

音による空間情報の表現に関する研究

立命館大学 ○鹿田 敦史, 櫻井 隆平, 李 周浩

The Research on Sound based Spatial Information Expression

○Atsushi SHIKATA, Ryuhei SAKURAI and Joo-Ho LEE, Ritsumeikan University

Abstract: Human achieves much information from vision rather than other mode. Next to vision we achieve lots of information from audito. If we can get visual information from audio it might be very useful. This paper proposes a method that transforms visual information into audio information. The method correlates a camera image of user viewpoint to sounds, and the user is able to recognize spatial information from the sounds.

1. 緒言

近年、メディア変換に関する研究が盛んに行われている。特定の感覚器の利用が制限されている状況下において、メディア変換を用いればその感覚器への情報を他の感覚器の情報に変換し伝達することが可能となる。視覚から聴覚へのメディア変換では、視覚障害者補助はもちろん、画像や光景を注視することなく視覚情報を取得することができる。

視覚から聴覚へのメディア変換を利用した従来研究として、全方位カメラを用いて全方位画像を取得し、輝度値をもとに音信号を得る手法[1]、同様に全方位画像内の多数の画像特徴を音楽特徴に対応付け音楽を生成する手法[2]がある。これらは画像と音を信号レベルで対応付けを行うため、直接的ではあるが抽象的な変換となってしまふ。本研究では視覚情報を聴覚情報へ変換する手法として、画像内の物体の位置情報などを音に変換し、対象者の視野方向に基づいた空間情報の提示を提案する。

2. 音による空間情報の提示

2.1. 概要

本研究では、ステレオカメラを用いて画像を取得し画像中の物体の大きさや距離という情報を音で提示する。このカメラ画像を人の視野と捉え、人物の正面の画像 Fig.1(a)を取得する。視野方向に対応した情報を提示するため、その画像を縦に分割し、それぞれの領域内で最も手前にある物体を Fig.1(b)で示されるように検出し、その情報を音に変換する。これにより、Fig.1(c)のように人が実際に見えている環境の情報を提示することができ、対象者と物体との位置関係の把握が可能となる。実験では、本手法に基づき複数の環境の画像から生成した音を被験者に聞いてもらい、音からイメージした空間と画像を対応付けられるかの検証を

行う。

2.2. 物体の検出

画像情報を音情報に変換する際、まず縦に分割したカメラ画像のそれぞれの領域ごとの最も手前の物体の大きさと距離を取得する。1つの領域に対し横に細分化を行いブロックに分け、画面下から順にブロック内の物体の情報を取得する。あるブロック内で物体が検出されると、実世界での水平面上の座標が同じデータを持つブロックまで走査し、物体の高さを算出する。これにより縦に区分された視野ごとに、利用者から見て最も近い位置にある物体を検出することができる。

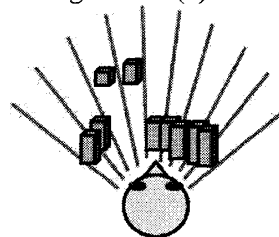
2.3. 画像情報から音情報

検出された物体の大きさは音を提示する時間に、距離は音量に対応付ける。音の周波数は基準値から時間ごとに上昇させる。それぞれの領域は視野の左右方向に対応しているため、情報の提示には左右のヘッドフォンの位相差と音量差を利用する。提示される音は、視野の左から右に向かって定位を変えながら、存在する物体の距離とその大きさを表す。



