



NTT Communication Science Laboratories

OPEN HOUSE 2026

こころと知性をつむぎあ
まじわりる世界を解き明かし

5/20^{Wed.}・21^{Thu.}・22^{Fri.}

Venue

NTT WEST QUIPTBRIDGE PRISM
4-15-82 Higashinodamachi, Miyakojima-ku, Osaka
10 min. walk from north exit of Kyobashi Station on JR Osaka Loop Line;
10 min. walk from west exit of Kyobashi Station on Keihan Main Line;
5 min. walk from Kyobashi Station on Nagahori Tsurumi-ryokuchi Line
on Osaka Metro

Registration
required

Register
here



Free
admission

Machine Learning Science

- 01 Seeking to capture the growth trajectory of cells Interactive hypothesis testing on cell differentiation
- 02 Estimating cellular interactions from patterns Modeling biological development with agent simulation
- 03 Towards energy-efficient AI models Integrating sparsity and quantization for model compression
- 04 Do not compute: Fast approach for vector search Accelerating ScaNN via pruning-based vector quantization
- 05 Fast and accurate event timing prediction Event time analysis via point processes and machine learning
- 06 Who is affected by this policy, and why? Accurate, interpretable statistical causal effect estimation
- 07 Evaluating uncertainty in the probability estimation Variance computation for probabilistic inference outcome

Media Information Science

- 08 Listening to what you want! Real-time selective listening of everyday sounds
- 09 Providing a diverse range of appealing voices An idol voice dataset for research on speech generation AI
- 10 Were you able to thoroughly disinfect the entire area? Evaluation of disinfection activities using thermal camera
- 11 3D world captured by humans and AI Comparing depth estimation bias using large-scale human data
- 12 Material perception through vision Vision-based methods for conveying softness and stickiness

Communication and Mathematical Science

- 13 Solving open problems in arithmetic dynamics Partial resolution of Morton-Vivaldi's conjecture
- 14 Can you trust these search results? Hub text identification for cross-modal embeddings
- 15 How is envy depicted in picture books? The effectiveness of picture books for emotional education
- 16 How was that story? An AI system for talking about picture book impressions
- 17 Measuring cognitive and affective empathy Dataset construction and analysis for both empathy types

Human Science

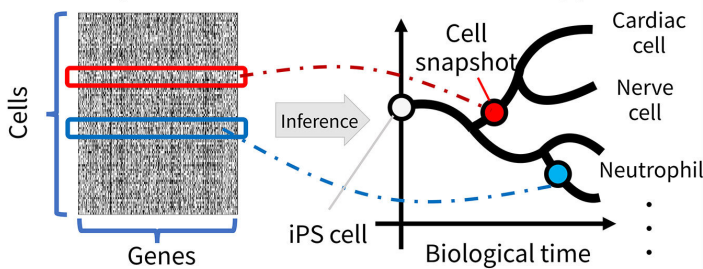
- 18 What do circular leg movements tell us? Smartphone-based foot skill and postural control evaluation
- 19 The essence of winning revealed through esports In-match biosignals predict win/loss and assess expertise
- 20 What mind-body states drive elite victory?! Individualized hormone-performance relationship & regulation
- 21 Understanding the mind from eyes and brain in the wild Scaling mind reading beyond the lab
- 22 How people make sense of non-realistic virtual worlds Examining how a sekai-kan is formed in the human mind

Abstract

Biological cells develop into various organs and tissues through differentiation. How individual cells undergo this differentiation process is a critical and universal question in the life sciences. To elucidate these differentiation mechanisms, it is essential to both generate hypotheses and test them. Therefore, we are constructing a new mathematical framework for hypothesis generation and testing. Specifically, this technology assists life science researchers in **interactively discovering hypotheses through AI, which aids in the visual design of differentiation structure hypotheses and in testing of their validity**. Deeper insights into the mechanisms of biological cell differentiation may enable us to induce differentiation phenomena involving various genetic and environmental factors. By investigating the causes of congenital diseases and the efficacy of medications for designated intractable diseases, we aim to contribute to the advancement of regenerative medicine and artificial organ technology.

Cell differentiation inference

Gene expression data Differentiation hypothesis



Purpose: Understanding of differentiation mechanism

Application: Regenerative medicine and gene therapy

Challenges of differentiation inference task

- **Current standard approach** : High-dimensional gene expression features are visualized in a lower dimension using dimension reduction techniques, and experts interpret them.

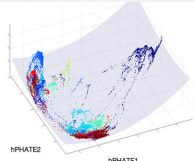
- #1: Unknown nature of genes responsible for local differentiation

Each specific gene is thought to play a major role in the local differentiation of cells. However, in order to verify the genes-cause-differentiation hypothesis, the differentiation structure itself is required.

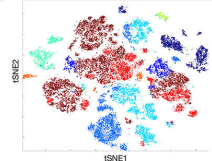
- #2: Lack of established metrics for assessing validity of visual rep

In general, visualization techniques carry the risk of removing important clues about the differentiation structure during compression, as well as the risk of introducing misunderstanding that were not present in the original data.

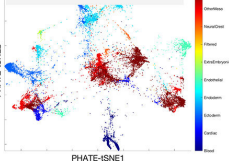
Hyperbolic PHATE [1]



t-SNE



PHATE+t-SNE

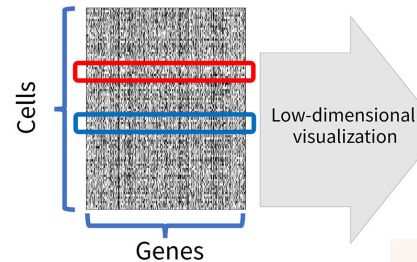


Examples of three types of visualizations for same gene expression data

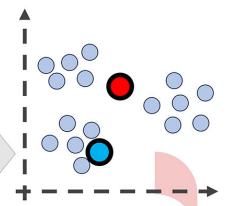
Data : Han+, Single cell transcriptomics identifies a signaling network coordinating endoderm and mesoderm diversification during foregut organogenesis, Nature Communications 2020.

Proposed method

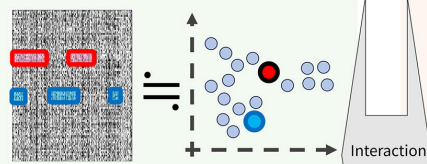
Gene expression data



Visualization

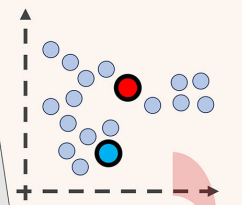


Support through AI

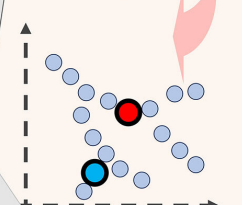


- **Role 1:** Inverse prediction of genes contributing to local differentiation
- **Role 2:** Validity assessment of visual representations using hypothesis testing

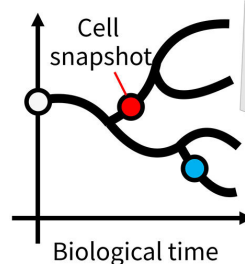
Manual editing



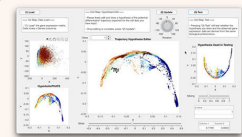
Editing repeats



Differentiation hypothesis



Differentiation inference through AI



We will demonstrate this process during the exhibition.

References

[1] M. Nakano, H. Sakuma, R. Nishikimi, K. Komiya, T. Iwata, K. Kashino, "HyperbolicPHATE: Visualizing Continuous Hierarchy of Latent Differentiation Structures," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2025.

Contact

Masahiro Nakano, Biomedical Informatics Research Group, Media Information Laboratory

Abstract

Self-organized multicellular patterns emerging from intercellular interactions are fundamental to diverse biological functions, such as detoxification in hepatic lobules and camouflage through stripe patterns in zebrafish skin. However, quantitatively inferring the underlying intercellular interactions that give rise to target multicellular patterns remains a **challenging inverse problem**. In this study, we propose a framework that estimates parameters governing adhesion and repulsive forces between cell types directly from target multicellular patterns. Our method extracts robust topological features across multiple spatial scales and employs a pretrained surrogate model to efficiently infer interaction parameters. We validate the framework using both simulated and image-based zebrafish pigment patterns. This framework provides an efficient and interpretable approach for linking multicellular pattern geometry to intercellular interactions and has potential applications in tissue engineering, including the **design and control of organoid morphogenesis**.

Multicellular Pattern Formation

Biological functions enabled by multicellular patterns

Representative multicellular patterns and associated functions

Liver: Detoxification

Hepatocytes, Endothelial cells, Kupffer cells, and etc...

Stripes patterns in zebrafish sides: Camouflage

Xanthophores and melanophores

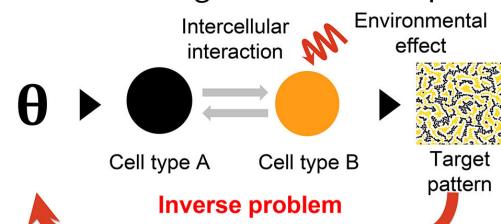


These patterns arise from intercellular interactions

Objective : Estimating intercellular interactions that enables target multicellular patterns
 →Applying the method to tissue engineering such as Organoids

Proposed method

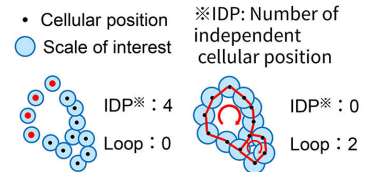
Estimate parameters of intercellular interaction from target multicellular patterns



θ : Parameters define adhesion and repulsive forces

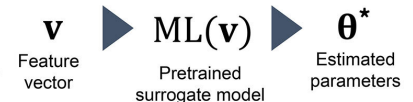
Topological features

Robust feature extraction by counting topological features at each scale.



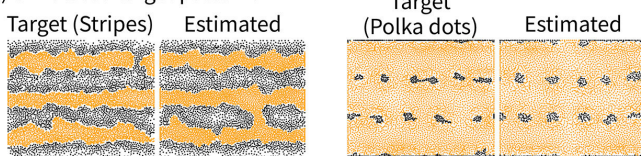
Inverse surrogate model

Efficient and direct estimation with surrogate modeling

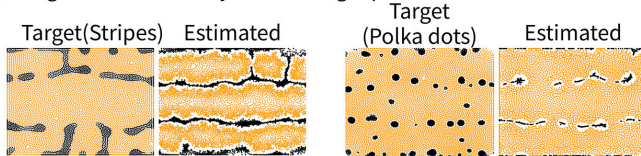


Validation with zebrafish pigment formations

(1) Simulated target patterns



(2) Image-based manually crafted target patterns



Result summary

Target pattern	Stripe	Polka-dots
(1) Simulated	○	○
(2) Image-based	△	△

○ : Visually and quantitatively match

△ : Roughly match, but there are some minor differences

(1) The framework successfully inferred intercellular interaction parameters from simulated target patterns.

(2) Minor discrepancies were found in image-based cases, but they could be used to refine the inverse surrogate model.

References

- [1] A. K. Jin, K. Komiya, R. Nishikimi, K. Kashino, "Topology Informed Surrogate Modeling for Parameter Optimization in Multicellular Models," in *Proc. International Conference on Biomedical and Health Informatics (BHI)*, 2025.
- [2] K. Komiya, A. K. Jin, R. Nishikimi, K. Kashino, "Learning Pairwise Potential via Differentiable Recurrent Dynamics," in *NeurIPS Workshop on Machine Learning and Physical Science*, 2025.

Contact

Kenji Komiya, Biomedical Informatics Research Group, Media Information Laboratory

Abstract

The **increasing performance of AI models** has led to **growing computational cost and energy consumption**, motivating efficiency techniques such as sparsity and quantization. However, conventional approaches often rely on their combined use, which can **restrict deployment to digital hardware** that efficiently supports sparse computation. In this work, we present a training framework that does not require learned sparsity, while remaining compatible with learned quantization. This **enables neural networks to be flexibly deployed** across a wider range of hardware platforms, **including energy-efficient analog devices** where sparsity is difficult to exploit. By decoupling model efficiency from specific structural constraints, our approach broadens the applicability of model compression. Ultimately, this work aims to **support the development of energy-efficient, reconfigurable AI systems that can operate across diverse computational substrates**, from digital to emerging analog hardware.

Neural Network (NN) Efficiency

- NNs require large computation and memory
- **Pruning** and **quantization** are key for efficiency

Precise and Dense NN

All connections, full-precision
→ computation and memory costs are high
e.g., 32bit x 9 weights

Pruning

Eliminate unnecessary connections
→ smaller cost
e.g., 32bit x 5 weights

Quantization

Reduce numerical precision
→ smaller cost
e.g., 3bit x 9 weights

Prun. + Quant.

