

# ウェアラブルのための聴覚定位によるインタラクション手法

広田 光一<sup>†</sup> 廣瀬 通孝<sup>†</sup>

計算機などの情報機器を身につけて利用するウェアラブルの考え方が注目されるようになってきた。ウェアラブルによる情報提示では、従来の情報機器と異なり、ながら作業やハンズフリーな状態での使用などの可能性が指摘されている。聴覚情報はヘッドホンのような日常的でウェアラブルな装備により提示が可能であることから、ウェアラブルのためのインタフェースとしての活用が期待されている。このような背景をふまえて、筆者らは聴覚定位を利用したウェアラブルとユーザとのインタラクション手法を提案する。具体的には、従来の GUI で用いられてきた視覚的なメニューやアイコンとポインタによる処理や対象の選択と同様の操作を、空間定位をとともなう聴覚提示によって実現することを検討している。本論文では、このようなインタラクションの基本となる聴覚的アイコンおよびポインタによるポインティング操作について、被験者を用いた評価実験を通して、精度と時間に関する特性を明らかにしている。具体的には、頭部伝達関数を利用して聴覚情報の方向定位を行うデバイス、頭部および手先の姿勢を計測するセンサ、これらを統括するノート PC より構成されるシステムを用いてウェアラブルな聴覚提示を実現し、これを用いて、定位方向の認識誤差、ポインティング時間、ポインティングにおけるノイズの影響などを定量的に評価している。また、これらの実験を通して、聴覚によるポインティングの実現可能性を確認している。

## Interaction with Wearable Systems Based on Auditory Localization

KOICHI HIROTA<sup>†</sup> and MICHITAKA HIROSE<sup>†</sup>

It is a merit of using wearable systems that we can use computers while doing other tasks. Even before the emergence of wearable computers, we were using auditory devices such as head phone stereos in a wearable style. The audio is hence thought to be a media that is most suitable for wearable interactions. We propose an interaction method between the user and the wearable systems based on the auditory localization. The concept of the method is based on the notion to introduce the selection mechanism of visual GUI into auditory interaction. In our approach, users interact with auditory icons using auditory pointers to select functions and objects. In this paper, we report the experiments on the efficiency and the accuracy of auditory pointing. We used a wearable auditory system, which consists of an auditory localization device based on HRTF, motion sensors to measure the angular posture of the head and the pointing device, and a notebook PC to control these device and sensors. We quantitatively evaluated the accuracy in the recognition of sound orientation, the time required for pointing the target icon, and the effect of noise during the pointing operation. Through these experiments, we confirmed the feasibility of the interaction method based on the auditory localization.

### 1. はじめに

計算機などの情報機器を身につけて利用するウェアラブルの考え方が注目されるようになってきた<sup>1)</sup>。このような利用形態の変化によって、情報機器の役割やユーザとのインタフェースは、従来とは異なったものとなることが予想され、とくに、他の作業をしながら情報機器を利用する、いわゆるながら作業の実現などが期待されている<sup>2)</sup>。このような中で、これまで視覚

情報に焦点をあてた研究が多くなされているが、聴覚情報についてはそれほど注目されてきたとはいえない。聴覚情報の利点として、ヘッドホンのように日常的に用いられている情報提示デバイスが利用可能で、また、アラームなどのように、ながら作業におけるインタラプティブな役割を持つという点をあげることができる。このような聴覚情報は、空間の中での出来事と関連を持つことから、音の発生源の方向の情報は聴覚情報の重要な一部分であると考えられる。これまで人工現実感を目的とした聴覚ディスプレイが開発されてきているが<sup>3),4)</sup>、近年の半導体デバイスの進歩によりウェアラブルとしての実現が可能になってきた。一方、聴

<sup>†</sup> 東京大学先端科学技術研究センター  
RCAST, the University of Tokyo

覚定位をウェアラブルシステムと人間とのインタラクションにどのように利用するかということについて、これまで十分な検討がなされていない。

本論文では、聴覚定位を利用したインタラクションの基本的な要素になると考えられるポインティングに着目し、被験者を用いた評価実験を通して、精度と時間に関する特性を明らかにする。このような特性は聴覚を利用したインタフェースの設計に不可欠のデータであると考えられる。

## 2. ウェアラブルのための聴覚インタフェース

聴覚情報の利用については、これまで主にデスクトップ環境における GUI のユーザビリティの向上および視覚障害者などに対する感覚代行を目的とした研究が行われてきている。この中では GUI 操作に現実とのアナロジーに基づいた音を付加する Auditory Icon の提案<sup>5)</sup>、ウィンドウシステムの GUI 要素の構造を聴覚表現することで視覚障害者の GUI 操作を可能にする試み<sup>6)</sup>、さらに聴覚定位を利用することで GUI 要素のレイアウト情報を聴覚的に表現する手法の提案<sup>7)</sup>などがなされている。一般に聴覚情報は言語的なものと非言語的なものとに大別される。言語的な情報は正確性が必要とされる情報伝達に適するとされ、これを実現する技術として音声認識および合成の研究がなされ、用途を限定すれば実用となる水準に達してきている。一方、聴覚情報の非言語的側面には音程音質や音響などがあり、音楽はこれらの側面に強く結びついた表現形態の端的な例である。人工現実感など空間型インタフェースによる聴覚情報提示においては 3 次元音響の利用が 1 つの研究の焦点となり、科学技術データなどを空間聴覚刺激にマップすることでデータの理解を効率化する可聴化 (Sonification) 技術<sup>8)</sup>や遠隔会議における聴覚定位の利用<sup>9)</sup>に関する検討、3 次元仮想空間での作業効率に聴覚情報が与える影響についての報告<sup>10)</sup>などが知られている。

