

A spontán beszéd prozódiai frázisszerkezetének modellezése és felhasználása a beszédfelismerésben¹

Pápay Kinga

DE BTK Általános és Alkalmazott Nyelvészeti Tanszék
4032 Debrecen, Egyetem tér 1.
kinga.papay@gmail.com

A spontán beszéd egységeinek jelölése, felismerése, illetőleg elkülönítése az automatikus beszédfelismerés egyik alapvető problémája nemzetközi szinten is [2], [6], [7], [9], [10]. Amellett, hogy a prozódiai és egyéb kulcsok a szöveg típusától függően különböznek, további problematikus pont, hogy a prozódiai megvalósítás és a szintaktikai szerkezet közötti kapcsolat feltérképezhetősége nemzetközi vita tárgya [3], [4], [5], [8]. Magyar vonatkozásban tovább nehezíti a fejlesztéseket a prozódiailag felcímkezett, országos nagyságú, spontán beszéd adatbázisok hiánya. Ugyanakkor megfelelő akusztikai előfeldolgozással (a szegmentális tartományban végzett lényegkiemelés jellemző vektorainak használata mellett a szuprasegmentális, prozódiai jellemzőkön alapuló lényegkiemeléssel), valamint a spontán beszéd szuprasegmentális jellemzőinek kutatásával a beszédfelismerő rendszerek hatékonysága növelhető: minél többet tudunk bevinni az emberi beszédfelismerési folyamat szintjei – akusztikai, fonetikai-fonológiai, szintaktikai, szemantikai, illetve pragmatikai szint – közül a gépi beszédfelismerésbe, annál biztosabb lesz a működése.

A kutatás célja a magyar spontán beszéd prozódiai frázisokra (IP-kre) bontása, a prozódiai határok megállapítása és ennek bekapcsolása a beszédfelismerő rendszerbe. A kutatás a spontán beszéd vizsgálatán keresztül járul hozzá a pontosabb ismeretekhez a prozodiáról, különös tekintettel a beágyazásokra – a beágyazott részek lokalizálásával, a tonális folytonosság szabályainak megállapításával és rendszerbe illesztésével kísérletet teszünk a felismerő hatékonyságának növelésére. Az elméleti nyelvészet prozódiaival kapcsolatos aktuális eredményeit használjuk fel. A prozódia ráillik a szintaktikai csoportosításra az alapvető tagolásban, de további, szemantikai és pragmatikai funkciói is vannak, amelyek ki vannak fejezve a prozódia egy származtatott szintjén. A prozódia az elsődleges elemét, a dallamot használja fel a csoportosításhoz; a dallamvariációk rekurzív használatára utal, hogy minél mélyebb a beágyazás, annál alacsonyabb frekvencián kezdődik a dallam. A prozódia reprezentálja a szintaktikai szegmentumok diszkontinuitását és a tonális kontúrok kapcsolódnak egymáshoz – a szintaktikai diszkontinuitás prozódiai reprezentációja az ún. könyvjelző-hatás. Ez a tulajdonság tágítja a hozzárendelések lehetőségét a szintaxis és a prozódia között, és a prozódiai frázisok kapcsolódása a felismerő szempontjából is modellezhető. A vizsgálatok a beágyazások, alárendelések és mellérendelések, illetve az újrakezdések és hezitálások prozódiai jellegzetességeire terjednek ki, különös tekintettel a tonális folytonosságra, a nem folytonos tonális összeállításra és a tonális rekonstrukció elvére [4], [5]. E prozódiai jellemzők felhasználásának eredménye lehet a keresési tér csökkené-

¹ A kutatás Az ember-gép kommunikáció technológiájának elméleti alapjai című, TÁMOP-4.2.2-08/1/2008-0009 jelű projekt keretein belül zajlik.

se (lehetőséget adhat a felismerés során futó Viterbi-algoritmus szakaszolására), zajos körülmények között robusztusabbá teheti a felismerő működését (ezáltal gyorsul és pontosabb lesz a felismerés), illetve felismerheti a megakadásjelenségeket (szintén a pontosabb felismeréshez járul hozzá).

