

Semester / Master Thesis:

Voicing Modelling in HMM-Based Speech Synthesis

Background: The research on speech synthesis is focused on how to make computers read arbitrary text as naturally as people do. The hidden Markov model (HMM)-based speech synthesis framework functions by training an HMM for each sound unit, connecting appropriate HMMs according to given text, and finally generating speech with the HMM sequence. HMM-based speech synthesis is very popular nowadays due to its flexibility.

Motivation & Task Description: There are two kinds of sound when people speak: *voiced* (big, cat, etc when your vocal cords vibrate) and *unvoiced* (pig, fast, etc when your vocal cords do not vibrate). The current HMM-based framework involves a *voiced-or-unvoiced* prediction model for each sound unit, such that we can know whether a sound to be synthesised should be made voiced or unvoiced. Unfortunately, due to the sophisticated training process of HMMs, it is not guaranteed that every voiced-or-unvoiced prediction model works properly when synthesising speech. Without a doubt, prediction errors degrade the quality of synthesised speech by causing audible unpleasant sound. In this task, we will be looking for an approach to improve the current voiced-or-unvoiced prediction mechanism.



Requirements: Some basic knowledge of speech processing would be of great help. Programming skills are a must.

In case of interest, you may contact us for more details about this particular task and have a look at <http://www.tec.ethz.ch/research.html> for more information about the entire research group.

Contacts

- Hui LIANG (梁晖): liangh@tik.ee.ethz.ch, ETZ D97.4
- Beat Pfister: pfister@tik.ee.ethz.ch, ETZ D97.6



[link to this document](#)