

Precisely annotated, parallel read speech database in Hungarian

ParallelSpeech-hu

The goal of the precisely annotated, parallel speech database was to construct a precisely annotated and labelled database collection for speech research and development.

Development

The construction of the database started in 2009. The recordings, pre-processing, annotation and labelling were performed by Mátyás Bartalis, Tamás Bóhm, Tamás Gábor Csapó, Klára Laczkó, Géza Németh, Gábor Olszy, Bálint Pál Tóth and Csaba Zainkó at the Budapest University of Technology and Economics, Hungary, Department of Telecommunications and Media Informatics.

The content of the database

The database contains read speech. Ten speakers read the same sentence corpus (2000 items). The sentence set is phonetically balanced (Vicsi-Vígh 1998. „Beszédkutatás” pp. 163-177.). Three file types (waveform, text and labels) were defined for each sentence.

Voices: 5 female and 5 male speakers

Annotation: the annotation was carried out by machine(ASR) and corrected by visual control. Each sound of the database was examined, so the sound boundaries are precisely determined within 10 ms range. The phonetic annotation contains the speech sounds of the waveform and also markers (start_sil, sil, end_sil) to show the non-speech parts as well.

The number of speech sounds in the database: 831941

The speech time: 25,5 hours

Directory and file structure

Main directory name: parallelspeech_hu (contains 10 subdirectories, one subdirectory for each speaker). Each subdirectory contains the voice of one speaker i.e. 36 sentences in each of 56 sub-sub directories.

File types

Three file types are presented in the speech database,

the waveform (xx.wav),

the orthographic text of the word with Hungarian letters in ISO-8859-2 (xx.txt),

the Praat textgrid file containing the sound boundaries and the sound symbols (xx.TextGrid)

Sound symbols

Special sound symbols are used in the phonemic description.

Hungarian vowels (letter= sound symbol): 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

Hungarian consonants (letter= sound symbol): 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. The phonological short/long contrast by the consonants is represented by the „:” character for long consonants (letter= sound symbol; for example: ssz=sz:, rr= r:)

The sound symbol with a „)” sign means that the sound is characteristic only in that place of the wave form (e.g. the sound is glottalised, or extremely short etc.).

Other symbols:

start_sil= the end of start_sil is the position of first speech sound of the sentence,

sil= represents non-speech acoustic content (silence, laugh, cough, etc.),

end_sil= the beginning of the end_sil is the end of the last speech sound of the sentence