Beszédinformációs rendszerek
6. gyakorlat

Beszédszintetizátorok a gyakorlatban és adatbázisaik

könyv 8. és 10. fejezet

Olaszy Gábor, Németh Géza, Zainkó Csaba

olaszy,nemeth,zainko@tmit.bme.hu

2018. őszi félév

Példamegoldási stratégia

- 1. Értelmezni kell a példa szövegét.
- 2. Fel kell mérni, hogy milyen tanult adatok, módszerek, fogalmak ismeretére van szükség, hogy a megoldás sikeres legyen.
- 3. Át kell gondolni a 2. pont elemei közötti összefüggéseket, rendszerjellemzőket
- 4. Mindezek alapján kell a válaszokat megszerkeszteni.

1. példa

Egy formáns TTS szintetizátor tervezéséhez kell az alapvető adatokra javaslatot tennie.

a) Adja meg, hogy milyen vezérlő paraméterekkel működjön!

zönge, zörej, F0, F1, F2, F3, hangerő

b) Adja meg a javasolt paraméterek fizikai érték tartományát!

Zönge (F0): 50-500 Hz

F1:200-1000 Hz, F2: 500-2000Hz, F3:2000-4000 Hz

Hangerő: 0-tól hangosig pl. 32 lépésben

c) Milyen sűrűn kell a vezérlő paramétereket frissíteni a beszéd gépi előállításánál?

A beszéd spektrális tartalma milyen sűrűn változhat? Minden zöngés periódusban. Tehát 5-10 ms-onként kell frissíteni.

2. példa

Diádos elemekből álló beszédelembázist készítünk gépi szövegfelolvasóhoz.

a) Mi a diád?

Két fél beszédhangnyi hullámforma rész

a) Adja meg a *bab* szó gépi előállításához szükséges diádokat betűjelekkel!

#b ba ab b#, tehát 4 db diád

b) A *mocsár* szóból diádokat készítünk. Hol célszerű elvágni a zár-rés hangot, hogy a lehető legkevesebb legyen a torzítás a későbbi elemösszefűzésnél?

A zárszakaszban, hiszen ott nulla az amplitúdó

c) A magyar nyelvre kb. hány diádot kell elkészíteni, hogy tetszőleges szöveget lehessen felolvastatni egy beszédszintetizátorral?

40 beszédhanggal tervezve 40x40=1600 db diád (a hosszú mássalhangzók időtartamát nyújtással valósítjuk meg)

3. példa

Diádos és triádos elemekből álló beszédelembázist készítünk hullámforma összefűzéses gépi szövegfelolvasóhoz.

a) Adja meg a *babáknak* szó felépítéséhez felhasználható CVC triádokat betűjelekkel!

bab bák nak

a) Hol célszerű vágást alkalmazni, ha triádos elemeket készítünk?

A mássalhangzó közepén (kivételek vannak)

a) A magyar nyelvre optimálisan kb. hány diádot és kiegészítő triádot kell elkészíteni, hogy tetszőleges szöveget lehessen felolvastatni a hullámforma összefűzéses beszédszintetizátorral.

1600 diád

25 C és 14 V esetén 25x25x14= durván 7000 triád

4. példa

Mit generálunk elsősorban a WaveNet-tel?

- Hangosságértékeket
- Ultrahangos felvételeket
- Hullámformát

  Gépi tanuláson alapul, mély neurális hálózatot használ.
- Beszédfelismerő nyelvtant
- Érzelmi címkéket
- Kivétel-szótárakat

5. példa

A beszédválaszú rendszerekben alkalmazott gépi beszédkeltő módszerekről tanultak alapján válaszoljon a következő kérdésre

a) Milyen TTS módszereket alkalmazna egy vakok számára készült képernyő felolvasó rendszerben? Adja meg a működésük lényegét és erőforrás igényüket!

amiben jól gyorsítható a beszéd, és a hangmagasság paraméterrel állítható

Példa alkalmazás: ROBOBRAILLE.org

fájl konverter

6. példa

Egy kis vállalat részére tervezzen kötött szótáras név szerinti tudakozó beszédválaszú rendszert (min. MOS 4,3 az elvárás).

Az előfizető telefonszámát kell gépi hangon elmondani.

- bemenet: 200 vezeték- és keresztnév (beszédfelismerő megvan)

- kimenet minden névhez a megadott 11 jegyű tagolt mobil telefonszám

A cég által megadott női bemondó hangján kell szólnia.

Tegyen javaslatot az elvégzendő munkafázisokra!

Az elhangzó hangüzenetek megtervezése, tipizálása.

A telefonszám bemondás elemeinek tervezése Pl. 06 30 34 76 34 4 (3 féle csoport).

Vivőmondatok megtervezése (kb. 100 vivő mondat).

Hangfelvétel, vágás.

