavételi frekvenciával: “szép idő van”.**

**a.) adja meg pontokba szedve, hogy miylen műáveletsorral lehet ebből a mondatból kérdő mondatot kialakítani a PSOLA algoritmus felhasználásával!**

todo

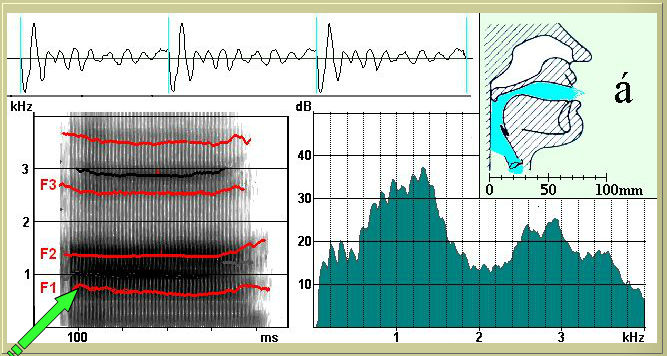
**b.) Adja meg törtvonalas közelítéssel (X=idő, y=alapfrekvencia), hogy milyen lesz az alapfrekvcencia változás az eredeti és az átlalakított mondatban**

todo

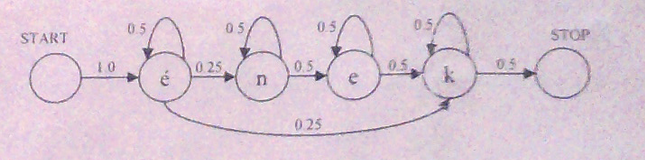
**Adja meg a törtvonalas közelítéssel (x=idő, y=intenzitás), hogy miylen lesz az intezitás-idő függvény az eredeti és az átalakított mondatban)**

todo

**3. Vázolja egy 100Hz-frekvenciájú szimmetrikus négyszögjel és egy ugyanilyen alapfrekvenciájú magánahazó spektrumának jemmelző tulajdonságait és imsertesse az ezzel kacpsoaltban tanult fogalmakat!**



Itt az Á hang. Érdemes a ms-es dolgokat felvinni, a négyszögjelnek csak egy formánsa van. A

4. 

**a.)Az ábrán látható rejtett markov modell hálózat milyen szavak felismerésére alkalmas, ha az egyes állapotok eloszlásfüggvényeit az azonosítő betűjelüknek megfelelő hangok jellemzővektiraival tanítottuk?**

- Tippre az “Ének” és “Ék” szavakéra.

- amúgy: [é|éne]k ismerhető fel, minden egyes szó végtelenszer ismétlődhet, de vissza nem lépünk, tehát:

ének, ééének, ééééénnnnnnneek, ééééékkkk, ék, stb. nem tudom hogy lehet szépen leírni :P

**b.) a fenti hálózatot izolát szavas beszédeflismerésre akarjuk használni. A beszédhangokat egy dimenziós, egység szórású Gauss-függvényekkel modellezük, melyek várható értékei a következőek: M”é” = 2.0; M”n” = 4.0; M”e”=1.0; M”k” = 3.0. Összesen 4db jellemzővektor érkezett a lényegkiemelőből: O1=1.0; O2=2.0; O3=2.0; O4=4.0; Melyik szó a felismerés eredménye?**

négy variáció lehet:

ének

Ékkk

ÉÉkk,

ÉÉÉk

ez jön ki nektek is az ÉÉKK pl.?: (Ugye a képletet mindenki tudja a korábbi feladatokból:P)

- M"é" + o1 = 0.4\*e^0.5 = 0.242

- M"é" + o2 = 0.4\*e^0 = 0.4

- M"k" + o3 = 0.5\*e^0.5 = 0.242

- M"k" + o4 = 0.5\*e^0.5 = 0.242

így 1.0 \* 0.242 \* 0.5 \* 0.4 \* 0.25 \* 0.242 \* 0.5\* 0.242 \* 0.5

ezek értékei:

...ének = 0,000023

ééék = 0,000288

éékk = 0,0001728

ékkk = 0,00010368

**5.**

**a.) Milyen követelményeket támasztana a felhasználandó bezédatabázissal szemben, ha az adatbázissal betanított beszédfelsimerőt egy szövegszerkesztőhöz kapcsolt irodai diktálórendszer részeként kívánják használni?**

- olyan minták kellenek, amik az irodai zajviszonyokban is használhatóak

- megfelelő beszédvariáltásg, ne egy ember alapján, hanem széles felhasználás

- mikrofon, sima szöveg, nem kell telefonhanghoz igazítani

- szótár lehet kissé specifikus, egyes területeken, irodai nyelvezeten több minta

- beszlőfüggetlenség

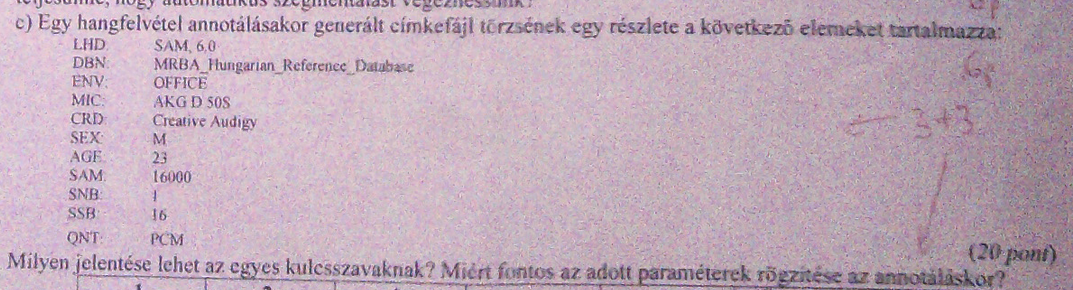
- beszédataptáns legyen

- adatbázis sok beszélővel legyen

- külö számokra érzékeny legyen

**b.) hogyan történhet a beszédatabzási automatikus szegmentálása? Mi az eljárás neve? Milyen előfeltéreleknek kell teljesülnie, hogy automatikus szegmentálást végezhessünk?**

passz

**c.)** 

# 

# 

# 

# 2010.01.07 vizsga

**1. feladat (v070525 - 6.feladat)**

**a, Mi a lényeges különbség a felhasználás szempontjából a beszélő-függő és a**

**beszélő-független beszédfelismerők között?**

A beszélőfüggetlen rendszereket bárki, bármikor használhatja előzetes betanítás

nélkül, viszont általában kisebb szótárral és megbízhatósággal rendelkeznek. A

beszélőfüggő rendszerek általában beszélőadaptívak is egyben, azaz

használatukhoz szükséges egy előzetes betanítási fázis, ezután azonban több

szót és jobb megbízhatósággal képesek felismerni, izolált szavak helyett akár

kapcsoltszavas vagy akár diktáló üzemmódban is.

