

コース	基盤科学コース
学籍番号	070225
氏名	岸野嵩久
指導教員	ルジェロ・ミケレット

Our purpose is to develop a new human interface for the visually impaired. We have designed several algorithms to translate an image into a sound. Then, we have demonstrated that a figure drawn on an image can be recognized using our sounds.

## 1. 研究背景と目的

視覚障害者のハンディキャップの解消や軽減を目指したインターフェース開発は従来から進められてきた。特に文字認識の分野では多くの研究が行われている。一方で、情報変換が容易な文字情報以外の、図形などの形の情報を変換することも重要な課題となる。図形情報を視覚以外の情報に変換する方法として、触覚ディスプレイなど、視覚の情報を触覚の情報に置き換える手法が多く取り入れられている。しかし、これらは図形を認識するために特殊なディスプレイが必要である。また、同時に複数の図形を認識することができず、認識に時間が必要になる。そこで、本研究では特殊なデバイスを必要とせず、複数の図形の同時認識を可能とし認識に必要な時間を最小限に抑えられる「音」への変換に着目した。特殊なデバイスを必要としない分、元の図形を認識しやすい音に変換するアルゴリズムが必要となる。

以上より、本研究では視覚障害者のためのインターフェースの開発を目的とした、「視覚の情報」を「聴覚の情報」に変換するアルゴリズムの設計を行った。このアルゴリズムでは学習効果による認識効率と応用性の向上を利用する。設計したアルゴリズムを用いて、変換された音から元の画像に描かれた図形を認識することができるかを検討した。

## 2. 方法

本研究では、情報変換の過程でコンピュータによる特定の形や文字の検出を行うような認識処理は行わない直接的な変換方法を用いた。このように信号に近いレベルで直接的に情報変換をすることで、認識誤りや決まった形へのパターン化をなくし、変換過程での情報の欠損を最小限に抑えることができる。

図形を音へ変換する方法として、図形の輪郭線を抽出しその情報を波形データに変

換した。波形データへの変換方法は①明度のグラフ化、②フーリエ変換、③位置に周波数を割り当てる、という異なる3種類の方法を試用した。

### 3. 結果

設計したアルゴリズムで変換した音の認識テストを行った。すべての認識テストは5人の晴眼者に対して行った。

初めに、変換された図形の音が元の図形の形状に依存して異なることを確認するための実験を行った。その結果、いずれの変換方法においても、図形ごとに変換された音が異なっていると認識できた割合は95[%]以上であった。したがって、図形の違いを音の違いとして被験者が識別できることを示した。

変換した音から図形を認識することは、輪郭線の変化を変換した音から認識することに等しい。すなわち、音によって線の変化をどの程度の分解能で認識できるかにより、図形を認識できる精度が決まる。そこで、直線について注目し、変換した音から直線の傾きをどの程度の精度で認識できるか調査する実験を行った。3種類の変換方法のひとつである「周波数割り当て」による結果では、被験者達が音から認識した直線の傾きは、実際の直線の角度との差は平均7.69[度]であった。また、この角度の差と被験者の回答時間はFig.1に示すように、テスト回数を重ねるごとに減少した。経験を積むことで学習効果によって認識精度の向上が期待できる。以上より、音から認識される直線の傾きの精度は高く、図形の直線的な輪郭線の変化を認識することが十分に可能であることが示唆された。

そこで、変換アルゴリズムを実際に図形に適用し、図形を変換した音から元の図形を連想する実験を行った。被験者による個人差が見られたものの、変換した音から図形を連想することは可能であった(Fig.2)。しかし、図形によっては完全な連想が困難な傾向も見受けられた。

### 4. まとめ

図形を直接的に音に変換するアルゴリズムを設計し、認識実験によりその認識効率の評価を行った。結果として、図形を直接的に音から元の図形を認識することは可能であることが示された。今後は、認識効率の向上を目指すために、ステレオ効果を用いるなどして変換アルゴリズムの改良を行う。

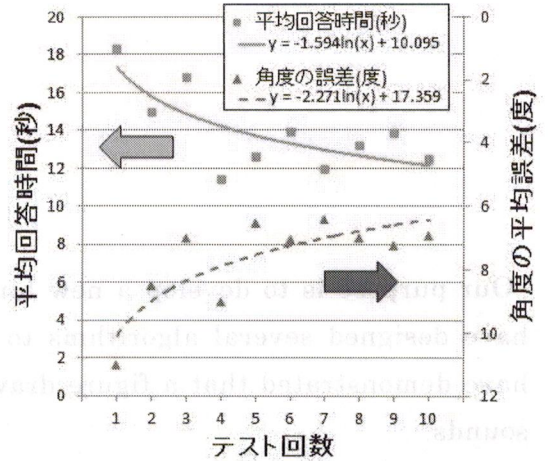


Fig.1: 「周波数割り当て」で認識した直線の角度の誤差(点線)と回答時間(実線)

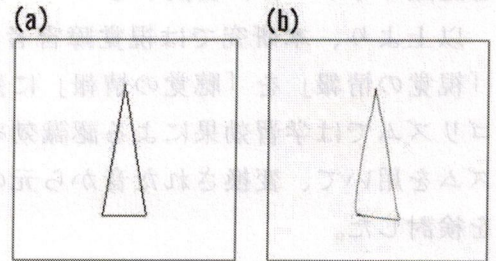


Fig.2 「周波数割り当て」で音に変換された図形(a)、音から被験者が連想して描いた図形(b)

<卒業論文>

視覚情報の音への変換と  
その認識に関する基礎的研究

横浜市立大学 国際総合科学部

国際総合科学科 基盤科学コース ミケレット研究室

070225 岸野 嵩久

# 目次

第1章 序論.....	1
1.1 はじめに.....	1
1.2 研究背景.....	2
1.2.1 クロスモーダル知覚.....	2
1.2.2 感覚情報の変換.....	2
1.2.3 視覚と聴覚.....	3
1.3 研究目的.....	4
第2章 画像の音への変換.....	5
2.1 はじめに.....	5
2.2 人間の視覚における形認識のメカニズム.....	6
2.3 音への変換アルゴリズム.....	8
2.3.1 画像の音への写像.....	8
2.3.2 波形データへの変換方法.....	9
第3章 予備実験.....	13
3.1 はじめに.....	13
3.2 実験1：形による音の違いの認識.....	13
3.2.1 目的.....	13
3.2.2 方法.....	14
3.2.3 結果と考察.....	16
3.3 実験2：音の印象からの形の連想.....	19
3.3.1 目的.....	19
3.3.2 方法.....	19
3.3.3 結果と考察.....	20
第4章 音の学習と解析.....	24
4.1 はじめに.....	24
4.2 実験1：直線の角度の学習とその音の解析.....	25

4.2.1	実験目的	25
4.2.2	方法	25
4.2.3	結果と考察	27
4.3	実験 2：図形の学習とその音の解析	30
4.3.1	実験目的	30
4.3.2	方法	30
4.3.3	結果と考察	33
<b>第 5 章 総合考察</b>		<b>38</b>
5.1	まとめ	38
5.2	今後の課題	39
5.3	応用的意義	40
<b>参考文献</b>		<b>41</b>
<b>謝辞</b>		<b>42</b>

# 第1章 序論

## 1.1 はじめに

近年、コンピュータ及び情報処理技術の進歩はめまぐるしく、様々なインターフェースは我々にとってより使いやすいものになった。しかしながら、それらのほとんどは3D映像やタッチパネルといった視覚を対象としたものである。これらと比較すると、盲人のような視覚に障害を持つ者、視力にハンディを持つ者に対するインターフェースの技術は未だ十分とは言い切れない。

視覚障害者の抱えるハンディキャップの解消や軽減を目指したインターフェースの開発は従来から進められてきた[1-4]。それらは主に、歩行誘導や文字認識についての研究であり、特に文字認識の研究は多く行われている[5,6]。文字はあらかじめパターンが決められているものであり、それゆえに認識が行い易い。また、文字自体に読み方があることから、使用者が文字の読み方(文字の名前)を知っているのならば文字自体を認識させることも容易である。これらのことから、視覚的に文字を認識し、その情報変換は容易であるといえる。

一方で、我々の日常生活では文字だけでは伝えきれない情報が存在する。そのひとつが図形などの「形」についての情報である。身の回りにある障害物を認識する際などには、形の情報を認識することが重要である。しかし、形に関する情報は文字のように決まったものではない。また、あらかじめ決められたパターンに分類することは困難である。これらのことから、「形」の情報を変換し認識させることはいまだ実用段階にあるものは少なく、重要な課題となっている。

## 1.2 研究背景

### 1.2.1 クロスモーダル知覚

我々人間を取り巻く世界は時間と共に常に刻々と変化し続けている。このような環境に適応してきていくためには周囲の変化を迅速かつ正確に知覚・認識することが求められる。視覚や聴覚などのいわゆる五感と呼ばれる感覚器が外部から受容した刺激(感覚情報)を信号化し、それを脳が解析することで我々は外部の様子を知覚、認識を行う。さらに、我々の脳は各感覚器からの感覚情報を統合して解析処理を行っている。代表的な現象として、マガーク効果が挙げられる。マガーク効果とは、たとえば「ば」という音声と同時に「が」と発音している映像を提示すると「だ」という音が知覚されるという、視覚情報が聴覚における知覚に影響を及ぼす知覚現象である。このように、単一の感覚器からの感覚情報のみを処理するよりも迅速かつ正確に知覚し、認識することが可能である。

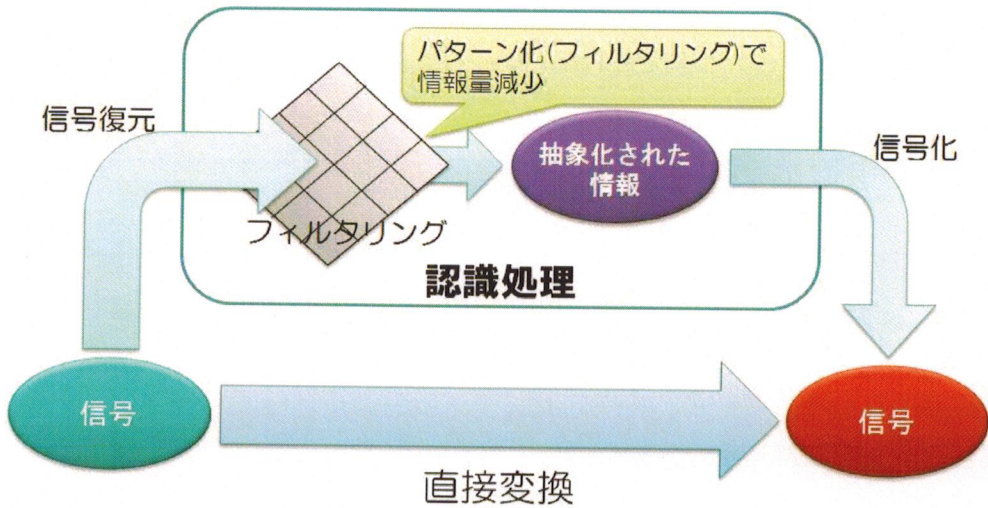
以上のような、異なる感覚情報がお互いに影響しあう知覚現象のことを「クロスモーダル知覚」と呼ぶ。クロスモーダル知覚による知覚の相互作用に注目することにより、異なる感覚器間の情報の変換処理を実現は十分に可能であるといえる。ある感覚器の使用が制限されている場合に別の感覚器の情報へ変換することにより、新しい感覚情報の伝達法やヒューマンインターフェースへの応用が期待される。

### 1.2.2 感覚情報の変換

感覚情報の変換の可能性に注目し、感性情報処理の分野を中心として研究が行われてきた。感覚情報変換の過程は大きく分けて Fig.1.1 に示されるような2種類の過程が考えられる。

ひとつは情報変換の過程で情報の認識処理を行い、変換の対象となる信号の情報をあらかじめ決まったパターンへの抽象化もしくは行うものである(パターン化)。この変換過程の問題点は、抽象化や記号化できない情報が抽象化の過程で失われることや、未知の認識対象に対して用意されていた認識パターンが適用できない点である。

