



注視点と視線の交点の深度差を用いた VR 空間におけるユーザの言語負荷の推定

Linguistic Load Estimation in VR Using Depth Difference between Gazed Point and Gaze Depth

シルム ヨハネス¹⁾, ペルスキアエルナンデス モニカ¹⁾, 磯山 直也¹⁾, 内山 英昭¹⁾, 清川 清¹⁾

Johannes SCHIRM, Monica PERUSQUIA-HERNANDEZ,

Naoya ISOYAMA, Hideaki UCHIYAMA, and Kiyoshi KIYOKAWA

1) 奈良先端科学技術大学院大学 先端科学技術研究科 (〒 630-0192 奈良県生駒市高山町 8916-5)

概要: 視線計測が可能な VR 用ヘッドマウントディスプレイが普及し、VR を体験中の視線データを用いてユーザの内部状態を推定する研究が盛んになっている。本研究では、注視点と視線の交点の深度差を用いて VR を体験中のユーザの言語負荷の推定を試みた。具体的には、VR を体験中に難易度の異なる音声を聞き取らせ、両眼の視線の交点が注視対象の物体表面からどの程度前後にずれるか (focus offset) を調査した。その結果、音声の主観的難易度と瞳孔径や focus offset の間に相関は見られなかった。しかし、一部の条件間に有意差が確認され、瞳孔径に加えて focus offset が実際の言語負荷と何らかの関係があることが示唆された。

キーワード: 視線情報, 認知負荷, バーチャルリアリティ

1. はじめに

本研究では、眼から得られる情報を用いてバーチャルリアリティ (VR) を体験中のユーザの認知負荷を推定することを目的としている。視線や瞳孔径、瞬目の頻度など、眼から得られる情報は人間の心理状態をある程度反映することが知られている。例えば、認知負荷が高まるにつれて瞳孔径が大きくなることはよく知られている。しかし、瞳孔径は入射する光量や年齢、集中度など様々な要因で変化するため、VR を体験中のユーザの認知負荷を安定して推測する目的には必ずしも適していない。

本稿では、注視点の深度に着目し、これを用いて認知負荷の推定を試みた評価実験について報告する。一般に VR のユーザは様々な三次元物体が配置された VR 環境に没入し、常に何かを注視して作業することが多い。我々は、認知負荷が低ければユーザは物体表面を正確に注視し、認知負荷が高ければ注視が不安定となりその前後で視線が変わると考えた。すなわち、両眼の視線の交点 (正確には各視線の最近傍点の midpoint) と注視点 (各視線と物体表面の交点の midpoint) の深度差 (focus offset) やそのバラツキが認知負荷の指標になると想定した。VR 用ヘッドマウントディスプレイ (HMD) を用いて視線情報を調査した研究例は急速に増えつつあるが [3], デスクトップモニタを用いた場合に比べてまだ報告が少なく、特に focus offset に関する調査はほとんどなされていない。

そこで、認知負荷以外の影響で focus offset が変化することを抑制するために、VR 環境中で聞こえてくる音声の内

容を理解させる聞き取りタスクを設定し、言語負荷が focus offset に与える影響を調査した。実験では、聞き取りタスクとして (日・英)×(難・易) の 4 通りの言語負荷の音声を扱い、11 人の参加者から得た視線情報を統計的に分析した。VR 用 HMD には視線計測が可能な HTC VIVE Pro Eye を用いた。その結果、一部の条件間に有意差が確認され、focus offset が言語負荷と何らかの関係がある可能性が示唆された。有意差が一部の条件間にしか見られなかった要因として、視線計測の校正精度が参加者ごとに大きく異なっていたことが推察された。以下、関連研究、実験の詳細、および結果と考察について述べる。

2. 関連研究

これまでに視線や注視点に関する様々な研究がなされており、いくつかの異なる指標が用いられている。eye vergence は最も一般的であり、両眼の視線のなす角度 (輻輳角) を表す [2]。Francis らは、考え事をしていない際の輻輳角は、その人にとって最も楽な視線の向きによって決定されると報告している [1]。gaze depth は眼から視線の交点までの距離を表す [8]。

Walcher らは、両眼の視線が注視する物体表面と交差する 2 つの交点を考え、それらの距離を fixation disparity と呼んでいる [7]。そして、fixation disparity は、瞬目の頻度や瞳孔径と並んで内発的注意と外発的注意の識別に有効であったと報告している。fixation disparity は、我々の focus offset と全く同様に、視線が物体表面上で交差すればゼロ