Simplify structure and values
→ smallest cost
e.g., 3bit x 5 weights

Sparse Computation on Analog Devices

- Pruning assumes **exact zeros in digital devices**, but fails in **analog devices** (e.g., optical) due to **noise**

Digital Computing

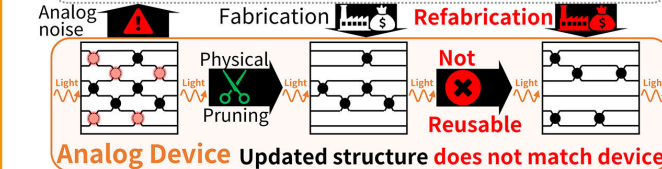
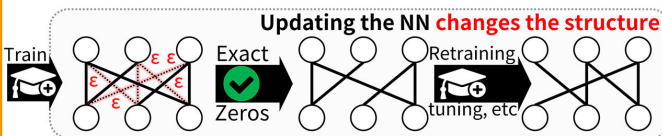
Pruned connections represented with exact zeros
→ accurate

Analog Computing

Pruned connections reactivated by noise
→ accuracy collapse

- **Conventional approach: physical pruning at fabrication**

- ✓ Remove connections from topology → exact zeros
- ✓ No need to store zeros → higher compression
- ✓ Reduced hardware → lower area, noise, power
- ✗ **Less reconfigurability** → limited practicality



Our Research: Towards Reusable Analog AI

- What to learn differs across hardware: digital prefers connections; analog prefers signs and magnitudes

Digital Computing

Weight values: **Fixed**
Structure: **Learned**
Efficient and sparse NN
Sparsity represented with exact zeros

AND

Flip

e.g., 1bit x 9 weights

Analog Computing

Weight values: **Learned**
Structure: **Fixed**
Efficient and sparse NN
Sparsity implemented with physical connections

AND/OR

Learn signs
learn magnitudes

Learn • Fix targets • flip

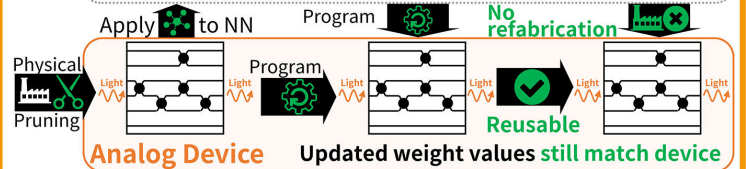
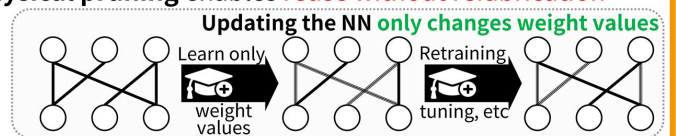
e.g., 1bit x 5 weights

e.g., 3bit x 5 weights

Proposed: Learn weight signs and magnitudes over a fixed topology

- ★ Quantization-aware training (QAT) under structural constraints
- ✓ Retains benefits of conventional method
- ✓ **Structure unchanged** → reusable
- ✓ No structure learning → can be determined by **hardware topology optimization**

- Learning only signs and magnitudes after applying **physical pruning** enables **reuse without refabrication**



References

- [1] Á. López García-Arias, Y. Okoshi, H. Otsuka, D. Chijiwa, Y. Fujiwara, S. Takeuchi, M. Motomura, "The Trichromatic Strong Lottery Ticket Hypothesis: Neural Compression With Three Primary Supermasks," *Workshop on Machine Learning and Compression, Conference on Neural Information Processing Systems (NeurIPS)*, 2024.
- [2] Á. López García-Arias, Y. Okoshi, H. Otsuka, D. Chijiwa, Y. Fujiwara, S. Takeuchi, M. Motomura, "The Trichromatic Strong Lottery Ticket Hypothesis: A Unifying View of Supermask-Based Learning," *IEEE International Joint Conference on Neural Networks (IJCNN)*, 2026.

Contact

Ángel López García-Arias, Recognition Research Group, Media Information Laboratory

Abstract

Quantization, which replaces vectors with codewords, is widely used to enable **fast, accurate inner-product-based similarity search over large-scale data**. ScaNN is a popular approach for quantization. ScaNN computes the quantization error between each vector and all possible codewords to select the codeword with the smallest error, achieving **high approximation accuracy**. However, since ScaNN requires error computation with all codewords, it incurs a high computational cost, making quantization extremely slow for large-scale datasets. The proposed approach uses upper bounds on quantization errors and efficiently evaluates them to **prune codeword candidates**. This significantly reduces the number of error computations required. As a result, the proposed approach can **substantially accelerate vector quantization while preserving ScaNN's search accuracy**. Consequently, it facilitates practical large-scale data processing in applications such as image retrieval and natural language processing.

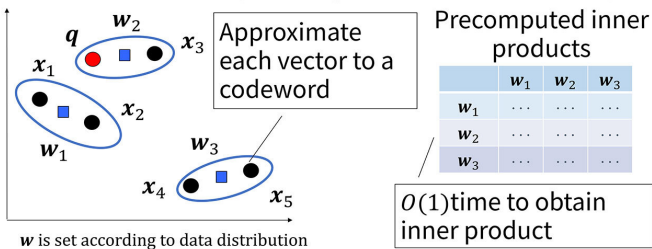
Maximum Inner Product Search

Find vector x of maximum inner product for query q

- Various application (e.g., image search and NLP)
- ✗ High CPU cost for inner products of the query and vectors in the database

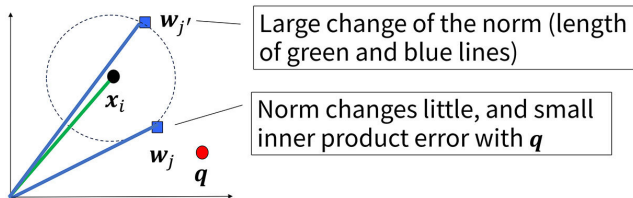
Quantization: approximate vector x with codeword w for search

- Improved efficiency
- ✗ Degraded search accuracy



Existing Approach: ScaNN

Idea: Preserve vector norms and reduce inner product error to improve search accuracy



Approach: Assign to the codeword minimizing error function $l(x_i, w_j)$

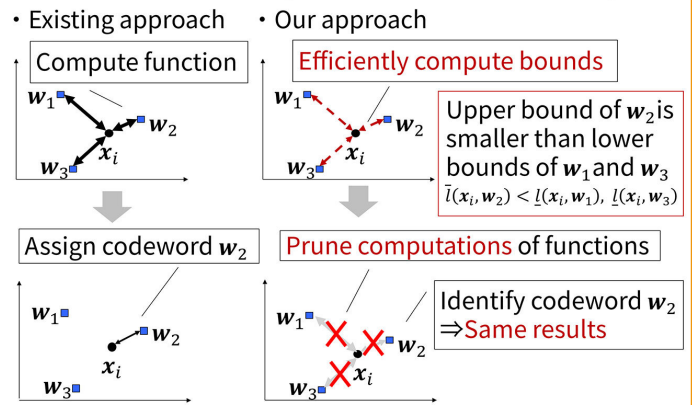
- Error function is low \Leftrightarrow Norm is preserved
- ✗ High CPU cost for error function

$$l(x_i, w_j) = \frac{h_i - 1}{\|x_i\|^2} (p(x_i, w_j))^2 - 2h_i p(x_i, w_j) + h_i \|x_i\|^2 + \|w_j\|^2$$

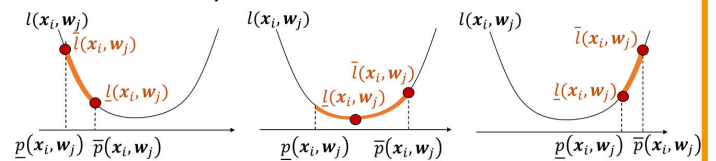
$h_i (> 1)$: hyper parameter, $p(x_i, w_j)$: inner product of x_i and w_j

Our approach: computation pruning

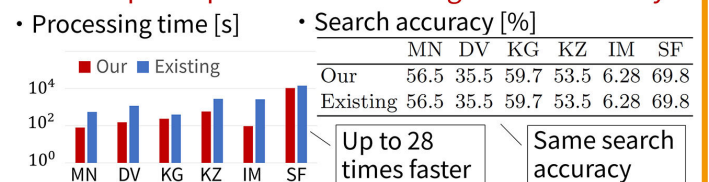
Prune unnecessary computations by efficiently obtaining upper bound $\bar{l}(x_i, w_j)$ and lower bound $\underline{l}(x_i, w_j)$



Function $l(x_i, w_j)$ is quadratic for inner product $p(x_i, w_j) \Rightarrow \bar{l}(x_i, w_j)$ and $\underline{l}(x_i, w_j)$ are obtained from upper/lower bounds of inner product



Achieve speedup without sacrificing search accuracy



Experimental data (Number of vectors \times Length of vectors)

MN: MNIST (70,000 \times 780) KZ: Kuzushiji (270,912 \times 784)
 DV: Devnagari-Script (92,000 \times 1,024) IM: imagenet (2,340,373 \times 150)
 KG: kaggle (200,000 \times 202) SF: sift10m (10,000,000 \times 128)

References

[1] Y. Fujiwara, Á. López, Y. Ida, A. Kumagai, M. Nakano, M. Nakatsuji, A. Kimura, "Fast Vector Quantization Algorithm for ScaNN", in *Proc KDD*, 2026.

Contact

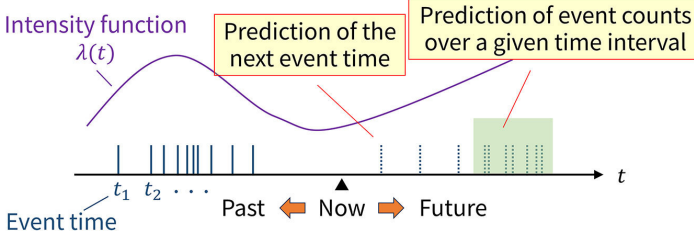
Yasuhiro Fujiwara, Recognition Research Group, Media Information Laboratory

Abstract

If the timing of future events can be predicted, risks can be mitigated through proactive preparation, or opportunities can be fully leveraged. In this study, **we present a method for efficiently predicting future event times** by leveraging point processes and machine learning. While existing approaches achieve high predictive accuracy, they often require substantial computational cost during training. To address this issue, we replace the conventional log-likelihood objective with the least squares contrast for point processes, **enabling up to several hundred-fold speedups in training while maintaining comparable predictive performance**. This improvement makes our method scalable to large datasets. As machine learning-based event prediction becomes increasingly accurate with the growing volume of data, the associated computational burden also continues to rise. Our approach, which efficiently handles large-scale event series data, **aims to support proactive decision-making** by accurately predicting when events such as equipment failures or demand fluctuations will occur.

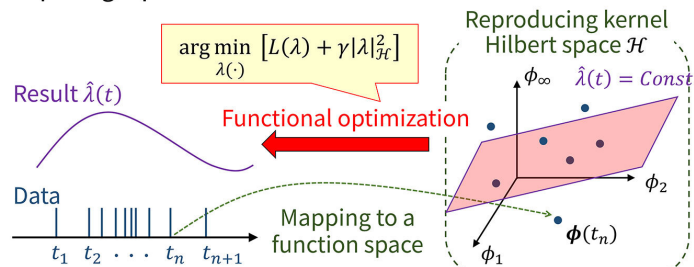
Event time series analysis

- Data that record **when and where events occur** are referred to as **event time series data**. e.g.) equipment failures, accidents, infections, SNS posts.
- Point processes provide a standard framework for analyzing such data. By learning **the intensity function** from data, they enable a wide range of analyses and predictions regarding the timing of future events.



Intensity function learning with kernel methods

Kernel methods are a class of machine learning techniques that enable flexible function learning from data without requiring a predefined functional form.



Challenge

For intensity function learning, the log-likelihood has been widely adopted as **the loss function** $L(\lambda)$ in kernel methods. However, this approach is **computationally expensive**, making it difficult to scale to large datasets.

Proposed method: Introducing least squares loss

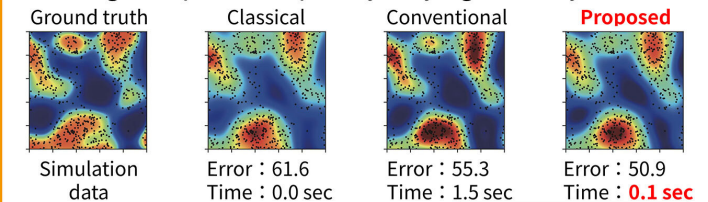
We propose fast kernel methods for estimating intensity functions that replaces the conventional log-likelihood objective with **the least squares loss for point processes**.

$$\text{Loss func.} \quad \begin{array}{l} \text{Negative log-likelihood} \\ \int_{\mathcal{T}} \lambda(t) dt - \sum_{n=1}^N \log(\lambda(t_n)) \end{array} \quad \begin{array}{l} \text{Least squares loss} \\ \int_{\mathcal{T}} \lambda(t)^2 dt - 2 \sum_{n=1}^N \lambda(t_n) \end{array}$$

Conventional methods require solving an optimization problem whose complexity can scale cubically with the size. In contrast, our method **eliminates the need for such costly optimization**, resulting in significant speedups.

■ Non-homogeneous Poisson processes [1]

Learning the spatio-temporally varying intensity function

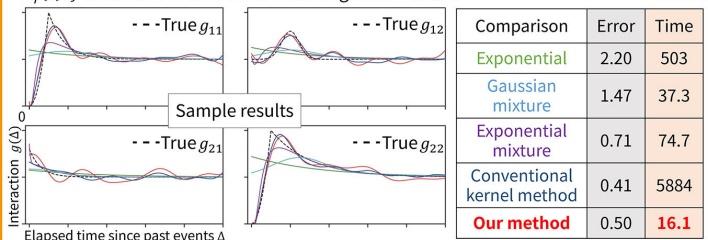


■ Hawkes processes [2]

Learning the interaction functions $g_{ij}(\Delta)$ governing dependencies among multiple event sequences.

$$\lambda_i(t) = \mu_i + \sum_{j \in \mathcal{U}} \int_{\mathcal{T}} g_{ij}(t-s) dN_j(s), \quad i \in \mathcal{U} := \{1, \dots, U\}$$

U : node number, μ_i : i -th node's base intensity, $N_j(s)$: j -th node's event count occurring before time s



References

- [1] H. Kim, T. Iwata, A. Fujino, "K2IE: Kernel method-based kernel intensity estimators for inhomogeneous Poisson processes," in *Proc. The 42nd International Conference on Machine Learning (ICML)*, 2025.
- [2] H. Kim, T. Iwata, "A representer theorem for Hawkes processes via penalized least squares minimization," in *Proc. The 14th International Conference on Learning Representations (ICLR)*, oral, 2026.