このような空間的な聴覚定位を実現する技術としては頭部伝達関数を利用する方法<sup>3)</sup>とキルヒホッフ積分方程式に基づく方法<sup>11)</sup>とが代表的なものとなっている。頭部伝達関数による方法は、人間は 2 つの耳により聴覚刺激を受容しこれらの刺激の違いから音源の方向を認識していることを利用する。両耳の刺激の違いは、主に頭部や身体による音の反射や回折により生じることから、これらの影響に相当する伝達関数 (頭部伝達関数) に音源から発せられる音を入力し、その出力をヘッドホンにより耳に直接提示することで音の空間定位が実現される。一方、キルヒホッフ積分方程式

に基づく方法は、広がりのある空間内の音場を再現することを可能とする技術であるが、このためには、この空間を囲う境界面における音圧変化をスピーカアレイなどにより制御する必要がある。したがって、本研究が指向するような装着型の実装は困難である。

ウェアラブルによる情報提示の 1 つのアプローチとして、ユーザの所在位置に応じてこれと関連する情報を提供する手法が多く検討されてきている。この中では屋外の現実環境で位置に応じた情報の提示を行うことで現実を情報的に修飾する領域型展示の概念<sup>12)</sup>などが提案されている。聴覚による情報提示も、博物館などで利用されているポータブル型の音声ガイドなど限定された目的においては、位置に応じた情報提示が実用化されている。これに対して、ウェアラブルとのインタラクション手法についてはいまなお試行錯誤の段階にある。身体の動きを計測する方法<sup>13)</sup>や画像処理を利用する手法<sup>14)</sup>などが提案されているが、聴覚情報を利用したインタラクションについては音声言語処理による方法を除いて十分な検討がなされていない。

筆者らは、聴覚定位を利用したウェアラブルとユーザとのインタラクション手法を提案する。これは、聴覚的なアイコンをユーザの周囲に提示し、これを聴覚情報をもとにポインティングすることで選択操作を行うインタフェースを想定している。アイコンは周囲の空間に配置し、したがって、頭部との相対的な動きを生じるものとする。一方、ポインタはユーザの手先の動きに追従するものとする。上述のようにウェアラブルとのインタラクションは他の作業と共存できる形態が望まれているが、筆者らのアプローチは聴覚のみを利用することで視覚をインタラクションから解放することをねらっている。また、ウェアラブルによる情報提示の 1 つの利点に現実環境と情報との関係づけの可能性が指摘されているが、聴覚定位によっても従来の視覚情報による現実環境の修飾と同様の効果を得ることができるのではないかと考えている。

たとえば、歩行時のナビゲーション情報の提示では、ユーザにとって視覚は環境を確認する最も重要な感覚であることから、情報提示によって視覚の注意を環境からそらすことは危険である。一方、聴覚定位を利用することで方向指示などの情報をより直感的に表現することが期待され、とくにユーザの視野の外を含む全周囲を提示範囲とすることができる。視覚情報の提示においても同様であるが、一度に提示できる情報には限りがあることから階層的な情報提示が必要になると予想される。具体的には同時に提示されている複数のアイコンから 1 つを選択することで、これに関係する

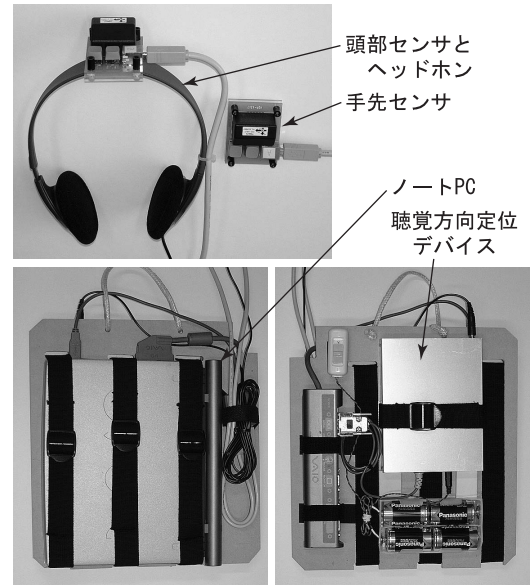
より詳細な情報を参照するなどの状況である．このような提示情報の制御はユーザとシステムのインタラクションの最も一般的な場合であると考えられる．筆者らはこのような選択操作においてはこれを直感的な方法でポインティングすることが適当であると考えた．また，このようなポインティングは選択操作だけでなくシステムに方向などの情報を直感的に伝える手段としても有用であると考えている．上述のように聴覚的なアイコンを空間定位などにより提示する手法は GUI の聴覚表現などですでに検討されているが，この手法をここで述べたような実世界指向のインタラクションへの適用することについては十分な検討がなされていない．

本研究では聴覚定位によるインタラクションの実現とその基本的な特性の評価を行う．聴覚情報は視覚情報に比較して一覽性に欠けると考えられること，アイコンに持たせることのできる情報量が少ないことなど，検討すべき課題は多いが，最も基本的な疑問はこのような系におけるポインティング動作の特性に関するものである．人間のポインティング動作については，これまで視覚的なフィードバックのもとで目標に対してポインタを移動する際の特性が検討され，ポインティングに要する時間が目標の大きさおよび目標までの距離などによってある程度推定できることが知られている<sup>15)</sup>．これに対して聴覚的なフィードバックによる系については，これまで上述のような応用が提案されてこなかったこともあり，ポインティング特性の検討の試みは知られていない．

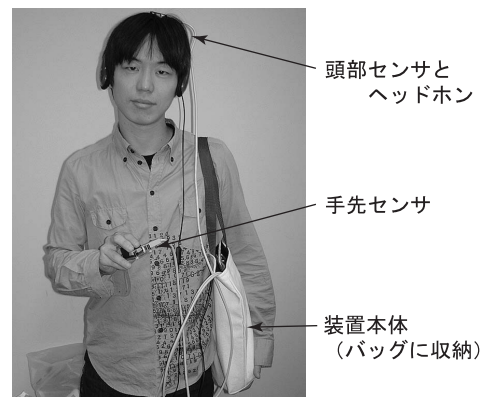
### 3. 聴覚提示システム

ポインティング特性の評価に関する実験では筆者らの試作したウェアラブルな聴覚提示システムを使用した．このシステムは，方向定位デバイス<sup>16)</sup>，頭部および手先の姿勢計測センサ，これらを統括するノート PC により構成されている（図 1 (a)）．