に、物体表面よりも奥で交差すれば正に、手前で交差すれば負になる。ただし、focus offset の場合は両眼の視線が平行に近くなるにつれて急激に増大するため、奥で交差することに対してより敏感な指標と言える。

また、坂本らは VR 環境における読書内容の主観的難易度を推定する研究において、我々の focus offset と全く同じ意味で「注視点と視線の交点の距離」に着目している。その結果、ユーザ依存の機械学習モデルの場合この指標を含めた方がそうでないものよりも推定精度が高く、ユーザ非依存の場合は逆の結果となったことを報告している [9]。

3. 実験

3.1 概要

本実験では、VR 環境中で聞こえてくる音声の内容を理解させる聞き取りタスクを用いて、言語負荷が focus offset に与える影響を調査した。実験に用いた VR 環境には図 1 に示す電話ボックスを配置した。実験開始時の初期位置から電話ボックスまでは約 2.6 m 離れており、参加者は物理的に歩いて電話ボックスに入り後述する 4 種類のいずれかの音声を聞く。電話ボックスはおよそ内径 1.0 m、高さ 1.95 m であり、物体への注視を促すために床から 1.5 m の位置の緑のランプが、見回しを促すために床から 1.2~1.4 m の範囲の壁面に数枚のポスターが配置されている。注視をさらに促すために音声が再生される間は緑のランプが点滅する。以上のレイアウトは、注視距離が安定するとされる 2~3 m の範囲を考慮して設定している [3]。

聞き取りタスクの音声には、以下に示す (日・英)×(難・易) の 4 種を用いた。

- JE (日・易) 教員による大学同窓会の趣旨説明
- JD (日・難) SF アニメの文学的な独話
- EE (英・易) 天気に関する簡単な会話
- ED (英・難) 満足遅延によるスキル習得とゲームの比較

音声の長さは条件によって約 45~63 秒であり、後述するように結果の解析には最初の 40 秒間のみを使用した。また、電話ボックスから聞こえる音声のリアリティを高め、かつ、聞き取りが簡単になりすぎないように、200~2000 Hz の周波数のみを使用した。

実験システムは Unity 2019.4.13f1 (LTS) で開発し、制御用 PC (OS: Windows 10 (1909), CPU: Intel Core i7-10700, GPU: GeForce GTX 1660 (6 GB), RAM: 16 GB (PC4/DDR4)) と視線計測可能な VR 用 HMD (HTC VIVE Pro Eye) を用いて実装した。

3.2 実験手順

実験は研究室内の静かな部屋で下記の実験手順に従って実施された。各参加者は音声の内容を理解すること以外に指示は与えられず、音声を聞いている間に自由に周囲を見回すことができた。

1. PC モニタ上で実験内容の説明書を読んだ後、同意書に同意し事前アンケートに回答する。
2. HMD を装着する。



図 1: 実験用 VR 環境。初期位置から 3m 弱歩いて電話ボックスの中に入ると受話器付近から様々な音声が 3D サウンドとして再生される。再生中は緑のランプが点滅する。

3. 視線計測のキャリブレーションや音量の調節などの初期設定を行う。
4. 電話ボックス内まで歩く。
5. 自動的に再生される音声を最後まで聞く。
6. HMD を外し、PC モニタ上で事後アンケートに回答する。
7. 再び HMD を装着し、各条件で手順 4~6 を繰り返す。

3.3 評価項目

事前アンケートでは性別や年齢を回答させたが、解析には用いなかった。条件ごとの事後アンケートでは、主観評価値として音声の主観的難易度を 5 段階のリッカート尺度で回答させた (1: 全く理解できなかった, 2: 少し理解できた, 3: 半分理解できた, 4: ほとんど理解できた, 5: ほぼ完全に理解できた)。また、客観的指標として各聞き取りタスク中の瞳孔径と focus offset を記録した。

既に述べたように、focus offset は両眼の視線の交点と注視点の深度差であり、正確には「各視線の最近傍点の midpoint」と「各視線と物体表面の交点の midpoint」の深度差として算出する。さらに、注視距離によるバラツキの影響を抑制するために、この深度差を眼から注視点までの距離で除した。電話ボックスの緑のランプを注視する場合の focus offset の例を図 2 に示す。図 2(a) のように正確に注視対象を見ている場合は focus offset はゼロになる。図 2(b) のように注視対象よりも奥で視線が交差する場合は focus offset は正になり、図 2(c) のように注視対象よりも手前で視線が交差する場合は focus offset は負になる。

