第 24 回日本バーチャルリアリティ学会大会論文集（2019 年 9 月）

VR 環境下における方向認識の視聴覚統合

Integration of visual and auditory signals for spatial localization under VR environment

高培鐘 [1)]，松本啓吾 [1)]，鳴海拓志 [1)]，谷川智洋 [1)]，廣瀬通孝 [1)]
Peizhong GAO, Keigo MATSUMOTO, Takuji NARUMI, Tomohiro TANIKAWA and Michitaka HIROSE

1) 東京大学 大学院情報理工学系研究科（〒 113-8656 東京都文京区本郷 7-3-1, {kou, matsumoto, narumi, tani, hirose}@cyber.t.u-tokyo.ac.jp ）

概要: 方向認識のタスクにおいて違う場所を示す視聴覚情報が提示された時のヒトの認知モデルには最尤推定モデル (maximum-likelihood estimation) が知られている。本研究では、先行研究で使用されたランダム・ドット・ステレオグラムの代わりに VR 空間内におけるオブジェクトを視覚刺激として使用する時のヒトの視聴覚統合特性について検討する。

キーワード： 視聴覚統合，方向認識，最尤推定モデル

1. はじめに

本研究は方向認識のタスクに注目し，VR 環境下における視聴覚統合の性質を最尤推定モデル (Maximum-likelihood estimation：MLE) と照らし合わして検討する．複数の感覚を用いて世界を認識するヒトの感覚統合特性は多く研究されてきた．一つの事象に対して感覚同士がお互い一致する場合の感覚統合がほとんどであるが，その中には感覚同士が違う情報を示す時のヒトの認識モデルも多数提案されている．その一つに最尤推定モデル (MLE) と呼ばれるヒトの各感覚にそれぞれ独立した認識結果を有し，感覚ごとの「信頼度」に従い統合された認識結果に各々が占める重みが決まるというモデルが知られている [1]．物の形や大きさを認識する際の視触覚統合は MLE に当てはまることが報告されている [1, 2]．方向認識のタスクにおける視聴覚統合にも同じ傾向が見られると言われている [3, 4]．本研究では VR 下での人の感覚統合モデルを方向認識タスクを用いることで検証する．

2. 背景

2.1 関連研究

Battaglia らの研究においては，RDS を視覚刺激として使用している [3]．RDS は立体視図形の一つで，左右の目用に二枚の画像があり，目の焦点をうまく合わせると立体が浮かび上がる画像である [5]．図 2(a) は本研究が実装した R 左目用の RDS の画像例である．しかし，RDS は奥行き情報の面で実際の 3D オブジェクトと差があることが報告されている [6]．次に，Helbig らの研究においては実世界の物体を視覚刺激として用いて実験を行なっていた [2]．VR における視覚情報はフレームレートや解像度の面で実世界と違うことが考えられる．以上のことから，VR 環境下における 3D オブジェクトを視覚刺激として提示した時の方向認識の視聴覚統合特性を調べる必要がある．

2.2 最尤推定モデル (MLE)

ヒトの脳は複数の感覚情報から最も信頼度の高い推測結果を得ようとすると考えられており，そのモデルは多くの研究者によって提案されてきた．ノイズの大きな環境において脳が各独立する感覚から認識する情報（以下は単に感覚情報と呼ぶ）は確率モデルで表され，特に正規分布に従うと考える場合の感覚統合モデルは最尤推定モデル (MLE) と呼ばれる．MLE では，各感覚情報の信頼度はその感覚が従う正規分布の分散によって決まり，感覚情報の分散が小さい場合は信頼度が高いと考えられている．さらに，感覚統合時の認識結果（以下は統合結果と呼ぶ）に各感覚情報が占める割合は各々の感覚情報の信頼度によって決まり，信頼度が高い感覚には大きい重みが割り振られる．

MLE は式 1 のように表される．$\hat{R}$ は統合結果が従う正規分布の期待値を表し，$\bar{R}_i$ は i 番目の感覚情報が従う正規分布の期待値であり，w_i はそれに対応する重みである．$\hat{\sigma}^2$ は統合結果の分散を意味し，その逆数が各感覚情報の分散の逆数の和で表される．

$$\hat{R} = \sum_i^n w_i \bar{R}_i \quad \text{and} \quad \frac{1}{\hat{\sigma}^2} = \sum_i^n \frac{1}{\sigma_i^2} \tag{1}$$