これらに対してもう一方の変換過程は、変換対象の情報の認識処理を行わない(図中下)。認識処理が行われていない段階の信号のまま情報を扱い、直接違う信号へ変換する変換過程となる(直接変換)。ふたつの情報を信号のレベルで関係付け・対応付けを行うことで、情報の損失を最小限に抑えた変換を実現することが可能であると考えられる。



**Fig.1.1 信号変換過程**：上側の矢印のように認識処理を行うと、パターン化により情報量が減少する。対して、下側の矢印のように直接変換を行うことによって情報量はほぼ維持される。

### 1.2.3 視覚と聴覚

我々が周囲の状況の知覚・認識を行うとき、視覚からの情報に頼るところが大きい。視覚情報の比重が高いことは、視覚からの情報を処理する大脳皮質視覚野が大脳皮質全体の約三分の一を占めているところからもいえる。視覚の代替となるために考えられる感覚器は、大脳皮質上の領域の割合の大きさから考えて主に触覚と聴覚のふたつになると考えられる。

視覚の機能の代替となるためインターフェースの研究は、触覚ディスプレイなど、視覚の情報を触覚の情報に変換する手法が多く取り入れられている[7,8]。しかし、これらはピンが突出するディスプレイなどの特別なデバイスを必要とする。さらに、変換された情報を認識するためには指などでデバイスを走査することが必要であり、認



識に時間がかかることが考えられる。また、図形などの形を認識することを考えるのならば、走査は一つずつの図形についてしか行えず、同時に複数の形を認識することは困難であることが考えられる。

一方で、聴覚の情報はスピーカーやヘッドフォンなどの一般に普及しているデバイスを利用し、「音」として出力が可能であるので特別なデバイスがなくとも使用することが可能である。また、我々の聴覚は複数の音を同時に聞き分けることができる。この特性を利用することで、複数の形を同時に認識できる可能性が考えられる。そして、指で走査して図形の形をひとつひとつ追跡することなく、音が再生される時間内ですべての図形の情報が表現される。これにより、認識に必要な時間はひとつひとつを走査する場合に比べて、短縮することが可能になる。

以上のことから、視覚の補助ないし代替として図形を認識するために聴覚を用いることで、触覚を利用するよりも優れたインターフェースを開発ができることが考えられる。

### 1.3 研究目的

以上より、本研究では視覚障害者の補助のためのインターフェースの開発を目的とする。本研究では、視覚からの情報として図形などの「形」について注目した。視覚からの情報を聴覚へ感覚情報の直接的な変換を行い、視覚の代行となるような情報伝達方法の実現を目指す。具体的には、画像を音である波形データに変換し、変換前の画像に描かれた図形を音から連想・認識できるような情報変換のアルゴリズムの設計を行った。そして、音から形を認識することが可能であることを実証する。また、変換した音の形の認識率などから、認識に影響を及ぼす諸要因について検討を行い、どのような条件下でより最適な認識に至るのかを検討した。

## 第2章 画像の音への変換

### 2.1 はじめに

本章では、画像を音に変換する上でのその概念と、情報変換アルゴリズムの設計の方向性を示す。

我々は五感と呼ばれる5つの感覚器から様々な信号を受容している。受容した信号は脳で解析され、特徴が抽出される。抽出された特徴の持つ意味、もしくは情報を起きている現象などからフィードバックする。これを繰り返し行うことで、信号の特徴とその特徴の意味する情報との関係性の学習が行われ認識に至るようになる。また、受容した信号が過去に学習したものと完全に一致しなくとも、信号の特徴と類似したものがあればその情報を類推、連想することが可能である。実際に、我々人間は生まれた時には行えなかった言語の理解や、複雑な行動が学習を重ねて知識が蓄積されていくとともに可能になる。また、同じ行動であっても状況によりその過程を変えることが可能になり、行動や判断に汎用性を備えるようになる。この一連の過程が我々人間の感覚器からの情報を認識するメカニズムとなる。この過程に則り、本研究では情報変換アルゴリズムによって変換された音を人間の認識メカニズムに沿って認識させる。人間の認識メカニズムに則ることで学習により認識の精度を向上させ、汎用性を持った認識を行えることが十分に期待される。

本研究では、使用者が与えられた音から画像に描かれた図形を認識する。この点を考慮すると、音は人間の脳が形を認識するために視覚の情報から抽出している情報を含んでいるなければならない。人間の認識のメカニズムにおいて形の認識に最も影響を与えている情報を、画像として音に変換することが必要となる。また、変換された音の学習が行えるように、形の特徴に依存して音の特徴も異なり、その音の違いを使用者が聞き分けられる変換アルゴリズムが適していると言える。

以上より、本研究では脳が形を認識する際に必要とするデータを画像から抽出し、音に変換する。また、学習により音からの情報の認識力を高め、未知の認識対象への応用を図る。

## 2.2 人間の視覚における形認識のメカニズム

我々人間にとって、視覚情報の入り口となる感覚器官は眼球である。外界からの光は眼球内の角膜、水晶体などを通過して最後に網膜に到達する。網膜上に光の映像が映し出され、光形式の信号となる。光化学反応過程を経由して、光形式の信号は電気形式の信号に変換される。変換された電気信号は視神経を通じ、外側膝状体を経由して視覚情報の処理を行っている大脳視覚野に伝えられる。大脳視覚野は大脳皮質上の位置や、受容野の特性による機能の違いに基づいて名称がつけられている。その中のV1野(第一次視覚野)に、網膜からの信号の多くが投射される。

V1野の受容体は刺激選択性を持ち、限られた刺激にのみ反応する。V1野における細胞の受容野特性についてはヒューベル(D. H. Hubel)とウィーゼル(T. N. Wiesel)によって、猫とサルを使って詳しく調べられている。その結果、V1野には単純型細胞と複雑型細胞が多くあることが明らかにされた。単純型細胞の受容野は細長い楕円、もしくは方形をしている。この受容野は、光が投射されているときに反応するものと、光が投射されていない時に反応するものの2種類で構成される。単純型細胞の構造はFig.2.1に示すように、中央帯とその両側で細胞の反応様式が異なるものと、片側とその反対側で反応様式が異なるものの2種類がある。両種類とも、反応様式をまたぐような光には反応せず、反応様式の境界線に沿った方向の光のみに反応する。このようにして、単純型細胞は特定位置における特定方向の直線や明暗の境界部分であるエッジの検出を行うことができる。複雑型細胞の受容野は単純型細胞と同様に方形をしている。しかし、単純型細胞とは異なり、位置に関係なく特定方向の直線やエッジにのみ反応する。これは、複雑型の受容野が単純型細胞を組み合わせることで構成されているものであるからだと考えられている。以上より、V1野には特定の傾きを持った細長いパターンやエッジに強く反応する細胞が多く存在することが明らかにされた。個々の細胞は最も強く反応する最適方位を持ち、方位選択型細胞と呼ばれている。

方位選択型細胞は規則的に配列されている。細胞の配列は大脳皮質内で3次元構造に組み立てられている。その様子をFig.2.2に示す。図のように、同じ方位に対して反応する細胞が皮質表面から皮質深層に向けて柱状に配列されている。このような構造を方位選択性コラムと呼ぶ。隣り合う方位選択性コラムは少しだけ違った最適方位を持ったものが配置されている。これは全体として見ると最適方位の変化の順序に対して周期的に配列されていて、1周期で最適方位の変化は180度となる。また、左目からの情報を主に処理する部分(左目コラム)と右目からの情報を主に処理する部分

(右目コラム)が交互に配置されている。

以上のような構造から、V1野では視野内のあらゆる方向の直線やエッジを検出する。V1野で検出された情報は、次に視覚連合野と呼ばれる領域に伝達され処理される。形は、V1野からの情報をもとに視覚連合野で認識される。

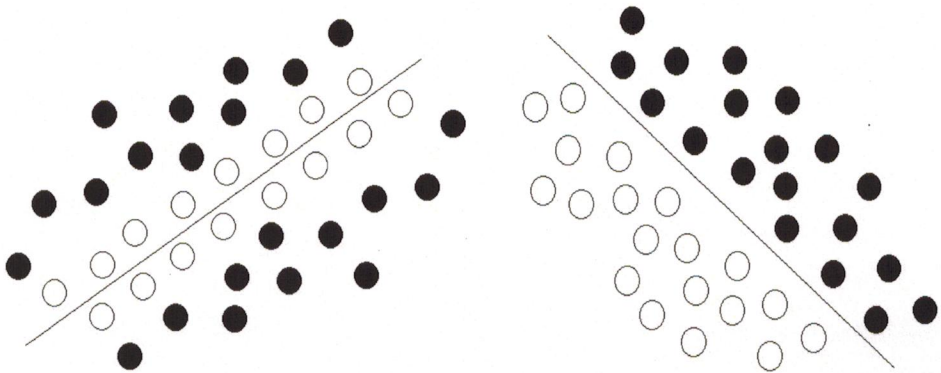


Fig.2.1 単純型細胞の受容野：

(a)中央帯とその両側とで反応様式が異なるもの (b)左右で反応様式が異なるもの

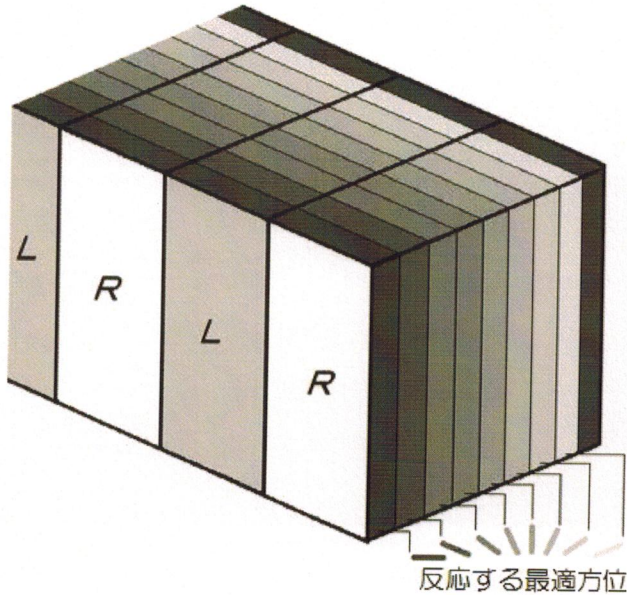


Fig.2.2 方位選択制コラム：方位選択型細胞は規則的に配列されている

## 2.3 音への変換アルゴリズム

### 2.3.1 画像の音への写像

#### (i)画像処理

前節の形認識のメカニズムより、本研究の情報変換では変換対象の画像に対する画像処理は V1 野で行われる処理であるエッジの抽出のみとし、輪郭抽出以外には画像処理や認識処理を行わない。つまり、視覚連合野で行われている V1 野からの情報をもとにした情報の解析・認識などの処理は使用者の解析能力と認識能力に依存する。輪郭を抽出された画像はその輪郭線の部分を白色(明度 255)、それ以外の部分を黒色(明度 0)であらわされる。エッジ抽出の処理前と処理後の画像を Fig.2.3 に示す。輪郭線は幅が 1[pixel]であり、抽出前の画像が幅 1[pixel]の線が描かれていた画像の場合は、その線の部分が白色になり背景部分が黒色になる。

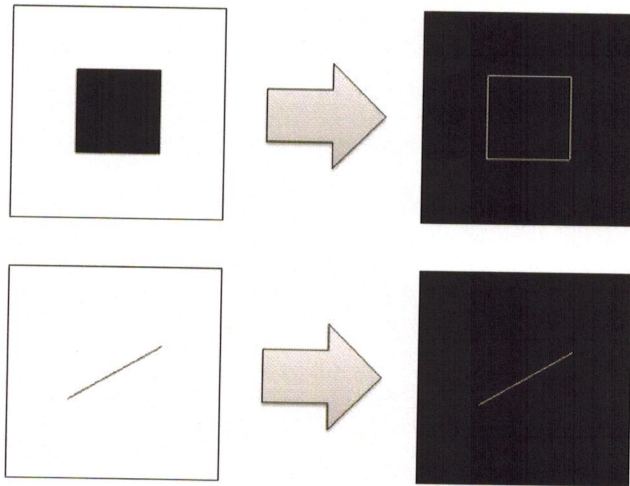


Fig.2.3 エッジ抽出前の画像と抽出後の画像：輪郭部分は幅 1[pixel]の白色線で描かれる。

#### (ii)画像の走査

前処理を行った画像を走査して、そのデータを取り出す。ここで取り出すデータは明度データである。その走査方法は Fig.2.4(a)に示すように、目線を画像の左端から

