NTT

NTT Communication Science Laboratories

OPEN HOUSE 2022

6/2 thu ◎ 3 fri

Jun 2nd 12:00- release

On the Web

Access from: https://www.kecl.ntt.co.jp/openhouse/2022/index_en.html

Welcome to “Open House 2022”

Futoshi Naya
Vice President, Head,
NTT Communication Science Laboratories

The COVID-19 pandemic of these last few years has drastically changed our lifestyles and social activities. Amid these changes, communication between people having diverse values, between people and computers, and between people and society as well as the technologies that support this communication are becoming increasingly important.

At NTT Communication Science Laboratories (CS Labs), we are promoting world-class basic research with the mission of constructing fundamental theories on the essence of human beings and information and creating innovative technologies that will bring about “heart-touching communication.”

CS Labs celebrated the 30th anniversary of its founding in July 2021, and in October 2021, the “Institute for Fundamental Mathematics” was established within our laboratories as a virtual organization for researching fundamental theories of modern mathematics and bolstering the role of mathematics as the “fountain of knowledge” in NTT R&D.

“Open House 2022” will introduce activities at the Institute for Fundamental Mathematics as well as the latest achievements in media processing, data and machine learning (AI), human sciences, and brain science through lecture videos, poster exhibits, and online demonstrations in an easy-to-understand manner.

Unfortunately, our open house this year will again be held online as a countermeasure to the COVID-19 pandemic. Nevertheless, we hope that it will provide opportunities for everyone to search out unknown truths and hold discussions and exchanges toward the creation of an even better society in the future while keeping in mind the dramatic changes now affecting people, society, and the environment. All of us look forward to welcoming many visitors to “Open House 2022.”

Science of Machine Learning

01	Distributed traffic coordination without traffic signals	Learning of collective intelligence via digital twins
02	Efficient training of photonic AI	Accelerated learning of fine-layered optical neural networks
03	Multiple AIs make better predictions	Bayesian ensemble learning for better generalization performance
04	Training fast & lightweight neural networks	Pruning neural networks with iterative randomization
05	How the events spread?	Learning Time-evolving States via Dynamic Hawkes Processes

Science of Communication and Computation

06	Rapid disaster recovery through efficient shelter management	Optimal shelter operation with sequential return of evacuees
07	Translating with your favorite expressions	Lexically constrained neural machine translation
08	Toward uncongested infrastructures under user-equality	Equilibrium optimization of combinatorial congestion games
09	Looking for mistakes in machine translation	Post-editing support based on source-target word alignment
10	Elderly-friendly speaking styles	Using voice and words for better understandability
11	Talking with AIs about views from a vehicle	Casual-dialog system based on scenery and nearby information
12	Toward secure cryptography against quantum attacks	Quantum algorithm for finding collisions of hash functions
13	Where does the wonder of numbers come from?	Finding new arithmetic phenomena via generalized motives

Science of Media Information

14	"Huh? What do you mean?" Summarize a long story short	Robust speech summarization against speech recognition errors
15	Flexible bokeh renderer based on predicted depth	Deep generative model for learning depth and bokeh effects only from natural images
16	Heart health monitoring with sounds and electric signals	Estimating heart activities from multichannel sounds and ECG signals
17	Controlling facial expressions in face image from speech	Crossmodal action unit sequence estimation and image-to-image mapping
18	Maintain comfortable visibility anytime, anywhere	Image blending with content-adaptive visibility predictor

Science of Human

19	Gazing and talking help infants learn	Elucidating effects of social cues on infants' object learning
20	Why do people hesitate to use contact tracing apps?	Social factors influencing adoption of COCOA
21	Is the rising fastball a perceptual illusion?	Modifying pitched ball perception by VR
22	Mental skills of esports experts revealed by brain measurement	The relationship between frontal neural oscillation and performance
23	Unveiling the auditory system with a neural network	Approaches to cochlear implant and binaural processing
24	How does mindfulness meditation reduce stress?	Autonomic and endocrinological variation by meditation style
25	Measuring well-being through diverse aspects	Well-being in terms of mental states, values, and idea of self
26	Faster walking by moving the wall forward	Vision-based speedometer regulates human walking
27	Fingertip illusions direct the mind	How the brain decodes pulling sensations
28	What do we want to touch?	Understanding of desire to touch using large-scale Twitter data
29	Eyes as a window of our mind	Pupil size tracks subjective perceptual changes

01

Learning of collective intelligence via digital twins

Distributed traffic coordination without traffic signals

Abstract

In the era of autonomous vehicles, traffic coordination systems using signals will be replaced. In IOWN's signal-free mobility, it is suggested that vehicles will autonomously transition their states (e.g., speed acceleration, handle steering, and position) via communication among vehicles. For signal-free mobility, a recurrent neural network (RNN) architecture is proposed which alternately iterates (i) communication between closely positioned vehicles (token exchange to prevent vehicle collisions) and (ii) local state updates. Since our method can be performed in a distributed manner, it is suitable to control a large number of vehicles in a city in real-time. Via training through digital twins (simulation system linked with the real world), we will obtain a collective intelligence model. We confirmed the overall efficiency of trained RNN through traffic coordination tests in digital twins and real experiments using real small vehicles.

Goal

The concept of signal-free mobility, in which a set of automated vehicles coordinates their traffic without using traffic signals, is shown in [1]. To realize this concept, we have studied on a distributed control problem to reduce travel/transportation time to the limit while vehicles are collision-free [2].

Constrained dynamics learning

Traffic coordination in which each vehicle updates its states (e.g., speed, position) while imposing constraints on them to prevent collisions can be represented by an ordinary differential equation (ODE).

$$\frac{d\boldsymbol{x}}{dt} = \overbrace{M_1(\boldsymbol{x}, t, \theta, \boldsymbol{A}, \boldsymbol{b})}^{\text{State update in local vehicle}} + \overbrace{M_2(\boldsymbol{x}, t, \boldsymbol{A}, \boldsymbol{b})}^{\text{Communication between vehicles}}$$

By discretizing this ODE, we constructed a recurrent neural network (RNN) in which V vehicles evolve their states K times. As shown in the figure below, this RNN consists of alternatingly repeat of (i) local state updates ($\boldsymbol{x}$), (ii) communication between vehicles to exchange token for satisfying collision-free constraints, and (iii) local updates of input/constraint parameters ($\boldsymbol{A}$, $\boldsymbol{b}$). The size of this RNN is huge with a width of V and a depth of K. However, it is composed of a set of operations that can be parallelized, allowing for real-time state updates as a forward propagation. Meanwhile for backward propagation, driving dynamics model ($\boldsymbol{\theta}$) is optimized to have a small loss score designed to increase the averaged vehicle speed.

Image of process at time t

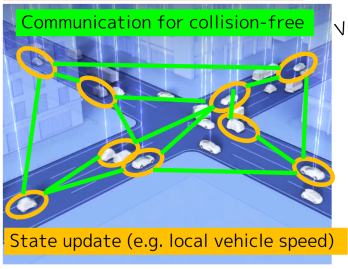
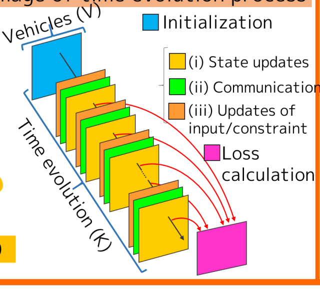

Communication for collision-free
State update (e.g. local vehicle speed)

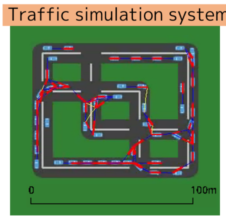
Image of time evolution process



Dynamics model training using digital twins

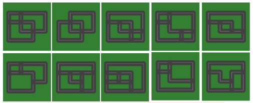
To efficiently train driving dynamics model, we constructed a traffic simulation system that evolves states in digital twins of V vehicles and roads linked to them in real world. By driving digital twins of vehicles on various road maps including virtual ones (see figure below), we can efficiently collect data sets. We optimized driving dynamics mode; though R=300 round iterations of simulation (forward propagations) and backward propagations.

