

A spontán beszéd prozódiai frázisszerkezetének modellezési lehetőségei

A spontán beszéd egységeinek jelölése, felismerése, illetőleg elkülönítése az automatikus beszédfelismerés egyik alapvető problémája nemzetközi szinten is (Rabiner 1993, Szaszák et al. 2004, 2005, 2007, 2009). 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 bevonni 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, illetve a spontánbeszéd-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, újrakezdések, változtatások a közlés közben, hevesebb érzelmek stb. méréseket, illetve az annotálást (szegmentálás, címkézés és feliratozás) a Praat hangelemző szoftver segítségével végezzük; ennek során az adott tagmondat hullámformájához rendeljük annak alapfrekvencia- és intenzitásgörbéjét. A vizsgálatok után következik a szabályalkotás, illetve a statisztikai modellezés, valamint ezek bekapcsolása a beszédfelismerő rendszerbe – az új modult a rendszerbe illesztve annak vizsgálata következik, hogy milyen mértékben javítható a beszédfelismerés hatékonysága. A kutatás első lépése a prozódiai frázis fogalmának meghatározása, és a lehetséges, illetve a vizsgálat szempontjából releváns frázistípusok felállítása – ehhez áttekintjük a témával kapcsolatos szakirodalmat, illetve az általunk gyűjtött multimodális HuComTech adatbázis spontán beszédet tartalmazó részét. Az így felállított típusokat használjuk fel az adatbázis (több szintű) annotációjához. A prozódiai frázisok szintjén a parantézisek, beágyazások, iterációk, újrakezdések vizsgálatán túl ígéretes a hezitációk kategóriája, amely megjelenhet külön frázisként, illetve a megelőző frázis folytatásaként is – a vizsgálatok során megkérdőjeleződik ezek „kitöltött szünetként” való kezelhetősége, illetve a frázisok szintjén túl az érzelmi vagy a kognitív állapot szintjén is értelmezhetők.

Hivatkozások:

- Borostyán, G., Szaszák, Gy., Vicsi, K. 2004. Folyamatos beszéd szószintű szegmentálása szuprasegmentális jegyek alapján. In: Alexin Z., Csendes D. (szerk.) II. Magyar Számítógépes Nyelvészeti Konferencia. Szeged: SZTE Informatikai Tanszékcsoport. 319 – 326.
- Németh, Zs., Szaszák, Gy., Vicsi, K. 2007. Prozódiai információ használata az automatikus beszédfelismerésben; mondatmodalitás felismerése. In: Alexin Z., Csendes D. (szerk.) V. Magyar Számítógépes Nyelvészeti Konferencia. Szeged: SZTE Informatikai Tanszékcsoport. 69 – 80.
- Rabiner, L. 1993. Fundamentals of Speech Recognition. Englewood Hills, NJ: Prentice Hall
- Szaszák, Gy. 2009. A szuprasegmentális jellemzők szerepe és felhasználása a gépi beszédfelismerésben. Budapest: PhD értekezés.
- Szaszák, Gy., Vicsi, K. 2005. Folyamatos beszéd szószintű szegmentálása szuprasegmentális jegyek alapján II. In: Alexin Z., Csendes D. (szerk.) III. Magyar Számítógépes Nyelvészeti

Konferencia. Szeged: SZTE Informatikai Tanszékcsoport. 360 – 370.