報告番号	※甲　　第　　号

主　論　文　の　要　旨

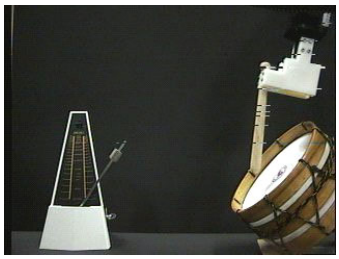
論文題目　　視聴覚事象の対応付けとそれに基づく概念の獲得

氏　　名　　西堀　研人

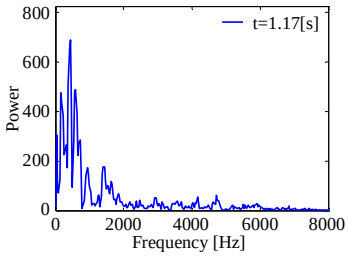
論　文　内　容　の　要　旨

人間は，周囲の環境に対して様々な働きかけを行い，物体の認識やシーンの理解を行っている．人間の感覚器官として，特に重要な視聴覚によって知覚される事象の対応付けは，人が生活する上で必要不可欠なものであり，これまでも多くの研究がなされてきた．人が視聴覚事象を対応付けるとき，経験や知識を用いる他に，赤ん坊のように経験や知識を用いず，生得的に備わった物理的な法則を理解する機能を用いていることが考えられている．また，幼児が対象に働きかけ，対象に対する理解を促進していることが報告されている．本研究では，視聴覚事象の対応付け手法の拡張として，物体操作と注意による対応付けを行い，また対応付いた視聴覚事象を蓄積し，その統計的な関係から概念を獲得する手法を提案する．以下にその手法について述べる．

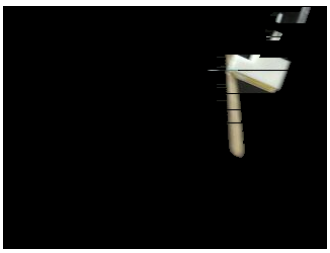
物体操作による視聴覚事象の対応付けについて提案する．物体操作による視聴覚事象の対応付けでは，物体固有の情報ではなく，一般的な法則（ゲシュタルトの群化の法則）を用いる．物体を操作しようとする脳から筋への信号の変化と，そのとき，視聴覚によって観測される音の変化および映像の変化の“同時性”，物体操作と視聴覚事象の間における繰り返しの“類似性”を手掛かりとして対応付ける．実験では，ロボットマニピュレータを用いてドラムをばちでたたいたり，ベルを振ったりして視聴覚事象を発生させ，マイクロフォンとカメラを用いて観測し，物体操作運動と視聴覚事象の対応付けを行った．聴覚処理では，対象音検出のため物体操作運動とは関係のない時刻から雑音を推定し，除去する．その後，混合音での音オンセット検出において，周波数バンドごとに分け，適応的に検出する．検出した音オンセットのスペクトル同士の類似性から音源ごとの音オンセット時系列を得る．一方，視覚処理では，物体の運動範囲を推定し，運動範囲ごとに得られる映像の重心軌跡から物体の運動方向変化の時系列を得る．統合処理により，運動と視聴覚事



