4E－11

物体の運動とそれに伴う音との対応付け

早川和宏* 向井利春‡ 大西昇†‡

*名古屋大学大学院工学研究科電子情報学専攻
†名古屋大学大学院工学研究科情報工学専攻
‡理化学研究所，BMC 研究センター

1 はじめに

人間は，複数の人が歩いているシーンにおいて，足音が誰の足の動きによるものかを視覚的に捉えることができる．すなわち，物体の運動とそれに伴う音との対応付けが可能である．この際用いられる対応付けの手掛かりの一つとして，経験に基づく予測が挙げられる．例えば，ヒールの足音ならばおそらく歩いている人は女性であろうといった予測（前情報）を持ち，視覚的に探索するというものである．

しかしながら，人間はこのような経験的知識を得る以前の段階においても，運動と音との対応付けを行うことができる．この時の手掛かりは，経験的に蓄積されている物体固有の情報ではなく，一般的に成立する物理法則（テンポや速度）であると考えられる[1]．

本研究では，このような一般的法則を手掛かりとして運動と音の対応付けを工学的に実現する手法[2][3]について述べる．

2 運動と音の関係

物体の運動とそれに伴う音との関係は**表 1** のように整理される．本研究では表 1 の網掛け部分を対象とし，特に周期性のある運動に着目する．

表 1: 運動と音の関係

		物体全体の移動あり	物体全体の移動なし
部分的な動作あり	発する音あり	足音をたてて歩く人	拍手
	発する音なし	水槽を泳ぐ魚	指揮者の手
部分的な動作なし	発する音あり	坂道を転がる石	ベルが鳴っている電話
	発する音なし	机上を滑べる氷	置物の人形

3 計測

異なる動作物体を二つ配置し，一方は動作に伴って音を発し，もう一方は動作のみとする．このようなシーンをカメラ一台，マイク一本によって動画および音を記録する．**図 1** は画面左でメトロノームを約 3[Hz]，右上でペンを約 1[Hz] で左右に振っているシーンを記録したものである．

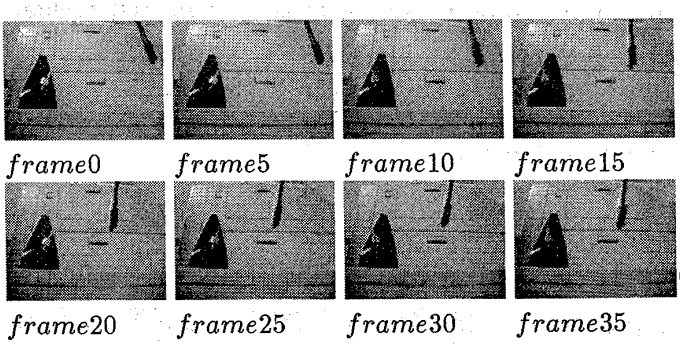


図 1: 計測したシーン

4 処理方法

計測された動画（記録時間は約 5 秒間）から画像処理部と音処理部で，それぞれについて周波数を求め，それらを対応付けの手掛かりとする．

4.1 画像処理部

1. 動画は 30[frames/s] で記録され，画像サイズは 320×240[pixels]，gray level は 8[bits] である．

Finding Correspondence Between Object Motions and Their Sounds
Kazuhiro Hayakawa*, Toshiharu Mukai‡, Noboru Ohnishi†‡
*Dept. of Information Electronics,Nagoya University
†Dept. of Information Eng.,Nagoya University
‡RIKEN BMC Research Center

2. 図 **2** のように，一枚の画像を 16 の部分画像（80×60[pixcels]）に分割し，それぞれの部分画像について画素値の総和をその時刻の値として持つ時系列を作成する．

