情報処理学会論文誌 Vol.56 No.1 306–315 (Jan. 2015)

三次元音響と周波数変調を組み合わせた被写体位置呈示のための音声フィードバック手法の提案

瀬古 圭一[1,a)] 福地 健太郎[1,b)]

受付日 2014年5月7日, 採録日 2014年10月8日

概要：ビューファインダを見ずに被写体を撮影する用途向けに，被写体の位置を音声フィードバックで呈示する新手法を提案する．従来の手法では，ビューファインダ上の被写体の水平座標を三次元音響による仮想音源の左右の位置で，垂直座標を周波数の高低でそれぞれ呈示していたが，垂直座標の推定精度が低いことが課題となっていた．本論文ではこれを解決するため，周波数の上昇と下降で画角の中心からの上下方向を示し，周波数の変化量で画角の中心からの距離を表す手法を提案する．実験により，提案手法は垂直座標の推定精度を向上させること，また水平座標の呈示と組み合せた場合でも二次元座標の位置推定精度が向上することを示した．加えて，実際にカメラを持たせた場合の条件において，1.84 度程度の誤差で被写体を中心に収めて撮影できることが分かった．提案手法は写真撮影への応用にとどまらず，二次元座標を音響的に呈示する手法として広く利用できる．

キーワード：可聴化，ウェアラブルコンピューティング，写真撮影，音響ナビゲーション

Auditory Feedback Interface Using Spatial Audio with Frequency Modulation for Object Tracking

Keiichi Seko[1,a)] Kentaro Fukuchi[1,b)]

Received: May 7, 2014, Accepted: October 8, 2014

Abstract: We introduce a novel auditory feedback technique for estimating the position of a photographic target for finderless camera. Previously our auditory feedback technique represented the horizontal axis of the target by spatial audio and the vertical axis by frequency, but its low accuracy of estimation of vertical position was a problem. In this paper, we propose a new method that represents the direction to the target by the direction of the change of the frequency, and the distance from the center of the view to the target by the pace of the change. We conducted two experiments and confirmed that our method improves the accuracy of estimation in both vertical-only and two dimensional navigation. In addition, we also conducted a user study with a DSLM camera and our method achieved a median error rate of 1.84 degrees. The proposed method can be applied not only to photographing but also to various two dimensional audio navigation.

Keywords: sonification, wearable computing, photographing, audio navigation

1. はじめに

写真や動画の撮影においては一般的にビューファインダにより被写体の位置を確認する．しかし，動く被写体を追いかけながら撮影する場合では，ビューファインダに注視していると周囲の環境に注意が及ばなくなり，衝突や転倒の恐れがある．また，スナップ撮影のような手軽な撮影条件では，ビューファインダで被写体を確認せずにおおよその位置で見当をつけてシャッターを切る場合もある．近年ではこうした撮影を想定したビューファインダを搭載しないカメラ（Sony DSC-QX100，GoPro, Inc. GoPro シリーズなど）が登場し，ビューファインダを使わない機会が増えてきている．こうした撮影において，被写体を画角内に

[1] 明治大学
Meiji University, Nakano, Tokyo 164–8582, Japan
[a)] keiichi.seko@gmail.com
[b)] kentaro@fukuchi.org

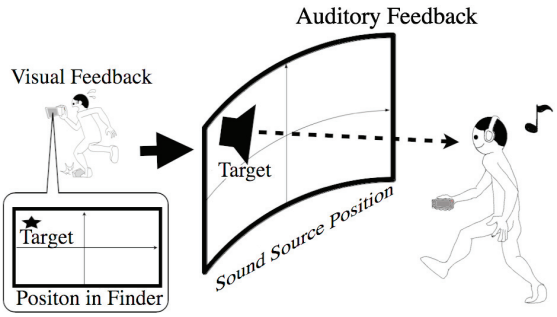


図 1 音声フィードバックを用いた撮影のコンセプト
Fig. 1 An illustration of audio navigation for finderless photographing.

確実に収めるためには適切な誘導が必要である．

我々はこうした撮影シーンや，視覚的に目標位置を確認するのが困難な状況に対応するため，音声フィードバックを用いて目標の相対的位置を呈示する手法をこれまでに提案した [1]．この研究では被写体の位置を三次元音響を用いて呈示することを当初検討したが，水平座標に比べて垂直座標の位置推定精度に問題があることが平原らの研究 [2] で分かっていたため，水平座標の位置呈示には三次元音響を，垂直座標の位置呈示には再生音の周波数の高低をそれぞれ割り当てていた．しかし実験の結果，この手法においても精度に課題を残すことが分かった [3]．特に，被写体が画角の中心から上下のどちらに位置しているのかの判断に迷い，座標推定が遅くなることが課題となっていた．

本論文では，垂直座標の呈示手法の改善策として，再生音の周波数を変調させることで，方向と距離を知覚しやすい形で呈示する手法を提案する．具体的には，基準となる周波数の音を再生し，被写体が画角の中心より高い位置にいれば周波数を上昇させ，低い位置にいれば周波数を下降させることで上下方向を示す．またその上昇・下降の変化量で画角の中心からの距離を示す．この周波数の変化を短時間に何度も繰り返し呈示する．水平座標については従来の手法と同じく三次元音響の手法を採用する．

本呈示手法の推定精度を調べた結果，周波数が上昇と下降のどちらに変化したかの認識精度が高く，周波数の変化量の認識精度も高いことが分かった．また，三次元音響を組み合わせた二次元座標呈示の精度を調べた結果，垂直方向と斜め方向のカメラ移動の精度が向上していることが分かった．また，ビューファインダを見ない状態での静止画撮影に提案手法を適用した場合には，被写体の位置合わせにおける角度誤差の中央値が 1.84 度と，高い精度で行えることを確認した．

2. 関連研究

カメラ撮影時に被写体の位置を音声情報で呈示するものとして，Jayant らは，「Top」「Bottom」「Right」「Left」などの言語情報によって被写体の位置を呈示する手法を提案している [4]．言葉を組み合わせて，たとえば「Top, Right」なども用いることで，計 9 カ所の位置が呈示できるが，詳細な座標を呈示することはできない．