Traffic simulation system

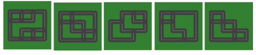


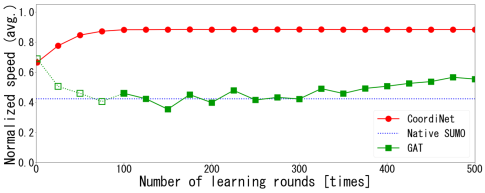
The proposed method showed an averaged vehicle speed improvement of about 30% compared to the initialization (random) (red line). The higher averaged speed compared to the unconstrained graph neural network (green line, GAT[3],) and the untrainable traffic simulator (blue dot line, SUMO[4]) confirm the effectiveness of the proposed method.

Road maps for training (10)



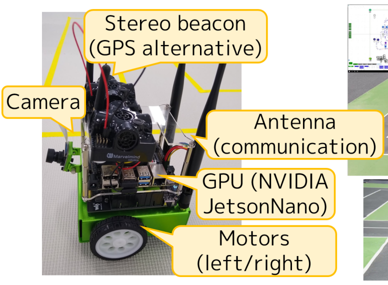
Road maps for testing (5)

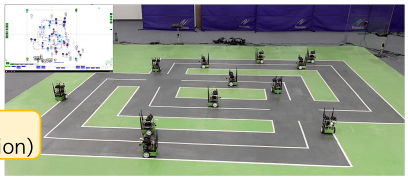


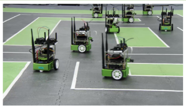


Feedback to real world system

We constructed a real world system of signal-free mobility using a set of small real vehicles (see figure below) and conducted experiments to feedback the optimized driving dynamics model to the real world. We confirmed that each vehicle autonomously run without collisions by exchanging tokens to each other.







References

[1] IONW conceptual video, “Mobility by IOWN,” YouTube, 2019

[2] K. Niwa, N. Ueda, H. Sawada, A. Fujino, S. Takeda, B. Kleijn, G. Zhang, “CoordiNet: Constrained dynamics learning for state coordination over graph,” in *Proc. the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD 2022)*, 2022 (under review).

[3] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, “Graph attention networks,” *arXiv preprint arXiv:1710.10903*, 2017.

[4] Simulation of Urban MObility (SUMO), https://www.eclipse.org/sumo/

Contact

Kenta Niwa / Learning and Intelligent Systems Research Group, Innovative Communication Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp

02

Accelerated learning of fine-layered optical neural networks

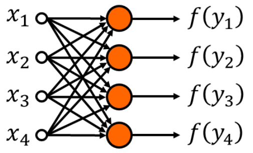
Efficient training of photonic AI

Abstract

An optical neural network (ONN) is a promising system due to its high-speed and low-power operation. The ONN has a multiple-layered structure of programmable Mach-Zehnder interferometers (MZIs). Due to this structure, it takes a lot of time to learn MZI parameters with a conventional automatic differentiation (AD). To solve the time-consuming problem, we develop a function module implemented in C++ to collectively calculate input-output values in a multiple-layered structure, where novel customized derivatives for an MZI are utilized in backpropagation. We demonstrate that our learning method works 50 times faster than the conventional AD when a pixel-by-pixel MNIST task is performed in a complex-valued recurrent neural network. Our approach supports ONN design and contributes to realize green-computing AI's instead of conventional ones consuming a lot of energy.

Optical Neural Network

Conventional Neural Network

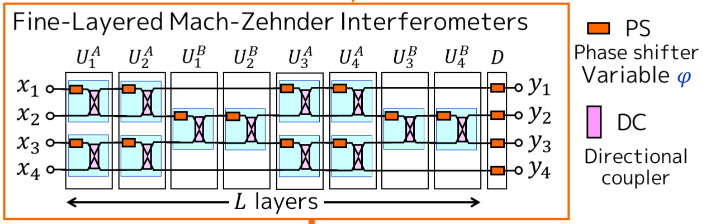


f: Nonlinear function

$$\boxed{y_j = \sum_{i=1}^{4} w_{ji} x_i}$$

Physical Layer

Input x → **Light** (Amp. : Phase) → **Analog multiplier** Weight and Input → Output y **Light** (Amp. : Phase)


Fine-Layered Mach-Zehnder Interferometers
U_1^A U_2^A U_1^B U_2^B U_3^A U_4^A U_3^B U_4^B D
x_1
x_2
x_3
x_4
y_1
y_2
y_3
y_4
L layers
PS
Phase shifter
Variable φ
DC
Directional coupler

Mathematical Model

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = D \cdots \overset{U_1^B}{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \multicolumn{2}{c}{U_{1[1]}^B} & 0 \\ 0 & & & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}} \overset{U_2^A}{\begin{pmatrix} U_{2[1]}^A & & 0 & 0 \\ & & 0 & 0 \\ 0 & 0 & U_{2[2]}^A & \\ 0 & 0 & & \end{pmatrix}} \overset{U_1^A}{\begin{pmatrix} U_{1[1]}^A & & 0 & 0 \\ & & 0 & 0 \\ 0 & 0 & U_{1[2]}^A & \\ 0 & 0 & & \end{pmatrix}} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \updownarrow dim$$

$$U_{p[q]}^R = \frac{1}{\sqrt{2}} \begin{pmatrix} e^{i\varphi_{p[q]}^R} & i \\ ie^{i\varphi_{p[q]}^R} & 1 \end{pmatrix}$$

U_p^R: Unitary matrix

$\varphi_{p[q]}^R$: Parameter

$R = A, B$

$1 \leq p \leq L/2$

$1 \leq q \leq dim/2$

Problem to Solve

Physical restriction: Difficulty in manufacture of large-scale circuits

→ Use of recurrent neural networks (**RNN**)

Fine-layered structure: One layer → One linear circuit

→ **Learning very deep neural networks**

A lot of computational time required by the conventional automatic differentiation (AD)

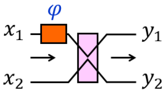
Accelerated Learning Method

Key 1. Customized derivatives: CD

Update of parameter: $\varphi \leftarrow \varphi - \eta \boxed{(\partial L/\partial \varphi)}$

L: Loss func.
η: Learning rate

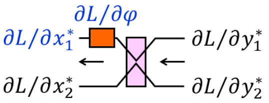
Forward



$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} e^{i\varphi} & i \\ ie^{i\varphi} & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Backward

Conjugate transpose


$\partial L/\partial \varphi$
$\partial L/\partial x_1^*$
$\partial L/\partial x_2^*$
$\partial L/\partial y_1^*$
$\partial L/\partial y_2^*$

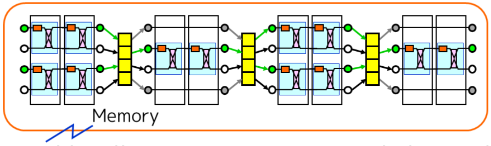
$$\begin{pmatrix} \partial L/\partial x_1^* \\ \partial L/\partial x_2^* \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} e^{-i\varphi} & -ie^{-i\varphi} \\ -i & 1 \end{pmatrix} \begin{pmatrix} \partial L/\partial y_1^* \\ \partial L/\partial y_2^* \end{pmatrix}$$

$$\boxed{\frac{\partial L}{\partial \varphi} = 2 \cdot \mathrm{Im}\left(x_1^* \frac{\partial L}{\partial x_1^*} \right)}$$

Update by multiplication of **only two values**

Key 2. Pointer Rewiring in C++: PR

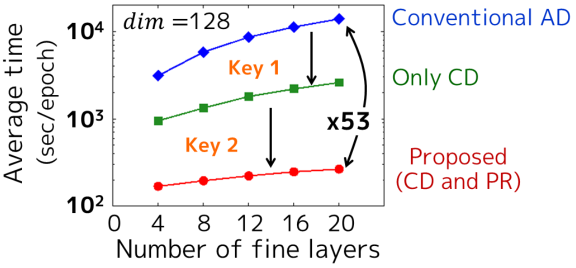
Fine-layered structure → one function module in C++



Data read by direct access to stored-data address

Experimental Results

- RNN with optical-circuit hidden unit
- Pixel-by-pixel MNIST classification task



References

[1] K. Aoyama, H. Sawada, “Accelerated method for learning fine-layered optical neural networks,” in *Proc. of IEEE/ACM the 40th International Conference on Computer-Aided Design*, 2021.