右端に向かって移動させる行為を真似て、画像を左から右へ走査する。画像を構成する画素の縦1列分の明度データを順番に取り出していき、画像全体の明度データを得る。Fig.2.4(b)に示すように、本手法では縦一列分の明度データから信号を生成、波形データに変換しそれをひとつの音要素とする。使用する画像の大きさは  $200 \times 200$ (画素数  $N = 40000$ [pixel])であるので、1枚の画像を音に変換すると200個の音要素から構成されることになる。走査に要する時間は1[sec]とし、音として再生される際にはこの操作方法と音の再生を対応させる。よって、音は画像の左時恥の音要素から順に再生し、画像全体の再生時間 $T$ は1[sec]となる。

画像の横幅 $w$ である200[pixel]を1[sec]で横断するので、ひとつの音要素が再生される時間 $\tau$ は5[msec]となる。グリーン(Green)の実験によると、聴覚が音の時間的変化を認識できる最短の限界は1[msec]から2[msec]程度であるとされている。これより、隣り合う音要素が異なる音であってもその変化を認識することは可能だと考えられる。

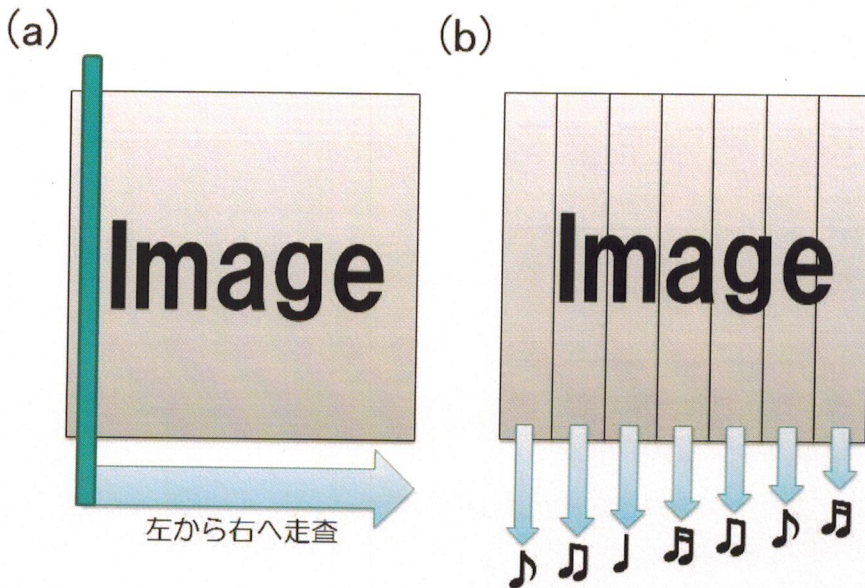


Fig.2.4 変換過程での画像の取り扱い：(a)画像の走査方法 (b)音の生成

### 2.3.2 波形データへの変換方法

先述したように、画像の縦1列分の明度データを波形データに変換してひとつの音要素にする。画像の縦幅を $h$ とすると、画像走査により取り出した縦列の明度データ  $I(y)$  ( $0 \leq y < h$ )はひとつの音要素  $S_x(t)$  ( $0 \leq t \leq \tau$ )を構成する波形データ  $I_x(n)$  ( $0 \leq$

$n < h$ )に変換される。同じ縦列に輪郭線の一部である点が複数存在している場合は、存在する点それぞれを反映させたもの、波形データを足し合わせた合成波などにしてひとつの音要素にする。変換された各音要素 $S_x(t)$ を $x = 0$ から $x = w - 1$ まで繋ぎ合わせることで、全体の音をあらわす波形データ $S(t)$  ( $0 \leq t \leq T$ )となる。

本研究では、波形データへの変換方法を3種類提案する。以下でそれぞれの方法について述べる。

### (i)変換方法 A : 明度の変化の波形化

この方法は、画像の縦1列における画素の明度変化に注目した方法である。Fig.2.5(a)のような画素列の位置に対する明度の変化を、Fig.2.5(b)のようにグラフ化する。グラフ化した明度データの位置パラメータ $y$ を Fig.2.5(c)のように時間パラメータ $t$ に変換させ波形データとして扱う。

$$S_x(t) = I_x(n) = I(y)$$

縦列に複数の点( $I(y) = 255$ )が存在する場合には、それぞれの波形データを足し合わせて合成波を形成する。この波形への変換処理は、画素の位置パラメータを波の時間パラメータに対応させるだけなので、変換過程での情報の欠損は全くない。画像情報を直接的に反映させた波形データであるといえる。

### (ii)変換方法 B : フーリエ変換による波形化

画素の縦列に対して逆フーリエ変換を行い、明度データ $I(y)$ を波形データ $I_x(n)$ に変換する。同じ縦列に複数の点( $I(y) = 255$ )が存在する場合、変換後の波形データはそれぞれの点の位置に対応した sin 波の合成された波となっている。波形データでは虚数を取り扱うことができないので、逆フーリエ変換の結果の虚数部は無視し、実数部のみを波形データに適用した。

### (iii)変換方法 C : 位置と周波数の対応付けによる波形化

画素の縦列の位置 $y$ に対して定めた周波数 $f_y$ を割り当てる。対応する位置に点がある場合にその周波数の sin 波を生成し、ひとつの音要素とする。音全体の再生時間が1[sec]であるので、サンプリングレート $f_s = N$ とする。

$$I_x(n) = I(y) \sin \frac{2\pi f_y n}{f_s}$$

$$S_x(t) = \sum_{y=0}^h I_x(n)$$

同じ縦列に複数の点( $I(y) = 255$ )が存在する場合は、上式からも分かるようにそれぞれの位置に対応する周波数の  $\sin$  波の合成波を生成しひとつの音要素とする。

割り当てる周波数は画像の最低位置を 200[Hz]とし、位置が高くなるにつれ割り当てられる周波数は大きくなるようにした。よって、位置が高くなるにつれて生成される音要素は高音のものになる(Fig.2.6)。これは、人間が高音を聞いたときには高い位置、低音を聞いた時には低い位置を連想するという心理学的な要因を考慮、利用したものである。位置に対して割り当てられる周波数の対応関係は、位置が 1[pixel]変換すると周波数が 200[Hz]変化するようにした。我々人間の聴覚の周波数分解能は指数関数的であり、高い周波数ほど弁別閾が広くなる。20000[Hz]付近ではその弁別閾は 200[Hz]程度となる。よって、本手法の変化量である 200[Hz]の周波数の変化を認識することは十分に可能であるので、僅かな高さの変化も認識できることが期待される。

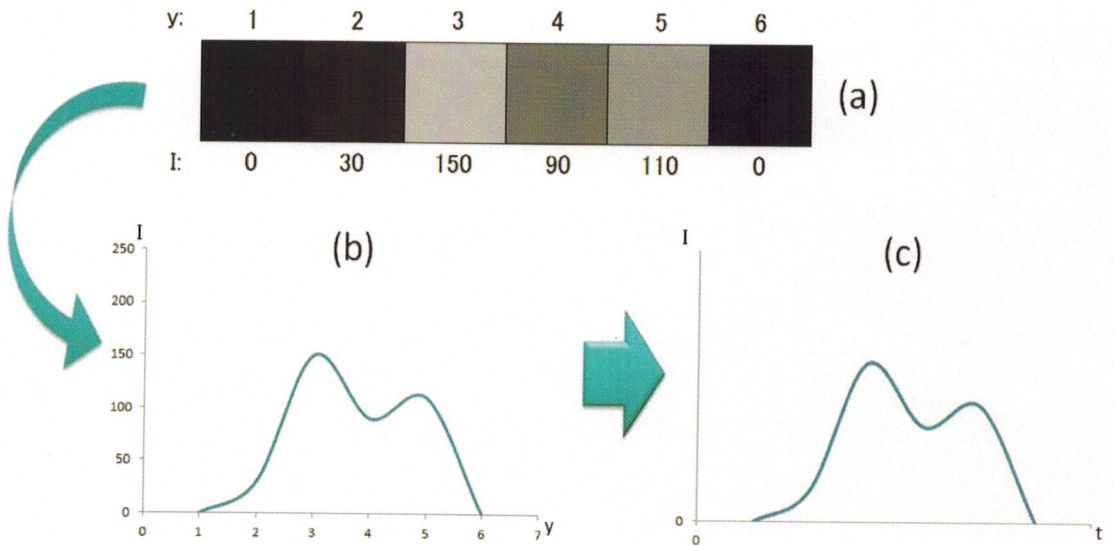


Fig.2.5 明度変化のグラフ化とパラメータ変換：

(a)画素列の明度変化 (b)明度変化のグラフ化 (c)時間へのパラメータ変換



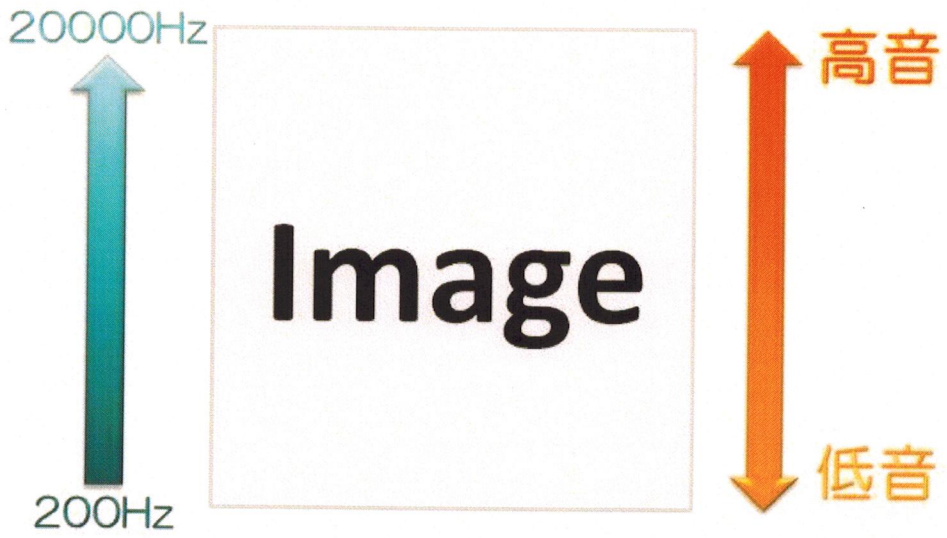


Fig.2.6 : 画像の位置と周波数の対応関係

# 第3章 予備実験

## 3.1 はじめに

本章では、画像を音へ変換する手法である、変換方法 A、変換方法 B、変換方法 C によって形を反映した音に変換されるか、またその音を使用者が識別することが可能であるかの確認を行い、アルゴリズムの信頼性を検証するための実験を行う。

音から形を認識するためには、形の輪郭の変化によって変化する音に変換されている必要がある。音の変化は形の輪郭の変化と何らかの関係性を持って変化しなければ音から形を認識することは難しい。当然のことではあるが、異なる形を変換した音同士を比較したのであれば異なる音でなければいけない。また、形ごとに異なる音が生成されたとしても、使用者がそれぞれの音の特徴を聞き分けることができることができなければならない。形ごとに特徴が見られた場合、その特徴の傾向が重要であると考えられる。各形を視覚的に見て受ける特徴の印象と、各形の音を聞いて受ける特徴の印象に関係はあるか。また、音から受ける印象は変換元の形が似ているとその特徴や印象も近くなるのか否かが、音からの形の正確な認識に大きく影響すると考えられる。

## 3.2 実験 1 : 形による音の違いの認識

### 3.2.1 目的

実験 1 の目的は、3 種類の変換方法を用いて異なる形が描かれた画像を音に変換した場合に、その形の違いを反映した音が生成されているかどうかを検討する。また、形の違いを反映した音が生成されているならば、その音の違いを被験者が認識可能かどうかを検証する。形の違いを反映した音に変換され、その音の特徴を使用者が認識できているのならば、それぞれの音を比較したときに音の特徴が同じかどうかを認識

できるはずである。本実験では、5種類の異なる形をした画像を用いて、その変換した音を比較してその音に違いがあるかどうかを変換方法 A、B、C のそれぞれについて検討した。

