

Ein quantitatives Tonmodell für Ibibio

Entwicklung eines Prädiktionsmoduls für das BOSS-Sprachsynthesesystem

Arne Bachmann

May 10, 2007

Abstract

In this Master's Thesis, a simple generic model for tone language synthesis is proposed and applied to an Ibibio corpus. While most parametric models (like *Tilt* and *PaIntE*) were designed for accent languages, or allowed for multiple solutions of their parameters (*Fujisaki*), the model proposed here consumes arbitrary tonal shapes while considering the tonological rules of the language. Additionally a software intonation module was developed to apply this model to the *Bonn Open Synthesis System* BOSS. Starting from theoretical considerations, tone contours are stylized syllable by syllable through polynomials. The whole corpus of those polynomial shapes is reduced by a simple vector quantization algorithm to a codebook of e. g. 64 codevectors called *tone templates*. By using about 20 common and easily extractable parameters from the foregoing modules, a classification tree predicts the likeliest tone template for each syllable. A standard unit selection algorithm selects segmental units with the best fitting contour shapes from the corpus. Thus the synthesis system aims to produce lexically and grammatically correct speech intonation. In this work, the problems and details leading to design decisions are discussed, the accuracy of the stylization, vector quantization and prediction are presented. Some proposals are made for further development of the model and software. The accompanying programming lead to the BOSS-IBB extension V 0.1 currently being integrated as module package V 0.2 into BOSS V 3.1.

Literature

Tilt model: Taylor (1998, 2000), Dusterhoff und Black (1997); Dusterhoff et al. (1999); Syrdal et al. (1998). PaIntE model: Möhler (2001); Möhler und Conkie (1998); Syrdal et al. (1998). Fujisaki model: Fujisaki (1988, 2004); Fujisaki et al. (1984); Fujisaki und Nagashima (1969); Fujisaki und Ohno (1996); Fujisaki et al. (1996); Gu et al. (2004); Hirose et al. (1983); Möbius et al. (1993). BOSS: Bachmann (2007); Breuer et al. (2005).

Vector quantization and LBG-algorithm: Linde et al. (1980), Möhler und Conkie (1998). Classification and regression trees (CART): Breiman et al. (1984), Lewis (2000); Tutz (2005).

References

- Bachmann, A. (2007): *BOSS-IBB. Speech Synthesis Module Documentation for the BOSS Ibibio module*. University of Bonn.
- Breiman, L.; J. H. Friedman; R. A. Olshen; C. J. Stone (1984): *Classification and Regression Trees*. New York: Chapman and Hall (Wadsworth, Inc.).
- Breuer, S.; P. Wagner; J. Abresch; J. Bröggelwirth; K. Stöber (2005): *Bonn Open Synthesis System (BOSS) Documentation and User Manual*. version 3.0 Aufl., <http://www.ikp.uni-bonn.de/dt/forsch/phonetik/boss/>.
- Dusterhoff, K.; A. W. Black (1997): "Generating F0 contours for speech synthesis using the tilt intonation theory." In: *Proceedings of the 1997 ESCA Workshop on intonation*, Athens, Greece.
- Dusterhoff, K.; A. W. Black; P. Taylor (1999): "Using Decision Trees within the Tilt Intonation Model to Predict F₀ Contours."
- Fujisaki, H. (1988): "A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour." In: Fujimura, O., Hg., *Vocal Physiology: Voice Production, Mechanisms and Functions*, New York: Raven Press, Kap. 30, 165–175.
- Fujisaki, H. (2004): "Information, Prosody, and Modeling – with Emphasis on Tonal Features of Speech." *Speech Prosody 2004* .
- Fujisaki, H.; K. Hirose; N. Takahashi; M. Yoko'o (1984): "Realization of accent components in connected speech." In: *Transactions of the Committee on Speech Research*, Acoustic Society of Japan, Bd. S84, 279–286.
- Fujisaki, H.; S. Nagashima (1969): "A model for the synthesis of pitch contours of connected speech." In: *Annual Report of the Engineering Research Institute*, University of Tokyo, Bd. 28, 53–60.
- Fujisaki, H.; S. Ohno (1996): "Prosodic Parameterization of Spoken Japanese Based on a Model of the Generation Process of F0 Contours." In: *Proceedings of the ICSLP*, Philadelphia, PA, Bd. 4, 2439–2442.
- Fujisaki, H.; S. Ohno; O. Tomita (1996): "On the levels of accentuation in spoken japanese." In: *Proceedings of the ICSLP*, Philadelphia, PA, Bd. 2, 634–637.

- Gu, W.; K. Hirose; H. Fujisaki (2004): “A Method for Automatic Tone Command Parameter Extraction for the Model of F_0 Contour Generation for Mandarin.” *Speech Prosody 2004* .
- Hirose, K.; H. Fujisaki; M. Yamaguchi (1983): “Automatic estimation of characteristic parameters of fundamental frequency contours.” In: *Reports of the 1983 Spring Meeting*, Acoustic Society of Japan, Bd. 1, 93–94.
- Lewis, R. J. (2000): “An Introduction to Classification and Regression Tree (CART) Analysis.”
- Linde, Y.; A. Buzo; R. M. Gray (1980): “An algorithm for vector quantizer design.” In: *IEEE transactions on communications*, Bd. 28, 84–95.
- Möbius, B.; M. Pätzold; W. Hess (1993): “Analysis and synthesis of German F_0 contours by means of Fujisaki’s model.” In: *Speech Communication*, Bd. 13, 53–61.
- Möhler, G. (2001): “Improvements of the PaIntE model for F_0 parametrization. Technical report (draft version).”
- Möhler, G.; A. Conkie (1998): “Parametric modeling of intonation using vector quantization.”
- Syrdal, A. K.; G. Möhler; K. Dusterhoff; A. Conkie; A. W. Black (1998): “Three methods of intonation modelling.”
- Taylor, P. A. (1998): “The tilt intonation model.” In: *Proceedings of the ICSLP*, 4, Sydney, 1383–1386.
- Taylor, P. A. (2000): “Analysis and Synthesis of Intonation using the Tilt Model.” *Journal of the Acoustical Society of America* 107(3), 1697–1714.
- Tutz, G. (2005): “CART – Classification and Regression Trees.” Folien-Präsentation.