2007 június 15-i vizsga megoldása

Feladatsor: a lap alján csatolva, vagy az infosite-on: InfoSite - 2007 június 15. (A Csoport)

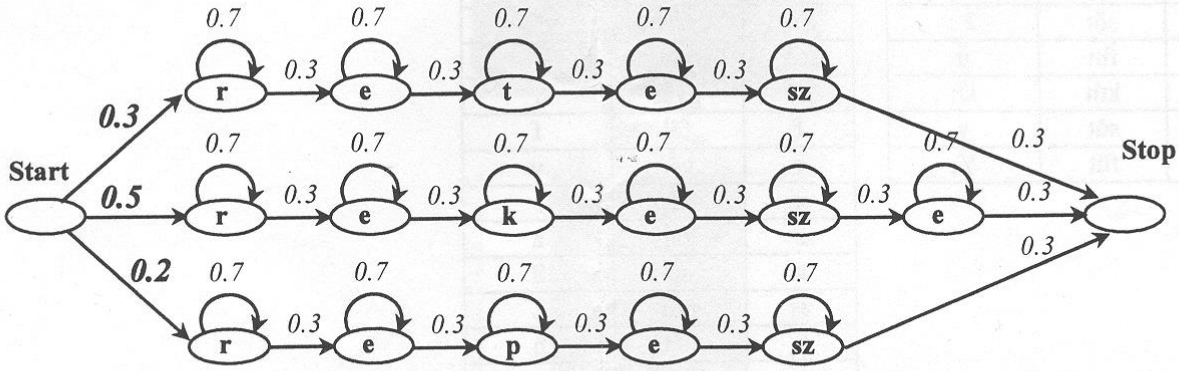
1. feladat

1.
a) Röviden válaszoljon az alábbi kérdésekre
   a) Milyen beszédkódolási eljárásokat ismer
   b) Milyen mintaillesztési eljárásokat ismer
   c) Milyen területeken használhatóak a beszédminősítő eljárások
   d) Mi a megfigyelési valószínűség?
   e) Mi a különbség az akusztikus és a nyelvi modell között?

a. PCM: Pulse Code Modulation (logaritmikus), ezen belül van az A-law (EU) és μ-law (USA). Lineáris kvantálás. LPC - lineáris predikció. MPEG (layer 3).
b. Szabályalapú, statisztikai alapú (HMM - Hidden Markov Model és ANN - Artificial Neuro Network) illetve sablon alapú (DTW).
c. Mindenhol, ahol hang- ill. beszédátvitel történik, így mobilhálózat, telefonvonal, VoIP, stb.
d. Azt az értéket adja meg, hogy mennyi annak a valószínűsége, hogy egy HMM rendszer x állapotában j jellemzővektort figyeljünk meg.
e. Az akusztikus modell az egyes beszédhangokra ad egy referencia-jellemzővektorokat, míg a nyelvi modell a beszédhangok kombinációs lehetőségeit adja meg szótárak segítségével, illetve akár a ragozáshoz nyújt megfelelő szabálybázist.

2. feladat

2.



Beszédfelismerési kísérletet végeztünk a fenti HMM hálózattal. Összesen 5 db jellemzővektor érkezett. Sorrendben a 3 megfigyelésének valószínűségei az egyes beszédhangmodellek esetén

$b_{"r"}(o_3) = 0.1, \quad b_{"e"}(o_3) = 0.6, \quad b_{"t"}(o_3) = 0.2, \quad b_{"k"}(o_3) = 0.8, \quad b_{"p"}(o_3) = 0.8.$

El tudjuk-e dönteni ez alapján, hogy mi a felismert szó? Ha igen, mi volt az, ha nem, miért nem? (15 p

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 o_3 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. feladat

3. a) Mit jelent egy beszédadatbázis szöveganyagának annotálása, és mit jelent a szegmentálása?
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) (15 por

a. 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. 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.
    - másegyéb?
    - odahívta --> odahífta: részleges has., zöngétlenedés

(Írásban nem jelölt) teljes hasonulásra példa: anyja --> annya, hagyja --> haggya másik irányban működő: község --> kösség, tizennyolc

4. feladat

A hangszalagrezgést elektroglottográf segítségével (10KHz-es, 16 bites lineáris mintavételezéssel) rögzítjük, majd visszajátsszuk. A beszélő a következő szöveget mondta: "Eljössz velem? Nem megyek. Nem? Bárcsak eljönnél, úgy szeretném!" (Volt ZH kérdés is - 2009 ősz)

a)Milyen beszédjellemzőket lehet meghallani egy ilyen hangszalagrezgésről készített hangfelvételből ami biztosan az elhangzott beszédhez tartozik? Legalább hármat soroljon fel.

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.

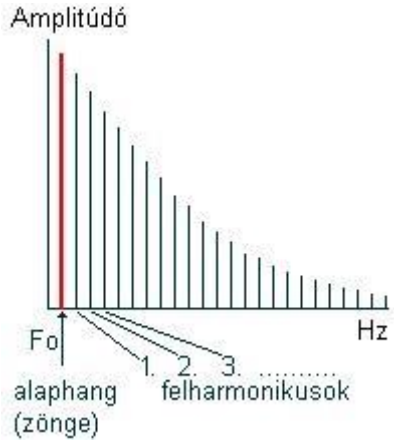
c) Hallható-e a beszéd szupraszegmentális elemei közül valamelyik? Ha igen, akkor melyik(ek). Ha nem, akkor miért nem?

Szupraszegmentális szinten: beszéddallam, hangsúlyok, esetleg ritmus, tempó.

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. feladat

5. Női hangot digitalizálunk 8 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?.
b) Mennyi lesz a jel/zaj viszonya az így elkészített beszédnek?
c) Mennyire sérül a beszéd dallama a hibás szűrő miatt? (15 *pon*

a. 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. Jel/Zaj viszony: SNR=1,74+n*6,02=1,74*16*6,02=98,06
c. 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. feladat

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! (Volt ZH kérdés is - 2009 ősz)

- 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! 🙂

-- Gabo - 2008.05.28.