3. 各時系列をフーリエ変換することにより，部分画像毎に運動物体の周期に対応した周波数が計算できる（図 **3**）．

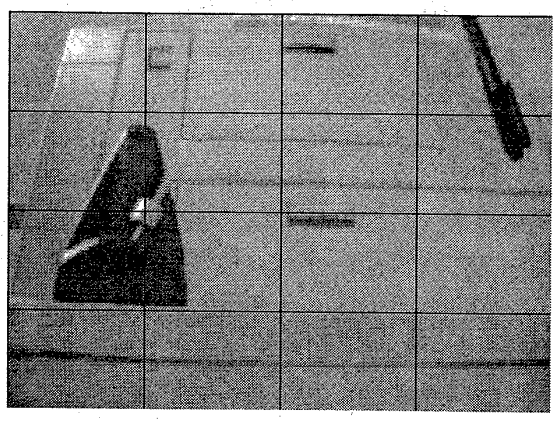
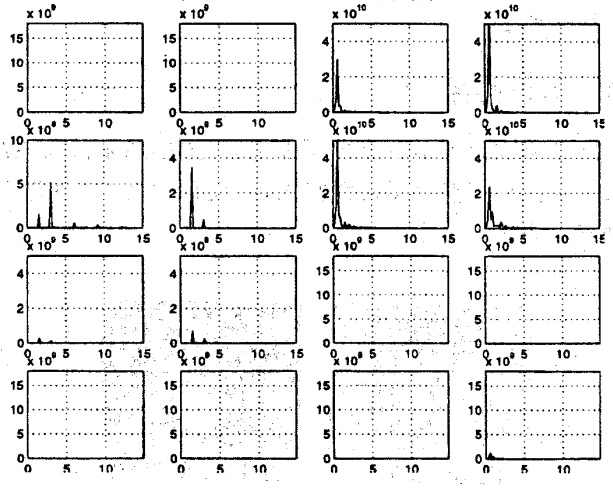


図 2: シーンの分割



横軸：周波数，縦軸：パワースペクトル密度

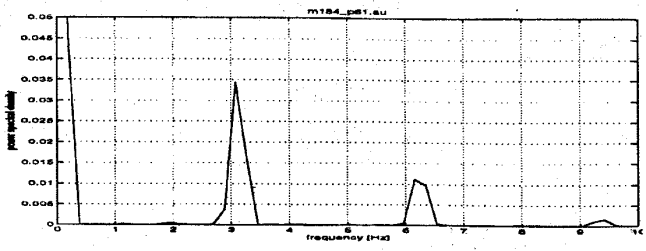
図 3: 部分画像毎の周波数

4.2 音処理部

1. 音はモノラル 44.1[KHz] で記録される．

2. 画像と音のサンプル数の関係から，音の値の絶対値をとった後 1470 点ずつ区切り，その中の最大値を一枚の画像に対応する音の値とする．

3. フーリエ変換することにより，音の周波数（テンポ）を得ることができる（図 4）．



横軸：周波数，縦軸：パワースペクトル密度

図 4: 音の周波数

5 実験結果

図 3 では画面左側で約 3[Hz]，右側で約 1[Hz] のピークが見られ，また図 4 では約 3[Hz] のピークが見られる．これらのことから，音を発しているのは左側の運動物体，すなわちメトロノームであることがわかる．

6 まとめ

周期的な運動に着目し，周波数を手掛かりに運動と音の対応付けが可能であることを示した．

今後の課題としては，部分画像の分割数により運動物体によっては半分の周波数が現れてしまうため効率的な分割方法を検討すること，また，運動物体が重なり合っているなど複雑な状況にも適応できるように拡張することが挙げられる．

参考文献

[1] P.K.Kuhl and A.N.Meltzoff: “The Bimodal Perception of Speech in Infancy,” Science, Vol. 218, pp. 1138-1141, 1982.

[2] T.Mukai and N.Ohnishi: “Grouping Corresponding Parts in Vision and Audition Using Perceptual Grouping among Different Sensations,” IEEE/SICE/RSJ International Conference on Multisensor Fusion and Intergration for Intelligent System, pp. 713-718, 1996.

[3] 早川，向井，大西: “信号の同期性に基づく視聴覚情報の対応付け,” 電気関係学会東海支部連合大会講演論文集,p301,Sep,1997.