空間情報を音声情報に変換して呈示する研究としては，Panëels ら [5]，Gonzalez-Mora ら [6]，Shoval ら [7] の研究があげられる．いずれも視覚障害者向けの誘導手法の研究で，障害物の回避や目的地への誘導を目的としている．これらの研究は三次元音響を用いて空間内の対象物の位置を呈示しているが，呈示する方向は水平面に限定されており，仮想音源の水平方向の位置変化と距離に応じた音量の変化を利用している．我々はカメラ撮影を目的としており，カメラの光軸に直交する二次元平面を対象としているため，これらの手法をそのまま適用することはできない．

3. 提案手法

3.1 音声フィードバックの要件

カメラ撮影時に被写体位置を音声フィードバックにより呈示する上で，それを有効に支援するための要件を下記にまとめ，あわせて本報告での達成目標について述べる．

(1) **被写体の位置推定の精度**
被写体を直接目視可能な状況で，音声フィードバックによる誘導をあわせることで，ビューファインダで画角内の被写体位置を確認するのと同等の精度を目標とする．

(2) **撮影者の状態**
撮影者がビューファインダを目視しながら撮影するのが困難な状況，あるいは被写体をファインダー越しではなく直接目視したいような撮影状況を対象とする．ファインダーの目視が困難な状況としては，スポーツやレクリエーションなどのシーンにおいて，撮影者・被写体とも移動している状況を想定している．ただし本報告での評価実験は，撮影者は室内で静止した状態で行っている．撮影者に運動をさせた場合については文献 [3] を参照されたい．

(3) **対象となる被写体**
(2) で述べたように，被写体は動いていることを想定している．主な場面としてはスポーツ中の人物を想定している．本報告ではまず位置推定の精度について評価しており，移動している被写体に対する有効性は未評価である．

(4) **撮影時間**
想定する撮影状況は，カメラによる動画撮影であり，被写体をビューファインダの内側に収め続けることが目標となる．したがって，撮影開始時に被写体をビューファインダに収めるまでの時間については，全体の撮影時間から考えれば，数秒の所要時間は許容範囲であると考える．ただし，提案手法を静止画（スチル）撮影に適用する場合には，できる限り素早く位置あわせ

できることが求められているため，本報告ではそうした応用が可能かどうかを議論するために，静止画撮影を想定した実験を実施した．

3.2 従来の手法

我々はこれまでに，ビューファインダを見ない撮影で被写体の位置を認識できるよう，ビューファインダ上の被写体の位置座標を三次元音響と周波数の高低によって呈示する音声フィードバックを提案している [1]．この手法では，まず被写体位置をカメラ画像から認識し，その垂直座標に比例して変化する周波数からなる正弦波を生成し，次にそれを三次元音響を用いて，水平座標に対応した位置に定位させる．この手法を評価した結果，垂直座標の推定精度が低いことがこれまでに分かっている [3]．精度が低下した原因として，被写体が画角の中心から上下どちらに位置しているのかの判断に迷いが生じやすいこと，また中心からの距離についても分かりにくいことがあげられた．そのため，垂直座標の呈示手法の改善が課題であった．

3.3 提案手法

上記の課題を解決するために，人は絶対的な周波数の知覚よりもその変化の知覚に優れていることを利用し，再生音の周波数を変化させる，周波数変調の手法を採用した．具体的には，被写体が画面中央より上に位置する場合には周波数を徐々に高くし（上昇音），下の場合には周波数を徐々に低くする（下降音）ことで，まず被写体が上下のどちらに位置するかが明確に知覚できるように設計した．また，周波数の時間あたりの変化量により，画面中央からどれだけ被写体が離れているかを示すこととした．この変化を短かい時間で繰り返すことで，定常的に被写体の垂直座標を呈示することを狙う．

以下に提案手法の詳細を述べる．再生音としては正弦波を用いる．1 回の周波数の変化で，最初に出力する，基準となる周波数を「始点周波数」，最後に出力する周波数を「終点周波数」とする．1 回の周期は 400 ms とし，200 ms の間正弦波を再生する．この間に周波数は始点から終点まで線形に変化させる．その後 200 ms の休止期間を置き，一周期を終える．周期が終わるたびに被写体の位置に応じて終点周波数を更新し，次の周期を開始する．周波数の変化の様子を**図 2** に示す．

再生する周波数の範囲は，可聴領域の中で最も支配的な領域の 300〜2,000 Hz の範囲 [8] に収まるようにする．また，Shoval らの検証によると，三次元音響において仮想音源の位置の誤認が少ないのは，400〜1,000 Hz 付近であることが分かっている [7]．本手法で再生する周波数の大部分がその領域に収まるよう考慮して，十二平均律音階の C5（523.2 Hz）を始点周波数とし，C5 から前後 1 オクターブ，C4（261.6 Hz）から C6（1,046 Hz）を終点周波数の範囲とした．垂直座標と終点周波数の関係については，音高の知覚が周波数変化に対数比例することを考慮し，指数関数的に対応させる．ただし今回は計算負荷を軽くするために，これを次式で表す二次曲線で近似した．

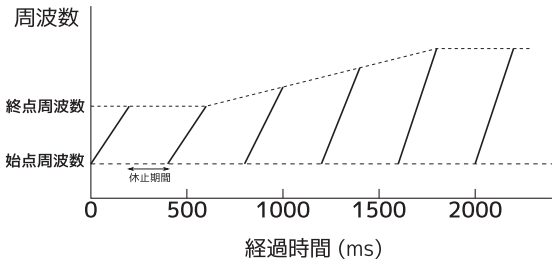


図 2 提案手法の始点周波数から終点周波数までの変化を模式的に表したもの．被写体の移動にあわせて終点周波数は変化する

Fig. 2 An illustration of the frequency modulation pattern. The frequency at the end of a cycle is changed according to the vertical position of the target.