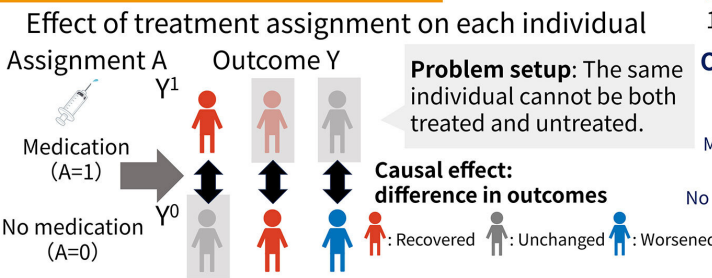
Contact

Hideaki Kim, Learning and Intelligent Systems Research Group, Innovative Communication Laboratory

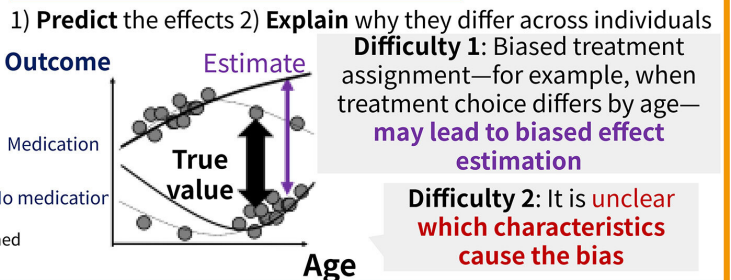
Abstract

To design important, individualized interventions such as medical treatments and targeted advertising, we need to understand who benefits from a policy, **by how much**, and **why**. We develop two methods for this goal. The first method focuses on **accurate effect estimation**. It aims to improve the prediction of outcomes with and without treatment **by detecting and correcting hidden correlations between treatment assignment and outcomes**—for example, when older patients are less likely to receive risky surgery and also tend to have worse outcomes. The second method focuses on **explanation**. It **identifies which personal characteristics account for differences in treatment effects across individuals**, while assessing the **statistical significance** of each characteristic. Compared with simple machine learning techniques that rely only on observed correlations, these methods aim to evaluate the treatment effects more accurately **by capturing cause-effect relationships**. By building accurate and interpretable causal effect estimation techniques that work even with limited data, our research aims to **support data-driven decision-making in high-stakes settings**. Ultimately, we hope this research will contribute to a future in which important decisions can be tailored **more precisely, reliably, and effectively to each individual**.

What is a causal effect?



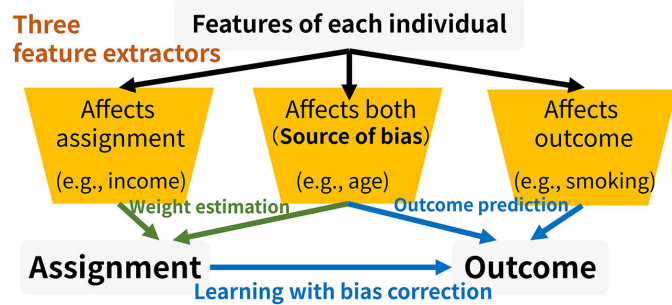
Two tasks for policy design



1. Debiased representation learning for effect prediction [1]

Feature representation learning for bias correction in high-dimensional data

- Learn debiased feature representations from data and **minimize prediction errors weighted by the degree of bias**
- Use a differentiable weight estimation technique** to efficiently correct bias and improve effect estimation



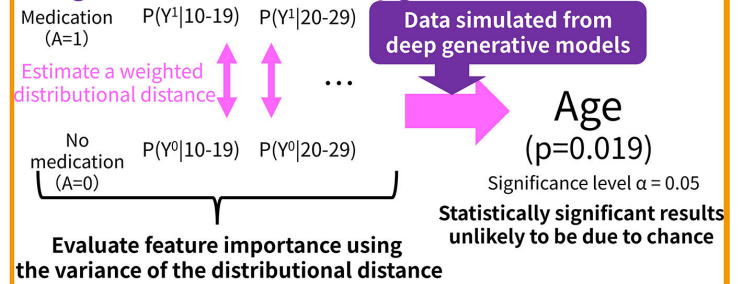
Effect prediction results Effect of reading device assignment (mobile vs. PC) on reading time

	Proposed	Existing (Representation learning)	Existing (Two regression models)
Test RMSE	2.10	2.38	2.55

2. Feature discovery for treatment effect heterogeneity [2]

More accurate identification of features that influence treatment effects than existing techniques

- Measure the feature importance by **how much the distance between the outcome distributions** with and without treatment **varies across feature values**
- Evaluate statistical significance (p-values) of each feature using data simulated from deep generative models**



Feature selection results U.S. electronic health record data on the treatment of systemic inflammation

Existing	Proposed
Age, Gender	Age, Gender, Smoking history
	Data-driven discovery of clinically important features reported in previous studies

References

[1] Y. Chikahara, K. Ushiyama, “Differentiable Pareto-Smoothed Weighting for High-Dimensional Heterogeneous Treatment Effect Estimation,” *Proc. of The 40th International Conference on Uncertainty in Artificial Intelligence (UAI '24)*, 2024.

[2] Y. Chikahara, M. Yamada, H. Kashima, “Feature Selection for Discovering Distributional Treatment Effect Modifiers,” *Proc. of The 38th International Conference on Uncertainty in Artificial Intelligence (UAI '22)*, 2022.

[3] S. Horii, Y. Chikahara, “Uncertainty Quantification in Heterogeneous Treatment Effect Estimation with Gaussian-Process-Based Partially Linear Model,” *Proc. of The 38th AAAI Conference on Artificial Intelligence (AAAI '24)*, 2024.

[4] T. Iwata, Y. Chikahara, “Meta-learning for heterogeneous treatment effect estimation with closed-form solvers,” *Machine Learning*, Vol. 113, pp. 6093-6114, 2024.

Contact

Yoichi Chikahara, Learning and Intelligent Systems Research Group, Innovative Communication Laboratory

Abstract

In probabilistic inference, which predicts the likelihood of an event based on data and the relationships among events, a lack of data increases the uncertainty of the prediction, affecting decision-making depending on it. This study introduces a technique for **quantitatively computing the "variance of inference outcome,"** an indicator of this uncertainty. It has been difficult to compute due to the enormous computational burden. To address this, we use the technique of counting the number of solutions that satisfy given logical formulas. Our study can be used to **make more robust decisions that could otherwise lead to serious consequences.** For example, in predicting availability of infrastructure equipment, even if the prediction satisfies safety standards, a large variance value indicates high uncertainty. In such cases, our approach enables operational decisions such as encouraging additional data collection to avoid excess failure rate that exceeds the prediction.

Probabilistic Inference and Its Uncertainty

Probabilistic inference predicts the probability of occurrence of an event from data and the relationship among events.

- Blackout will increase the probability of disconnection.
- How much is the probability of disconnection considering both normal and blackout situations?

	Prob. of situations	Prob. of disconnection	
Normal	99%	0.5% when normal	Normal & disconnected $99\% \times 0.5\% = 0.495\%$
Blackout	1%	30% when blackout	Blackout & disconnected $1\% \times 30\% = 0.3\%$
			Disconnection = 0.795%

Predicted probability (**inference outcome**) will be used for, e.g., decision-making.

When our goal is $\leq 1\%$ prob. of disconnection, We need not update equipment because the inference outcome is below the goal.

Lack of data causes uncertainty in the parameters used for inference. This leads to an **uncertain inference outcome.**

	Prob. of situations	Prob. of disconnection	
Normal	$(99 \pm 0.5)\%$	$(0.5 \pm 0.2)\%$ when normal	Disconnection $(0.795 \pm ??)\%$
Blackout	$(1 \pm 0.5)\%$	$(30 \pm 14)\%$ when blackout	

Since uncertainty in the inference outcome could affect the decision-making, we consider computing the **"variance of inference outcome,"** an index for this uncertainty.

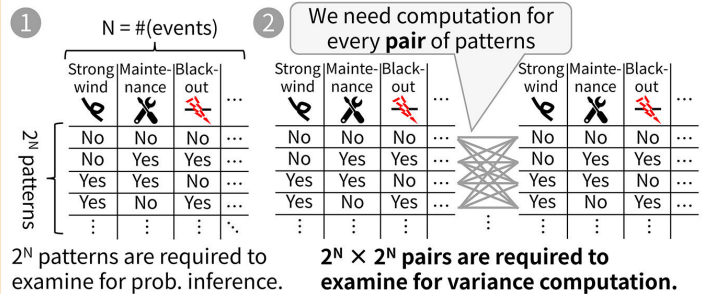
If the prob. of disconnection is $(0.795 \pm 0.1)\%$, We need not update equipment because the inference outcome is safely below the goal.

If the prob. of disconnection is $(0.795 \pm 0.4)\%$, We **should** update equipment because there's some possibility that the prob. of disconnection exceeds 1%...

Difficulty of Variance Computation

- 1 For probabilistic inference, we need to examine all patterns on the occurrences of the related events.
- 2 To compute the variance of inference outcome, we need more complicated calculation on these patterns.

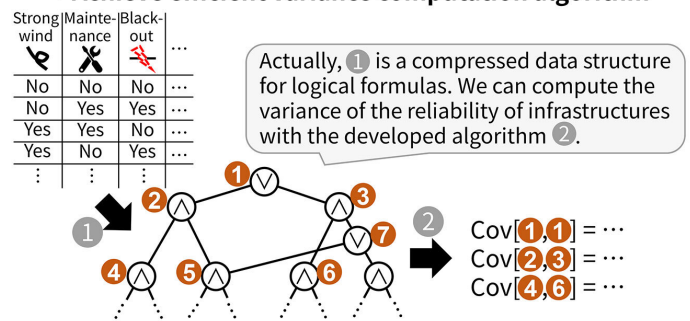
Traditionally, we can compute the variance in a reasonable time only when **the number of related events is up to 10.**



Solution by Counting on Logical Formulas

We proposed an algorithm to compute the variance in a reasonable time **even when #(related events) ≥ 100 .**

- 1 Transform the patterns on the occurrences of related events into a compressed data structure
 - 2 Develop an algorithm to compute the variance of inference outcome on the built data structure
- **Achieve efficient variance computation algorithm**



References

[1] K. Nakamura, M. Nishino, N. Yasuda, "Variance computation for weighted model counting with knowledge compilation approach" , in Proc. The 40th Annual AAAI Conference on Artificial Intelligence (AAAI ' 26), 2026.

Contact

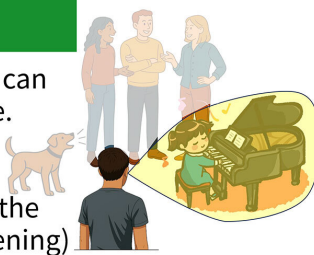
Kengo Nakamura, Linguistic Intelligence Research Group, Innovative Communication Laboratory

Abstract

Humans can selectively listen to a target sound even when many sounds overlap. This research brings that capability to computers by **developing real-time target sound extraction that isolates desired audio from mixed signals on general-purpose PCs** while maintaining high accuracy. By incorporating an audio foundation model with general sound representations developed at NTT, the method further improves extraction accuracy and sound quality. We also implement binaural processing to estimate the direction of arrival, making the system closer to human listening. Ultimately, the technology **lets users flexibly hear or suppress sounds depending on the context**, for example, by reducing household noise in remote-work meetings while preserving meaningful sounds during family calls, enabling more comfortable and effective communication.

Selective Listening

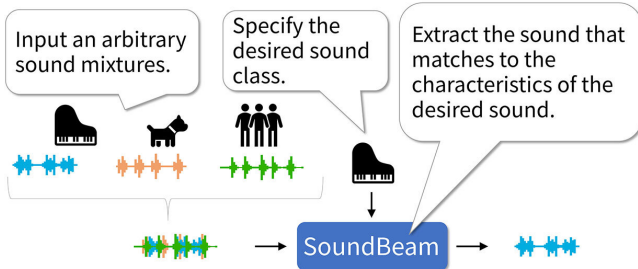
- In daily life, many sounds can be heard at the same time.
- Humans can focus only on the sounds they want to listen to depending on the situation. (= Selective listening)



Goal: Realizing computational selective listening!

SoundBeam Mechanism[1]

SoundBeam realizes selective listening of arbitrary sound using a neural network (NN).



By changing the desired sound class, we can extract various kinds of sounds.

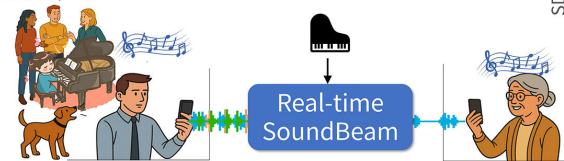


→ We introduce three new functions and performance enhancements of SoundBeam for selective listening toward various applications!

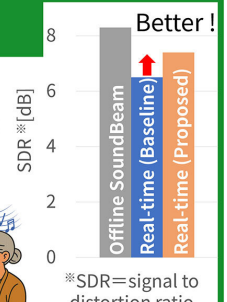
Some parts of the figures were created using generative AI.

① Real-time processing [2]

We devised an NN methodology enabling effective selective listening for real-time processing simultaneously with recording. (Collaboration with CD and HI labs)

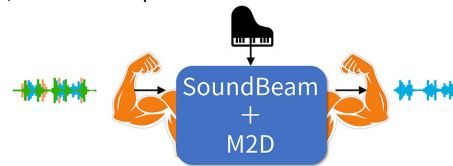


Application: Selectively transmit only the sounds you want others to hear in web conferences.



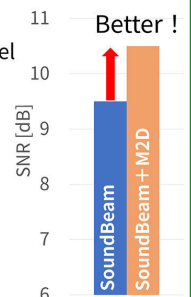
② High-Quality Sound Extraction [3]

We improved the accuracy of selecting the desired sound and enhanced the quality of the extracted sound, by combining SoundBeam with NTT's audio foundation model M2D^{**}, which can capture detailed audio characteristics.



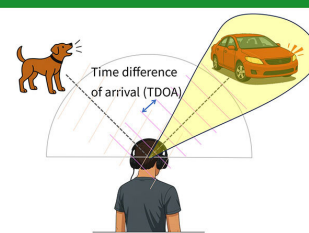
※ M2D=Masked Modeling Duo
NTT's learning method for audio foundation models

Application: High-quality post-production for recorded audio such as movies, music, and home videos.