## 3.2.2 方法

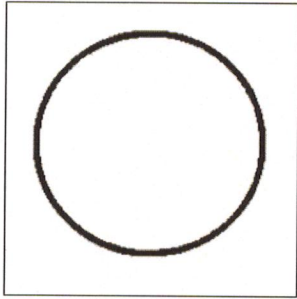
### (i) 実験被験者

実験の被験者は成人男性 5 名であった。いずれの被験者も視覚機能に異常のない晴眼者であり、また聴覚にも障害のない健聴者である。5 人のうち 3 人は楽器の演奏経験がある。

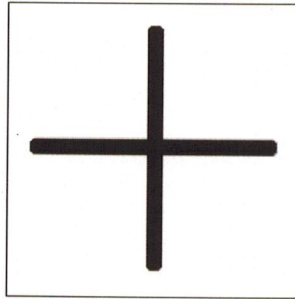
### (ii) 手順

Fig.3.1 のようなそれぞれ異なった形の図形が描かれた 5 種類の画像を音に変換し、それを聴覚刺激とした。5 種類の刺激の内からふたつをランダムに選出し、被験者に連続して与える(同じ刺激がふたつ与えられることもある)。被験者の課題は、与えられたふたつの音を聴き比べふたつの音が異なる音か、同じ音かまたは判別不能かを分かり次第素早く回答することであった。この時与える刺激は聴覚刺激である音のみであり、視覚刺激である音に変換される前の画像を提示することはない。また、被験者はこれらの音は何の画像が変換されたものであるかは知らされなかった。実験中の音はヘッドフォンを使用して聞かせ、回答はキーボードによる入力で行った。

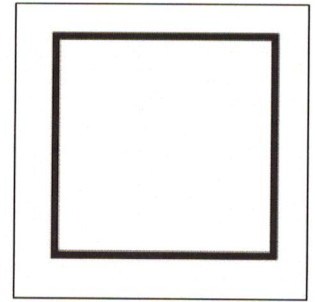
実験は、変換方法 A、B、C 各 3 種類の方法それぞれにつき 25 問出題し、全 5 回の施行を行った。なお、被験者には各変換方法のアルゴリズムについては、画像の左側から右からに向かって走査しているということ以外は知らされていない。



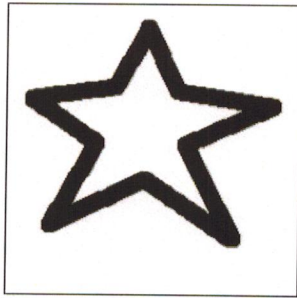
Circle



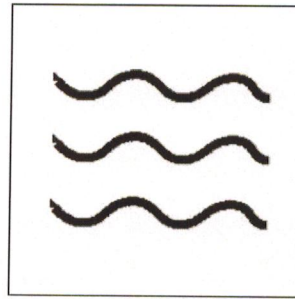
Plus



Square



Star



Wave

Fig.3.1 実験1に使用した5種類の図形の描かれた画像

### 3.2.3 結果と考察

変換方法 A、B、C それぞれにおける、全被験者の回数別の識別率を Table.3.1 に示す。Table.3.1(a)より変換方法 A については、被験者による個人差は見られるがいずれの被験者も識別率が 84[%]を下回ることはなかった。各被験者の標準偏差は最大で 6.4 であった。このことより、形の違いにより音に違いが表れていること、形に依存した音の違いは被験者にとって識別すること可能な範囲内で表現されているということが分かる。Table.3.1(b)より、変換方法 B の場合は識別率が 84[%]下回ることはなく被験者別の標準偏差も最大で 4.8 であった。Table.3.1(c)変換方法 C については、識別率が最小で 88[%]、被験者別の標準偏差は最大で 4.38 であった。以上より、形によって異なる音が生成されているということ、本実験で使用した図形程度に異なる形から生成された音の違いを被験者が識別可能であることがそれぞれ示された。よって、3 種類の変換方法について、変換アルゴリズムの信頼性が示されたと言える。

また、各変換方法の平均認識率を Fig.3.2 に、その値と標準偏差値を Table.3.2 に示す。平均認識率は変換方法 A が 96.8[%]、変換方法 B が 96.8[%]、変換方法 C が 99.2[%]であった。標準偏差はそれぞれ変換方法 A が 4.8、変換方法 B が 4.23、変換方法 C が 2.52 であった。ここからも、音を聞くことで形の違いを判別することは可能であることが示唆される。また、3 種類の変換方法の中でも変換方法 C は他の変換方法と比べて平均認識率が 2.4[%]高かった。変換方法 C の標準偏差は変換方法 A と比べると 2.28、変換方法 B と比べると 1.71 だけ小さい。このことより、変換方法 C は回数や被験者によるばらつきも少ないということが分かる。以上より、変換方法 C は 3 種類の変換方法の中で最も音の違いを聞き分けることに適した変換方法である可能性があるということが示唆される。

Table.3.1 被験者の回数別識別率(%): (a)変換方法 A (b)変換方法 B (c)変換方法 C

(a)

	被験者SO	被験者TK	被験者KN	被験者RM	被験者NH
1回目	100	100	88	100	88
2回目	96	100	92	100	96
3回目	96	100	100	100	96
4回目	96	100	84	100	100
5回目	100	100	100	88	100
標準偏差	1.95959	0	6.4	4.8	4.38

(b)

	被験者SO	被験者TK	被験者KN	被験者RM	被験者NH
1回目	92	100	96	88	84
2回目	100	100	100	100	96
3回目	100	100	96	96	92
4回目	100	100	96	100	96
5回目	96	100	100	100	92
標準偏差	3.2	0	1.959	4.8	4.38

(c)

	被験者SO	被験者TK	被験者KN	被験者RM	被験者NH
1回目	100	100	100	96	100
2回目	100	100	100	100	100
3回目	100	100	100	96	100
4回目	100	100	100	100	100
5回目	100	100	100	88	100
標準偏差	0	0	0	4.38	0

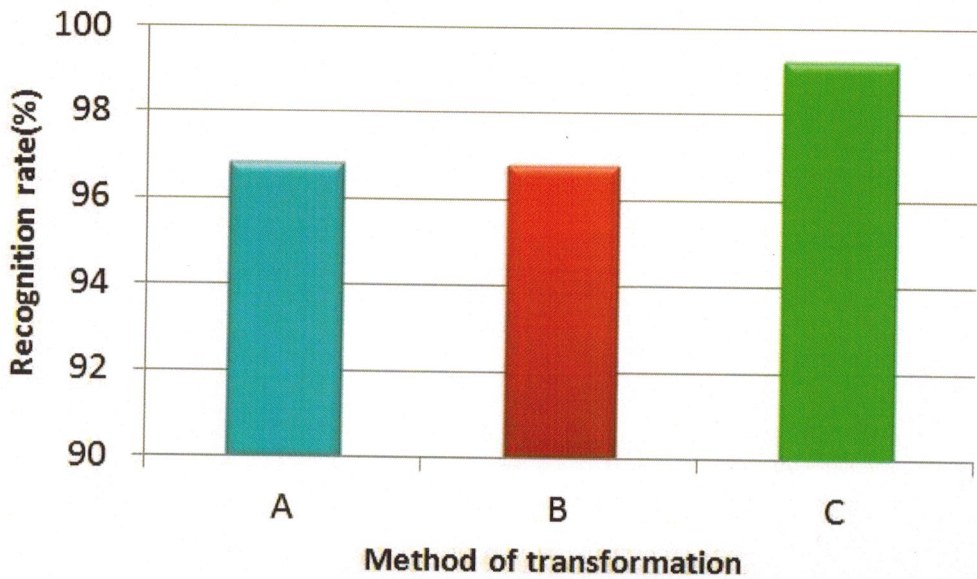


Fig.3.2 変換方法ごとの平均認識率(%)：全被験者の各回テストの認識率から変換方法ごとに平均を算出した。

Table.3.2 変換方法ごとの平均認識率(%)と標準偏差：Fig3.2の平均値とその標準偏差。

	A	B	C
平均正解率(%)	96.8	96.8	99.2
標準偏差	4.8	4.23	2.52

## 3.3 実験 2 : 音からの形の連想

### 3.3.1 目的

形の違いは音に反映され、異なる音として表現されることが実験 1 により示された。そこで本実験では、異なる音を聞くことで異なる形を連想することができるかを検討する。

音から形を連想するためには、その音の特徴を聞き取り視覚的な特徴と結びつけられなければならない。あるふたつ音の変換元の形の視覚的な特徴が近い場合、その音の特徴も似た傾向になると考えられる。図形の視覚的な特徴を考えると、輪郭線が変化しない線か、または曲線や複数の様々な角度の線によって変化するかどうかで特徴が分かれる。輪郭線の変化が大きく激しいものは音の変化量も大きく、輪郭線の変化が小さく一定なものは音の変化も小さく一定音のようになる。よって、音の変化の様子に依存して輪郭線の変化が大きい図形と少ない図形に連想の分布に傾向が表れると予想される。

本実験では、5 つの異なる音を聞き、提示されている形の中から音から連想できたイメージを選んでもらった。連想する傾向の分布を記録し、特徴の聞き分けについての検討を行った。なお、本実験も変換方法 A、B、C の 3 種類のそれぞれに対して行われた。

### 3.3.2 方法

#### (i) 実験被験者

実験の被験者は、実験 1 の被験者とは異なる 4 人の成人男性であった。いずれも聴覚に障害のない健聴者である。4 人のうち 2 人は楽器の演奏経験がある。

#### (ii) 手順

本実験の被験者の課題は、5 つの異なる音を被験者に聞いてもらいその音から連想できる形を与えられた図形から選択し解答してもらった。聴覚刺激として与えられる音は、選択肢として提示される 5 種類の画像を変換したものである。5 種類の画像は実験 1 と同じ画像を用いた。音を与えられるときには、被験者には何の音を与えられ



ているのかは知らされない。また、視覚刺激としての画像は常に被験者に提示されている。音は全種類5回ずつ与えられ、与えられる運版はランダムに決定される。実験中の音はヘッドフォンより与えられ、回答はキーボードから入力で行った。実験の試行回数は5回で、全125問になる。

以上のような条件で、変換方法 A、B、C の3種類それぞれに対して同様に行った。実験1と同様に、被験者には変換方法のアルゴリズムに関しては画像の走査法以外は知らされていない。

### 3.3.3 結果と考察

使用した図形の中でも、Square と Plus は輪郭線が直線で構成されているので、視覚的な特徴として輪郭線の変化が小さいものに分類できる。一方で、Circle や Star、Wave は輪郭線が曲線や複数種の傾きの直線で構成されているため、輪郭線の変化が大きいものに分類できる。

変換方法 A において、各図形の音に対する被験者全員の連想した図形の割合を Table.3.3(a) に示す。Circle の音に対する被験者が連想したイメージで最も多かったのは Circle で 38[%]、次いで Star が 31[%] であり、Wave にも 17[%] の分布が見られた。これら3種類の図形に分布が見られたのは、Circle は輪郭の変化が大きい図形であったことが要因として推定される。これに類似した傾向が Wave にもみられる。Wave の音に対する回答は最大で Wave の 47[%]、その他にも Circle に 19[%]、Star に 14[%]、Square に 12[%] の分布が見られた。Square にも分布が見られたのは、曲線の音を直線の音と認識してしまった場合が推定される。Star の音に対する回答の割合は、Star が 48[%]、Wave が 40[%] であった。Star の輪郭線は、視覚的に見て Circle のようなひとつの弧を描くような輪郭の変化よりも鋭利で変化の上下の変化が激しい。よって、Star の音は Circle の音よりも変化が激しくなり、Circle の視覚的な特徴と結びつかなかったことが考えられる。一方で、輪郭線の変化がほぼ見られない図形である Square と Plus の音から連想される図形は、Square の音に対して Square の図形を連想した割合が 76[%]、Plus の音に対して Plus の図形を連想した割合が 85[%] であった。輪郭線の変化が大きかった図形の音に比べ、音の変化ないことがこのような分布の要因であると考えられる。

変換方法 B についての回答の割合を Table.3.2(b) に示した。各図形の音によって連想できる図形の傾向は変換方法 A での傾向とほぼ同様の分布であることがうかがえ