**b, Betanításnál milyen típusú adatbázist kell az egyik és a másik rendszerhez?**

Beszélőfüggetlen rendszer esetén több beszélőtől szükséges hanganyag, hogy

ebből közös jellemző vonásokat tudjunk kivonni a betanítás során a minél

robusztusabb működéshez. Beszélőfüggő rendszer esetében pedig a hangok

paraméterbecslésére nincs szükség (vagy jóval kisebb adatbázis is elegendő),

hiszena betanítási fázis során pont ezeket a paramétereket hangoljuk az adott

beszélő alapján. Minden más vonatkozásban (szótár felépítése, nyelvi modellek

stb) a két megoldás nem különbözik, illetve max. a szavak számában.

**c, Milyen egyéb szempontokat kell figyelembe venni?**

Szótárméret, tematika, a hangkörnyezet (zajos utca v csendes iroda),

beszédmodor (spontán vagy dialógusszerű), stb...

**2. feladat**

**a, Gépi beszédminősítő rendszert hogyan ellenőrizné.**

Gépi == objektív beszéd minősítés. Hatékonyság mértéke: korreláció MOS-sel

vagy minimális négyzetes eltérés MOS-től. Tehát a gépi minősítés ellenőrzéséhez

el kell végezni (gondolom szúrópróba szerűen) a MOS minősítést is.

**b, VoIP esetén mi okozza a legnagyobb problémát a beszédminőség mérése során?**

A válasz: Jitter (késleltetés-ingadozás). (A hálózat paramétereinek nem stabil

volta. Teljesen más minőséget kapunk ha kis illetve szélessávon mérünk, illetve

változatos kapcsolat (műholdas, kábel, adsl) esetén is jelentős eltéréseket

tapasztalhatunk a beszéd minőségében, a hálózatforgalmi szituációkat nem is

említve (pl. ha közben töltünk is).) Nem a rizsára voltak kíáncsiak.

**5. feladat**

Na itt def, ábra, példa kellett:

**LPC**

Lineáris predikció. Bizonyos esetekben ha nem a mintát, hanem a minták

különbségét kvantáljuk, akkor kevesebb biten tudjuk átvinni ugyanazt az

információt (tömörítés).

**F1**

Az első formáns. Magánhangzók jellemzéséhez a formánsokat

használjuk. A formánsok a spektrális felbontás burkológörbéjének

maximumai. (A következő definíciót vastag vonallal fogják lesatírozni-> Az

alapfrekvencia az F0, ennek egészszámú többszörösei a formánsok. F1 az

F0-hoz legközelebb eső formáns.)

**pitch**

Az alapfrekvencia, azaz F0. A PSOLA (Pitch Synchronous Overlap Add

Method) eljárás az alapfrekvenciát változtatja a beszéd dallamának

változtatása érdekében.

**ergodikus**

A folyamat tulajdonsága. Ha a folyamat ergodikus, akkor 1 folyamat is

magában hordozza a sokaság tulajdonságait.

**VXML**

Voice eXtended Markup Language - dialógusok leírására alkalmas

**son**

Relatív hangosság. Hányszorosa a hangosság a 40 phon hangnak?

Képlettel: N = 2 ^ [(L - 40) / 10] L: phon , hangosságérzet. 40 phon = 1

son, 50 phon = 2 son , 60 phon = 4 son stb.

**aliasing**

Ez a jelenség azt eredményezi, hogy a mintavételezett hullámforma

torzulni fog visszaállításkor. Pl.: a Nyquist kritérium szerint 10 000 Hz

mintavételezési frekvencia mellett az 5000 Hz feletti frekvenciájú

hangkomponensek torzulni fognak visszaállításkor. Ezt úgy kezelik, hogy egy alul áteresztő anti-aliasing filterrel kiszűrik az 5000 Hz feletti

komponenseket.

**HMM**

Hidden Markov Model - Rejtett Markov Model. Statisztikai alapú

beszédfelismerési módszer.

**triád**

Egy triádos adatbázisban három hosszú hangkombinációkra tároljuk a

hullámformákat. A tárolt adatok mennyisége ezért a fonémák köbével

arányos. A tárolt hang az első fonéma közepénél kezdődik és a harmadik

közepénél fejeződik be.

**MP3**

MPEG-1 Audio Layer 3. (nem MPEG-3) Veszteséges tömörítés. Az emberi

fül számára nem hallható hangokat eldobja.

Részletesebben:

Érzeti (részsávos kódolás) – percetptual (subband) coding

Frekvenciamaszkolási jelenség (elfedés a frekvenciatartományban)

1)minden sávra megállapítjuk, hogy milyen energiájú összetevok vannak a

jelben

2)az elfedett összetevok kihagyása

3)a kvantálási zaj növelésének lehetősége: úgy kvantálunk, hogy a

kvantálási zaj ne legyen nagyobb, mint az elfedési szint (itt a tömörítési

lehetőség)

**1. Feladat:**

1.1: Folyamatos, nagyszótáras, beszélőfüggetlen felismerő betanításához készülő beszédadatbázisában jellemzően..

**Sok beszélő működik közre** | diádokat és triádokat rögzítenek. | CVC hang?? rögzítenek | **fonetikailag gazdag mondatokat rögzítenek** | inkább női beszélők működnek közre. | jól képzett beszélők működnek közre.

1.2: Hogyan fejezhető be a mondat, hogy igaz legyen: A kvantálás…

sűrűsége függ a mintavételi frekvenciától. | csak egyenletes közökkel használják a beszédre. | veszteséggementesen visszalakakítható. | **független a beszéd alapfrekvenciájától** | csak 8, 16 és 32 bites lehet. | sűrűsége nem befolysolja a hangminőséget.

1.3: Melyik hang(ok) zöngltlen(ek)? (A hangok közelítő betűképükkel jelöltük)

g | z | d | dz | **sz** | za

1.4: Egy nyelv 24 hangot (6 magánhangzó és 18 mássalhangzó) tartalmaz. A tanult megoldás esetén minimum hány diád és CVC triád elem kell a nyelv minden hangkapcsolatának lefedéséhez?