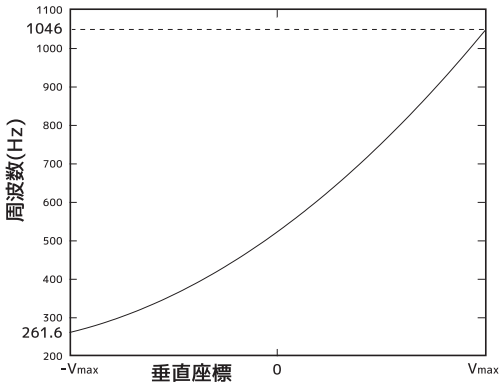


図 3 被写体の垂直座標と終点周波数の関係

Fig. 3 The relationship between the vertical position of the target and the frequency at the end of a cycle.

$$f_{\mathrm{end}} = \frac{f_{\mathrm{o}}}{4V_{\mathrm{max}}^2} v^2 + \frac{3f_{\mathrm{o}}}{4V_{\mathrm{max}}} v + f_{\mathrm{o}} \qquad (1)$$

ここで f_{end} は終点周波数，f_{o} は始点周波数，v は垂直座標の中心を o としたときのそこからの相対垂直座標，V_{max} は垂直座標の最大値を示す．

3.4 被写体のトラッキング手法

被写体の座標をフィードバックするためには，対象となる被写体を撮影者が指定し，またその位置をカメラ側で認識し続ける必要がある．本論文は音声フィードバックの改善を主題としているため，被写体の指定手法とトラッキング手法については取り扱わない．被写体を指定する手法としては，画面をタッチしたりレーザーポインタで指し示すなどの方法が考えられる．トラッキング手法としては，指定した被写体の色相ヒストグラムを参照しながらトラッキングする CAMSHIFT [9] 法やパーティクルフィルタを使った手法 [10] があげられる．本研究では実験 3 において CAMSHIFT 法を採用して被写体をトラッキングしている．

4. 実験 1：垂直座標の推定精度の検証

垂直座標を周波数変調により呈示する提案手法を，一定の周波数で呈示する従来手法と比較する実験を行った．この実験では，呈示された音響フィードバックから垂直座標をどれだけ正確に推定できるかをそれぞれの手法で計測し比較する．

4.1 周波数パターン

実験では垂直座標を示す周波数のパターンを 2 種類使用する．従来手法と同様に，一周期の間再生周波数が変化しないものをパターン A，提案手法によるものをパターン B とする．本実験では垂直座標を −100%〜100%の範囲で表し，その範囲を 13 段階に分割し，各段階での音を被験者に呈示し回答させる．パターンごとの，各段階における再生周波数を**表 1** に示す．ここで “-M” の ID が付された段階は画面中央位置に対応する．

4.2 実験システム

実験用アプリケーションについて，画面（**図 4** 左）に沿って説明する．タスクの開始時にはダイアログが表示され，音が再生されることを予告する．被験者によってボタンが押されると，先に上げた 13 段階の呈示音のうち 1 つが再生される．被験者はその音がどの垂直座標を示しているかを判断し，回答用バーの該当箇所をタッチして回答して，タスクは終了し，次のタスクを開始する．呈示音はボタン操作で何度でも再生できるようにした．

次節で詳述するが，練習フェイズにおいては，被験者は目安となる音をボタン操作により再生し，目安を確認することができる．回答用バー右側に，「100%」「50%」「0%」「−50%」「−100%」の 5 つのボタンが配置され，それぞれ H6，H3，M，L3，L6 の音に対応している．

実験用アプリケーションは，タッチスクリーンを搭載したノート PC（SONY SVT131B11N）上で実行した．音の出力はヘッドフォン（SONY MDR-CD900ST）により，音圧レベルを 70 dB に設定し，実験は静かな環境下で実施した．

表 1 パターン A，B の始点周波数と終点周波数（左：パターン A，右：パターン B）

Table 1 Frequency tables for frequency pattern A (left) and B (right).

音 ID	始点周波数 (Hz)	終点周波数 (Hz)
A-H6	1046	1046
A-H5	932.3	932.3
A-H4	830.6	830.6
A-H3	739.9	739.9
A-H2	659.2	659.2
A-H1	587.3	587.3
A-M	523.2	523.2
A-L1	466.1	466.1
A-L2	415.3	415.3
A-L3	369.9	369.9
A-L4	329.6	329.6
A-L5	293.6	293.6
A-L6	261.6	261.6

音 ID	始点周波数 (Hz)	終点周波数 (Hz)
B-H6	523.2	1046
B-H5	523.2	932.3
B-H4	523.2	830.6
B-H3	523.2	739.9
B-H2	523.2	659.2
B-H1	523.2	587.3
B-M	523.2	523.2
B-L1	523.2	466.1
B-L2	523.2	415.3
B-L3	523.2	369.9
B-L4	523.2	329.6
B-L5	523.2	293.6
B-L6	523.2	261.6

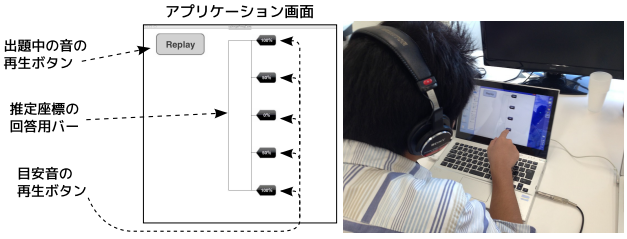


図 4 実験 1 用アプリケーション画面（左）および実験の様子（右）

Fig. 4 The application of the experiment #1 (left) and its overview in use (right).

4.3 実験手順

各被験者に対しては，練習フェイズと本番フェイズの 2 つを続けて実施した．練習フェイズでは，目安となる音をセッション中何度でも確認できるようにした．本番フェイズでは，目安音の確認は不可とした．両フェイズとも，呈示音は前掲の 13 種類の音の中からランダムに 20 回再生する．ただし，13 種類の音はすべて一度以上再生されるよう調整している．

被験者として，健聴者の大学生 6 名に参加してもらった．実験の様子を図 4 右に示す．

4.4 結果

図 4 に，両パターンにおいて，呈示音ごとに回答された座標の平均値と標準偏差を示す．パターン A（図左）では，再生音の周波数が高くなるのに比例して，推定座標も下から上へと滑らかに推移するのが理想だが，A-M から A-H1 の間，A-H2 から A-H3 の間にかけて，周波数と回答座標の平均値との関係が逆転している．一方，提案手法であるパターン B（図右）においては，B-H2 から B-H3 の間，B-H4 から B-H5 の間で，回答座標が横ばいになっているが，パターン A で見られたような逆転関係は生じなかった．逆転が見られる箇所においては，2 つの隣接した箇所の再生音の高低が正しく識別できなかった被験者が多かったことを示しており，座標の呈示手法として問題があるが，提案手法はこれを改善できることが示された．

