

3次元聴覚提示によるメニュー選択インタフェース

2007年3月修了 人間環境学専攻 56813 渡邊 洋介
指導教員 広田 光一 助教授

Abstract: 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. In this research, we propose an auditory interface for menu selection using three-dimensional sound. The concept of the method is based on the notion to introduce the selection mechanism of visual GUI into auditory interaction. We report the experiments on the efficiency and the accuracy of auditory pointing.

Key Words: Aural menu interface, Wearable computing, Sound localization

1. 緒言

近年の情報技術の進展により、計算機の小型化・高性能化が急速に進み、その使い方が多様化している。計算機を身につけて利用するウェアラブルコンピューティングもそのひとつである。これまで視覚情報に焦点を当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言えない。聴覚情報はヘッドホンのような簡単な装備により提示が可能であることから、ウェアラブルコンピュータのためのインタフェースとして、その活用が期待されている。また、聴覚情報は注目していなくても受け取ることができる受動的で通知性の高い情報ため、視覚情報に比べて不便さや疲労、危険を伴いにくいという利点もある。一方で、聴覚提示をウェアラブルコンピュータ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は、これまで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ティブな音声提示型のインタフェースを提案し、被験者による評価実験を通してその操作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ひとつの指針を与え得ると考えられる。

2. 関連研究

聴覚インタフェースの研究については、GUIメニューをレベル差ステレオを用いて3つまで同時に提示する方法^[1]や、肩に装着したスピーカーで聴覚情報を提示する方法^[2]などの試みがある。しかしながら、これらは音源の分離が困難であることや、人間の動作と連動した直感的な操作が困難であるという問題点がある。また、実世界情報提示に関する研究では、ウェアラブルコンピュータの提案^[3]や博物館ガイドのためのユビキタス・デバイスの提案^[4]などがあるが、これらは視覚的なメニュー項目の提示が主なテーマとなっており、音声のみで操作できるインタフェースではない。したがって、実世界情報提示を主としたインタフェースは、音声においてはまだ一般化されていないと言える^[5]。また、聴覚に関する基礎研究で本研

究と関連が深いものでは、雑音環境で聴覚定位が有効であったという研究^[6]や、空間定位を用いて音声単語を提示する際に同時話者4人で40~50%の正答率であったという研究^[7]がある。しかしながら、聴覚をインタラクティブな手法で使用した場合の操作特性に関してはまだ明確な知見が得られていない。

本研究では聴覚定位を利用したインタラクティブな聴覚インタフェースを提案し、その時間と精度に関する操作特性を評価することを目的とする。

3. メニュー選択インタフェース

3.1. 3次元聴覚提示装置の構成

聴覚情報を用いたインタフェースの特性評価に関する実験では、ウェアラブルな聴覚提示システムを使用した^[8]。このシステムは、聴覚定位装置、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(Fig.1(左))。聴覚定位装置は頭部伝達関数^[9]を利用して音を方向定位させる装置である。頭部伝達関数とは、頭部での音の反射・回折を考慮した関数である。姿勢計測センサはMDP-A3U7(NEC TOKIN)を用いた。ノートPCはCF-W4(Pentium 1.30GHz, 松下電器産業)で、実験に関わる処理を実行する。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。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(Fig.1(右))。この装置を用いて聴覚インタフェースを提示する。



Fig.1 Wearable auditory device

3.2. インタフェースの構築

本研究ではインタラクティブな聴覚インタフェースを

提案する。ユーザの周りに 3 次元空間を構成し、それを経度と緯度で区切り、各区分に音声を配置する。ユーザは手元のセンサを動かすことで、聴覚情報を得ることができる(Fig.2).



Fig.2 The concept of auditory interface

本研究で提案したインタフェースにおいて、メニューとは、上述の 3 次元聴覚提示装置により生成し、ユーザを中心とする空間内に提示される音声の集合である。音声メニューは、同時に 14 チャンネルまで提示することができ、配置位置も自由に設定することができる。ユーザは手元のセンサを動かしてヘッドホンから音源を聞きながらメニューを選択する。選択とは、ユーザが任意の音声メニューの位置で手先センサのボタンを押すことをいう。また、ユーザが手先センサを動かして音声メニューを閲覧している状態をポインティングという。これは、GUI 環境でマウスポインタを動かす動作に相当する。ポインティングの際、ユーザは連続した電子音(クリックノイズ)によって手元センサが指している位置のフィードバックを受ける。また、メニュー選択を補助する機能として次の 3 つを構築した。

機能 1：聴覚定位による空間認識の補助

定位とは、メニューの配置が空間に固定されていること。定位がない場合、メニューの配置がユーザの頭部の方向と共に移動することになる(Fig.3).