24^2 | **24^2 + 6\*18^2** | 24^2 | 24^2+18\*6^2 | 6\*18\*24 | 6^2+18^2

1.5: Mihez kapcsolódik az ITU-T P.800-ban szabvány?

A-law, u-law kódoláshoz |Objektív veszédminősítéshez |HMM beszédfelismerési eljáráshoz | PSOLA algoritmushoz | Korpuszos TTS adatbázisához | **Szubjektív beszédminősítéshez**

1.6: Kempelen Farkas a beszédkutatás mely területén alkotott jelentőset?  
 **Beszélőgép (1791) (?)**

1.7: 15 ember közül az aktuálisan beszélőt kell felismernie beszéde alapján. Ha nem ismert a tartalom, melyik technilógiát használná ehhez?

**Beszélő azonosítás**

1.8: Egy igényes kötött szótáras magyar nyelvű számfelolvasó rendszer hullámforma elembázisa hány építőelemet tartalmazhat, ha a rendszer 1 és 1 milliárt közötti számokat tud felolvasni?

1 milliárd | 10 | 2500 | 25 |**250** |1500

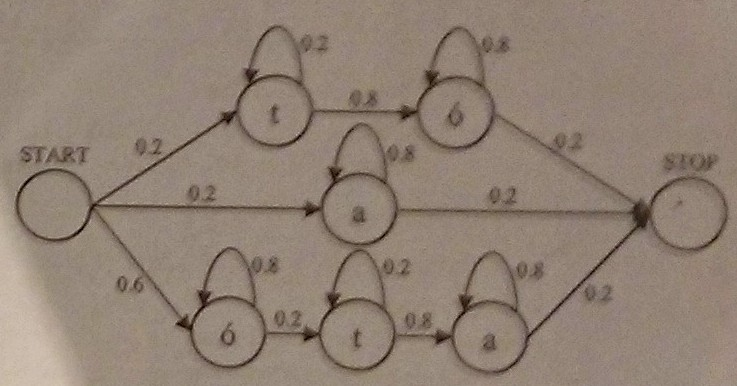
1.9: Hogyan fejezhető be a mondat, hogy az állítás igaz legyen? Előszűrű nélkül 10kHZ-en mintavételezve a beszédjelet, majd 4,5kHZ-es aluláteresztővel visszaállítva…

...9kHZ körüli beszédkomponens biztosan kiszűrhető. | ...a hangminőség a telefonosnál biztosan jobb lesz. | ...tökéletesen visszaállítható a jel minden esetben. | **...többnyire érhető, de zajos lesz a jel.** | ..biztosan érthetetlen, használhatatlan lesz a kimenet |...az átlapozódást kivédtük.

1.10: A “áttelel” szó kiejtésében a “t” hangra milyen tulajdonságok jellemzőek?

némafázisa 100us hosszú | felpattanással kezdődik | **összetett szerkezetű hang** | nagy energiájú hang | **kb 150-200 ms hosszú a “t” hang** | fojtott zönge meghosszabbodik a hosszú “t” miatt.

**2. Feladat**



Az ábrát újra rajzolom hamarosan olvashatóbbra.

a, Az ábrán látható rejtett Markov-modell hálózat milyen értelmes, magyar szavak felismerésére alkalmas, ha az egyes állapotok eloszlásfüggvényeit az azonosító betűjüknek megfelelő hangok jellemzővektoraival tanítottuk? (2 pont)

b, A fenti hálózatot izolált szavas beszédfelismerésre akarjuk használni. A beszédhangokat egy dímenziós, egység szorású Gauss-függvényekkel modellezzük, melyek várható értékei a kovetkezők: m(t)=1.0; m(ó)=4.0; m(a)= **3.0;** Összesen 2 db jellemzővektor érkezett a lényegkiemelőből, sorrendben: O1=2.0; O2=4.0; Melyik szó felismerés eredménye? (8 pont)

két jellemzővektor, és át kell érnünk startból stopba -> vagy aa, vagy tó.

t(o1) = 0.2\*0.4\*0.606

ó(o2) = 0.8\*0.4\*1

a(o1) = 0.2\*0.4\*0.606

a(o2) = 0.8\*0.4\*0.606

tó = (még 0.2vel szorozva) 0,00310272

a = (szintén, Stop miatt) 0,00188024832

tehát a TÓ a megoldás.

**3. Feladat**

Hahyományos CD lemezen (44100 Hz mintavételi frekvencia, sztereó felvétel, 16 bites lineáris kvantálás) áll rendelkezésre egy 3 perc hosszú zeneszám. Szeretnénk belőle csengőhangot készíteni egy olyan mobiltelefonra, ami 11025 Hz-es mintavételi frekvenciával tud mono, 8 bites, A-törvényű logaritmikus kvantálású mintákat lejátszani.

a, Milyen műveleteket kell az átalakító programnak elvégezni? A választ illusztrálja ábrákkal.

(8 pont)

b, Mekkora tárhelyet igényelnek az egyes formátumok és miben tér el a minőségük? (6 pont)

c, Vissza lehet-e állítani az eredeti felvételt a telefonos formából? Ha igen, hogyan? Ha nem, miért nem? (6 pont)

**a) Ábrákkal illusztrálja az átalakítás folyamatát! (8 pont)**

Ábrák helyett az egyes lépések (kis dobozkákat rajzolnék egymás után, bennük az egyes

lépések neveit írnám):

• Visszaállítom a kvantált, mintavételezett jeleket (sztereó!) analóggá.

• Átlagolom a két jelet időtartományban, amplitudó szerint 1 mono jellé.

• Aluláteresztő szűrő, mely 5 kHz-ig engedi át a jelet, persze 5 kHz körül lineáris

gyengítéssel.

• Mintavételezés 11,025 kHzen.

• Kvantálás 8 biten.

**b) Hány zeneszám van a lemezen? Valamennyi zeneszám átalakítható-e? Ha nem,**

**mi lehet a megoldás? (6 pont)**

1 sec hanganyag tárigénye: 44,1kHz mintavételezés, 16 bit, sztereó hangsávok:

44100\*16\*2= 1,411,200 bit = 172kbyte. 800Mbyte/172kbyte= 4763, azaz 4763 sec

hanganyag tárolható, ami kb 79 perc. Ez 3 perces zeneszámokkal számolva 26 zeneszám.

