Real-time talking head animation and face analysis research at MARCS Auditory Laboratories (UWS) and the Institute for Cognitive Systems (TUM)

Takaaki Kuratate, Institute for Cognitive Systems, TU Muenchen

Abstract:
   In this talk, I will present an overview of talking head animation and 3D face analysis research running at MARCS Auditory labs, University Western Sydney, Australia and collaborative research at ICS, Technical University Munich, Germany.

   At MARCS Auditory labs, directed by Prof. Denis Burnham, various auditory-visual speech research themes are examined along with other speech and music related research. Since 2007, MARCS lab has been the leading institute of the "Thinking Head" project funded by the Australian Research Council and National Health and Medical Research Council Special Initiatives. This project involves multiple universities in Australia and overseas partners in Europe and North America, and has significantly expanded computer engineering-related research themes at MARCS lab, as they work to develop a plug-and-play modular face animation system.   Also MARCS lab has multiple 3D face and speech articulator measuring devices, such as a 3D face scanner which can continuously record 3D face data at 60Hz (3dMD's 4D System), 3D magnetometer systems (EMA, Wave system), and motion capture systems (OPTOTRAK, VICON), helping to keep them as one of the top auditory-visual speech research institutes in the world.

   Also at Technical University Munich, the Institute for Cognitive Systems (ICS) headed by Prof. Gordon Cheng was recently started, including the Social Robotics Communication lab, which collaborates with MARCS labs and the Intelligence System Research Institute at AIST, Japan.   The talking head system developed at MARCS lab is combined with the OpenHRI system developed at AIST. This combined system can recognize speech input both in English and Japanese and synthesize speech response together with face animation. The preliminary system which is setup as a simple command-and-action system which answers with pre-defined responses by text-to-speech synthesis with talking head animation according to the specific speech input, will be extended to add a basic conversational module for further human-robot/computer interaction research.