

Mihajlik Péter „Spontán magyar nyelvű beszéd gépi felismerése nyelv-specifikus szabályok nélkül” című pályaművének összefoglalása

A szerző a magyar nyelvű gépi beszéd felismerés gyakorlati alkalmazhatósága érdekében végzett munkásságát, azon belül a hozzá legszorosabban köthető, egymásra épülő tudományos téziseit mutatja be.

Elsőként tömör szakmai összefoglalót nyújt a gépi beszéd felismerés korszerű módszereiről. Részletesebben ismerteti a viszonylag új WFST (súlyozott véges állapotú transzdúcer) keretrendszert, melyet magyarországi kutatóhelyen elsőként ő alkalmazott gépi beszéd felismerésre.

Az első tézis a magyar nyelvű fonetikai koartikuláció, azaz a beszédhangok artikulációs egymásrahatásának vizsgálatával foglalkozik beszéd felismerési alkalmazásban. A szerző kísérleti úton megmutatja, hogy a magyar nyelvtechnológiában méltatlanul elhanyagolt trifón modellezés minden esetben drasztikusan jobb felismerési eredményeket nyújt, mint a mások által általánosan alkalmazott környezetfüggetlen beszédhangmodellezés. Két módszert is bemutat a trifón (hármashang) modellek hatékony kialakítására, melyek közül a nagyobb mértékben adatvezérelt megoldás szignifikánsan jobb felismerési eredményeket szolgáltat.

A második tézis a magyar nyelvű fonológiai koartikuláció, azaz a hasonulási szabályok szóhatárokon is átívelő modellezésének kérdésével foglalkozik a gépi beszéd felismerés szempontjából. A szerző által bemutatott kísérletek megmutatják, a tipikus hasonulási jelenségek komolyabb elméleti háttérrel és eszköztárral igénylő explicit modellezése nem javítja érdemben a felismerés pontosságát ahhoz képest, mintha ezen jelenségek meglététől eltekintենek. Ennek gyakorlati haszna, a beszéd felismerő rendszerek minőségromlás nélküli egyszerűbb konstrukciója. A legjobb eredmények eléréséhez szükség van az előző tézisbeli, adatvezérelt *fonetikai* koartikuláció modellezésre.

Az első két tézisben említett kísérletek alkalmazás-közeli, minden korábban publikálnál jóval nagyobb magyar nyelvű telefonos beszéd adatbázisokkal készültek.

A következő két tézis a folyamatos spontán nagyszótáros magyar nyelvű beszéd felismeréssel foglalkozik, melynek vizsgálata nemzetközileg ismert és elismert beszéd adatbázison történt.

A harmadik tézisben a szerző javaslatot tesz a magyar nyelvű spontán nagyszótáros beszéd felismerésénél a ragozásból, hajlításból adódó morfológiai változatosság kezelésére. A javasolt megoldásban szavak helyett morfém szerű egységek képzik mind a nyelvi, mind az akusztikai-kiejtési modellek alapjait. A dolgozat a nemzetközi szakirodalomban az első között mutat fel spontán nyelvű beszéd felismerésnél statisztikai alapokon származtatott „morf” lexikai egységekkel a hagyományos szó egységekhez, illetve a nyelvi elemzővel kapott morféma egységekhez viszonyítva szignifikánsan jobb eredményeket.

A 4. tézis lényege, hogy az adott spontán beszéd adatbázis esetén az ortografikus leírt alakról az elhangzott „fonetikus” alakra történő leképezéshez semmilyen nyelvi szabályt, vagy kiejtési szótárat nem használva nem romlanak lényegesen a felismerési eredmények. Ezt a betűk (grafémák) közvetlen akusztikai modellezésével éri el a szerző. A tézis gyakorlati szempontból nagy mennyiségű fáradtságos munkától kímélheti meg a beszéd felismerő

rendszer készítőit, illetve lehetőséget ad más nyelvű beszédfelismerő rendszerek költségkímélő kifejlesztésére.

Végül a szerző kísérletileg megmutatja, hogy az összes nyelvspecifikus szabály (morfoszintaktikai, kiejtési, graféma-fonéma átalakítási, fonetikai) elvételevel tisztán az adatokra építve a beszédfelismerési modelleket nem változik szignifikánsan a felismerési hatások a „state-of-the-art” beszédfelismerő rendszerhez képest. A hagyományos szó és (környezetfüggő) fonéma-alapú megközelítéshez képest – melyet nemzetközileg elterjedten használnak – a javulás pedig szignifikáns.

A tézisek termékekben, szolgáltatásokban részben hasznosulnak, részben hasznosítás alatt állnak, valamint elméleti szempontból is széleskörű érdeklődésre tarthatnak számot.