

## 頭部方向に応じた音声メモの配置とブラウジング

米澤 朋子<sup>†</sup>, 山添 大丈<sup>†</sup>

<sup>†</sup> ATR 知能ロボティクス研究所

本研究では、ホワイトボードや通常のメモなどペン型記述メモの従来手法に頼らず、音声情報を用いて備忘録を実現することを目的とする。その際、ユーザがホワイトボードの領域の自由な場所にメモを書き込むのと同様な思考空間を考慮し、ユーザの発声する音声メモなどを記録しユーザの相対的頭部方向に応じて配置したり、録音された音声メモのブラウジングを行う際に記録・配置された各音声メモの位置と現在のユーザの頭部方向に応じて音声メモの存在を表す音が聞こえるよう提示するブラウジングシステムを提案する。

## Space assignment and browsing of voice memo corresponding to user's head direction

Tomoko Yonezawa<sup>†</sup> Hirotake Yamazoe<sup>†</sup>

<sup>†</sup> ATR Intelligent Robotics and Communication Lab.

In this paper, we propose an intuitive interface of both the recorder and browser of the personal voice memo using the user's head directions directions. This system aims to realize the useful memo space with using the user's voice and head instead of pen-type descriptions such as the white boards or paper memorandums. The system starts at deciding the criterial direction as absolute degrees of a 3D magnetic sensor, and memorizes each utterance of the user with the relative direction to the criterial direction. While the user browses the voice memo space, each voice memo is converted into "voice-icons" to express the existences of the memos, that are layouted in the sound space. When the user looks at the direction of the certain voice memo, the memo makes its original voice.

### 1 はじめに

人は何らかの複雑な問題を思考するときや複数の絡み合った問題を整理するとき、ホワイトボードやメモ用紙など、実在する世界の表面に何らかの描画や短い記述などのメモを書き出す。それは時に整列した箇条書きであったり、無作為な位置に様々な記述を繰り返された結果落書きの集合のように見えることもある。またその一方で、その思考を再度整理するために、記述や図画を位置関係を考慮しながら描きなおすような作業を通じて、思考過程や自分が後に理解し易いように配置をするケースもある。最近では、視覚的メモのPDAのペン入力装置、電子ホワイトボードなど、ポータビリティや電子化にも展開されつつあり、思考過程の保存を要するユーザの需要を満たしつつある。

このような視覚的なメモ手法は、書き込める領域の大きさや状況に応じて用途の目的が異なるが、共通して、入力状況が限定的である。例えば、駅の構内でふと思いついたことを書きとめようとす

ると、「何に」書くべきかを考え、その道具を自分の荷物から探すことから始める。持ち帰れて、かつ、白紙であるもしくは白紙領域がある、それ自体に重要性のない紙を探す。また、ペンなどのその場で使いやすい筆記用具を探す。運良く双方ともに見つかったとしても、机の上で何かものを書くのとは違い、手を下敷きにして不安定な文字を書く状況も生まれる。また、例えばベッドで横になっており、頭は働くが体がだるくて動かないときなど、わざわざメモを取る行為までたどり着けず、次の日には何を思いついたか、何をメモした買ったかすら忘れていたような状況もしばしば起こる。

本研究では従来のメモ手法に音声メモを取り入れ、上記のような様々な状況下で容易にメモを行うための手法を検討する。そこでは、ユーザの正面を基準とした頭部方向の相対座標に基づき仮想音響空間を構築し、音声メモを貼り付けていくという概念を実現する(図1参照)。さらに、的確に

