

カメラ映像の特徴点追跡を用いた動的音高変化による 方向提示手法の提案

大沼 怜生^{1,a)} 下田 康太¹ 伊勢 侑将¹ 大澤 怜二² 半田 篤史² 川口 一画^{3,b)}

概要：遠隔作業指示においては、実空間上の操作対象に対する注意誘導が必要となるが、通常のビデオ通話システムではその達成が困難である。このような課題に対して、頭部伝達関数を用いた立体音響によって注意誘導を行う手法も提案されているが、汎用的な頭部伝達関数では精度が低くなるという課題があった。そこで本研究では、空間音響を用いずに、利用者の頭部方向と目標角度の差分に応じて動的に変化する非空間音響を用いて注意誘導を行うシステムを提案する。提案手法では、利用者の頭部に装着したカメラの映像から特徴点を追跡し、利用者の頭部方向および目標角度との差分を計算してヘッドフォンから提示する聴覚情報の音高を変化させる。本研究では、映像の特徴点追跡を用いた音高変化による方向提示システムのプロトタイプを実装した。

1. はじめに

ビデオ通話を用いた遠隔会議は広く普及しているが、通常のビデオ通話システムでは注視や指差し等の空間的参照の伝達が困難であり、円滑なコミュニケーションの達成が困難な状況が存在する。例として、工場等で遠隔地にいる指示者が現地側作業者に機械等の組み立てや操作手順を指示する遠隔作業指示が挙げられる。遠隔作業指示においては、実空間上の操作対象に対する注意誘導が必要となるが、通常のビデオ通話システムではその達成が困難である。このような課題に対して、ARを用いて実空間中の指示対象物上に指差しやポインタ等の指示情報を重畳表示し、注意誘導を行う手法が提案されている [3], [12], [15]。このような手法では、視覚的情報により指示位置が一意に定まる一方で、HMD やタブレット端末等の機材が必要になること、HMD の着用やタブレットの把持が作業の妨げになる等の制約があった。このような課題に対して、視覚情報を用いず空間音響を用いて注意誘導を行う手法が提案されている [4], [13]。空間音響を用いた注意誘導では、頭部伝達関数を用いた立体音響によって疑似的に特定の方向から音が到来する状況を再現する。ただし、利用者毎に頭部伝達関数を測定することは困難であり、事前に記録された汎用的

な頭部伝達関数では精度が低くなるという課題があった。

このような課題に対して、著者らは先行研究において、空間音響を用いずに、利用者の頭部方向と目標角度の差分に応じて動的に変化する非空間音響を用いて注意誘導を行うシステムを提案した [16]。ただし、頭部方向の検出のためにモーションキャプチャカメラの設置およびマーカの装着が必要であり構成が複雑であった。そこで本研究では、遠隔作業支援場面を想定し、利用者の頭部に装着したカメラ映像を用いてモーションキャプチャシステム等を用いることなく注意誘導を実現するシステムを提案する。システムの利用状況として、遠隔指示者が利用者の一人称視点映像を見ながら、映像の中にて特定のターゲットを指定することを想定している。具体例として、部屋のレイアウト変更時に遠隔から移動したい対象物を指定したり、その移動先を指示したりする際に利用することが考えられる。想定する作業環境およびシステムの挙動を図1に示す。

提案手法では、利用者の頭部に設置されたカメラによって取得した映像を用いて特徴点を追跡し、目標角度との差分に応じてヘッドフォンから提示する聴覚情報の音高を変化させる。そのため、提案手法ではカメラ映像の特徴点追跡によって利用者の頭部方向を算出した。また、利用者は音高の変化を手掛かりに提示方向を動的に探索する。本研究では、提案手法が利用者の注意誘導に与える影響を調査するため、システムのプロトタイプを実装した。システムには、今後注意誘導の影響調査のための実験を行うことを想定して、提案手法、音を用いた既存の注意誘導手法である空間音響、提案手法による音高変化および空間音響によ

