

## **DIPLOMATERV összefoglaló**

**Készítette: Csapó Tamás Gábor**  
2008. tavaszi félévében

### **Változatos prozódia megvalósítása szövegfelolvasó rendszerekben**

Az információs társadalom kialakulása során elengedhetetlen az ember-gép kapcsolat folyamatos fejlesztése. A beszédtechnológiai alkalmazások, ezen belül a beszéd-szintézis elterjedése is ebbe a folyamatba illeszkedik, hiszen sok esetben a felhasználó és a gép között beszéd segítségével megvalósuló kommunikáció nélkülözhetetlen.

A beszéd-szintézis rendszerek minőségét az alapján ítélik meg, hogy az általuk keltett beszéd mennyire hasonlít az emberire. A létrehozott hang érthetősége a mai szövegfelolvasókban már nem jelent problémát. A jelenlegi rendszerek többsége egy szabályrendszer segítségével a nyelvi elvárásoknak megfelelő, adott szöveghez mindig azonos prozodiát (intonáció, hangsúlyozás, ritmus) rendel. Ugyanakkor ahhoz, hogy a gépi megoldás ne tűnjön monotonnak, az emberhez hasonlóan változatosságot kell létrehozni, azaz ugyanazt a mondatot nem mindig ugyanúgy kell bemondania a rendszernek. Az elmúlt években fokozatosan előtérbe került ez a kulcskérdés.

A dolgozatban először áttekintjük a beszéd-szintézis szakirodalmát, a prozódia modellezésének lehetőségeit részletesen ismertetve. Bemutatjuk az emberi beszéd változatosságának vizsgálatára és modellezésére tett kísérleteket. Megtervezünk és implementálunk egy módszert, amely alkalmas a korábbiaknál változatosabb mesterséges beszéd előállítására. Ezt oly módon végzi, hogy a bemeneti szöveghez természetes beszédből származó minták alapján határozza meg a dallamot. A változatosságot a dallamminta választásának bizonyos fokú véletlenszerűsége biztosítja. A megvalósított rendszer minőségét a különböző szempontok szerint elvégzett elemzések eredményei tanúsítják.

Az ily módon természetesebb hangzásúvá tett szövegfelolvasó rendszer számos gyakorlati alkalmazásban használható, mint például SMS-, e-levél-, könyv-felolvasó, vagy telefonos tudakozó. A változatosabb prozódia főleg hosszú szövegek felolvasása esetén előnyös, hiszen ekkor zavaró leginkább a beszéd-szintetizátor monotonitása.