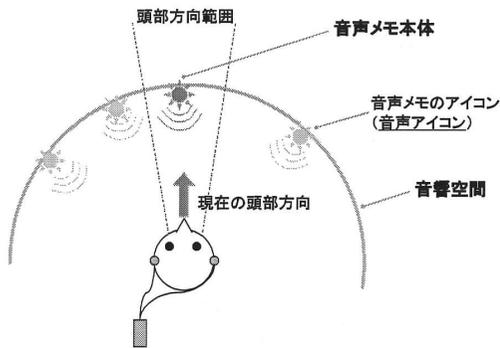


図1 システム概念図

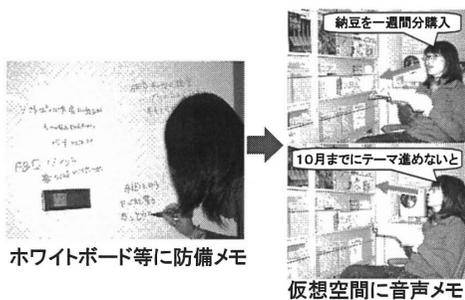


図2 利用イメージ

それらの音声メモを直感的にブラウジングしたり整理したりできるシステムの構築を狙う。

既存のICレコーダや携帯電話の録音機能などを用いた音声メモ手法は、現在一部のユーザの間では重宝されており、音声によるメモの需要は確認されている。防備録を入力するだけでなく、作曲家が思い浮かんだフレーズを吹き込むなど多岐にわたる用途がある。それだけ、人間の外部記憶装置としての役割を音声メモが果たし得ることが伺える。しかし一方でこれらの音声メモは、複数個の音声メモを含むときに探索が容易ではない。ICレコーダであれば時間系列に基づいたファイルの羅列であり、すべてを同時に効率的に探索することは音声の性質上不可能である。

これに対し、我々の提案する音声メモの配置・ブラウジングシステムでは、ユーザを軸とした相対的な仮想音響空間と音声メモの位置関係を関連付け、仮想的に音声メモを配置することをユーザ発声時の頭部方向情報を用いた録音により実現する。

さらに、複数音声メモの同時ブラウジングにおいては、音声アイコンという概念を導入することで、音声の重複による混乱を防ぐ。そして、ユーザの頭部方向に応じて音声メモの存在する方向の音源定位で各音声アイコンが聞こえ、ユーザの頭部方向の $\pm 5^\circ$ の範囲に入った音声メモは現音声で再生するという設計を提案する。図2にを示す。仮想音響空間に対応付けた音声によるメモを実現することにより、様々な状況下で、思考空間を有効活用することができるようになったり、後天的に視力を失ったユーザにも目もスペースを提供ができるようになると期待できる。

## 2 関連研究

提案手法の関連研究としては、メモなどのコンテンツを仮想空間・実空間・メディアコンテンツ中の任意の位置に対してアノテーション（注釈情報）として貼付し、それら位置情報と関連付いたアノテーションを情報共有・検索、情報提供などに利用するシステムが多く提案されている [1-6]。

例えば、荒屋らは、仮想美術館の任意の位置に音声アノテーションを付加し、展示案内に利用するシステムを提案している [2]。

垂水らは、実空間にコンテンツを配置しておき、その特定エリア内にいるユーザに対してのみコンテンツの配信を行うなど、エリア・時間などを限定した配信を行うシステムを提案している [3]。

暦本らは実環境に対し音声メモと画像を添付し、AR(augmented reality)によりユーザに提示することで、情報共有を行うことを目指したシステムを提案している [4]。また、ARによりアノテーションの提示を目指したシステムも多く提案されている [5,6]。

これらのシステムでは、主に情報共有やコミュニケーションを目的として、実世界やビデオなどのメディアコンテンツに関連付けた情報として空間やコンテンツ中の特定の位置に対して音声メモを貼付する。それに対し、提案システムでは、ユーザにとっての相対的な仮想空間を用いて音声メモの配置を可能とすることでユーザ個人の思考や記憶をサポートすることを目的としている。ユーザ自身の正面を基準とした相対方向に着目し、相対方向に基づく仮想音響空間に対し音声メモを貼り付けていくことで、常に再現できるユーザにとっての相対的な音声メモ付き仮想空間を実現するこ

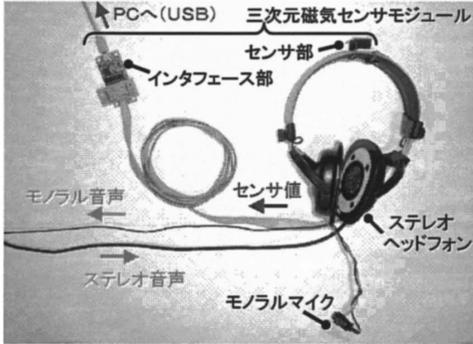


図3 頭部方向推定装置付きヘッドセット

とを目指す。

ユーザの頭部方向に応じた音声再現に関しては、矢入らによる頭部運動感応型聴覚ディスプレイ [7] があるが、遅延に対する検証結果から、少しでも遅延を少なくすることが音源の正確な再現における重要課題とされている。本稿における音声メモの配置では、センサの値の遅延を考慮し、複雑な計算による精緻な音源定位を目指すのではなく、音源の時間差と強度差における単純な処理により頭部周囲の円上に音源定位を再現する。