原点付近での挙動にも特徴が見られた．パターン A の原点（A-M）での回答座標は平均値で 11.7%と，0%から大きく外れている．また，正の座標である A-H1 に対する回答座標において，22%の回答が負の座標であった．これらは，原点付近で判断を間違う被験者が多かったことを示している．一方で提案手法においては，原点である B-M での平均値は −0.17%で，標準偏差は 2.43 であった．タッチパネル上での被験者の指の接触面の大きさが換算で 1.7%分に

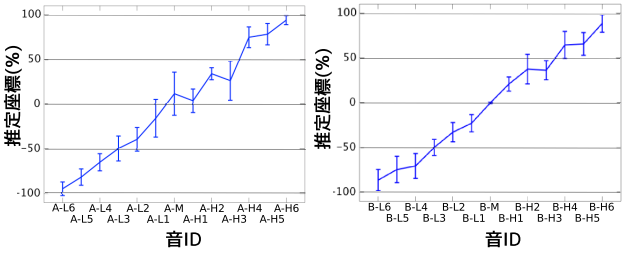


図 5 各パターンの周波数と推定座標の関係
（左：パターン A，右：パターン B）

Fig. 5 Estimated positions for every sound ID. Left: pattern A. Right: patterb B.

相当することを考えると，これ以上の精度での観測は本実験手法ではできないが，十分に小さな誤差に抑えられていると考える．原点に隣接する B-L1 および B-H1 での回答は，いずれもすべて原点より下および上の座標と回答されており，原点付近での上下の判断で間違いにくい手法であることが分かった．

4.5 考察

提案手法の用途はナビゲーションであり，現在位置が目標付近にいることが正しく知覚されることが重要である．呈示音により目標からのずれが示された際に，上下のどちらに移動すればそのずれを減少させることができるのか，の判断が迷いなくできることは，それだけ迅速に位置合わせ操作を実行できることにつながる．従来手法ではこの点に課題を抱えていた．

また，本研究においては，水平座標と垂直座標は異なる手法によりフィードバックされるため，垂直座標が合っている状態で水平座標を調整する場合などに，それ以上垂直座標を動かす必要がないことが利用者に伝わることが求められる．そのため，原点にいることが利用者に正しく知覚されることが重要となる．

実験により，提案手法は従来手法と比べて，原点付近での操作で迷いを生じにくい，より目的に適した呈示手法であることが示された．

5. 実験 2：二次元座標の推定精度の検証

提案手法により，垂直座標の推定精度は改善された．これを水平座標の呈示と組み合わせた際に，二次元座標の推定精度が改善されるかどうかを，位置合わせにかかる時間やその過程を分析することにより検証した．

5.1 実験概要

この実験では，被験者の操作履歴を正確に記録し検証することを目的としているため，被験者には，本来の応用であるカメラの操作のかわりに，デスクトップ PC 上で，マウスにより模擬カメラを操作させた．マウスポインタをカメラ画角の中心とし，カメラを振る操作はマウスの前後左右の操作で行う．被写体は画面上に，被験者には見えないよう仮想的に設置される．カメラ画角はマウスポインタを中心とした矩形により設定され，画角内の被写体の位置に応じた音声フィードバックが被験者には呈示される．被験者はフィードバックを頼りに，画角の中心を被写体に重ねることが求められる．

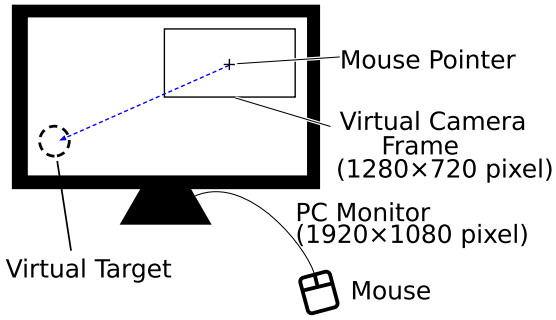


図 6 実験 2 用システムの構成図

Fig. 6 System settings for the experiment #2.

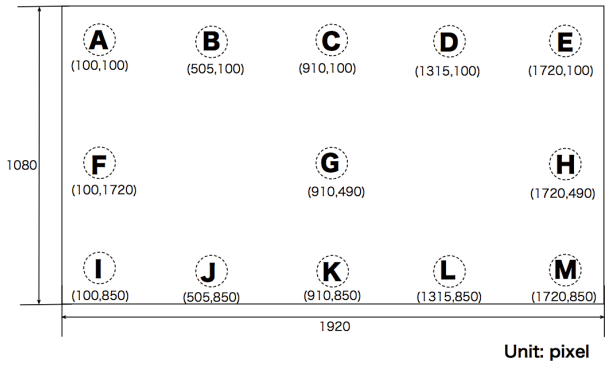


図 7 被写体の配置パターン

Fig. 7 Positions of the virtual targets.

本実験では，被写体までの移動にかかる時間と，その過程におけるマウスの動きを分析することで，従来手法と提案手法との差を検証し，また提案手法の特徴について分析を行った．

5.2 実験方法

本実験で使用したシステムの構成を図 6 に示す．PC モニタは実寸 60 cm × 33.8 cm，解像度 1,920 × 1,080 pixel のものを使用した．画面上には，クロスカーソルで示されたマウスポインタを中心に，カメラ画角を模した矩形がつねに表示されており，その大きさは 1,280 × 720 pixel である．

被写体は，図 7 に示した 13 カ所のいずれかに設定される．ただし，被験者には音声フィードバックのみを手がかりに被写体位置を推定してもらうため，画面上には被写体は示されない．音声フィードバックの呈示手法については次節で説明する．

被写体の大きさは，直径 100 pixel に設定されている．マウスポインタがこの領域に入ると，被写体に到達したことを被験者に伝える効果音を再生し，タスクは終了する．そ

の後被写体の場所が変更され，次のタスクを開始する．1セッションで20回のタスクが与えられる．被写体の位置はランダムで与えられるが，被験者間での軌跡の比較のために，すべてのセッションでA→M，M→A，I→E，E→I，C→K，K→C，F→H，H→Fの8経路が必ず含まれるようにした．

以上のセッションを，従来手法（A）と提案手法（B）のそれぞれの垂直座標呈示手法ごとに1回ずつ，計2セッションを被験者1人あたりにつき実施した．被験者として，健聴者である大学院生9名に参加してもらった．

5.3 音声フィードバック

カメラ画角を模した矩形は，マウスポインタの移動にあわせて移動する．このカメラ画角内に被写体の中心位置が入っている間，被写体中心の画角中心からの相対座標に応じて，音声フィードバックが生成される．