¹ 筑波大学大学院 理工情報生命学術院 システム情報工学研究群 情報理工学位プログラム

² 筑波大学 情報学群 情報メディア創成学類

³ 筑波大学 システム情報系

^{a)} onuma@iplab.cs.tsukuba.ac.jp

^{b)} kawaguchi@cs.tsukuba.ac.jp

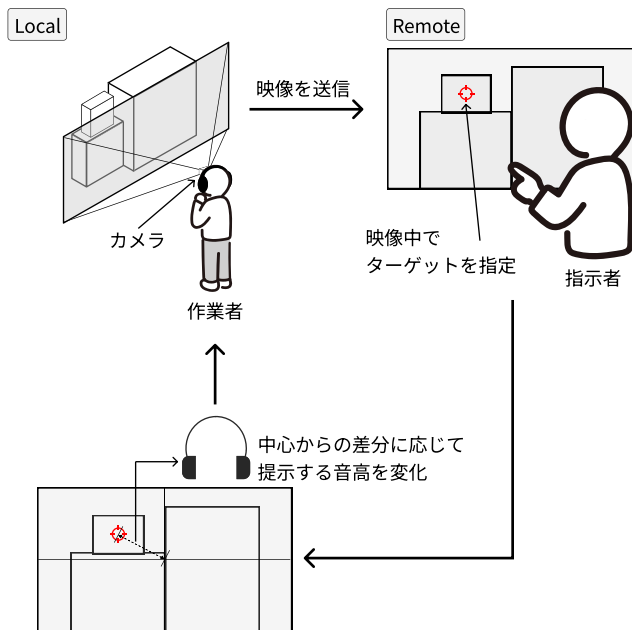


図 1 想定する作業環境およびシステムの挙動

る音源位置変化の複手法の 3 手法を実装した。

2. 関連研究

2.1 視覚情報を用いた注意誘導

遠隔作業指示において、特定の方向に注意を誘導するための手法として、AR を用いて現地環境に視覚情報を付与し注意誘導を行う手法が提案されている。酒田らは、現地側作業者の肩にカメラおよびレーザーポインタを装着し、遠隔指示者がカメラ映像から現地の様子を確認してレーザーポインタで視線を誘導するシステムを設計した [9]。Gauglitz らのシステムでは、遠隔指示者がタブレット端末で撮影された現地側映像に指示のための図形を描画、現地側作業者はタブレット端末越しに現地側の空間に AR 表示された指示を確認することが出来る [3]。Kangas らのシステムでは、複数のカメラおよび HMD を用いて、遠隔支持者が現地の作業環境を 360° 映像で確認出来る [6]。大概らのシステムでは、深度カメラを用いて現地側環境の 3D モデルを作成し遠隔指示者に VR 提示、遠隔指示者は VR 空間内で現地側環境 3D モデルに対する指差し等の動作を行う。遠隔支持者の動作は 3D アバタの動作として現地側環境に重畳され、現地側作業者は HMD 越しにそれを確認することが出来る [15]。Amores らは、遠隔指示者が現地側作業者の一人称視点を見ながら、ハンドジェスチャを現地側作業者に共有するシステムを作成した [1]。これらのシステムでは視覚的情報により指示位置が一意に定まる一方で、タブレット端末や HMD 等の機材が必要になること、タブレットの把持や HMD の着用が作業の妨げになる等の制約があった。

2.2 空間音響を用いた注意誘導

視覚的情報を用いて注意誘導を行う手法の制約（必要機材および作業性の低下）に対して、空間音響を用いて注意誘導を行う手法が提案されている。聴覚情報を用いた注意誘導では、頭部伝達関数を用いた空間音響によって疑似的に特定の方向から音が到来する状況を再現する手法が用いられる。広田らは頭部伝達関数を用いた空間音響により空間中に聴覚的アイコンを配置し、ポインタを用いたポインティングにより選択を行うインタフェースを提案した [13]。Majdak らは利用者ごとに頭部伝達関数を測定し、空間音響を用いて音源の位置特定タスクを行った。その結果、頭部方向を用いたポインティングは手を用いたポインティングと同等のポインティング能力があることが示された [8]。また、Chung らは、視覚障害者向けの音を利用したナビゲーションにおいて、空間音響が有効であることを示した [2]。空間音響を用いた注意誘導は、ヘッドフォンを中心とする簡易的な機材で実現可能であり、かつハンズ／アイズフリーでの注意誘導が可能になる。ただし、利用者毎に頭部伝達関数を測定することは困難であり、事前に記録された汎用的な頭部伝達関数では精度が低くなるという課題があった。