### 3 装置構成

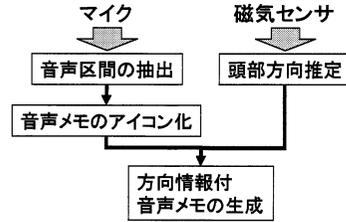
#### 3.1 システム概要

図3に提案システムの構成を示す。頭部方向とは、ユーザのある時点における頭部方向を正面とした、相対方向の角度を示す。提案システムは、ステレオヘッドフォンとモノラルマイク、3D磁気センサモジュールからなっており、マイク・磁気センサはともにヘッドフォンに固定されている。

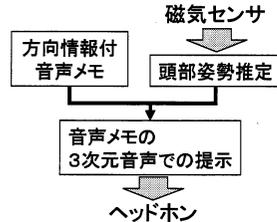
次に、図4に処理の流れを示す。処理は音声メモ作成時(図4-A)とブラウジング時(図4-B)に大きく分けられる。

音声メモ作成時には、まずマイクで得られた音声データから、音声区間を検出し、音声メモ・音声アイコンを生成する。同時に、磁気センサにより頭部方向を推定し、音声区間における頭部方向をもとに、頭部方向情報付きの音声メモを作成する。

ブラウジング時には、磁気センサによる現在の頭部方向と音声メモに付加された方向情報をもとに、各音声メモの3次元音声を生成することで、頭部方向に応じた音声メモの提示を実現している。



A. 音声メモ記録部分



B. 音声ブラウジング部分

図4 システムの処理フロー

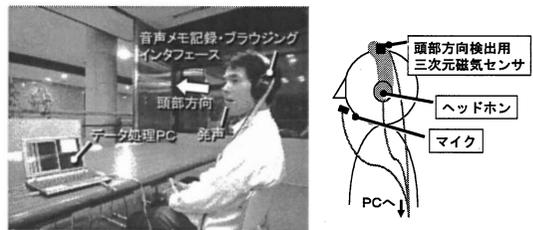


図5 利用風景

図5にシステムの利用風景の例を示す。

次節以降では、各処理の詳細について述べる。

#### 3.2 頭部方向推定装置

まず、頭部方向推定について述べる。頭部方向の推定には、パイテック製3DセンサモジュールTDS01Vを用いている。TDS01Vはセンサ部とセンサデータをUSBに変換するインタフェース部からなっているが、図3に示すように両モジュール間をケーブルで接続し、センサ部のみをヘッドフォンの頭頂部に取り付けている。

TDS01Vは、3軸地磁気と3軸加速度からなっており、両方のデータを用いることで、3軸の頭部方向が推定可能であるが、本稿では、左右の頭部方向のみについてを検討するため、以下では3

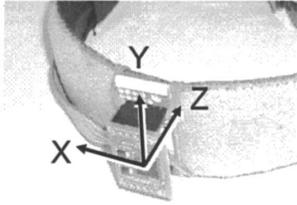


図 6 頭頂部に設置した地磁気センサの各軸方向

軸地磁気のみを利用する。

基準方向を向いた時の地磁気センサの出力を  $[I_x, I_y, I_z]'$  とし、計測時の地磁気センサの出力を  $[M_x, M_y, M_z]'$  とする。図 6 に各軸方向を示す。

この時、相対頭部方向  $\theta$  [deg] は、次式で計算できる。

$$\theta_t = \tan^{-1} \left( \frac{M_y}{M_x} \right), \quad \theta_i = \tan^{-1} \left( \frac{I_y}{I_x} \right) \quad (1)$$

$$\theta = \begin{cases} \theta_t - \theta_i & (-180 \leq \theta_t - \theta_i \leq 180) \\ \theta_t - \theta_i + 360 & (\theta_t - \theta_i \leq -180) \\ \theta_t - \theta_i - 360 & (\theta_t - \theta_i \geq 180). \end{cases} \quad (2)$$