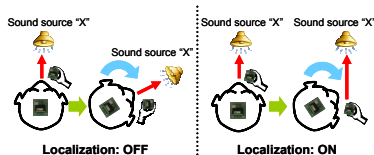


Fig.3 Localization

機能 2：余韻による音声認識の補助

余韻とは、手元センサを素早く動かしても、選択されたメニューを最後まで読み上げる機能をいう。余韻がオフの場合は、常にセンサが指しているメニューを読むことになる(Fig.4).

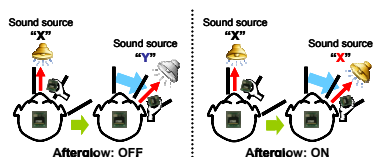


Fig.4 Afterglow

機能 3：同時提示による実環境の再現

上述した聴覚提示は、ユーザが音声メニューに対してランダムにアクセスできる探索的な提示・選択方法である。同時提示による聴覚提示は、この探索的な提示・選

択方法にさらに機能を追加し、ユーザに対して音を多方向から同時に提示することで、あたかも商店街を歩いているかのような音環境を計算機で仮想的に表現した機能である。ユーザは手元のセンサを動かさなくても、多方向から提示される音の方向によって、メニューの位置を知ることができる。また、ユーザがポインティングしている位置のメニューは、音声ピッチが通常の 2 倍となる。このようにピッチを変化させるハイライト効果で、ユーザはポイント中のメニューのフィードバックを受ける。

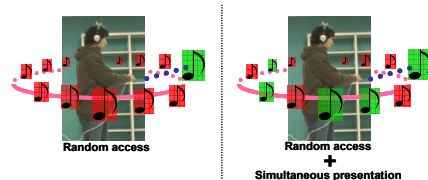


Fig.5 Random access and simultaneous presentation

4. 評価実験

4.1. 実験の概要

提案したインタフェースの精度と時間に関する特性を明らかにするために、被験者による 3 つの評価実験を行った。実験では、選択すべき音源(ターゲット)を 3 次元聴覚提示装置から被験者に対してランダムに指示し、被験者がメニューの中からターゲットを選択するまでの時間と間違えた回数を記録した。

4.2. 実験 1：音声アルファベット選択実験

4.2.1. 実験方法

提示音声：A から Z までのアルファベット 26 文字。

- ターゲットは、ランダムに抽出したアルファベット 1 文字を頭部正面方向から音声で提示(Fig.6).
- メニューは、以下の条件に従って構成。

条件の組合せ：36 通り。順序効果を排除するために、

条件をランダムに変更し、各被験者に提示。

- 選択肢の数：4, 8, 12 [個]
- 方位角：10, 20, 30 [deg]
- 定位：ON, OFF
- 余韻：ON, OFF

計測回数：それぞれの条件で、10 回ずつ計測(被験者 1

人につき計 360 回計測)し、合計で 1080 回計測。

被験者：正常な聴覚を有する成人 3 名。

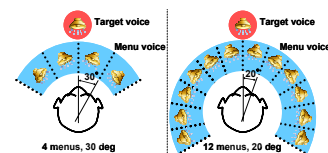


Fig.6 Target and menu voice

4.2.2. 計測結果および考察

全体的な傾向として、選択肢の数が増加すると、ターゲットを選択するまでに要する時間が増大し、選択肢の方位角が大きくなるほど、選択時間が減少することが観察される(Fig.7)。検定の結果、選択肢の数、および方位角の大き

さについて有意な差が認められた($\alpha=0.05$). 一方で, 定位と選択肢発音の余韻に関しては, それらの有無により, 選択時間の平均値に有意な差を認めることはできなかった($\alpha=0.05$)(Fig.8). また, 正答率は, メニューの数を 4, 8, 12 と増加させてもそれぞれ同程度に減少しており, 60%程度であった(Fig.9). 同様に, 選択肢の方位角の大きさ, 選択肢とポインタの定位の有無, 選択肢発音の余韻の有無の各条件について, 正答率との関係を調べた. 方位角を大きくした場合や, 定位や余韻の有無についても, Fig.9 のように, それぞれ同程度の正答率であった.

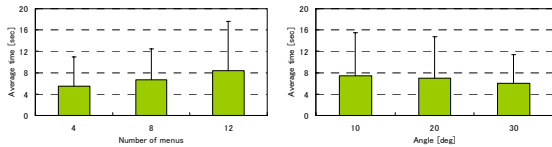


Fig.7 Number of menus and angle

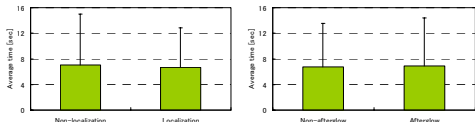


Fig.8 Localization and afterglow

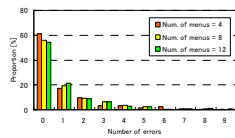


Fig.9 Number of errors

4.3. 実験 2 : 音声キーボード実験

4.3.1. 実験方法

提示音声 : A から Z までのアルファベット 26 文字.