相対座標の垂直成分に対しては，従来手法であれば正弦波の周波数の高低で，提案手法であれば式(1)によって変調された正弦波によって，それぞれ再生音を生成する（4.1節を参照）．ただし，プログラムの間違いのため，実際に呈示された周波数の範囲は277.2 Hz～1,108 Hzになってしまっていたが，違いはわずかであったため，ここでは結果に影響を与えないと判断し，そのまま使用した．

相対座標の水平成分に対しては，上で生成された正弦波を三次元音響技術を用いて，水平方向の定位を与える．三次元音響技術としては，Pure Dataのearplugモジュール[11]を使用した．これはKEMARダミーヘッドを用いて計測した頭部伝達関数（HRTF: Head Related Transfer Function）を基に作成されたもので，被験者ごとのHRTFを計測・使用するものではない．

5.4 結果と分析

5.4.1 移動方位の精度

提案手法のナビゲーション手段としての有効性を検証するために，まずマウスの移動方向の精度を分析する．そのために，マウスポインタの全軌跡を対象に，マウスの移動方向が正しく被写体に向かっているかどうかについて，全移動時間に対して正しく動いている時間の割合を算出した．

まず，400ミリ秒ごとのマウスの座標を利用し，各期間でのマウスの移動方向θを算出する．このとき，区間頭でのマウスの座標から目標となる被写体までの方位角をϕとし，このときの$|\theta - \phi|$を移動角の誤差として扱う．この誤差が22.5°に収まっている区間を，マウスが正しい方向に移動できている区間と定義する．

この誤差について，各区間における被写体までの方位角ϕを8方向に分割し，それぞれについて正しい方向に移動できている割合を算出し，図示したのが**図8**である．たとえば被写体が上方向（$-22.5° < \phi < 22.5°$）に位置する区間では，従来手法（A）では約50%の時間で正しく上方向（$-22.5° < \theta < 22.5°$）に移動しているのに対し，提案手法ではその割合が約65%に改善されている，ということがこの図には示されている．

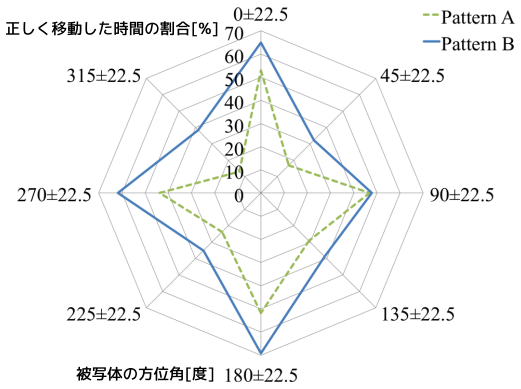


図8 正しい方向に移動した時間の割合
Fig. 8 Correct motion ratio of the expreriment #2.

図8より，正確な方向へ誘導できている割合は，被写体が右方向（$67.5° < \phi < 112.5°$）にあるとき以外では，提案手法の方が従来手法よりも高いことが分かった．特に，垂直方向への移動については，実験1の結果から予期されたように，推定精度が改善できていることが分かる．また，その他の方向での結果からも，提案手法は水平座標のフィードバック手法と組み合わせた場合に従来手法よりも有効であることが示された．

一方，被写体が右方向にある場合には，従来手法と提案手法との間に差が見られなかった．これについては詳細な原因は判明していない．左右で改善傾向に差がある非対称の原因としては，被験者は全員右利きであり，マウスの操作上の非対称性が生じていたことが考えられる．また，三次元音響は被験者ごとに調整されたものではないため，細かい調整が求められる段においてはその差が不利に働くこともあげられる．加えて，「利き耳」に相当する特性があることも考えられよう．いずれも確かなことは分かっていない．

被写体が斜め方向にあるとき，提案手法による改善は見られるものの，正しく移動できた割合は40%前後であり，垂直および水平方向への移動での割合（60～70%）と比べると低い．提案手法による垂直座標の推定精度の向上が改善に寄与している一方で，組合せによる斜め方向への移動の呈示は，垂直座標の推定精度の改善のみでは解決できないことが示唆される．

5.4.2 到達時間

マウスポインタの移動ルート，A→M，M→A，E→I，I→E，C→K，K→C，F→H，H→Fについて，マウスポインタが被写体に到達するまでの時間の分析結果を**図9**に示した．箱は第三四分位から第一四分位数，箱内の横線は中央値，上下の髭は最大値，最小値を表している．

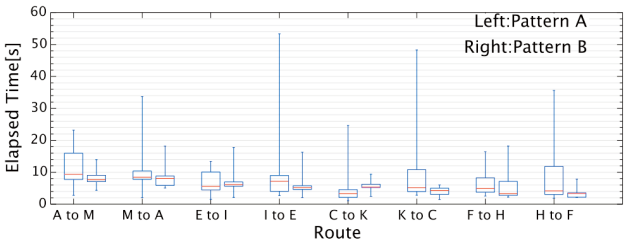


図 9 到達時間の比較

Fig. 9 Comparison of task completion time.

到達時間を比較した場合，中央値と最小値については，すべての移動経路で呈示手法間に大きな差が見られなかった．しかし，到達時間の分散具合については，A → M，E → I，I → E，K → C，H → F の移動経路で提案手法のほうが従来手法よりも四分位範囲（IQR）が小さい．これらの経路について，F 検定により等分散性を検定した結果，A → M（$P < .05$），I → E（$P < .001$），K → C（$P < .001$），H → F（$P < .001$）の経路において等分散であることが否定された．このことから，これらの経路においては安定して短時間に被写体まで到達していることが分かる．一方，分散具合に大きな差が見られなかった M → A，C → K の経路に関しては，最大値は提案手法の方が従来手法よりも小さく，ワーストケースの改善が見られる．

提案手法による結果のうち，中央値から 3IQR 以上の時間がかかったものが A → M a の 13.9 秒，M → A の 17.7 秒，E → I の 18.2 秒，I → E の 16.2 秒であった．これらのケースにおいて，どうしてこのような時間がかかっているのかについて分析した結果を次項で述べる．