方向定位デバイスは，上述の頭部伝達関数を利用する方法によって実現されたもので，伝達関数の計算を FIR フィルタにより実現している．音源波形および FIR フィルタの係数は ROM に格納されており，精度 16 bit，サンプリング周波数 48 kHz で表現されている．フィルタのタップ数は 128 であり，これは残響を 2.5 msec まで考慮することに相当する．デバイスは同時に 4 つの音源のそれぞれに異なる伝達関数をかける処理を行うことができる．ノート PC との通信は RS232C によるシリアル回線を通して行われ，この伝送に要する時間はおよそ 8 ms，デバイスが音源波形お



(a) 聴覚提示システムの構成



(b) 実験における装着の状態

図 1 システムと装着状態

Fig. 1 Wearable auditory device.

よびフィルタ係数のデータを受信してからこれが音声出力に反映されるまでの時間はおよそ 0.1 ms である．

フィルタの係数は MIT Media Lab Machine Listening Group の Bill Gardner 氏および Keith Martin 氏により計測され公開されているもの<sup>17)</sup>を 48 kHz で再サンプリングしたものをを用いた．これにより，下方 45 度より下の方向を除く全天周 711 方向について伝達関数が定義される．定位方向はフィルタの切替えにより実現されるため不連続なものとなり，個々のフィルタの担当する方向の範囲はその範囲の中心（その範囲を代表する方向）から 6 deg 未満である．なお，この伝達関数は，頭部から音源までの距離が 1.4 m の場合についてダミーヘッドを用いて計測されたものであ

るが、後述の実験では方向定位のみを利用するものとし、距離定位については扱わない。

姿勢計測センサには 3D モーションセンサ (MDP-A3U7, トーキン) を用いた。これは、3 軸の姿勢を地磁気、重力加速度、ジャイロモーメントを利用して計測するもので、オイラー角での精度はロール、ピッチ、ヨーそれぞれ 15 deg (最悪値) で、また、データ更新速度は 125 Hz である。このセンサを頭部および手先にそれぞれ 1 つずつ用いることで空間に対する各々の絶対的な姿勢を計測する。頭部のセンサはヘッドホンに固定し、手先のセンサは把持して操作するものとした。また、個々のセンサには押しボタンがあり、その押し下げの情報もセンサデータとともに送出される。なお、ノート PC とのインタフェースは USB である。

ノート PC は PCG-505S (Pentium 166 MHz, ソニー) で、実験にかかわる処理を実行する。後述の実験では、頭部に対する音源の方向が頭部および手先の運動により変化することから、それぞれの音源の頭部に対する方向を計算し、この方向に定位させるためのフィルタを選択する処理が主要な部分を占めている。この処理はおよそ 40 Hz (実測) で実行されており、したがって、処理にともなう提示の遅れは最大で 25 ms であると推定される。

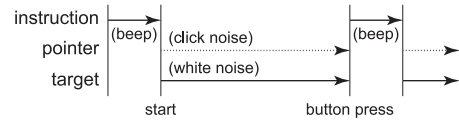
実験ではこれらの要素より構成される聴覚提示システムを装着した状態で使用する (図 1 (b))。このシステムで頭部および手先の運動から聴覚情報が更新されるまでの遅れはおよそ 41 ms 以内と推定され、また、方向提示の誤差はセンサの誤差と伝達関数を離散的に切り替えていることから生じる誤差とが合成されたものとなる。なお、装着重量は本体 3.5 kg、ヘッドホンおよび頭部センサ 170 g、手先センサ 100 g である。

## 4. 聴覚インタフェースの操作特性評価

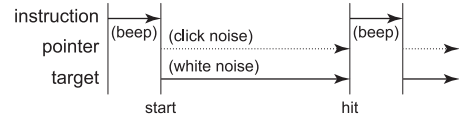
### 4.1 実験の目的と計画

試作した聴覚提示システムを利用して上述のような選択操作によるインタラクションを実現した場合の操作性に関する評価を試みた。操作性に関連して特に重要なのは、ポインティングの精度とこれに要する時間であり、また、その基礎となるのが音源方向の認識の精度である。

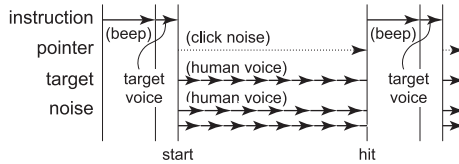
頭部伝達関数は頭部形状などに依存することから、本来はユーザの頭部形状に相当する伝達関数を利用することが望ましい。しかしながら、計測は必ずしも容易ではなく、一般には別の人の頭部形状に基づく伝達関数が使用される。したがって、定位には必然的に誤差が含まれる。



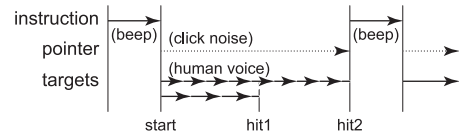
(a) 実験 1



(b) 実験 2



(c) 実験 3



(d) 実験 4

図 2 提示刺激の時間シーケンス

Fig. 2 Time sequence of auditory presentation.