Contact

Kazuo Aoyama / Learning and Intelligent Systems Research Group, Innovative Communication Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp

03

Bayesian ensemble learning for better generalization performance

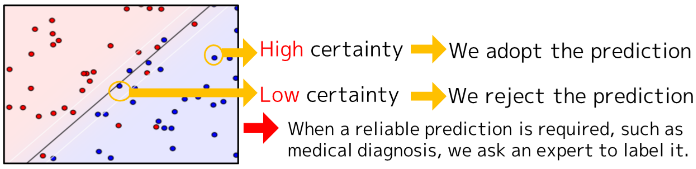
Multiple AIs make better predictions

Abstract

Evaluating the certainty of the prediction is essential for machine learning tasks. For example, certainty is required to assess the predictions' reliability, decision-making, and experimental design problems. We developed a method to efficiently calculate the certainty for a large model such as a neural network using an ensemble of models. Although evaluating the certainty of predictions using an ensemble of models has been widely used in existing work, it was theoretically unclear how to prepare ensembles. Our research theoretically derived an algorithm for preparing ensembles for expressing the certainty of prediction using multiple models. Evaluating the uncertainty is important to make machine learning reliable. We can easily evaluate the certainty using an ensemble of models and expand the range of machine learning applications by proceeding with this research.

The certainty of the prediction

When applying machine learning, the prediction as well as the "**certainty**" of how likely the prediction is essential for some tasks.


High certainty → We adopt the prediction

Low certainty → We reject the prediction

→ When a reliable prediction is required, such as medical diagnosis, we ask an expert to label it.

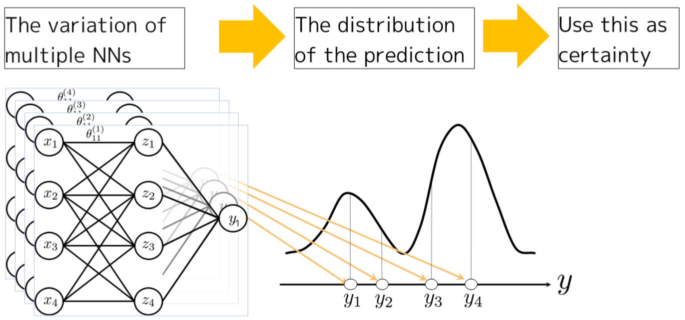
Applications using the certainty
- **Experimental design**. Gathering data that is particularly useful for the training.
- **Decision-making problems**. Determining what action to be taken next.

✓A widely used method of obtaining certainty is to use "**Bayesian inference**" to obtain a distribution of predictions.

✓For large models such as neural networks (NN), Bayesian inference requires approximation to perform.

Existing study: The variability of multiple models

We prepare multiple NNs, and approximate the distribution of predictions by the variation of their predictions.


The variation of multiple NNs → The distribution of the prediction → Use this as certainty

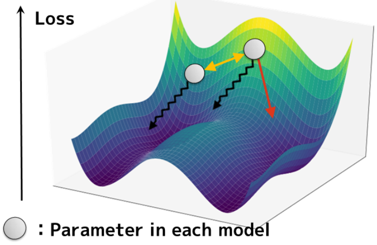
✓When we train NNs in the same way, NNs become similar.

✓How to prepare multiple NNs for the certainty ?

Our study: How to prepare multiple models

Incorporates a "**repulsion term**" that makes models differ from each other into the objective function.

Conceptual diagram of our approach when using gradient descent



○ : Parameter in each model

Update direction of params = Gradient of the loss + Repulsion term

Contribution : The models obtained by our method are guaranteed to approximate a distribution better than without a repulsion term theoretically.

Numerical experiments: Decision-making problem

Based on the mushroom features given at each time point, we decide whether to eat or not repeatedly.

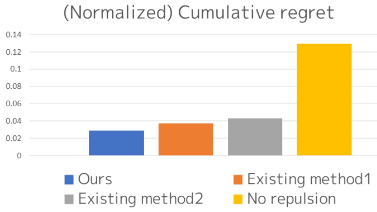
Feature 22dims

Safe → Reward ＋５point

Poison → Reward －３５point（with５０%）
Reward ＋５point（with５０%）

✓ The goal is to maximize total reward. Only reward information is a learning cue (no labels are given as in classification problems).

✓ Using the certainty, it is necessary to control the trade-offs between exploitation and exploration for gathering information while taking actions that maximize the reward.



✓ The graph shows the cumulative regret when 50000 decisions are made. Thus, the small regret indicates better performance.

✓ We normalized each regret by setting the regret of the completely random decision as one.

✓ We used 20 NNs.

References

[1] F. Futami, T. Iwata, N. Ueda, I. Sato, M. Sugiyama, "Loss function based second-order Jensen inequality and its application to particle variational inference," in *Proc. Neural Information Processing Systems (NeurIPS)*, 2021.

Contact

Futami Futoshi / Ueda Research Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp

04

Pruning neural networks with iterative randomization

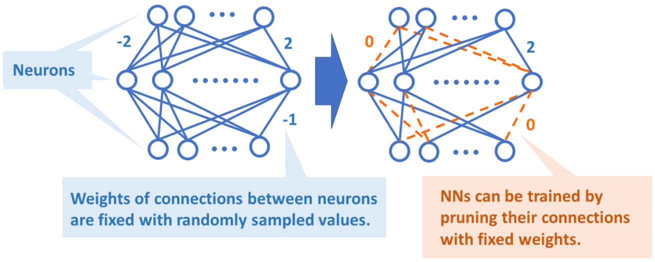
Training fast & lightweight neural networks

Abstract

Weight-pruning optimization is a new learning mechanism for neural networks. By this mechanism, we can train neural networks while keeping it as quantized and sparse ones. However a major challenge of weight-pruning optimization is its memory & computational cost during training. In this study, we developed a novel technology called iterative randomization to greatly reduce the costs. We both empirically and theoretically showed that our technique resolves the memory & computational challenge of weight-pruning optimization. By advancing this study, we will make AI technologies more affordable and energy-efficient.

Weight-Pruning Optimization of Neural Networks

□ Recent study [Ramanujan+,2020] has shown that neural networks (NNs) can be trained **by pruning their connections with fixed weights (weight-pruning optimization)** instead of directly optimizing the weights like standard training methods.



Neurons

Weights of connections between neurons are fixed with randomly sampled values.

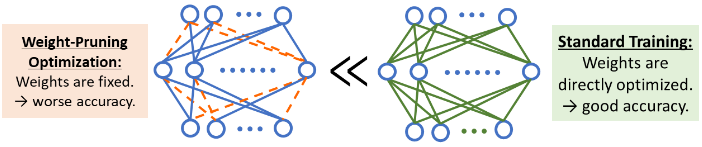
NNs can be trained by pruning their connections with fixed weights.

□ It leads to **lightweight & fast NNs** because

- ➢ Weights can be fixed with binarized/quantized values throughout training & inference.
- ➢ The resulting network is already spase.

Problem: Memory & Computational Cost

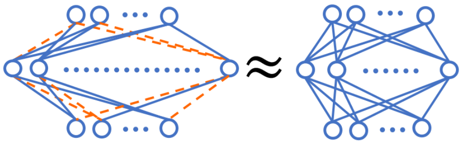
□ Due to the fixed weights, the accuracy of the NN trained with **weight-pruning optimization** is **typically worse** than the ones trained with **the standard training methods**.