る。しかし、輪郭線の変化が多くみられる図形である Star と Wave の音からは Wave が連想される割合が 50[%]以上を占めた。

変換方法 C における回答の割合を Table.3.2(c)に示す。各図形の音に対して連想した図形の割合の最大値は Circle が 77[%]、Square が 75[%]、Plus が 88[%]、Star が 63[%]、Wave が 65[%]となった。これは他のふたつの変換方法の場合と比べると偏りが大きいと言える。しかし、Circle の音に対して Star の図形に 18[%]、Wave の音に対して Star の図形に 22[%]の連想が分布している。連想される図形に偏りが生じるが、音から連想される図形の傾向は他のふたつの変換方法の場合と同様の傾向が見られた。これより、変換方法によって偏りや強さは異なるが、変換前の図形と輪郭線に類似性がある図形が連想される傾向があるといえる。






各図形の音について、変換された元の図形を正しく連想することができたかを被験者全体の認識率の平均として Fig.3.3 に示す。まず変換方法 A と変換方法 B について、Square と Plus の認識率は最低でも 75[%]以上であった。これに対して、輪郭線の変化が大きい Circle や Star、Wave の認識率はすべて 52[%]を下回った。よって、この 2 種類の変換方法は輪郭線の変化が少ない図形を連想する精度に対して、輪郭線の変化がある図形を連想する精度は減少することがうかがえる。一方で、変換方法 C は Square を除いたすべての図形の音において、他の 2 種類の変換方法よりも各図形の認識率が高かった。このことより、3 種類の変換方法ではこの変換方法 C が最も形の連想を行いやすく、元の形を認識しやすい音に変換する方法であることが考えられる。特に、Circle に対する認識率は 77[%]と他の変換方法と比べると約 40[%]高く、Circle のような緩やかな曲線程度の輪郭線の変化ならば認識が十分可能であることが示唆される。しかし、Square や Plus、Circle に比べて Star と Wave の認識率は低いことから、輪郭線の変化が大きい図形の認識率が他の図形に比べて減少する傾向は変換方法 A、B と同じく存在することがうかがえる。

以上より、輪郭線の変化が少ない図形は音の変化も少ないので音の特徴を聞き分けることを容易にし、図形の連想をしやすいことが示唆される。対して、輪郭線の変化が大きくなると音の変化の特徴から輪郭線の変化の特徴に連想しづらくなり認識率が減少すると考えられる。また、変換方法 C の認識率が全体的に高いことより、変換方法 C を用いて変換された音は本来の形を連想するのに十分な音的な特徴を持っていると考えられる。






**Table.3.3** 各図形の音に対する被験者の回答の分布の割合(%) :

(a)変換方法 A (b)変換方法 B (c)変換方法 C






(a)

					
circle	77	1	0	18	4
square	9	75	14	0	2
plus	2	8	88	0	0
star	11	2	2	63	22
wave	1	6	5	23	65

(b)

					
circle	29	5	7	43	16
square	10	75	10	1	4
plus	23	0	76	1	0
star	11	1	3	32	52
wave	5	5	5	34	51

(c)

					
circle	77	1	0	18	4
square	9	75	14	0	2
plus	2	8	88	0	0
star	11	2	2	63	22
wave	1	6	5	23	65

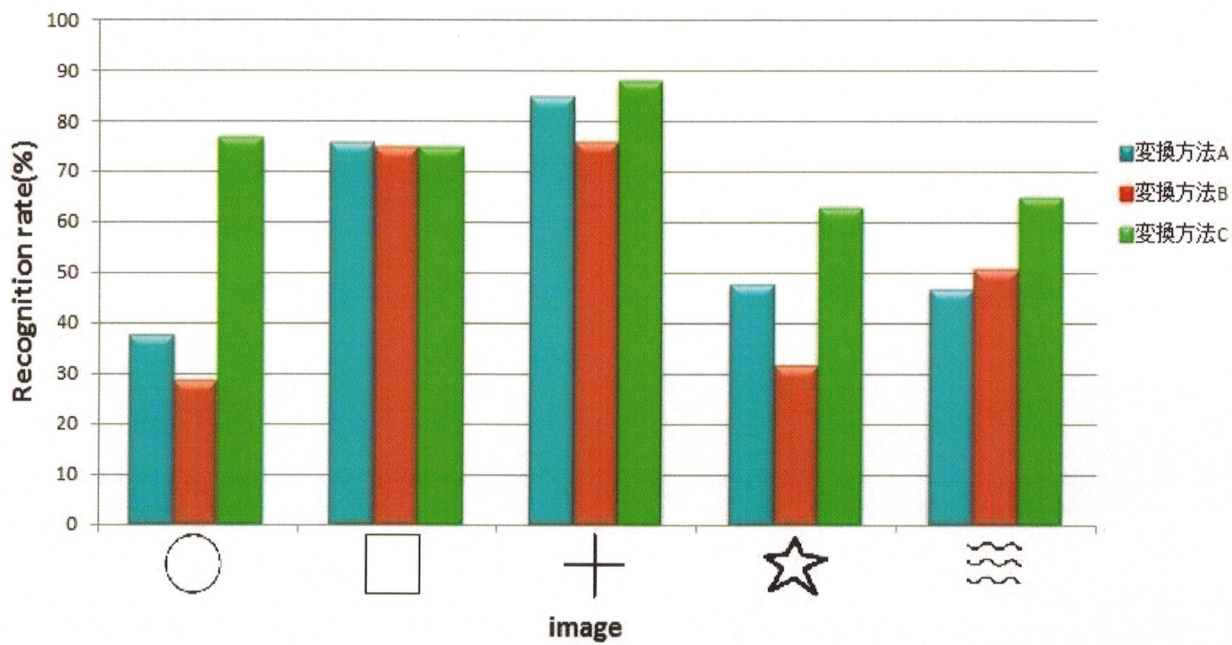


Fig. 3.2 変換方法ごとの平均正答率(%)

# 第4章 音の学習と解析

## 4.1 はじめに

本章では、画像を変換した音から変換前の図形を連想できるかの検討を行う。また、音の学習を行うことの有効性と学習による認識の応用性、各変換方法の認識率についての検討を行う。

第3章にて、描写されている図形によって変換された音は異なり、また音を聞いてそれから連想される形は音の特徴によって異なるということが示された。そこで、学習という過程を用いて特定の形の視覚的特徴とそれが音に変換された時の聴覚的特徴を結びつける。これにより、学習したことのある音であればその特徴から正しい形を連想できるようになる。その後、学習した図形と似ているが一度も聞いたことがない図形の音を聞かせる。学習した特徴を含んでいるが、学習した形とは正確には異なる未知の形の音を聞いたとき、被験者たちは学習結果を応用して未知の形を連想・認識することができるかどうかを検討する。

我々人間にとって学習と応用の過程は日常では当然のこととして処理されている。このことを鑑みれば学習という過程を行うことは可能だと予想されるが、音からの形の認識という処理に対してどの程度の学習効果と応用力が見られるのかは未知数である。被験者の未知の形の認識力からこれらについての検討を行う。

また、我々人間は視覚的に形を認識する際に次のような2通りの方法を行っている。ひとつは、形の構成する線の空間的配置とその傾きを認識・解析し、組み合わせることで形を認識するという方法である。もうひとつは、形を構成する要素に分解・解析することはせずに、形全体をひとつのパターンとして認識する方法である。これは全体の構成要素ひとつひとつを解析した結果、その組み合わせをひとつのパターンとして認識した場合も含める。ここで前者を局所的な視点、後者を大局的な視点として定義する。本実験ではこのふたつの視点に則り、局所的な視点での学習・応用と、大局的な視点での学習・応用についての検討をそれぞれ行う。

## 4.2 実験 1 : 直線の角度の学習とその音の解析

### 4.2.1 実験目的

本実験は局所的な視点での音の学習と解析することについての検討を目的とする。形を構成する局所的な要素のもっとも単純なものとして、今回は 1 本の直線の角度について学習・解析する。

直線の傾きに対しての学習が行われるのならば、いくつか基準となるような角度の直線を学習することで、それを基とした応用的な解析による認識が可能になると予想される。学習した音と未知の音を比較することで、音の変化の仕方の違いなどから角度の違いを解析し、直線を認識することができるようになると予想される。これらについて検討する。

8 種類の基本となる角度の直線を学習した後に 6 つの直線についてそれぞれ解析を行い、局所的な視点での学習と解析の検討をする。

### 4.2.2 方法

#### (i) 実験被験者

実験の被験者は、第 3 章の実験 1 に参加した被験者 5 人であった。

#### (ii) 手順

本実験では被験者に与えられる課題は、直線の音を聞きその角度を解析して解答してもらうことである。基準とする角度の直線を学習した後に角度を知らされていない直線の音を聞き、その角度を解析した直線をイメージする。イメージした直線を解答用紙に描く形で回答してもらった。

認識テストの前に行われる学習過程では、角度は解析する指針となるように Fig.4.1 に示した 8 通りの角度の直線を学習することとした。学習プログラムではこれらの直線の音をランダムな順番で被験者に与えられる。各直線の音は被験者によって任意の回数だけリピートして聞くことが可能であるが、次の順番の直線の音に移ってしまうと以前に聞いた直線の音を聞くことはできない。また、学習過程の間は各直線の画像を視覚刺激として提示されている。認識テストでは、角度がランダムに決定

された長さが一定な直線の音を聞き、学習結果を基にその直線を解析して解答用紙に描いてもらった。直線の音は解答中であれば被験者の任意で何度でもリピート再生が可能であった。1回のテストで出題される数は6問とした。認識テストでは学習過程で使用した視覚刺激は提示されない。ただし、認識テスト終了時には被験者はどの角度の直線が出題されたか提示される。実験中の音はヘッドフォンより出力され、学習プログラム、認識テストともにリピートなどの操作はキーボードより行う。実験回数は10回であり、変換方法A、B、Cそれぞれに対して行われた。また、被験者は各変換方法のアルゴリズムについて知らされている。

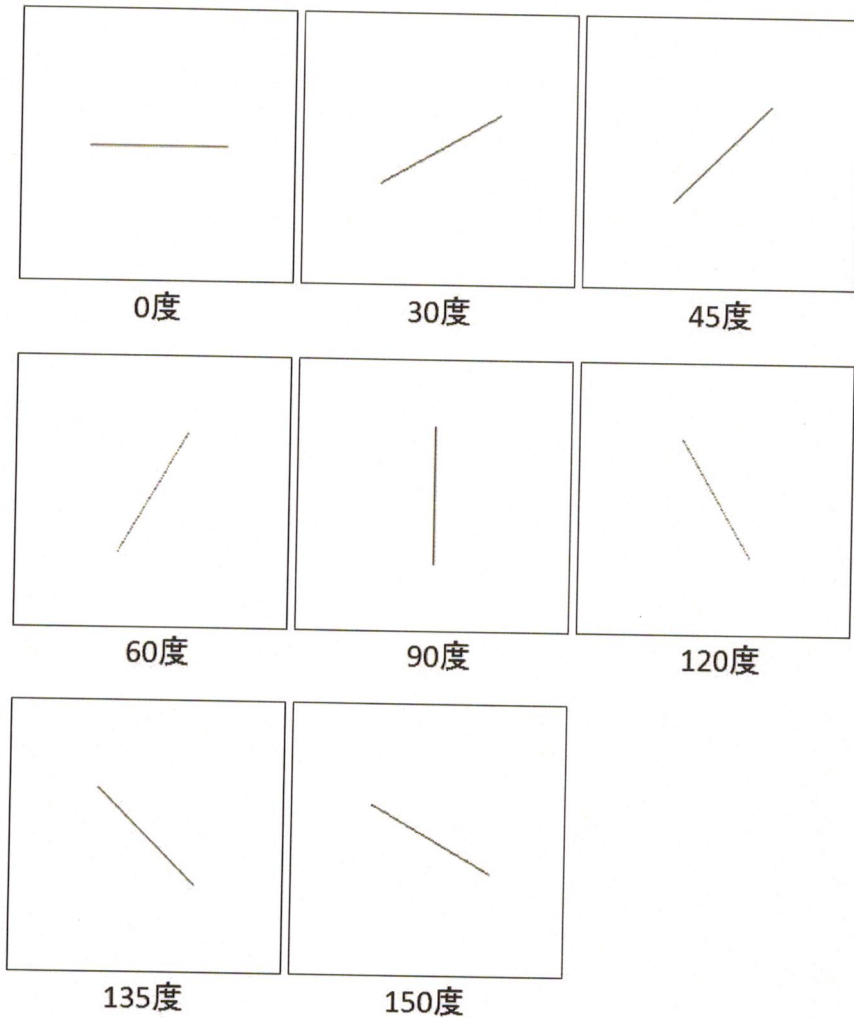


Fig.4.1 実験1の学習過程で使用した直線

## 4.2.3 結果と考察

Fig.4.2 にテスト回数ごとの被験者の応答時間の平均を示す。本実験における応答時間は、音の再生が開始されてから被験者が用紙に直線を描画し終えるまでの時間である。いずれの変換方法でも回数を重ねることで応答時間は減少している。このことから、音から直線を認識するために必要な時間は回数を重ねることにより短くなっていると言える。十分な回数の学習やテストを行うことで相応の時間での認識が可能になることが予想される。

