

Magyar olvasott diád hangsorok adatbázisa TTS fejlesztéshez

diads-hu

A magyar olvasott hangsorok adatbázisa olyan hangsorokat tartalmaz, amelyekből diádni hullámforma részek vághatók ki és használhatók fel gépi szövegfelolvasó szoftver hangadatbázisának elkészítéséhez. Ez az adatbázis tartalmazza az alább megadott hangok összes kapcsolódását.

Megvalósítás

A hangsorok összeállítása, feldolgozása, bemutatásra való előkészítése: Bartalis Mátyás, Csapó Tamás Gábor, Németh Géza, Olasz Gábor, Tóth Bálint Pál, Zainkó Csaba (BME Távközlési és Médiainformatikai Tanszék).

Az adatbázis felépítése

Az adatbázis felolvasott rövid hangsorokat tartalmaz strukturált formában. Cél, hogy a magyar beszéd minden hangkapcsolatát lefedje. Minden hangkapcsolatot egy adott hangsor képvisel (például: *abáka=bá* hangkapcsolat, *kaáka=aá* magánhangzó-kapcsolat). Ebből a hangorból kell a hangkapcsolat hullámformáját kivágni és adatbázisba rendezni, hogy a későbbi beszédszintézis során a megfelelő hullámforma elemeket összefűzve a kívánt beszéd hullámformája előálljon.

Elemtípusok. A CV, VC, VV hangkapcsolódásokra három szótagból álló logatomokat, tartalmaz az adatbázis 23 mássalhangzóra és 14féle magánhangzóra. A CC kapcsolatokra, rövid szavakat tartalmaz az adatbázis (például: *labda=bd* mássalhangzó-kapcsolat).

Bemondó neve: nő

Elemszám: 37 beszédhangra és a szünet elemre összesen 1455 db hangsor

Annotálás: Az adatbázis annotálása során kézi feldolgozással meghatároztuk a hanghatárokat

Könyvtár és fájlstruktúra

A főkönyvtár neve: *diphone_hu*, mely 8 db alkönyvtárt tartalmaz.

Az alkönyvtárak tartalmazzák az adatbázis elemeket különböző fájlformátumokban (ld. alább). Az egyes elemek fájlnevei ugyanazok, csak a kiterjesztés változik. Például: *diphone_female_01_000.wav*, *diphone_female_01_000.TextGrid*.

File típusok

Az adatbázisban minden wav fájlhoz tartozik egy TextGrid kiterjesztésű szöveges adatfájl, valamint a felolvasott hangsor szövegfájlja. A fájlnevek egy-egy csoportban azonosak

1. Hangfájl: *xx.wav* (a hang meghallgatható hullámformája),
2. A felolvasott szó TextGrid jellegű feldolgozása a hanghatárok és a hangok időbeni meghatározásához: *xx.TextGrid*
3. A felolvasott szó ortografikus formája: *xx.txt*

A beszédhangok jelölése

A TMIT egyedi hangjelölését alkalmaztuk (számítógépes karakterek). Ezek a karakterek minden esetben

hangot jelentenek, nem betűt. A hangkódok többnyire megegyeznek a hang betűkódjával.

Magyar magánhangzók (betű= hangjelölés): a=a, á=a1, e=e, é=e1, i=i, í=i1, o=o, ó=o1, ö=o2, ő=o3, u=u, ú=u1, ü=u2, ű=u3

Magyar mássalhangzók (betű= hangjelölés): b=b, d=d, g=g, gy=gy, p=p, t=t, k=k, ty=ty, m=m, n=n, ny=ny, j=j, h=h, v=v, f=f, sz=sz, z=z, zs=zs, s=s, c=c, cs=cs, l=l, r=r. A hosszú mássalhangzókat „:” jel hozzáadásával adjuk meg (betű= hangjelölés: ssz=sz:, rr= r:)