3.4 仮説

本実験で検証する仮説を以下に示す。H1 は先行研究と同様の結果を期待して設定したものであり、本実験の主たる関心は H2 にある。

- H1 聞き取りタスクの主観的難易度が高くなるほど、瞳孔径やそのバラツキが大きくなる。
- H2 聞き取りタスクの主観的難易度が高くなるほど、focus offset やそのバラツキが大きくなる。

4. 実験結果

実験参加者 実験には情報科学を専攻する日本人の大学院生 11 人 (男性 10 人, 女性 1 人; 平均年齢: 22.8 歳, SD: 0.8)

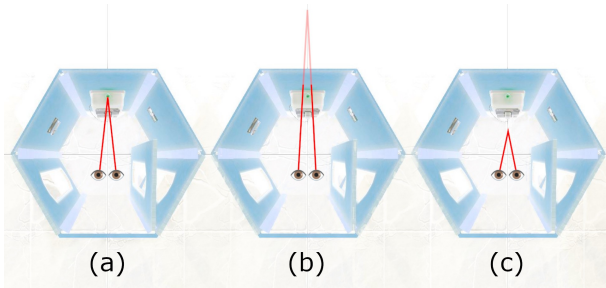


図 2: 電話ボックスのランプを注視する例。(a) 一般に、物体を注視する場合 focus offset がほぼゼロになる。(b) 認知負荷が高い場合は focus offset が正となる(視線が注視対象よりも奥で交差する)頻度が高まると予想した。(c) 個人差や校正誤差などにより focus offset が負となる場合もあると予想した。

が参加した。全員著者の所属する研究室の学生であり、報酬として菓子を配った。

主観的難易度 4種の音声の主観的難易度について、間隔尺度とみなして算出した平均値と標準偏差はそれぞれ、JE: 4.9 (SD: 0.3), JD: 3.1 (SD: 1.5), EE: 3.6 (SD: 0.9), ED: 1.6 (SD: 0.7) となった。

計測データの事前処理 90 Hz で記録した計測データから R の GazePath パッケージ [4] を用いて瞬目やサッカードのデータを除去し、生じた欠損部を線形補間で補完した。瞳孔径は音声条件ごとに最初の 2 秒のデータの平均値を減じて正規化し、平滑化した。また、フレームごとの両眼の瞳孔径の平均を最終的な瞳孔径として算出した。一方、focus offset はフレームごとに算出し、窓サイズ 201 の Savitzky-Golay フィルタで平滑化し、参加者ごとに中央値を用いて正規化した。

以上の定量的データを統計的に分析するために、データを 10 秒単位に分割し、瞳孔径と focus offset のそれぞれについて平均値と分散値を算出した。各音声条件ごとの 40 秒間のデータからそれぞれ 4 つの値が得られるため、11 人の参加者に対して各音声条件ごとにそれぞれ 44 個の値が得られた。

瞳孔径 図 3 に瞳孔径の平均値を示す。Kruskal-Wallis 検定の結果、瞳孔径の平均値に有意差 ($p < 0.01$) が確認された。Wilcoxon の順位和検定と BH 法によって下位検定を実施したところ、EE と JE ($p < 0.05$)、EE と JD ($p < 0.01$)、EE と ED ($p < 0.01$) の間にそれぞれ有意差が確認された。EE では JE, JD, ED の場合に比べて瞳孔径の平均値がそれぞれ 0.072 mm, 0.095 mm, 0.103 mm 大きかった。しかし、主観的難易度と瞳孔径の平均値の間の Spearman の順位相関係数は $\rho = 0.09$ で相関は見られなかった。

図 4 に瞳孔径の分散値を示す。Kruskal-Wallis 検定の結果、瞳孔径の分散値には有意差 ($p < 0.01$) が確認された。Wilcoxon の順位和検定と BH 法によって下位検定を実施し