以下の実験は年齢が 20 代および 30 代の男性の被験者 10 人 (a) ~ (j) と記述する) を用いて行った。ポインティングのターゲットは上述のアイコンを想定したもので、その方向は空間に対して定義される。したがって、頭部に対するターゲットの方向は頭部の運動によって変化する。一方、ポインタの方向は、頭部と手先との相対的な運動によってのみ変化する。

### 4.2 実験 1: 認識誤差の評価

ターゲットとポインタにはそれぞれホワイトノイズおよび 50 ms 間隔で繰り返されるインパルス波形を用いた。これらの音源はポインティングの開始と同時に提示され終了まで継続的に提示される (図 2 (a))。被験者はポインタを移動しターゲットに最も近づいたと感じたところでボタンを押して通知する。空間の座標系で仰角  $-30, 0, 30$  deg について方位角 30 deg 間隔、仰角 60 deg について方位角 60 deg 間隔の、合計 42 点からランダムにターゲットを選択し、実験を通して各点につき 5 回の計測を行うものとした。

実験の条件として、頭部の運動を許す場合と許さない場合を仮定した。ウェアラブルとしての使用では、上述のように頭部の運動を許すことが自然であるが、ここでは、比較のために許さない場合についても計測を行った。運動を許さない条件は、被験者の頭部を厳

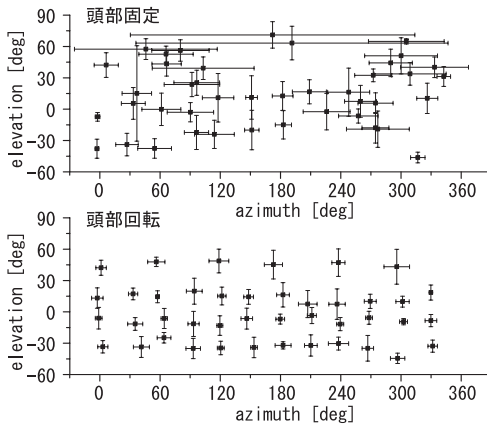


図3 方向定位認識誤差  
Fig. 3 Pointing error.

密に固定することは必ずしも容易でないため、頭部のセンサからのデータを無視することで同様の効果を実現している。

頭部の回転を許す場合と許さない場合の方向認識特性の違いを被験者(a)の場合について図3に示す。頭部固定の条件では頭部が動かないことを前提に頭部と空間の座標系が一致しているのに対して、頭部回転の条件では両者が相対的に回転する。図では各々のターゲット方向について空間の座標系で観測した仰角および方位角について平均値を中心とし標準偏差をエラーバーにより表現している。なお、方位角は前方が0 degで、上から見て時計回りに測るものとする。

頭部を固定した場合は、方向によって認識精度に違いがあるのに対して、頭部の回転を許した場合には方向によらずほぼ一定で比較的高い精度が得られている。これは、音源が頭部に対して認識精度の高い方向にくるよう頭部を動かすことや、運動によってフィルタの離散的な切替えによる精度限界よりも高い精度で認識が可能になることなどによるものと考えられる。また、頭部伝達関数による聴覚定位で比較的多く見られる前後の誤判断が頭部の運動を可能とすることで減少する効果も働いていると考えられる。なお、頭部を可動とすることで、理想的には頭部固定の場合の結果の中で最も精度の高い場合以上の精度が得られると期待されるが、現実にはもう少し大きな誤差を生じている。

10人の被験者について、ポインティングの誤差を整理した結果を表1に示す。頭部回転を許す条件については各々のターゲットに対する5回の計測から求められるRMS誤差を、また、許さない条件では、各々のターゲットについてポインティング方向に偏りが生じることから、そのオフセット誤差とその偏った点を

表1 方向定位認識誤差の評価  
Table 1 Evaluation of pointing error.

単位: deg

被験者	頭部回転		頭部固定
	RMS	Offset	RMS
(a)	16.3	18.3	21.5
(b)	24.6	42.3	35.9
(c)	47.2	53.5	39.3
(d)	36.9	49.3	47.6
(e)	46.9	45.9	47.0
(f)	38.0	34.0	28.4
(g)	31.9	70.3	39.2
(h)	31.8	56.8	30.5
(i)	42.0	42.8	33.2
(j)	18.8	24.4	22.3
avg.	33.4	43.8	34.5

中心とするRMS誤差を求め、表にはその平均値を示した。なお、オフセットの方向はポイントされた方向のベクトル平均として計算している。

この結果から、ポインティングの精度は個人差が非常に大きいことが明らかとなった。この理由としては、上述のように、頭部伝達関数に個人差があるため、実験に使用したものと適合の度合いが低い場合には方向定位に誤差を生じてしまうためであると考えられる。このような状況においても、頭部の回転を許す条件では、許さない条件に比較して方向定位の精度が向上することが確認された。一方、頭部回転の条件での結果について各仰角ごとの誤差の平均値の比較を試みた。仰角がそれぞれ-30, 0, 30, 60 degの条件ですべての被験者について得られたRMS誤差の平均は36.8, 32.9, 31.9, 30.9 degであった。検定によれば仰角-30 degの条件での誤差が他の条件に比較してわずかに大きく( $\alpha = 0.05$ )、仰角によるポインティング姿勢の違いが作業のしやすさの違い、ひいては精度に影響を与えたのではないかと考えられる。

#### 4.3 実験2: ポインティング時間の評価

ポインティングに要する時間とターゲットまでの角度距離および許容される誤差との関係を次の実験により評価する。被験者はターゲットをできるだけ早くポイントするように指示される。実験の条件として、ターゲットの出現範囲と、ターゲットの大きさを変化させた。出現範囲は全周(仰角-45~60 deg, 全方位)および前方(仰角-30~60 deg, 方位角-45~+45 deg)を仮定した。また、ターゲットの大きさはここではポインティングにおける許容角度を意味し、すなわち、ターゲットの方向とポインタの方向とのなす角がこの大きさ以下になった場合にポイントされたと思なすということである。この実験では6, 12, 24 degを仮定

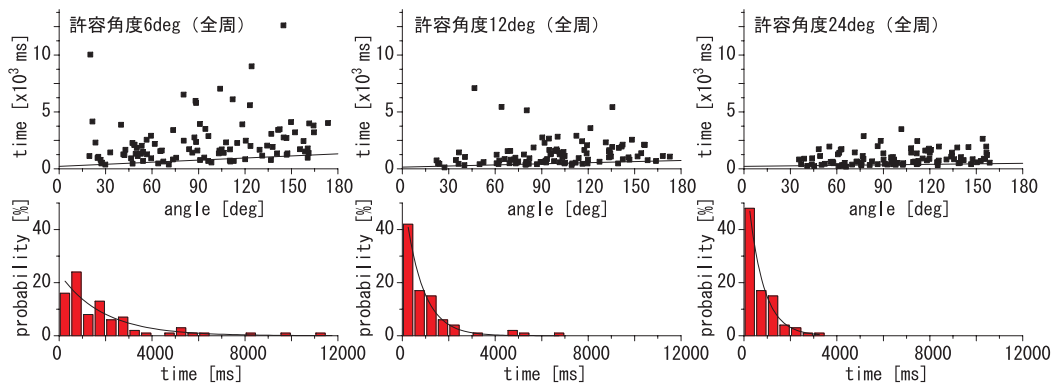


図4 ポインティング時間

Fig.4 Pointing time.