5.4.3 マウスポインタの軌跡の分析

到達時間が長くなった前述の 4 件で，なにが要因となって余分な時間が経過したかを分析するために，それぞれの被験者によるマウス操作の履歴を分析した．マウスポインタの座標は 5 ミリ秒ごとに記録されており，それらをプロットしたものを**表 2** に示す．図中，青い丸で被写体領域を，赤い点でマウスポインタの軌跡を示している．500 ミリ秒ごとに経過時間を付記している．赤い点が重なって色が濃くなっている箇所では，マウスが停留していることを示している．

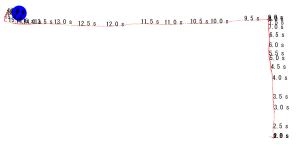
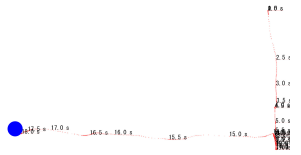
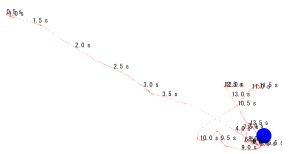
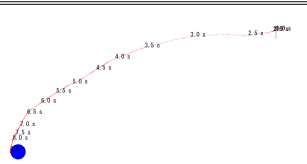
4 件のデータを概観すると，I → E，M → A，E → I の 3 件については，まず垂直移動してからその後に水平移動し，全体として直角を描くような軌跡を残している．また，A → M の経路では，被写体付近で酔歩しているような軌跡が見られる．我々はこれらの 2 種類の軌跡をそれぞれ「傾向 1」「傾向 2」と呼び，2 つを分けて議論することとした．

傾向 1：水平座標と垂直座標を逐次的に調整する

M → A，E → I，I → E の 3 件は，水平方向と垂直方向とを逐次的に移動させている．この 3 件は同一の被験者によるものであった．比較のため，逐次的な調整ではなく水平方向と垂直方向とを同時に操作している例を表 2 の最下行に示しているが，多くの場合はこのような軌跡が描かれる．M → A で 17.7 秒かかった軌跡では，8 秒かけて上方向に移動し，さらに 10 秒かけて左方向に移動している．また，E → I で 18.2 秒かかった軌跡では，5 秒かけて下方向へ移動してからその場に 10 秒ほどとどまった後，3 秒かけて左方向に移動している．この被験者は他のルートでも同様にまず先に垂直方向に移動した後に，水平方向に移動していた．

表 2 到達時間の外れ値の軌跡

Table 2 Trajectories of the outliers from the experiment #2.

経路	時間	マウスポインタの軌跡	軌跡の傾向
I → E	16.2 s		傾向 1
M → A	17.7 s		傾向 1
E → I	18.2 s		傾向 1
A → M	13.9 s		傾向 2
E → I	8.5 s		参考

この被験者に実験後，音声フィードバックの使用感を尋ねると「三次元音響の音が認識しづらいので，まずは垂直方向へ移動してから水平方向へ移動した」と回答した．原因は不明だが，被験者の個人差や，使用した三次元音響モジュールが被験者になじまなかった，などの原因が考えられる．

傾向 2：被写体付近で座標推定の精度が低くなる

A → M の場合では，被写体に近づくにつれて位置推定の精度が低くなっている．**図 10** は，被写体付近でのマウスポインタの軌跡を拡大した図である．被写体の近く，中心から左に 130 pixel 程度までは 4 秒で到達しているが，その後 5 秒間ほど，その周辺をさまよっている．その後，一

時被写体中心から左上 250 pixel ほど離れたところまで移動した後，12 秒の時点から右下方向へ移動し，2 秒後に被写体までたどり着いている．

この被験者の報告によると，「被写体に近づくにつれて，座標推定が難しくなった．一時被写体から離れた理由は，座標推定しやすくするため」という回答が得られた．位置推定を難しく感じていたのは 4.0〜9.0 秒の間で，遠くに離れて座標推定をしやすくなったのは，12 秒の時点に該当するものと考えられる．4.0〜9.0 秒の期間では，水平方向の移動が主である．同被験者は実験 1 も受けており，その結果は他の被験者と同様，垂直座標の推定に特に問題は生じていない．そのため，この期間に垂直座標の推定に問題が生じたとは考えにくい．一度被写体周辺から離れることによって位置を推定しようとした理由としては，水平座標のフィードバックが十分でないこと，特に原点付近でのフィードバックで問題が生じていることが原因として考えられる．

5.5 考察

本実験の結果について，移動方位の誤差および到達時間についての分析結果を考察する．

マウスポインタの移動方位の精度は提案手法による改善が確認できた．しかし斜め方向の正確な移動の割合は，垂直方向のものに比べて低い水準であった．この原因として，垂直座標と水平座標とで異なる性質の変化を組み合せたため，それぞれの座標ごとの逐次的な調整を誘発している可能性が考えられる．ここにさらに斜め方向に被写体があることを示す信号を重ねても，情報過多で利用者がよりとまどう恐れがある．

到達時間の分析で，特に時間のかかった 2 つの傾向について軌跡を分析した結果，両者とも水平座標の推定精度の低さが原因ではないかと考えられた．この主な原因は使用している三次元音響の質による．たとえば利用者の耳の寸法を元に HRTF を調整する Zotkin らの手法 [12] を応用することで水平座標の推定精度を改善できる可能性がある．

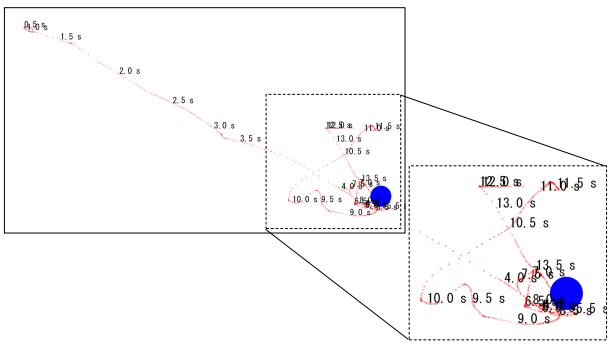


図 10 A → M で 13.9 s かかった軌跡の被写体付近の拡大
Fig. 10 A magnified view of the trajectory of the outlier case (A → M).