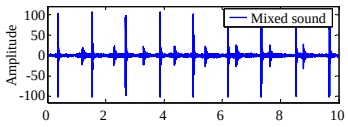
(a1) シーン



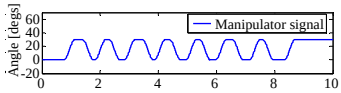
(b1) 音源 1 の周波数成分



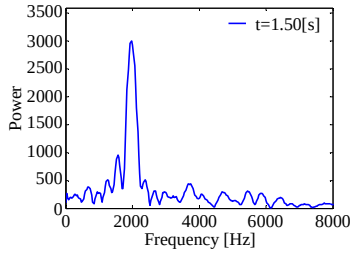
(c1) 音源 1 に対応付いた画像



(a2) 混合音信号



(a3) 肘関節角度の制御信号



(b2) 音源 2 の周波数成分

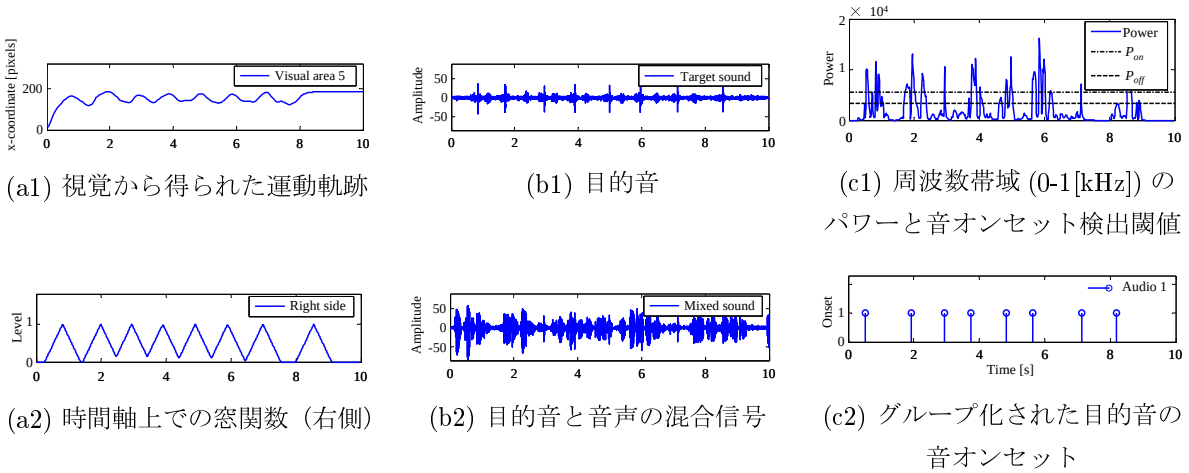
(c2) 音源 2 に対応付いた画像

図 1: 運動範囲と視聴覚事象の対応

象時系列を同一のサンプリング周期に統一し，事象間の類似性を比較する．相関の高い事象同士をグループ化することで，運動によって生じた視聴覚事象について対応付ける．

図 1 は，実験の結果得られた視覚における運動範囲と視聴覚事象の対応を示す．(a1) は，カメラで観測したシーンであり，(a2) は，マイクロフォンで計測した混合音信号であり，(b1) と (b2) のように 2 つの音源グループに分けられた．(c1) と (c2) は音源 1 と音源 2 に対応付いた画像である．(a3) は，ロボットマニピュレータの肘関節角度を制御する値である．(a3) と相関が高い (b1) の音と (c1) の画像が物体操作によって発生した事象であることがわかった．

次に，選択的注意による視聴覚事象の対応付け手法を提案する．人間は，目的に応じて視覚と聴覚を組み合わせ，周囲の環境から必要な情報を選択的に取得し，感覚器からの大量の情報を瞬時に処理している．聴覚的注意は，視覚において動きを知覚した場合に，その動きがある時点に聴覚の注意を集中し，雑音環境において目的音を検出することとした．実験では，目的音に音声やホワイトノイズを雑音として付加し，その混合音からの目的音の検出を時間軸上での窓関数を用いて行った．時間軸上での窓関数は，視覚における物体の運動方向が変化した時刻を中心に作成した，三角窓のフィルタである．**図 2** に，視聴覚情報を用いた音オンセット検出について示す．(a1) は，視覚から得られた対象物の運動軌跡であり，運動方向が変化する時刻から (a2) の時間軸上での窓関数を作る．(b1) は，運動によって発生した目的音であり，それに (b2) のように音声を雑音として混合した．(b2) に (a2) の時



(a1) 視覚から得られた運動軌跡　(b1) 目的音　(c1) 周波数帯域 (0-1 [kHz]) のパワーと音オンセット検出閾値

(a2) 時間軸上での窓関数（右側）　(b2) 目的音と音声の混合信号　(c2) グループ化された目的音の音オンセット

図 2: 視覚情報を用いた音オンセット検出

間軸上での窓関数を通過させた後，(c1) のように周波数帯域において，音オンセットを検出する．検出した音オンセットの音源グループと，(a1) の運動方向変化の時刻の相関が高いものから，(c2) の目的音の音オンセット時系列が得られる．視聴覚事象の対応付けの成功率は，時間軸上での窓関数を用いないときの 78.6% から，用いたときの 95.2% へ上昇した．

視覚的注意は，聴覚において音が知覚されるが，視覚情報が得られない場合に，音源方向への頭部の回転や，視線方向と明るさの調整により，音に対応する運動を探すこととし，実験を行った．**図 3** は，視覚的注意により検出した対象物を示す．(a)〜(c) にかけて，薄暗い穴の中にカメラの中心を向け，明るさの調節を行う．(d1) の中央部の矩形領域のヒストグラムは (d2) であり，(e2) のように線形変換する．これによって，(e1) のようにコントラストが高くなり，物体領域が抽出しやすくなる．(f1) の音源 1 に対応付いた画像として (f2) が得られた．視聴覚事象の対応付けでは，成功率は 93.3% であり，本研究の有効性が示された．

最後に，視聴覚事象の対応付けに基づく概念の獲得の手法を提案する．人は，視聴覚事象を対応付けし，事例として蓄積したものを体系化し，概念として獲得している．これらの処理を工学的に実現することを目標とした．視聴覚事象についての概念を自律的に獲得するため，対応付けられた音と画像の事例の集合を用い，正準相関分析により両者の統計的関係を学習する．音と画像の特徴ベクトルを正準空間へ写像し，K-means 法による類別を行い，概念を獲得する．概念を獲得した後に，カテゴリ未知の入力事例に対応する概念の，同じモダリティ（視覚，聴覚）における中心的な事例を提示することを事例の識別とした．また，カテゴリ未知の入力事例に対応する概念の，異なるモダリティにおける中心的な事例を提示することを事例の想起とした．