した．なお，ターゲットの方向は乱数により決定されるが，直前のポインタの方向からターゲットの大きさ +5 deg 以内の範囲には出現させないものとした．

ターゲットとポインタにはそれぞれホワイトノイズおよび 50 ms 間隔で繰り返されるインパルス波形を用いた．これらの音源は開始から終了までの間継続的に提示され，ターゲットがポイントされた場合にはピープ音により完了をユーザに知らせ，次のターゲットを提示する(図2(b))．ターゲットの提示からポインティング完了までの時間をポインティングに要した時間としてミリ秒単位で計測した．全周および前方の条件は実験者により事前に被験者に知らされている．

結果をポインティング角度距離と時間についてプロットしたところ，ポインティング時間に比較的大きなばらつきがあること，その中で比較的短時間でポイントできた場合についてはポインティング時間が距離の増加に対して増加する傾向があることが明らかになった．視覚に基づくポインティングでは，ターゲットの大きさおよび距離と時間との間に上述のようにフィッツの法則が知られているが<sup>15)</sup>，聴覚を手がかりとするポインティングに同様の法則が成立するかどうかは知られていない．また，上述のようにこの実験で得られた結果はばらつきが大きく，フィッツの法則を検証することは困難であると考えられる．

本実験では，むしろユーザによるポインティング時間の評価を目的としていることから，この結果を次のように整理する．まず，比較的短時間でポイントできた場合の角度距離と時間との関係を 1 次式(時間 =  $A \times$  角度距離 +  $B$ ) で記述し，次に，これに対する遅れのばらつきを指数減少の分布(頻度 =  $D \times \exp(-\text{時間}/T)$ ) として記述する．この整理により，各実験条件での結果を 1 次式の係数と指数関数の時定数として表現することができる．具体的には次の手順で処理を行った．

まず，データを角度距離について 10 deg 間隔でグループ分けし，各々の最小のデータをそのグループの代表データとする．ただし，グループに含まれるデータが 3 点未満の場合には代表データの信頼性が不十分であると考え，これを無視するものとした．この代表データに対して最小二乗法により直線をあてはめることで，上述の 1 次式を求める．すべてのデータからこの 1 次式より得られる遅れを差し引くことで遅れのばらつきの時間を求める．このばらつき時間を 500 ms 刻みにグループ分けすることで遅れ時間に対する確率分布を求め，これに指数関数を最小二乗法によりあてはめる．被験者(a)について提示範囲を全周とした場合の結果にこの処理を行った様子を図4に示す．この図から，直線および指数関数のあてはめがほぼ妥当なものであることが理解される．

同様の処理をすべての結果について行うことで得られた係数を表2に整理する．なお，表中の  $C_{60}$  は 1 次式に角度 60 deg を代入したもので，角度距離 60 deg のポインティングについて最短の所要時間を推定したものと考えることができる．この値は時間軸切片 B に比べて安定して求まることから，所要時間の指標として記載した．一方，指数関数の時定数  $T$  の値は，所要時間のばらつきに指数分布を仮定した場合に，結果のおよそ 63% がこの時間の範囲内に含まれることを意味している．

この表から，ターゲット(許容角度)が大きくなるに従い，ポインティング時間のばらつきが小さくなり，すなわちより確実なポインティング操作を行うことができるようになることが観察される．表の  $C_{60}$  および  $T$  の平均値について差の検定を行ったところ，全周と前方のそれぞれの条件でターゲットの大きさの違いにより有意な差があることが確認された( $\alpha = 0.02$ )．また，それぞれのターゲットの大きさの条件のもとで

表2 ポインティング時間の評価  
Table 2 Evaluation of pointing time.

提示範囲：全周					単位 A: ms/deg, B-T: ms							
許容角度	6 deg				12 deg				24 deg			
被験者	A	B	C <sub>60</sub>	T	A	B	C <sub>60</sub>	T	A	B	C <sub>60</sub>	T
(a)	6.11	223	590	1875	3.24	152	346	767	1.53	214	305	647
(b)	5.92	2371	2727	17525	5.23	161	475	1461	2.45	253	400	896
(c)	15.93	-109	847	6948	3.97	426	664	2749	7.44	-261	185	851
(d)	-1.99	1933	1814	3377	0.9	601	655	883	1.12	117	184	658
(e)	17.31	599	1638	10107	4.9	174	468	4112	2.84	114	285	1191
(f)	28.6	1185	2901	25673	31.33	-709	1170	5468	8.26	130	625	898
(g)	2.33	1764	1904	9867	5.41	226	551	1735	9.08	-236	309	603
(h)	17.42	886	1931	7030	16.2	-478	494	1567	6.33	75	454	1351
(i)	18.96	479	1616	9193	6.51	507	897	1667	4.25	152	407	742
(j)	3.76	426	652	1333	1.22	323	397	999	2.22	87	221	429
avg.	11.43	976	1662	9293	7.89	138	612	2141	4.55	64	338	827