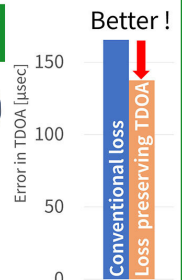


③ Reproduction of Sound Direction [4]

By accurately extracting the interaural time difference of sounds reaching the left and right ears, we achieved selective listening that precisely preserves sound source direction information.



Application: Hear important sounds such as warnings, including their directional information while using earphones.



References

- [1] M. Delcroix, J. B. Vázquez, T. Ochiai, K. Kinoshita, Y. Ohishi, S. Araki, "SoundBeam: Target sound extraction conditioned on sound-class labels and enrollment clues for increased performance and continuous learning," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 31, pp.121-136, 2022..
- [2] K. Wakayama, T. Ochiai, M. Delcroix, M. Yasuda, S. Saito, S. Araki, A. Nakayama, "Online target sound extraction with knowledge distillation from partially non-causal teacher," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 561-565, 2024.
- [3] C. Hernandez-Olivan, M. Delcroix, T. Ochiai, D. Niizumi, N. Tawara, T. Nakatani, S. Araki, "SoundBeam meets M2D: Target sound extraction with audio foundation model," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2025.
- [4] C. Hernandez-Olivan, M. Delcroix, T. Ochiai, N. Tawara, T. Nakatani, S. Araki, "Interaural time difference loss for binaural target sound extraction," in *Proc. 18th International Workshop on Acoustic Signal Enhancement (IWAENC)*, pp. 210-214, 2024. IEEE.

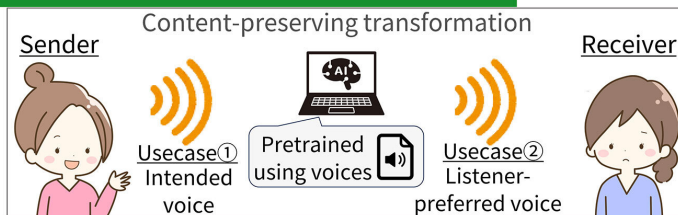
Contact

Marc Delcroix, Media Information Laboratory, Signal Processing Research Group

Abstract

Research data for speech generation AI are often biased toward acted speech produced by professional speakers, such as voice actors. As a first step toward enabling AI systems to generate speech that is personalized to individual users' preferences and perceived as appealing, we construct **the Japanese Idol Speech (JIS) corpus, a multi-speaker speech dataset featuring live idols as speakers with a wide range of vocal characteristics.** JIS is the first large-scale idol speech dataset with properly established contractual arrangements for research use, comprising over 200 speakers and approximately 30 hours of audio. In addition to reading speech and everyday conversational utterances commonly found in existing speech datasets, JIS includes distinctive idol-specific speech styles, such as utterances simulating Instax photo session events. By learning from the voices of diverse individuals, **we aim to develop an AI system that enables anyone to flexibly and effectively refine their own voice.**

Communication augmentation through speech generation AI



Future goal : **Generating attractive voices tailored to fit everyday life (Voice makeup)**

What is a fascinating voice ?

- ① Differ from listener to listener
 - ② Jointly determined by voice quality and style
- } Diverse

As a first step, **we collect diverse fascinating voices**

	JVS	CSJ	JIS (Prop.)
Speaker	Voice actors / Narrators	Non-professional	Live Idols
Scale	~30 h 100 people	~660 h 1417 people	~30 h 204 people
Natural-ness	△	○	○
Attractive-ness	○	△	○

※ JVS: Japanese Versatile Speech corpus, CSJ: Corpus of Spontaneous Japanese

Selected speaker genre

Live idol: Idols focused on live performances

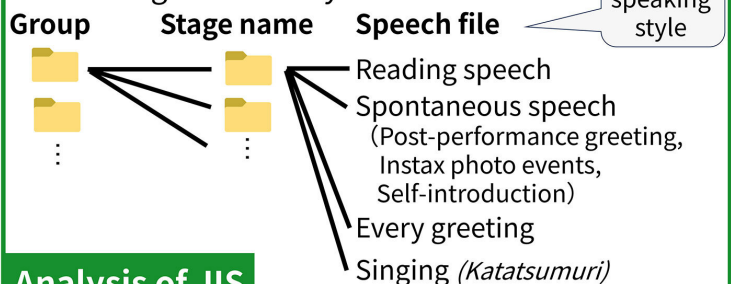
Benefits for Research

- **Balancing privacy protection and fan accessibility (via stage names)**
- **Detailed subjective evaluations by fans who understand speakers' appeal**
- **Feasible voice collection from many speakers**



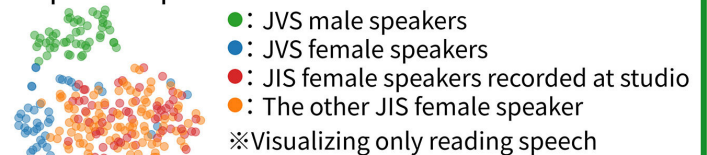
Japanese Idol Speech Corpus (JIS)

- A large-scale dataset of speech recordings from many female live idols



Analysis of JIS

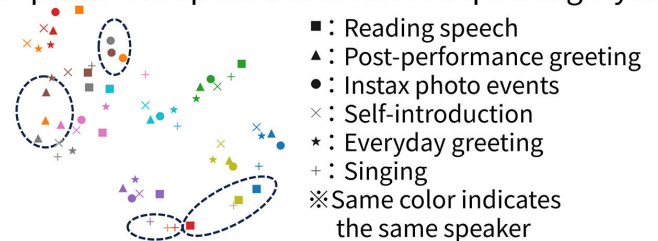
- 2D plot of speaker acoustic features



Separation between JVS female and JIS speakers

➡ **JIS covers speaker types insufficiently covered by JVS**

- 2D plot of JIS speech with various speaking styles



Not only the speaker differences

but also style-specific features (○) are confirmed

➡ Suggesting that differences in speaking style are related to listener impressions (e.g., ▲ ⇒ resonant)

➡ **Potential for scenario-aware voice generation**

References

[1] Y. Kondo, H. Kameoka, K. Tanaka, T. Kaneko, "JIS: A Speech Corpus of Japanese Idol Speakers with Various Speaking Styles," in *Proc. INTERSPEECH*, pp. 4783-4787, 2025.

Contact

Yuto Kondo, Computational Modeling Research Group, Media Information Laboratory

Abstract

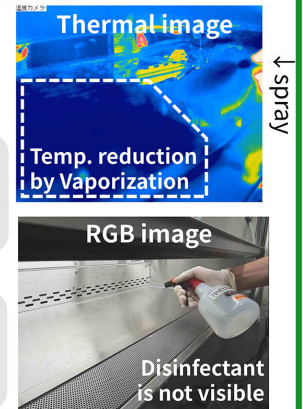
We propose a visualization system to verify wipe disinfection status on surfaces such as a safety cabinet. In wipe disinfection, disinfectants such as alcohol are sprayed onto the surface and wiped off with sufficient pressure to remove cells and bacteria. Our proposed method **evaluates wipe disinfection process** using a **thermal camera** and **pressure sensors**. The thermal camera captures sprayed disinfectant on surfaces by detecting a reduction in surface temperature due to its evaporation. The pressure sensor measures the applied pressure during wiping. By combining these data, our system visualizes the area that was sufficiently sprayed with disinfectant and wiped with enough pressure. In medical situations, including regenerative treatment, complete disinfection of the entire area is essential. However, wipe disinfection has been performed according to the manual, and methods for verifying disinfection results are limited. In contrast, our system enables users to **improve disinfection treatment by understanding disinfection** results and to achieve an environment where users can work safely to disinfect. We aim to contribute to improving **medical safety and efficiency**.

Wipe disinfection treatment

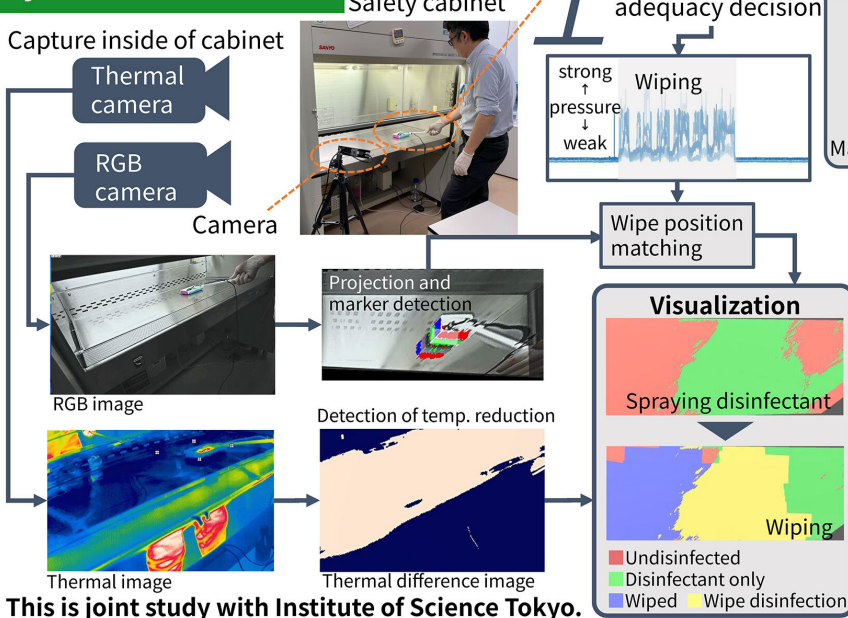
- Disinfection is essential treatment for preventing infection
 - Disinfection is mandatory for each patient in regenerative medicine
 - Preventing contamination
- Challenge 1: Insecurity** in disinfection
Thoroughly disinfected?
Difficulty to improve efficiency
- Challenge 2: Inefficiency** in disinfection
Discontinuation of use until all cells and bacteria are eliminated

Visualization of wipe disinfected area

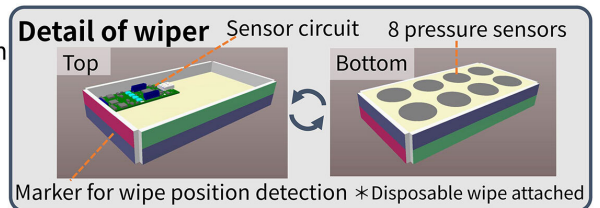
- Contributing to improve treatment safety and reduce workload
- Detection method of disinfected area**
- Disinfection:** Consistent disinfectant
- ➔ Detection of temperature reduction area by thermal camera
 - ➕ ➔ Sufficient wipe disinfection
- Wipe:** Throughout enough pressure
- ➔ Install pressure sensor on wiper
 - ➔ Record its position by RGB camera



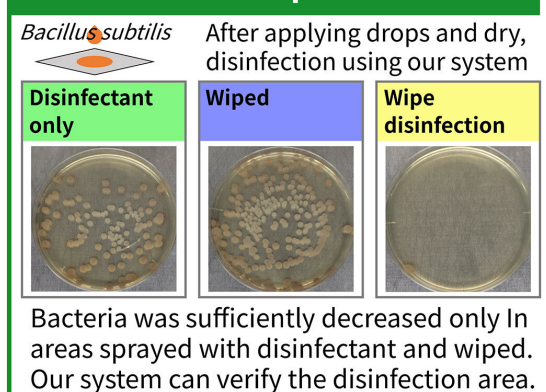
System architecture



This is joint study with Institute of Science Tokyo.



Verification of wipe disinfection



References

[1] Mitsuru Mizuno, Yasue Kishino, Yoshinari Shirai, Junpei Matsuda, and Ichiro Sekiya, "Assessment of wipe disinfection effectiveness on safety cabinet work surfaces using a wipe disinfection evaluation system" *The 24th Congress of the Japanese Society for Regenerative Medicine*, p.706, 2025. (in japanese)

Contact

Yasue Kishino, Interaction Research Group, Innovative Communication Laboratory

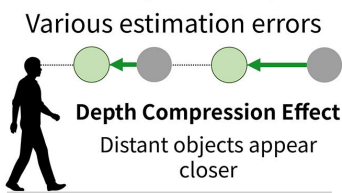
Abstract

Humans can naturally estimate 3D structures from 2D images, and recent advances in artificial intelligence (AI) have enabled physical devices to develop similar abilities. **Our research investigates whether these systems rely on the same visual cues as humans in depth estimation.** To this end, we collected large-scale human-annotated data for indoor and outdoor images and compared them with predictions from various AI models. Our results show that **many AI models exhibited estimation biases similar to humans** (e.g., perceiving distant objects as closer than they physically are). Additionally, **we identify an accuracy-similarity trade-off**: highly accurate AI models often behave less like humans. By precisely modeling human-like error patterns, our work contributes **to the development of AI models that better align with human perception.** This may support safer and more intuitive applications, such as remote robot operation, where visual misunderstandings can lead to accidents.

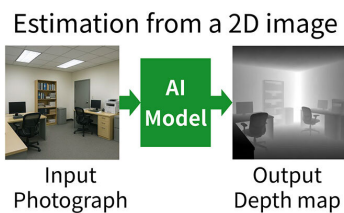
Background

Humans can estimate 3D structure using various cues from the information projected onto their retinas.

Human Depth Perception



AI Depth Perception

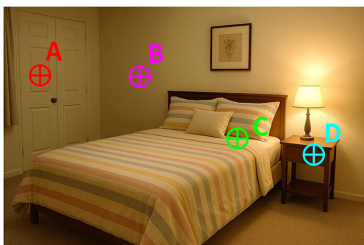


Q. How similar are the error patterns of human and AI depth perception?

Why important? Human-centered information presentation by precisely predicting human error patterns in 3D perception

Human-annotated Data Collection

We collected a **large-scale human-annotated dataset in depth perception** through online experiments.