Hangelembázis tervezése, programozása,elkészítése.

A rendszer vezérlő (elemösszefűző) programjának elkészítése.

6. Példa folytatás

Egy kis vállalat részére tervezzen kötött szótáras név szerinti tudakozó beszédválaszú rendszert (min. MOS 4,3 az elvárás).

Az előfizető telefonszámát kell gépi hangon elmondani.

- bemenet: 200 vezeték- és keresztnév

- kimenet minden névhez a megadott 11 jegyű tagolt mobil telefonszám

A cég által megadott női bemondó hangján kell szólnia.

Tegyen javaslatot az elkészítés időtartamára!

1-2 hónap

a) Milyen vivőmondatokat alkalmazna?

A keresett személy telefonszáma: 06-30-……

b) Milyen mestermondatot javasolna a fejlesztéshez?

NEM KELL MESTERMONDAT, MERT EGY NAP ALATT FELVEHETŐ A HANGANYAG

c) Adja meg, hogy milyen eszköz és szakember igényre lenne szüksége a munka sikeres elvégzéséhez!

Hangstúdió, bemondó, hangeditáló-vágó, programozó, integrátor

7. példa

a) Melyik mai módszerhez áll legközelebb Kempelen Farkas beszédkeltő gépe?

Ez a gép az artikulációt utánozza. Ehhez legközelebb a formáns szintézis áll

a) Mikor és ki adta be a világ első szabadalmát tetszőleges szöveg reprodukálására alkalmas beszélőgépre? Melyik mai módszerre hasonlít?

Bánó Miklós magyar mérnök 1916-ban

A diád építőelemes beszédszintetizáló rendszerre hasonlít (Profivox BME TMIT)

8. példa

A Budapest Népliget autóbusz-pályaudvarra tervezzen magyar nyelvű hangos utastájékoztató rendszert. Kellemes női hang a követelmény.

a) Milyen gépi beszédelőállítási technológiát alkalmazna ebben a rendszerben?

Korpusz alapú elem összefűzés

b) Mik az előnyei?

Szép hangon szólal meg, természetes hangzású.

c) Mik a hátrányai?

Nagy szakértelmet kíván mind az elkészítése, mind a működtetése.

d) Mi a mestermondat szerepe a rendszer bővítésénél?

Biztosítja a hangszínezet megtartását új hangfelvételnél.

e) Hány óra beszédet célszerű optimálisan rögzíteni egy ilyen rendszerhez?

Sok órányi beszéd. Tervezett szövegbázist kell felolvastatni.

9. példa

Egy meglévő városi navigációs rendszert fejlesztenek tovább felhő alapú beszédszintetizátor alkalmazásával Budapest területére.

Feladat: Az utcaneveket kell felolvasnia a helyesírással megadott formájú adatbázisból a felhasználó gombnyomásos kérésére a GPS koordináták alapján egy jó minőségű gépi hanggal. A korábbi rendszerből csak 20 percnyi hanganyag áll rendelkezésre, az eredeti bemondó már nem elérhető. A továbbfejlesztett rendszernek azonban hasonló hangúnak kell lennie a korábban elkészített promtokhoz.

a) Milyen gépi beszédelőállítási technológiát javasolna?

Beszédhang adaptálásra is szükség van, ezért HMM alapú adaptálásos TTS.

a) Adja meg a rendszer elkészítésének lépéseit!

HMM alapú TTS rendszer elkészítése, majd adaptálás a 20 percnyi beszéd alapján.

a) Adja meg az alkalmazott technológia előnyeit és hátrányait!

A HMM alapú minden hangkapcsolódást jó akusztikai formában elő tud állítani.

Hátrány: nagy tudást igényel a fejlesztés és csak team munkában lehet eredményes

10. példa

Egy magyar nyelvű időjárás jelentést felolvasó alkalmazáshoz korpuszos beszédszintetizátor beszédadatbázisát kell elkészítenie.

a) Hogyan készítené el az alkalmazás alapját képező szövegkönyvet?

Minden évszakban gyűjtenék szövegeket, sok szöveget.

a) Milyen gépi feldolgozás(oka)t alkalmazna és milyen sorrendben, miután a kiválasztott bemondó felolvasta a téma lefedését biztosító szövegkönyvet?

fonetikai átirat készítése, hanghatárok bejelölése, mondat szintű beszédadatbázis elkészítése, a vezérlő, válogató program elkészítése, tesztelése

a) Adja meg a beszédadatbázis elkészítési műveletsorának lépéseit!

Szöveg gyűjtés, hangfelvétel (Mester mondattal), mondatokra vágás, fonetikai átirat és hanghatár bejelölés minden mondatra.

a) Milyen prozódiai modellt használna ebben a rendszerben?

Helyzet alapút a mondaton belül (kezdő, belső, mondatvégi elemek szerinti válogatás címkézés)