

Tóth Bálint Pál „Rejtett Markov-modell alapú gépi beszédkeltés” című pályaművének rövid összefoglalása

A szerző dolgozatában a szöveg-beszéd átalakítás (*Text-to-speech, TTS*) statisztika alapú megközelítésével, ezen belül a rejtett Markov-modell alapú szövegfelolvasással (*Hidden Markov model based Text-To-Speech, HMM-TTS*) kapcsolatos tudományos téziseit mutatja be. A rejtett Markov-modell alapú szöveg-beszéd átalakítás az elmúlt évtizedben talán a legnagyobb érdeklődést váltotta ki a tudományterületen. A szerző tudomása, és a hazai illetve nemzetközi publikációk alapján Magyarországon elsőként foglalkozott a tématerülettel.

Az értekezés egy közel húsz oldalas elméleti összefoglalóval kezdődik, majd a tudományos célkitűzések és a kutatói munka során használt eszközök, módszertan bemutatásával és a beszédatadabázisok ismertetésével folytatódik. Ezt követően a szerző három témakörbe foglalva ismerteti új tudományos eredményeit.

Az **első témakör** a rejtett Markov-modell alapú szövegfelolvasó magyar nyelvre való kialakításával és továbbfejlesztésével foglalkozik. A szerző első lépésként létrehoz egy magyar nyelvű, beszélőfüggő HMM-TTS rendszert. Meghallgatásos teszttel igazolja, hogy ennek minősége szignifikánsan nem tér el a korábban legjobb minőségűnek tartott téma-függő korpusz alapú szövegfelolvasó minőségétől, mindamellett a HMM-TTS rendszer erőforrás igénye lényegesen kisebb és téma-független működésre képes. Következő lépésként megkülönböztető jegyek bevezetésével minőségjavulást ér el a korábbi rendszerhez képest, melyet szintén meghallgatásos teszttel igazol. Ezek után magyar nyelvű, beszélőadaptált HMM-TTS rendszert hoz létre és megmutatja, hogy a beszélőfüggő rendszernél ennek szignifikánsan jobb a minősége, miközben új beszédhangok létrehozásához jelentősen kevesebb hanganyagra van szükség. Az első témakör negyedik részében empirikus úton igazolja, hogy a tanító adatbázis nagy pontosságú kézi javításának elhagyása nem okoz szükségszerűen szignifikáns minőségromlást beszélőfüggő és beszélőadaptált HMM-TTS-ek esetén.

A **második témakörben** a szerző azzal a kérdéssel foglalkozik, hogy hogyan tud emberi beavatkozás nélkül minél jobb minőségű gépi beszédhangot létrehozni. A dolgozatban bemutatásra kerül egy eljárás, amely képes a beszédfelismerő kimenetén alapuló felügyelet nélküli beszélőadaptációra. A szerző kísérleti úton igazolja, hogy az eljárás segítségével lehetséges a jelentős mennyiségű emberi munkát megkívánó felügyelt eset minőségétől szignifikánsan nem eltérő minőségű gépi beszéd előállítására. A második témakör második részében egy olyan eljárást kerül bemutatásra, mely képes a felügyelet nélküli beszélőadaptáció által elérhető gépi beszéd minőségének javítására. Az eredmények ebben az esetben is meghallgatásos tesztekkel vannak alátámasztva.

Az értekezés **harmadik témaköre** a rejtett Markov-modell alapú szövegfelolvasás korlátozott erőforrású eszközökre (pl. okostelefonok) való illesztésével foglalkozik. Első lépésként a szerző modellt dolgoz ki a HMM-TTS korlátozott erőforrású eszközökön való megvalósítására és mérésekkel igazolja, hogy a gépi beszéd ezzel a

módszerrel való előállítása szignifikáns minőségromlás nélkül képes szignifikánsan gyorsabb működésre korlátozott erőforrású eszközökön. Ezt követően megvizsgálja a számításigényes részeit a HMM-TTS-nek és elméleti és gyakorlati javaslatot tesz ezek párhuzamos működésére. A dolgozatban mérések támasztják alá, hogy a párhuzamos működés bevezetésének hatására azonos beszédminőség mellett a szövegfelolvasó válaszideje szignifikánsan javul.

A szerző minden eredményének megvalósíthatóságát és működését kísérleti mintarendszerekben alkalmazta. A közeljövőben várhatóan a kutatás részeként létrejött magyar nyelvű HMM-TTS megjelenik a Google Play alkalmazás áruházban.