

1. feladat

a. Formáns

- Röviden: A zöngéjelből az artikulációs csatorna üregrendszere által felerősített felhangnyaláb. [forrás: Beszéd CD, 7. oldal]
- Hosszabban: A zöngés beszédhangok létrehozásához két független építőelemre van szükség: a gerjesztő jelre (zöngé, alaphang, alaphangfrekvencia: F_0) és az artikulációs csatornára, amelyik a zöngé jelét átformálja. Az átformálás során a zöngé adott felharmonikusait az üregrendszer rezonanciái felerősítik. Ezeket a felerősített felhangnyalábokat formánsoknak nevezik. [forrás: Beszéd CD, 84. oldal]
- Például az "a" hang formánsai: F_1 500-600Hz, F_2 900-1100Hz, F_3 : 2200-2400Hz

b. Gerjesztés: 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. [forrás: Beszéd CD, 62. oldal]

- 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 szerkezetű a beszédhang, ha időben periodikus vagy állandó.

Összetett szerkezetű, ha belső időszerkezettel is rendelkezik, ez írja le a hangon belüli akusztikai jelenségek időtartam-értékeiket, 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. 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**: A prozódia a beszéddallam, a hangsúly, a ritmus, a hangero, a tempó és a hangszínezet nyelvi használata, a beszédképzés szupraszegmentális szintjének része. [forrás: Beszéd CD, 7. oldal]

Komponensei:

- Dallam
- Hangsúly (ezen belül alkotórészek: alaphangfrekvencia, intenzitás, időtartam)
- Ritmus
- Hangszín (?)

2. feladat

a. Milyen szempontok szerint lehet minősíteni a beszédfelismerő rendszereket?

- 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?

A modell azért rejtett, mert egy megfigyelés esetén nem lehet egyértelműen meghatározni, hogy melyik állapot generálta azt. [forrás: Beszéd CD, 7. oldal]

3. feladat

Kis bevezető:

A HMM feladatok lényege, hogy a START állapotból a STOP állapotba vezető utak közül meghatározzuk a legvalószínűbbet. Egy út valószínűsége nem más, mint az út során meglépett átmenetek valószínűségének és az az út által bejárt állapotokban megfigyelt események valószínűségének a szorzata. Tehát ha elindulok a START állapotból, és először az "ő", utána pedig a "k" állapotba érek, a következő módon történik az út valószínűségének a számítása:

- átmenet STARTból "ő"-be: 1.0
- "ő" állapotban az σ_1 eseményt megfigyelni: 0.242
- "ő" állapotból "k" állapotba lépni: 0.1
- "k" állapotban σ_2 eseményt megfigyelni: 0.054
- "k" állapotból a STOP állapotba érni: 0.2

Ezek szorzata pedig 0.00267, azaz az út valószínűsége 0,267%

Triviális dolgok de inkább leírom mégegyszer:

- Az út felváltva áll lépésekből és megfigyelésekből
- A START állapotból a STOP állapotba kell eljutni
- MINDEN eseményt meg kell figyelni valamely állapotban, azaz annyi közös állapotot kell érinteni START és STOP között, ahány megfigyelt vektorunk van

- Ezen belül az összes lehetséges út közül kell kiválasztani a legvalószínűbbet

Általános esetben dinamikus programozás szerű módszerrel lehet a legjobb utat megtalálni (táblázat kitöltése, alulról felfelé a megfigyelt események szerint sorban haladva vízszintesen, felfele meg a lehetséges állapotok. Balra nem léphetünk, mert újra nem figyelünk meg eseményt, a simán eggyel jobbra való lépés meg a helybenmaradással ér fel, stb, ha az utolsó eseménynél nem értük el a STOP állapotot, az nem út, a STOP állapotig elért utak közül a legjobbat választjuk ki), Viterbi algoritmus segítségével, de ezek annyira egyszerű példák h ránézésre megmondod azt a 2 utat ami lehetséges, kiszámolod melyik mennyi, és az nyer aki előbb..

Megoldás:

Valószínűségek:

- ő állapotban σ_1 -t megfigyelni: 0.242
- ő állapotban σ_2 -t megfigyelni: 0.054
- k állapotban σ_1 -t megfigyelni: 0.054
- k állapotban σ_2 -t megfigyelni: 0.242

Két út lehetséges, ezek valószínűsége:

- 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.

Ezt még annyival egészíteném ki - szerintem egyáltalán nem triviális, bár ki lehet sakkozni -, hogy a megfigyelést úgy kapod, hogy a vizsgasoron megadott Gauss-függvénybe behelyettesítgatsz, m lesz az állapotok/betűk m-je, x pedig az o megfigyelésvektorok. Tehát ha annak a vg-t akarod kiszámolni, hogy az ő állapotban σ_1 -et figyeled meg, akkor a függvénybe $m = m = 5,0$ -t és $x = \sigma_1 = 4,0$ -t helyettesíted be.

