Advanced Signal Processing Seminar SS 2010

Signal Processing in Text-to-Speech Synthesis

Harald Romsdorfer

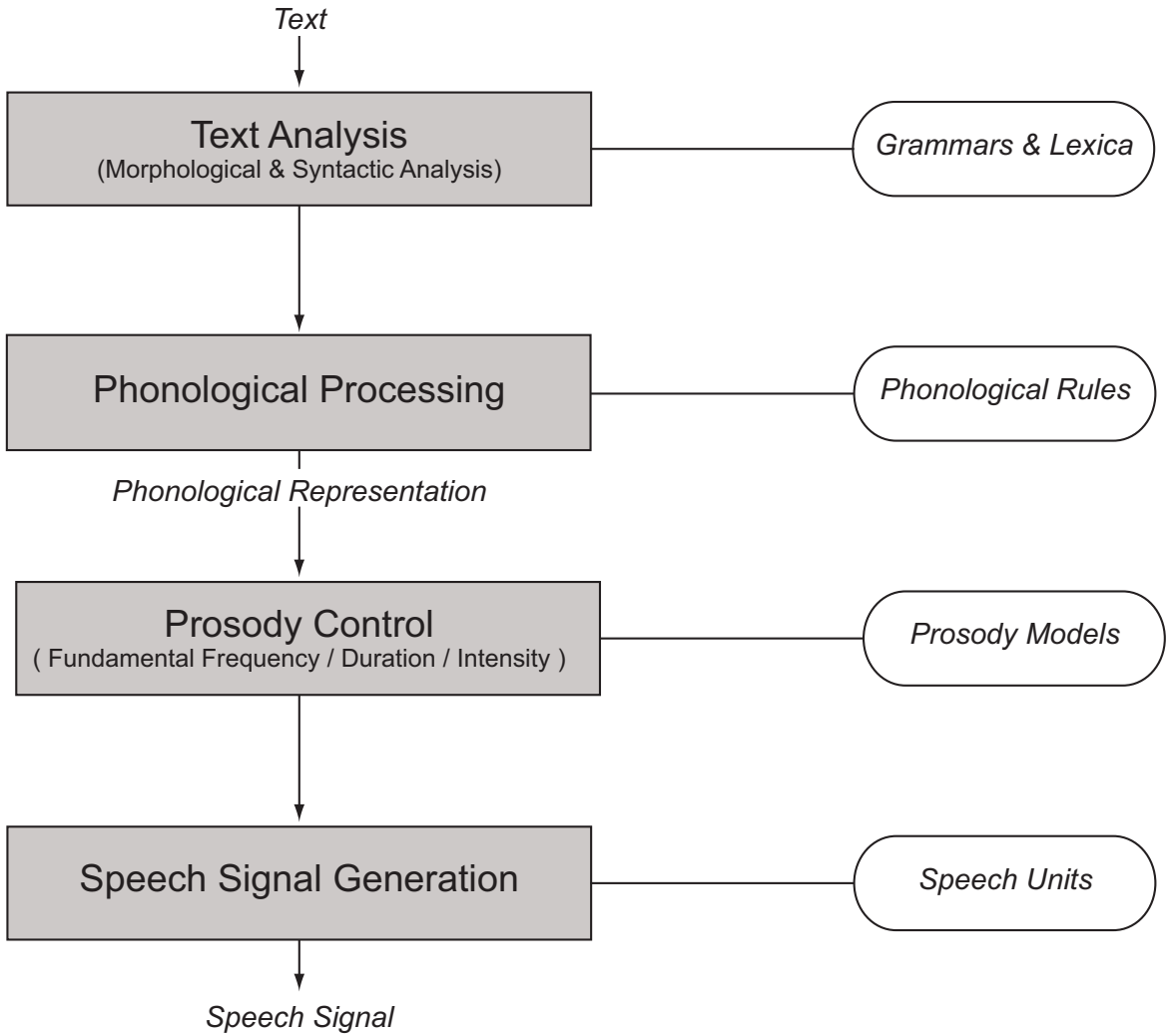
Signal Processing and Speech Communication Laboratory

16.3.2010

Administrative Information

- **Instructor:** Harald Romsdorfer
- **Meeting Date/Time:** to be defined
- **Mode:** Each group of 1-2 students should select one topic. They should give an in-depth presentation of this topic and the referenced work therein (for about 1 - 1.5 hours). Following the presentation, we would like to discuss the presented topic. Therefore, it is necessary that each participant reads the presented article.
- **Grading:** Grades are given based on the presentation and the participation in discussions (50% :: 50%). The presentation slides must be sent before the presentation to romsdorfer@tugraz.at.

