3. gyakorlat

Ma megnézzük, hogy hogyan tudunk spektrogramot (és néhány egyéb dolgot) számítani Matlab-bal. Más környezetet is választhattunk volna, pl. python-t, az óra végén ehhez is adok tippet.

1. Matlab hozzáférés legálisan:

Egyetemünk hallgatói és dolgozói számára a Matlab ingyenesen hozzáférhető. Eléréséhez EDUID azonosító kell. Akinek nincs, itt szerezhet (pl. neptun azonosítóval): <https://www.eduid.u-szeged.hu/reg/>

A Matlab elérésének leírása:  
<https://u-szeged.hu/download.php?docID=105051>

Letölthetjük a telepítőt vagy használhatjuk online. Én az online használatot fogom választani, ebben az esetben az első regisztráció után itt tudunk majd újra belépni: <https://www.mathworks.com/academia/tah-portal/university-of-szeged-31495897.htmlA>

A spektrorgmok számításáahoz a Voicebox nevű régi függvénykönyvtárat fogom használni, amit innen lehet letölteni:   
<http://www.inf.u-szeged.hu/~tothl/speech/voicebox.zip>

Csomagljuk kivalahova a fenti zip-et, majd indítsuk el a Matlab-ot. Ha online verziót használunk, akkor indítás után fel kell töltenünk a voicebox fájljait a Matlab drive-jára. Ha offline telepítettük, akkor a kicsomagolás helyét kell beállítanunk munkakönyvtárnak.

1. Hangfájl beolvasása és spektrogram kirajzolása

Hangfájl beolvasása:

[wav,FS,WMODE,FIDX]=readwav('SA1.WAV');

Input: SA1.WAV: a beolvasandó hangfájl

Output: Y: a beolvasott minták

FS: sampling rate

WMODE, FIDX: további beolvasott paraméterek (nem fognak kelleni)

Spektrogram kirajzolása:

spgrambw(wav, FS); (paraméter nélküli eset)

spgrambw(wav, FS, ‘’); ahol az aposztrófok közé írható:

i: invertált szürke skála

j: színes skála

m: logaritmikus (mel) frekvenciaskála

w: a hullámformát is kirajzolja

Természetesen szükség esetén lehet egyéb paramétereket is állítani (pl. ablak mérete, lépésköz)

3. A fül felbontásához és frekvenciaskálájához igazított spektrális elemzés késztése

Ha a spektrogramot az alábbi módon számoljuk, akkor B-ben megkapjuk a spektrális adatokat, T-ben és F-ben pedig a tömb két méretét:

[T,F,B]=spgrambw(wav, FS, ‘’);

Láthatjuk, hogy 1760 spektrális vektort kaptuk, a magasság (spektrális felbontás) pedig 257.

A lineáris frekvenciaskáláról átválthatunk a fülünknek jobban megfelelő mel-skálára az m kapcsolóval, azonban ekkor is 257 spektrális értéket kapunk. Előadáson megbeszéltük, hogy a beszédfelismeréshez nem lesz szükség ilyen részletes felbontásra, elég lenne durvább is, akár csak 20-40 érték. Megtehetnénk, hogy csak minden 4-5-dik értéket tartunk meg, de akkor a minták közé eső információ elveszne. Ezért azt fogjuk csinálni, hogy súlyozott összegzéssel összevonunk szomszédos sávokat. Ha ezt a lineáris frekvenciatengelyű spektrogramon végezzük, akkor magasabb frekvenciákon egyre kevesebb mintát kell vennünk, azaz egyre szélesebb súlyozást fogunk használni. Ezekkel a súlyozásokkal tkp. egy speciális. a szőrsejtek hangológörbéit utánzó szűrősor hatását imitáljuk.

Tehát először csináljuk egy standard spektrogramot:

[T,F,B]=spgrambw(wav, FS, '');

A súlyozások kirajzolása ill. eltárolása (előbbi 20 sávval a láthatóság kedvéért, utóbbi 80 sávra):

melbankm(20,512,FS);

m=melbankm(80,512,FS);

m méretét megnézve láthatjuk, hogy m-be 80 darab 257 méretű vektor került, ezek a 80 sávhoz tartozó 257 darab súlyértékek, mivel a spektrogramunk 257 magas

A súlyozás elvégzése ezután egy mátrixszorzás lesz:

melspectrogram=m\*B';

Kirajzolás (az amplitúdókat logaritmikus skálán véve – spgramw-nél ez a default):

imagesc(log(melspec));

A kép fejjel lefele van, úgyhogy Y tengely megfordítása: set(gca,'YDir','normal')

Megjegyzés: a Matlab beépített melSpectrogram függvénye ugyanezt tudja, csak be akartam mutatni a súlyozást:

melSpectrogram(wav, FS,'NumBands',80)

A gyakorlatban ez a melSpectrogram használható a beszédfelismerő inputjaként – főleg neuronhálós rendszereknél szokás ezt a reprezentációt használni inputnak. Hagyományosan 26-40 magasságú melSpectrogram-ot szoktak használni, bár újabban a 80 is elterjedt.

4. Beszédelemzés (gammatone) szűrősorral:

A spektrogram sávjainak összevonása csak egy gyors közelítése a hallásunk felbontását imitáló szűrők működésének. Profibb (de lassabb) megoldás ez speciális szűrősor létrehozása, majd az eredeti hangfájl megszűrése a szűrősor szűrőivel. Egy ilyen jól megtervezettt szűrősor az ún. gammatone filterbank:

A szűrők együtthatóinak leszámolása (1 ERB szélességű szűrőkkel fedjük le a 100-8000Hz frekvenciatartományt 0.35 ERB lépésközzel):

[b,a,fx,bx,gd]=gammabank(0.35, FS, '', [100 8000]);

a ill. b mérete mutatja, hogy 83 szűrőt kaptunk.

A szűrés végrehajtása:

filterbank(b,a,wav,gd);

5. Mel-spektrogram készítése HTK-val

A következő órától elkezdünk ismerkedni a HTK beszédfelismerővel. A HTK-nak saját modulja van a jellemzőkinyerésre, ezzel egy hangfájlból így nyerhetjük ki a melspec paramétereket:

HCopy.exe –C preprocess.config SA1.wav SA1.htk

Ahol a config paraméterek most úgy vannak beállítva, hogy egy 100 szűrőből álló mel-szűrősor eredményét kapjuk meg (lényegében a fentebb ismertetett háromszöges-súlyozós eljárással).

melspec=readhtk\_lite('SA1.htk');

A toolbox segít megnézni az eredményt (ha online Matlab-ot hesználunk, akkor ne felejtsük el előbb feltölteni SA1.htk-t a felhőbe…):

imagesc(melspec');

set(gca,'YDir','normal')

6. Alapfrekvencia görbe kinyerésére 2 módszer:

fxrapt(wav,FS);

fxpefac(wav,FS,'','g',''); vagy fxpefac(wav,FS,'','G','');

7. Formánskövetés

Mustafa-Bruce formant tracker (nem része a Voicebox-nak):

mb\_ftracker(wav, FS);

Hasonló beszédfeldolgozó eszközök python-hoz:

librosa (for python), pyAudioAnalysis, SurfBoard