

聴覚によるパターン認識のための表示平面の検討*

伊 東 一 典**

1. ま え が き

一般に、種々の画像信号は実平面の光学的変化として表示され、視覚により認識される場合が多いが、我々はこの信号を聴覚により認識させる方法について検討し、その可能性を追求している^{(1),(2)}。

感覚情報には空間的要素と時間的要素があり、視覚が双方をかなりの程度に感知できるのに対し、聴覚は時間的要素が主で、空間的要素の把握能力は少ないといわれている。しかし空間に存在する音源位置を知ることができる音源定位⁽³⁾は空間的要素にはかならず、両耳への入力信号に位置に関する情報が含まれているためである。

そこで、我々は逆に、この入力音信号を人工的に合成すれば聴覚により認識される空間(聴覚空間)内の任意の場所に点音源を設定したり移動させたりすることができるものと考えた。更にこの方法によれば、合成点音源を画素としてこの空間に任意のパターンを表示することが可能であると思われ、視覚を用いない画像表示、認識技術として、視覚障害者のための視覚代行⁽⁴⁾も含めて広い応用分野が開けるものと思われる。

ところで、画像信号を聴覚に認識させる試みは種々行われており、古くは d'Albe の研究⁽⁵⁾、最近では Beddoes らの装置⁽⁶⁾があるが、これらは文字信号伝達に限定され文字の特徴をあらかじめ文字と対応させた符号化音に変換する方式を用いているため任意のパターン表示は行えない。

我々は前述の発想の基に、数年来任意のパターンを表示する方法について検討を重ねており、既に片仮名などの文字表示を含めパターン表示の可能性を見出している^{(1),(2)}。ところが、ごく最近、この種の考えを適用した影像知覚補助装置が Fish⁽⁷⁾によって試作された。しかし、システムの完成を急いだためか、画像入力系については種々検討がなされているものの、最も重要と思われる画像出力系に関する検討がほとんどなされていない。すなわち、水平方向は両耳への入力音レベル差、垂直方向は周波数変化を用いて表示平面を構成しているが、その要素量と表示点位置の関係をはじめ、構成に必要な基本的事項に関する検討がないため効果的表示が行われているとは思われないのである。

本論文は聴覚に画像表示を行うための表示音源要素、特に、表示平面構成要素の検討に関するもので、合成点音源を画素とする画像表示点数の測定など心理物理的手法による新たな試みについて述べている。

* 昭和50年9月 電子通信学会信越支部大会
昭和51年10月 電子通信学会電気音響研究会において発表

** 電気工学科 助手
原稿受付 昭和52年9月28日

2. 表示面の構成要素について

音源の方向識別がなぜできるのか、すなわち音源の定位要素について、古くは心理学の立場から、最近では神経生理学の面⁽⁸⁾からも明らかにされつつある。

音源位置の把握は両耳により行われることから聴覚空間を図1のように頭を原点として左右(X成分)、上下(Y成分)、遠近(Z成分)の3感覚に分解して考えることができる。

X成分(左右感)は両耳への入力音レベル差、位相差、到達時間差の単独かあるいは複合した適当な要素量で決定される。本実験に採用したヘッドホン受聴では各要素量を単独に変化できるが、このとき、音源は要素量の変化に伴い後頭部周囲を回転して知覚されることから座標のX成分は頭を中心にして正後面を零度とする方向角で表示するのが現実に即している。従って、以後は図1に示すような角度表示とする。X成分の各要素についてはステレオ技術に関連してその定位特性が種々報告されており⁽⁹⁾、自由音場では各要素の定位への寄与は音源周波数により異なり、約1,500 Hz以下では位相差が、それ以上では回折、陰影効果が顕著となるためレベル差が重要な要素になると考えられている。

Y成分(上下感)についてはどのような要素が定位に寄与するのか明確でない。聴取者が周囲環境固有の音の反射、回折特性を経験、体得してはじめて形成されるものといわれ、その手がかりが耳介にあるとの報告⁽¹⁰⁾もあるが、音源、聴取者が共に静止している場合は不正確であるといわれている。従って人為的にこの音源要素を合成することは今のところ不可能である。一方、あまり研究がなされていないが心理的な上下感覚の存在も重要である。これは高い周波数の音は高い位置に、低い周波数の音は低い位置に音源を感じるものであるが、