Nem alakíthatók át azok a számok, mely 5kHznél magasabb frekvenciakomponenseket

tartalmaznak. Megoldás erre a fentebb már említett aluláteresztő szűrő.

**c) Vissza lehet-e állítani az eredeti felvételt a telefonos formából? Ha igen, hogyan?**

**Ha nem, miért nem? (6 pont)**

Nyilván nem lehet visszaállítani a telefonos formából, ennek több oka is van. Egyrészt a

monó hang átlagolással készült a sztereó hangsávokból, ezt lehetetlen visszaszűrni. (2 és

6 átlaga 4. 4 melyik két szám átlaga?). Másrészt az alacsony mintavételezés miatt

elvesztjük az 5kHz feletti komponenseket, ezeket sem tudjuk visszanyerni. Harmadrészt

pedig a 8 bites logaritmikus kódolás nem arányos a lineáris 16 bitessel, ezért főleg a

magasabb tartományokban nagyobb lesz a kvantálásból eredő zaj nagysága

**4. Feladat**

Adja ma a következő fogalmak és rövidítések jelentését és válaszoljon a feltett kérdésekre.

a, Mi a SAMPA, az ATM, a PHON, az F0 és a PSOLA jelentése és kapcsolata? (4 pont)

* **SAMPA**: Speech Assessment Methods Phonetic Alphabet. Beszédhangok jelölése 7 bites ASCII karakterekkel. A SAMPA-val a beszédhangok egyértelműen leírhatók, segíthet a beszédértés
* **F0:** alapfrekvencia
* A **PSOLA** (Pitch Synchronous Overlap Add Method) eljárás az alapfrekvenciát változtatja a beszéd dallamának változtatása érdekében.
* **PHON:** hangosságérzet *mértékegysége.* A Phon görbe az azonos hangosságérzetű görbék serege, ahol a referencia-frekvencia az 1 kHz. Azaz 1kHz-es hangok esetén a Phon érték megegyezik az akusztikai dB-el.
* **ATM:** Atmoszféra, az úgynevezett ideális légkörben közepes tengerszinten mért légnyomás értéke, 1 atm = 101325 Pa

Mind a beszédakusztika fizikai értékeihez kapcsolódnak

b, Mi a teljesítmény sűrűség spektrum és az autokorrelációs függvény jelentése és kapcsolata? (4 pont)

* Az akusztikai dB-ből visszakövetkeztethetünk a hangjel amplitudójára (10-es hatványraemelés), az így kapott időjel négyzete a teljesítmény sűrűség spektrum. (ha jól mondom :] )
* [wat](http://itl7.elte.hu/jelfel/node22.htm)

c, Mi a “Hamming-ablak” és az “ovális-ablak” jelentése és kapcsolata? (4 pont)

* A **Hamming-ablakot** a jelre illesztve egy véges időtartományban kell csak elvégezni a Fourier-integrálást. A szerepe az, hogy adott időpillanatban releváns frekvenciákat felerősítse, a távoliakat gyengítse hogy adott időpillanatra jó spektrumot kapjunk a Fourier-integrálás után.
* Az **Ovális ablak** az a rész, ahol a csigához kapcsolódk a kengyel a fülben.

szerintem nincs köztük kapcsolat, csak a témakör.

**5. Feladat**

a, Mit mutat egy vonalas spektrum a beszédre vonatkoztatva? (2 pont)

**Az egyes vonalak tavolsaga megegyezik az alapfrekvencia (F0) ertekevel. A zonge spektruma a vonalas spektrummal jol kozelitheto, mivel a hangszalagokkal kepzett zonge kvazi-peridoikus.**

b, Milyen Fourier-eljárással határozható meg elméletileg a vonalas spektrum? (2 pont) **Fourier-sor**

c, Rajzolja le a hangszalagok felett a gégében mérhető hangnyomás-idő függvény közelítő spektrumát, ha nő ejti ki az “á” hangot. (4 pont)

d, Hány felharmonikust tudunk megszámolni a c, válasz ábráján a 4000 Hz-es pontig?

(4 pont)

e, Rajzolja fel az ajkakról kisugárzott “á” hang hangnyomás-idő függvénye vonalas spektrumának közelítését (a magánhangzó háromszögről tanultakat is vegye alapul. (6 pont)

**CD**

2. Megoldások

# **1. Bevezetés**

**1. Mi az artikuláció?",**

["az adott nyelvre jellemző beszédképző mozgások automatikus létrehozása","a száj mozgása","a beszéd megértése"]),

**2-Jelölje a hamis állítást!",**

["a beszédhangokat leggyakrabban a tüdőből kiáramló levegő segítségével hozzuk létre","a levegőt beszívása közben nem vagyunk képesek artikulációs mozgásokat végezni","a csettintő hangokat a magyarban állathívogatásra használjuk"]),

**3. A mássalhangzók alábbi három osztálya közül melyik a leggyakoribb a világ nyelveiben?",**

[“zöngés zárhangok","zöngétlen zárhangok","nazálisok"]),

**4. "Mi a beszéd dekódolása?",**

["az elhangzott közlés megértésének folyamata","a hangszínkép megfeleltetése a hangsornak","az elhangzott közlés észlelésének folyamata"]),

**5. Hány hangból építkezik a Hawaii nyelv, amelyik a legkisebb ismert hangrendszer?",**

["5 magánhangzóból és 8 mássalhangzóból","5 mássalhangzóból és 8 magánhangzóból","3 mássalhangzóból és 3 magánhangzóból"]),

**6. "Melyik állat hangját utánozza a 'kwa-kwa' szó?",**

["béka","kacsa","az anyanyelvtől függ, a lengyel anyanyelvűek számára a kecskéjét"])];

# 2. Beszédkeltés, hallás

**1. Hány részből áll a légzőrendszer?",**

["tüdő és rekeszizom","tüdő, rekeszizom, garat, hörgők","tüdő, légúti rendszer, rekeszizom"]),

**2. Milyen átlagos hosszúságú és átmérőjű a gége?",**

["6 cm hosszú és 4 cm átmérőjű","11 cm hosszú és 2,5 cm átmérőjű","18 cm hosszú és 2,5 cm átmérőjű"]),

**3. Melyek a gégét felépítő fontosabb porcok?",**

["pajzsporc, gyűrűporc, kannaporc, gégefedő","pajzsporc, nyelvcsont, hangszalagok","gégefedő, gyűrűporc, kannaporc"]),