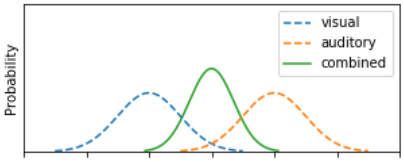
各感覚が占める重み w_i は式 2 のように，各々の感覚情報の分散 σ_i^2 の逆数が統合結果の分散の逆数に占める割合で計算できる．

$$w_i = \frac{1/\sigma_i^2}{\sum_j^n 1/\sigma_j^2} \left(= \frac{1/\sigma_i^2}{1/\hat{\sigma}^2} \right) \tag{2}$$

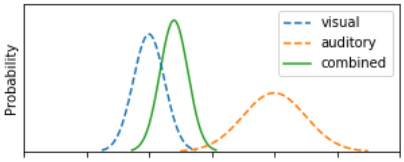
すなわち，統合結果は各々の感覚情報の線型結合で表現できる．さらに，統合結果の分散がいずれの感覚情報の分散よりも小さいことから，脳は複数の感覚を統合することによって認識の信頼度をあげていることになる．

視聴覚統合の場合，つまり $n = 2$ の時の MLE を表しているのが図 1 である．青い破線は視覚情報の確率分布，黄

色い破線は聴覚情報の確率分布，緑の実線は視聴覚統合結果の確率分布を表す．図 1(a) の場合，視覚情報と聴覚情報は同じ分散を持つため，統合結果に各々が占める重みは等しい．しかも統合結果はいずれの独立確率分布よりも小さい分散を持つことになる．また，図 1(b) の場合，視覚情報の分散が小さいため，統合結果にはより大きい重みを占めることになり，統合結果も視覚の方に偏っている．



(a) 独立確率分布が同じ場合



(b) 独立確率分布が異なる場合

図 1: 視聴覚統合における最尤推定モデル

3. 視聴覚統合実験

VR 環境下における 3D オブジェクトを視覚刺激として使う場合のヒトの視聴覚統合特性を MLE モデルを用いて調べる．そのために，視覚刺激として Battaglia らの実験 [3] に使われた RDS の両眼用の画像を本実験で用いたヘッド・マウンティッド・ディスプレイ (HMD) 用に作成する．それと VR 環境下に作成した 3D オブジェクトを視覚刺激として用いる場合の視聴覚統合実験を行い，それらの結果を MLE に照らし合わし比較する．

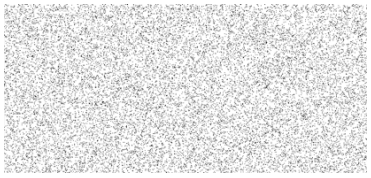
3.1 実験参加者

女性 2 名，男性 2 名（年齢 20 - 30）を対象に実施した．実験参加者の視力または矯正視力は全員正常値だった．いずれの被験者も今回の実験について事前知識はなかった．被験者には実験協力金として学内の規定に基づき，一人あたり 5,000 円のアマゾンギフト券を渡した．

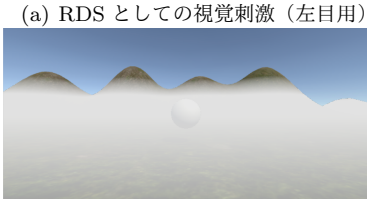
3.2 感覚刺激とシステム構成

実験では聴覚刺激としてホワイトノイズを用いた．Steam® Audio SDK[7] を用いて頭部伝達関数を適応することで聴覚刺激に被験者との相対位置の空間的情報を付与した．音源の位置は全部で七つあり，被験者の真正面に一つと左右に三つずつ配置されている．各々の隣り合う音源同士の間には 1.5° の視野角の差があった．すなわち，{-4.5°，-3°，-1.5°，0°，1.5°，3°，4.5°} の七つの視野角（これを集合 P とする）の位置に音源が配置されていて，いずれか一つの位置から音が出るようになっている．音は VIVE PRO という HMD のヘッドフォンを通じて被験者に提示した．

