

授業科目名	音声情報処理特論						
英語名	Speech Processing (Advanced)						
担当教員名	奥乃 博, 河原 達也, 正木 信夫, 尾形 哲也, 西村 竜一						
配当学年		単位数	2	開講期	後期	曜時限	火 4
授業種別・ 授業形態	専攻専門科目 講義			授業言語	英語		
<b>【授業の概要・目的】</b>							
<p>本講義では、音について基礎理論から最先端の音の認識技術とその応用について学ぶ。</p> <p>第1部では、音の基礎理論と音声生成機構および聴覚機構について述べる。まず音声に関わる音の物理特性を明かにし、音声生成の理論（ソース・フィルタ理論）と近年の観測技術の進歩に基づく新たな知見について述べる。聴覚機構については、人間の聴覚特性について明かにした上で、それらの特徴を利用したデジタル信号処理技術について紹介する。</p> <p>第2部では、音声の基本的な性質と特徴量について述べ、代表的な音声分析手法を解説する。音声符号化、音声合成、及び音声認識についてその概要を述べる。音声認識においては、代表的な音響モデル・言語モデルについて解説する。</p> <p>第3部では、音声に限らず楽音や環境音を含む音環境理解について述べ、両耳聴の原理から空間的な知覚の原理を探り、視覚と聴覚の情報統合による認知機能について説明する。また、具体的な応用としてロボット聴覚の課題と現状、音声模倣発達についても取り上げる。</p>							
<b>【授業計画と内容】</b>							
<ul style="list-style-type: none"> <li>・音・音声・脳（正木）：音の物理と音声生成機構、脳機能計測と音声情報処理機構</li> <li>・ヒトの聴覚と応用技術（西村）：人間の聴覚機構、音声生成・聴覚特性に基づくデジタル信号処理</li> <li>・音声の分析・認識・合成（河原）：音声に含まれる情報、線形分離等価モデル、音声分析、音声合成・音声符号化・音声認識の概要、音声認識のための音響モデル言語モデル</li> <li>・音環境理解とロボット聴覚（奥乃）：音環境理解とは、音源定位、音源分離と分離音認識、音楽情報処理、動物音響学</li> <li>・マルチモーダル情報処理とロボットによる情報生成（尾形）：人工神経回路による音声模倣発達モデル、人工神経回路による視聴覚変換モデル</li> </ul>							
<b>【履修要件】</b>							
<b>【成績評価の方法・基準】</b>							
5回のレポート試験を中心に、授業中の発表の評価、質疑応答などを勘案して、到達目標の達成度を総合的に判断する。							
<b>【教科書】</b>							
配布資料を使用							
<b>【参考書等】</b>							
<ul style="list-style-type: none"> <li>・板橋秀一編：『音声工学』（森北出版, 2005）</li> <li>・Moore, B.C.J.: "An Introduction to the Psychology of Hearing", 4th Ed., Academic Press, 1997. (大串監訳 『聴覚心理学概論（第3版）』（誠信書房), 1994)</li> <li>・Bregman, A.: "Auditory Scene Analysis" (MIT Press, 1990)</li> <li>・Rosenthal, D. and Okuno, H.G. (eds.): "Computational Auditory Scene Analysis" (Lawrence Erlbaum Associates, 1998)</li> </ul>							
<b>【その他（授業外学習の指示・オフィスアワー等）】</b>							
<ul style="list-style-type: none"> <li>・電子メール：{okuno, kawahara,, ogata}@i.kyoto-u.ac.jp, masaki@atr.jp, ryou@nict.go.jp</li> <li>・講師の都合により、一部順序の変更がありうる。</li> <li>・レポートは講師に直接電子メールで提出すること。</li> </ul>							

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		
<b>Course Description (overview, purpose)</b>							
<p>This lecture elaborates on issue in sound signal processing from theories to applications.</p> <p>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.</p> <p>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.</p> <p>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.</p>							
<b>Course Schedule</b>							
<ul style="list-style-type: none"> <li>• Sound/speech signal, Brain (Masaki): Physics of sounds, speech articulation mechanism, brain imaging and speech recognition function</li> <li>• Human's audition and the applications (Nishimura): Digital signal processing based on human auditory function</li> <li>• Speech analysis/recognition/synthesis (Kawahara) : Information in speech signal, linearly separated equivalent circuit model, Speech analysis/synthesis/coding, recognition, acoustic model, language model</li> <li>• CASA and robot audition (Okuno): Computer scene analysis, sound localization, sound separation, music information processing, animal acoustics</li> <li>• Multimodal processing and cognitive robotics (Ogata): Phoneme acquisition/imitation model and multimodal mapping model using neuro-dynamical system</li> </ul>							
<b>Prerequisites and Course Requirements</b>							
<b>Grading Methods and Evaluation Criteria</b>							
Grading will be determined by submitted reports; the questions will be given by lecturers.							
<b>Textbooks</b>							
<b>References</b>							
<ul style="list-style-type: none"> <li>• Moore, B.C.J.: "An Introduction to the Psychology of Hearing", 4th Ed., Academic Press, 1997.</li> <li>• Bregman, A.: "Auditory Scene Analysis" (MIT Press, 1990)</li> <li>• Rosenthal, D. and Okuno, H.G. (eds.): "Computational Auditory Scene Analysis" (Lawrence Erlbaum Associates, 1998)</li> </ul>							
<b>Miscellaneous (homework assignment, office hours etc.)</b>							
<ul style="list-style-type: none"> <li>• {okuno, kawahara, ogata}@i.kyoto-u.ac.jp, masaki@atr.jp, ryou@nict.go.jp</li> <li>• The order of lecture is subject to change.</li> <li>• Reports should be submitted through e-mail.</li> </ul>							