Weight-Pruning Optimization:
Weights are fixed.
→ worse accuracy.

Standard Training:
Weights are directly optimized.
→ good accuracy.

□ Thus **larger NN is required** for weight-pruning optimization to achieve similar accuracy with the standard ones. Therefore its **memory & computational cost** is a big challenge of weight-pruning optimization.

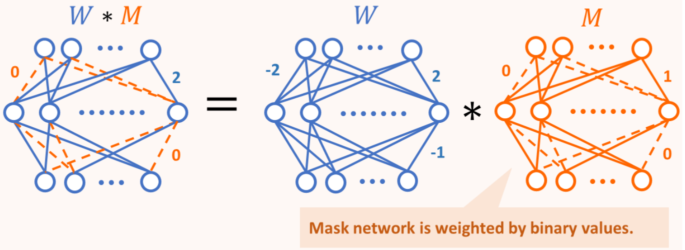


Pruning NNs with Iterative Randomization

Idea: **Searching higher accuracy solution by randomizing weights of pruned connections** during weight-pruning optimization.

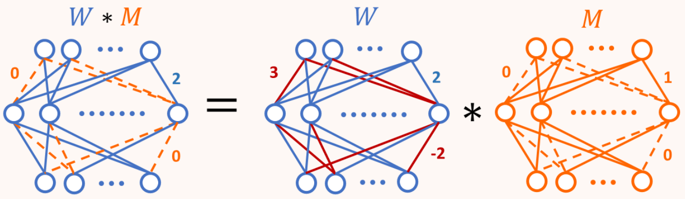
Summary of Algorithm

Step 1. For **neural network W** with fixed weights, optimize the corresponding **mask network M** (= pruned structure of W).



Mask network is weighted by binary values.

Step 2. **Replacing the weights of pruned connections in W** (i.e. the corresponding weights in M is 0) **with random values.**



Step 3. Back to Step 1. (Repeat)

□ **Our algorithm makes it possible to achieve higher accuracy with weight-pruning optimization** because once pruned connections can be revived with another weights if necessary.

□ **We theoretically proved that, with our algorithm, the larger NN is no longer required** to achieve similar accuracy as the standard methods. Hence **the challenge of memory & computational cost is now resolved.**

□ Moreover, it requires **almost no additional computational cost** under GPU environments.

References

[1] D. Chijiwa, S. Yamaguchi, Y. Ida, K. Umakoshi, T. Inoue, "Pruning randomly initialized neural networks with iterative randomization," *Advances in Neural Information Processing Systems 34*, 2021.

[2] V. Ramanujan, M. Wortsman, A. Kembhavi, A. Farhadi, M. Rastegari, "What's hidden in a randomly weighted neural network?," in *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11893–11902, 2020.

Contact

Daiki Chijiwa / Computer and Data Science Laboratories
Email: cs-openhouse-ml@hco.ntt.co.jp

05

Learning time-evolving states via dynamic Hawkes processes

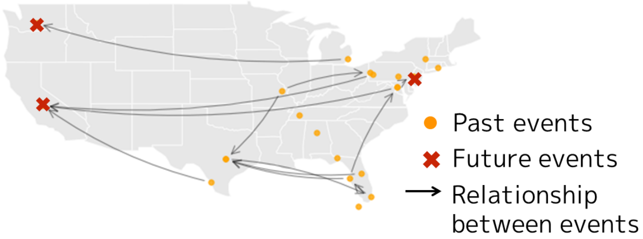
How the events spread?

Abstract

Sequences of events including infectious disease outbreaks, social network activities, and crimes are ubiquitous and the data on such events carry essential information about the underlying diffusion processes between communities (e.g., regions, online user groups). Modeling diffusion processes and predicting future events are crucial in many applications including epidemic control, viral marketing, and predictive policing. Diffusion processes depend not only on the influences from the past, but also the current (time-evolving) states of the communities, e.g., people's awareness of the disease and people's current interests. We propose a novel Hawkes process model that is able to capture the underlying dynamics of community states behind the diffusion processes and predict the occurrences of events based on the dynamics. The proposed method offers a flexible way to learn complex representations of the time-evolving communities' states, while at the same time it allows to computing the exact likelihood, which makes parameter learning tractable.

Diffusion process

Various social phenomena can be described by diffusion processes among multiple communities.
E.g., Demonstrations that started in large cities have spread to dozens of cities across the country.

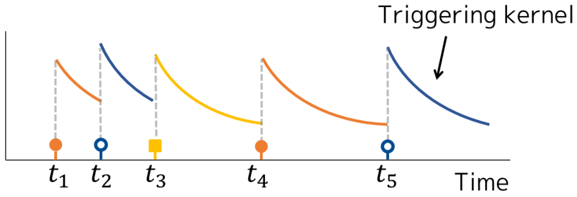


Demonstrations in United States.

Understanding diffusion mechanism and predicting future events are crucial in many applications such as epidemic control and predictive policing.

Baseline: Hawkes processes

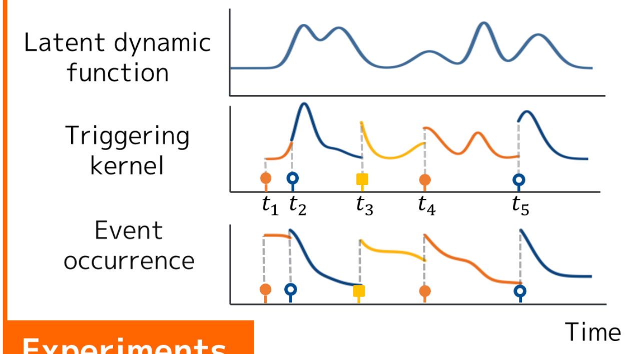
Capture the impact of past events on the event occurrence in each community by *triggering kernel*.



Limitation: Focus on learning the static influence of the past events on the current event, thereby overlooking the factor of time-evolution.
E.g., Expansion of demonstrations depends on motivations for participation of community population.

Proposal: Dynamic Hawkes processes

- Introduce *latent dynamics function* for each community that represents its hidden dynamic states.
- Model the triggering kernel by using latent dynamics function and its integral.



Experiments

- Evaluate the prediction performance of the proposed method on four real-world datasets.
- Use MAPE between the predicted number of events and the ground truth as metric.

	Reddit	News	Protest	Crime
Homogeneous point process	0.553	0.6	0.345	0.144
Hawkes process	0.458	0.471	0.415	0.179
Reinforced process	0.595	0.481	0.581	0.175
SelfCorrecting process	0.475	0.452	0.524	0.123
RMTPP	0.311	0.446	0.639	0.302
Proposed method	**0.305**	**0.442**	**0.318**	**0.117**

Proposed method outperforms the five existing methods across all the datasets.

References

[1] M. Okawa, T. Iwata, Y. Tanaka, H. Toda, T. Kurashima, H. Kashima, "Dynamic Hawkes processes for discovering time-evolving communities' states behind diffusion processes," in *Proc. of the 27th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD21)*, pp. 1276–1286, 2021.

Contact

Maya Okawa / Human Informatics Laboratories
Email: cs-openhouse-ml@hco.ntt.co.jp

06

Optimal shelter operation with sequential return of evacuees

Rapid disaster recovery through efficient shelter management

Abstract

Shelters are provided to evacuees whose homes have been destroyed in a disaster. In a recovery phase, efficient operation of the shelters is necessary to restore the facilities to their original use. In this study, we proposed a method to minimize the total cost of operating shelters and the burden of relocating evacuees between shelters by utilizing the return home time of evacuees. Our method allows the shelters to be used for their intended purpose as soon as possible after a disaster, thus enabling rapid recovery. Even when the number of evacuees is large, we introduced a variable that represents the number of evacuees grouped by the return home time so that the calculation can be performed efficiently. We also proposed a method to estimate the burden of relocating evacuees between evacuation shelters, thus achieving a balance between the operation costs of shelters and the relocation costs of evacuees. When disaster simulations are used to select response measures, it is not efficient to run through all the patterns of response measures exhaustively. By developing our method further, we aim to establish a simulation infrastructure that solves not only disasters but also various social issues through simulation.

