

Kurzusleírás távoktatáshoz

Oktató: Szekrényes István

Oktató elérhetősége (egyetemi email): szekrenyes.istvan@arts.unideb.hu

Kurzus kódja, címe: BTAL3013-K2, Számítógépes beszédfeldolgozás

Kurzus órarendi ideje: Kedd 18:00

1. A kurzus rövid leírása

A kurzus célja, hogy a hallgatók gyakorlati betekintést nyerjenek az számítógépes nyelvészet egy speciális, interdiszciplináris területébe. A beszédakusztika alapjainak (tiszta és zenei hangok elemzése, Fourier transzformáció, spektrografikus ábrázolás, gerjesztett-szűrő modell, formánsok, szupraszegmentális jellemzők) és a feldolgozáshoz szükséges számítógépes eljárások megismerése után néhány konkrét alkalmazási lehetőséget (annotáció, szünet és beszéd detektálása, beszélődetektálás, beszédfelismerés) is sorra veszünk a beszédtechnológia, a korpusznyelvészet és a bűnügyi nyelvészet területéről.

2. Szakirodalom

Németh Géza – Olaszgy Gábor (szerk.): A MAGYAR BESZÉD. Bp. Akadémiai Kiadó, 2010.

<http://smartlab.tmit.bme.hu/downloads/pdf/A-magyar-beszed.pdf>

Beke András: Gépi beszélődetektálás magyar nyelvű spontán társalgásokban. Bp. ELTE, 2015.

http://www.eltereader.hu/media/2017/06/Beke_READER.pdf

Pápay Kinga: Valószínűségi skálák az igazságügyi nyelvészetben. In Várad Tamás (szerk.)

I. Alkalmazott Nyelvészeti Doktorandusz Konferencia. Budapest: MTA Nyelvtudományi Intézet. 102-113. <http://www.nytud.hu/alknyelvdok07/proceedings07/Papay.pdf>

.

3. A gyakorlati munkához használható számítógépes eszközök

3.1. Praat (elsődleges)

<http://www.fon.hum.uva.nl/praat/>

Magyar nyelvű leírás: <https://altnyelv.unideb.hu/compling/praat/praat-tutorial.html>

3.2. Python (választható)

<https://www.python.org/>

<http://people.csail.mit.edu/hubert/pyaudio/docs/>

3.2. Kaldi (választható)

<https://kaldi-asr.org/doc/>

<https://www.eleanorchodroff.com/tutorial/kaldi/introduction.html>

<http://jrmeyer.github.io/>

3. A kurzus teljesítésének feltételei

Az adott hétre kiadott gyakorlati feladatok és ellenőrző kérdések határidőre történő megoldása, megválaszolása. A határidő után történő beküldés az adott hétről való hiányzásnak minősül. A határidő minden héten: péntek éjféli. A megoldások beküldése egyelőre e-mailben történik, de későbbiekben lehet áttérünk egy felöltő-rendszer használatára. Az e-mail tárgyában kérem, mindenki jelölje meg a kurzus nevét és az adott hét sorszámát (pl. „Számítógépes beszédfeldolgozás, 8. hét”). Az ellenőrző kérdésekre saját szavaival válaszoljon! A 8. héttől elérhetővé válnak olyan bónusz feladatok is, amelyek megoldása nem kötelező, de ha valaki teljesíti, akkor azzal kiváltja a következő hét összes feladatát.

4. Technikai információk

A gyakorlati feladatok megoldásához szükséges anyagok egy jelszóval védett tárhelyen lesznek elérhetők. A letöltéshez szükséges adatokat (felhasználónév, jelszó) az oktató neptunüzenetben küldi el a hallgatónak. A feladatleírásban hivatkozott fájlok minden héten ezen a címen érhetők el:

<https://altnyelv.unideb.hu/szekrenyes/e-learning-2020/Szamitogepes-beszedfeldolgozas/data/>

5. A kurzus részletes tematikája és feladatai

7. hét (március 23 – 27)

Téma: A kurzus első óráin a beszédakusztika alapjaival és a beszédaktivitás detektálásának témakörével foglalkoztunk. Ezekkel a témákkal kapcsolatban a 7. héten 3 ellenőrző kérdésre kell válaszolni és két gyakorlati feladatot megoldani.

Kapcsolódó szakirodalom:

<http://smartlab.tmit.bme.hu/downloads/pdf/A-magyar-beszed.pdf>: 3.3 fejezet

<https://altnyelv.unideb.hu/compling/praat/praat-tutorial.html#picturew>

Ellenőrző kérdések:

1. Mi a Fourier transzformáció kimenete?
2. Mi jellemző a periodikus jelek frekvenciaspektrumára?
3. Mely beszédhangok jönnek létre az ún. lökéshullámszerű gerjesztéssel?