提示範囲：前方					単位 A: ms/deg, B-T: ms							
許容角度	6 deg				12 deg				24 deg			
被験者	A	B	C <sub>60</sub>	T	A	B	C <sub>60</sub>	T	A	B	C <sub>60</sub>	T
(a)	5.67	56	396	554	6.66	-64	336	402	7.06	-182	242	86
(b)	1.77	367	473	1740	2.88	179	352	595	2.82	66	235	314
(c)	7.62	233	690	2456	6.46	-97	291	1085	5.14	-148	161	501
(d)	22.32	-456	883	922	4.83	174	464	533	4.92	-61	234	275
(e)	15.56	-339	595	3052	9.37	-242	321	1138	2.34	-13	128	487
(f)	2.98	1021	1200	7370	24.47	-545	923	1903	19.87	-688	504	613
(g)	10.59	176	812	2367	4.08	91	336	566	4.16	-19	230	382
(h)	-2.89	1177	1003	4037	11.22	-116	557	1444	5.65	-130	209	457
(i)	32.08	-434	1491	5082	1.31	422	501	1577	3.88	-18	215	425
(j)	2.8	242	410	768	4.28	2	259	463	6.14	-197	172	86
avg.	9.85	204	795	2835	7.56	-20	434	971	6.2	-139	233	362

は、提示範囲を前方に絞ることで全周に比較して C<sub>60</sub> および T の小さくなっていることが観察され、これは  $\alpha = 0.05$  で有意であった。この理由としては、ターゲットが前方に限定されることで体全体の動きをとまなうような頭部回転あるいはポインティング動作が少なくなり、これにまつわるばらつきが減少したものと考えられる。

#### 4.4 実験3：ノイズの影響の評価

ポインティングによるインタラクションの利用では、選択肢などの情報が音声として同時に複数の方向から提示されることを想定している。また、ウェアラブルとしての現実環境中での使用ではユーザはウェアラブルによる聴覚情報と環境音との両方を受け取ることになる。このように同時に提示される聴覚情報の中から目的とするものを選択的にポイントする際のインタフェースとしての特性を評価するための基礎的な実験を行った。また、この実験は、上述の実験2の結果と比較することで、ターゲットの音源波形にホワイトノイズではなくヒューマンボイス(人の声)を用いた場合の特性を調べることも意図している。

ポインティングに要する時間とターゲットまでの距離および許容される誤差との関係を次の実験により

評価した。被験者はターゲットをできるだけ早くポインタするように指示される。ターゲットの出現範囲は全周(仰角 -45 ~ 60 deg, 全方位)とし、ターゲットの大きさ(許容角度)として 6, 12, 24 deg を仮定した。ターゲットの方向は乱数により決定されるが、直前のポインタの方向からターゲットの大きさ +5 deg 以内の範囲には出現させないものとした。ターゲットとは別にノイズ2点をランダムな方向に提示する。音源波形はいずれもヒューマンボイスであり、アルファベットの 'a' ~ 'z' の発音からランダムに互いに異なるものを3つ選択して用いた。ターゲットがいずれの波形であるかをポインティング開始の直前に上方(空間の座標系で仰角 90 deg)に定位した音声により示し(図2(c)), 被験者は提示される3点からターゲットを選択してポイントする。ポインタには 50 ms 間隔で繰り返されるインパルス波形を用いた。ターゲットとノイズの音源は開始から終了までの間繰り返し継続的に提示され、ターゲットがポイントされた場合にはピープ音により完了をユーザに知らせ、次のターゲットとノイズの組を提示する。ターゲットの提示からポインティング完了までの時間をポインティングに要した時間としてミリ秒単位で計測した。

表3 ノイズの影響の評価

Table 3 Evaluation of the effect of noise.

提示条件：ノイズなし					単位 A: ms/deg, B-T: ms							
許容角度	6 deg				12 deg				24 deg			
被験者	A	B	C <sub>60</sub>	T	A	B	C <sub>60</sub>	T	A	B	C <sub>60</sub>	T
(a)	-0.16	1304	1295	3564	4.67	221	501	1270	3.22	200	393	638
(b)	13.92	1638	2473	7648	6.26	413	789	2803	3.66	484	704	1324
(c)	10.68	344	985	5398	6.36	360	742	2800	2.83	555	725	1302
(d)	23.34	674	2074	5517	17.41	-142	902	2134	4	262	502	1292
(e)	5.22	1424	1737	8603	11.35	236	917	3608	6.07	-46	318	1160
(f)	12.68	1370	2131	6949	1.16	1662	1732	4203	2.99	717	897	2227
(g)	0.24	2137	2151	3183	6.92	380	795	1796	4.29	503	761	1374
(h)	21.03	331	1593	2785	5.92	784	1139	2544	8.14	-126	363	810
(i)	-2.56	2193	2039	1046	4.12	458	705	1675	6.91	-209	206	1276
(j)	2.73	471	635	1787	3.93	228	464	846	2.15	259	388	409
avg.	8.71	1189	1711	4648	6.81	460	869	2368	4.43	260	526	1181
提示条件：ノイズあり					単位 A: ms/deg, B-T: ms							
許容角度	6 deg				12 deg				24 deg			
被験者	A	B	C <sub>60</sub>	T	A	B	C <sub>60</sub>	T	A	B	C <sub>60</sub>	T
(a)	15.22	189	1102	6283	7.31	918	1357	2139	5.64	376	714	1974
(b)	40.85	396	2848	14370	6.8	1689	2097	4030	2.11	1011	1138	2146
(c)	5.44	2029	2356	8762	6.74	745	1149	4270	4.99	440	739	2502
(d)	12.68	1171	1932	5709	24.25	-640	815	3693	13.51	220	1030	2964
(e)	1.28	1846	1922	7314	0.65	1238	1277	3715	1.71	736	839	1386
(f)	18.77	559	1685	6499	14.97	820	1718	7077	1.74	1325	1430	4095
(g)	20.71	1698	2940	8878	9.11	817	1363	3681	8.69	133	655	2391
(h)	25.66	968	2508	8268	10.32	915	1535	4702	11.15	207	876	3542
(i)	40.99	-621	1838	10012	13.24	187	982	3419	12.83	-651	119	2812
(j)	2.69	1233	1394	3594	4.74	512	796	1755	3.55	348	561	1166
avg.	18.43	947	2053	7969	9.81	720	1309	3848	6.59	414	810	2498