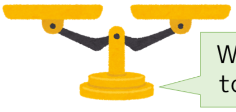
Motivation

Evacuation Shelters are often set up in Schools, so they must be closed before schooling resumes

Shelters can be closed early if evacuees decreasing are relocated into shelters remaining

Evacuees' relocation cost
"Want to continue to stay in neighbor shelter"

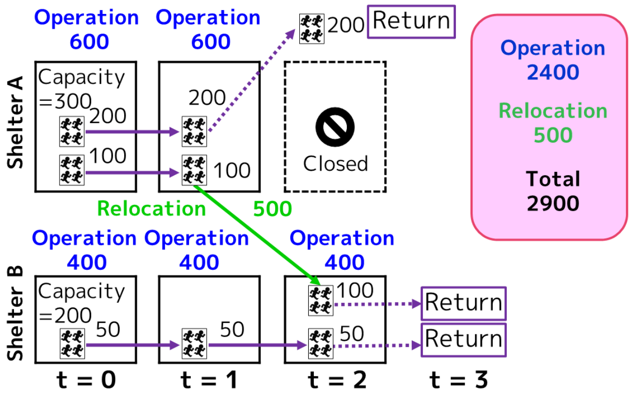
Shelters' operation cost
"Want to close shelters and recover early"



We developed a method to minimize total costs

Key Point 1

The amount of calculation does not depend on # evacuees by grouping them with the same return time



Shelter A

Operation 600 | Operation 600

Capacity =300, 200, 100 (t = 0) → 200, 100 (t = 1) → 200 Return; Closed (t = 2)

Relocation 500

Shelter B

Operation 400 | Operation 400 | Operation 400

Capacity =200, 50 (t = 0) → 50 (t = 1) → 100, 50 (t = 2) → Return, Return (t = 3)

t = 0 | t = 1 | t = 2 | t = 3

Operation 2400
Relocation 500
Total 2900

Key Point 2

Relocation cost is estimated with historical disaster data (Kobe Earthquake) and the following assumptions

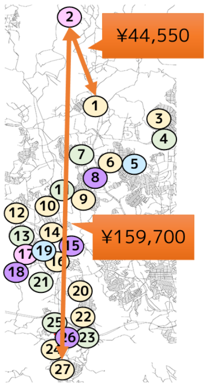
Assumptions

1. No relocation across the district
2. Proportional costs to the distance
3. Best operations at each time

Historical data
- # evacuees
- # shelter operated

Estimate

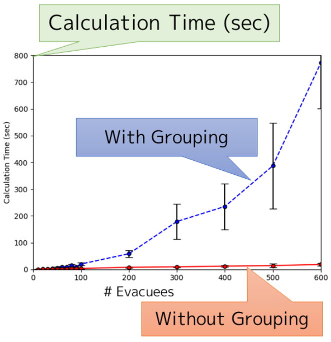
Relocation Cost (Burden):
10,000 yen/km per capita

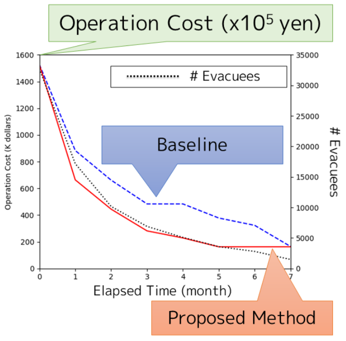


Experiments

Simulation Experiment of earthquake in Ikoma City

Proposed method reduced Calc. Time & Objective Cost





Methods	Baseline	**Proposed**
Operation Cost	¥4.9 x10^8	**¥3.5** x10^8
# Relocations	8259	3611
Relocation Cost	¥8707 x10^4	¥5383 x10^4

29% Cost cut in Operation

Reduced Relocation

References

[1] H. Shimizu, H. Suwa, T. Iwata, A. Fujino, H. Sawada, K. Yasumoto, "Evacuation shelter scheduling problem," in *Proc. the 55th Hawaii International Conference on System Sciences (HICSS 2022)*, pp. 5705–5714, 2022.

Contact

Hitoshi Shimizu / Learning and Intelligent Systems Research Group, Innovative Communication Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp

07

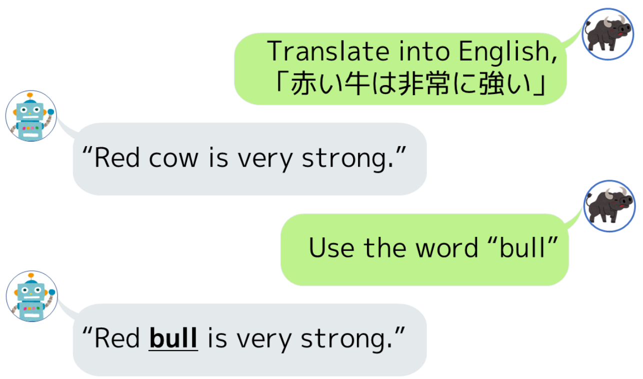
Lexically constrained neural machine translation

Translating with your favorite expressions

Abstract

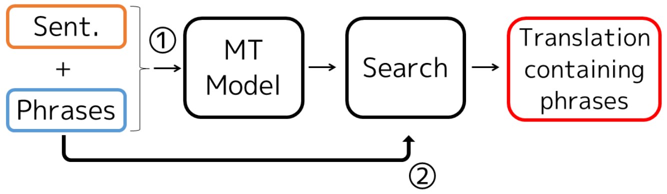
Although the recent neural machine translation achieves excellent performance, controlling its output expressions is still challenging. We propose a lexically constrained neural machine translation, a method whose translations contain user-specified phrases. Our method improves translation performance while saving inference time and was ranked first in the international competition at WAT 2021. When translating documents in domains such as legal, patent, and scientific, the translation of proper nouns and technical terms is strongly required to be the same expressions throughout the document. Our method will contribute to ensuring consistency in translation by user-specifying expressions.

Lexically Constrained Machine Translation



- Translating with favorite expressions is essential.
- For translation of patent and technical papers, the translation of proper nouns is required to be the same expression throughout the document.
- **Controlling outputs** of machine translation (MT) is still challenging.
  → The **controllability** of MT needs to be improved.
- We propose an MT method whose **translation contains given specified phrases**.
  - It achieves high translation accuracy and works fast.
  - It won **1st place** in the competition at WAT 2021.

Proposed Method



① Input a sentence and **specified phrases** into the model, and train the model **to output given phrases**.

② Search a translation **containing given all phrases** based on model outputs.

Point:

Learning the model to output phrases (①) makes the latter search step (②) **more efficient**.

- Only ① cannot guarantee that the translation contains all given terms.
- Only ② is less accurate and works slower.

Experiment

w/ scientific paper dataset

- Evaluate the performance when a human gives the appropriate expression for technical terms as specified phrases.

Translation Accuracy
BLEU (higher is better)

Method	En→Ja	Ja→En
General MT†	44.64*	29.30*
Only ①	53.79*	41.88*
Only ②	45.38	23.22
Proposed	**55.49**	**43.33**

† w/o phrases information
* translations do not contain all given phrases

- By combining ① and ②, our method can **specify phrases and improve translation accuracy**.
  → Our method can **yield a human-parity score** when we specify the appropriate phrases.
- Comparing two methods whose translations contain all given terms, we confirmed our method works **more than three times faster** than only ②.

References

[1] K. Chousa, M. Morishita, "Input augmentation improves constrained beam search for neural machine translation: NTT at WAT 2021," in *Proc. of the 8th Workshop on Asian Translation (WAT2021)*, pp. 53–61, 2021.