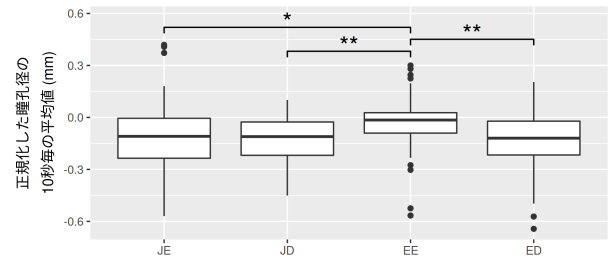


図 3: 正規化した瞳孔径の 10 秒ごとの平均 (mm)。

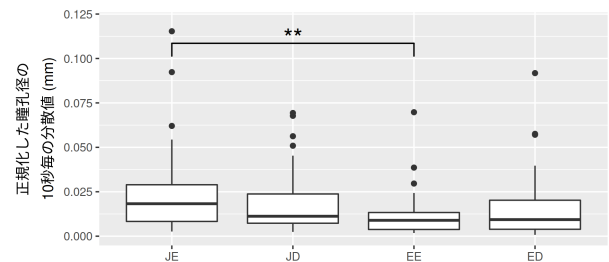


図 4: 正規化した瞳孔径の 10 秒ごとの分散 (mm)。

たところ、EE と JE の間 ($p < 0.01$) に有意差が確認された。JE では EE の場合に比べて瞳孔系の分散が 2.5 倍大きかった。しかし、主観的難易度と瞳孔径の分散値の間の Spearman の順位相関係数は $\rho = -0.13$ で相関は見られなかった。

Focus Offset 図 5 に focus offset の平均値を示す。Kruskal-Wallis 検定の結果、focus offset の平均値に有意差 ($p < 0.01$) が確認された。Wilcoxon の順位和検定と BH 法によって下位検定を実施したところ、JD と EE ($p < 0.01$) および JD と ED ($p < 0.05$) の間にそれぞれ有意差が確認された。EE と ED では、JD の場合に比べて focus offset が平均してそれぞれ 7.0% と 6.5% 大きかった。しかし、主観的難易度と focus offset の平均値の間の Spearman の順位相関係数は $\rho = 0.04$ で相関は見られなかった。一方、Kruskal-Wallis 検定の結果、focus offset の分散値に有意差は見られなかった ($p = 0.056$)。

5. 考察

実験結果からは、一部の条件間で興味深い有意差が見られたものの、全体として H1 と H2 はいずれも支持されな

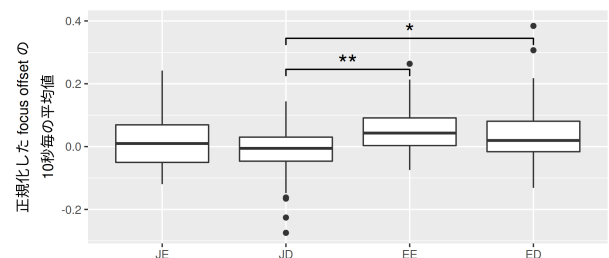


図 5: 正規化した focus offset の 10 秒ごとの平均。

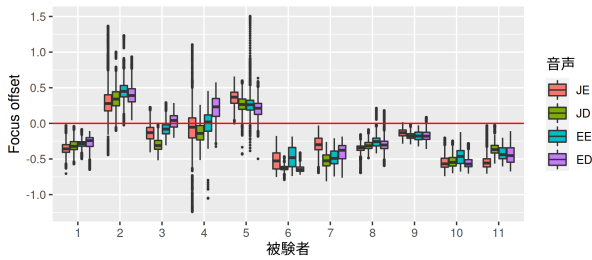


図 6: 正規化する前の各参加者の focus offset.

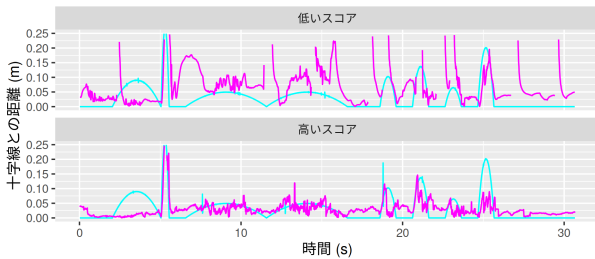


図 7: 2名の参加者の指標追跡タスクのパフォーマンス。水色は指標の移動速度、紫は指標と注視点の距離。参加者によってパフォーマンスが著しく異なる。

かったと言える。その理由として、主観的難易度と実際の言語負荷レベルが必ずしも正の相関になかったことが考えられる。参加者の多くは、EEは努力すれば内容が理解できる難易度であると感じており、EDは努力しても理解できない難易度である（ため理解を諦めた）と感じていた。従って、実際の言語負荷レベルはEEが最も高く、EDはそれより低かったことが考えられる。

実際、瞳孔径の平均値ではEEとそれ以外の3条件間にそれぞれ有意差が見られ、EEの場合に瞳孔径が際立って大きかったことが示されている。また瞳孔径の分散についてもEEではJEの場合よりも瞳孔径の変化が少なかったことが示されている。