これにより、100[msec] 毎に頭部方向のデータを取得し（図 7 参照）、音声メモの記録装置とブラウジング装置におけるリアルタイムインタラクションに用いる。

次に、頭部方向の推定精度を測るため、以下の実験を行った。壁面上の  $-30 \sim 30$  [deg] の  $10$  [deg] 刻みの位置にマーカを貼り、順に顔を向けた時の推定結果を図 7 に示す。データ取得試行の結果より、被験者 A（図 7-A）では比較的正確な角度が得られた一方で、被験者 B（図 7-B）では全体的に推定角度が大きくなる、という被験者により特徴が異なる結果が得られた。

### 3.3 頭部方向に応じた音声メモ

#### 3.3.1 方向情報付音声メモ記録装置

音声メモを記録する際には、ユーザの発声区間を検出する必要がある。常に  $5$  [sec] の音声をバッファに書き換え続け、音声発声区間（発話開始・発話終了）を検出したとき、音声メモとして保存する。音声発声区間を検出する手法としては、閾値を設けて ON/OFF の時間長で次のように判定す

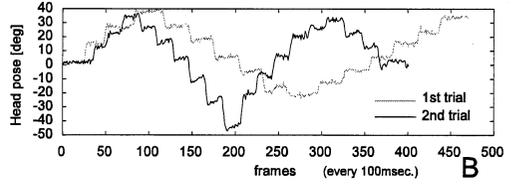
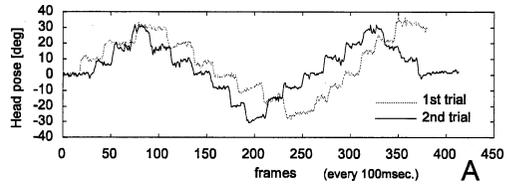


図 7 地磁気センサによる頭部方向推定例

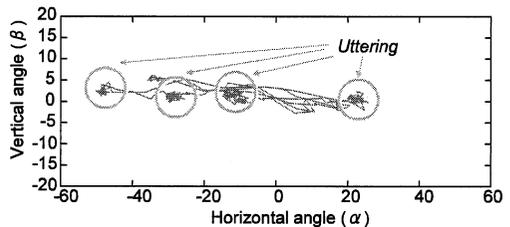


図 8 頭部方向の値と音声メモの対応例

る。音声発声が  $0.5$  [sec] を超えた時に音声発話開始とみなし、その後音声停止区間が  $1.0$  [sec] を超えたときに音声発話停止とみなす。ただし、発話開始判定後に音声停止区間が  $2.0$  [sec] 以上のときは、発話開始判定を覆し発声区間とみなさない。このような音声区間の検出の自動化が不可能な状況、例えば距離の近い自分以外の発話者が多い等であれば、発話区間判定を手動スイッチなどに切り替えることも考えられるが、ユーザビリティやスムーズな入力を想定すると、便宜上は自動であることが望ましい。

この間、音声発声区間を検出した際に用いるため、頭部方向推定結果の値も常に  $5$  [sec] のバッファに記録し書き換え続ける。頭部方向のセンシング結果例から、一方向を見て音声を入力したつもりでも、 $5 \sim 7^\circ$  程度の  $\alpha, \beta$  のずれが確認できた（図 8 参照）。よって、音声メモ入力開始時や終了時の値ではなく、発声区間中の  $\alpha, \beta$  の平均値を用いることとする。音声メモの配置は正面を設定した初期値に対する  $\pm 90^\circ$  に行うこととする。

### 3.3.2 音声アイコンの生成

音声メモのブラウジングの際に、音声そのものを複数同時に聞くことは、カクテルパーティー効果をも前提としても、そのメモの内容を聞き取ることに集中する上であまり望ましくないと考えた。よって、抽象化した音声を音声メモの音声アイコンとして用いることで、聞き取り対象の音声メモとそれ以外の音声メモの存在を同時に聴取することを実現する。