Contact

Katsuki Chousa / Linguistic Intelligence Research Group, Innovative Communication Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp

08

Equilibrium optimization of combinatorial congestion games

Toward uncongested infrastructures under user-equality

Abstract

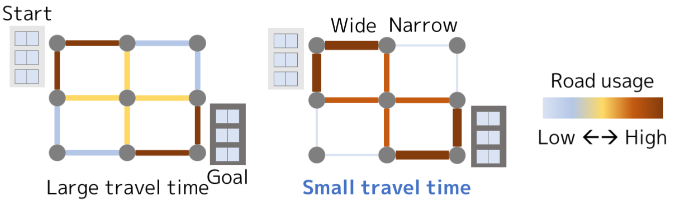
In social infrastructures such as road and telecommunication networks, a link is congested and incurs more cost if many people use it. We introduce a method to compute better social design where users' cost lowers even when each user chooses a path or a combination of links selfishly. We develop a new method to compute the difference in cost when we modify the social design using a differentiable computation technique. Moreover, we compress a massive number of available paths into a data structure called a binary decision diagram, enabling us to deal with broader settings in a reasonable time. Our approach can contribute to reducing the congestion of people's flows and telecommunication networks by designing infrastructures, e.g., improving roads and expanding the bandwidth of links. Moreover, the proposed method is versatile and thus may be applicable for broader areas such as machine learning problems containing combinatorial optimization tasks.

Social Design

Adjustable elements in designing infrastructures (e.g., road width, speed limit, bandwidth of communication link)

We want to prevent congestion of infrastructures by adjusting them

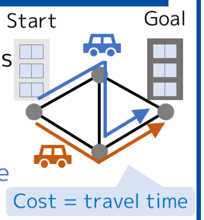
Ex.) Adjusting road widths to decrease travel time



Congestion Game and Equilibrium State

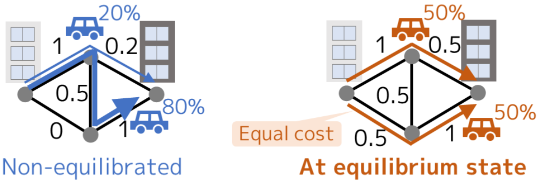
Modelling each player's selfish behavior of using infrastructures such as road networks

- There are infinitely many peoples
- Each people choose a path or a combination of links with smaller cost
- Link cost increases with higher link usage



Equilibrium state

Final result of people's trial to decrease his/her own cost = State where every player's cost is equal and the smallest



Equilibrium optimization

Computation of social design that make each people's cost smaller under some constraints such as budget

- There are so many paths that make even computing equilibrium state of a fixed social design challenging
- Moreover, we need to find out better social design

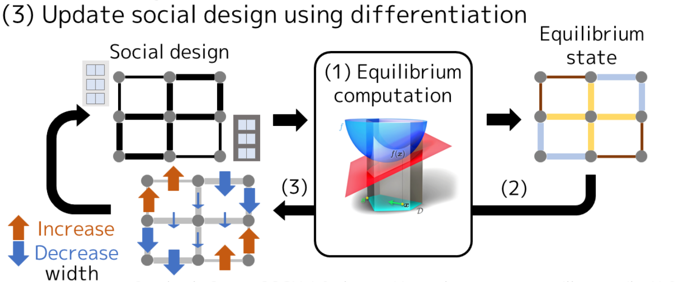
Essence of Proposed Method

Repeat:

(1) Compute equilibrium with fixed social design

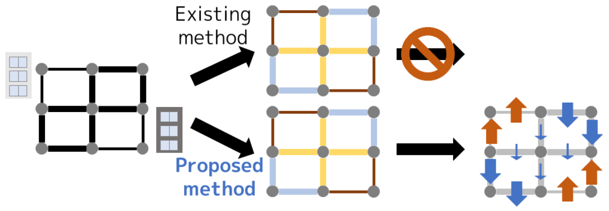
(2) Compute difference of cost w.r.t. modification of social design (**differentiation**)

(3) Update social design using differentiation



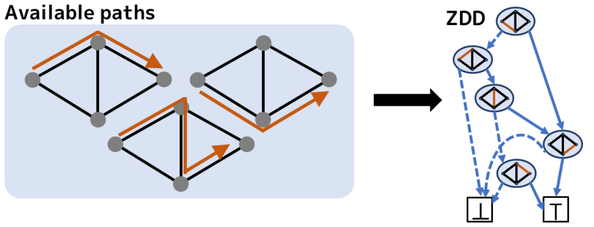
Stephanie Stutz, CC BY 4.0 <https://creativecommons.org/licenses/by/4.0>, via Wikimedia Commons https://commons.wikimedia.org/wiki/File:Frank-Wolfe_Algorithm.png

Essence 1: New equilibrium computation method that can also compute differentiation information, **enabling us to deal with broader setting**



Essence 2: Usage of zero-suppressed binary decision diagram (ZDD) that compresses available paths, **enabling fast computation**

Ex.) Representing 8 quadrillion paths with **less than 1MB**



References

[1] S. Sakaue, K. Nakamura, "Differentiable equilibrium computation with decision diagrams for Stackelberg models of combinatorial congestion games," in *Proc. 35th Conference on Neural Information Processing Systems (NeurIPS)*, 2021.

Contact

Kengo Nakamura / Linguistic Intelligence Research Group, Innovative Communication Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp

09

Post-editing support based on source-target word alignment

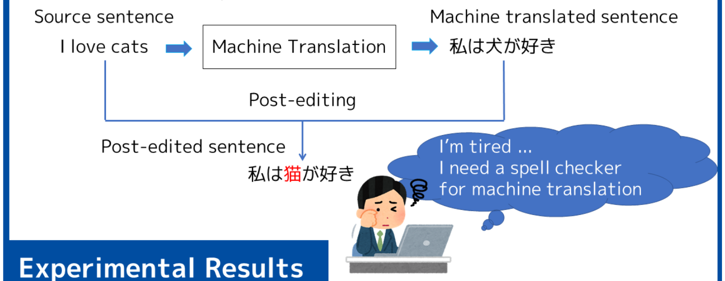
Looking for mistakes in machine translation

Abstract

Neural machine translation has the problem of generating fluent translations that do not necessarily match the content of the source text. We present technology that supports "post-editing," in which humans and machines cooperate to detect and correct errors in machine translation. We have developed a method to obtain word alignment between source and target sentences that are not necessarily semantically equivalent due to translation errors. It can present the user with the editing operations necessary to correct errors in the output of machine translation. We aim to realize interactive machine translation as easy to use as a spell checker.

Post-Editing for Machine Translation

Neural networks have greatly improved the accuracy of machine translation, but they will never eliminate machine translation errors. In fields where errors are not allowed, such as medicine and patents, post-editing (error detection and correction by humans) is essential.



Experimental Results

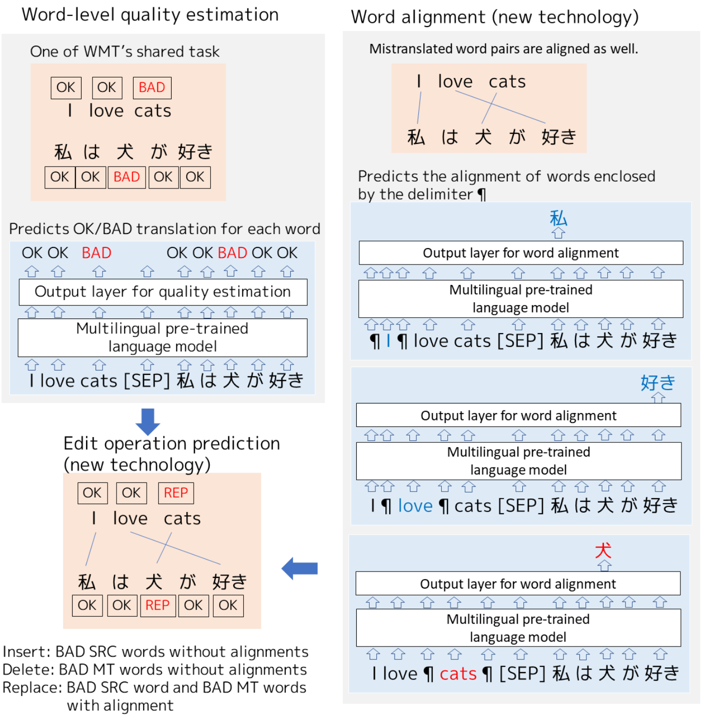
Quality Estimation Task datasets in WMT-2020