被験者らが描いた直線と実際の直線との誤差の様子を観察するために、両直線の角度差の平均がテスト回数によってどのように変化するかを Fig.4.3 に示した。線形近似曲線より、変換方法 B と変換方法 C はテスト回数が多くなるにつれて角度差が減少していくと分かる。これより、このふたつの変換方法は音から直線の傾きの度合いを認識する精度が上がっていることが示唆される。毎回の回答後に出題された直線を確認する作業により、回答で描いた直線との角度差を修正することが可能であったからだと考えられ、変換方法 B と変換方法 C はこの効果が表れやすかったことが示唆される。

また、被験者らの描いた直線から認識率を評価するにあたり、一貫した評価基準を設定し被験者らの回答から認識率を算出した。変換元の直線と比較して、直線の傾きの方向(右上がり又は左上がり)が合っているか、傾きの度合いが合っているかの 2 点を直線認識の評価の基準の対象として得点化した。基準を満たしている場合は 1 点加点し、1 問につき最高 2 点の得点が与えられる。1 回のテストでの最高得点は 12 点となる。なお、回答の直線と実際の直線の傾きの差が $\pm 5[^\circ]$ 以内であれば回答が手書きであることによる誤差の範囲内とし、角度は合っているとみなした。各回数における被験者らの得点が最大の場合の合計を認識率 100[%]と定義した。Fig.4.4 に被験者らの回数ごとの平均認識率の変化を示す。全 10 回のテストでの平均認識率は変換方法 A が 48.50[%]、変換方法 B が 47.67[%]、変換方法 C が 73.67[%]であった。3 種類の変換方法の中では変換方法 C の認識率が最も高いと言える。各方法の線形近似曲線については、変換方法 A と変換方法 C のふたつは近似曲線の傾き $\alpha$ がそれぞれ $\alpha_A = 0.9798$ 、 $\alpha_C = 0.7475$ であった。このことより、変換方法 A と変換方法 C は回数を重ねると認識率が上昇すると言えるので、改めて回数が重なるとともに認識精度の向上があることが示唆される。



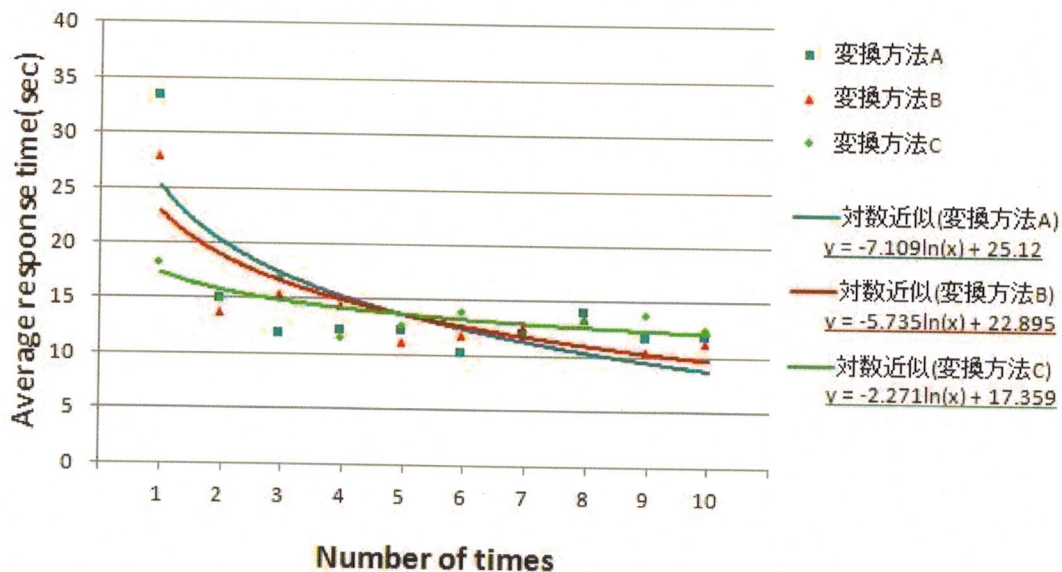


Fig.4.2 テスト回数ごとの平均応答時間(sec)

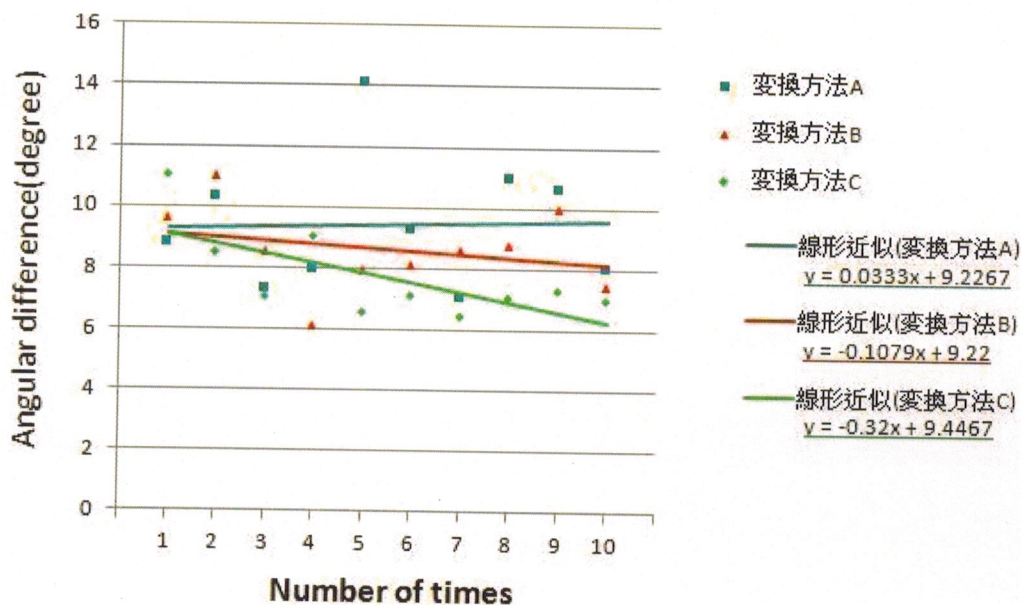


Fig.4.2 テスト回数ごとの平均角度差(度)

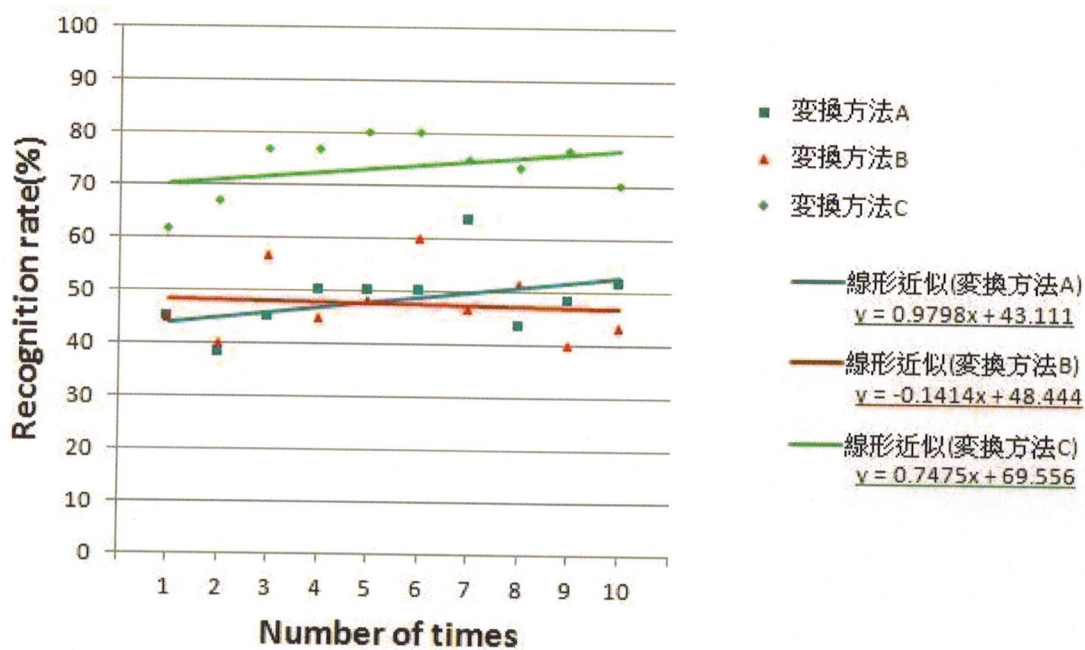


Fig.4.2 テスト回数ごとの認識率(%)

## 4.3 実験 2 : 図形の学習とその音の解析

### 4.3.1 実験目的

本実験は大局的な視点での音の学習と解析を行うことの検討を目的とする。大局的な要素としていくつかの簡単な図形を使用した。図形の音を学習し、新たに提示された画像の音について解析をする。

学習による応用が可能ならば、いくつかの図形を学習することで学習した図形の音を応用して未知の図形の形の違いや変化を認識することが可能であると考えられる。本実験では変化の元となる音のパターンを学習し、学習した図形の縦横の長さなどを変形させたことを音から認識することに応用できるかどうかを検証する。図形の変化は音への変換アルゴリズムから連想してもらう。ゆえに、変化を認識するための尺度は変換方法ごとによって異なる部分が存在する。よって、本実験の結果は各変換方法の特性も大きく影響しており、変換方法ごとの有効性を検証することが可能であると考えられる。

### 4.3.2 方法

#### (i) 実験被験者

実験の被験者は実験 1 に参加した被験者 5 人であった。

#### (ii) 手順

本実験では、被験者に図形を変換した音を聞いて変換前の図形を連想、解答してもらった。学習過程で被験者には Fig.4.5 の図形を変換した 4 種類の音を与えた。視覚刺激の提示、各音の与えられる順番、リピートの方法は実験 1 と同様の方法であった。認識テストでは、学習過程で使用した図形を変形し音に変換した全 16 個の図形を与えた(Fig.4.6)。1 回のテストで出題される問題数は 4 問とし、各問の回答中であれば被験者の任意で音のリピート再生は可能であった。また、テスト中は学習過程で使用した視覚刺激は与えない。音の学習を行うために、テスト終了後に被験者には出題された音の変換前の図形を提示した。音はヘッドフォンより出力され、リピートなどの操作はキーボードより行う。実験回数は 8 回であり、各変換方法に対して行われた。

