TTMER16 – Automatikus beszédfelismerés

1. ***A beszédfelismerők mely három nagy csoportba oszthatók?***

**Izolált szavas** – Rövid szünetekkel elválasztott szavakat vagy szócsoportokat tudnak feldolgozni.

**Kapcsolt szavas** – Szavak sorozatával dolgozik, mely a szótárkészlet szavainak tetszőleges sorrendű, korlátozott hosszúságú sorozata lehet. Nem feltételez szüneteket a kapcsolt szavak, csak a sorozatok között. A referencia szótárkészlet elemei általában izolált szavak.

**Folyamatos** – Legelterjedtebben rejtett Markov modelles felismerő. A mondatokat, szavakat fonéma, félszótag vagy szó szinten állapotok sorozatával írják le. A bejövő jellemzővektor sorozatot a Viterbi algoritmussal értékelik ki és meghatározzák, hogy az egyes modellek milyen valószínűek a megfigyelés alapján. Beszélőfüggetlen felismerés lehetősége.

1. ***Mit végez minden beszédfelismerő a bejövő beszédjelen legelőször?***

**Lényegkiemelés** meghatározza a beszéd tartalmát hordozó mennyiségeket és kiküszöböli a felismerés szempontjából érdektelen információkat (zaj, torzítások). A digitalizált beszédjelből egy diszkrét idejű, adott dimenziójú jellemzővektor sorozatot állít elő. (eljárások: lineáris predikció, Fourier transzformáció, kepsztrális analízis)

1. ***Miért bontjuk keretekre a bejövő beszédjelet? Mekkora a keretek mérete (nagyságrendben)?***

A beszédjel folyamatosan változik, és ezen változásokat szeretnénk nyomon követni. Túl nagy keretméret esetén a gyors változások kiátlagolódnak, túl kis keretméret esetén a beszéd jellemzőit pontatlanul tudjuk csak meghatározni. A keretméret 32 ms, 50%-ban átlapoltak.

1. ***Lényegkiemelésnél a DFT előtt miért használnak Hamming-ablakot?***

A DFT előtt ablakolni kell, mert nem periódikus jellel dolgozunk. A négyszögablak nem jó, mert pl. tiszta szinuszjel esetén az előállított spektrum „szétkent” lesz, nem csak a szinuszjel frekvenciájának megfelelő helyen találunk nullától különböző értéket. A Hamming ablak azonban a spektrumot „élesíti”, mert szűri a spektrumot.

1. ***Mi a mel-skála?***

Az emberi hangérzékelés a frekvencia függvényében változik. Magasabb frekvenciákon szélesebb frekvenciaközű hangokat képes csak megkülönböztetni, mint alacsonyabb frekvenciákon. A frekvencia felbontás 1000 Hz alatt közelítőleg lineárisan változik, efölött logaritmikusan.

1. ***Mi a lineáris idővetemítés? (Rajzolhat ábrát is!)***



A vastag vonal jelzi azt az utat, amely mentén haladva egyenletesen nyújtjuk ill. zsugorítjuk a bemenő vektorsorozatot az összehasonlításhoz.

1. ***A mérésben szereplő beszédfelismerő az időillesztésre milyen módszert használ? Néhány szóval írja le, hogy időillesztésre miért van szükség!***

Dinamikus idővetemítés: bemenete két vektorsorozat, kimenete pedig a lehető legjobb idővetemítés mellett a két vektorsorozat megfelelő vektorai közötti távolságok összege.

A bemondott szóhoz tartozó jellemzővektor-sorozat és az összes tárolt szó vektorsorozata közötti távolságot számolja ki, és a lehető legkisebb távolságút adja meg, mint felismert szót.

Cél: a különböző beszédsebességekből adódó eltérések korrigálása

1. ***Két jellemző vektort hogy hasonlíthatunk össze?***

Két vektor távolságát azok euklidészi távolságával adhatjuk meg:

**

1. ***A mérésben szereplő beszédfelismerővel hogyan lehet beszédfelismerési kísérleteket automatikusan elvégezni?***

A parancsállomány segítségével. A parancsállomány egysoros parancsokat, utasításokat tartalmaz, amelyeket egy értelmező végrehajt. A végrehajtás során jegyzőkönyv állomány is keletkezik.

1. ***A mérésben szereplő beszédfelismerő milyen párbeszédes alkalmazást valósít meg?***

Beszéddel vezérelt telefonos tárcsázás. A hívott fél telefonszámát, vagy egy korábban betanított előfizető nevét kell bemondani. Új felhasználó esetén a rendszert be kell tanítani.