2006 június 2-i vizsga megoldása

Feladatsor: InfoSite - 2006 június 2.

A csoport

1. feladat

a. *Milyen jellel mérjük a beszédátviteli rendszerek minőségét?*
Természetes emberi beszéddel, de érdektelen felvételeket kell felolvastatni az alanyokkal! (nem vagyok teljesen biztos h ezt kérdezik..)

b. *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.

c. *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)

d. *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). Bővebben: jegyzet

2. feladat

a. Mikor és ki készítette az első beszédkeltő gépet a világon? Hol látható?
Kempelen Farkas, *1791*-ben. Ma az MTA Nyelvtudományi Intézetében látható. (legalábbis ajánlom neki h ottlegyen..) !! Update: "Az egyetlen megmaradt példány ma a müncheni Deutsches Museumban van."
Forrás: http://hu.wikipedia.org/wiki/Kempelen_Farkas

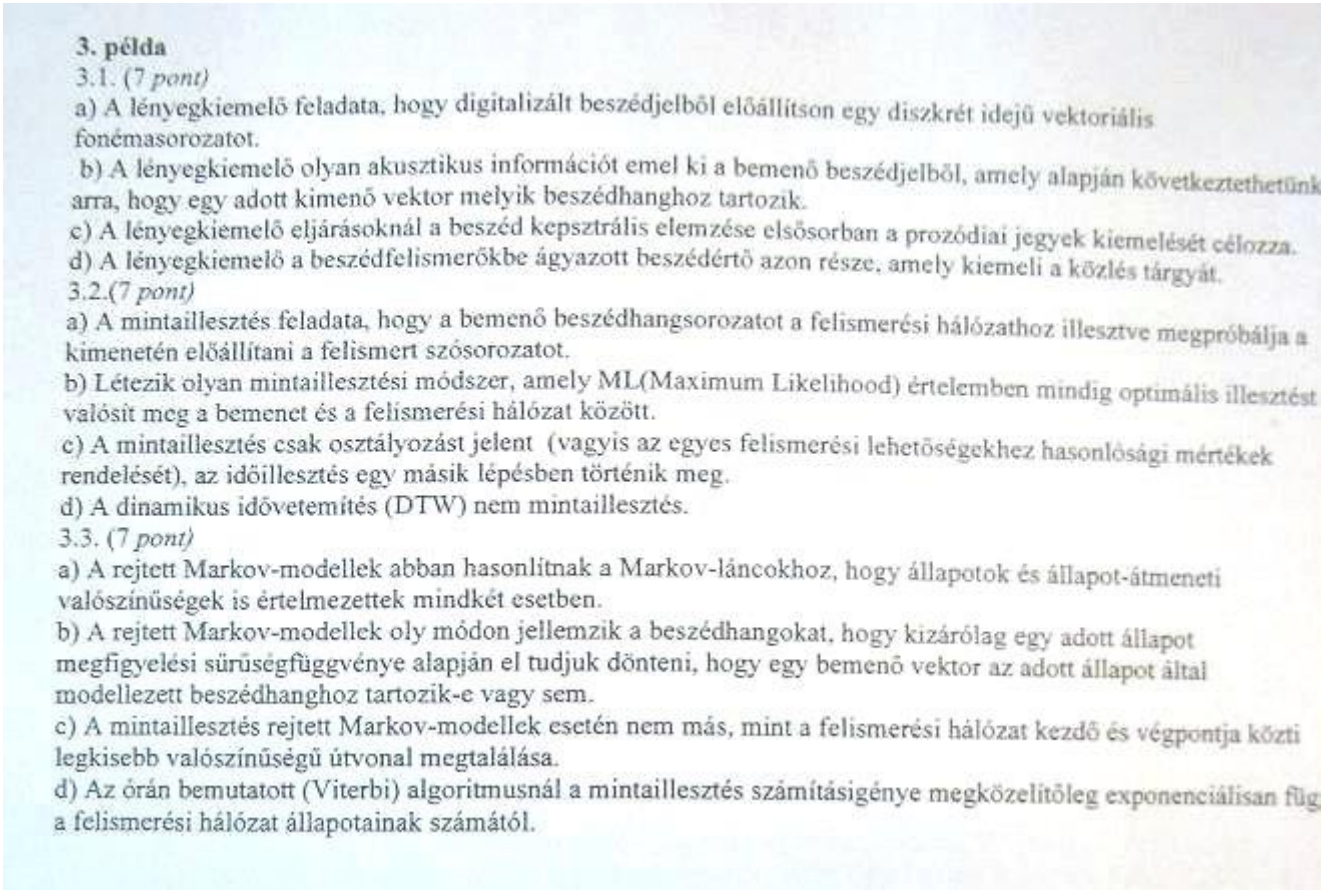
b. Mikor és ki adta be a világ első szabadalmát tetszőleges szöveg felolvasására alkalmas beszélőgépre?
Bánó Miklós, *1916*-ban.

c. 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

d. 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)
e. 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 állítható elő egyértelműen/hűségesen.
   - Megelőzhető megfelelő karakterisztikájú aluláteresztő szűrővel a bemeneten. (Sávkorlátozás)
   - Példát mindenki remélem tud adni ezek alapján :]
f. 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.

3. feladat



3. példa

3.1. (7 *pont)*

a) A lényegkiemelő feladata, hogy digitalizált beszédjelből előállítson egy diszkrét idejű vektoriális fonémasorozatot.

