

Párhuzamos, precíziós magyar olvasott beszédadatbázis kollekción ParallelSpeech-hu

A párhuzamos, precíziós magyar olvasott beszédadatbázis kollekción létrehozásakor 10 beszélő ugyanazt a szöveganyagot olvasta fel. A hangfelvétel precíziós annotáláson esett át, amely gépi és kézi elemekből állt. Ezzel egy megbízható adatokkal rendelkező, 99,9%-os pontosságú beszédadatbázis kollekción áll a kutatók rendelkezésére.

Megvalósítás

Az adatbázis felvétele, feldolgozása, bemutatásra való előkészítése 2009 óta folyik. Bartalis Mátyás, Böhm Tamás, Csapó Tamás Gábor, Laczkó Klára, Németh Géza, Olaszgy Gábor, Tóth Bálint Pál, Zainkó Csaba (BME Távközlési és Médiainformatikai Tanszék).

Az adatbázis felépítése

Az adatbázis beszédanyaga fonetikailag kiegyenlített mintegy 2000 mondatból álló szöveg (Vicsi-Vígh 1998 „Beszédkutatás” pp. 163-177.) felolvasásából jött létre. Tíz beszélő olvasta fel ugyanazt a közel 2000 mondatot. Minden mondatához egy hangfájl, egy txt és egy textgrid kiterjesztésű fájl tartozik.

Bemondók: 5 nő és 5 férfi

Annotálás: Az adatbázis precíziós annotáláson esett át, ami gépi és kézi feldolgozásból állt. A kézi javítás során minden beszédhang vizuális ellenőrzésen esett át, a hanghatárok pontossága 10 ms-on belül van. A fonetikai átírat pontosan megfelel a hanghullámnak.

Feldolgozott beszédhangok száma: 831941

A feldolgozott beszéd időtartama: 25,5 óra

Könyvtár és fájlstruktúra

A főkönyvtár neve: parallelspeech_hu, mely 10 db alkönyvtárt tartalmaz, minden alkönyvtáron belül 56 belső könyvtár található, ezekben vannak a feldolgozott mondatok, minden ilyen belső könyvtárban 36 db mondat. Az azonos mondatához tartozó fájlok nevei megegyeznek.

File típusok

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

1. Hangfájl: xx.wav (a hang meghallgatható hullámformája),
2. A felolvasott mondat szövegfájl alapú (TextGrid) feldolgozása a hanghatárok (és, szóhatárok) időpontjának adataival, valamint a fonetikai átírat hangjelöléseivel, továbbá a beszédhez nem tartozó szakaszok megjelölésével: xx.TextGrid
3. A felolvasott mondat ortografikus formája: xx.txt

A beszédhangok jelölése

A TMIT egyedi hangjelölését alkalmazzuk (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, ú=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, dz=dz, dzs=dzs, c=c, cs=cs, l=l, r=r. A hosszú mássalhangzókat : jel hozzáadásával adjuk meg (betű= hangjelölés: ssz=sz:, rr= r:)