この結果について上述の実験2と同様の処理を行ったものを表3に示す。この表から、全般的な傾向としてノイズの存在によりC<sub>60</sub>およびTの平均値が増大することが観察される。検定の結果、Tの平均値については各ターゲットの大きさ(許容角度)の条件について有意な差が認められた( $\alpha = 0.02$ )。また、C<sub>60</sub>についてはターゲットの大きさが12 degの結果についてのみ有意な差があると判定された( $\alpha = 0.02$ )。一方、実験2の結果(表2)の結果と比較すると、ターゲットの波形にヒューマンボイスを用いた場合はホワイトノイズを用いた場合に比べて、平均時間、ばらつきともに増加する結果となっているが、C<sub>60</sub>およびTの平均値についてはいずれのターゲットの大きさの条件についても有意な差を認めることはできなかった( $\alpha = 0.02$ )。なお、この実験においてもノイズありとなしのそれぞれの条件でターゲットの大きさ違いによりC<sub>60</sub>およびT平均値に有意な差があることが確認された( $\alpha = 0.02$ )。

被験者からは、ノイズの存在によってポイントング開始時にターゲットの方向を認識するのに時間がかかること、'c'と'p'と't'、'f'と'nなどの発音が紛らわしい場合に混乱を生じることなどが報告されている

が、その影響を定量的に明らかにできなかった。また、この実験ではノイズとターゲットに同じ人の声を用いたが、声質を変えることにより相互の区別が容易になることが予想される。一方、現実の音環境ではノイズが2点というのは必ずしも十分に複雑な状況とはいえず、今後より現実的な条件での検討が必要であると考えている。

#### 4.5 実験4：音源の方向関係の把握

方向定位により音源方向の関係の把握がなされていることを確認するために次の実験を行った。ターゲットを2点提示し、被験者にこれらをできるだけ早くポイントするように指示する。このような場合、現在のポインタ位置から近いターゲットを先にポイントすることで移動角度距離を短くすることができ、また、音源方向の関係が把握されていれば近いものからポイントすることが自然な動作であると考えられる。

ターゲットの出現範囲は全周(仰角 -45 ~ 60 deg, 全方位)および前方(仰角 -30 ~ 60 deg, 方位角 -45 ~ +45 deg)を仮定し、ターゲットの大きさ(許容角度)は12 degとした。ターゲットの方向は乱数により決定されるが、直前のポインタの方向からターゲットの大きさ +5 deg 以内の範囲には出現させないものと



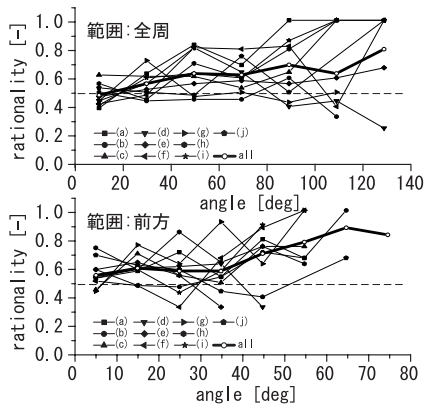


図5 音源方向とポイント順序の関係  
Fig. 5 Rationality in pointing order.

した．2つのターゲットの音源波形はいずれもヒューマンボイスであり，アルファベットの‘a’～‘z’の発音からランダムに互いに異なるものを2つ選択して用いた．ポイントには50ms間隔で繰り返されるインパルス波形を用いた．

それぞれのターゲットの音源は開始からポイントされるまでの間繰り返し継続的に提示され，両方のターゲットのポイントが完了した時点でピーブ音により完了をユーザに知らせ，次のターゲットを提示する(図2(d))．

この結果をポイント開始時のポイントの方向から2つのターゲットまでの角度距離の差と近いターゲットを先にポイントした割合との関係として整理した．具体的には，ターゲットまでの方向距離の差を全周の条件では20deg間隔，前方の条件では10deg間隔でグループ分けを行い，各々のグループで近いターゲットを先にポイントした割合を求めている．なお，グループに含まれるデータが3点未満のものについては確率の計算に信頼性がないと考えて結果から除いている．

得られた関係をプロットしたものを図5に示す．この図から，角度距離の差が大きくなるに従い上述の傾向は増加することが観察される．このことは，方向定位によって方向関係が直感的に理解され自然な経路選択が行われることを裏づけるものであると考えている．なお，提示範囲が全周の場合の確率にばらつきが大きくなっているのは，ポイント開始時にターゲットが頭部の後方など，認識の精度が低い領域に提示される可能性があり，この結果として上述の仮定に反するような動作を生じる割合が増えるためと推察される．

## 5. おわりに

本論文ではウェアラブルとユーザとの聴覚を利用したインタラクションの一手法として聴覚的なアイコンとポイントによる選択操作を提案した．また，このインタフェースの基本的な特性を被験者を用いた実験により定量的に評価した．この実験により得られたポイントング時間に関する知見は，聴覚的なアイコンなどの設計の基礎的なデータを提供するものと期待している．さらに，これらの実験を通して筆者らの提案する手法の実現可能性を明らかにすることができたと考えている．上にも述べたように，ポイントングの特性については個人差が非常に大きく，その原因の1つは頭部伝達関数の適合の程度の違いにあると考えられる．

評価に用いたシステムについても改良の余地があると考えている．装着重量3.8kgは日常生活での利用に不適当であることはもとより，評価実験においても被験者の疲労の原因となっている．姿勢センサの精度および時間遅れは必ずしも十分なものとはいえない．実際，空間に固定されているはずの音源が移動している感覚を生じる場合があることが被験者から報告されている．さらに，ここで用いた聴覚提示システムは4つの音源までしか定位させることができなかったが，より多くの聴覚情報が同時に提示される状況におけるユーザの振舞いも興味深く，したがってより多くの定位を行うことのできるデバイスが必要である．