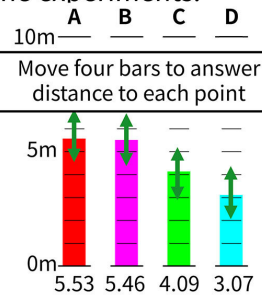


*Mimicking stimuli generated by GPT-4.0

■ Procedure

- Participants answered perceived distances in meters.
- We collected eight or sixteen responses per image.

■ **Stimuli** Indoor Images / NYU Depth V2 dataset
Outdoor Images / KITTI dataset



Error Comparison of Humans and AIs

■ AI Models: 69 pre-trained AI models

- We employed and trained 69 pre-trained AI models for physical depth estimation based on prior work.
- **The models vary in training strategies, datasets, and architectures.**

■ Metrics: Partial correlation (PC)

- We removed the effect of physical depth values and **compared the error patterns between humans and AI models.**

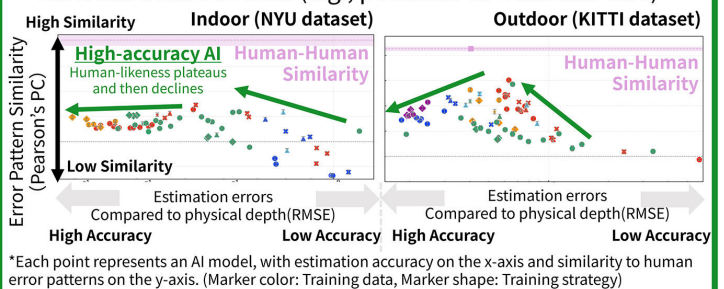
Estimates of an AI model

High PC = High Similarity

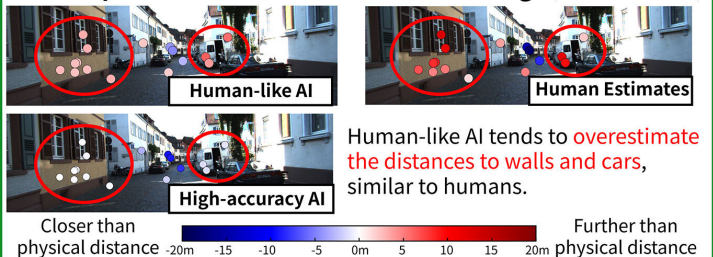
Estimates of Humans

Model Accuracy and “Human-likeness”

- **Many AI models exhibited estimation biases similar to those of humans.** (e.g., depth compression effect)
- We found **the trade-off between accuracy and human-likeness**: highly accurate AI models exhibit tendencies different from humans. (e.g., positions of walls and cars)



■ Examples estimations for an outdoor image (KITTI dataset)



References

- [1] Y. Kubota, T. Fukiage, “Human-like monocular depth biases in deep neural networks,” *PLOS Computational Biology*, Vol. 21, No. 8, e1013020, 2025.
- [2] Y. Kubota, T. Fukiage, “Accuracy does not guarantee human-likeness in monocular depth estimators,” *arXiv*, 2512.08163, 2025.
- [3] Y. Kubota, T. Fukiage, “Benchmarking human and DNN biases in monocular depth estimation,” under review, 2026.

Contact

Yuki Kubota, Sensory Representation Research Group, Human Information Science Laboratory

Abstract

Augmented reality (AR) can easily reproduce the motion and deformation of objects; however, assigning convincing material qualities to those objects is often difficult through simple physical simulation alone. **In this study, we present an approach for conveying material perception in AR by leveraging visual illusions based on apparent surface properties.** We developed methods to represent softness and viscosity through visual cues. For the illusion of softness, the perceived compliance is controlled by adjusting the amount of indentation and the range of deformation in AR space. For the illusion of sliminess, viscosity is expressed by modulating the deformation of the region presented between the user's fingers. Our interface enables rich material perception using vision alone, without the need for haptic devices, significantly enhancing both the realism and expressive flexibility of AR experiences. This approach has the potential to enrich everyday scenarios, such as supporting perceptual evaluation during online product selection.

Material perception transmission via visual media

Challenges

- Methods for conveying material qualities remotely have not yet been established.
- Display devices for presenting material qualities (e.g., 3D displays, haptic devices) are not widely adopted.

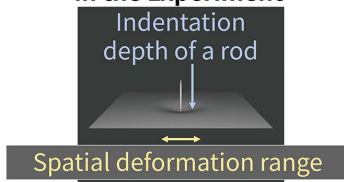
Approach

- Scientifically understand visual information related to material perception and use it to convey material qualities.

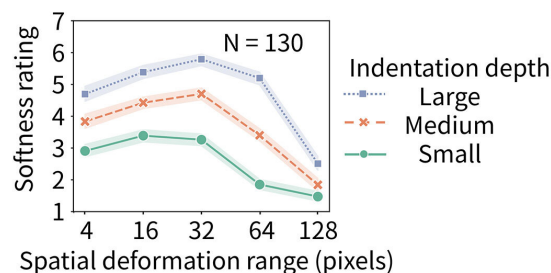


Illusion of softness: Perceived softness is determined by both the amount of indentation and the range of deformation

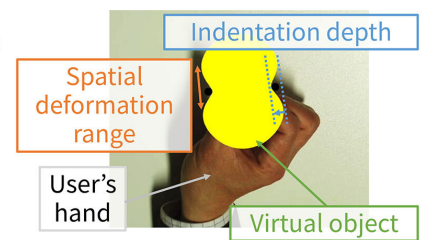
Examples of Stimuli Used in the Experiment



Experimental Results

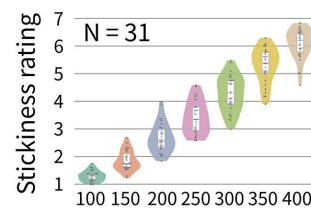


Applied to expressing softness of virtual objects in augmented reality



Illusion of stickiness: Greater break distance between the fingers leads to stronger perceived stickiness.

When the fingers are separated, the region between the fingers stretches, then breaks.



Distance until the stretched region breaks (pixels)

References

- [1] T. Kawabe, Y. Ujitoko, "Visual features involved in determining apparent elasticity elicit touch desire," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 31(10), pp. 9530 – 9536, 2025.
- [2] T. Kawabe, T. Morisaki, Y. Ujitoko, "Unbreakable bond: induced viscosity between the fingers," *The 47th European Conference on Visual Perception*, 2025.

Contact

Takahiro Kawabe, Sensory Interface Research Group, Human Information Science Laboratory

Abstract

Chaotic behavior such as that observed in weather systems can be studied by approximating time as a discrete variable. **Parabolic parameters** are key quantities that govern chaos in such discretized settings; they are roots of polynomials with rational coefficients known as **Delta factors**. Recent progress on the number-theoretic properties of Delta factors is presented in this work. Except for trivial cases, it has long remained unresolved whether Delta factors become polynomials with integer coefficients after specific transformations, or whether they admit nontrivial factorizations. By uncovering a new connection between parabolic parameters and Euler's totient function from number theory, this study provides a breakthrough on these problems. The research area linking complex dynamics and arithmetic properties of numbers is known as arithmetic dynamics and has developed rapidly over the past two decades. We aim to further elucidate number-theoretic phenomena arising in dynamical systems through a bidirectional use of methods and insights from both fields.

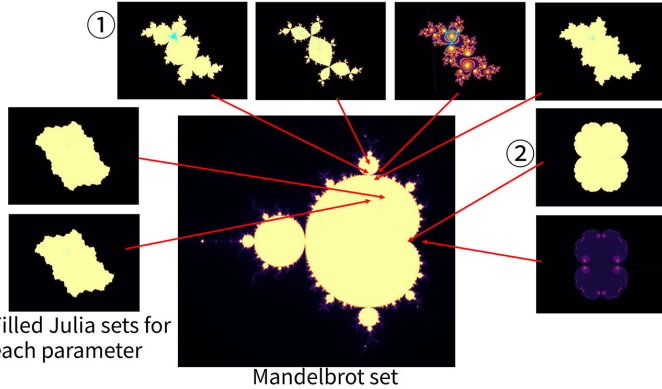
Iteration of polynomial

c : complex parameter

Iteratively substitute a complex number into a polynomial $f_c(x) = x^2 + c$. Is the orbit periodic, convergent, or divergent?

Example: The f_c -orbit of 0 for $c = \frac{1}{4}$
 $0 \mapsto \frac{1}{4} \mapsto \frac{5}{16} \mapsto \frac{89}{256} \mapsto \frac{24305}{65536} \mapsto \dots \frac{1}{2}$

Filled Julia set $K_c = \{\text{complex number } x \text{ with bounded } f_c\text{-orbit}\}$
Mandelbrot set $M = \{\text{complex parameter } c \text{ with connected } K_c\}$
 $= \{\text{complex parameter } c \text{ such that } f_c\text{-orbit of } 0 \text{ is bounded}\}$



For a parameter c in the interior of the Mandelbrot, the f_c -orbit of 0 converges to a periodic orbit. The period of this orbit is constant on each domain. Across boundaries such as the "base of a cactus" (e.g., ①) or the "bottom of an indentation" (e.g., ②), the period changes by an integer multiple, leading to a drastic change in the behavior of the dynamical system

Parabolic parameters

The parameters at the "base of a cactus" and the "bottom of an indentation" are called **parabolic parameters**.

Parabolic parameters play a central role in studies of the Mandelbrot set and complex dynamics. Though geometrically special, they often exhibit unique number-theoretic features that may link dynamics and arithmetic.

Delta factors

Fix a positive integer m and a multiple n of m . The parabolic parameters c lying between

the "period- m component" and the "period- n component" are given as the roots of a polynomial $\Delta_{n,m}(c)$ with rational coefficients [4]. These polynomials are called **Delta factors**.

degree	(n, m)	$\Delta_{n,m}(C/4)$ for $f_c(z) = z^2 + c$	
1	(2, 1)	$C - 1$	
	(2, 1)	$C + 3$	
	(4, 2)	$-C - 5$	
	(3, 3)	$C + 7$	
2	(3, 1)	$C^2 + C + 7$	
	(4, 1)	$C^2 - 2C + 5$	
	(6, 1)	$C^2 - 3C + 3$	
	(6, 2)	$C^2 + 9C + 21$	
	(8, 2)	$C^2 + 8C + 17$	
	(12, 2)	$C^2 + 7C + 13$	
3	(6, 3)	$C^3 + 8C^2 + 18C + 81$	
	(4, 4)	$C^3 + 9C^2 + 27C + 135$	
	(4, 4)	$C^3 + 9C^2 + 27C + 135$	
4	(5, 1)	$C^4 + C^3 + C^2 - 9C + 31$	
	(8, 1)	$C^4 + 2C^2 - 16C + 17$	
	(10, 1)	$C^4 - 3C^3 + 9C^2 - 17C + 11$	
	(12, 1)	$C^4 + 2C^3 - C^2 - 14C + 13$	
	(10, 2)	$C^4 + 17C^3 + 108C^2 + 313C + 341$	
	(16, 2)	$C^4 + 16C^3 + 96C^2 + 256C + 257$	
	(20, 2)	$C^4 + 15C^3 + 85C^2 + 215C + 205$	
	(24, 2)	$C^4 + 16C^3 + 95C^2 + 248C + 241$	
	5	\emptyset	\emptyset
		\emptyset	\emptyset

When $\Delta_{n,m}$ is defined, the following conjecture was proposed.

Conjecture [Morton-Vivaldi, 1995]

- (1) Let $C = 4c$. Delta factors are polynomials in C with integer coefficients and \pm monic.
- (2) All delta factors are irreducible (i.e., they admit no nontrivial factorizations).

Results of this work

Since the conjecture was proposed, the problem has remained largely

unexplored, except for the trivial cases $m = 1, 2$.

In this study, we achieve a breakthrough by discovering a new relationship between parabolic parameters and Euler's totient function.

Our Theorem (1) Conjecture 1 is true [2],[5]

(2) Conjecture 2 is true for $m = 3$. [3]

(3) Only when $\Delta_{n,m}$ has degree d ($d = 1, 2$), it is divisible by a degree- d polynomial with rational coefficients.

Moreover, this occurs only for the following pairs (m, n) .

$d = 1$: $(n, m) = (1, 1), (2, 1), (4, 2), (3, 3)$ [1],[5]

$d = 2$: $(n, m) = (4, 1), (8, 2), (3, 1), (6, 1), (6, 2), (12, 2)$ [5]

References

[1] X. Buff, S. Koch, "Totally real points in the Mandelbrot Set," preprint, arXiv: 2211.15725, 2022, under review.
 [2] V. Huguin, "Unicritical Polynomial Maps with Rational Multipliers," *Conformal Geometry and Dynamics*, Volume 25, pp. 79-87, 2021.
 [3] J. Koizumi, Y. Murakami, K. Sano, K. Takehira, "Irreducibility of polynomials defining parabolic parameters of period 3," *Acta Arithmetica*, Volume 221, pp. 253-270, 2025.
 [4] P. Morton, F. Vivaldi, "Bifurcations and discriminants for polynomial maps," *Nonlinearity* 8, No.4, pp. 571-584, 1995.
 [5] Y. Murakami, K. Sano, K. Takehira, "Arithmetic properties of multiplier polynomials for certain polynomial maps," preprint, arXiv:2403.17315v3, 2024, under review.

Contact

Kaoru Sano, NTT Institute for Fundamental Mathematics/ Computing Theory Research Group, Media Information Laboratory

Abstract

Cross-modal embeddings, which map images and texts into a shared space, enable cross-modal retrieval. However, “hub” embeddings that exhibit spuriously high similarity to many queries regardless of true relevance are widely observed and degrade retrieval reliability. To analyze the nature of hubs, we propose a method for **identifying “hub texts,”** which show **unreasonably high similarity to many unrelated images**. We demonstrate that these texts significantly degrade retrieval performance in practice. This identification is **essential for understanding the behavior of hub texts and is a key step toward mitigating their impact**. Despite recent advances in AI, including embedding models, reliability remains an open challenge. In particular, the conditions under which models exhibit unexpected behavior are not yet well understood. Our findings contribute **to a deeper understanding of model behavior and reliability in modern AI systems**.