視覚刺激は二種類用いた．一つ目は Battaglia らが使用した RDS であり（図 2(a)），球体が背景平面から浮かび上がるように見える．球体は P の一つの位置から出現するようになっている．視野の中のドットを一定確率でずらすことによって RDS にノイズが入っており，全部で五つのノイズレベル（確率）がある．それぞれ 10%, 23%, 36%, 49%, 62% である．二つ目は VR 環境中の 3D オブジェクトである（図 2(b)）．この場合の刺激対象も球体であり，P から出現する．VR のシーンにおける五つのノイズは「霧」を使うことで実装した．両方の視覚刺激とも VIVE PRO の両眼ディスプレイを通じて被験者に提示した．両眼ディスプレイはそれぞれ 1440 × 1600 の解像度を持ち，90Hz のフレームレートを有する．



(a) RDS としての視覚刺激（左目用）



(b) 3D オブジェクトとしての視覚刺激

図 2: 二種類の視覚刺激

3.3 実験手順

被験者は方向認識のタスクが与えられて，P から連続して二回現れる刺激対象の左右関係を答えるタスクが課される．その内一回は標準刺激でもう一回は比較刺激であり，現れる順番はランダムである．実験は刺激対象の種類をもって五つに分けられる（表 1）．実験 1 は聴覚の，実験 2 と実験 3 はそれぞれ RDS と 3D オブジェクトを視覚刺激として使う場合の単一感覚実験である．実験 4 と実験 5 はそれぞれ RDS と 3D オブジェクトを視覚刺激として使う場合の統合感覚実験である．実験 1 から実験 3 の標準刺激は 0° の位置から出現し，比較刺激は集合 P から出現する．実験 4 と実験 5 の標準刺激の視覚対象は $-1.5°$ から，聴覚対象は 1.5° から出現し，比較刺激の視覚と聴覚対象は P から同じ位置に出現する．ノイズごとに標準刺激と比較刺激が提示されると 1 試行とみなし，各試行は全部で 15 回繰り返される．よって各被験者で実験 1 は 105 回，それ以外の実験は 525 回行われて，全部で 2205 試行行われる（各実験内の試行順番はランダム）．各試行ごとに被験者には「二回目が一回目の左右のどちらに現れたか」と尋ね，比較刺激の出現する七つの位置ごとに「比較刺激が標準刺激の右に現れた」と回答した確率 p_{right} を算出する．実験は二日に分けて実行した．

表 1: 実験の種類と条件．WN は White Noise，RDS は Random-dot stereogram，3D は VR 環境下の 3D オブジェクトを意味する．P は刺激が出現する位置の集合 {-4.5°，-3°，-1.5°，0°，1.5°，3°，4.5°} を指す．音源の位置は全部で七つ，ノイズの種類は五つある．

実験 No.	実験種類	刺激内容	標準刺激位置	比較刺激位置 (c)	ノイズ数	比較位置数
1	聴覚	WN	0°	$c \in P$	1（なし）	7
2	視覚	RDS	0°	$c \in P$	5	
3		3D	0°	$c \in P$		
4	視聴覚	RDS & WN	$-1.5° \,\&\, 1.5°$	$c_{WN} = c_{RDS}$ かつ $c_{WN}, c_{RDS} \in P$		
5		3D & WN	$-1.5° \,\&\, 1.5°$	$c_{WN} = c_{3D}$ かつ $c_{WN}, c_{3D} \in P$		

3.4 解析方法

結果解析には二つのフェーズがある．フェーズ 1 では，各感覚独自の性質を調べるため視覚か聴覚刺激だけを用いた実験 1 から 3 のデータを用いる．フェーズ 2 では，感覚の統合性質を調べるため視聴覚両方の刺激を用いた実験 4 と 5 のデータを解析する．具体的にはフェーズ 1 では，全部の比較位置の p_{right} に対して累積正規分布関数のフィッティングを行い，確率が 50% に対応する主観的等価点 (PSE) の値と 84% に対応する点と PSE との差（閾値）を算出し，それぞれを感覚情報の期待値と標準偏差とみなす．次に，得られた感覚ごとの正規分布から最尤推定モデル（式 1）に基づき，統合結果の視覚重み $\hat{w}_{pre}$ と標準偏差 $\hat{\sigma}_{pre}$ を予測する．フェーズ 2 では p_{right} を調べることで実験上で得た統合結果の視覚重み $\hat{w}_{emp}$ と標準偏差 $\hat{\sigma}_{emp}$ を算出する．最後に，$\hat{w}_{pre}$ と $\hat{w}_{emp}$，$\hat{\sigma}_{pre}$ と $\hat{\sigma}_{emp}$ を比較検討する．