2.3 本研究の指針

視覚情報を用いた注意誘導、および空間音響を用いた注意誘導の課題を解決するため、本研究では空間音響を用いずに利用者の頭部方向と目標角度の差分に応じて動的に変化する非空間音響を用いて注意誘導を行う手法を提案する。提案手法では、カメラを用いて利用者の頭部方向を取得し、目標角度との誤差に応じてヘッドフォンから提示する聴覚情報を動的に変化させる。動的に変化させる聴覚情報のパラメータとして、本研究では音高（周波数）を用いる。これは、音高は他の聴覚情報のパラメータ（音圧等）と比較して検知閾が小さく、約 0.2% の変化を知覚可能である [14] ためである。

我々が過去に行った研究 [16] では音高を動的に変化させ注意誘導を行う際、音高を変化させる性状において、比例的に変化させる性状、反比例的に変化させる性状、および対数的に変化させる性状の 3 性状間に顕著な差は見られなかった。そのため、人間の聴覚が音の変化を対数的に知覚する [5] ことから、本研究では、音高を対数的に変化させる性状を用いた。また、先行研究では Optitrack V120:Duo を用いて頭部方向情報を検出し、目標角度と頭部方向の水平方向の差分に応じて音高を変化させていた。本研究では、実際の遠隔指示を想定しカメラ映像の特徴点によって水平および垂直方向の頭部方向を認識するシステムを実装した。

3. 提案システム

本研究では、カメラを用いた特徴点追跡によってカメラ画角内の特定の位置をターゲットとして設定し、ターゲットとカメラの角度差に応じて音高を動的に変化させることにより、ユーザの頭部方向をターゲットに誘導するシステムを実装した。本章では、実装したシステムについて述べる。なお、今後注意誘導の影響調査のための実験を行うことを想定して、提案手法だけでなく、空間音響を用いる手法および提案手法と空間音響を組み合わせた手法も実装した。

3.1 ハードウェア



図 2 ハードウェア構成

本システムのハードウェア構成を図 2 に示す。本システムではカメラを密閉型ヘッドホンに装着したデバイスを用いた。カメラには Luxonis 社の OAK-D Pro W [10] を用いた。OAK-D Pro W は広視野角の高解像度カラーカメラやアクティブステレオ深度カメラ、夜間視認などの機能を有する AI カメラであり、本システムにおいては特徴点追跡の機能を用いた。カメラの画角は水平 127 度、垂直 79.5 度であり、この範囲内でターゲットを指定し追跡することができる。ヘッドホンには Logicool 社の PRO-X [7] を用いた。またヘッドホンに OAK-D Pro を固定するためのフレームは独自に設計したものを 3D プリントにて出力し使用した。システムの制御にはノートパソコン (mouse, G-Tune H5-TGLBBW11, CPU: Core i7 11800H, メモリ: 64GB, GPU: NVIDIA GeForce RTX 3070 Laptop GPU, OS: Microsoft Windows 11 Pro 10.0.22631) を用いた。