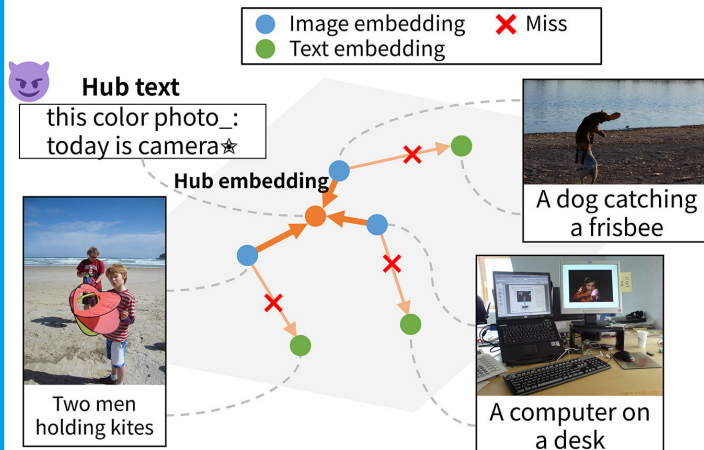
Embedding

Represent data as their feature vectors

- Vector distances represent similarity
 - Closer vectors indicate higher similarity
 - Enables comparisons across modalities
 - e.g., image ↔ text
- Widely used in information retrieval, etc.

Problem in embedding: Hubness

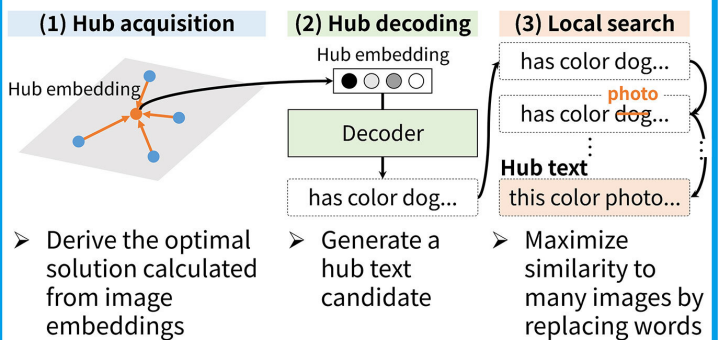
- Hub embeddings: Embedding vectors that exhibit high similarity to many unrelated items
- Hub texts: Texts that are mapped to hubs
 - When included in a database, they consistently lead to irrelevant search results regardless of the query



- Issues: The causes and fundamental properties of hubs remain unclear
 - **What kinds of vectors become hub embeddings?**
 - **What kinds of texts become hub texts?**
 - Naive identification method requires exhaustively checking all possible texts, which is infeasible

Hub text identification

- Objective: Understand the fundamental properties of hubness by identifying hub texts
- Proposed method: Derive hub embeddings and invert them into hub texts



- Derive the optimal solution calculated from image embeddings
 - Generate a hub text candidate
 - Maximize similarity to many images by replacing words
- Contribution: First identification of hub texts in cross-modal embeddings
 - Hub texts show higher similarity than human captions for up to 90% of images

Examples from image-text embedding model “CLIP”

	Text	Score	Image
Human caption	<i>Two dogs are playing on the beach catching a Frisbee.</i>	65.7%	
Hub text	<i>today color photo_ : dishstaged mms middle], croc ée * trot maker gely bw 8 oarded<U+FE0F>: garethapproached cision</i>	70.0%	
Human caption	<i>Two computers are sitting on top of the desk.</i>	62.8%	
Hub text	<i>today color photo_ : dishstaged mms middle], croc ée * trot maker gely bw 8 oarded<U+FE0F>: garethapproached cision</i>	69.5%	

- Future direction: Analyze identified hub texts and develop methods to mitigate their impact

References

- [1] H. Deguchi, K. Chousa, Y. Sakai, “One Single Hub Text Breaks CLIP: Identifying Vulnerabilities in Cross-Modal Encoders via Hubness,” in Proc. The 64th Annual Meeting of the Association for Computational Linguistics (ACL2026), 2026. (to appear)
- [2] H. Deguchi, K. Chousa, Y. Sakai, “Hacking Neural Evaluation Metrics with Single Hub Text,” in Proc. The 19th Conference of the European Chapter of the Association for Computational Linguistics (EACL2026), pp. 198-206, 2026.

Contact

Hiroyuki Deguchi, Linguistic Intelligence Research Group, Innovative Communication Laboratory

How is envy depicted in picture books?

Abstract

Envy can motivate self-improvement, but it can also lead to low self-esteem and interpersonal problems. It is particularly **important to teach children how to handle envy appropriately**. However, adults are unlikely to express envy in daily conversation, making it difficult for children to acquire the language used to describe it. **We examined how picture books use words related to envy** and evaluated their effectiveness as tools for emotional education. Previous research primarily focused on analyzing the frequency of emotion words. Our database of full-text picture books enables us to **analyze their contexts, such as events that trigger envy**. Our findings revealed that the various situations in which envy arises in daily life are also depicted in picture books. **Picture books help children develop emotional understanding and interpersonal skills** by allowing them to imagine characters' feelings while learning to express and respond to their own.

The acquisition of emotion words

Innate emotions

Basic Emotions

Sadness
Happiness
Anger
etc.

Emotions arising from the presence of others

Social Emotions

Envy
Pride
Guilt
etc.

The unpleasant emotion that can occur when an individual lacks something that another person has.

- Acquiring words that express social emotions such as **envy** is important for maintaining one's mental health and interpersonal relationships.
- Picture books are rich in emotion words.



Can children learn words that express envy and the situations in which to use them through picture books?

Analyses of how envy is depicted in picture books

Analysis 1: Frequency of words expressing envy in the database of 8,860 picture books

	Frequency (Token)	Volumes (Types)
Urayamashii	309	225
iinaa	211	149
Zurui	58	50
Netamashii	25	22
Shitto	8	5
Senbo	1	1

The words *iinaa* and *zurui* have multiple meanings besides envy

Analysis 2: The target object of envy in sentences containing the word *urayamashii*

	Frequency (Token)	Volumes (Types)
Possession	59	19.1%
Experience	54	17.5%
Status, life, and/or money	49	15.9%
Interpersonal relationships and/or family	48	15.5%
Ability, talent, and/or personality	39	12.6%
Temporary food	26	8.4%
Physical attractiveness	22	7.1%
Introduction of the word or proverb	11	3.6%
Others	1	0.3%
Total	309	100.0%

Picture books contain a variety of words that express envy and the contexts in which they are used, suggesting that they may be useful for emotional education.

References

[1]A. Nakai, S. Fujita, N. Watanabe, T. Hattori, T. Kobayashi, "How is envy depicted in picture books?: An examination of the frequency and context of envy-related words," Proceedings of the 33rd annual meeting of Japan Society for Research on Emotions, PS1-21, 2025. (in Japanese)

Contact

Ayaka Nakai, Child Development Research Group, Innovative Communication Laboratory

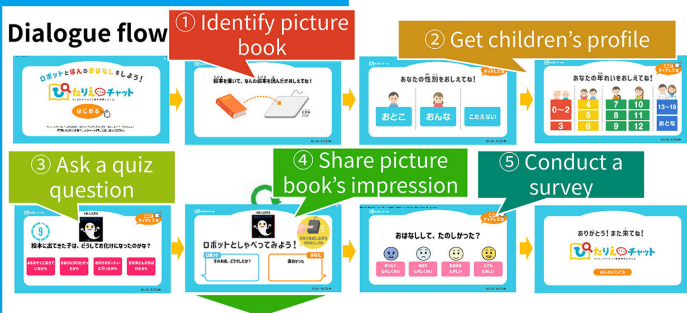
Abstract

Dialogues about picture books are known to enhance children’s interest in reading and support their language development. However, parents around them cannot always secure sufficient time for such interactions. To address this issue, we are developing a dialogue AI system, “PitarieChat,” that enables children to discuss their impressions of picture books. By using a large-scale picture book and children’s book corpus constructed by NTT, we achieved empathetic dialogue that elicits children’s impressions and enhances their motivation to read. Furthermore, by adopting tsuzumi 2 for the dialogue engine, the system enables natural conversations grounded in the attributes of the robot and the user, as well as the content of the picture books. In future work, we aim to develop personalization methods based on children’s age and individual characteristics, as well as dialogue control techniques to realize a conversational AI that enables each child to enjoyably develop their thinking and expressive abilities.

PitarieChat

- Conversations about picture books not only enhance children’s interest in reading but also contribute to their language and emotional development
 ⇒ Securing sufficient time is not easy
- Constructs PitarieChat for talking about picture book impressions
 - Utilizes NTT’s proprietary picture book corpus
 - 9,000 books, with ongoing expansion
 - Employs a robot† that engages children in dialogue to encourage spontaneous impressions about picture books, shows empathy, and deepens their reflections

System overview



Implemented two types of systems

- Scenario-based**
 - Rule-based empathetic dialogue
 - Small language model (SLM)-based chat-chat
 - Uses Japanese Dialogue Transformers
- LLM-based**
 - Empathetic dialogue grounded in picture book content using a large language model (LLM)
 - Uses NTT’s LLM tsuzumi 2
 - Basic information is provided in the prompt:
 - Picture book title and text
 - Brief profiles of the robot and the user

Field experiments

- Iwata City Childcare Support Center**
 - Scenario-based
 - 2023/10/11-2025/4/3 (about 18 months)
 - Number of users: 5,970*
 - Okinawa Prefectural Library**
 - LLM-based
 - 2025/10/27-2025/11/3 (7 days)
 - Number of users: 148人*
- *Only children aged 12 or under were counted.

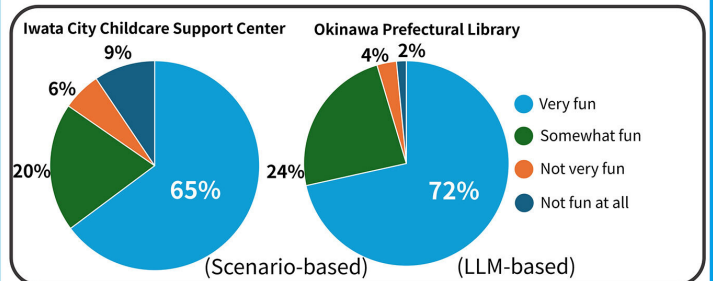


Dialogue example (LLM)

Robot: Okay, let's talk about *Harry the Dirty Dog*. How was the story?
User: It was fun.
Robot: It was fun seeing Harry get all muddy and go on lots of adventures! What part did you like the most?
User: The part where he got muddy.
Robot: You liked the part where Harry got all muddy and played a lot the best. Playing in the mud like Harry is really fun, isn't it? Let me know if there were any other scenes you liked.
User: The part where he became clean again.
Robot: The part where Harry came back clean was nice too. When everyone brushed and washed him, his white fur showed again, and he looked really happy.

User evaluations

Q. Did you enjoy talking with me?



Q. Want to read more picture books?

Very much (Scenario-based) **63%** (LLM-based) **79%**

- Improvements in response performance by the LLM significantly increased ratings of enjoyment and reading motivation (χ^2 test, $p < 0.001$).
- In future work, we plan to investigate personalization methods based on children’s age and individual characteristics.

† Sota: <https://www.vstone.co.jp/products/sota/>

References

[1] NTT West, Inc., “Exhibiting an AI-powered system at the Okinawa Prefectural Library to support new encounters with picture books – Toward fostering reading habits among local children –,” 2025, https://www.ntt-west.co.jp/newscms/okinawa/14861/1_20251024.pdf (in Japanese).
 [2] Mynavi News, “NTT West, Inc. exhibits AI reading impression dialogue system “PitarieChat” in Okinawa,” <https://news.mynavi.jp/article/20251107-3624023/> (in Japanese).

Contact

Yuya Chiba, Interaction Research Group, Innovative Communication Laboratory

Abstract

Empathy involves both **cognitive empathy**, understanding another person’s feelings, and **affective empathy**, physically resonating with them, yet their relationship remains poorly understood because datasets capturing both are scarce. In this study, we **simultaneously measured subjective ratings of emotional intensity and physiological responses** such as heart rate **from both senders and receivers**, and analyzed their links. To enable this, we built a **unique multimodal dataset** containing time-series ratings and physiological signals from both actors and observers. Our analyses showed that **receivers with a stronger perspective-taking tendency understood emotions more accurately**, while their physiological responses were **less likely to synchronize with those of the sender**; this pattern also varied by emotion type. By separately measuring these two forms of empathy, this work could support dialogue assistance, remote stress-sign detection, and burnout support.

Two types of empathy

Cognitive empathy

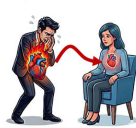
Understanding another person’s emotion with the mind



Sender Receiver

Affective empathy

Another person’s emotion “spreads” to the body



Sample images created by generative AI

Their distinct mechanisms require separate treatment.

Dataset constructed for a multifaceted understanding of two types of empathy



Dataset HP

Sender (empathizee)

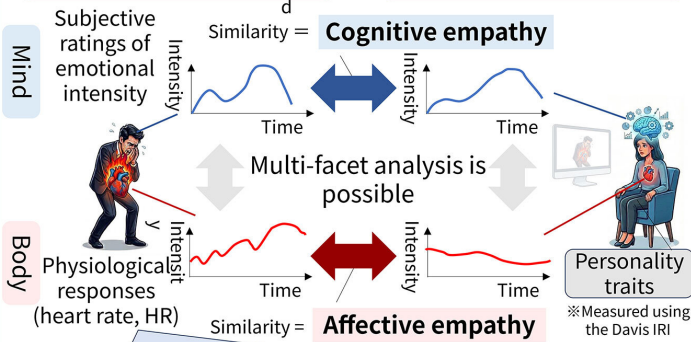
- 8 professional actors
- 7 types of emotions
- 56 videos, each 1–3 min long



Sample image created by generative AI. Real videos were actually used. Publish

Receiver (empathizer)

- Experiment 1**
- 12 participants (6 per video)
 - Emotion category unknown
- Experiment 2**
- 87 participants (43 per video)
 - Emotion category known
- To be released in the future



Advantages of using actors

Strong emotions can be reproduced more stably, and clear sender-side ratings and physiological responses can be obtained, making it easier to capture differences between the two types of empathy and individual differences among receivers.