4. 結果

実験 1 から実験 3 の，視聴覚を独立して提示した場合に得られた p_{right} に累積正規分布をフィッティングした時の結果は図 3 である．横軸は視野角で，縦軸は「比較刺激が標準刺激より右にある」と答えた確率 p_{right} を示す軸である．黒い線は聴覚を，青い線は視覚を，色が薄いほどノイズ率が大きいことを意味する．単一感覚実験の場合，視覚も聴覚情報期待値はゼロ付近に存在する．さらに，ノイズ率が大きいほど，期待値付近の接線の傾きが小さくなることから感覚情報の分散が大きくなる傾向が見られる．

実験 4 と実験 5 の，視聴覚が刺激が同時に与えた場合にフィッティングした累積正規分布を図 4 に示す．曲線の色が薄いほど視覚のノイズが大きいことを意味する．視聴覚統合の場合，ノイズが大きくなると分散が大きくなる傾向が見られる．

視覚が占める重みとノイズ率の関係を示したのが図 5 である．黄色い点は実験で得られた視覚の重みで，青い陰影の部分は MLE によって予測した視覚が占める重みである．ノイズ率が上がると視覚の重みに減る傾向が見られて，基本的に MLE の予測範囲におさまることがわかる．

最後に，確率分布の閾値（あるいは標準偏差）とノイズ率との関係を示したのが図 6 である．青い点は視覚の標準偏差で，黄色の点は視聴覚統合時の標準偏差を示し，黒い点線は聴覚の標準偏差を示す．青い陰影の部分は単一感覚実験の結果から MLE に基づいて予測した各ノイズ率における標準偏差の予測値の範囲である．全体的にノイズが上がると，標準偏差も上がる傾向が見られる．ノイズ率が一番大きい時の視覚の標準偏差は聴覚を上回ったことも見られる．RDS を視覚刺激として使う場合（図 6 左上），統合された視聴覚の標準偏差はいずれが独立した場合よりも小さくなり，MLE が予測した範囲内におさまっていることもわかる．それに対し，3D オブジェクトを視覚刺激として使う場合 (図 6 大)，一番ノイズ率が高い場合における統合結果の標準偏差は感覚が独立だった場合よりも高く，MLE が予測した範囲内に入らない点も存在する．

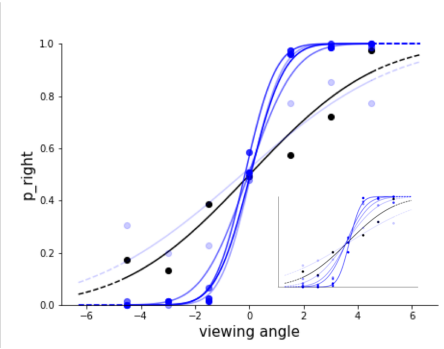


図 3: 単一感覚の場合の聴覚と五つの違うレベルのノイズが入った視覚情報の正規分布．黒い曲線は聴覚で，青い曲線は視覚を表し，色が薄いほどノイズレベルが高い．大きいグラフは 3D オブジェクトが視覚刺激の場合で，右下の小さいグラフは RDS が視覚刺激の場合を指す．

5. 考察

図 5 により，視覚重みの面からすると VR 環境下での 3D オブジェクトを視覚刺激として用いる場合は RDS を視覚刺激に使用した Battaglia らの結果とほぼ同じ傾向が見られる．すなわち，視覚刺激の種類が RDS でも VR 下のオブジェクトでも，方向認識のタスクでは，視覚の不確かさが増えるにつれ視聴覚統合が行われた際に脳はより聴覚情報を頼ることになり，この側面では最尤推定モデルに即する．

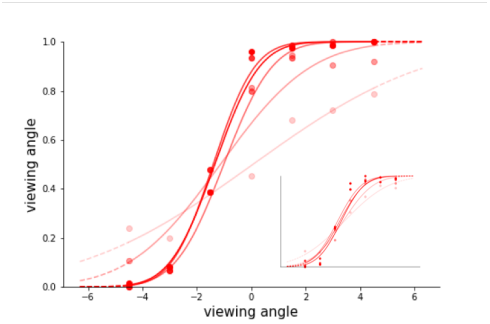


図 4: 統合感覚の場合の五つの違うレベルのノイズが入った統合結果の正規分布．色が薄いほどノイズレベルが高い．大きいグラフは **3D** オブジェクトが視覚刺激で，右下の小さいグラフは **RDS** が視覚刺激の場合を指す．

