

# Beszéd PótZH(?) - 2002.05.10.

## 1. feladat

**a) Osztályozza a beszédhangokat a létrehozásukhoz használt gerjesztés szempontjából! Jellemezze a beszédhangokat akusztikai, spektrális-, intenzitás-, ido-) szerkezetük szempontjából! (10 pont)**

Gerjesztés szerint:

- zöngés: az összes magánhangzó, b, d, g, gy, v, j, m, n, ny, l, r
- zörejes: p, t, ty, k, c, cs, f, sz, s, j\*, h
- kevert: dz, dzs, z, zs

Akusztikai szerkezet szerint

- Egyszerű: az összes magánhangzó, v, f, z, sz, zs, s, j, h, m, n, l
- Összetett: b, p, d, t, g, k, gy, ty, c, cs, dz, dzs, ny, r

Specifikus időtartamok szerint:

- Magánhangzók: i,u,ü,o,a,e,ö,é, á (70 és 160ms között rendre)
- Mássalhangzók 40ms: r; 50ms: n,l; 60ms: z, zs, réshangok; 70ms: p, t, k, ty; 80ms: f, sz, s ; 90ms: c, cs

**b) Mi az idoablak szerepe a beszéd szinképi elemzésében? Mi az elonye és mi a hátránya a rövid és a hosszú ablaknak? (8 pont)**

## 2. feladat

**Mi a pszichoakusztikus modellezés célja? (10 pont)**

Pszicho-akusztikus modell:

- 1: Idő-frekvencia leképezés
  - Keretekre vágás: rövid idejű (15-50 ms), átlapolódó (50%) keretek
  - Ablakozás
  - Fourier transzformáció
- 2: Pszicho-akusztikus érzeti modellezés
  - Az emberi hallás modellezésén alapul, célja a hallható különbségek kiemelése, és a nem észlelhetőek elnyomása
  - Monoton legyen a kapcsolat a belső távolság és az MOS között

Pszicho-akusztikus modell elemei:

- 1: Transzformálás az érzeti tartományra 1a:nemlineáris frekvencia skálák (mel, bark...)
- 2: Frekvencia elfedés
  - Közeli frekvenciák esetén az erősebbik elnyomja a gyengébbet
- 3: Időbeli elfedés
  - Egymás utáni rövid impulzusokat egynek hallunk

- Egy erős hang elnyomja a környező gyengébbeket
- 4: Pszicho-akusztikus hangosság
  - Jel energia és hangosság kapcsolata nemlineáris

### 3. feladat

**Egy 8kHz-es mintavételi frekvenciával és az alábbi,  $H(f)$  karakterisztikájú visszaállítóval működő mintavételező rendszer bemenetére a 's', majd a 'f' hangsor kerül egymás után férfi ejtésben, állandó alapfrekvenciával ( $F_0: 125\text{Hz}$ ).  $H(f) = 1$ , ha  $1 < \text{abs}(f) \leq 3.5$ ;  $(4 - \text{abs}(f))/0.5$ , ha  $3.5 < \text{abs}(f) < 4$ ; 0 egyébként (A frekvencia mértékegysége: kHz)**

**a) Megkülönböztethető-e a két visszaállított hangsor hangzása? Miért? (8 pont)**

F: zöngétlen réshang, nincsenek zörejjócok, egyenletes eloszlás a 1000-10000Hz frekvenciatartományban. A környezetében levő magánhangzó formánsaira csak kis mértékben van hatással. S zöngétlen réshang: zörejelemek 1800-6500Hz között, intenzív zörejjóc ált. 2500-3500Hz között. Az 's' hangot követő magánhangzó formánsaiban kismértékű mozgás van jelen az átmeneti fázisban. Metalogika alapján: a /c miatt nem a válasz! Indoklás: mivel mindkettő hang zöngétlen, azaz gerjesztése zörejes (fehérzaj szerű), ezért spektrumukban mindenféle frekvenciakomponens előfordul, és egész magas frekvenciákon is vannak fontos komponensek, ezeket ez a mintavételezés (telefon) nem viszi át, ezért az 'f' és 's' nehezen megkülönböztethető, a kis mintavételezési frekvencia miatt fellép az átlapolódás jelensége is

**b) Mi változik, ha a rendszer bemenetére is egy  $H(f)$  karakterisztikájú szűrő kerül? (7 pont)**

Megszűnik az átlapolódás jelensége, az 's' zörejjóca így könnyebben kivehető és megkülönböztethető az 'f' hang egyenletes frekvenciaeloszlásától. (ebben nem vagyok biztos)

**c) Javasoljon egy olyan mintavételi frekvenciát és összetett simító karakterisztikát, amely a fenti hangsorokat helyesen és elfogadható komplexitással megvalósítva átviszi! (7 pont)**

22kHz mintavételezéssel, és egy darab hasonló szűrővel 1 és 11kHz között (egyenletes meredekségű) a probléma megoldható. Szerintem nem átlapolódásról van szó, hanem Aliasról. A jelenség megszüntetése Anti Aliasing Filterrel lehetséges (egy aluláteresztő szűrő). ha a két fogalom ugyanazt jelentené, akkor én kérek elnézést... Aliasing def: Ez akkor lép fel, ha a mintavevő-tartóra ráengedünk a mintavételi frekvencia felénél nagyobb komponenseket is, amelyek így spektrális átlapolódásba kerülnek a hasznos jel periodikus spektrumával, és megjelennek olyan „ál”

### 4. feladat

**Milyen típusú beszédfelismerot építene egy mobiltelefonba, milyet egy személy- ill. tehergépkocsiba és miért? (15 pont)**

#### **5. feladat**

**Egy szövegfelolvasó (TTS) rendszereket fejlesztő cég legfontosabb alkalmazási területe a telefonos információ szolgáltatás. Ennél elterjedten használt mintavételi frekvenciája a 8kHz. A cég az akusztikus adatbázisait mégis 22kHz-cel veszi fel. Indokolt-e ez, vagy erőforrás pazarlás? (10 pont)**

#### **6. feladat**

**Tervezze meg a legfeljebb 1500 tanulóval rendelkező Birka Iskola telefonos könyvtár elérési rendszerét. A rendszerben a felhasználók az azonosítójuk megadásával érdeklődhetnek a kölcsönzött könyvek adatairól (pl. lejárat dátuma, késedelemért fizetendő összeg, stb.) Az előjegyzett könyvekről is kérhetnek adatokat és az előjegyzett könyv megérkezésekor telefonos értesítéshez telefonszámot adhatnak meg. Ismertesse a rendszer felépítését, az alkalmazandó beszédtechnológiai elemeket és azok főbb specifikációs adatait. Gondolkozzon kreatívan és széles látókörrel! A kérdésekre több jó válaszgyűjtés is adható! Indokolja válaszait (25 pont)**

-- [Swacsa](#) - 2010.01.13.