What the data analysis revealed

Overall Trend

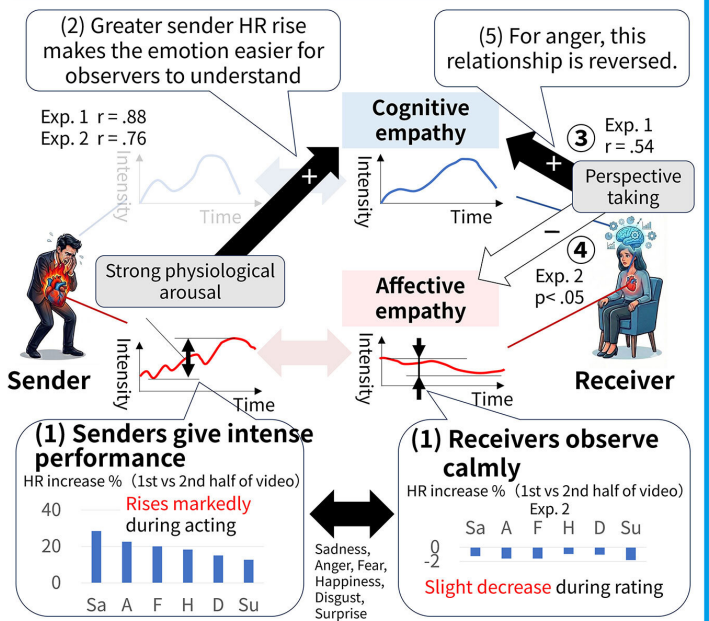
Senders’ intense performance and the receivers’ calm observation support accurate emotion understanding

- (1) Senders’ HR increases during acting, while receivers’ HR slightly decreases during rating.
- (2) More intense sender performance leads to more accurate emotion understanding in receivers.

Influence of Individual’s Personality Traits

Trade-off between “understanding” and “catching” emotions. Receivers with a stronger tendency to take another person’s perspective (measured by questionnaire):

- (3) understand emotions more accurately, while
- (4) their physiological responses differ more from those of the sender. However,
- (5) for anger, this relationship is reversed.



References

[1] A. Ota, S. Kumano, A. Murata, A. Nakane, S. Shimizu, “EMPAC: A Multimodal Dataset for Bridging Affective and Cognitive Empathy,” *Frontiers in Psychology* (under review), *bioRxiv*.
 [2] A. Ota, H. Kuze, A. Nakane, S. Kumano, A. Murata, S. Shimizu, “Evaluating Empathy from Cognitive and Affective Perspectives Using Empathic Video Stimuli,” *The Institute of Electronics, Information and Communication Engineers HCG Symposium 2024*, 2024.

Contact

Airi Ota, Digital Twin Computing Laboratory, Human Informatics Laboratories, and Shiro Kumano, Sensory Resonance Research Group, Human Information Science Laboratory

Abstract

Lower-limb motor control is closely related to health and fall risk, yet its quantitative assessment typically requires **expensive and specialized equipment**. Here, we present a simple method for evaluating foot dominance and postural stability using the **variability of circular leg movements measured with a smartphone**. Participants performed fifteen-second circular leg movements while the smartphone recorded acceleration data from which movement variability was quantified. The variability measure showed good agreement with conventional measures of postural stability obtained using dedicated equipment. Compared with the conventional approach, the proposed method is **low-cost and halves measurement time**. Our work contributes to the development of practical techniques for estimating motor function from everyday body movements, supporting a future in which individuals can **monitor physical ability, health status, and fall risk** using widely available devices.

Importance of postural stability

- Walking is essential, and age-related falls are a major issue
- To estimate the risk of falling, postural stability (ability to stand upright) must be quantified



Postural stability is not easy to measure

Conventional measure of postural stability

- Using a force plate, the area of the body's Center of Pressure (COP) is calculated
- Issues
 - Equipment is expensive
 - Multiple 30-60s measurements are required
 - Significant barriers to wide adoption



Expensive force plate

Circular movement variability method

Proposed method

- Smartphone strapped to the leg records variability during 15s circular movement
- Variability of leg movement was compared to conventional measure of postural stability

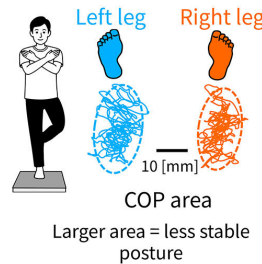


※ Each leg was measured three times

Smartphone

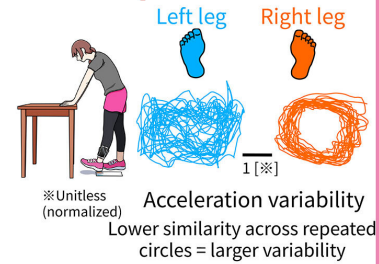
Leg movement variability can be used to assess postural stability

Conventional method



Larger area = less stable posture

Proposed method

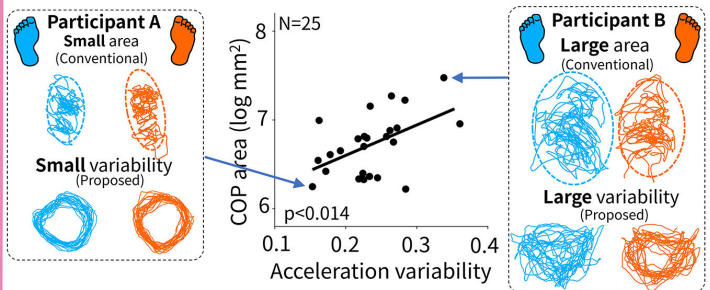


※ Unitless (normalized)

Acceleration variability
Lower similarity across repeated circles = larger variability

Larger acceleration variability indicates less stable posture

Acceleration variability and COP area were significantly related



Proposed method more reliable than conventional one

- Intra-class correlation analysis showed higher repeatability for the proposed method in both legs.

※ Mean (95% confidence interval) ※ More repeatable if closer to 1	Intra-class correlation coefficient	
	Non-dominant	Dominant
COP area (Conventional)	0.62 (0.40-0.79)	0.48 (0.25-0.70)
Acceleration variability (Proposed)	0.84 (0.72-0.92)	0.80 (0.65-0.90)

Summary

Relative to conventional method, the proposed one is:

- ✓ Fast
- ✓ Affordable
- ✓ More reliable

References

[1] A. Takagi, N. Abekawa, "Variability of circular leg movements is related to footedness and postural stability," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 34, pp. 198–204, 2026.

Contact

Atsushi Takagi, Sensory and Motor Research Group, Human Information Science Laboratory

Abstract

Improving competitive performance requires objective evaluation of the mental states and levels of expertise that lead to victory. However, lacking sufficient scientific support, conventional assessments still depend heavily on coaches' experience and intuition. This study measured athletes' physiological states during esports competition and examined whether brain activity could predict match outcomes, as well as differences in physiological responses between advanced and intermediate players. We found that **pre-match electroencephalograms could predict match outcomes with about 80% accuracy**, that **advanced players showed increased muscle activity at decisive moments**, and that **pronounced cardiac synchronization emerged when advanced players competed against one another**. These findings enabled us to quantify, for the first time, the internal bodily states that influence winning and losing, and to establish a new metric for evaluating depth of expertise. This study may support **evidence-based training**, **skill assessment**, and **mental-control technologies** applicable to medicine, construction, and other high-pressure settings.

Understanding winning players' brains and bodies

- In top-level matches between evenly matched players, the mental factor is often said to decide the outcome
- Although this factor has been studied from many dimensions, what happens in the brain and body remains unclear

Focusing on esports

- More dependent on mental than physical
- Accurate bio-signal measurement during matches

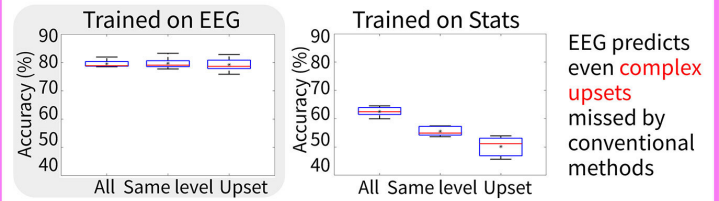
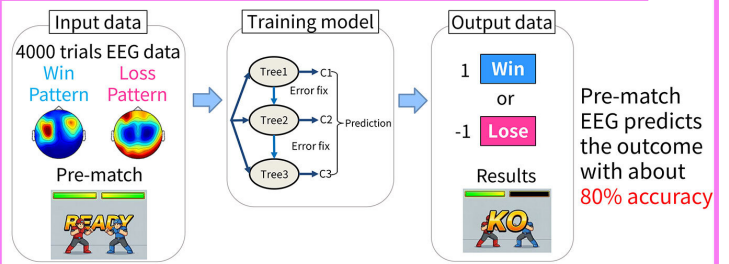


Our study

- Physiological recording during fighting video game play
- We tested whether **pre-match EEG¹ predicts Win/Lose**
- We examined mental-related **EMG² and HR³ dynamics** that **distinguish advanced from intermediate players**

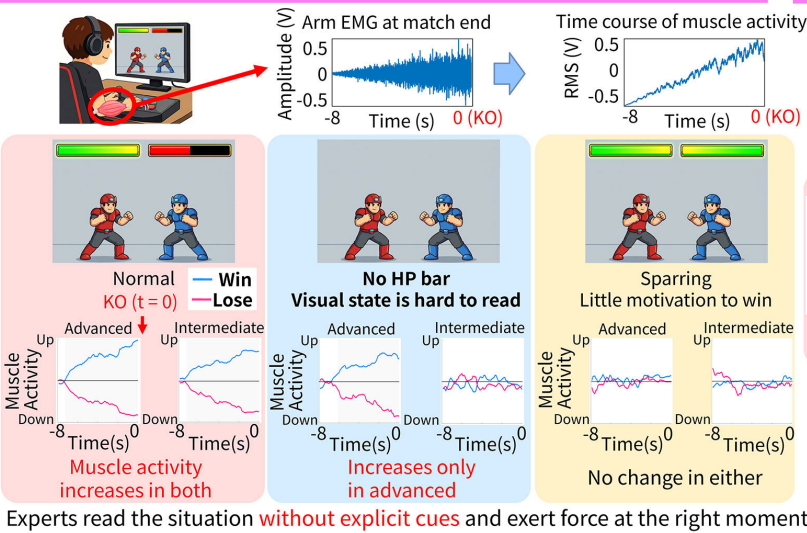
1: Electroencephalogram, 2: Electromyography, 3: Heart rate

Revealing the optimal brain state for winning

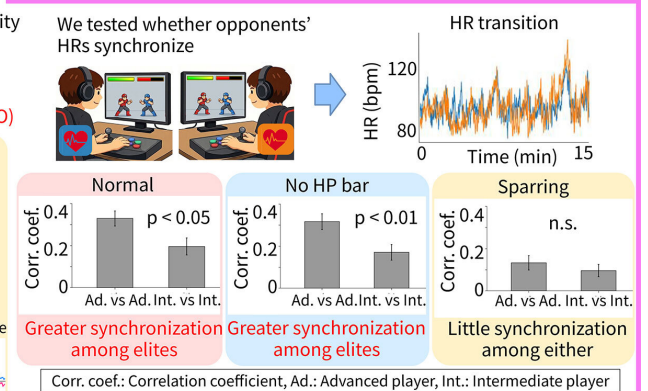


An optimal pre-match brain state helps decide the match results

Muscle activity reveals how experts read key moments



HR synchronizes more easily in elite matches



Clutch ability and expertise level can now be measured objectively

References

[1] S. Minami, H. Koyama, K. Watanabe, N. Saijo, M. Kashino, "Prediction of esports competition outcomes using EEG data from expert players," *Comput. Hum. Behav.*, Vol. 160, 108351, 2024.
 [2] S. Minami, K. Watanabe, N. Saijo, M. Kashino, "Task context and player expertise modulate arm EMG linked to win-loss outcomes in esports," *Comput. Hum. Behav.*, Vol. 175, 108850, 2026.
 [3] K. Watanabe, N. Saijo, S. Minami, M. Kashino, "The effects of competitive and interactive play on physiological state in professional esports players," *Heliyon*, Vol. 7, No. 4, e06844, 2021.

Contact

Sorato Minami, Embodied Intelligence Research Group, Human Information Science Laboratory

Abstract

We investigated the mind–body states that support winning in elite athletes by examining how pre-race hormonal states relate to performance in professional Formula car drivers in real competitions. **Hormonal profile linked to high performance varies across individuals**, suggesting that there is no single optimal competitive state for all athletes. Using field data collected at actual race events, we also established a practical method to **predict and adjust competition-relevant hormone states**. In one case, where lower cortisol was associated with better performance, a short pre-race nap successfully reduced the athlete's cortisol level. **These findings may enable individualized psychophysiological adjustment strategies** for high-pressure performance settings with potential benefits for athletic performance and coaching theory.

Hormonal States That Support Winning in Elite Athletes

Bold and **cautious** behaviors can affect competitive outcomes, and hormones help shape them [1,2].



Testosterone: A male hormones associated with **bold behavior (risk-taking)**.



Cortisol: A stress-related hormone associated with **cautious behavior (risk avoidance)**.

However, the optimal hormone states for high performance remain unclear, and practical adjustment methods are still lacking.