Gyakorlati feladatok:

1. Olvassa be a Praat programba a *piano_A4.wav* és a *noise.wav* hangfájlokat. Hajtson végre rajtuk Fourier transzformációt („To Spectrum...”). Az eredményül kapott „Spectrum” objektumokat a „Draw...” paranccsal rajzoltassa ki a Praat Picture ablakban (a kettő között a rajzoló palettát az *Edit→Erase all* paranccsal tudja üríteni), majd mentse el őket PNG képfájlként. Írjon rövid szöveges elemzést a két frekvenciaspektrum közötti különbségekről!

2. A *persian1.wav* nevű hangfájlt olvassa be a Praat programmal, majd próba-hiba alapon kísérletezve találja meg, hogy milyen beállításokkal produkálja a legoptimálisabb eredményt a program szünetdetektálásra szolgáló eljárása (*To TextGrid (silences)*..). E-mailben küldje el az eredményül kapott TextGrid fájlt („Save as text file..”) és a felhasznált paraméterek értékét.

8. hét (március 30 – április 3.)

Téma: Ezen a héten a hangfájlok manuális annotációját, átíratozását gyakoroljuk, amelyhez itt is a Praat program szolgál munkaeszközként. Az annotáció célja, hogy a beszélők egyes megnyilatkozásai írásban és időben is rögzítve legyenek a hangfelvétel időtengelyén. Az eredményül kapott annotáció (TextGrid objektum - lemezre mentve egyszerű szövegfájl) a filmfeliratoknál is használt SRT fájlok szerkezetéhez hasonló (mettől - meddig - mi hangzik el).

Kapcsolódó segédanyagok:

<https://altnyelv.unideb.hu/compling/praat/praat-tutorial.html#anno>

<https://www.youtube.com/watch?v=64cclyKVJZ4>

Gyakorlati feladat:

A Praat programmal olvassa be a *talking.wav* nevű hangfájlt. A hangfájlhoz készítsen annotációt a „To TextGrid...” parancs használatával az alábbi beállításokkal:

All tier names: speakerA speakerB ←--- két annotációs szintet hozunk létre a két beszélőhöz

which of these are point tiers: ←--- ez maradjon üresen

Jelölje ki a hangfájlt és a TextGrid objektumot, majd nyissa meg a szerkesztőablakot („View & Edit”). A első annotációs szinten („speakerA”) jelölje be („Boundary→Add on tier 1”) a férfi beszélő megszólalásainak elejét és végét (szünettől szünetig vagy beszélőváltásig), az így kapott intervallumokban szövegesen jegyezze le, amit a beszélő mond.

A második annotációs szinten („speakerB”) a feladat ugyanez a női beszélőre vonatkozóan.

Az munka befejezése után mentse el az eredményül a TextGrid fájlt, majd e-mailben küldje el az alábbi kérdésre is válaszolva:

Vajon miért, milyen beszédjelenség kezelése érdekében volt célszerű két külön annotációs szinten átíratozni a két beszélő beszédét?

Bónusz feladat:

Írjon Praat szkriptet, ami az eredményül kapott annotáció alapján kiszámolja, hogy az egyes beszélők hány másodpercet beszélnek összesen a felvételen. A szkript forráskódját és a kapott eredményt e-mailben küldje el!

9. hét (április 6 – 10.)

Téma: A kényszerített illesztés („Force alignment”) célja, hogy a hangfájlokhoz készített átíratokat egy automatikus eljárással kisebb, időben is illesztett beszédegységekre (szavakra, beszédhangokra) szegmentáljuk.

Kapcsolódó szakirodalom:

<http://www.inf.u-szeged.hu/~kocsor/publications/Papers/2004/Conf-2004-MSZNY-TL/Web/ZTK0>

Gyakorlati feladat:

Készítsen szöveges átíratot a *sample.wav* hangfájlhoz. Az átírat készítéséhez most ne a Praat programot, hanem valami egyszerű szerkesztőprogramot (pl. Notepad) használjon! A szövegfájlt mentse el *sample.txt* néven.

A *sample.wav* és a *sample.txt* fájlok feltöltése (Drag & drop technikával kell behúzni őket a kijelölt mezőbe) és a magyar nyelv kiválasztása után indítsa el („Run Webservice”) a CLARIN D projekt *WebMAUS Basic* szolgáltatását:

<https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSBasic>

A kimenetként kapott TextGrid fájlt töltsse le, majd a *speech.wav* hangfájllal együtt olvassa be a Praat programba. Az objektumokat jelölje ki, nyissa meg a szerkesztőablakot. Ellenőrizze, szükség esetén a határjelölő pálcák csúsztatásával (rákattintás után az bal egérgombot lenyomva tartva lehet mozgatni őket) javítsa a szavak határainak jelölését. A javított TextGrid fájlt e-mailben küldje el!

Bónusz feladat:

Írjon Praat szkriptet, ami másodpercben kilistázza az egyes szavak időbeli hosszát! A forráskódot e-mailben küldje el!

10 - 11. hét (április 13 - 24.)

Téma: A beszélődetektálás vagy beszélő diarizálás („speaker diarization”) alatt egy olyan nyelvfüggetlen eljárást értünk, ami egy több beszélő hangját is tartalmazó hangfelvételt kap bemenetként, a kimenet pedig azt mutatja meg, hogy a különböző beszélők a felvétel mely részein szólnak meg.