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

c) A lényegkiemelő eljárásoknál a beszéd kepsztrális elemzése elsősorban a prozódiai jegyek kiemelését célozza.

d) A lényegkiemelő a beszédfelismerőkbe ágyazott beszédértő azon része, amely kiemeli a közlés tárgyát.

3.2.(7 *pont)*

a) A mintaillesztés feladata, hogy a bemenő beszédhangsorozatot a felismerési hálózathoz illesztve megpróbálja a kimenetén előállítani a felismert szósorozatot.

b) Létezik olyan mintaillesztési módszer, amely ML(Maximum Likelihood) értelemben mindig optimális illesztést valósít meg a bemenet és a felismerési hálózat között.

c) A mintaillesztés csak osztályozást jelent (vagyis az egyes felismerési lehetőségekhez hasonlósági mértékek rendelését), az időillesztés egy másik lépésben történik meg.

d) A dinamikus idővetemítés (DTW) nem mintaillesztés.

3.3. (7 *pont)*

a) A rejtett Markov-modellek abban hasonlítnak a Markov-láncokhoz, hogy állapotok és állapot-átmeneti valószínűségek is értelmezettek mindkét esetben.

b) A rejtett Markov-modellek oly módon jellemzik a beszédhangokat, hogy kizárólag egy adott állapot megfigyelési sűrűségfüggvénye alapján el tudjuk dönteni, hogy egy bemenő vektor az adott állapot által modellezett beszédhanghoz tartozik-e vagy sem.

c) A mintaillesztés rejtett Markov-modellek esetén nem más, mint a felismerési hálózat kezdő és végpontja közti legkisebb valószínűségű útvonal megtalálása.

d) Az órán bemutatott (Viterbi) algoritmusnál a mintaillesztés számításigénye megközelítőleg exponenciálisan függ a felismerési hálózat állapotainak számától.

3.1

a. HAMIS. Nem fonémasorozatot kell előállítania, hanem egy olyan 10-40 dimenziós vektort, melyeknek kicsi az intraindividuális és az interindividuális jellemzője.
b. IGAZ.
c. HAMIS. Semmi köze a prozódiához, a beszéd kisebb egységeinek kezelésében segíti munkánkat.
d. HAMIS. No comment 🙂

3.2

a. IGAZ. Kicsit furán van megfogalmazva, de szerintem jó.
b. IGAZ.
c. HAMIS. A mintaillesztés egyik lényege hogy a különböző ritmusú ejtések között is tudjon mintailleszteni.
d. HAMIS. Igaz csak sablonalapú és a legegyszerűbb fajta, de mintaillesztési eljárás.

3.3

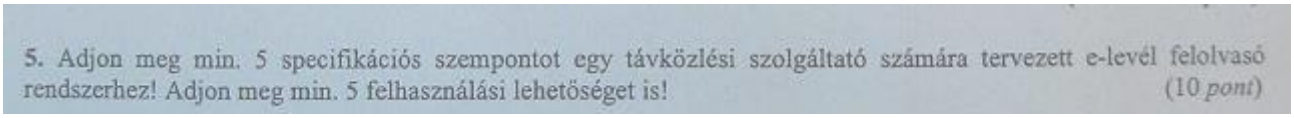
a. IGAZ.

b. HAMIS. Valószínűségekkel dolgozik a HMM, így teljes biztonsággal sosem tudja megmondani, hogy egy megfigyelés adott állapothoz tartozik vagy éppen nem tartozik.
c. HAMIS. *Legnagyobb* valószínűségi útvonalat keres.
d. HAMIS. Mert lineárisan, lásd dinamikus programozás.

4. feladat

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.
d. Mi a screen reader és a TTS kapcsolata?
    - A screen reader csak egy illesztő alkalmazás a képernyő és a TTS között, a képernyőn található információt adja át felolvasásra a TTS számára.

5. feladat


5. Adjon meg min. 5 specifikációs szempontot egy távközlési szolgáltató számára tervezett e-levél felolvasó rendszerhez! Adjon meg min. 5 felhasználási lehetőséget is! (10 pont)

5 specifikációs szempont:

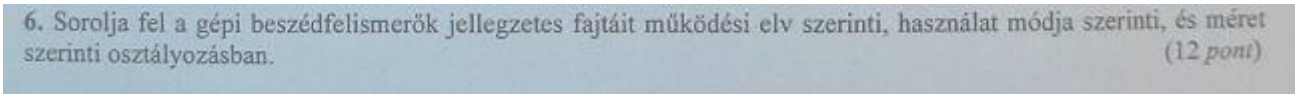
- Nyelv
- Operációs rendszer
- Beszéd minősége : érthetőség, természetesség
- Milyen hangokon szólaljon meg (ffi/női)

- 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. feladat


6. Sorolja fel a gépi beszédfelismerök jellegzetes fajtáit müködési elv szerinti, használat módja szerinti, és méret szerinti osztályozásban. (12 pont)

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

B csoport

csak az eltérő kérdésekre kitérve:

3. feladat

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. feladat

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.

b. lásd A csoport, 4/d.

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: $beszedhangok \cdot maganhangzok \cdot beszedhangok$, ennél némileg kevesebb mivel nem fordul elő minden hármas + a szükséges diádok: $beszedhangok \cdot beszedhangok$ (szerintem a tisztán triádos adatbázis egyszerűen a fonémák köbével arányos. Az már a kevert adatbázis ahol diádok is vannak. Vagy?)

-- Gabo - 2008.05.28.

-- Csapszi - 2008.05.28.

-- Maco - 2010.01.06.