一方、focus offsetの平均値について、瞳孔径と同様に4条件の中でEEが最も高い平均値となっているが、有意差が見られたのはJDとEE、JDとEDの間のみであった。従って、focus offsetは言語負荷の指標となる可能性はあるが、さらなる調査が必要であると考えられる。

このような結果となった要因として、視線計測の校正精度の個人差が大きく、focus offsetの算出が不正確でバラツキも大きかったことが考えられる。図6に参加者ごとの正規化をする前の各参加者のfocus offsetを示す。参加者ごとに平均値が大きく異なることや、一部の参加者と音声条件でバラツキが極端に大きいことが見て取れる。本実験では、視線計測の精度を確認するために、画面上を動く十字図形を注視し続ける指標追跡タスクを校正直後に設けていた。図7に示すように、指標追跡タスクのパフォーマンスは参加者によって著しい差があった。今後は、校正精度を高める工夫や校正精度の高いデータのみを用いて解析する必要があると考えられる。

6. おわりに

本稿では、注視点と視線の交点の深度差 (focus offset) を用いてVRを体験中のユーザの言語理解時の認知負荷の推定を試みた評価実験について報告した。実験結果からは、瞳孔径の平均や分散に加えてfocus offsetについても聞き取りタスクの主観的難易度との相関は見られなかった。しかし、いくつかの音声条件間でfocus offsetに有意差が見られたため、実際の言語負荷レベルと何らかの関係がある可能性はある。今後は、一部の参加者で精度が低かった視線計測の校正方法などの実験設定を見直し、参加者数を増やした上で引き続き調査を行う予定である。

参考文献

- [1] Ellie L. Francis, Bai-Chuan Jiang, D. Alfred Owens, and Richard A. Tyrrell: Accommodation and Vergence Require Effort-to-See, *Optometry and Vision Science*, Vol. 80, No. 6, pp. 467–473, 2003.
- [2] Michael Xuelin Huang, Jiajia Li, Grace Ngai, Hong Va Leong, and Andreas Bulling: Moment-to-Moment Detection of Internal Thought during Video Viewing from Eye Vergence Behavior, *ACM International Conference on Multimedia*, pp. 2254–2262, 2019.
- [3] Julie Iskander, Mohammed Hossny, and Saeid Nahavandi: Using biomechanics to investigate the effect of VR on eye vergence system, *Applied Ergonomics*, Vol. 45, No. 102883, 2019.
- [4] Daan R. van Renswoude, Maartje E. J. Raijmakers, Arnout Koornneef, Scott P. Johnson, Sabine Hunnius, and Ingmar Visser: Gazepath: An eye-tracking analysis tool that accounts for individual differences and data quality, *Behavior Research Methods*, Vol. 50, No. 2, 2018.
- [5] Takumi Toyama, Jason Orlosky, Daniel Sonntag, and Kiyoshi Kiyokawa: A Natural Interface for Multi-Focal Plane Head Mounted Displays Using 3D Gaze, *International Working Conference on Advanced Visual Interfaces*, pp. 25–32, 2014.
- [6] Lisa-Marie Vortmann, Jannes Knychalla, Sonja Annerer-Walcher, Mathias Benedek, and Felix Putze: Imaging Time Series of Eye Tracking Data to Classify Attentional States, *Frontiers in Neuroscience*, Vol. 15, No. 664490, 2021.
- [7] Sonja Walcher, Simon M. Ceh, Felix Putze, Marvin Kampen, Christof Körner, and Mathias Benedek: How Reliably Do Eye Parameters Indicate Internal Versus External Attentional Focus?, *Cognitive Science*, Vol. 45, No. 4, 2021.
- [8] Martin Weier, Thorsten Roth, André Hinkenjann, and Slusallek Philipp: Predicting the Gaze Depth in Head-Mounted Displays Using Multiple Feature Regression, *ACM Symposium on Eye Tracking Research & Applications*, No. 19, pp. 1–9, 2018.
- [9] 坂本 賢哉, 白井 詩沙香, 武村 紀子, オーロス キーソン, 長瀧 寛之, 上田 真由美, 浦西 友樹, 竹村 治雄: 視線情報に基づくVR空間でのマンガ教材読書時の主観的難易度推定, 日本バーチャルリアリティ学会 複合現実感研究会, Vol. 19, No. 2, pp. 51–56, 2021.