**4. "Mit nevezünk toldalélcsőnek?",**

["a tüdő és a hangszalagok közötti részt","a hangképző üregrendszert","az orrüreget"]),

**5. A hangképző rendszerben mettől meddig tart a szájüreg?",**

["a garatüregtől az ajkakig","a kemény szájpadtól az ajkakig","a hangszalagtól az ajkakig"]),

**6. Mit nevezünk artikulációs csatornának?",**

["a beszédszerveknek a hangréstől az ajkakig terjedő szakaszát","a szájpadlás és az ajkak közti részt","a lágy szájpad és az orrüreg közti részt"]),

**7. Mit nevezünk hangrendszernek?",**

["a magas és mély magánhangzók váltakozását","az adott nyelv beszédhangjainak összeségét","az adott nyelv zöngés hangjainak számát"]),

**8. Hány fő részből áll a fül?",**

["2","4","3"]),

**9. A fülnek melyik részében található a kalapács, az üllő és a kengyel?",**

["külső fül","belső fül","közép fül"]),

**10. A beszéd agyi központjai melyik féltekén találhatók?",**

["mindkettőben","bal agyféltekében","jobb agyféltekében"])];

# 3. Akusztikai alapfogalmak

**1. Hogyan jellemezné a periodikus rezgésformát?",**

["zörej érzetét keltik és több nem harmonikus rezgésből állnak","azonos időközökkel megszólaló zörejelemek","a hangrezgés fázisai periodikusan visszatérnek és zenei hangok érzetét keltik"]),

**2. A 3,5 kHz-es szinuszhang hangszínképén a 0-8 kHz-es tartományban hány vonal jelenik meg?",**

["kettő: 3,5 és 7 kHz-nél","egy: 3,5 kHz-nél","három: 1,25 kHz-nél, 3,5 kHz-nél és 7 kHz-nél"]),

**3. Mikor lép fel rezonancia?",**

["amikor egy közeg a másik közeg által keltett hangrezgéstől rezgésbe jön","amikor a hangszalagok a kiáramló levegőtől rezgésbe jönnek","amikor a hangszalagok rezgése következtében felhangok keletkeznek"]),

**4. Mit nevezünk egy rezonátor sávszélességének?",**

["azt a frekvenciasávot, ahol a maximális amplitudóérték jobb és baloldalt 70,7 százalékra csökken","azt a frekvenciasávot, amin belül a rezonátorüreg rezonancia-frekvenciája jobb és bal oldalra megváltozik","azt a frekvenciát, amelyen a rezonancia bekövetkezhet"]),

**5. Ha egy 100 Hz-es, egy 500 Hz-es és egy 1000 Hz-es szinuszrezgést összekeverünk, mennyi lesz a rezgés alapfrekvenciája?",**

["500 Hz","1000 Hz","100 Hz"]),

**6. 100 Hz-es frekvenciájú hangnál hány periódust számlálhatunk meg 50 ms időtartam alatt?",**

["5 periósust","2 periódust","50 periódust"]),

**7. Mettől meddig tart az infrahang tartománya?",**

["1 Hz-től 20 Hz-ig","20 Hz-től 500 Hz-ig"," 20 kHz-től 30 kHz-ig"]),

**8. 6 dB-nyi hangerőemelkedés hányszoros intenzitásváltozásnak felel meg?",**

["hatszoros"," tízszeres","kétszeres"]),

**9."Az 1 kHz-es rezgésszámú négyszögjel hangszínképén hány vonal jelenik meg a 0-8 kHz-es tartományban?",**

["8","1","4"]),

**10. Az 1 kHz-es rezgésszámú fűrészfogjel hangszínképén a 0-8 kHz-es tartományban hány vonal jelenik meg?",**

["8","4","1"])];

# 4. Beszédakusztika

**1. A beszédhangok közötti időtartam-arányok a beszédre jellemző időtényezők melyik csoportjába sorolhatók?",**

["a nyelvi időtényezők csoportjába","a fizikai időtényezők csoportjába","mindkettőbe"]),

**2. A t hang belső időszerkezetének változása a beszéd során befolyásolja-e a hang hangzását?",**

["nem","igen","csak a hang hosszúságára van hatással"]),

**3. Megnyúlhat-e egy hangsoreleji magánhangzó időtartama akár 1,8-szeresére a hangsor hosszától függően?",**

["nem","a magánhangzó időtartama független a hangsor hosszától","igen"]),

**4. "Mit nevezünk formánsnak?",**

["a zöngének a rezonanciaüreg által felerősített felhangnyalábját","a zöngés beszédhangokat","az alaprezgés egész számú többszörösénél mérhető felhangokat"]),

**5. Az s hangnál a 0-4 kHz-es frekvenciatartományban hány formánst lehet kimutatni?",**

["egyet sem, mert az s hang formánsai magasabban helyezkednek el","a zörejhangoknak nincsenek formánsai, hanem zörejgócai vannak, tehát egyet sem","egyet, mert az első formáns 3800 Hz-nél van"]),

**6. 200 Hz-es zöngerezgés esetén hány vonal jelenik meg a hangszínképen a 0-1 kHz-es tartományban?",**

["5","2","4"]),

**7. Hány szintet különböztetünk meg fonetikailag a beszéd elemzésekor?",**

[" 2 szintet ( szegmentális és szupraszegmentális)","1 szintet (szupraszegmentális szint)","A beszéd fonetikai elemzésekor a monoton beszédjelet külön vizsgáljuk, tehát 3 szint van"]),

**8. "Melyik fonetikai szinthez tartozik a beszéd ritmusa?",**

["szegmentális szinthez","szupraszegmentális szinthez","egyik szinthez sem"]),

**9. Szintetizált hangsorok segítségével vizsgálható-e külön a beszéd szupraszegmentális szerkezete?",**

["nem, mert a beszédről a szupraszegmentális összetevőket leválasztva értelmetlen hangsort kapunk","igen","nem, mert a szintetizált beszéddel csak a szegmentális szerkezet vizsgálható"])];