A szupraszegmentális hangszerkezet egyes elemei, a prozódiai jegyek lényegében a három akusztikai jellemző különböző időtartományokra érvényes – szó- vagy mondat szintű – kombinációi. A beszéd alapfrekvencia-, energia- és időviszonyainak vizsgálatát statisztikai módszerekkel végezzük magyar nyelvű, megfelelő spontán beszéd adatbázison. A spontán beszéd adatbázis gyűjtése és felhasználása specifikusabbá teszi a felismerőt, hiszen a spontán beszédben még gyakoribbak azok a jelenségek, amelyek az automatikusan futó algoritmust megzavarhatják: szótévesztések, javítások, újrakezdekés, változtatások a közlés közben, hevesebb érzelmek stb. A méréseket, illetve az annotálást (szegmentálás, címkézés és feliratozás) a Praat hangelemző szoftver [1] segítségével végezzük; ennek során az adott tagmondat hullámformájához rendeljük annak alapfrekvencia- és intenzitás görbéjét. A vizsgálatok után következik a szabályalkotás, illetve a statisztikai modellezés, valamint ezek bekapcsolása a HTK beszéd felismerő rendszerbe [11] – az új modult a rendszerbe illesztve annak vizsgálata következik, hogy milyen mértékben javítható a beszéd felismerés hatékonysága. Statisztikai modellezés esetén a betanítás során az adatbázis hangfájlaiból az előfeldolgozással nyert szupraszegmentális jellemző vektorok, valamint az adatbázis szegmentálási és címkézési adatai használhatók fel a prozódiai modellek felépítéséhez. A prozódiai szegmentálás ismeretében a hipotézis gráfok újrasúlyozhatók, így a végeredmény kiértékelését már a prozódia alapján nyert információ is befolyásolja [2], [6], [9], [10].

Hivatkozások

1. Boersma, P., Weenink, D.: Praat: doing phonetics by computer 5.1.14. Institute of Phonetic Sciences, University of Amsterdam (2009) <http://www.praat.org>
2. Borostyán G., Szaszák Gy., Vicsi K.: Folyamatos beszéd szószintű szegmentálása szupraszegmentális jegyek alapján. In: Alexin Z., Csendes D. (eds.): II. Magyar Számítógépes Nyelvészeti Konferencia. SZTE Informatikai Tanszékcsoport, Szeged (2004) 319 – 326
3. Chomsky, N., Halle, M.: The Sound Pattern of English. Harper and Row, New York (1968)
4. Hunyadi, L.: Grouping, the cognitive basis of recursion in language. In: Kertész, A. (ed.): Argumentum, 2. Kossuth Egyetemi Kiadó, Debrecen (2006) 67 – 114
5. Hunyadi L.: Cognitive grouping and recursion in prosody. In: van der Hulst, Harry (ed.): Recursion and Human Language. Mouton de Guyter, Berlin New York (2009)
6. Németh Zs., Szaszák Gy., Vicsi K.: Prozódiai információ használata az automatikus beszéd felismerésben; mondatmodalitás felismerése. In: Alexin Z., Csendes D. (eds.): V. Magyar Számítógépes Nyelvészeti Konferencia. SZTE Informatikai Tanszékcsoport, Szeged (2007) 69 – 80
7. Rabiner, L.: Fundamentals of Speech Recognition. Prentice Hall, Englewood Hills, NJ (1993)
8. Selkirk, E. O.: Phonology and Syntax: The Relation between Sound and Structure. MIT Press, Cambridge (1984)
9. Szaszák, Gy.: A szupraszegmentális jellemzők szerepe és felhasználása a gépi beszéd felismerésben. PhD értekezés, Budapest (2009)

10. Szaszák, Gy., Vicsi, K.: Folyamatos beszéd szószintű szegmentálása szupraszegmentális jegyek alapján II. In: Alexin Z., Csentes D. (eds.): III. Magyar Számítógépes Nyelvészeti Konferencia. SZTE Informatikai Tanszékcsoport, Szeged (2005) 360-370
11. Young, S. et al.: The HTK Book (for version 3.4). Cambridge University, Cambridge (2009)