また，特に原点付近での判断に迷う場合が見受けられるため，水平座標が原点付近にいることを示す情報を重ねる方法を検討したい．

6. 実験 3：被写体の位置合わせ精度の検証

本実験では，実際のカメラを用いて，提案手法による二次元座標の誘導の有効性を検証した．

6.1 実験手法

図 11 左に実験風景を示す．壁に同図右に示した的を被写体として貼っており，そこから 2,100 mm 離れた場所に被験者には立ってもらった．的の一番小さい円は内側が赤く塗られており，その大きさは 64.7 mm であり，被験者の立ち位置からの画角は 1.76 度に相当する．

被験者はカメラによりこの的をできるだけ画角の中心にとらえるようにして写真を撮影することが求められる．被験者は被写体を直接目視できるが，カメラのビューファインダを見ることはできない．音声フィードバックのない状態では，カメラの姿勢から推測して撮影することとなる．音声フィードバックがある場合は，画角内の的の位置に応じたフィードバックが，提案手法により呈示されるため，それを頼りにカメラを操作することとなる．

実験には，健聴者の大学生・大学院生 7 名が参加した．

6.2 実験システム

カメラにはズームレンズをマウントした Sony NEX-5R を使用し，焦点距離は 18 mm に設定した．HDMI 出力を PC でキャプチャし，取得したフレームバッファから，CAMSHIFT 法を使用して的の赤丸の位置を追跡した．カメラにはマウスをテープでとめ，マウスのボタンをシャッターがわりに利用した．被験者はカメラを構え，的を中心にとらえたと思ったら，任意のタイミングでマウスのボタンをクリックして撮影することができる．

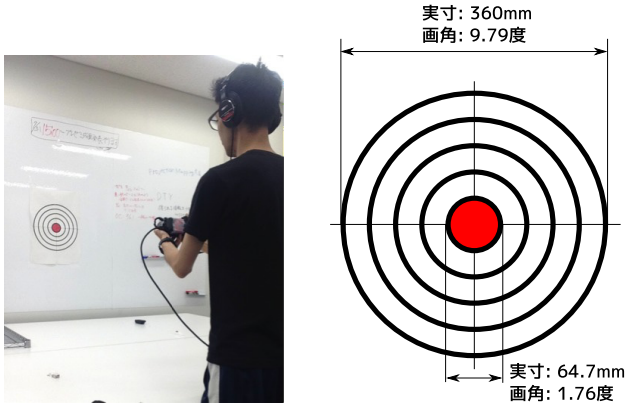


図 11 実験 3 の実施環境（左）および的の寸法（右）
Fig. 11 Experimental environment (left) and the specification of the target (right) for the experiment #3.

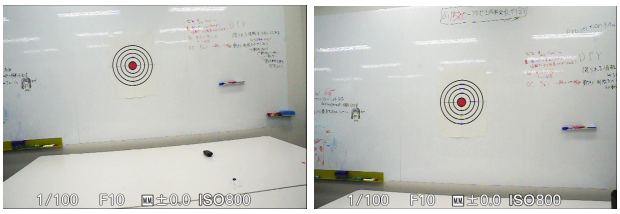


図 12 撮影写真の比較（左：音声フィードバックなし，右：音声フィードバックあり）

Fig. 12 Some result photo images from the experiment #3 (Left: without feedback. Right: with feedback).

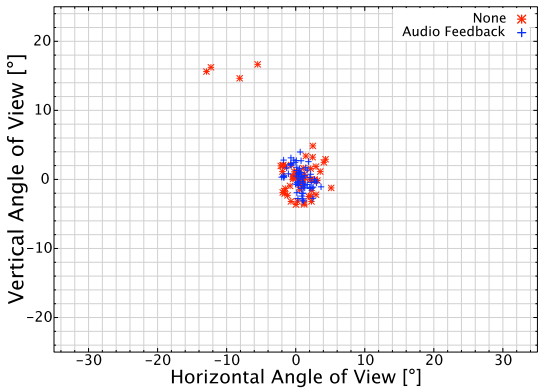


図 13 撮影画像中の被写体の位置

Fig. 13 The positions of the target of photo images.

実験手順としては，まず音声フィードバックのない状態で 5 枚の写真を撮影してもらい，次に音声フィードバックのある状態でさらに 5 枚の写真を撮影してもらった．1 枚写真を撮影する度に，カメラは下に降ろした後に構え直してもらい，続けて写真が撮影できないようにした．撮影された写真の例を 図 12 に示す．

6.3 結果と分析

撮影された 70 枚の写真の，被写体位置をプロットしたものを図 13 に示す．画角中心から被写体までの距離の分布について，図 14 に箱髭図を示した．中央値（図箱内横線）で比較すると，音声フィードバックなしでは 2.59 度，ありでは 1.84 度となった．中心からの距離の平均値は，音声フィードバックがない場合には 3.44 度，音声フィードバックがある場合は 1.95 度であり，t 検定の結果，有意に改善されたことが分かった（$p < .01$）．音声フィードバックにより，分散が小さく，より中央に被写体を寄せた写真を撮影することができることが示された．

カメラを構えてからシャッターを押すまでの時間の分布を，図 15 に示した．平均値では，音声フィードバックなしでは 9.23 秒かかったのに対し，音声フィードバックありでは 26.0 秒であった．中央値はそれぞれ 6.88 秒と 24.3 秒であった．

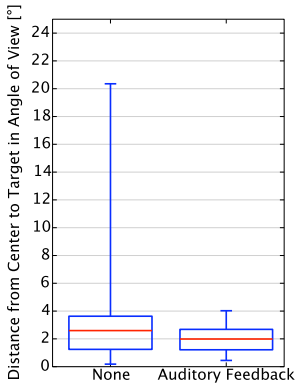


図 14 撮影画像の中心から被写体の中心までの距離

Fig. 14 Distance from the center of the view.

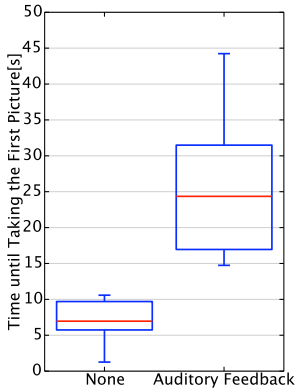


図 15 カメラを構えてシャッターを切るまでの時間

Fig. 15 Time taken to adjust the camera.