- ターゲットは, ランダムに抽出したアルファベット 1 文字を頭部正面方向から音声で提示(Fig.10).
- メニューは, 標準 JIS キーボード配列に従って配置.

条件の組合せ : 2 通り. 順序効果を排除するために, 条件をランダムに変更し, 各被験者に提示.

- 選択肢の数 : 26 文字(標準 JIS キーボード配列)
- 方位角 : height 5, width 10[deg] / 1 word
- 定位 : ON, OFF
- 余韻 : OFF

計測回数 : それぞれの条件で, 10 回ずつ計測(被験者 1 人につき計 80 回計測)し, 合計で 720 回計測.

被験者 : 正常な聴覚を有する成人 9 名.

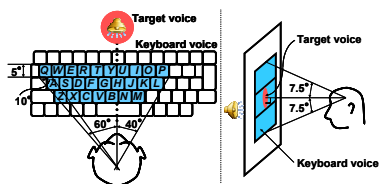


Fig.10 Target and keyboard voice

4.3.2. 計測結果および考察

各被験者が選択に要した時間の平均は約 6 秒だった

(Fig.11(左)). 1 分間で約 10 文字の入力が可能である. また, 定位の有無に関わらず, 個人によって平均時間にはばらつきがあることが明らかとなった(Fig.11(右)). これは, キーボードの配列を覚えている被験者とそうでない被験者との差から生じたものと推測される. 検定の結果, 定位の有無によって, 選択に要した時間に有意な差を認めることはできなかった($\alpha=0.05$). 正答率は 81%であった(Fig.12).

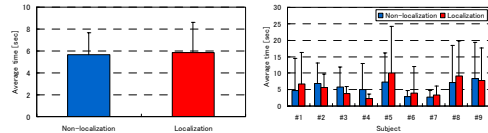


Fig.11 Average time of keyboard experiment

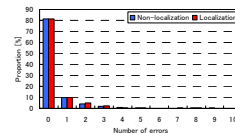


Fig.12 Number of errors

4.4. 実験 3 : 音声メニュー選択実験

4.4.1. 実験方法

提示音声 : 音声合成ソフトで作成した日本語音声単語.

- ターゲットは, ランダムに抽出した音声単語を頭部正面方向から音声で提示(Fig.13).
- メニューは, 以下の条件に従って被験者の周囲に構成.

条件の組合せ : 4 通り. 順序効果を排除するために, 条件をランダムに変更し, 各被験者に提示.

- 選択肢の数 : 12 [個]
- 方位角 : 30 [deg]
- 定位 : ON
- 余韻 : OFF
- 提示単語 : 野菜の名称(Ve), 山手線の駅名(Ya)
- 提示方法 : ランダムアクセス(RA), ランダムアクセス+同時提示(RA+SP)の 2 通り. 同時提示では, 平均 6.0[sec], 標準偏差 2.0[sec]のガウス乱数に従って周囲からランダムにメニューを提示.

計測回数 : それぞれの条件で, 20 回ずつ計測(被験者 1 人につき計 80 回計測)し, 合計で 400 回計測した.

被験者 : 正常な聴覚を有する成人 5 名.



Fig.13 Menu

4.4.2. 計測結果および考察

各被験者が選択に要した時間の平均は, ランダムアクセス+同時提示の方法で野菜の名称を提示した場合(以下 "RA+SP,Ve")は 7.3[sec], ランダムアクセス+同時提示の方

法で山手線の駅名を提示した場合(以下"RA+SP,Ya")は5.0[sec], ランダムアクセスで野菜の名称を提示した場合(以下"RA,Ve")は5.2[sec], ランダムアクセスで山手線の駅名を提示した場合(以下"RA,Ya")は3.1[sec]となった(Fig.14(左)). "RA+SP,Ya"と"RA,Ve"の組合せ以外において, それぞれ有意な差があることが確認された($\alpha=0.05$). メニューの位置が分かっている場合は, そうでない場合と比べて, メニューの選択が短時間で可能だと言える. また, 被験者によって, 選択時間の平均値にばらつきがある. 例えば, 被験者#3は"RA+SP,Ya"において, 他の場合に比べて最も早い平均値をとっている(Fig.14(右)). 選択精度は, ランダムアクセス型の提示方法では82%, ランダムアクセス+同時提示の場合は70~75%であった.

