Course Title	Speech Processing (Advanced)						
Instructor(s)	Hiroshi G. Okuno, Tatsuya Kawahara, Shinobu Masaki, Tetsuya Ogata, Ryuichi Nishimura						
Assigned Grade		Units	2	Semester	Fall semester	Time	Tue 4
Course Category & Course Type	Specialized Lecture	Language		English			

Course Description (overview, purpose)

This lecture elaborates on issue in sound signal processing from theories to applications.

First, we explain the basic theories and the speech-auditory mechanisms. The physical properties of sound generation are clarified, then, the speech generation mechanism (i.e. source filter theory) and novel findings obtained from recent brain imaging technologies. Furthermore, we explain the techniques of digital signal processing based on human's auditory mechanism.

Second, we explain the basic features of sound signals and the sound analysis techniques. The outlines of sound encoding, voice synthesis and sound recognition is also shown. Concerning sound recognition, we expound the representative acoustic models and language models.

Third, we explain computer auditory scene analysis (CASA) including environmental and musical sounds. The spatial perception using binaural theory is investigated and recognition function integrating visual-audio sensory input is explained. Moreover, we introduce some applications such as the robot audition technologies and the developmental imitation models of phoneme acquisition from the viewpoint of cognitive robotics.

Course Schedule

・Sound/speech signal, Brain (Masaki): Physics of sounds, speech articulation mechanism, brain imaging and speech recognition function
・Human's audition and the applications (Nishimura): Digital signal processing based on human auditory function
・Speech analysis/recognition/synthesis (Kawahara): Information in speech signal, linearly separated equivalent circuit model, Speech analysis/synthesis/coding, recognition, acoustic model, language model
・CASA and robot audition (Okuno): Computer scene analysis, sound localization, sound separation, music information processing, animal acoustics
・Multimodal processing and cognitive robotics (Ogata): Phoneme acquisition/imitation model and multimodal mapping model using neuro-dynamical system

Prerequisites and Course Requirements

Grading Methods and Evaluation Criteria

Grading will be determined by submitted reports; the questions will be given by lecturers.

Textbooks

References

・Moore, B.C.J.: "An Introduction to the Psychology of Hearing", 4th Ed., Academic Press, 1997.
・Bregman, A.: "Auditory Scene Analysis" (MIT Press, 1990)
・Rosenthal, D. and Okuno, H.G. (eds.): "Computational Auditory Scene Analysis" (Lawrence Erlbaum Associates, 1998)

Miscellaneous (homework assignment, office hours etc.)

・{okuno, kawahara,, ogata}@i.kyoto-u.ac.jp, masaki@atr.jp, ryou@nict.go.jp
・The order of lecture is subject to change.
・Reports should be submitted through e-mail.