Text-to-Speech Synthesis



Topics

- Analysis of Speech Signals (Feature Extraction)
  - STFT, MFFC, RASTA, PLP, LSF
  - Cepstral Smoothing
  - Phone Clustering
- Speech Corpus Segmentation
  - HMM-based Forced Alginment
  - DTW-based Pattern Matching

  [YEH+02, Hos08]
- Prosodic Modification of Speech Signals
  - Source-Filter Models: LPC, Spectral Coeffs (STRAIGHT)
  - TD-PSOLA
  - FD-PSOLA

  [MC90, KEF01]

Topics

- Prosody Generation
  - F0 Modeling
  - Segment Duration Modeling
  - Intensity Modeling

  [BH96, YKK08, Rom09]
- Concatenative Speech Synthesis
  - Diphone Synthesis
  - Unit Selection Synthesis
  - Source-Filter Synthesis: LPC, Fourier, MFCC

  [BC95, HB96, vS97, Dut08]
- Statistical Parametric Speech Synthesis
  - HMM-based Speech Synthesis: MFCC, LPC

  [Yam06, Tay09]

Topics

- Voice Transformation Algorithms
  - Statistical Parametric Speech Synthesis
  - Concatenative Speech Synthesis

  [KM98, KM01, Yam09]
- Language Transformation Algorithms
  - Statistical Parametric Speech Synthesis
  - Concatenative Speech Synthesis

  [LIF05, LIF06]
- Polyglot Speech Synthesis
  - Foreign Inclusion Detection
  - Polyglot Speech Synthesis: Concatenation-based, HMM-based

  [SO94, BL04, BS06, Rom09, LIF05]

Books

T. Dutoit.

Corpus-based speech synthesis.
In J. Benesty, M. Sondhi, and Y. Huang, editors, Handbook of Speech Processing, chapter 23, pages 471–487. Springer, Berlin Heidelberg, 2008.

P. Taylor.

Text-to-Speech Synthesis.
Cambridge University Press. 2009.

S. Young, G. Evermann, T. Hain, D. Kershaw, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, and P. Woodland.
The HTK Book.
Cambridge University Engineering Departement, Cambridge, 2002.

Bibliography

A. W. Black and N. Campbell.

Optimising selection of units from speech databases for concatenative synthesis.
In Proceedings of Eurospeech'95, pages 581–584, Madrid, Spain, September 1995.

A. W. Black and A. Hunt.

Generating F_0 contours from the ToBI labels using linear regression.
In Proceedings of ICSLP'96, pages 1385–1388, 1996.

A. W. Black and K. A. Lenzo.

Multilingual text-to-speech synthesis.
In Proceedings of the ICASSP 2004, Montreal, Canada, 2004.

A. W. Black and T. Schultz.

Speaker clustering for multilingual synthesis.
In ISCA Tutorial and Research Workshop on Multilingual Speech and Language Processing (MultiLing 2006), Stellenbosch, South Africa, April 2006.

A. Hunt and A. Black.

Unit selection in a concatenative speech synthesis system using a large speech database.
In Proceedings of ICASSP'96, pages 373–376, Atlanta, Georgia, USA, 1996.

J.-P. Hosom.

Speaker-independent phoneme alignment using transition-dependent states.
Speech Communication, 2008.

Bibliography

H. Kawahara, J. Estill, and O. Fujimura.

Aperiodicity extraction and control using mixed mode excitation and group delay manipulation for a high quality speech analysis, modification and synthesis system STRAIGHT.
In Proceedings of 2nd MAVEBA, pages 13–15, Firenze, Italy, September 2001.

M. Kain and M. W. Macon.

Spectral voice conversion for text-to-speech synthesis.
In Proceedings of ICASSP 1998, pages 285–288, 1998.

M. Kain and M. W. Macon.

Design and evaluation of a voice conversion algorithm based on spectral envelope mapping and residual prediction.
In Proceedings of ICASSP 2001, Salt Lake City, Utah, USA, May 2001.

J. Latorre, K. Iwano, and S. Furui.

Cross-language synthesis with a polyglot synthesizer.
In Proceedings of Interspeech 2005, pages 1477–1480, Lisbon, Portugal, September 2005.

J. Latorre, K. Iwano, and S. Furui.

New approach to polyglot synthesis: How to speak any language with anyone's voice.
In ISCA Tutorial and Research Workshop on Multilingual Speech and Language Processing (MultiLing 2006), Stellenbosch, South Africa, April 2006.

E. Moulines and F. Charpentier.

Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones.
Speech Communications, 9(5–6):453–467, December 1990.

Bibliography

H. Romsdorfer.

Polyglot Text-to-Speech Synthesis. Text Analysis and Prosody Control.
PhD thesis, No. 18210, Computer Engineering and Networks Laboratory, ETH Zurich (TIK-Schriftenreihe Nr. 101), January 2009.

R. W. Sproat and J. T. Olive.

A modular architecture for multi-lingual text-to-speech synthesis.
In Proceedings of ESCA/IEEE Workshop on Speech Synthesis, pages 187–190, New Paltz, New York, September 1994.

J. P. H. van Santen.

Combinatorial issues in text-to-speech synthesis.
In Proceedings of Eurospeech'97, pages 2511–2514, Rhodes, Greece, September 1997.

J. Yamagishi.

An introduction to HMM-basd speech synthesis.
Technical report, Tokyo Institute of Technology, October 2006.

J. Yamagishi.

Thousands of voices for HMM-based speech synthesis.
In IEEE Audio, Speech, & Language Processing, 2009.

J. Yamagishi, H. Kawai, and T. Kobayashi.

Phone duration modeling using gradient tree boosting.
Speech Communication, 50(5):405–415, May 2008.