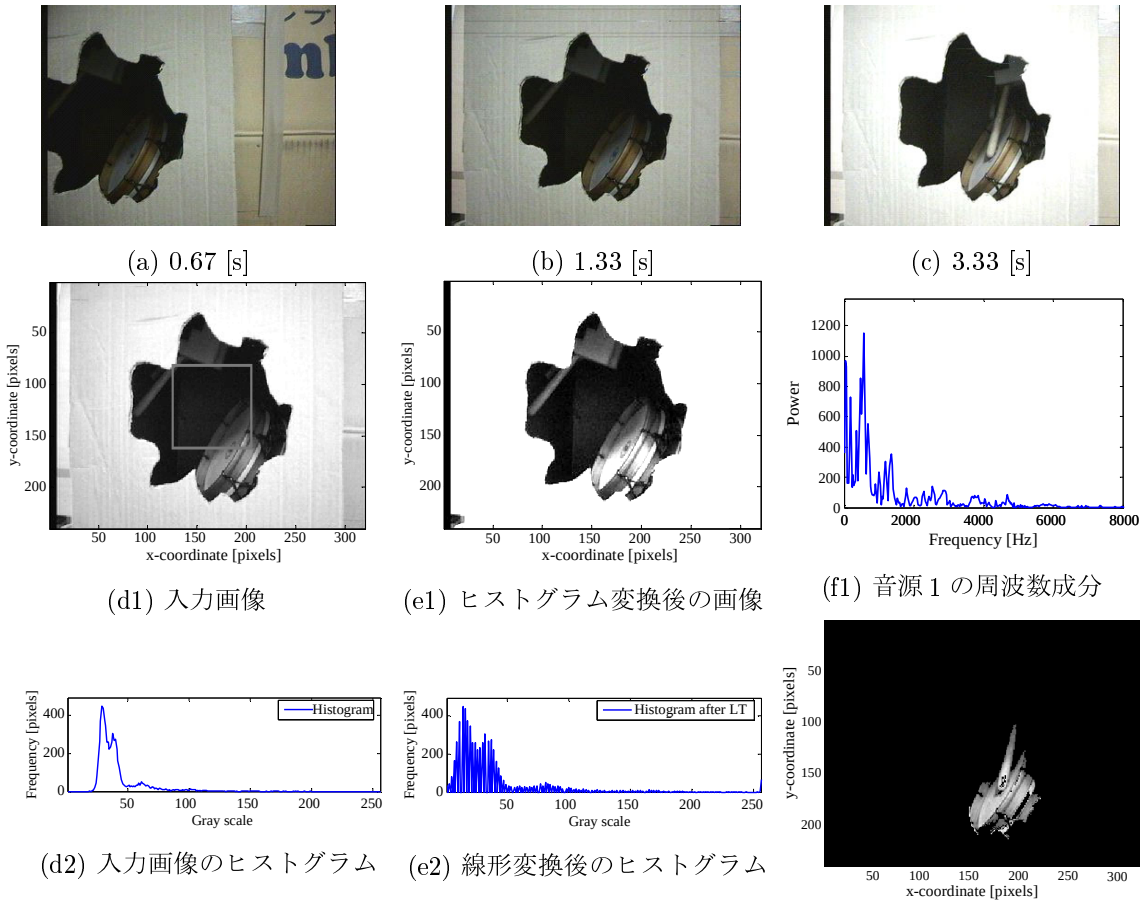


図 3: 視覚的注意により検出した対象物

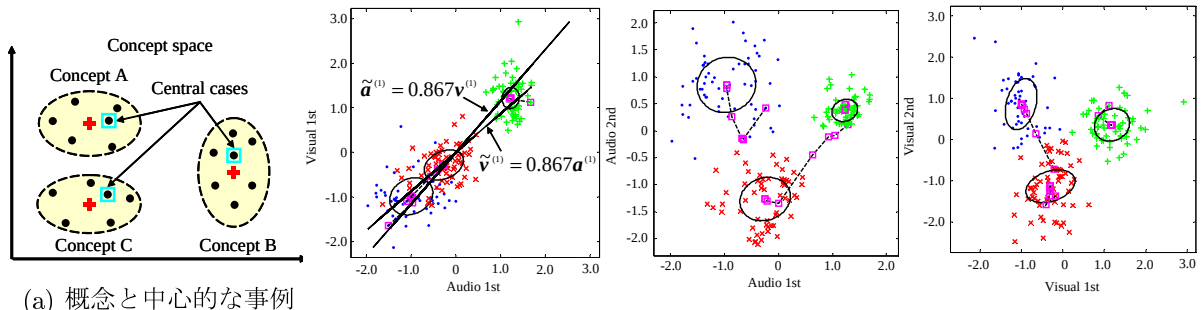
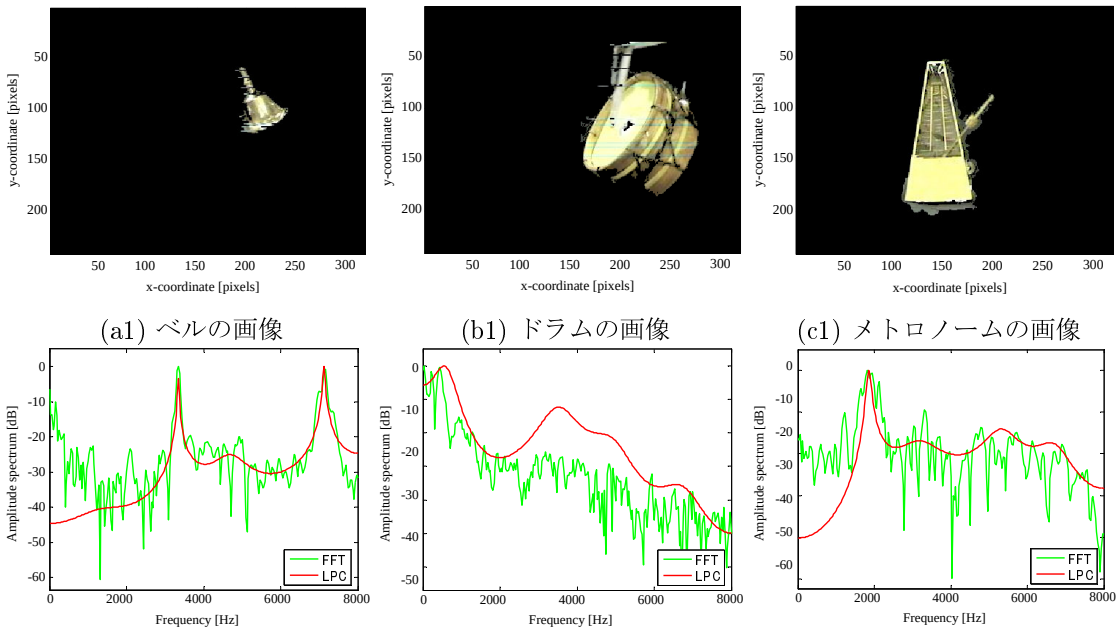


図 4: 概念と中心的な事例の関係，およびカテゴリ未知の場合における特徴分布



(a1) ベルの画像　(b1) ドラムの画像　(c1) メトロノームの画像

(a2) ベル音の周波数成分　(b2) ドラム音の周波数成分　(c2) メトロノーム音の周波数成分

図 5: 視聴覚事象の中心的な事例

図 4 は，概念と中心的な事例の関係，およびカテゴリ未知の場合における特徴分布である．(a) に示すように，概念空間において，“ • ”が事例，“ + ”が獲得された概念のクラス中心であるとき，中心的な事例とは，クラス中心から最も距離が近い事例である．概念は，共通の特徴を持つ事例の集合体からなり，中心的な事例を概念に属する最も典型性の高い事例として表現する．