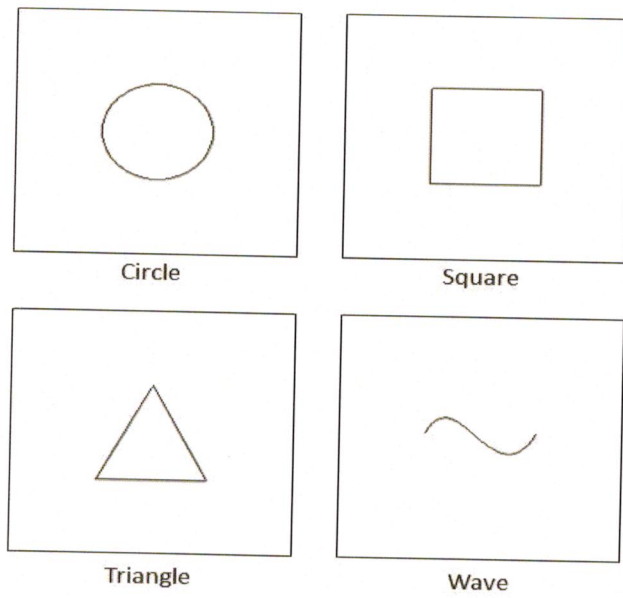


Fig.4.5 実験2で使用された学習用の図形

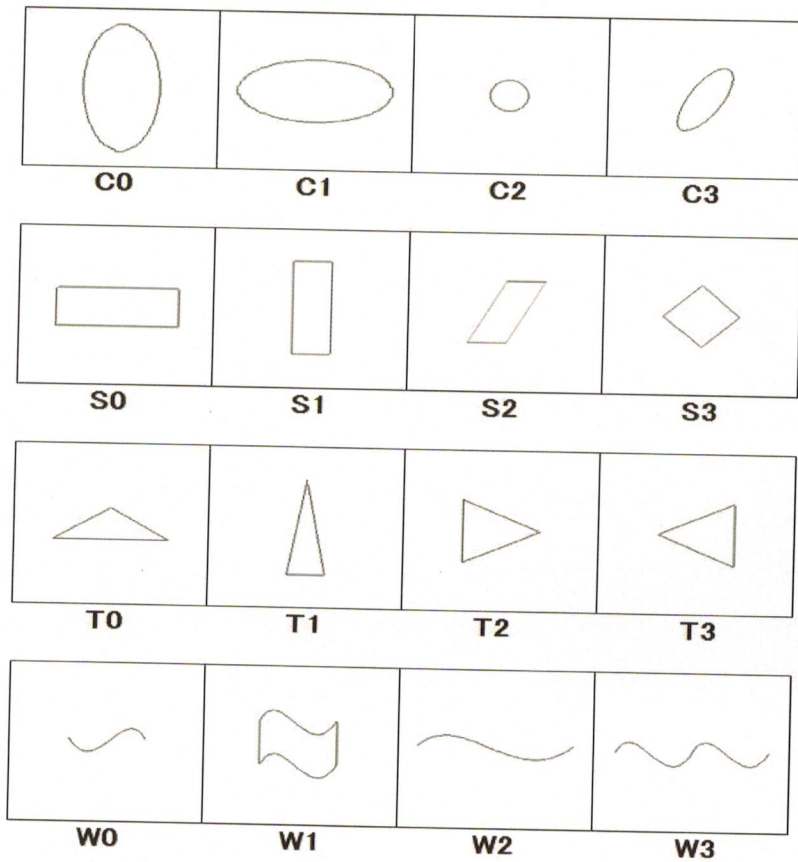


Fig.4.6 実験2で使用されたテスト用に變形された図形：各画像の下は画像名。

### 4.3.3 結果と考察

本実験では、実験 1 と同様に認識率を評価するために被験者の回答に対する評価基準を定めた。変形した図形に対して、その変換方法(回転した、引き伸ばされた等)が認識され図形の特徴となる部分をとらえられていた場合に 1 点、完全に図形を認識できていた場合に 2 点を加点するものとした。ひとつの図形に対して被験者全員が 2 点であった場合を認識率 100[%]とした、各図形の認識率を Fig.4.7 に示す。各変換方法の認識率の平均値は、変換方法 A が 57.19[%]、変換方法 B が 48.44[%]、変換方法 C が 58.13[%]であった。

変換方法によって認識率にばらつきが見られるものの、本実験で出題された図形に対してはある程度の認識率の偏りが窺える。C1、S0、W2、W3 の図形は 3 種類の変換方法ともに認識率がそれぞれ 70[%]以上であり、他の図形に比べて認識率が高い。S0、S1 の図形は四角形の縦と横の長さを変更されたもので、横方向に対して平行な線と垂直な線の 2 種類の線によって構成されている。平行線は音に変化がないので、音の特徴から平行線で構成された図形だと連想することは容易であると考えられる。このことは第 3 章の実験 2 の結果からも示唆されている。

図形の横幅は音の鳴っている時間によって判断される。横幅の変化は学習過程で与えられた図形の音と比較することによって、認識することが可能であったことが示唆される。また、C1、S0、W2、W3 は他の図形に比べて横の幅が大きい形をしている。横の幅が広い分だけ、音が鳴っている時間は他の図形の音よりも長くなる。これにより音がどのように変化しているかを識別し、形を認識することが他の図形に比べて容易であったことが考えられる。同じ四角形でも W0 と S1 では、変換方法 B を変換方法 C の 2 種類の変換方法において S0 の認識率が高かったことからこのことが示唆される。しかし、T0 は比較的横幅が広い図形なのにも関わらず、その認識率は変換方法 B 以外の 2 種類の変換方法において平均値を下回った。ここで、被験者の変換方法 A、変換方法 C における T0 に対する回答をそれぞれ Fig.4.8 に示す。横幅が広がったことで斜辺の角度が小さくなり音の変化が緩やかになったことから、ふたつの斜辺をひとつの弧を描く曲線と誤って認識してしまったことが考えられる。対して、C2 や C3 の音はそれぞれ Fig.4.9、Fig.4.10 のような図形が回答された。これらの画像は図形の横幅が狭く、音の鳴る時間が短くなるために音が急激に変化する。これにより、曲線の音を直線の音と誤認してしまったことが考えられ、幅が狭い図形の音の場合は輪郭線が直線か曲線化を識別することが困難になることが示唆される。

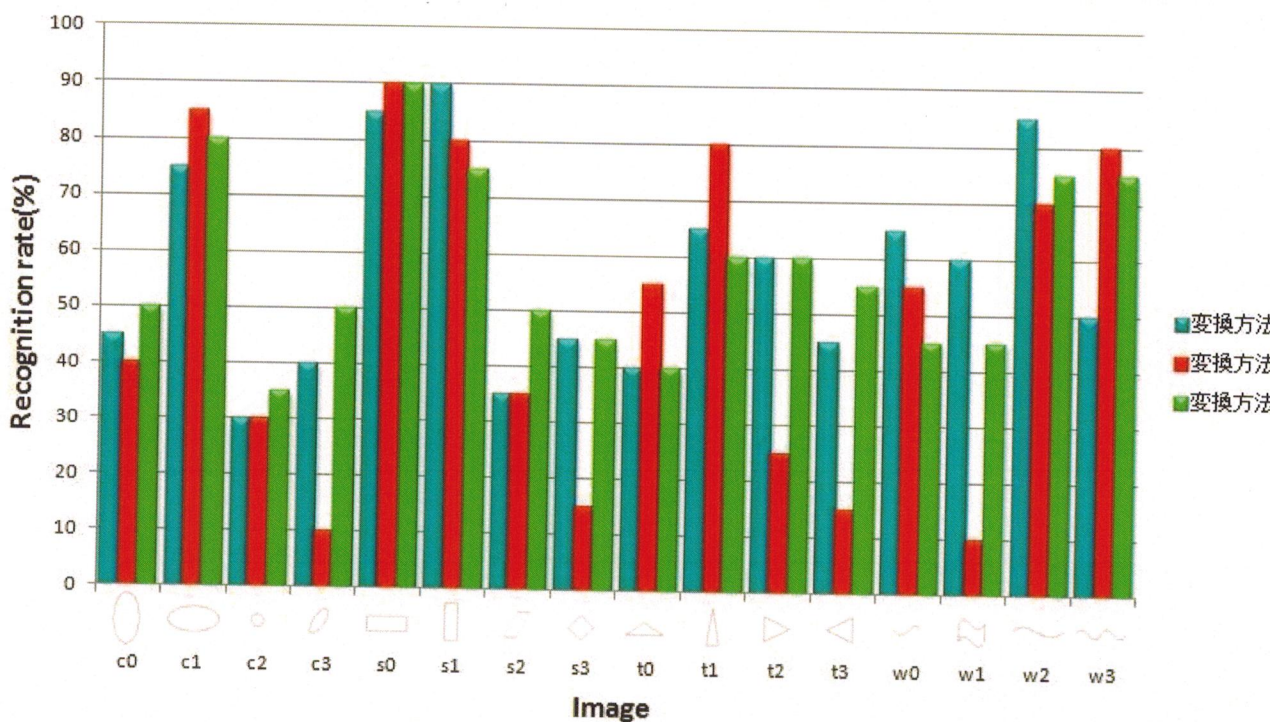


Fig.4.7 各図形の認識率(%)

S3 のように、異なる傾きの線が同時に存在するような図形ではどちらか一方しか認識できないことが各変換方法にそれぞれ見受けられた(Fig.4.11)。これは、我々の耳の感度が周波数に対して一様ではなく、周波数に依存して最小可聴値が異なることが考えられる。よって、同じ振幅の波であっても周波数により我々が感じる音量は異なる。加えて、聴覚フィルタによるマスキング効果が考えられる。感度が高く最小可聴値が低い周波数であった方の線がマスキャーとなり、もう一方の線の周波数の最小可聴値が上昇して聞こえなくなったことが示唆される。

テスト回数による認識率の変化を Fig.4.12 に示した。各変換方法の結果に対する線形近似曲線について、変換方法 A と変換方法 C の線の傾きが正の値であり、その傾き  $\alpha$  はそれぞれ  $\alpha_A = 0.5655$ 、 $\alpha_C = 1.1905$  であった。よって、テスト回数を重ねることで認識率が向上していることが示唆された。これは、テストの過程でいくつかの変形された図形の音を聞き学習した図形の音との違いや音の変化の特徴を学習したことで、他の図形の連想にも応用されていたことが考えられる。

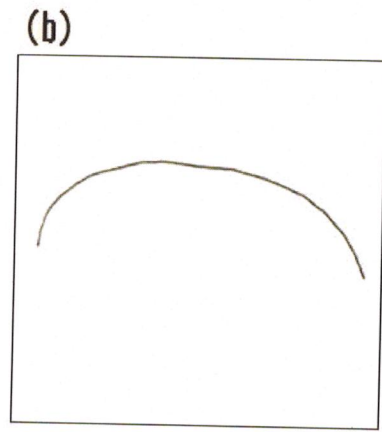
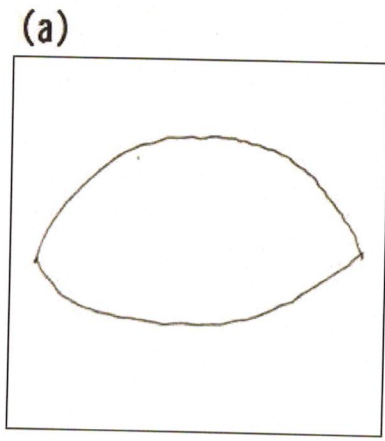


Fig.4.8 T0 に対する被験者の回答 : (a)変換方法 A (b)変換方法 C

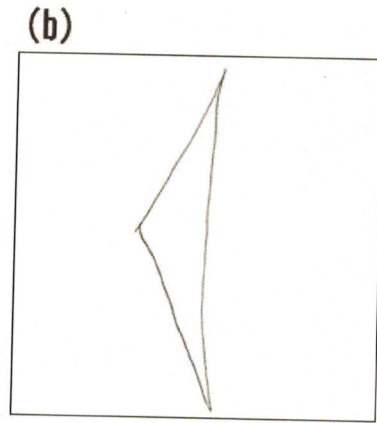
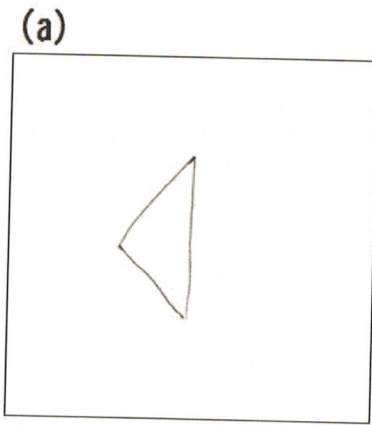