音声アイコンを準備する際の抽象化にあたっては様々な手法が考えられるが、ユーザにとっての正面に対する相対方向に応じ各音声メモの存在を示すことを目的として、左右音量差により音源定位しやすい1000Hz以上を含むことを考慮し、単純な鐘の音の楽音を用いることとした。そして、ユーザの頭部方向（水平方向 $\alpha$ ）の相対角度（0~90）の絶対値に応じて、音声の周波数変換を、1.0倍~2.0倍の間で

$$speed = (12\sqrt{2})^{|\alpha|/90*12} \quad (3)$$

によって速度を変換し行う。これにより、最大で1オクターブ差のある音声アイコンが生成される。

### 3.3.3 方向情報付音声ブラウジング装置

本システムでは提案する音声メモブラウジングの一手法として、一次元的な配置を左右の音量差および時間差を考慮した単純な処理で音源定位を調整できる範囲で再現する。

まず、再現にあたりユーザの正面がどこであるかをシステム開始時のセンサ値により決定する。次に、磁気センサから得られたユーザの頭部方向の絶対角度を、あらかじめ決定された「正面」に対する相対角度 $(\alpha, \beta)$ に変換する。ユーザの頭部方向範囲を相対角度の $\pm 5^\circ$ とし、頭部方向範囲に音声メモが存在する場合、録音した音声メモをそのまま再生する。また、頭部方向範囲外に音声メモが存在する場合は、現在の頭部方向に対する相対角度 $\alpha'_n$ をそれぞれリアルタイムに算出し、現在頭部方向に対する各相対角度に応じ、音声アイコンの音源定位を決定する。

音声メモは正面を設定した初期値に対する $\pm 90^\circ$ に存在するが、ユーザの頭部方向が正面以外を向き、現在の頭部方向範囲 $\pm 90^\circ$ の音声メモおよび音声アイコンを再生対象とする。各音声は、500[msec]の無音区間をはさみながら各音源定位で繰り返し

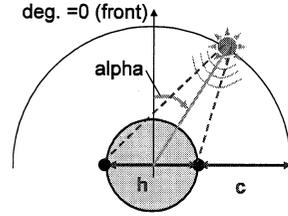


図9 音源定位の設定

再生する。

時間差と音量差は次のように設定する。時間差の計算においては、まず、音源から各耳への距離 $dL, dR$ [cm]を導出する。ここでは、頭部の幅を $h$ 、頭部中心から音源を置く円までの距離を $c + \frac{h}{2}$ として、

$$dL = \sqrt{\{(c + \frac{h}{2}) \cos \alpha'' + \frac{h}{2}\}^2 + \{(c + \frac{h}{2}) \sin \alpha''\}^2} \quad (4)$$

$$dR = \sqrt{\{(c + \frac{h}{2}) \cos \alpha'' - \frac{h}{2}\}^2 + \{(c + \frac{h}{2}) \sin \alpha''\}^2} \quad (5)$$

(ただし $\alpha'' = \alpha + 90$ )

により計算する(図9参照)。そして、時間遅れ $delayL, delayR$ [msec]を音速34[msec/cm]を用いてそれぞれ $delay = d/34$ により導出する。音量差は、通常の音源であれば常に同じ音量であっても定位が変わると左右の耳に届く音量[db]の総合は常に同じとは限らないが、今回の実装では回り込み音声などの計算はしない。よって単純に距離による減衰 $atn$ のみ扱うこととし、 $ATN_{init} = 0.8c^2$ として

$$atn = \frac{ATN_{init}}{d^2} \quad (6)$$

により減衰率を決定する。ただし $d_{min} = c$ とする。実施の際には $c=15, h=15$ としたため、 $ATN_{init} = 180$ となっている。

## 4 考察

### 4.1 センサ値からの頭部方向推定の問題

前節の頭部方向推定の予備実験では、比較的良い推定が得られる被験者と誤差の多く含む結果になってしまう被験者が存在する、という結果が得

られた。この原因としては、あるマーカの方向に顔を向けたつもりでも、実際の頭部方向は、必ずしも意図した方向とは一致しないという可能性が考えられる。これについては、さらなる被験者実験が必要である。

現状の実装は左右方向のみを考慮するという予備的なものであり、上下方向など、それ以外の動きがある場合に誤差を生じていることも考えられる。音声メモ配置のための空間を有効的に利用するという点からも、3軸方向全てについて姿勢が得られることが望ましい。また地磁気のみでは周辺環境に影響されやすいという問題もある。

