1 Course Syllabus

- **Weeks 1-2: Introduction** (16.10.03 – 23.10.03)
  Topics: acoustics, digital signal processing, TTS architecture
  Reading: (Dutoit, 1997, Chapters 1–3)
- **Weeks 3-4: Morpho-Syntactic Processing** (30.10.03 – 06.11.03)
  Topics: Tokenization, PoS-Tagging, Chunking, Parsing.
  Reading: (Dutoit, 1997, Chapter 4), Church (1988); Brill (1992); Roche and Schabes (1995)
- **Weeks 5-7: Phonetization and Prosody** (13.11.03 – 27.11.03)
  Topics: Phonetic transcription, Prosody generation.
  Reading: (Dutoit, 1997, Chapters 5–6), Klatt (1987), Liberman and Church (1992), van den Bosch and Daelemans (1993), Naval Research Laboratory (1976) ...
- **Week 8: Concept-to-Speech Synthesis** (04.12.03)
  Topics: Prosody generation, interface requirements.
  Reading: Youd and Fallside (1989), Pan and McKeown (1997), ...
- **Weeks 9-12: Speech Synthesis** (11.12.03 – 15.01.04)
  Topics: Formant-, concatenative-, articulatory-, spectral-synthesis, LPC, PSOLA.
  Reading: (Dutoit, 1997, Chapters 7–10), Black and Taylor (1994), Klatt (1980), Holmes et al. (1964), ...
- **Week 13-14: Systems and Synthesizers** (29.01.04 – 05.02.04)
  Topics: rsynth, festival, mbrola, txt2pho, mary
  Reading: Ing-Simmons (1994), Black and Taylor (1997); Taylor et al. (1998)

2 Administrivia

2.1 General Information

Organizer: Bryan Jurish (`moocow@ling.uni-potsdam.de`)
Course Web Site: www.ling.uni-potsdam.de/~moocow/class/spsyn
Office: II.24.152
Telephone: (0331) 977-2641
Office Hours: Thursdays, 14:00 – 15:00, or by appointment

2.2 Grading Policies

Proseminar: *Referat*
Hauptseminar: *Referat & Ausarbeitung*

Students interested in presenting a *Referat* should arrange a topic with me, and meet with me at least one week before the presentation to discuss the topic.

3 Introduction

3.1 What is TTS?

- **Task:** Text input, spoken audio output.
- **Related Tasks:** Restricted-domain TTS, Concept-to-Speech Synthesis (CTS)
- **Applications:**
  - Aids to persons with disabilities
  - Automatic translation (speech-to-speech)
  - Telecommunications services
  - Entertainment
  - *etc.*

3.2 Acoustics

- Speech is sound.
- Sound waves are fluctuations in (air) pressure.
- Some acoustic phenomena:
  - Periodic signals
    - Frequency (*F0*): pitch
    - Amplitude (peak, peak-to-peak, RMS): intensity, volume
    - Phase (radians, degrees, periodic fractions)
    - Spectral envelope (timbre)
      - Peaks: formants
      - Troughs: antiformants
  - Aperiodic signals (noise)
  - Filters (low-pass, high-pass, band-pass)
  - Audition (logarithmic)
- Some acoustic characteristics of speech:
  - Periodic signals: vowels, sonorants
  - Aperiodic signals: fricatives, transients

3.3 Digital Signal Processing

- A digital audio signal is a stream of discrete amplitude values (*samples*)
  - A sample represents instantaneous signal amplitude (air-pressure, voltage) at a particular point in time (and space).
  - Characteristics of a sample stream:
    - Sampling rate (aka sampling frequency)
    - Sample width (aka sample size, range, quantization factor)
  - Phenomena:
    - Aliasing
    - Clipping
    - Sample-width noise

3.4 TTS Architectures

- **Symbolic NLP**
  - Input: Raw electronic text.
  - Tasks: Tokenization, PoS tagging, morphological analysis, parsing, semantic analysis, phonetification, prosody generation, *etc.*
  - Output: narrow phonetic transcription
- **Digital Signal Processing (DSP)**
  - Input: narrow phonetic transcription
  - Tasks: Database lookup, prosody matching, parametric speech modelling, signal generation.
  - Output: audio signal.

References

G. Bailly and C. Benoit, editors. *Talking Machines: Theories, Models, and Designs*. North-Holland, Amsterdam, 1992.

A. Black and P. Taylor. CHATR: a generic speech synthesis system, 1994. URL `citeseer.nj.nec.com/black94chatr.html`.

A. Black and P. Taylor. Festival speech synthesis system: system documentation. Technical Report HCRC/TR-83, University of Edinburgh, Centre for Speech Technology Research, 1997. URL `http://www.cstr.ed.ac.uk/projects/festival`.

E. Brill. A simple rule-based part-of-speech tagger. In *Proceedings of ANLP-92, 3rd Conference on Applied Natural Language Processing*, pages 152–155, Trento, IT, 1992. URL `http://citeseer.nj.nec.com/brill92simple.html`.

K. W. Church. A stochastic parts program and noun phrase parser for unrestricted text. In *Proceedings of the 2nd Conference on Applied Natural Language Processing*, pages 136–143, 1988.

T. Dutoit. *An Introduction to Text-to-Speech Synthesis*. Kluwer, Dordrecht, 1997.

G. Fant. *Acoustic Theory of Speech Production*. Mouton, The Hague, 1960.

J. N. Holmes, I. Mattingly, and J. Shearme. Speech synthesis by rule. *Language and Speech*, 7:127–143, 1964.

N. Ing-Simmons. rsynth version 2.0, 1994. URL `ftp://svr-ftp.eng.cam.ac.uk/pub/comp.speech/synthesis/rsynth-2.0.tar.gz`.

D. Klatt. Review of text-to-speech conversion for english. *Journal of the Acoustical Society of America*, 82(3):737–793, 1987.

D. H. Klatt. Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, 67(3):971–995, 1980.

M. J. Liberman and K. W. Church. Text analysis and word pronunciation in text-to-speech synthesis. In S. Furui and M. M. Sondhi, editors, *Advances in Speech Signal Processing*. Dekker, New York, 1992.

Naval Research Laboratory. Automatic translation of english text to phonetics by means of letter-to-sound rules. Technical Report 7948, Naval Research Laboratory, Washington, D.C., 1976.

S. Pan and K. R. McKeown. Integrating language generation with speech synthesis in a concept to speech system. 1997.

E. Roche and Y. Schabes. Deterministic part-of-speech tagging with finite-state transducers. *Computational Linguistics*, 21(2):227–253, 1995. URL `http://citeseer.nj.nec.com/roche95deterministic.html`.

P. Taylor, A. Black, and R. Caley. The architecture of the the festival speech synthesis system. In *Third International Workshop on Speech Synthesis, Sydney, Australia, November 1998*, 1998. URL `citeseer.nj.nec.com/article/taylor98architecture.html`.

A. van den Bosch and W. Daelemans. Data-oriented methods for grapheme-to-phoneme conversion. In *Proceedings of the 6th Conference of the EACL*, pages 45–53, 1993. URL `ftp://ilk.kub.nl/pub/antalb/eacl-93.ps.gz`.

N. J. Youd and F. Fallside. Driving a speech synthesizer from conceptual input in the context of a voice dialogue system. In *Proceedings of Eurospeech '89*, pages 514–517, Paris, 1989.