Our Study To examine hormone–performance relationships in competitions and explore an on-site adjustment method

Athlete-Specific Hormonal States Linked to High Performance in Real Competitions

Study Overview

We focused on motorsport, in which boldness and caution can affect performance, and collected data from **two elite professional drivers in the Super Formula, the highest tier of formula racing in Japan**.

Hormone Analysis

Saliva samples were collected before driving, frozen, and subsequently analyzed for hormone concentrations.



← Saliva collection before driving.

Performance Analysis

Lap time improvement between the two qualifying runs was analyzed.



Course Record : T_r [s] Best Time in Each Qualifying Session Time Relative to the Course Record

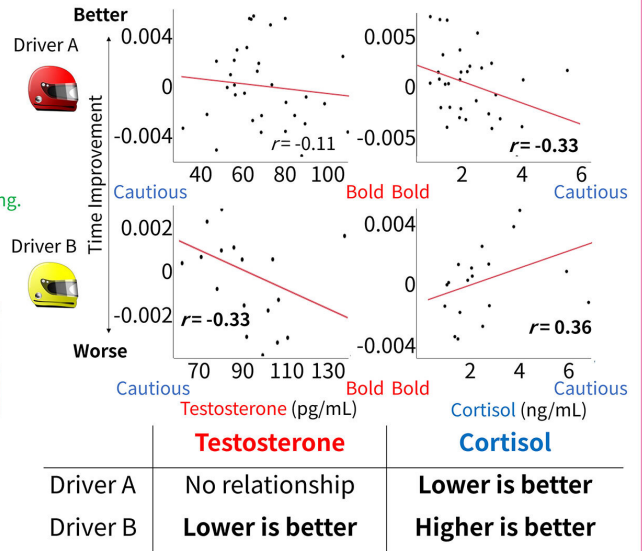
Qualifying 1	All Drivers (n = 20–22)	T_1 [s]	→	$S_{p1} = T_r / T_1$
Qualifying 2	Top 12 Drivers in Qualifying 1	T_2 [s]	→	$S_{p2} = T_r / T_2$



↑ Performance was evaluated using best lap time.

Time Improvement: $S_{p2} - S_{p1}$ adjusted for environmental factors

* Positive values indicate improved performance on Qualifying 2.



Predicting and Adjusting Hormone States in Competition Settings

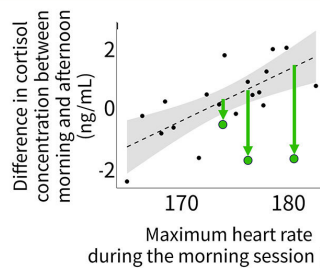
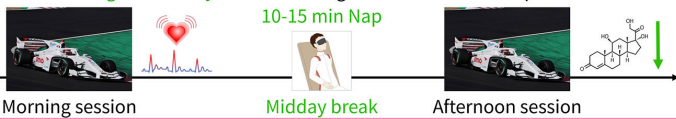
Case Study in Driver A: Lower Cortisol Linked to Better Performance

Prediction of afternoon cortisol:

Afternoon cortisol changes were predicted from the maximum heart rate during the morning session.

On-site adjustment:

Based on evidence that napping can reduce cortisol levels[3,4], napping was introduced during the midday break when high cortisol levels were predicted.



- Higher morning max heart rate predicted higher afternoon cortisol.
- **The nap successfully lowered the cortisol level below the predicted value.**
- : Nap before the afternoon race

References

[1] J. G. Johnson, “Decision making in sports,” *Psychology of Sport & Exercise*, Vol. 80, 102919, 2025.
 [2] P. H. Mehta, K. M. Welker, S. Zilioli, J. M. Carré, “Testosterone and cortisol jointly modulate risk-taking,” *Psychoneuroendocrinology*, Vol. 56, pp. 88–99, 2015.
 [3] A. N. Vgontzas, S. Pejovic, E. Zoumakis, H. M. Lin, E. O. Bixler, M. Basta, J. Fang, A. Sarrigiannidis, G. P. Chrousos, “Daytime napping after a night of sleep loss decreases sleepiness, improves performance, and causes beneficial changes in cortisol and interleukin-6 secretion,” *American Journal of Physiology - Endocrinology and Metabolism*, Vol. 292, No. 1, pp. E253–E261, 2007.
 [4] J. K. Devine, J. M. Wolf, “Determinants of cortisol awakening responses to naps and nighttime sleep,” *Psychoneuroendocrinology*, Vol. 63, pp. 128–134, 2016.
 [5] Y. Ooishi, S. Matsumura, M. Kashino, N. Saijo, “Role of testosterone and cortisol in formula car drivers for achieving high performance in a real competitive situation,” in *Proc. FENS Forum 2024*, 2024.

Contact

Seiji Matsumura, Embodied Intelligence Research Group, Human Information Science Laboratory

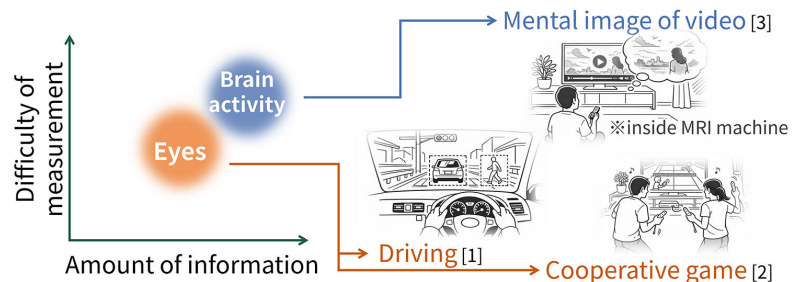
Abstract

Humans act in rich and complex everyday environments, but it is still difficult to understand how the brain works in such situations and whether a person's mental state can be estimated from natural behavior. This exhibit introduces a series of studies that **extend mind reading beyond the laboratory by using eye and brain signals measured during daily activities such as driving, cooperative gameplay, and video viewing**. Based on knowledge from controlled laboratory experiments, we show that it is **possible to estimate higher-level mental information in more realistic settings**. Examples include detecting attentional state and concentration during natural driving, finding cooperative states from eye-related signals during team play, and decoding perceived or imagined video content from brain activity. In the future, this approach may support **better communication and adaptive assistance by adjusting information presentation and interaction to each person's state**. In driving, for example, it may help driver monitoring systems detect inattentive states.

Mind Reading in the wild

Building on methods from conventional laboratory experiments

- Can we read good and bad states from the **eyes**, which are relatively easy to measure?
- Can we also decode nonverbal cognitive content by measuring **brain activity**?



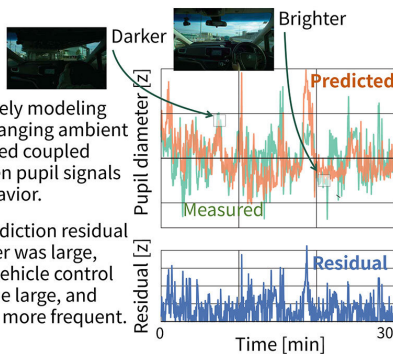
Reading unsafe driving states from the eyes [1]

Coupled patterns in pupil signals and driving behavior

Joint study with DENSO COROPRATION

- We identified pupil-based markers linked with unstable operation in natural driving.

Vehicle-operation data and pupil diameter during natural driving under safe conditions



After appropriately modeling the effects of changing ambient light, we extracted coupled patterns between pupil signals and driving behavior.

→ When the prediction residual of pupil diameter was large, fluctuations in vehicle control also tended to be large, and gaze shifts were more frequent.

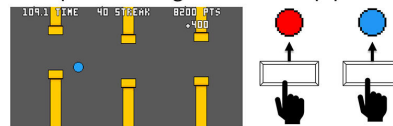
- Useful for improving safety, for example by detecting inattentive driving states

Reading cooperative states from the eyes [2]

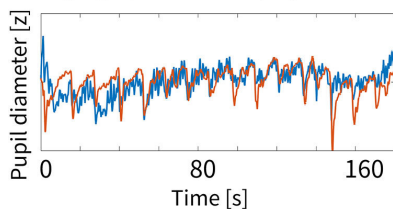
Team performance and synchronization of the eyes

- Pupil diameters of two players in a game requiring cooperation are more strongly synchronized when the team performed better.

Measured pupil diameter while two players took turns controlling a ball in an experimental game to avoid pipes.



Example of strong pupil synchronization during high-performance play



Reading mental content from brain activity [3]

Mind Captioning

- Generating texts that explain contents of viewed/recalled video from brain activity

Viewed/recalled video



Generated text from brain activity during viewing

Two young men and woman are dancing in unison and joking around.

Generated text from brain activity during recall

Four young men are standing together, laughing and doing hand motions.

- Showcased the feasibility of non-verbal thought-based brain-to-text communication, potentially aiding individuals with difficulties in language expression such as aphasia

References

- [1] J. Yamashita, Y. Suzuki, F. Nakanishi, M. Yamataka, H. Terashima, K. Maruya, "Finding Workload in the Wild: Recovering Latent Cognitive Dynamics Jointly from Eye Metrics and Driving Telemetry," in Proc. *The 2026 ACM Symposium on Eye Tracking Research & Applications (ETRA2026)*, accepted.
- [2] H.-I. Liao, M. Montemayor, K. Haly, M. Kashino, S. Shimojo, "Interpersonal pupil synchronization during high-engagement team plays," in Proc. *The 47th European Conference on Visual Perception (EVP2025)*, 2025.
- [3] T. Horikawa, "Mind captioning: evolving descriptive text of mental content from human brain activity," *Science Advances*, Vol. 11, No. 45, eadw1464, 2025.

Contact

Hiroki Terashima, Sensory Representation Research Group, Human Information Laboratory

Abstract

People encounter events that violate physical laws in virtual worlds, yet they do not simply treat them as something they cannot understand. Instead, they make sense of how that world works by drawing on past experiences and evaluations. This study investigates **how people form an understanding of virtual worlds**. We developed a questionnaire scale to measure player experience in video games that contain many non-realistic events, **enabling us to empirically examine the process of understanding**. Using this scale, we found that **experiencing a new game can reshape how previously experienced worlds are evaluated**, a phenomenon we call a retrospective effect. These findings provide **a basis for designing virtual worlds that people can naturally engage with**, supporting more immersive learning and training experiences that promote lasting behavioral change.

What is “Sekai-kan” (a world)?

One’s understanding of how a world works.

■ Non-realistic worlds

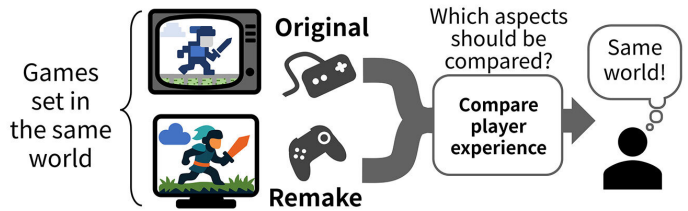
Games and VR offer many diverse worlds beyond reality.
→ Useful for studying how people understand a world.

■ How people understand and create non-realistic worlds

By examining which aspects of game experience make non-realistic worlds feel similar to each other, we aim to clarify how people form an understanding of non-realistic worlds.

- Magic ✨
- Space 🌌
- Horror 👁️
- etc.

■ Player experience in video game remakes
Remakes reconstruct the world of the original.
Comparing player experience in remakes and the original reveals key aspects of world understanding.



Structure of world understanding

■ Based on questionnaire ratings of player experience across a wide range of games, we successfully estimated the underlying structure of how people understand a world.

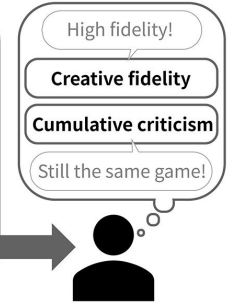
- Development of questionnaire
- Online-survey
- Factor analysis



Aspects compared



World understanding



Finding 1-1

We identified two independent components: “in-world” and “real-world”.

Finding 1-2

Higher “in-world” and “real-world” ratings lead to higher fidelity.

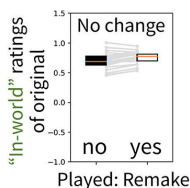
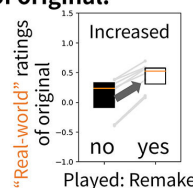
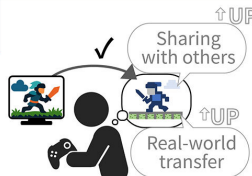
※1 Real-world transfer: Degree to which the experience supports real-world behavior and understanding
※2 Sharing with others: Degree to which the experience can be shared and discussed with others

Updating world understanding

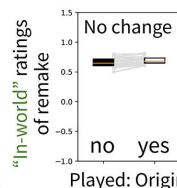
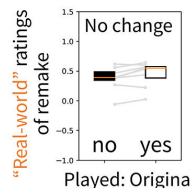
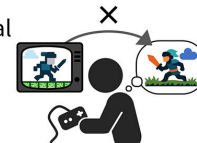
■ Between games set in the same world, how does experience in one affect evaluation of the other?

=For original-remake pairs, does having experience with one work affect evaluation of the other?

Experience with remake increases “real-world” ratings of original.



Experience with original does NOT change ratings of remake.



Finding 2

“Real-world” ratings of previously experienced worlds can be updated through new experiences.

► Evaluation of a world is not fixed but can be reinterpreted afterward.

References

[1] T. Yokosaka, K. Miura, Y. Isogaya, T. Ohtani, K. Maruya, “Factors of player experience in describing the relationship between remade and original works,” 2024 IEEE Conference on Games (CoG), pp. 1–4, 2024.
[2] T. Yokosaka, Y. Isogaya, T. Ohtani, K. Maruya, “Video-Game Retrospective Effect: How Playing Remakes Alters Experience Evaluations for Original Works,” In Companion Proceedings of the Annual Symposium on Computer-Human Interaction in Play, pp. 203–208, 2025.

Contact

Takumi Yokosaka, Sensory Representation Research Group, Human Information Laboratory