- 8,000 tuples of source, machine-translated, and post-edited sentences with OK/BAD translation tags for each word
- Of these, 1,000 tuples are manually word-aligned
- The accuracies (F1) of edit tags and word alignment are evaluated using the remaining 7000 + 800 tuples as training and 200 tuples for test data

	English to German			English to Chinese		
	SRC edit tag	MT edit tag	Word alignment	SRC edit tag	MT edit tag	Word alignment
Baseline	0.626	0.767	0.828	0.360	0.733	0.739
Proposed	0.755	0.827	0.916	0.849	0.897	0.888

アプローチ

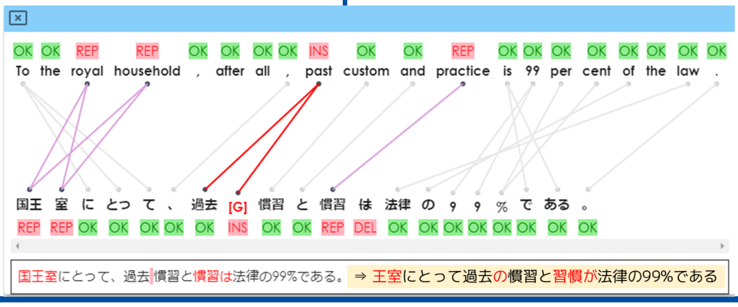
Predicts editing operations in the post-editing by combining word-level quality estimation (OK/BAD) and word alignment


Word-level quality estimation
One of WMT's shared task
OK OK BAD
I love cats
私 は 犬 が 好き
OK OK BAD OK OK
Predicts OK/BAD translation for each word
OK OK BAD OK OK BAD OK OK
Output layer for quality estimation
Multilingual pre-trained language model
I love cats [SEP] 私 は 犬 が 好き
Edit operation prediction (new technology)
OK OK REP
I love cats
私 は 犬 が 好き
OK OK REP OK OK
Word alignment (new technology)
Mistranslated word pairs are aligned as well.
I love cats
私 は 犬 が 好き
Predicts the alignment of words enclosed by the delimiter ¶
私
Output layer for word alignment
Multilingual pre-trained language model
¶ I ¶ love cats [SEP] 私 は 犬 が 好き
好き
Output layer for word alignment
Multilingual pre-trained language model
I ¶ love ¶ cats [SEP] 私 は 犬 が 好き
犬
Output layer for word alignment
Multilingual pre-trained language model
I love ¶ cats ¶ [SEP] 私 は 犬 が 好き

Insert: BAD SRC words without alignments
Delete: BAD MT words without alignments
Replace: BAD SRC word and BAD MT words with alignment

User Interface

The post-editor edits the machine translated sentence referring to the word alignment and edit tags


OK OK REP REP OK OK OK OK INS OK OK REP OK OK OK OK OK OK OK OK

To the royal household , after all , past custom and practice is 99 per cent of the law .

国王 室 に とっ て 、 過去 [G] 慣習 と 慣習 は 法律 の 9 9 % で ある 。

REP REP OK OK OK OK OK INS OK OK REP DEL OK OK OK OK OK OK OK OK

国王室にとって、過去慣習と慣習は法律の99%である。 ⇒ 王室にとって過去の慣習と習慣が法律の99%である

Information display area

Edit tags

Source sentence

Word alignment

Machine translated sentence

Edit tags

Edit area

References

[1] M. Nagata, K. Chousa, M. Nishino, "A supervised word alignment method based on cross-language span prediction using multilingual BERT," in *Proc. the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020.

[2] Y. Wei, T. Utsuro, M. Nagata, "Word-level quality estimation for machine translation based on source-MT word alignment," in *Proc. 27th Annual Meeting of the Association for Natural Language Processing*, 2021. (Joint Research with Tsukuba University)

Contact

Masaaki Nagata / Linguistic Intelligence Research Group, Innovative Communication Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp

10

Selecting voice and words for better understandability

Elderly-friendly speaking styles

Abstract

We investigate both word and voice selection to clarify elderly-friendly speaking styles. Previously, there has been no recommendation beyond "speak loudly and slowly" nor any explanation of where and how to make such changes. We select exemplar speakers, whom the elderly consider easiest to understand, from among elderly service workers and qualified personnel. Through the interviews with the exemplar speakers and analysis of elderly directed speech data uttered by the exemplar speakers, we clarify some of the detailed features of elderly-friendly speaking styles. This work provides important new insight into the practice of elderly-friendly speaking. We aim to clarify knowledge about elderly-friendly speaking styles that might be tacit knowledge among the exemplar speakers and open the knowledge to everyone for practical use. Moreover, we aim to realize richer communication between the elderly and the artificial intelligence (AI) that has learned elderly-friendly speaking styles.

Issues and Goals

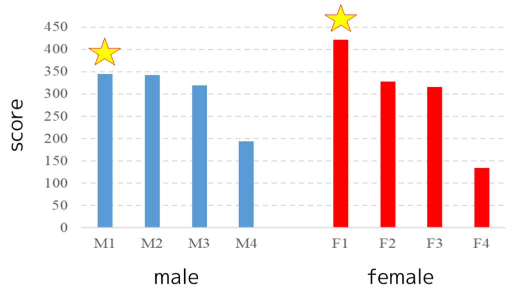
- Speech-based assistive technologies for the elderly are limited to hearing aids, uniform speed elongation etc..
- Reports on how to speak to the elderly (as a speech-based assistive technology) have only recommended "speak loudly and slowly," as unorganized or tacit knowledge

Goals

Formulate “elderly-friendly speaking styles" and make them specific and practical

Approach

1) Select exemplar speakers whom the elderly consider easy to understand: Elderly subjects compare pairs of speeches uttered by two speakers among elderly service workers and qualified persons and give scores to the winner to select the highest-scoring male and female speakers (M1 & F1 marked with ☆ in figure below)
2) Obtain tips that the exemplar speakers consciously make in practice through post-experiment interviews
3) Find subconscious tips for the easy-to-understand speech by analyzing speech, such as comparing between-sentence pause lengths



Scores for each speaker
when selecting exemplar speakers

Current status and Future work

Gradually clarified elderly-friendly speaking styles
(blue: conventional; underlined: new)

From interviews

On the linguistic side

Rephrasing into concise syntax in familiar and unambiguous words

On the acoustic speech side

loudness: enough for the listeners to hear
frequency: generally at lower pitch, higher when conveying emotion
speed: basically slow and constant, and even slower in important areas
separation: no need for as much separation as that for the hearing impaired or in noise

⇒ Much of “where?” and “to what extent?” the elderly-friendly style emerged is subconscious

thus analyzed

From speech analysis

pause length: differences found according to document structure and expression

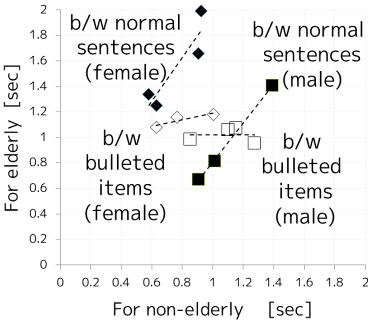


Figure shows how the length of pauses between bulleted items (◇□) in speech for the elderly aligns around 1 second on the vertical axis (dashed line is regression line, ◆/■ denotes length of pause between normal sentences)

Future work

- Describing concretely the remaining observations of "where?" and "to what extent? the elderly-friendly speaking style emerges
- Clarifying similarities and differences of elderly-friendly speaking style with other styles (for infants and non-Japanese, etc., or made in practice by announcers)

References

[1] H. Nakajima, Y. Aono, “Collection and analyses of exemplary speech data to establish easy-to-understand speech synthesis for Japanese elderly adults,” in *Proc. 23rd Conference of the Oriental COCOSDA International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA)*, pp. 145–150, 2020 (https://ieeexplore.ieee.org/document/9295000).