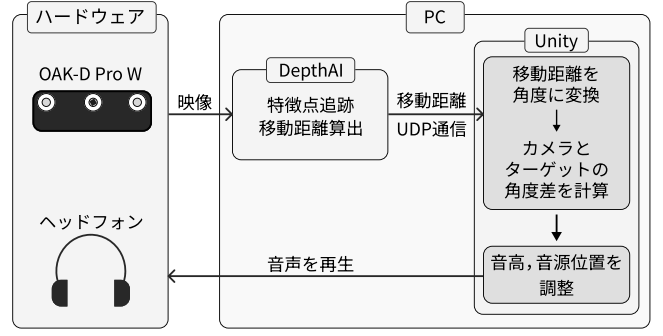


図 3 ソフトウェア構成

3.2 ソフトウェア

本システムのソフトウェア構成を図 3 に示す。ソフトウェアシステムは Python と Unity[11] を用いて実装した。OAK-D の左モノクロカメラから得られたフレームを受信し、DepthAI を用いて 50 個の特徴点を抽出および追跡する。ここで、1 フレームごとに 50 個の特徴点の移動距離の平均を取り、カメラの移動距離とする。このカメラの移動距離を UDP 通信で Unity に送信する。Unity ではまず、受け取ったカメラの移動距離を角度に変換する。カメラの移動範囲が水平方向に 1280 [px]、垂直方向に 800 [px] であるため、得られた値を 10 で割ることで角度とした。そして、常にターゲットとカメラの角度差を計算し、水平方向の角度差と垂直方向の角度差の二乗和平方根 (以降, ang) によって変化させた音を提示する。音の提示に際して、基本音源として常に 1000 [Hz] の sin 波の音源ファイルを Unity 内でループ再生する。再生音高の設定は Unity の AudioSource.pitch を調節することで実装し、(再生音高) = 1000 × pitch とした。また, ang が 4.5 度未満の位置に 1 秒以上滞留した際に正解音 (pitch = 2.5) が鳴る設定とした。提案手法では、ターゲットとカメラの角度差が変化すると音高を変化させ、ang が小さくなるほどに音が高くなるような実装を行った。音高 pitch は, ang が [0, 150] の範囲で式 (1) に基づいて変化した。

$$\text{pitch} = \begin{cases} 2.5 & (0 \leq \text{ang} < 4.5) \\ -0.43 \log \left(\frac{\text{ang} - 4.5}{175} \right) + 0.01 & (4.5 \leq \text{ang} \leq 150) \end{cases} \quad (1)$$

2.3 節にて述べたように、人間の聴覚は音高の変化を対数的に知覚する特性を持つことが知られている。この特性に基づき、音高の変化が自然に感じられるよう、本研究では対数関数を用いて音高を調整する式を設計した。

また、今後の比較実験を想定して提案手法を用いた条件 (SoundPointer 条件) だけでなく、空間音響のみを用いる条件 (Stereo 条件)、空間音響および提案手法を組み合わせた条件 (Stereo + SoundPointer 条件) の 3 条件を、ターゲットとカメラの角度差が小さくなったことを提示する手法として実装を行った。Stereo 条件は、ターゲットと

カメラの角度差が変化するとともに空間音響を用いてターゲットの位置を提示するシステムを実装した。空間音響の実装には、Unity のオーディオ機能に標準搭載されている機能を利用した。Unity では、音源の持つパラメータのうち AudioSource.spatialBlend の値を 1.0 とすることで、音源の空間音響化が可能となる。この Unity 標準機能を用いて、Stereo 条件の空間音響を実装した。また、この時ターゲットとカメラの角度差による音高の変化は発生しない。Stereo + SoundPointer 条件は、先の Stereo 条件における音源の位置の変化と SoundPointer 条件における音高の変化をともに取り入れた実装を行った。

これに加えて、実験実施のための機能として、ターゲットが、 $x[-45, 45]$, $y[-30, 30]$ の範囲に 15 度刻みでランダムに生成されるようにした。この時、長時間正解を得られなかった場合に次のターゲットに進み、正解を得られなかったターゲットを最後の試行の後に追加する機能を実装した。これは、システムの設計上時間経過につれてターゲットおよびカメラ方向に誤差が蓄積してしまう可能性があるためである。

4. おわりに

本研究では、カメラ映像の特徴点を追跡して、利用者の頭部方向および目標角度の差分に応じて動的に音高を変化させる注意誘導手法を提案した。本稿では、提案手法を検証するためシステムのプロトタイプを実装した。今後は提案手法が利用者の注意誘導に与える影響を調査するための実験を行う。実験は提案手法、音を用いた既存の注意誘導手法である空間音響、提案手法および空間音響による音源位置変化の複合手法の 3 手法を比較し、利用者の頭部方向が目標角度と一致するまでの時間への影響、およびユーザビリティについて調査する。