実空間の位置と音源周波数が絶対的に対応している人は少なく、多くの場合相対的な感覚として捕えられている。

Z成分(遠近感)は入力音の相対レベルが重要な定位要素であり、このほかに直接音と間接音の強度差、周波数スペクトルの変化なども定位に関与しているといわれている。この定位も相対的なものである。

ところで、我々の目的である平面的な画像の表示のためにはこれら3つの感覚の中から2つを選んで表示平面を構成すればよい。そこで種々の音源要素により表示点音源を合成してそれぞれの要素の定位効果を検討し、最も効果的な表示平面の構成要素を見出すことにした。

実験には音源合成の容易性から表示音として純音を用い、X成分は両耳への入力音レベル差、位相差および到達時間差、Y成分は音源周波数変化による心理的上下感、Z成分は入力音の相対レベルを検討要素とした。又、表示はすべて完全密閉形ヘッドホン(パイオニアSE405)により行い、表示音は表示周波数 f (Hz)、表示時間 T (sec)、立上り立下り時間 τ (sec)、ラウドネスレベル L (phon)で特徴付けた。

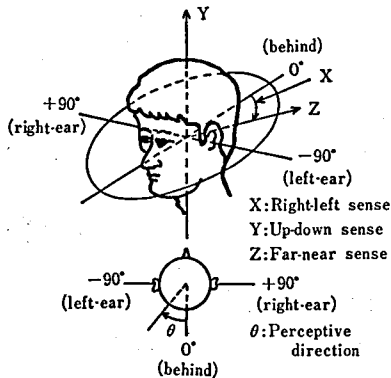


図1 聴覚空間の3成分

3. X成分の基本特性

前述のように既知の定位要素の中では、絶対的特性を有するのはX成分のみである。そこで、我々はまずこの成分を基本として検討を進めることにした。

初めに入力音レベル差、位相差および時間差について、それぞれの要素量と表示点位置の対応関係を5名の被験者（男性20代）について調べた。測定は両耳への入力音間に各要素による差を付加して被験者に聞かせ、この合成音が到来すると感じる方向を指示してもらった。付加する各要素量によりこの方向は異なるが、被験者に聞かせる音の順序は乱数表により定め、特定の順序による特性の出現を防いだ。表示音は70 phon一定で、表示間隔は前表示音の影響がないよう配慮して5秒とした。又、同一要素量の表示点を一被験者当たり5回聞かせ、その平均値を求めている。なお表示の前に特に基準音を聞かせることはしなかった。

図2(a)は両耳への入力音レベル差と知覚された音源方向の関係である。特性は正後面を中心にほぼ左右対称のため左耳側のみを示した。この場合、左耳入力>右耳入力であり、レベル差の増加と共に知覚方向角度は増大するが、約12 dB SPLで-80度付近に飽和していることが分かる。又、表示音の周波数と共に知覚方向角度はわずかに小さくなる傾向を示していることが分かる。これは表示時間は異なるが永井ら⁽⁹⁾の結果とほぼ一致している。

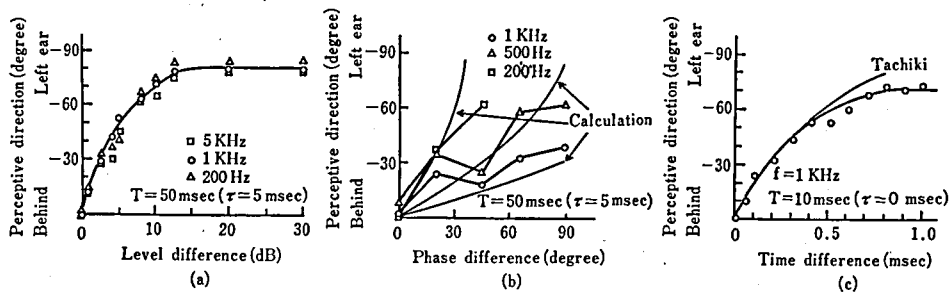


図2 両耳への入力音要素量と音源知覚方向

図2(b)は両耳への入力音位相差と知覚される音源方向の関係である。前図と同様にはほぼ左右対称となるので左耳入力が進相の場合について示した。図中の計算値は自由音場において音源までの距離を無限大としたときの入力音の位相差 $\Delta\varphi$ と音源の方向角 θ の関係を示す次式から求めたものである。