-- Zsoltsi - 2007.05.30.

4. feladat

Suttogó beszéd: $1000 * 10^{-6}P = 10^{-3}P$

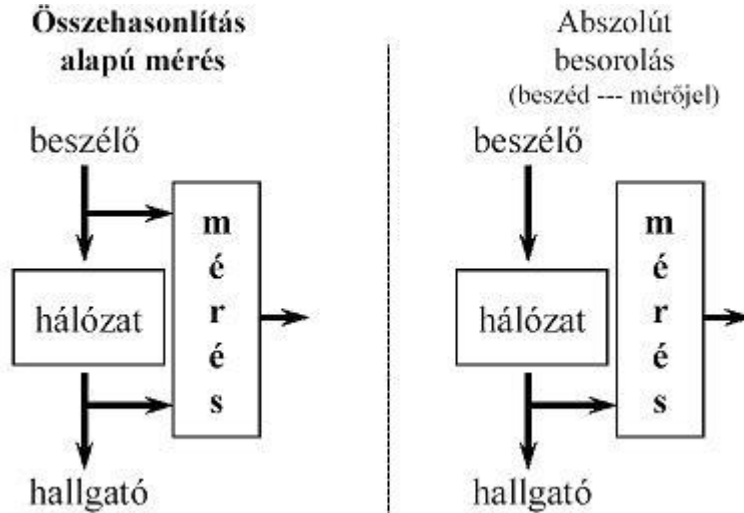
Kiabálás: $1000000 * 10^{-6}P = 1P$

Dinamikatartományuk különbsége: $20 * \log\left(\frac{1P}{10^{-3}P}\right) = 60dB$ (mivel 1000x hangosabb, ez is a jó eredmény -- $20*3$ az 60)

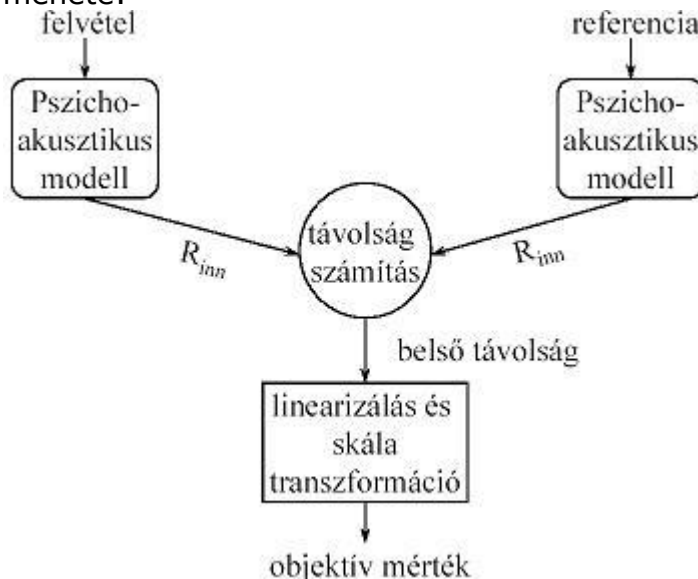
Különbség abban rejlik hogy suttogó beszédben nincs zöngés gerjesztés, így alapprofundencia és formánsok sem, tehát a magánhangzók vonalas színeké helyett is folytonos színeképet kapunk spektrális elemzésnél. (ehhez még sztem lehetne írni..)

5. feladat

- Beszédminősítő rendszer blokkvázlata I.: összehasonlítás alapú mérés:



- Beszédminősítő rendszer blokkvázlata II.: a jelfeldolgozás általános menete:



A hálózat az az átviteli technológia vagy kódolás, amelynek beszédminőségét minősíteni szeretnénk. A mérés menete a második ábrán van kifejtve, ezen belül a psichoakusztikus modell az, ami az ember számára lényeges, hallható részek kiemelése és a nem érzékelhetőek, észlelhetőek elnyomása, majd ezekből ablakozással és Fourier transzformációval jellemző vektorokat ad ki keretként. 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.

(Ez így korrekt?)

6. kreatív feladat

- első lépésként 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 elkezd mondani a gép hogy miket lehet venni: (vonali)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/nyugdíjas](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), "előlről" paranccsal pedig újratekesheti 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 előlről ha nem tetszik valami. Pl.: "Két darab vonaljegyre dobjon be 300 forintot. Ha meggondolta magát vagy félreértettem, mondja vagy nyomja hogy vissza vagy kezdjük előlről" (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.

-- RGabo - 2006.05.27.