これに関しては、強い磁気を用いるため周辺環境からの影響を比較的受けにくい3軸磁気センサ(Polhemusなど)を用いるなど、様々な方法が考えられるが、ポータビリティに重点を置いた構成を考慮し、地磁気センサと加速度センサ・ジャイロの組み合わせなどについて検討している。

#### 4.2 音声メモのブラウジング時における音源定位

音源定位では、頭部が固定された状況での聴取は音源位置にセンシティブではないが、頭部動作により変化が起こるため、情報が増え、結果として音源位置がはっきりしやすいことがわかった。つまり本システムにおける頭部方向を逐次変えるような用途の場合、ひとつの方向に停止する状況が少なく、問題はあまり顕在化しなかった。しかし、実際の頭部方向と、音源定位の知覚にずれがあると、本システムを利用するときのメリットである直感性に大きな損傷を生ずる。そのため、音声メモの音源定位の明瞭度の向上、もしくは、最小弁別角度 MAA (minimum audible angle) など、音源位置の明瞭度を考慮した音声メモの配置自動調節機能、などが必要だと考えられる。

一方、周辺情報の提示用の音声アイコンが、ユーザの頭部方向範囲に音源がある状況になった時、音声メモ自体に切り替えられる設計であったが、システムの試行において、スムーズで違和感のない切り替えのための音声アイコンのデザインの検討が必要であることが示唆された。

### 5 おわりに

本稿では、ユーザの頭部方向に応じ仮想空間の位置に関連付けた音声メモの録音と、音源定位を用いた複数の音声メモのブラウジング手法を提案

した。そして、頭部方向を推定する磁気センサを備えたステレオヘッドホンによる音場表現を用いて、仮想空間の次元を限定し、提案システムを試作した。そして、3次元磁気センサモジュールを用いた頭部方向の検出と、音源定位を用いた音声メモの複数配置と音声アイコン化による仮想空間メモシステムの動作を確認した。今後は、磁気センサ以外を用いた頭部方向推定による音声仮想空間ブラウジング手法についてや、ポータビリティのための追加検討、および、音声入出力関係のハードウェアの検討(音声雑音に影響されないスロートマイクもしくはNAMマイクの使用や、音像の作りやすいヘッドフォンの開発)、もしくは普及率の高い通常のイヤホンなどによる、ステレオサウンドでより音源定位を感じやすい音声加工上のソフトウェアでの工夫などを検討していきたい。

#### 謝辞

本研究の機会を下さった ATR 知能ロボティクス研究所の萩田紀博所長、同研究所コミュニケーション支援研究室の安部伸治室長、内海章氏に感謝する。

#### 参考文献

- 1) Katashi Nagao, Shigeki Ohira, and Mitsuhiro Yoneoka. "Annotation-based multimedia summarization and translation." International Conference On Computational Linguistics 2002, Volume 1, pp.1-7, 2002.
- 2) 荒屋真二, 佐藤誠司, 三宅芳博, "多情報仮想環境のための音声情報ボックス," 電子情報通信学会論文誌 D Vol.J89-D No.8 pp.1897-1900, 2006.
- 3) 垂水浩幸, 森下健, 上林弥彦, "SpaceTag のアプリケーションとその社会的インパクト," 情報処理学会グループウェア 33-6, pp.31-36, 1999.
- 4) 暦本純一, 綾塚祐二, 林一輝, "Augment-able Reality: 実空間と情報空間を融合した情報交流," WISS98, pp.115-124, 1998.
- 5) 寺田智裕, 神原誠之, 横矢直和, "拡張現実感を用いた車載型注釈提示システムの構築," MIRU2002, Vol.II, pp.359-364, 2002.
- 6) 興梠正克, 蔵田武志, 坂上勝彦, 村岡洋一 "パノラマ画像群を位置合わせに用いた注釈つき映像の実時間提示システム: パノラマベースト・アノテーション" 信学技報 PRMU99-64, pp.1-8, 1999.
- 7) 矢入聡, 岩谷幸雄, 鈴木陽一, "頭部運動感応型ソフトウェア聴覚ディスプレイの開発," TVRSJ vol.11, No.3, pp.437-445, 2006.