$$\Delta\varphi = 2\pi fl \sin \theta / v$$

ここで音源周波数 f 、両耳間隔 l （平均17cm）、音速 v （340m/sec）である。測定値は大きな傾向として計算値と合うのみで固有の癖を有し、位相差のみによる定位効果のあいまいなことを示している。なお、永井らによる定常状態信号に対する実験結果⁽⁹⁾においてもばらつきが多く一貫性のある傾向は見出されていない。

図2(c)は両耳への入力音の時間差と音源の知覚方向の関係である。この場合もほぼ左右対

称であり左耳入力が先行する場合を示した。入力音時間差約 0.8 msec 以上で知覚方向角度は飽和している。なお、図中の細線は立木ら⁴⁾が音源から両耳への送話管の長さを変えて時間差を作り、不規則な破裂音を用いて行った実験結果で小角度では一致していることが分かる。

以上の結果から、要素量と知覚方向の関係が比較的単調でばらつきの少ない入力音レベル差および時間差が我々の応用に適していることが明らかとなった。ここで有効な表示範囲は両者ともに、ほぼ -80 度から +80 度となっており、適当なレベル差あるいは時間差を付加することにより一義的にこの範囲内の任意の場所に表示点が構成できる。

そこで次に、この2つの要素についてそれぞれの要素量の表示音を聴取したときに知覚される方向の角度分布を測定した。

まず、前述の結果より、2要素量それぞれについて -80 度から +80 度の範囲に等角度に 51 個の表示点を設定した。この中から数個の測定点を選び、1 点ずつ多くの任意点の中に入れて表示してその測定点が 51 種の表示点の中のどの点として知覚されるかを調べた。実際には、任意点 900 個、測定点 100 個の組合せを 200 個を 1 問とする 5 問構成にして表示を行い知覚方向角の分布を求めた。又、各問間に 20 分の休みをおいた。実験は 3 名の被験者に試み、表示音周波数は 1 KHz で音圧は 70 dB(c) 一定とした。表示音は無作為な順序で 2 秒間隔に呈示した。

図 3(a)は要素を入力音レベル差としたときの 2つの測定点、すなわち左耳入力/右耳入力が 0 dB 及び 2 dB について得られたそれぞれの知覚方向分布を例示したもので、ほかの測定点の特性もほぼ同様であり、ほぼ正規分布と見なせることを確かめている。また分布の特性は入力音時間差についても同様であった。このような分布特性を X 成分の 2 要素について幾つかの測定点ごとに得て、各々の分布の 50% となる角度幅を方向弁別角度幅と名付け画像表示密度の目安にすることにした。

図 3(b), (c)は各々の入力音レベル差及び時間差に対して得られた方向弁別角度幅と音源知覚方向の関係を示したものである。これは表示時間が 1 msec 及び 10 msec の結果を示したもので、ほぼ左右対称の特性となるので正後面より左耳側のみを示した。両図とも関数の形

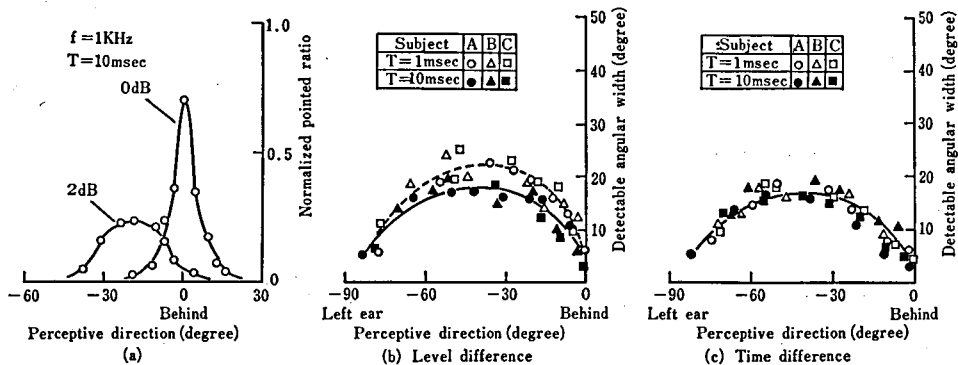


図3 音源知覚方向弁別角度幅

に幾分の差異はあるものの3名の被験者ともほぼ同様な特性を示し、方向弁別角度幅は正後面、左耳側で小さく、中間部では大きい傾向を示している。レベル差では表示時間による差異が顕著であるが時間差ではほとんど見られない。