(a) Original image

(b) Detection image



(c) Person viewpoint

Fig.1 Detection of object in eyeshot

Table 1 Verification items

検証項目	
1	音の提示時間
2	物体の位置や大きさの正確性
3	全体的な聞きやすさ, 空間の把握しやすさ

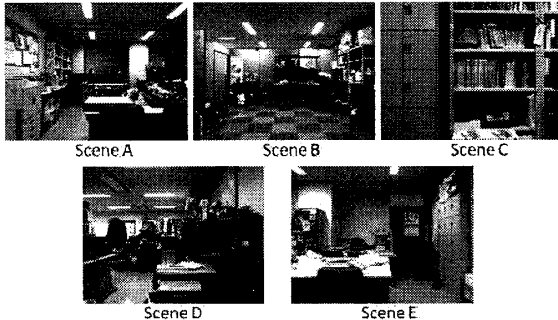


Fig.2 Five images used for an experiment

3. 実験

本研究では, ステレオカメラ(Bumblebee2, ViewPLUS 社)を用いた。

3.1. 実験内容

3.1.1. 視野の分割数

画像(視野)を縦に分割する際に最適な分割数を検証した。正面からの情報提示は左右の定位の基準となるため, 検証する分割数は奇数とした。3人の被験者に対し本研究の手法を説明し2種類の環境を撮影した画像と4種類の分割数で生成した音を提示し, Table 1に示す3項目について10段階で評価してもらった。

3.1.2. 空間の認識

13人の被験者に対し, 画像情報を音情報に変換していることを説明し, 物体の大きさと距離に応じてどのような音が提示されるかを示したサンプル音を聞かせた。その後, Fig.2で示す5つの環境で撮影した画像から生成した音だけを聞いてもらい, それぞれの音が表す空間をイメージしてもらった。そして5つの環境の画像を見せ, それぞれの音がどの場所を表わしているかを選択してもらった。

3.2. 実験結果

3.2.1. 視野の分割数

各分割数における評価の平均値はTable 2のようになった。5分割では物体検出の精度が低くなってしまい, 9分割や11分割になると音の提示時間が長くなってしまったため, 7分割が最適であると結果を得た。

3.2.2. 空間の認識

実験の結果, それぞれの環境についての再現率, 適合率をTable 3に示す。平均再現率は77%となり, 本研究の手法を用いて変換した音から基本的

Table 2 Value about division number of image

	画像の分割数			
	5	7	9	11
項目1	4.7	5.5	4.4	3.1
項目2	5.5	8.3	7	8.3
項目3	3.7	3.7	3.8	3.7
total	4.6	5.8	5	5

Table 3 Result about spatial recognition

	A	B	C	D	E	recall	precision
A	10	0	0	3	0	77%	77%
B	1	11	1	0	0	85%	85%
C	0	0	10	1	2	77%	77%
D	2	1	0	9	1	69%	69%
E	0	1	2	0	10	77%	77%

な空間の認識はできていると考えられる。最も認識率が高かったのは空間の広がりを感じられるScene Bだった。また, 最も多かった誤りはScene AをScene Dと認識したものである。これは, 音量により物体までの距離感を認識した際に, 空間の形状が非常に似ていたため誤認識が増えたと考えられる。これらから被験者は空間内に配置された物体の大きさよりも距離感といった空間の広がりを重視し, 空間を認識したものであると思われる。つまり本研究で示した手法を用いることによって, 被験者視点での空間の形状把握には十分に効果があることが確認できた。

4. 結言

本研究では音による空間情報の表現として, ステレオカメラを用いて画像中の物体の位置や大きさを検出し, 周波数, 振幅, 左右の位相に対応付け音を生成し提示を行い, 音から元の空間をイメージすることができるか検証を行った。実験により, 本研究で提案した手法を用いて変換した音を聞くことにより, 元となった空間の物体の位置関係などを認識できることを実証した。今後の課題としては, 画像中の輝度や色相などの視覚情報を音に変換し付加することや, 現行ではステレオカメラを人の視野として捉えていたが環境内に配置されたカメラから人の視野情報を取得することなどがあげられる。

参考文献

- [1] 港隆史, 関戸智史, 石黒浩, 河原英紀: “全方位視覚の特性を利用した画像から音信号への変換”, 日本ロボット学会学術講演会予稿集, 2002.
- [2] 池田徹志, 室田健吾, 石黒浩: “全方位映像から音楽情報へのメディア変換に基づく視覚情報の伝達”, 情報処理学会論文誌, vol.48, No.1, pp274-pp283, 2007.