本論文では聴覚提示のみに焦点をあてた実験を行ったが，現実環境においてはユーザは環境から視覚的な刺激も同時に受けることになる．視覚と聴覚についてはさまざまな感覚間相互作用があることが知られており<sup>18)</sup>，たとえば音像の方向が視覚刺激に引き寄せられるなどの現象が知られている．上述の実験は視覚的刺激の変化の少ない室内で行っているが，環境に対して視覚的注意を要求される現実的な環境において視覚が聴覚的なポイントングの精度などに与える影響の評価は興味深い課題である．

一方，本論文ではポイントングの枠組みの議論だけを行い，これをウェアラブルとして情報提示に適用するための意味論やコンテンツについては触れなかった．具体的に，街中であるいは作業現場でどのような情報をどのように提示し，選択操作によって何をシステムに伝えるのか，さまざまなケースについて検討していく必要があると考えている．

謝辞 本研究は科学技術振興事業団の戦略的基礎研究推進事業(CREST)の一部として行われた．

## 参 考 文 献

- 1) Mann, S.: Wearable Computing: A first step toward Personal Imaging, *IEEE Computer*, Vol.30, No.3, pp.25-29 (1997).
- 2) 廣瀬通孝: ウェアラブル・コンピュータの展開, 情報処理, Vol.40, No.9, pp.873-877 (1999).
- 3) Wenzel, E.M., Stone, P.K., Fisher, S.S. and Foster, S.H.: A System for Three-Dimensional Acoustic 'Visualization' in a Virtual Environment Workstation, *Proc. Visualization '90*, pp.329-337 (1990).
- 4) 廣瀬通孝: バーチャル・リアリティ, 産業図書 (1993).
- 5) Gaver, W.: The SonicFinder: An Interface That Uses Auditory Icons, *Human Computer Interaction*, Vol.4, pp.67-94 (1989).
- 6) Mynatt, E. and Edwards, W.K.: Mapping GUIs to Auditory Interfaces, *Proc. ACM UIST'92*, pp.61-70 (1992).
- 7) 在塚俊之, 畑岡信夫: 視覚障害者の GUI アクセスを助ける音響技術, 日本音響学会誌, Vol.54, No.5, pp.393-398 (1998).
- 8) 小木哲朗, 廣瀬通孝: 科学技術データ提示における多感覚の統合効果, 日本機械学会論文集 (C 編), Vol.61, No.584, pp.322-328 (1995).
- 9) Aoki, S., Cohen, M. and Koizumi, N.: Design and Control of Shared Conferencing Environments for Audio Telecommunication, *Presence*, Vol.3, No.1, pp.60-72 (1994).
- 10) 赤羽 歩, 波多野哲朗, 平田幸広, 原田哲也, 佐藤 誠: 仮想空間内での対象物選択作業における 3次元聴覚情報の役割, 日本 VR 学会論文集, Vol.5, No.3, pp.975-980 (2000).
- 11) 浅山 宏: バーチャルリアリティのための音響モデリング, 計算工学, Vol.4, No.2, pp.72-78 (1999).
- 12) Ueoka, R., Hirose, M., Kuma, K., Sone, M., Kohiyama, K., Kawamura, T. and Hiroto, K.: Wearable Computer Application for Open Air Exhibition in EXPO 2005, *IEEE Pacific Rim Conference on Multimedia*, pp.8-15 (2001).
- 13) 前田太郎, 安藤英由樹, 渡邊淳司, 野村宣邦, 三木健: 行動モデル化のためのウェアラブルロボティクス, 日本 VR 学会第 6 回大会予稿集, pp.153-154 (2001).
- 14) Ashbrook, D., Auxier, J., Gandy, M. and Starner, T.: Experiments In Interaction Between Wearable and Environmental Infrastructure Using the Gesture Pendant, *Proc. HCI2001*, pp.818-822 (2001).
- 15) MacKenzie, I.S.: Movement time prediction in human-computer interfaces, *Proc. Graphics Interface '92*, pp.140-150 (1992).
- 16) 広田光一, 廣瀬通孝: ウェアラブルのための聴覚提示デバイスの試作, ヒューマンインタフェース学会研究報告集, Vol.3, No.1, pp.9-11 (2001).
- 17) Gardner, W.G. and Martin, K.D.: HRTF measurements of a KEMAR dummy head microphone, MIT Media Lab Perceptual Computing Technical Report # 280 (1994).
- 18) 小宮山摂: 視覚と聴覚による音像知覚, 日本音響学会誌, Vol.52, No.1, pp.46-56 (1996).

(平成 13 年 10 月 1 日受付)

(平成 14 年 10 月 7 日採録)



広田 光一 (正会員)

1988 年東京大学工学部産業機械工学科卒業, 1990 年同大学院産業機械工学専攻修士課程修了, 1994 年同博士課程修了, 1995 年豊橋技術科学大学情報工学系助手, 2000 年先端科学技術研究センター助教授. 主にヒューマンインタフェースの研究に従事. 日本 VR 学会, 日本機械学会, 計測自動制御学会各会員.



廣瀬 通孝 (正会員)

1977 年東京大学工学部産業機械工学科卒業, 1997 年同大学院修士課程修了, 1982 年同博士課程修了, 工学博士, 同年東京大学工学部産業機械工学科専任講師, 1983 年同助教授, 1999 年東京大学大学院工学系研究科機械情報工学専攻教授, 同年東京大学先端科学技術研究センター教授, 現在に至る, 主にシステム工学, ヒューマンインタフェース, バーチャルリアリティの研究に従事. 著書に『技術はどこまで人間に近づくか』(PHP 研究所), 『バーチャル・リアリティ』(産業図書), 『岩波講座現代工学の基礎システムの構造と特性 < 設計系 IV >』(岩波書店) 等. 日本機械学会, 計測自動制御学会, 日本バーチャル・リアリティ学会各会員.