

Beszéd szintetizátor prozódiai változatosságának növelése

Csapó Tamás Gábor, V. Info., csapszi@sch.bme.hu

Konzulensek:

Dr. Fék Márk, TMIT, fek@tmit.bme.hu

Dr. Németh Géza, TMIT, nemeth@tmit.bme.hu

Napjainkban közösen éljük meg az információs társadalom kialakulását. Ehhez elengedhetetlen az ember-gép kapcsolat folyamatos fejlesztése, ugyanis széles rétegek számára kell elérhetővé tenni az új technológia által kínált lehetőségeket. Ebbe a folyamatba illeszkedik a beszédtechnológiai alkalmazások, ezen belül is a beszéd szintézis elterjedése. A felhasználó és a gép közötti beszéd segítségével megvalósuló kommunikáció nélkülözhetetlen, ha a felhasználó keze és látása lekötött (pl. autóvezetés közben), illetve sérülés miatt nem használható (pl. látássérültek esetében), továbbá ha az igénybe vett szolgáltatás telefonvonalon keresztül érhető el.

A beszéd szintézis rendszerek minőségét az alapján ítélik meg, hogy az általuk keltett beszéd mennyire hasonlít az emberi beszédre. A jelenlegi rendszerek többsége egy szabályrendszer segítségével a nyelvi elvárásoknak megfelelő, adott szöveghez mindig azonos prozódia (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 év során jelentős kezdeti eredményeket értünk el a beszéd szintetizátorok prozódiajának változatosabbá és természetesebbé tétele területén nagyméretű természetes beszéd korpusz felhasználásával. A jelen dolgozatban először a témához tartozó szakirodalmat tekintjük át, majd a korábban kidolgozott módszer továbbfejlesztési irányait ismertetjük. Egyrészt finomítottunk a távolságmértéken, amelyet a bemeneti szöveg és a beszéd korpuszban szereplő természetes mondatból származó prozódia társításához használunk. Másrészt a korábbi módszerek eredményességét nagyobb beszéd korpuszon is vizsgáltuk. Érdekes kérdés az is, hogy természetes prozódia egy az egyben történő másolásához képest javít-e a rendszeren egy absztrakciós szint bevezetése.

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

Feldolgozott irodalom:

1. Olaszy, G., Kovács, M., Nikléczy, P., Gósy, M., Magyar nyelvi beszédtechnológiai alapismeretek. (600 oldal CD-ROM-on). Szerk.: Olaszy G., Nikol Kiadó, Budapest, 2002, <http://alpha.tmit.bme.hu/pub/beszinf/start.html>
2. Raux, A., Black, A., „A Unit Selection Approach to F0 Modeling and its Application to Emphasis”, ASRU 2003, pp. 700-705.
3. Dong, M., Lua, K., T., „An Example-based Approach for Prosody Generation in Chinese Speech Synthesis”, International Symposium on Chinese Spoken Language Processing, Beijing, 2000, pp. 303-307.