

2002 január 23-i vizsga megoldása

1. feladat

Gerjesztés szerint lehet:

- 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

[Forrás: CD, 138. oldal]

Specifikus **időtartamok**:

- 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

Intenzitás:

- I_{min} : h
- I_{max} : á,e
- Magánhangzók csökkenő sorrendben: á,e, a, é, ö,o, i, ü, u

[Forrás: CD, 93. oldal]

(Ha valaki tudja, pontosabban mire gondolhattak, az szerkessze át, gondoltam mégsem kérik az összes hang kifejtését egyenként...)

2. feladat

Kötöttszótáras rendszer ésszerű.

1. tematika meghatározása -> számfelolvasás
2. felhasználók osztályozása: felhasználó, laikus
3. üzenetek meghatározása: mivel HIFI minőség kell, így igényesebb rendszereknél ajánlott az elemet megelőző és követő

hangelemeknek megfelelő változatait is letárolni pl gyezer meg nyezer, nem csak ezer (így is max pár100 elem lesz), állandó meg tartalom nincs a specifikáció szerint

4. felolvasandó szöveg megtervezése, vivőmondatok: utóbbiak nincsenek, tehát elég az összes lehetséges hangelemekapcsolódásnak megfelelő számosságú üzenetet kiválasztani
5. bemondó kiválasztása: mindegy, csak ne a palik 😊
6. hangfelvétel, HI-FI minőségben!
7. digitalizálás, ügyelni kell arra hogy a hangminőség ne romoljon, megfelelő bitráta stb. megválasztása
8. adatbázis elkészítése, elemek kivágása
9. próbaüzem, akusztikai csiszolás
10. rendszerintegrálás

3. feladat

F: zöngétlen réshang, nincsenek zörejgócok, egyenletes eloszlás a 1000-10000Hz frekvenciatartományban. A környezetében levő magánhangzó formánsaira csak kis mértékben van hatással.

S zöngétlen réshang: zörejelemek 1800-6500Hz között, intenzív zörejgóc ált. 2500-3500Hz között. Az s hangot követő magánhangzó formánsaiban kismértékű mozgás van jelen az átmeneti fázisban.

- a. 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. Megszűnik az átlapolódás jelensége, az s zörejgóca így könnyebben kivehető és megkülönböztethető az f hang egyenletes frekvenciaeloszlásától. (ebben nem vagyok biztos)
- c. 22khz mintavételezéssel, és egy darab hasonló szűrővel 1 és 11khz között (egyenletes meredekségű) a probléma megoldható.

Szerintem nem átlapolódásról van szó, hanem Aliasingról. A jeneség 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” komponensek, amelyek az eredetiben nem voltak benne. -> ez szvsz ugyan az 😊 -- [TitCar](#) - 2007.05.25. Ez bizony ugyanaz 😊 -- [Maco](#) - 2010.01.06.

4. feladat

Leginkább a zár- és zárréshangok torzulnak az átvitel során, mivel a spektrális szerkezetükben ezek a hangok rendelkeznek nagyon magas frekvenciaösszetevőkkel (4kHz fölött is), amit a telefonvonal szűrője levág. A gyakorlatban azért nem zavaró, mivel a hangok 4kHz alatti komponensei is adnak némi támpontot, valamint az ember a magasabb értelmezési szinteken a hang- és szövegkörnyezetből is következtetni tud arra, hogy milyen hang lehetett ott.

(Más ötlet?)

5. feladat

Valószínűségek:

- Ő állapotban θ_1 felismerésére: 0.242
- Ő állapotban θ_2 felismerésére: 0.054
- T állapotban θ_1 felismerésére: 0.054
- T állapotban θ_2 felismerésére: 0.242

2 lehetséges út (felismerés) van, valószínűségük:

- START-Ő-T-STOP: $1.0 * 0.242 * 0.1 * 0.242 * 0.3 = 0.00176$, azaz 0.17 %
- START-Ő-Ő-STOP: $1.0 * 0.242 * 0.5 * 0.054 * 0.4 = 0.00261$, azaz 0.26 %

vagyis a felismerő *"őő"*-t ír ki.

6., kreatív feladat 😊

WAP:

- mivel az adatátvitel szűkös beszédátvitelhez (adatforgalom+beszéd?), ezért lehetne az, h sms-ben elküldik a kért tőzsdeinformációkat, aztán az sms-t felolvassa az SMSmondó
- vagy eleve egy telefonon futó kliensalkalmazás leszedi a szükséges adatokat a szerverről WAPon keresztül és olvasná fel így integrált módon egymaga oldaná meg a problémát
- mindkét esetben egy egyszerű beszédszintetizátor jöhet csak szóba a mobiltelefonok (mondjuk azt hogy) egyelőre még szűkös kapacitása miatt, tehát formáns vagy diádalapú megoldás az elsődleges jelölt.
- mindkét esetben a kliensprogram nyelvét a felhasználó telepítésnél választhatná ki

Telefon:

- Itt már a vezérlés is lehet, sőt ajánlottan beszédalapú. Gondosan megtervezett, dialógusszerű, 2-3 hierarchiaszintes szerkezetben közölhetné igényét a felhasználó, pl: Mire kíváncsi? Értékpapír. Melyik cég papírja? Manchester United. Milyen mutatójára? Értékére. 4.52 penny.
- vagy kötetlen kérdésfeltevés után kérdezne rá a program a bizonytalan vagy hiányos részletekre: Hogy állnak a Manchester papírjai? 4.52 penny.
- a válaszok vegyes (TTS+kötött) felolvasó rendszer segítségével generálódnának, dinamikus rész lenne a számok, dátumok, esetleg cégnevek felolvasása, a többi statikus, vivőmondatok
- ügyelni kell hogy többnyelvű legyen a rendszer, hisz ez mégis a világ legnagyobb tőzsdéje, alából lehetne angol, és ha nem reagál a felhasználó, akkor mondaná adott nyelveken hogy melyik gombot nyomja meg ha héberül/kínaiul/xu! vagy egyéb nyelven szeretné hallani a frankót
- a beszédfelismerő beszélőfüggetlen legyen, a beszédgenerátor triadosként lenne optimális, esetleg diados
- a hang minősége 3,7khz, ezt mind beszédgenerálásnál (minták minősége), mind beszédfelismerésnél figyelembe kell venni

PC:

- számítógéppel a billentyűzet, de főleg az egér legalább olyan gyors kommunikációt biztosít mint a beszéd, kivéve kereséseknél
- tehát adott, kevés alternatíva közül egér segítségével, sok (cégek neve, dátum keresése) lehetőségnél beszéddel választanánk
- a felismerő futhatna a felhasználó gépén, csökkentve a szerver terheltségét, ugyanakkor nem szabad túl nagyok lennie, valamint az sem jó ha be kell tanítani minden szóra (DTW kiesett), tehát HMM jön szóba, moderált méretű paraméteradatbázissal, esetleg beszélőadaptív lehetne!
- a beszédgenerátort mindeképp a felhasználó gépén lenne érdemes megoldani, ahol alapértelmezetten/adott parancsra felolvasná a lekérdezés eredményét
- a program nyelvét a felhasználó választaná ki telepítésnél, de a kliens természetesen nyelvfüggetlen módon, csak adatok formájában kommunikálna a szerverrel

(további ötletek nyugodtan jöhetnek, ez egy kreatív feladat!!)

-- RGabo - 2006.05.26.