ベル (図中の青色 •)，ドラム (赤色 ×)，メトロノーム (緑色+) の視聴覚事象を，事前知識なしで正準空間においてクラスタリングした結果は，(b) の視聴覚成分，(c) の聴覚成分，(d) の視覚成分となり，**図 5** の視聴覚事象の中心的な事例を得た．未知の視聴覚入力を，獲得した概念のいずれかに識別する成功率は 83.2% 以上であった．図 4(b) の視聴覚成分間の回帰直線を用いて，未知の視覚または聴覚入力に対応する概念の，聴覚または視覚の事例として想起する成功率は 81.5% 以上であり，本手法の有効性を確認した．

今後の課題として，以下の事柄が挙げられる．物体操作による視聴覚事象の対応付けでは，実時間での対応付けや視聴覚事象が不良であるときに，物体操作を手掛かりとして，視聴覚事象間の対応付けを行うことが考えられる．注意による視聴覚事象の対応付けでは，視聴覚事象が両方とも不良な状況において，視聴覚の相互作用によって対応付け精度が向上することが望まれる．また，概念の獲得研究については，まず本手法とニューラルネットワーク等の他手法で概念を獲得した場合との成功率の比較を行うことや，より多くのカテゴリにおける概念獲得実験を行うことが課題である．また，ロボットシステムにおいて，概念の追加的な獲得や，概念の階層構造の構築を行うことが，より複雑な知識の獲得のために必要と考えられる．