# 5. A Magyar Beszéd

**1. "A magyar beszédhang-állományban hány magánhangzó van?",**

["9","5","14"]),

**2. A magyar beszédhang-állományban hány mássalhangzó van?",**

["25","27","26"]),

**3. A képzés helyétől függően az 'u' hang hova tartozik?",**

["hátul képzett","elöl képzett","középen képzett"]),

**4. Melyik az a magyar magánhangzó, amelynek formánsértékei a következők: F1 240 Hz, F2 1750 Hz, F3 2600 Hz?",**

["é","ü","u"]),

**5. Melyik az a magyar magánhangzó, amelynek formánsértékei a következők: F1 600 Hz, F2 1100 Hz, F3 2300 Hz?",**

["a","ü","e"]),

**6. Melyik az a magyar mássalhangzó, amelyiknek van képzési hely szerinti párja?",**

["b","m","h"]),

**7. Az á magánhangzót hogyan képezzük az ajakműködés szempontjából?",**

["ajakréssel","ajakkerekítéssel","ebből a szempontból semleges képzésű"]),

**8. Mi a közös a b és a g mássalhangzókbanl?",**

["zárképzés","résképzés","néma fázis"]),

**9. Mi az n hang zöngétlen párja?",**

["t","d","nincs zöngétlen párja"]),

**10. "Hol van a képzés helye a v és f hangoknál?",**

["ajak-fog ","elülső szájpad","fogmeder"]),

**11. Hol van a képzés helye az l és r hangoknál?",**

["két ajak ","gége","fogmeder"]),

**12. Hol van a képzés helye a j hangnál?",**

["elülső szájpad ","fogmeder","gége"]),

**13. Mit nevezünk artikulációs sebességnek?",**

["az időegység alatt elhangzó beszédhangok számát","az időegység alatt elhangzó beszédhangok és szünetek számát","az időegység alatti ajakmozgás változásának sebességét"]),

**14. "Melyik hangnak van néma fázisa?",**

["g","sz","t"]),

**15. Melyik hangnak van zárfelpattanása?",**

["s","f",**"**k"]),

**16. Mit jelent a prozódia szó?",**

["a szép beszéd művészete","a beszéd dallama, hangsúlyozása és ritmusa","a beszélő szónoki képessége"]),

**17. Mi az a hangsebészet?",**

["gégeoperáció","a beszédhangok részeinek kivágása egy jelből","két beszélő hangjának összekeverése"]),

**18. Milyen hangsor lesz az -asa- hangsorból, ha az s hangot kivágjuk?",**

["aa","ara","ata"])];

# 6. Adatbázisok

**1. Mire szolgálnak a beszéd-adatbázisok?",**

["különböző beszédtechnológiai rendszerek támogatására","nyelvtanításra","beszédhibák javítására"]),

**2. Mit jelent a címkézés?",**

["címkék elhelyezése a szövegben","a beszédjel feldolgozása során nyelvi jelzések elhelyezése","beszédhibák megjelölése a beszédjelben"]),

**3. Milyen adatbázis a BABEL?",**

["analitikus-diagnosztikus","általános","specifikus"]),

**4. Milyen adatbázis a SPEECHDAT?",**

["analitikus-diagnosztikus","általános","specifikus"]),

**5 .Mit jelent a diád szó?",**

["dupla hang","két félhang","két hang kapcsolata"]),

**6. Mit jelent a triád szó?",**

["három beszédhang sorozata","két félhang között egy egész hang","három magánhangzó találkozása"]),

**7 .A hullámforma-szintetizátor 1500 építőelemből dolgozik a szöveg felolvasása során. Milyen építőelemekből áll az adatbázisa?",**

["triád","két félhang","a magánhangzók és a mássalhangzók"]),

**8. Hány hanghatár van egy diádban?",**

["1","2","3"]),

**9. Hány hanghatár van egy triádban?",**

["1","2","3"])];

# 7. Beszédszintézis

**1. Mikor mutatta be Kempelen Farkas a mechanikus beszélőgépét?",**

["1802 ","1798","1791"]),

**2. Mikor készült az első magyar nyelvű szövegfelolvasó rendszer?",**

["1990 ","1982","1988"]),

**3. Mi az első lépés egy kötött szótáras beszédszintetizáló elkészítésénél?",**

["az üzenetek szövegét kell meghatározni","hangfelvételt kell készíteni","meg kell határozni a felhasználandó beszédhangok csoportját"]),

**4 .Mikor adta be Bánó Miklós a Tetszőleges szöveg reprodukálására alkalmas beszélőgép című szabadalmát?",**

["1896","1916","1932"]),

**5. Hogy hívták az első magyar nyelvű számítógépes szövegfelolvasó rendszert?",**

["Pannonvox","Hungarovox","MULTIVOX"]),

**6. Mi a neve az első magyar, emberi hangú, könyvfelolvasó szintetizátornak amit 2002-ben állítottak üzembe?",**

["Profivox","Hangos-könyv","Mondom-2002"]),

**7. "A formáns kódolású beszédszintetizátoroknál mi képezi a szintézis alapját?",**

["a beszédhangok szegmentális tényezőinek mért, számszerű adatai ","a beszédből kiszegmentált hangok, hangkapcsolatok ","a beszéd alapfrekvenciájának értéke"]),

**8. A legújabb fejlesztésű beszédszintetizátorok formáns- vagy élő beszéd alapúak-e?",**

["formáns","élő beszéd","vegyes"]),

**9. Mely összetevők figyelembevétele teszi élethűbbé a monoton hangzású hullámforma-szintetizátor hangját?",**

["a szupraszegmentális összetevők","a formánsösszetevők ","a magasabb frekvenciaértékű beszédösszetevők"]),

**10. Mi a jellemzője a kötött szótáras szintetizátornak?",**

["a kimondandó üzenetek előre ismertek","kötött módszerű az elemhatárok kijelölése","csak meghatározott számítógépen használhatók"]),

**11. Közelítőleg hány hullámforma-építőelemet kell meghatározni egy jó minőségű számfelolvasóhoz?",**

["220","4000","25"]),

**12. Hány formánsparamétert jelenít meg a Multivox szintetzátor fejlesztői rendszere?",**

["2","4","6"]),

**13. Milyen gyakorlati alkalmazása van a beszédsszintézisnek orvosi területen?",**

["felolvassa az orvosnak a diagnózist","hallásvizsgálat","újságfelolvasás a betegeknek"])];

# 8. Beszédfelismerés

**1. Mit jelent az izolált szavas beszédfelismerés?",**

["csak egy szót ismer fel","csak számokat ismer fel","egymástól szünettel elválaszva ejtett szavakat ismer fel"]),

**2. A beszédtechnológia mely területén használják leginkább a Markov-modellt?",**

["beszéddetekció","beszédszintézis","beszédfelismerés"]),

**3. "Mit jelent az idővetemítés?",**

["más-más időablakot használunk","a különböző tempóval ejtett hangsorokat egy közös időtengelyre képezzük le","megszorozzuk a jel hosszát egy konstanssal"]),

