Milan Sečujski, Darko Pekar, Dragan Knežević, Vladimir Svrkota
Faculty of technical sciences, Novi Sad

Prosody prediction in speech synthesis based on regression trees

The paper examines the possibility of automatic prediction of prosodic parameters, namely, f0 contour and phonetic segment duration, which are of utmost importance for synthesis of highly intelligible and natural speech. Within this research, f0 contours and durations of phonetic segments in specific phonetic and prosodic contexts are predicted using regression trees trained on a speech corpus containing approximately 4 hours of previously recorded and annotated speech in the Serbian language. The speech corpus consists of utterances by a single female speaker whose voice is used for speech synthesis as well. Since it is known that both phonetic and prosodic context of a particular speech segment affect its f0 contour and duration, the speech database was annotated with both phonological and phonetic markers (phonemic identity and specific information related to the manner of articulation) as well as markers related to prosody (types and positions of phrase breaks and sentence focus). In cases of phones whose articulation consists of more than one phonetically distinct phase (such as occlusion and explosion of stops or vocalic and non-vocalic segments of the vibrant R), annotation was carried out on a sub-phonemic level.

The regression tree method is based on creating tree structures based on the samples in the training set (training phase) and using them for prediction of the behaviour of an unknown parameter whose value is known for all samples in the training set. Having been trained, the tree is able to predict an unknown quantity in a sample that does not correspond to any of the samples in the training set, but is drawn from a population with the same statistical distribution. By estimating the duration of phonetic segments in particular contexts using regression trees (as is the case with any other regression-based prediction) one avoids the need for explicit mathematical modelling of the influence of relevant linguistic factors, which is known to be an extremely complex problem. Beside actual values of segment durations, the method also offers insight into the complex way that particular linguistic factors interact when affecting the behaviour of fundamental frequency of speech and its rhythm.

Breiman et al. 1984: L. Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone. *Classification and Regression Trees*. Chapman & Hall/CRC, Boca Raton, London, New York, Washington D.C.

van Santen et al. 1990: J. P. H. van Santen, J. Olive. *The Analysis of Contextual Effects on Vowel Duration*, Computer, Speech & Language, Elsevier, 359–390.

Campbell et al. 1991: N. Campbell, S. Isard. *Segment Durations in a Syllable Frame*, Journal of Phonetics, Elsevier, 19: 37–47.

van Santen 1992: J. P. H. van Santen. *Contextual Effects on Vowel Duration*, Speech Communication, Elsevier, 11: 513–546.

van Santen 1993: J. P. H. van Santen. *Timing in Text–to–Speech Synthesis*. EUROSPEECH, Berlin.

Dutoit 1997: T. Dutoit. *An Introduction to Text–to–Speech Synthesis*. Dordrecht/Boston/London: Kluwer Academic Publishers.

Hardcastle et al. 1999: W. J. Hardcastle, J. Laver. *Handbook of Phonetic Sciences.* Wiley–Blackwell.

Sečujski et al. 2002: M. Sečujski, R. Obradović, D. Pekar, Lj. Jovanov, V. Delić. *AlfaNum System for Speech Synthesis in Serbian Language.* TSD, Brno, 237–244.

Delić et al. 2006: V. Delić, M. Sečujski, D. Pekar, N. Jakovljević, D. Mišković. *A Review of AlfaNum Speech Technologies for Serbian, Croatian and Macedonian*. IS–LTC, Ljubljana, 257–260.

Delić 2007: V. Delić. *A Review of R&D of Speech Technologies in Serbian and Their Applications in Western Balkan Countries*. SPECOM, Moskva, 64–83.

Lazaridis et al. 2007: A. Lazaridis, P. Zervas, N. Fakotakis, G. Kokkinakis. *A CART Approach for Duration Modeling of Greek Phonemes*. SPECOM, Moskva, 287–292.

Sečujski et al. 2009: M. Sečujski, A. Kupusinac. *Određivanje položaja akcenta u govornoj bazi podataka primenom stabala odluke.* INFOTEH, Jahorina, Bosna i Hercegovina.