Fig.4.9 C2 に対する被験者の回答 : (a)変換方法 A (b)変換方法 C

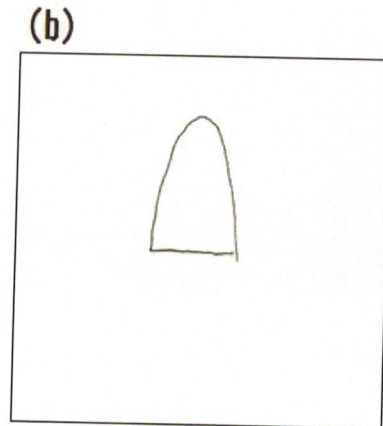
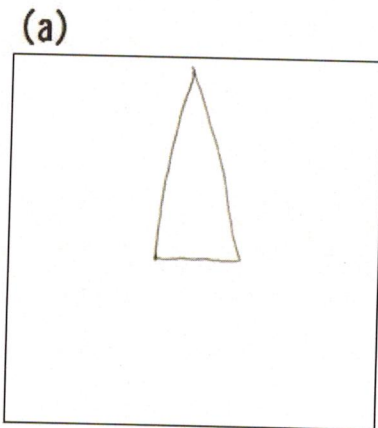


Fig.4.10 C3 に対する被験者の回答 : (a)変換方法 A (b)変換方法 C

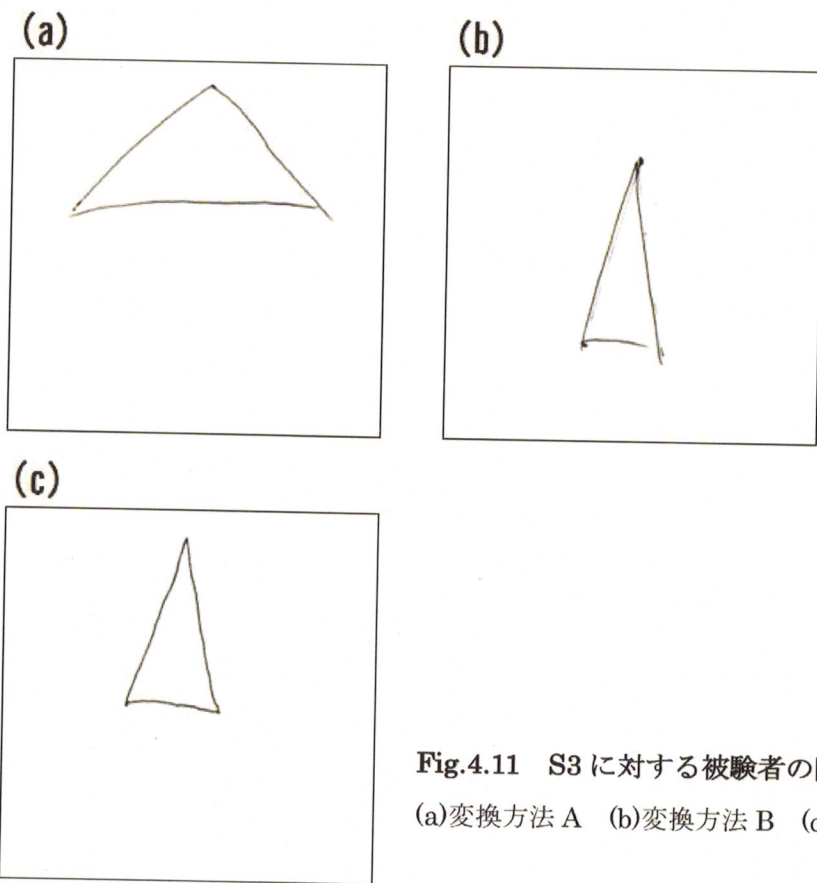


Fig.4.11 S3に対する被験者の回答：  
 (a)変換方法A (b)変換方法B (c)変換方法C

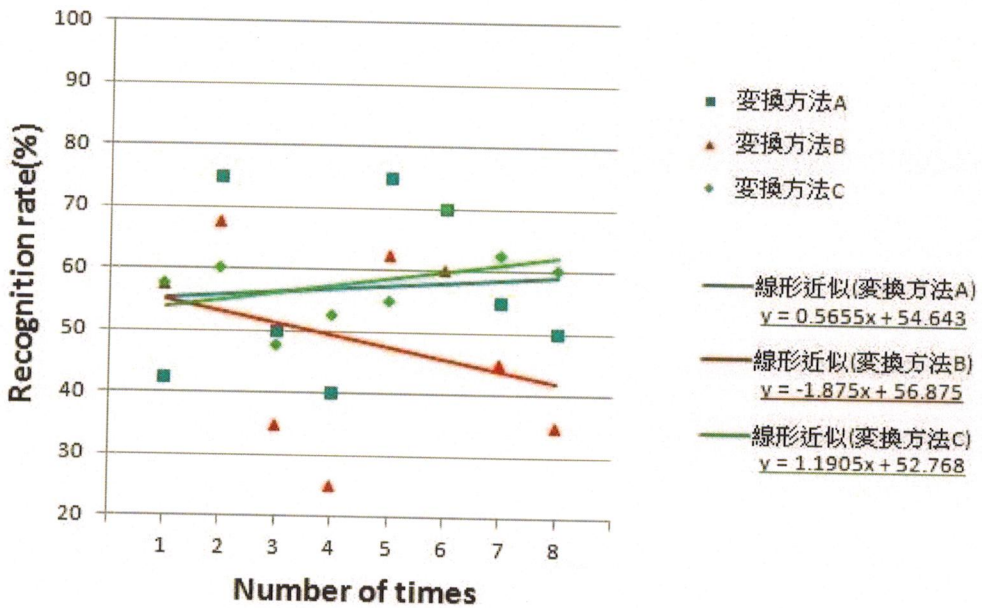


Fig.4.12 テスト回数による認識率の変化(%)



# 第5章 総合考察

## 5.1 まとめ

本研究では、感覚情報の直接的な変換を用いて画像を音に変換する方法を3種類考案し、それぞれの方法について変換した音から形を認識できる可能性を検討した。

初めに、画像に描かれた形に依存して生成される音が異なり、その音の違いを識別することに対して**95[%]以上の識別率が実験より得られた**。このことより、使用者が異なる形の音を識別することが十分に可能である結果が得られた。また、音から連想される形は音の特徴によって異なった。音から連想される形と実際の形との間に関連性が見られた。これらのことから、本手法による画像から情報変換された音は形によって異なる音になり、その音の違いは被験者にとって識別可能であるという、変換アルゴリズムの信頼性が示された。

次に、直線の傾きを音から認識する実験より、変換方法によって精度に差は出たものの直線の傾きを認識することが十分に可能であると考えられる結果が得られた。これにより、線形な輪郭線の変化を音から推測することが可能であることが示唆される。また、**聞いたことがない形の音であっても、知っている形の音を基に連想することが可能である**という結果が得られ、本手法の情報変換で生成された音から汎用性を持った認識が行える可能性があることが示唆される。大きさや輪郭線の変化の大きさなどの形の特徴によって認識率は異なったが、認識の行いやすいものは認識率が**90[%]**になり十分な連想、認識が行われていることが示された。よって、**学習によって応用的な認識を行う効果が見られた**といえる。また、認識率には音に変換される前の形の特徴が大きく影響していることが明らかになった。

変換方法によって違いが生じたものの、回数をこなして認識を行うことによって精度に向上が見られた。応用的な側面だけでなく、精度の面でも学習による向上が示唆された。よって、さらに回数を重ねることで反応時間、**認識効率のさらなる向上が期待できる**。

本研究では画像を音に変換する方法として考案した、3種類の方法をそれぞれにつ

いてその有効性の検討を行った。本研究でのすべての実験を通して、認識率、学習による結果の向上の双方の面において、変換方法 C は 3 種類の変換方法の中で高い結果を得た。特に認識率の平均値については、いずれの実験でも最も高い結果を得ることができた。これらの結果から、変換方法 C が 3 種類の変換方法の中で最も有効性が高いことが示唆された。

以上より、本研究ではあらゆる形の認識に対して十分な有効性を持つには至らないまでも、**本手法で変換した音から画像に描かれた形を認識することが十分に可能であることを実証した**。また、視覚情報を聴覚情報に変換、認識させるような**インターフェースの開発が十分に可能である可能性を示した**。

## 5.2 今後の課題

本研究が今後解決してゆくべき課題について述べる。変換した音から画像に描かれた形を認識するに十分な有効性を得るためには、形の特徴による認識率のばらつきを改善する必要があると考えられる。あらゆる形の大きさ、輪郭の変化にも対応出来るパラメータの最適化が求められる。特に再生時間と音を表現する周波数帯の最適化が重要であると考えられる。

本変換手法での再生時間は画像を走査する速さに対応し、再生時間が短いほど画像の走査は速くなる。認識率の低かった図形は画像の走査速度を減少させることにより、輪郭の変化を認識しやすくなり図形の認識率に向上が見られることが予想される。ただし、走査速度が遅すぎると全体の形が認識しづらくなることが考えられる点と、本研究の認識実験でも緩やかな傾きの直線を曲線と誤認識してしまうことが見られた点からも、単に走査速度を減少させるのではなく最適な速度を検証することが必要になることが考えられる。また、縦に複数の線が存在する場合に正確に複数の線を認識しづらいことが示唆されたが、聴覚の感度が良い周波数帯の範囲で画像を表現することにより緩和できることが考えられる。一様に最小可聴値が低い帯域であれば周波数による音圧の差が低いことが予想される。ただし、マスキング効果は周波数が近いほど効果が大きいので、その点を考慮した音波の振幅を設定する必要があることが考えられる。

視覚障害者のためのインターフェースとしての用途を考えると、現在の静止画に対するアルゴリズムを動画に応用させることが必要と考えられる。

## 5.3 応用的意義

聴覚から視覚の情報を得ることの有用性は視覚障害者に限られたことではない。超からの情報は、何らかの動作をしながら自然と得ることができる。よって、何らかの視覚を使った動作をしながらも、視線とは別の方向の情報を得ることが可能である。例えば、自動車を運転しながら運転手の死角となる部分の情報を音から得ることでき、歩行者の死角からの飛び出しなどの不意の事故などを防ぐことが期待される。視覚情報を聴覚情報に変換するようなインターフェースが開発されることによって我々人間にとって十分に有用な影響があることが考えられる。

また、本研究で行われた情報変換は画像に対する前処理も少なく、高性能な処理装置や大がかりな機器を使用することはない。よって、処理能力が低い装置でも十分に利用できる可能性があり、小型のデバイスで利用できることが期待できる。

# 参考文献

- [1] 河合良浩, 富田文明: “視覚障害者のための三次元物体認識支援システム”, 電子技術総合研究所彙報, 62, 4, 179-186, 1988
- [2] 小林真, 太田道男: “音像定位を利用した能動的情報取得が可能な視覚代行装置”, バイオメカニズム学会誌, 21, 1, 39-42, 1977
- [3] 中村和弘, 青野嘉幸, 田所嘉昭: “視覚障害者用誘導型歩行支援システム”, 電子情報通信学会論文誌 D-II, J79-D-II, 9, 1610-1618, 1996
- [4] カルワハンディ ササダラ, 田所嘉昭: “画像処理を用いた携帯型視覚障害者歩行支援システムの検討”, 映像情報メディア学会誌, 55, 11, 1499-1505, 2001
- [5] 鈴木悠司, 平岩裕康, 竹内義則, 松本哲也, 工藤博章, 大西昇: “視覚障害者のための環境内の文字情報抽出システム”, 電気学会論文誌 C, 124, 6, 1280-1287, 2002
- [6] 田中誠, 後藤英明: “視覚障害者用のウェアラブルな文字認識デバイス”, 情報処理学会研究方向くコンピュータビジョンとイメージメディア, 2008-CVIM-165, 125-130, 2008
- [7] 渡辺哲也, 久米祐一郎, 伊福部達: “触覚マウスによる図形情報の識別”, 映像情報メディア学会誌, 54, 6, 840-847, 2000
- [8] 渡辺哲也, 小林真: “視覚障害者間の図形情報伝達用電子レーズライタの開発”, 電子情報通信学会技術研究報告, HCS200-10, 27-32, 2000

# 謝辞

本研究を進めるにあたって、指導教員であるルジェロ・ミケレット准教授には、大変なご指導をいただきました。心から感謝いたします。ありがとうございました。

指導教員でないにもかかわらず様々なご助言、ご指導を下さった基盤科学コースの先生方には心からお礼申し上げます。ありがとうございました。

また、本研究の被験者を引き受けていただいた方々にお礼申し上げます。ご協力ありがとうございました。