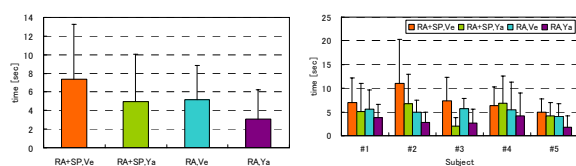


Fig. 14 Average time of menu selection experiment

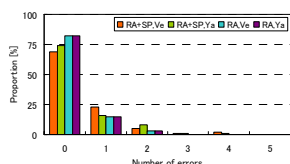


Fig. 15 Number of errors

4.5. 評価実験の総括

実験1では, アルファベットの選択に平均約5秒かかった. 時間効率を向上させるためには, 選択肢の数を少なくし, 方位角を大きくすることが確認された. 一方で, 定位と余韻の有効性は確認できなかった. また, 正答率は約60%であった. 実験後の被験者へのインタビューでは, アルファベット1文字だと聞き取りやすい場合とそうでない場合があるという回答があった.

実験2では, 選択時間は平均約6秒であった. また, 個人によって選択時間にばらつきがあることが確認された. 一方で, 定位の有効性は確認できなかった. 正答率は81%であった. 被験者へのインタビューでは, キーボードの配置を覚えていると短時間で選択可能であるという回答があった.

実験3では, ランダムアクセス型の提示方法の場合, 選択時間の平均は約3~5秒であった. 一方, ランダムアクセス+同時提示の場合は, 平均約5~7秒であった. また, 正答率は, ランダムアクセスは82%, ランダムアクセス+同時提示では70~75%であった. アンケートによる評価では, メニューの位置を覚えていると選択が容易であるという回答や, 同時提示が有効な場合もあるという回答があった. しかし, 同時提示の有効性を統計的に示すような結果は得られていない.

5. 結論

3次元聴覚提示による聴覚インタフェースを提案した. また, 被験者を用いた評価実験を実施し, このインタフェースの特性について評価した. その際, 提示する音声メニューの位置関係や意味を考慮した実験を行い, インタラクションに要する時間や精度に関する特性を明らかにした. これらの実験により得られた知見は, 基礎研究として, 聴覚インタフェースの設計にひとつの指針を与え得るものと期待している.

実験では, 定位の有無について, 選択時間に有意な差を確認することはできなかった. しかし, 動的な変化の少ない実験室内での実験においては, 頭部の動きが少なく, 定位の有無による影響が微小であった可能性もある. 実験室内とは異なり, 現実的な環境においては, 日常の動作で頻繁に頭部が動くため, 定位の有無が選択の時間や精度に影響を与える可能性もあると考えられる.

聴覚情報を利用したウェアラブルコンピュータとユーザとのインタラクションの特性を明らかにしたことで, 例えばeメールクライアントやスケジューラなどのアプリケーションを対象として, 具体的なシステムへの適用と評価を検討することができると考えている. 一方で, 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. メニューについては階層的に選択するための入力手法を検討する必要がある. キーボードについては, 身体の正面の位置のキーは範囲を狭くし, 外側に向かうに従ってキーの範囲を広くしていくなどの改良が考えられる.

参考文献

- [1] 河本敏志, 竹村治雄, 片山善章, 萩原兼一, 横矢直和: 音声メニュー同時提示方法の提案と評価, 情報処理学会研究報告, HI59-7 (1995).
- [2] Sawhney,N., Schmandt,C.: Nomadic radio:speech and audio interaction for contextual messaging in nomadic environments, ACM Transactionson Computer-Human Interaction, Vol. 7, pp.353-383 (2000).
- [3] Mann,S.: Wearable Computing: A first step toward Personal Imaging, IEEE Computer, Vol.30, No.3, pp.25-29 (1997).
- [4] 檜山敦, 山下淳, 西岡貞一, 葛岡英明, 広田光一, 広瀬通孝: 「ユビキタスゲーミング」位置駆動型モバイルシステムを利用したミュージアムガイドコンテンツ, 日本バーチャルリアリティ学会論文誌, Vol.10, No.4, (2005).
- [5] 池井寧, 山崎仁志, 広田光一, 廣瀬通孝: ウェアラブル指向音声メニュー多重化提示手法 vCocktail に関する研究, ヒューマンインタフェース学会論文誌, Vol.7, No.4, (2005).
- [6] Ericson,M.A., McKinley,R.L.: The intelligibility of multiple talkers spatially separated in noise, Binaural and spatial hearing in real and virtual environments, (Ed. Gilkey, Anderson), Erlbaum, NJ,pp. 701-724 (1997).
- [7] Drullman,R., Bronkhorst,A.W.: Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation, J. Acoustical Society of America, Vol. 107, No. 4, pp. 2224-2235 (2000).
- [8] 広田光一, 廣瀬通孝: 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, 情報処理学会論文誌, Vol.44 No.1, pp.1234-1243 (2003).
- [9] Gardner,W.G., Martin,K.D.: HRTF measurements of a KEMAR dummy head microphone, MIT Media Lab Perceptual Computing Technical Report #280 (1994).