

## Zöngétlen beszédhangok modellezése a rejtett Markov-modell alapú beszéd szintézisben

Csapó Tamás Gábor<sup>1</sup> – Németh Géza<sup>2</sup>

<sup>1,2</sup> Budapesti Műszaki és Gazdaságtudományi Egyetem, Távközlési és Médiainformatikai Tanszék

<sup>1</sup> csapot@tmit.bme.hu, <sup>2</sup> nemeth@tmit.bme.hu

A rejtett Markov-modell alapú beszéd szintézis (HMM-TTS) során beszédkódoló eljárással állítjuk elő a szintetizált beszédet a modellezett paramétereiből. A szakirodalomban eddig a zöngés beszédhangok megfelelő modellezésére fordítottak nagyobb figyelmet (Zen et al. 2009; Drugman & Raitio 2014). Jelen előadásban bemutatjuk a zöngétlen hangok modellezésére tett kísérleteinket. A beszédkódoló alap változata (Csapó et al. 2015) folytonos alapprofundencia (F0) modellezésen alapul, így nem különbözteti el a zöngés és zöngétlen beszédhangokat F0 szempontjából. Két alternatív megoldást javasolunk a zöngétlen hangok modellezésére, 1) adatvezérelt módon, 2) nyelvi szabályokkal (Csapó et al. 2016). Az adatvezérelt módszer a gépi tanulás előtt, az analízis lépésben újszerű módon számítja a paramétereiket. A szabály alapú módszer a szintetizálandó szöveg fonémáin alapul.

Egy objektív paramétereket összehasonlító mérés szerint az első alternatívával szintetizált zöngétlen beszédhangok közelebb vannak a természetes beszédhez az alaprendszerhez képest. Egy angol nyelvű mintákat tartalmazó meghallgatásos teszt szerint mindkét módszer jobb, mint az alaprendszer. Egy másik, magyar nyelvű mintákat tartalmazó meghallgatásos tesztben azt vizsgáltuk, hogy hat férfi és hat női beszélő szintetizált beszéde közül melyik preferált. Az eredmények szerint a professzionális beszélők szintetizált hangját kellemesebbnek találták a tesztelők.

### Irodalom

- Csapó, T. G., Németh, G., Cernak, M. 2015. Residual-based excitation with continuous F0 modeling in HMM-based speech synthesis, SLSP 2015 (3rd International Conference on Statistical Language and Speech Processing), Budapest, Magyarország, Lecture Notes in Artificial Intelligence 9449, pp. 27–38.
- Csapó, T. G., Németh, G., Cernak, M., Garner, P. N. 2016. Modeling Unvoiced Sounds In Statistical Parametric Speech Synthesis with a Continuous Vocoder, EUSIPCO 2016, beküldve.
- Drugman, T., Raitio, T. 2014. Excitation Modeling for HMM-based Speech Synthesis: Breaking Down the Impact of Periodic and Aperiodic Components, Proc. ICASSP 2014, pp. 260–264.
- Zen, H., Tokuda, K., Black, A. W. 2009. Statistical parametric speech synthesis, Speech Communication, 51(11), pp. 1039–1064.