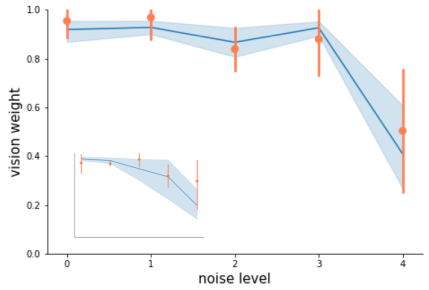


図 5: 視覚重みと視覚に入っているノイズレベルとの関係．黄色い点感覚統合実験から得られた結果で，青い陰影は **MLE** によって予測された予測範囲である．大きいグラフは **3D** オブジェクトが視覚刺激の場合，左下の小さいグラフは **RDS** が視覚刺激の場合を指す．

しかし，視覚刺激が RDS の場合に見られる視聴覚統合後の認識結果の信頼度の増加は VR におけるオブジェクトを視覚刺激として用いる場合では見られていない．すなわち，

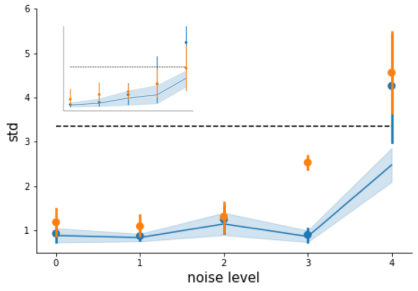


図 6: 視聴覚を独立して提示した場合と同時提示した場合の正規分布の標準偏差と視覚ノイズレベルとの関係．青い点は視覚の独立提示時，黄色い点は視聴覚同時提示時，黒い点線は聴覚の独立提示時に得られた標準偏差である．青い陰影は **MLE** によって予測された統合結果の標準偏差の範囲である．大きいグラフは **3D** オブジェクトが視覚刺激の場合，左上の小さいグラフは **RDS** が視覚刺激の場合を指す．

MLE モデルに従っていないことから,VR 環境下における視聴覚は統合されているものの脳がこの統合された認識結果を信用していないか，統合がうまく行われていない可能性があると考えられる．今回実験で用いた VIVE PRO の場合各レンズの画素数はおおよそ 250 万であるのに対し，人間の片目は何億もの画素数を持っており中心視野だけでも約 1000 万の画素数を持っていると言われている．すなわち，現実のオブジェクトをみる際の解像度は VR 環境を遥かに上回ることになる．視野角からしても VIVE PRO は 110° の視野角しか持っておらず，200° 近くの視野角を持っているヒトと比べると狭い．それ以外に，ヘッド・マウンティッド・ディスプレイを装着することで，現実ではないという先入観が入ることも感覚認識や統合の違いに寄与したと考えられる．

6. まとめ

本研究では VR 環境下において 3D オブジェクトを視覚刺激として用いる時の視聴覚統合特性を調べた．統合された視聴覚の統合結果の分散が予測値を上回ったことから，VR 環境下における方向認識のタスクでは視聴覚統合はうまく行われていないことが示唆された．なお，単一感覚の信頼度により統合結果の重み変化は VR 環境にも見られることから，VR 下の多感覚情報を使った行動誘導やリダイレクションの研究に応用できると考える．

参考文献

[1] Marc O. Ernst and Martin S. Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, Vol. 415, No. 6870, pp. 429–433, 2002.

[2] Hannah B Helbig and Æ Marc O Ernst. Optimal integration of shape information from vision and touch. pp. 595–606, 2007.

[3] Peter W. Battaglia, Robert A. Jacobs, and Richard N. Aslin. Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A*, Vol. 20, No. 7, p. 1391, 2003.

[4] W. David Hairston, M. T. Wallace, J. W. Vaughan, B. E. Stein, J. L. Norris, and J. A. Schirillo. Visual localization ability influences cross-modal bias. *Journal of Cognitive Neuroscience*, Vol. 15, No. 1, pp. 20–29, 2003.

[5] Bela Julesz. Binocular depth perception of computer-generated patterns. *Bell System Technical Journal*, Vol. 39, No. 5, pp. 1125–1162, 1960.

[6] P. O. Bishop. Can random-dot stereograms serve as a model for the perception of depth in relation to real three-dimensional objects? *Vision Research*, Vol. 36, No. 10, pp. 1473–1477, 1996.

[7] Steam® audio, copyright 2017 – 2019, valve corp. all rights reserved.