**4. Milyen elv alapján működnek a mai beszédfelismerők?",**

["összehasonlítás","fonetikai osztályozás","artikulációs leképezés"]),

**5. Mi a mintaillesztés lényege beszédfelismerési eljárásnál?",**

["a beszélő hangjából képzett vektorokat összehasonlítja előre tárolt mintákkal","a beszélő hangjához illeszti a felismerőt","a beszélő hangjából mintát vesz a felismerő, és egy konstanssal megszorozza"]),

**6. Mit jelent a MEL-skála?",**

["hangosság szerinti ábrázolás","a spektrum átlagolt logaritmikus ábrázolása","a spektrum 1 kHz-ig lineáris, felette logaritmikus osztású ábrázolása"]),

**7. Mit jelent a kapcsolt szavas beszédfelismerés?",**

["összekapcsolja az izoláltan ejtett szavakat","a beszélő folyamatosan beszélhet","rövid szövegrészeket (több szó) folyamatosan ejthet"])];

# 9. Jelfeldolgozás

**1. Az analóg jelek digitális jellé alakításakor mit nevezünk mintavételezésnek?",**

["azt a műveletet, amikor az analóg jelből bizonyos időközönként mintát veszünk ","azt a műveletet, amikor mérjük, hogy az analóg jel periodikus vagy nem periodikus","azt a műveletet, amikor mintánként mérjük, hogy a jel intenzitása az előbbi értékhez képest változott-e"]),

**2. Mit tudunk meghatározni az FFT-vel?",**

["az összetett rezgés összetevőinek frekvenvcia- és intenzitásértékét","az artikulációs csatorna átviteli karakterisztikáját","a hangok belső időszerkezetét"]),

**3. A beszédhangok elemzésénél LPC-analízissel lehetséges-e az adott beszédhangra jellemző artikulációs csatorna átviteli karakterisztikájának meghatározása?",**

["nem","igen","igen, de csak a zöngés hangoknál"])];

3. Random jó ha tudod infók

formánsok a vokális traktus (rezonátorüreg) által felerősített felhangnyalábok

FONTOS: NEM az alaphang és annak felharmonikusai a formánsok, hanem a rezonátorüreg erősítési helyei!

Szegmentális és szupraszegmentális szerkezet:

Szegmentális

A beszédjel vizsgálati szintje: beszédhangonként vagy beszédhang részletenként

Szupraszegmentális

A beszédjel vizsgálati szintje: több beszédhang egyszerre (frázis, mondat, szöveg, …)

Vizsgált jellemzők

dallam, hangsúly, tempo, szünet, ritmus, hangerő

A hangok jelölése: V=magánhangzó, C= mássalhangzó

Fo = alapfrekvencia= a hangszalagok nyitódásának, záródásának gyakorisága: 50-500 Hz. Ez

adja a beszéd dallamát, része van a hangsúlyozásban is. Férfiaknál átlagosan 100 Hz, nőknél 200

Hz.

Formáns= az artikulációs csatorna rezonanciafrekvenciaáin felerősített felhangnyaláb. Jelölése:

F1, F2, F3, F4, F5, Formáns sávszélesség: B1=30-50 Hz, B2=100-300 Hz, B3= 400-700 Hz

A formánsok hatóköre. F1= 250-800 Hz, F2= 500-2300 Hz, F3=200-4500 Hz.

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

ered. A zönge frekvenciája az Fo, alapfrekvencia. Ez a vonalas spektrum alapharmonikusa.

Fojtott zönge= zöngés zár- és zárrés hangok zárszakaszi építőeleme. Formánsai nincsenek.

b,d,g,gy,dz,dzs

Néma fázis= a zöngétlen zár- és zárrés hangok zárszakaszi építőeleme. Hang nem keletkezik a

néma fázisban. p,t,k,ty,c,cs

Zárfelpattanás= zárhangok zárszakaszi eleme utáni hangrész, amely befejezi a zárhangot

(lökéshullám).

Specifikus időtartam= a hangra jellemző alapidőtartam a hangkörnyezet függvényében

Lásd részletesen az MNYBA CD-ről

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

Lásd részletesen az MNYBA CD-ről

Locus=a mássalhangzó artikulációját jellemző frekvencia koncentrációk összessége a frekvencia

tengelyen (Főleg a zárhangokra vonatkoztatják a V F2-jéből következtethető ki).

7. FONÉMA= a sokféle beszédhangból általánosított elméleti egység, amely megkülönböztető

funkciót tölt be a nyelv hangsorainak jelentéstartalmában. Például: bár – pár; már- vár.

A hosszú-rövid hangok is külön fonémaként kezelendők a magyarban. Például: hal - hall; sok –

sók; tör – tőr. A szintézisben a hosszú-rövid magánhangzókat külön kell kezelni, a

mássalhangzók esetén csak a rövid hang időtartamát kell megnövelni (hangfüggő a nyújtási

eljárás).

8. Hanghelyzet. A beszédhangoknak különböző szerkezeti megvalósulásai lehetnek

hangsorkezdő, hangsorbelseji és hangsorzáró helyzetben

9. Artikulációs sebesség (hang/s). Lásd részletesen az MNYBA CD-ről

10. Beszédsebesség (hang/s) Lásd részletesen az MNYBA CD-ről

11. VOT= zönge kezdési idő a zöngétlen zárhang zárfelpattanásától a zönge megindulásáig mért

idő. (Orvosi diagnosztikai eljárásokban is használható.) Értéke nyelvenként változik.

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

• Beszéd

Lineáris kódolás: 8kHz 16bit : 128 kbps

– Telefon (nemlineáris: A/µ law) 8kHz 8bit (12bit) : 64kbps

– VoIP (LPC, CELP…) 4-64 kbps

• Zene

– CD minıség: 44,1kHz, 16bit (stereo) : 1411 kbps

– MP3: 64-128-256-384 kbps

– Realaudio: 32kbps

Adatbázisok:

- A gyakorlatban az adatbázisok tervezésénél figyelembe kell venni, hogy az adatbázis

létrehozása nem más, mint a véletlenszerű folyamat egyes megvalósulásainak (realizá-

cióinak) összegyűjtése. B

- Analitikus – diagnosztikus adatbázis: nyelvi és fonetikai kutatások segítését 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 beszé

A fonetikai átírásnak számos szintje létezik:

Analitikus – diagnosztikus adatbázis: nyelvi és fonetikai kutatások segítését 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

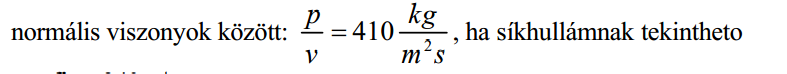
adatbázis.