謝辞 本研究は JSPS 科研費 20K19839 の助成を受けた。

参考文献

- [1] Amores, J., Benavides, X. and Maes, P.: ShowMe: A Remote Collaboration System that Supports Immersive Gestural Communication, *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, New York, NY, USA, Association for Computing Machinery, p. 1343–1348 (online), DOI: 10.1145/2702613.2732927 (2015).
- [2] Chung, S., Lee, K. and Oh, U.: Investigating Three-dimensional Directional Guidance with Nonvisual Feedback for Target Pointing Task, *2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, Los Alamitos, CA, USA, IEEE Computer Society, pp. 206–210 (online), DOI: 10.1109/ISMAR-Adjunct51615.2020.00061 (2020).
- [3] Gauglitz, S., Nuernberger, B., Turk, M. and Höllerer, T.: World-Stabilized Annotations and Virtual Scene Navigation for Remote Collaboration, *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, New York, NY, USA, Association for Computing Machinery, p. 449–459 (online), DOI: 10.1145/2642918.2647372 (2014).
- [4] Holland, S., Morse, D. R. and Gedenryd, H.: AudioGPS: Spatial Audio Navigation with a Minimal Attention Interface, *Personal Ubiquitous Comput.*, Vol. 6, No. 4, p. 253–259 (online), DOI: 10.1007/s007790200025 (2002).
- [5] Inui, K., Urakawa, T., Yamashiro, K., Otsuru, N., Nishihara, M., Takeshima, Y., Keceli, S. and Kakigi, R.: Non-linear laws of echoic memory and auditory change detection in humans, *BMC Neuroscience*, Vol. 11, No. 1, p. 80 (online), DOI: 10.1186/1471-2202-11-80 (2010).
- [6] Kangas, J., Sand, A., Jokela, T., Piippo, P., Eskolin, P., Salmimaa, M. and Raisamo, R.: Remote Expert for Assistance in a Physical Operational Task, *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, New York, NY, USA, Association for Computing Machinery, p. 1–6 (online), DOI: 10.1145/3170427.3188598 (2018).
- [7] Logicool: Blue VO!CE マイク テクノロジー搭載 Logicool PRO X ゲーミング ヘッドセット. <https://gaming.logicool.co.jp/ja-jp/products/gaming-audio/pro-x-gaming-headset-blue-voice-mic-tech.981-001297.html> (Accessed on 12/16/2024).
- [8] Majdak, P., Goupell, M. J. and Laback, B.: 3-D localization of virtual sound sources: Effects of visual environment, pointing method, and training, *Attention, Perception, & Psychophysics*, Vol. 72, No. 2, pp. 454–469 (online), DOI: 10.3758/APP.72.2.454 (2010).
- [9] Sakata, N., Kurata, T., Kato, T., Kourogi, M. and Kuzuoka, H.: WACL: supporting telecommunications using - wearable active camera with laser pointer, *Seventh IEEE International Symposium on Wearable Computers, 2003. Proceedings.*, pp. 53–56 (online), DOI: 10.1109/ISWC.2003.1241393 (2003).
- [10] The4: OAK-D Pro W – Luxonis. <https://shop.luxonis.com/products/oak-d-pro-w?srsltid=AfmB0oqlDmdipejZQinY2lKR1-jgyrg89V8zq-YHlEeFqs8SpyHjqbmY> (Accessed on 12/16/2024).
- [11] Unity Technologies: Unity Real-Time Development Platform - 3D, 3D, VR & AR Engine (2024). <https://unity.com> (Accessed on 12/24/2024).
- [12] 巻瀧有哉, 小林達也, 加藤晴久, 柳原広昌: HMD キャリブレーションとオンサイト学習によるハンズフリー遠隔作業支援 AR システム, 情報処理学会研究報告, No. 2 (2015).
- [13] 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. 156–165 (2003).
- [14] 大串健吾: 聴覚心理学概論 (1995).
- [15] 大槻麻衣, 山口央士朗, 今田昇吾, 葛岡英明, 鈴木雄介: 遠隔作業指示における身体動作と視線提示の効果, 情報処理学会研究報告, No. 5 (2018).
- [16] 飯塚陸斗, 大山智弘, 杉山将利, 下田康太, 田之頭吾音, 井原圭一, 石橋遼, 川口一画: 動的音高変化による方向提示手法設計のための基礎的調査, インタラクション 2023 論文集, 情報処理学会, pp. 486–490 (2023). <https://www.interaction-ipsj.org/proceedings/2023/data/pdf/2A-08.pdf>.