図4は図3(b), (c)における方向弁別角度幅の最大値について表示時間依存性を測定した結果であり、3名の被験者についての測定値およびその平均値を示している。方向弁別角度幅の最大となる音源知覚方向角度は図3(b), (c)で明らかなように個人差が若干見られるが、ほぼ -45 度付近であった。図の左目盛は最大方向弁別角度幅を示しており、右側に目盛った値は全表示範囲(-80 度 $\sim +80$ 度)に最大方向弁別角度で等間隔に表示し得る表示点数である。この表示点数は少なくとも各点の表示において、その点が50%以上の認識率で聴取者に弁別され得るので有効表示点数と呼ぶことにする。

図4において最大方向弁別角度幅の値には個人差がほとんど見られず ± 2 度程度となっている。又、入力音時間差の特性が表示時間にほとんど依存しないではほぼ一定であるのに対し、レベル差の特性は10 msec以下で表示時間の減少と共に最大方向弁別角度幅が大きくなっており、この範囲で定位が急にあいまいになることが分かる。これは時間差が過渡的な信号で位置情報を伝達できるのに対して、レベル差ではある程度の表示時間を必要とすることを意味している。又、この図からX成分の有効表示点数はごく短い表示時間領域を除けば、いずれの要素でも約9点であることが分かる。

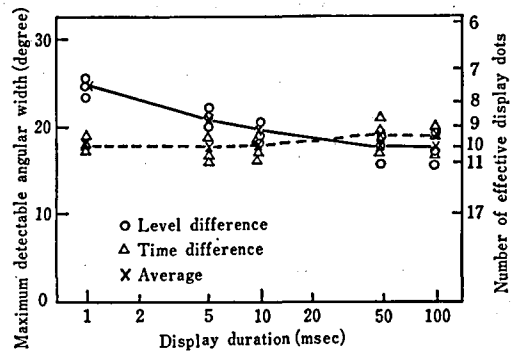


図4 最大方向弁別角度幅と有効表示点数

4. 表示平面の検討

表示平面の検討のためには、Y, Z成分についてもX成分と同様に要素量と表示点位置の関係が明らかになればよいのであるが、既に述べたように、両成分は相対的な感覚であり絶対的に表示点位置を決定することが困難であり、前述のような試みを適用できない。しかし、相対的な感覚であっても幾つかの表示点の弁別の可能性は考えられるので、とりあえずX成分について得られた有効表示点数9点をそのままY, Z成分にも適用して、各成分におけるこれらの点の認識の程度からその表示平面構成成分としての評価を行うことにした。

まず、両成分について、これらの点の最適抽出方法を検討し、続いてX成分も含めた3成分について、表示時間と表示点認識率の関係などの測定を行い、各成分の特性の比較検討を行った。

4.1 実験方法

実験は9個の各表示点に記号を付してこの点の中から任意の20点を乱数表により抽出して2秒間隔で被験者に表示し、その記号を筆答してもらった。その結果から、正答の割合、すなわち認識率を求め表示効果の評価基準とした。認識率は表示した点を正しく指摘した場合

を正認識率とし、更に正解の両隣の点を指摘した場合も含めた認識率を準認識率と名付けた。正解とする範囲から考えて、準認識率の値はおよそ正認識率の2倍になると考えられる。

ところで、簡単な文字である片仮名やアルファベットの表示には 5×5 程度の描点が必要である。9点表示における準認識率はほぼこの4~5点の認識率に相当するものであり、又、正認識率50%は4~5点の認識に十分な値と考えられるので画像表示の可能性を論じる上での1つの目安となる。

