**TTMER32: Dialógustervezés VoiceXML-ben**

**Ellenőrző kérdések**

1. ***Mi a beszédszintézis célja?***

A beszédszintézis nem más, mint az emberi beszéd előállítása mesterséges módon, tipikusan számítógép segítségével. Amennyiben a bemenet írott szöveg, beszédfelolvasóról (Text-to-Speech) beszélünk.

1. ***Melyek a legjelentősebb beszédszintetizátor-technológiák?***

Formánsszintézis, Elemösszefűzés, Elemkiválasztás, Rejtett Markov modell

1. ***Írd le az elemösszefűzéses (pl. diád alapú) beszédszintetizátor alapelvét!***

Az elemösszefűzéses beszédszintézis során természetes beszédből kivágott hullámforma elemeket fűznek össze. A mesterségesen előállítható beszéd érthetőségéért a hangátmenetek természetessége a felelős, és nem maguk a hangok. Emiatt kiemelten fontos a hangátmenetek megfelelő modellezése. Az egyes elemösszefűzéses rendszerek elemszámuk alapján kerülnek megkülönböztetésre, pl diád rendszer 382 = 1444, míg a triád 383 = 54872 mintát tartalmaz. Pl. a busz szó két különböző hangkörnyezetből kivágott diád elem egymás után helyezésével jön létre. Az előálló beszéd megfelelő prozódiájáról is gondoskodni kell jelfeldolgozási módszerek segítségével. Az így előálló beszéd érthető, de messze nem természetes.

1. ***Írd le az elemkiválasztásos (pl.korpusz alapú) beszédszintetizátor alapelvét!***

Az elemkiválasztásos beszédszintézis az elemösszefűzéses továbbfejlesztése, nagyobb beszédadatbázis (korpusz) áll rendelkezésre, amelyben egy-egy elem többször is előfordulhat különböző környezetben, ráadásul az elemek hosszabbak, szavak ls szókapcsolatok is lehetnek. A kimeneti beszéd létrehozása során a rendszer minél hosszabb olyan elemeket keres a korpuszban, amelyek a bemeneti szöveghez illeszkednek. Mivel az elemek hosszabbak, kevesebb összefűzési pont lesz a beszédben. A szintetizált beszéd minősége javítható azáltal, hogy a lehető legtermészetesebb beszédelem kerüljön kiválasztásra. A rendszer minőségét befolyásoló tényező beszédkorpusz és a bemeneti szöveg témájának közelsége.

1. ***Mi a beszédfelismerés célja?***

A beszédfelismerés célja, az akusztikai beszédjelet szöveggé alakítani, ezzel lényegében a beszédszintézis inverz folyamatát valósítják meg.

1. ***Melyek a beszédfelismerés fő lépései és mit eredményeznek?***

Lényegkiemelés: a jelfeldolgozás során kinyerjük a beszédjelből a beszéd tartalmára vonatkozó paramétereket. Eredménye egy standardizált, a következő lépésekhez optimalizált diszkrét idejű jel.

Mintaillesztés: A mintaillesztés feladata jellemzővektor sorozat leképzése egy szótári elemre, vagy azok sorozatára. Napjainkban a legelterjedtebb a statisztikai alapú felismerők alkalmazása, ahol a vektorsorozatot egy tanítóadatok alapján becsült, HMM alapú valószínűségi modell struktúrához illesztik. A mintaillesztés folyamatának végeredménye a legjobb illeszkedéshez tartozó nyelvi modell kimeneti címkéit, és elhangzásuk időpontját tartalmazza.

1. ***Ismertesse az MFCC számításának lépéseit!***

MFCC: Mel Frequency Cepstral Coefficient

Az FFT spektrális komponensekre alkalmazott mel-skálás átlagolásban a komponensek összegzésére használatos ablak méretét 1kHz felett exponenciálisan növeljük, kompenzálva a kisebb információsűrűséget. A mel összegzett komponensekből teljesítmény spektrumot számítunk zajelnyomási megfontolásból, majd az így kapott értékeke logaritmizáljuk illeszkedve az inger és érzet között általában fennálló kapcsolathoz.

1. ***Mi az akusztikai és nyelvi modell közti különbség?***

Az akusztikai modellben a rendelkezésre álló tanító hanganyag és annak szöveges leirata alapján becsüljük a jellemzővektor sorozat eloszlását az elhangzó fonémákra nézve. Az így létrejött statisztikai modellel becslést lehet adni egy jellemzővektor sorozat és egy fonéma sorozat közötti akusztikai illeszkedés mértékére. Izolált szavas beszédfelismerésnél erre támaszkodunk. Míg a nyelvi modell határozza meg, hogy milyen módon és valószínűséggel kapcsolódhatnak össze a felismerő szótári elemei. Folyamatos beszédfelismerés során a tanítószöveg alapján becsüljük meg a szótári elemek kapcsolódásának valószínűségeit és az akusztikai modell bizonytalanságát is kivédjük ezzel a modellel.

1. ***Mely paranccsal van lehetőség a VoiceXML-ben egy adott szöveg felolvasására?***

A <prompt> *üzenet* </prompt> címke használatával.

1. ***Egészítsd ki a következő VoiceXML kódot, hogy köszöntse a rendszert felhívó személyt!***

<prompt>

Üdvözlöm Önt rendszerünkben.

</prompt>

1. ***Mely paranccsal lehet VoiceXML-ben nem felismert beszéd esetén automatikus hibaüzenetet visszaadni?***

A <nomatch> címke határozza meg mi történjen, ha nem sikerült felismerni a beszédet.

1. ***Mely paranccsal lehet VoiceXML-ben válasz hiánya/csend esetén automatikus hibaüzenetet visszaadni?***

A <noinput> címke határozza meg mi történjen, ha nem volt válasz, vagy nem volt elég hangos a felismeréshez.

1. ***Milyen paranccsal lehet VoiceXML változót deklarálni?***

A <var> címke egy változót deklarál, pl. <var name=”result”/>

1. ***Mely parancsokkal tudunk VoiceXML-ben feltételes kifejezéseket létrehozni?***

<if> majd a cond tulajdonság vagy feltétel, ha teljesül akkor végrehajtódik az <if> után következő kódrészlet, ha nem akkor a rendszer átugrik további kódrészletekre amíg <else> vagy <elsif > címkét nem talál.

1. ***Mely paranccsal van lehetőség VoiceXML-ben beszédfelismerő nyelvtan definiálására?***

<grammar> paranccsal van erre lehetőségünk.

1. ***Egészítsd ki a következő VoiceXML kódot úgy hogy felismerje a Hírek / Mozi / Sport szavakat!***

<one-of>

<item>hírek</item>

<item>mozi</item>

<item>sport</item>

</one-of>