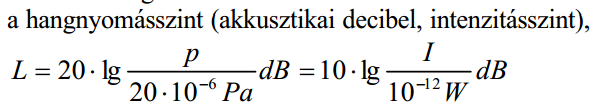
Adatbázisokra jellemző, hogy milyen nyelvi egységekből épülnek fel, pl. izolált szavakból,

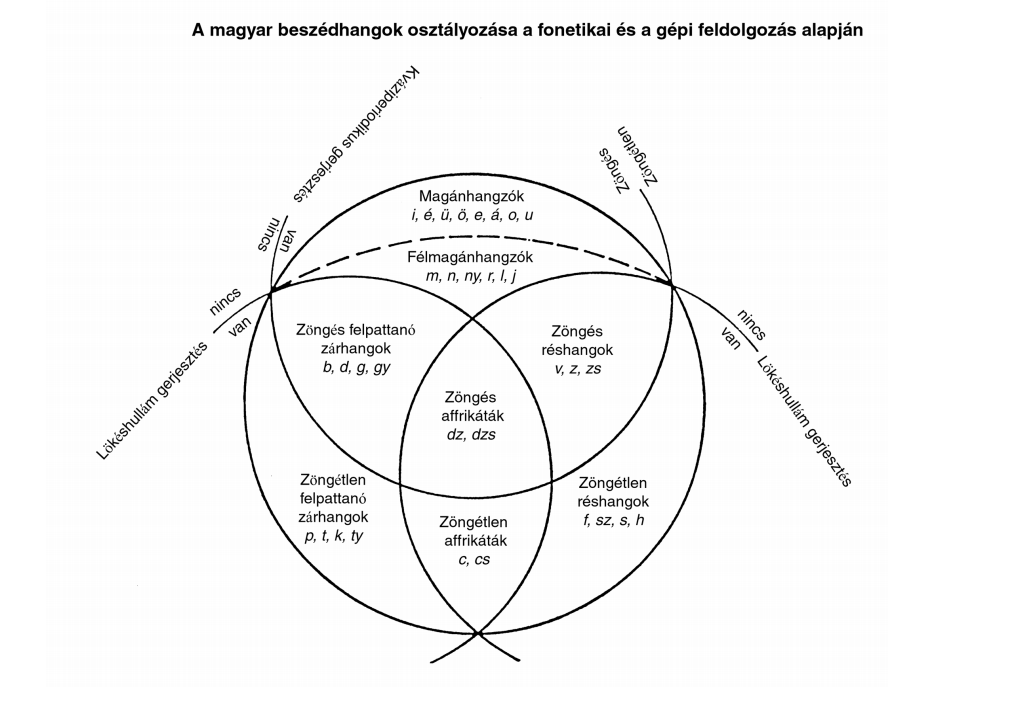
mondatokból stb., továbbá a bemondás módja szerint lehet olvasott szöveg, spontán beszé

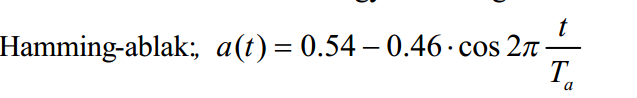
**összefoglalóból:** [**http://wiki.sch.bme.hu/images/f/fc/Beszed\_jegyzet\_teljes.pdf**](http://wiki.sch.bme.hu/images/f/fc/Beszed_jegyzet_teljes.pdf)

a hangot le lehet írni nyomással (p) illetve térfogatsebességgel (v)









nAZÁLISOK Magánhangzóknál

: 250-300 hZ KÖRNYÉKÉN f1, MAGASABBAN A FORMÁNSOK

Likvodák: van formáns, de nem jellegzetzes, kisebb energia mint magánhagnzónál

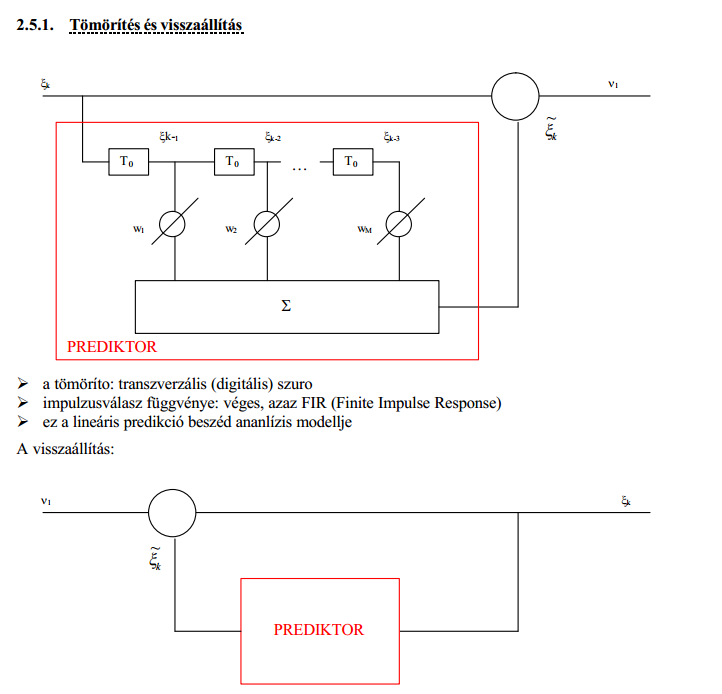
Felpattanó zárhang: néma fázis mad folytott zöngezöngétlen, madj zeárfelpattanás

Réshang: 3.5kHz felett van a spekturm, telefonban nem érted melik

zár-rés: néma - folyottt - zárfelpattan - réshang

A 12 bites lineáris kvantálással 96 kbit/s, míg a logaritmikus(PCM) kvantálással 64 kbit/s átviteli sebességre

van szükség. A fenti tömörítési módszert hívják érzeti tömörítésnek.



gépi beszéd hármom kaegória: kötött, kötetelen és vegyes szókészlet

kötött lépések: Tematika felderítése - bemondanadó szöveg tervezése - szótár kialakítása - bemondó választása - akusztikai bázis elkészítése - rendszerbeillesztés

szubjektív beszédiminősítő endszer: ITU-T P.800