ここでは繰返し多人数に実験できるように表示音をテープ録音して用いている。実験は両耳とも健全な10名の被験者(男性20代)を対象とした。

4.2 表示音周波数依存性

図5はX成分における表示点認識率の周波数依存性を示したものである。9つの表示点は、レベル差0, ± 1 , ± 3 , ± 6 , ± 12 dB及び時間差0, ± 0.1 , ± 0.3 , ± 0.5 , ± 0.7 msecの要素量で構成した。これは図2(a), (c)から分かるように、正後面を中心にして両耳間にはほぼ等角度(20度)に9点を選んだことに相当する。ここで表示時間はレベル差では50 msec, 時間差では10 msecとしているが、これは図4から両者が同程度の有効表示点数となる表示時間である。レベル差表示において正認識率は100 Hzから2 KHzで比較的平坦な特性を示し50%を越えており、準認識率は100 Hzから5 KHz程度まで90%近い値を示し周波数依存性がほとんどないといえる。一方、時間差でも特性は比較的平坦であるが、ほぼ1 KHz以上の周波数で著しく表示効果が低下していることが分かる。なお図では平均値のみを示したが、個人差は全周波数領域で $\pm 10\%$ 程度見られた。

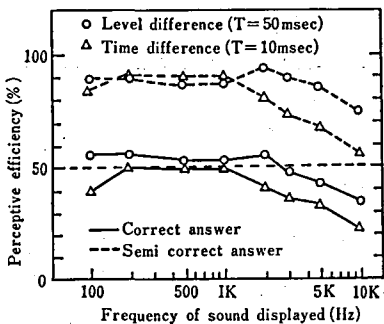


図5 表示音周波数と認識率(X成分)

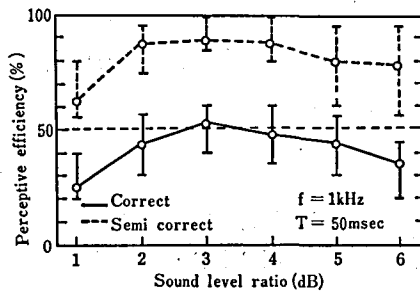


図6 表示音圧差と認識率(Z成分)

図6はZ成分についての結果である。Z成分においては入力音強度の相対値が定位要素であることから、表示点の抽出は両耳への入力信号を同一として音圧差のステップ点列とした。ここでは常に基準を70 dBに取り、これより上下に種々の音圧差でそれぞれ4つの表示点を構成し合計9点とした。図6は表示点抽出音圧差と認識率の関係を示したもので、正認識率、準認識率ともに類似の特性を示している。音圧差3 dBが最適で正認識率が50%を越えており、認識率の幅も測定音圧差のうちで最も小さく個人差が少ないことを意味している。なお表示音周波数特性は各音圧差とも比較的平坦でほとんど周波数に依存しなかった。

図7はY成分の周波数変化による心理的上下感覚による結果である。表示点は被験者の音

感を狂わせないよう配慮して身近な音階を用いることにした。又、前述の2成分の認識率の表示音周波数依存性の結果も考慮して、261.63 Hz (中央ハ) を最低音として周波数比 $f_n : f_{n+1} = 1 : 1.12$ (長2度), $1 : 1.26$ (長3度), $1 : 1.32$ (完全4度) からなる3種の平均律音階を用いて各々9つの表示点を構成した。図から明らかなように認識率は音程に依存しており、正認識率が50%を越えて、かつ個人差が最も少ない長3度が最適音程となっている。

4.3 表示時間依存性

次に前節の結果から比較的認識率の高い要素、すなわちX成分はレベル差と時間差、Y成分は長3度音程、Z成分は音圧差 3 dB により、それぞれ前節と同一の9つの表示点を用いて表示時間と認識率の関係を調べた。

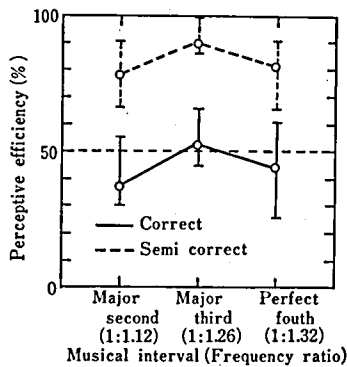


図7 音程と認識率 (Y成分)

