Lecture "Foundations: Speech Synthesis"
Dr. M. Schröder

Due: Friday 2005-12-23

Exercise

1. Generate the expression "happy new year" [h { p i n u: j I r=] from diphones.
   a) Which diphones are needed?
   b) Generate a list of carrier words from which the required diphones can be taken.

2. The SSML tag for emphasising words is:
   `<emphasis level="strong|moderate|none|reduced">`
   Use it to annotate the sentence:
   "No, I want the blue car!"
   such that it is an appropriate answer to the question: "So you want the red car?"

3. An experimental psychologist wants to investigate the role of different acoustic/prosodic/vocal parameters on emotion perception. He wants to use speech synthesis to create the stimuli; his requirements are that the technology should "yield highest naturalness and allow for full control over voice quality and prosody". He asks for your advice on the choice of synthesis technology to use – what will you tell him?

4. Prosody prediction. In many languages, content words are often accented while function words are not. Which module within a TTS system typically provides that information? Can you propose a minimalistic algorithm for implementing that module for a new language?

5. A TTS system pronounces the string "It does cost 29000 Euros." wrong – the output sounds like:
   "It doe's cost two-nine-zero-zero-zero oo-rose."
         (1)   (2)   (3)
   Explain the three errors: In which module within the TTS system did each of them occur? How could the error be fixed?