

## 1. feladat (v070525 - 6.feladat)

### a, Mi a lényeges különbség a felhasználás szempontjából a beszélő-függő és a beszélő-független beszédfelismerők között?

A beszélőfüggetlen rendszereket bárki, bármikor használhatja előzetes betanítás nélkül, viszont általában kisebb szótárral és megbízhatósággal rendelkeznek. A beszélőfüggő rendszerek általában beszélőadaptívak is egyben, azaz használatukhoz szükséges egy előzetes betanítási fázis, ezután azonban több szót és jobb megbízhatósággal képesek felismerni, izolált szavak helyett akár kapcsoltszavas vagy akár diktáló üzemmódban is.

### b, Betanításnál milyen típusú adatbázist kell az egyik és a másik rendszerhez?

Beszélőfüggetlen rendszer esetén több beszélőtől szükséges hanganyag, hogy ebből közös jellemző vonásokat tudjunk kivonni a betanítás során a minél robusztusabb működéshez. Beszélőfüggő rendszer esetében pedig a hangok paraméterbecslésére nincs szükség (vagy jóval kisebb adatbázis is elegendő), hiszen a betanítási fázis során pont ezeket a paramétereket hangoljuk az adott beszélő alapján. Minden más vonatkozásban (szótár felépítése, nyelvi modellek stb) a két megoldás nem különbözik, illetve max. a szavak számában.

### c, Milyen egyéb szempontokat kell figyelembe venni?

Szótárméret, tematika, a hangkörnyezet (zajos utca v csendes iroda), beszédmodor (spontán vagy dialógusszerű), stb...

## 2. feladat

### a, Gépi beszédminősítő rendszert hogyan ellenőrizné.

Gépi == objektív beszéd minősítés. Hatékonyság mértéke: korreláció MOS-sel vagy minimális négyzetes eltérés MOS-tól. Tehát a gépi minősítés ellenőrzéséhez el kell végezni (gondolom szűrőpróba szerűen) a MOS minősítést is.

### b, VoIP esetén mi okozza a legnagyobb problémát a beszédminőség mérése során?

A válasz: Jitter (késleltetés-ingadozás). (A hálózat paramétereinek nem stabil volta. Teljesen más minőséget kapunk ha kis illetve szélessávon mérünk, illetve változatos kapcsolat (műholdas, kábel, adsl) esetén is jelentős eltéréseket tapasztalhatunk a beszéd minőségében, a hálózatforgalmi szituációkat nem is említve (pl. ha közben töltünk is).) Nem a rizsára voltak kíáncsiak.

## 3. feladat

HMM ábra START, sil, egyik út é, g, vagy a másik út é, s, sil STOP; plusz visszacsatolás az utolsó sil-ből az elsőre. Meg voltak adva az átmeneteke valószínűségei, az msil, mé, mg,...stb értéke és az o1,o2, o3, o4 értéke. Valamint meg volt adva, hogy egyenletes Gauss eloszlású 1 dimenziós és a szórása 2 volt, ha jól emlékszem. Plusz az ábrán a g és sil közt egy "alobha" szó behúzva, másiknál s és sil közt "a" betű. A feladat kb az volt, hogy adj meg a felismerő által felsimert szöveget/eket, vagy szót.

#### 4. feladat

"sás" akusztikai elemzése: Na ez abcd...részekre tagolva...kb ezeket kérték: sás szó felrajzolása hullámformaként dB, idő tengelyen. Külön a msh, mennyi idő a kiejtése, hol képezzük, gerjesztése, képzési mód. Ugyenez a mgh-ra is. CSAK az "s" és "a" hangra konkrétan. Majd ezek spektrumát is fel kellett rajzolni. Á-t 200-1000HZ és dB tengelyen, S-et pedig 200-4000Hz dB tengelyen. Valamint milyen frekvenciát javasolna a mintavételezéshez, hogy visszaállításkor ne vesszen el lényeges komponens. Ja, plusz a sás szóban mennyi a legnagyobb frekis komponens.

#### 5. feladat

Na itt def, ábra, példa kellett:

- **LPC**  
Lineáris predikció. Bizonyos esetekben ha nem a mintát, hanem a minták különbségét kvantáljuk, akkor kevesebb biten tudjuk átvinni ugyanazt az információt (tömörítés).
- **F1**  
Az első formáns. Magánhangzók jellemzéséhez a formánsokat használjuk. *A formánsok a spektrális felbontás burkológörbéjének maximumai.* (A következő definíciót vastag vonallal fogják lesatírozni-> Az alapfrekvencia az F0, ennek egészszámú többszörösei a formánsok. F1 az F0-hoz legközelebb eső formáns.)
- **pitch**  
Az alapfrekvencia, azaz F0. A PSOLA (Pitch Synchronous Overlap Add Method) eljárás az alapfrekvenciát változtatja a beszéd dallamának változtatása érdekében.
- **ergodikus**  
A folyamat tulajdonsága. Ha a folyamat ergodikus, akkor 1 folyamat is magában hordozza a sokaság tulajdonságait.
- **VXML**  
Voice eXtended Markup Language - dialógusok leírására alkalmas
- **son**  
Relatív hangosság. Hányszorosa a hangosság a 40 phon hangnak? Képlettel:  $N = 2^{[(L - 40) / 10]}$  L: phon, hangosságérzet. 40 phon = 1 son, 50 phon = 2 son, 60 phon = 4 son stb.
- **aliasing**  
Ez a jelenség azt eredményezi, hogy a mintavételezett hullámforma torzulni fog visszaállításkor. Pl.: a Nyquist kritérium szerint 10 000 Hz mintavételezési frekvencia mellett az 5000 Hz feletti frekvenciájú hangkomponensek torzulni fognak visszaállításkor. Ezt úgy kezelik, hogy

egy alul áteresztő anti-aliasing filterrel kiszűrjük az 5000 Hz feletti komponenseket.

- **HMM**

Hidden Markov Model - Rejtett Markov Model. Statisztikai alapú beszédfelismerési módszer.

- **triád**

Egy triádos adatbázisban három hosszú hangkombinációkra tároljuk a hullámformákat. A tárolt adatok mennyisége ezért a fonémák köbével arányos. A tárolt hang az első fonéma közepénél kezdődik és a harmadik közepénél fejeződik be.

- **MP3**

MPEG-1 Audio Layer 3. (nem MPEG-3) Veszteséges tömörítés. Az emberi fül számára nem hallható hangokat eldobja.

Részletesebben:

Érzeti (részsávós kódolás) – perceptual (subband) coding

Frekvenciamaszkolási jelenség (elfedés a frekvenciatartományban)

1) minden sávra megállapítjuk, hogy milyen energiájú összetevők vannak a jelben

2) az elfedett összetevők kihagyása

3) a kvantálási zaj növelésének lehetősége: úgy kvantálunk, hogy a kvantálási zaj ne legyen nagyobb, mint az elfedési szint (itt a tömörítési lehetőség)

## 6. feladat

**Kreatív feladat 😊 Automatikus telefon-behajtó rendszer. Ehhez kellett 10 beszédtechnológiai eljárást írni és min 5 felhanszálási területet.**