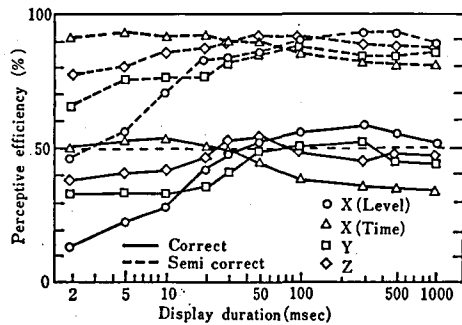


図8 3成分における表示時間と認識率

図8はその結果をまとめて示したものである。X成分のレベル差と時間差の特性は図4の表示時間と最大方向弁別角度幅の結果と類似した特性を示しており、正認識率が約30 msecを境に短時間表示と長時間表示で逆転している。一方、Y、Z成分は短時間表示領域で認識率は若干低下するが、比較的平坦な特性を示している。特にZ成分は30msec程度で正認識率が50%を越えており、この要素の情報が短時間で得られていることを示している。

4.4 考察

図5に示すX成分の表示音周波数依存性の結果において、レベル差の表示効果が表示音周波数にあまり依存しないのはヘッドホン受聴のため頭や耳介による音波の回折、陰影効果をほとんど受けないためで、我々の応用には好ましい結果となっている。一方、時間差の特性は高周波領域で認識率が次第に低下しているが、これは両耳に近い表示点に関して過渡的な時間差の部分で知覚される音源方向と、その後続く表示音の定常状態部分の位相差により知覚される音源方向との間に相違を生じるために起るものと思われる。

ところで、Z成分においては音圧差の検出感度が表示音の大きさに依存するものであり、表示音の総合音圧が良好な可聴範囲内で、かつ各表示点の音圧差が検出感度を上回っていることが表示音構成の条件と思われ、図6において音圧差 3 dB で最大の認識率を得た原因と考えられる。これは表示音周波数とその周波数差検出についてもいえることであり、図7で長3度音程が最適な表示音周波数差、すなわち表示間隔となっている。従って、Y、Z成分

では可聴範囲、特にその中でも不快感の少ない聴感度のすぐれた範囲において長3度あるいは音圧差3dBで構成できる表示点数が画像表示の1つの目安になる。

図8に示した認識率の表示時間依存性の結果は各成分の性格の差異を示している。X成分は表示時間依存性が顕著で、特にほぼ30msecを境にしてレベル差は短時間表示領域、時間差は長時間表示領域ではほかの2成分に比べて正認識率が小さな値となっている、この特性は準認識率においても顕著であり、表示点の知覚方向が大幅に乱れていることを示すもので、これらの領域での表示点の位置決定の困難なことを表している。一方、Y、Z成分はほぼ同様に比較的平坦な特性を示しており、短時間表示領域でも認識率の急激な低下は見られない。

これは、X成分は感覚的な量のみであるため情報の記憶効果が少なく、表示時間内に位置情報を分析して位置判定する必要があるため表示時間依存性が著しいが、Y、Z成分は心理量により構成されているため表示時間内に既成記憶情報との照合が行えれば一種の記憶効果となり、表示時間外にも位置判定を行えるという感覚要素の本質的な差異に起因するものではないかと思われる。

なお各成分の認識率が長時間表示領域でわずかに減少傾向を示しているが、これは順応効果によるものと考えられる。

以上、各成分の要素量により構成したそれぞれ9表示点の1次元認識特性の結果から、Y、Z成分にとりあえず適用した9点は最適抽出条件の下では、X成分と同様な表示点認識率を納めることが明らかとなり、Y、Z成分の有効表示点数もほぼ9点と考えて良いと思われる。これにより9×9画素の表示パネルが構成できるものと思われ、最終的な画像表示を考えても簡単な文字、図形はこの程度の表示点数で表示が可能である。

従って、平面的な画像の表示のためには、これら3成分の内の最適2成分を用いて表示平面を構成すればよいことが明確になった。前述までの結果から、表示平面の構成要素として少なくとも1成分は絶対的な表示点位置が把握できる要素、すなわちX成分のレベル差あるいは時間差を適用するのが好ましいと考えられる。そこで表示平面としてX-Y面およびX-Z面の可能性が考えられる。