Kapcsolódó szakirodalom:

http://www.eltereader.hu/media/2017/06/Beke_READER.pdf: Bevezetés

Ellenőrző kérdések:

1. Milyen fő lépésekre tagolhatók fel a beszélő detektáló/dirazáló eljárások?
2. Milyen akusztikai jellemzőket használhatunk fel a beszélőváltás detektálásához?
3. Miben különböznek a „fentről lefelé lebontó” és az „alulról felfelé” egyesítő klaszterezési technikák?

Gyakorlati feladat:

A *persian.wav* nevű hangfájl feltöltése és a megfelelő beállítások (Output format: TextGrid, sampling rate: 44100) tesztelje és röviden értékelje a CLARIN D projekt beszélődiarizáló szolgáltatását:

<https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/SpeakDiar>

Bónusz feladat: Ugyanezen a hangfájlon (*persian.wav*) tesztelje a **VBDIAR** python vagy egy tetszőleges választott másik diarizáló alkalmazás kimenetét. A kimenet ellenőrzéséhez készítsen referenciaanyagot RTTM formátumban és mérje le a DER (Diarization Error Rate) hibaarány értékét az alábbi szkript használatával:

<https://github.com/nryant/dscore>

12. hét (április 27 – május 1.)

Téma: A kriminalisztikai célú beszélőazonosítás

Kapcsolódó szakirodalom:

<http://www.nytud.hu/alknyelvdok07/proceedings07/Papay.pdf>

Ellenőrző kérdések:

1. Mi a különbség a szakértő által hozott lehetőségi, valószínűségi és a kategorikus ítéletek között?
2. Rövidesen ismertesse azt a logikai problémát, amire a szakértői vizsgálat kiértékelésekor tekintettel kell lennünk.
3. Mi az LR-teszt lényege és mit mutat meg az eredménye?

Gyakorlati feladat:

Vizsgálja meg, hogy a **krimi** mappában található hanganyagokból mért alapfrekvencia értékek átlaga segíthet-e eldönteni, hogy három gyanúsított közül melyiktől származik a bizonyítékként használt hangfelvétel.

Bónusz feladat:

Ugyanezen az anyagon, az alapfrekvencia értékek alapján implementálja az LR-tesztet is! Háttérmodellként használja a MUSAN korpusz beszédet tartalmazó részét (vagy egy másik tetszőleges, ingyenesen hozzáférhető beszédadatbázist):

<https://www.openslr.org/17/>

Az adatok mérésére és kiértékelésére itt célszerű egy Praat szkriptet készíteni. A kimenetben mind a három gyanúsítottra számolja ki az LR-teszt értékét.

----- a végzős hallgatóknak itt ér véget a félév -----

13. hét (május 4 – 8.)

Téma: Beszédfelismerő rendszerek.

Kapcsolódó szakirodalom:

<http://smartlab.tmit.bme.hu/downloads/pdf/A-magyar-beszed.pdf> (333 – 375)

Ellenőrző kérdések:

1. Mi a különbség az izolált szavas, a kapcsoltzavas és a folyamatos beszédfelismerő rendszerek között?
2. Mi az a három beszéddel kapcsolatos fő probléma, ami a beszédfelismerő rendszereknek kihívást jelent?
3. Milyen metrikával szoktuk megadni egy beszédfelismerő rendszer pontosságát?
4. Mi az a beszélőadaptáció?

5. Mik azok a statisztikai N-gramm modellek?

Bónusz feladat:

Készítsen 10 hangfelvételt, amelyeken 5x kimondja azt, hogy „igen”, majd 5x azt, hogy „nem”!

Az alábbi példák és leírások alapján a hangfelvételek mellé készítse el azokat a szövegfájlokat is (wav.scp, text, utt2spk, segments, lexicon.txt, nonsilence_phones.txt, optional_silence.txt, silence_phones.txt), amivel a Kaldi beszéd felismerő rendszert taníthatjuk be a magyar „igen” és „nem” szavak megkülönböztetésére:

<https://github.com/kaldi-asr/kaldi/tree/master/egs/yesno>

https://kaldi-asr.org/doc/kaldi_for_dummies.html

<https://www.eleanorchodroff.com/tutorial/kaldi/overview.html>

Ha Unix rendszeren dolgozik, akkor az első példaszkriptet írja át úgy, hogy a saját tanítóanyagával is működjön (ez a rész opcionális). Ha ki is akarja próbálni, ügyeljen rá, hogy az mfcc.conf fájlban a „sample-frequency” kapcsoló a felvételek tényleges mintavételezési frekvenciájához legyen igazítva.

14. hét (május 11 - 15.)

Téma: a beszéd dallamának modellezése és lejegyzése.

Gyakorlati feladat:

Értékelje az alapprofrekvencia értékek alapján generált, a beszéd dallamát reprezentáló vizualizációk pontosságát (elég az 1. forduló):

<https://altnyelv.unideb.hu/pttest/>

Bónusz feladat:

Ugyanezen a felületen csinálja végig a 2. fordulót is!