6.4 考察

本実験から，カメラの実機を用いた場合でも，提案手法による音声フィードバックが位置合わせの精度を向上させる効果があることが分かった．今回の実験で，ビューファインダを用いた場合の精度は調査していないが，平均値で 1.95 度という精度は，画像中の的の赤丸の大きさが 1.76 度であることを考えると，十分に高い精度であるといえる．

一方で，位置合わせにかかる時間が増加している．20 秒近い増加は，気軽な撮影に向くとはいえない．もちろん時間制約のある条件下では音声フィードバックを補助的に用いて位置合わせをしつつ，制約時間を経過したらシャッターを切ればよいが，音声フィードバックがない条件でもそこそこの位置合わせ精度ができており，増加時間と改善精度とのバランスは考慮する必要があろう．

要件 (4) で示したように，動画撮影を主とした場合にはこの程度の増加時間は無視できるものとして扱ってよいが，素早く動く被写体や，現れたり隠れたりするような被写体を相手にする場合には，位置合わせの速度への需要は高く，今後の検討が必要である．

7. まとめ

本研究では，ビューファインダを用いない撮影において，被写体の位置を把握するための音声フィードバックの手法について検討した．従来手法では垂直座標の呈示手法についてその精度に課題があった．本論文ではそれを改善する手法として，再生音の周波数の変化方向で上下の方向を，変化量で中心からの距離を呈示するものを提案し，その有効性を実証した．

7.1 改善できた点

提案手法では，垂直座標の推定精度が向上したのを確認した．また二次元座標の推定で，垂直方向と斜め方向の位置合わせの精度が向上していることが分かった．カメラ実

機による検証においても精度の高い位置合わせが可能であることが示された．

7.2 今後の課題

7.2.1 撮影者や被写体が移動する場合の座標推定の検証

今回の実験では，音声フィードバックが有効に機能するための要件として 3.1 節で示した，撮影者が運動している場合と，被写体が移動している場合の被写体の位置推定の検証を行っていない．これらの条件での提案手法の有効性を検証する必要がある．吉崎らの研究によれば，三次元音響は受聴者の自発的な運動がともなうことで，音像を定位しやすくなる [13]．撮影者が運動をともなうことで，被写体の位置推定の精度が変わる可能性がある．一方で，被写体の移動速度によっては位置合わせが追いつかない可能性も示唆された．

7.2.2 斜め方向への正確なカメラ移動

垂直座標と水平座標とで異なる性質の変化を呈示しているためか，斜め方向の位置合わせに依然課題が残る．被写体の方位を呈示する性質の音声情報を追加する手法が考えられるが，すでに複数の情報呈示手段を組み合わせており，これ以上の追加は混乱を招く恐れもある．

7.2.3 提案手法の応用

提案した音声フィードバックはカメラ撮影以外の分野にも応用できると考えている．たとえば視覚障害者向けのナビゲーションシステムは，ウェアラブルコンピューティングへ応用し，視野に干渉しない形での作業支援に使えることが期待できる．こうした応用においての問題点を探っていきたい．

謝辞 本研究は JSPS 科研費 26730106 の助成を受けたものである．

参考文献

[1] Seko, K. and Fukuchi, K.: A guidance technique for motion tracking with a handheld camera using auditory feedback, *Proc. UIST '12*, pp.95–96 (2012).
[2] 平原達也，大谷 真，戸嶋巌樹：頭部伝達関数の計測とバイノーラル再生にかかわる諸問題，電子情報通信学会 Foundamentals Review，Vol.2, No.4, pp.68–85 (2009).
[3] 瀬古圭一，福地健太郎：音声フィードバックを用いたカメラ撮影のための動体追跡支援システムの研究，情報処理学会研究報告，Vol.2013-HCI-152, No.9, pp.1–9 (2013).
[4] Jayant, C., Ji, H., White, S. and Jeffrey, P.: Supporting blind photography, *Proc. ASSETS '11*, pp.203–210 (2011).
[5] Panëels, S.A., Olmos, A., Blum, J.R. and Cooperstock, J.R.: Listen to it yourself!: Evaluating usability of what's around me? for the blind, *Proc. CHI '13*, pp.2107–2116 (2013).
[6] Gonzalez-Mora J.L. et al.: Seeing the world by hearing: Virtual Acoustic Space (VAS) a new space perception system for blind people, *Proc. 2nd ICTTA'06*, pp.837–842 (2006).
[7] Shoval, S., Borenstein, J. and Koren, Y.: Auditory guidance with the navbelt – A computerized travel aid for the blind, *IEE Trans. Systems, Man, and Cybernetics, Part C: Applications and Reviews*, Vol.28, No.3, pp.459–467 (1998).
[8] Ritsma, R.J.: Frequencies Dominant in the Perception of the Pitch of Complex Sounds, *The Journal of the Acoustical Society of America*, Vol.42, No.1, pp.191–198 (1967).
[9] Foley, J.D. et al.: Computer Graphics: Principles and Practice (2nd edition), Addison-Wesley (1990).
[10] Bradski, G.R.: Computer Vision Face Tracking For Use in a Perceptual User Interface, *Intel Technology Journal*, pp.214–219 (1998).
[11] Xiang, P., Camargo, D. and Puckette, M.: Experiments on Spatial Gestures in Binaural Sound Display, *Proc. International Conference on Auditory Display*, pp.1–4 (2005).
[12] Zotkin, D.N., Hwang, J., Duraiswami, R. and Davis, L.S.: HRTF Personalization using anthropometric measurements, *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp.157–160 (2003).
[13] 吉崎大輔，平原達也：ハンドル操作で回転制御したダミーヘッドで収録した動的バイノーラル音による水平面音像定位，日本バーチャルリアリティ学会論文誌，Vol.17, No.4, pp.327–331 (2012).

瀬古 圭一 （学生会員）

1983 年生．2006 年よりソニーエンジニアリング株式会社にて，組込みソフトウェア開発に従事し，2012 年退社．2013 年明治大学大学院理工学研究科修士課程了．2013 年より同研究科博士課程に在籍，同大学理工学部助手を兼務し，現在に至る．ウェアラブル端末のユーザインタフェースに興味を持つ．

福地 健太郎 （正会員）

明治大学総合数理学部准教授．2004 年東京工業大学大学院情報理工学研究科博士後期課程単位取得退学．博士（理学）．電気通信大学大学院情報システム学研究科助教，独立行政法人科学技術振興機構 ERATO 五十嵐プロジェクト研究員，明治大学理工学部特任准教授を経て，2013 年より現職．ユーザインタフェースやエンタテインメント応用，音楽・映像分野との協調に興味を持つ．ACM，VR 各会員．2002 年 FIT 船井ベストペーパー賞，2010 年日本 VR 学会論文賞受賞．