まずX-Y面については、X成分としてレベル差を用いた場合、レベル差による表示点認識特性がほとんど表示音周波数に依存しないことから十分構成の可能性を持つと思われる。一方、X成分に時間差を用いた場合には表示音周波数範囲がレベル差に比べて限定される。表示音周波数範囲は図5から分かるように前者ではほぼ100Hzから2KHz、後者ではほぼ200Hzから1KHzの範囲で認識特性が平坦で最適である。又、Y成分はこれらの範囲内に長3度音程で9点程度の構成は可能である。ところでFishのシステム⁽⁷⁾の垂直方向の周波数範囲は最初のシステムでは70Hzから7KHzが、その後のシステムでは200Hzから5KHzが用いられているが、これは我々の得た最適表示音範囲に比べてやや広い領域を用いていることになる。

次に、X-Z面については、X成分にレベル差を用いた場合はZ成分の入力音強度が小さくなるにつれ、両耳への入力音レベル差の検出能力が低下するため位置情報の伝達が困難になり適切な表示平面の構成は不可能と思われる。しかし、X成分が時間差の場合はこのような問題はなく構成可能と考えられる。

以上、表示平面としてX-Y面とX-Z面の可能性について述べたが、この2平面の内ど

これらの要素の組合せが最適表示平面となるかは、今後、更に2次元の認識特性、1次元、2次元の繰返し表示、連続表示効果など画像表示に関する総合的な検討の必要性がある。

5. あとがき

従来、種々のパターンは実在する平面に表示され、主として視覚、ときには触覚により認識されるのみであったが、聴覚の空間知覚である音源定位効果に着目して、合成点音源を画素として聴覚により認識される空間にパターンを表示する新しい画像表示技術の可能性について基礎的検討を行った。その結果、以下に述べることが明らかになった。

- (1) X成分の有効表示点数は表示音に純音を用いた場合約9点/(-80度~+80度)である。
- (2) 上記の表示点数はY, Z成分にも十分適用可能である。
- (3) 表示音の最適周波数範囲は各成分で若干の違いはあるがほぼ200Hzから2KHzである。
- (4) X-Y面およびX成分に時間差を用いるとX-Z面が表示平面としての可能性を持つ。終りに、表示点の認識特性には個人差が見られ、被験者の中にはすべての成分について80%程度の高い認識率を示すものもあった。今後、個人差の原因の解明、学習の効果などについて検討する必要がある。なおデータの処理に当ってはFACOM230-25を使用した。

謝辞 本研究の共同研究者であり、常に御指導賜る信州大学工学部米沢義道助教授に厚く御礼申し上げる。又、御助言頂いた信州大学工学部柳沢武三郎教授に深謝する。

本研究の一部は放送文化基金および長野県科学振興会の助成によるものである。

参 考 文 献

- (1) 伊東, 米沢: "聴覚へのパターン表示方式の研究", 信学技報, MBE75-8 (1975-06).
- (2) 伊東, 米沢: "聴覚へのパターン表示", 信学技報, EA76-32 (1976-10).
- (3) 森本政之: "音の方向定位について", 音響技術, 7, p.19 (1974).
- (4) P. W. Nye and J. C. Bliss: "Sensory aids for the blind: A challenging problem with lessons for the future", Proc. IEEE, 58, 12, p. 1878 (1970).
- (5) E. E. F. d'Albe: "The type-reading Optophone", Nature, 94, 3, p. 4 (1914).
- (6) M. P. Beddoes and C. Y. Suen: "Evaluation and a method of presentation of the sound output from the Lexiphone-A reading machine for the blind", IEEE Trans., BME-18, 2, p.85 (1971).
- (7) R. M. Fish: "An audio display for the blind", IEEE Trans., BME-23, 2, p.144 (1976).
- (8) 勝木保次: "聴覚系における情報処理", 医用電子と生体工学, 10, 2, p.17 (1972).
- (9) 永井, 佐藤, 岩崎: "立体録音に於ける方向感の実験的検討", 音響誌, 10, 3, p.149 (1954).
- (10) C. L. Searle, L. D. Braida, D. R. Cuddy and M. F. Davis: "Binaural pinna disparity: another auditory localization cue", J. Acoust. Soc. Amer., 57, 2, p. 448 (1975).
- (11) 電子通信学会編: "聴覚と音声", p.134, コロナ社 (1966).