[2] H. Nakajima, N. Miyazaki, S. Sakauchi, “Pause length analysis between utterances to elderly people,” in *Proc. 2015 Autumn Meeting Acoustical Society of Japan*, pp. 399–400, 2015.

Contact

Hideharu Nakajima / Interaction Research Group, Innovative Communication Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp

11

Casual-dialog system based on scenery and nearby information

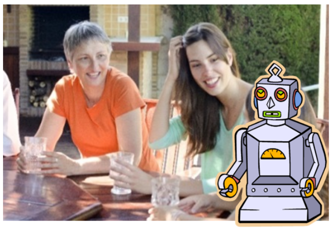
Talking with AIs about views from a vehicle

Abstract

This is a study of a partner dialogue system for mobile vehicles that uses the ever-changing, real-time view from the car as a topic of conversation. This system uses a deep-learning based dialogue model using the largest scale of Japanese dialogue data developed by NTT to realize natural dialogue. This system integrates scenery images from vehicle and spots around the car's location to talk about scenery around the vehicle. By sequentially incorporating information about the area around the vehicle's location, we are realizing a new experience of driving while enjoying the pleasure of sharing the "now" with a knowledgeable dialogue system.

Dialogue systems as our chatting partner

The dialogue system is becoming a pleasure partner for chit-chat.

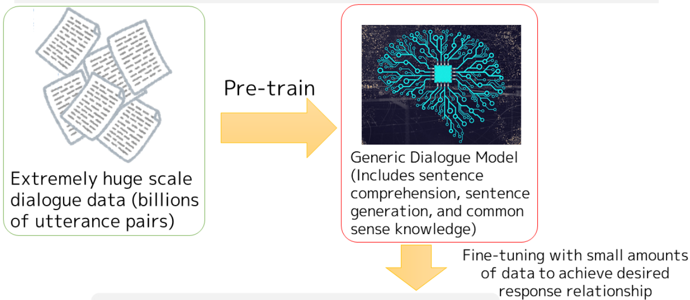


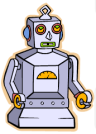
- **Anytime**
  (No restrictions on time or place)
- **Easily**
  (No need to be hesitant)
- **Deeply**
  (Deal with detailed hobbies and discuss private matters)

Dialogue system with pre-training

Large-scale pre-training based methods rapidly improve the performance of chatting systems.

*Pre-training: A method in which the system learns the naturalness of sentences and rough response patterns in advance using large-scale data.


Extremely huge scale dialogue data (billions of utterance pairs)
Pre-train
Generic Dialogue Model (Includes sentence comprehension, sentence generation, and common sense knowledge)
Fine-tuning with small amounts of data to achieve desired response relationship



System

User

Hello, I'm looking forward to meeting you. Let me ask you right away, what do you do in your free time?

Hello, nice to meet you. I enjoy watching "Wednesday Night Live".

I love it. Which series is your favorite?

I like the Confrontation Islands series.

Confrontation Islands is a masterpiece! I like European Revenge. Do you have a favorite episode?

We achieve natural dialogues in text-closed world.

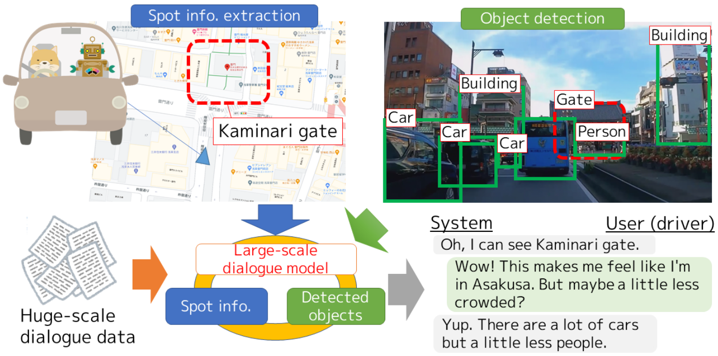
Robot that chats while "watching" the scenery

Problems with text-closed chats

- System cannot interact based on its own surrounding context.

This exhibition

- Realization of a chatting robot that recognizes the ever-changing scenery and geographic information seen from a car while driving.



System | User (driver)

Oh, I can see Kaminari gate.

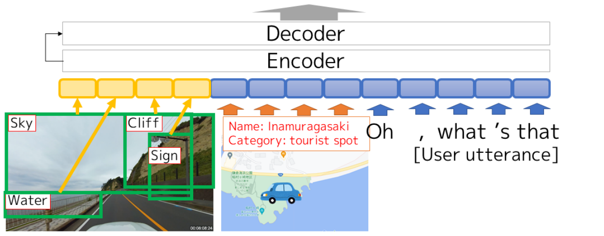
Wow! This makes me feel like I'm in Asakusa. But maybe a little less crowded?

Yup. There are a lot of cars but a little less people.

Model Structure

- Introduce image information and external knowledge into text-based chatting system.
- Input image features of detected objects.
- Input textual information on spots around the vehicle's location.
- Selects the most topical utterance from a group of utterances expressing impressions of a sequence of images obtained from video images.

System utterance candidates
1: I heard that area is a tourist spot called Inamuragasaki.
2: That looks like a cliff that could collapse at any moment.



Input info. into model

[Scenery image] [Spot information]

This research was supported by Grant-in-Aid for Scientific Research (KAKENHI), "Behavioral decision model estimation research group" (Assignment No. 19H05693)

References

[1] H. Sugiyama, M. Mizukami, T. Arimoto, H. Narimatsu, Y. Chiba, H. Nakajima, T. Meguro, "Empirical analysis of training strategies of Transformer-based Japanese chit-chat systems," *arxiv:2109.05217*, 2021.

[2] K. Mitsuda, R. Higashinaka, H. Sugiyama, M. Mizukami, T. Kinebuchi, R. Nakamura, N. Adachi, H. Kawabata, "Fine-tuning a pre-trained Transformer-based encoder-decoder model with user-generated question-answer pairs to realize character-like chatbots," *IWSDS*, 2021.

Contact

Hiroaki Sugiyama / Interaction Research Group, Innovative Communication Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp

12

Quantum algorithm for finding collisions of hash functions

Toward secure cryptography against quantum attacks

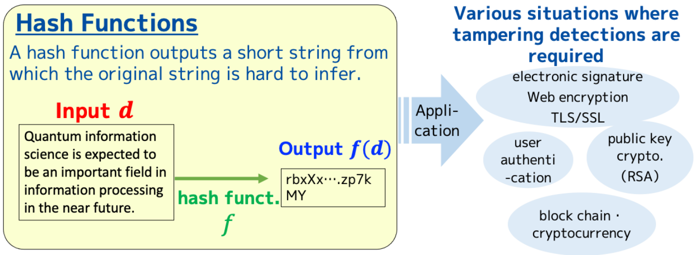
Abstract

Recently, the security analysis of ciphers against quantum attacks is rapidly growing in importance, since quantum computers could make strong attacks on them in the future. For such a security analysis, it is crucial to evaluate how fast quantum computers can solve the problems used to break ciphers. Among others, it is one of the major problems to find a multi-collision of random hash functions, essential primitives used ubiquitously in cryptosystems. In this work, we provide a novel quantum algorithm that solves this problem. This algorithm is the fastest among all possible ones in the sense that it achieves the theoretical limit. Our result would contribute to enhancing the security of hash-based ciphers in the quantum-computer era.

Background and Our Result

- **The security of cryptosystems is based on how much time is required to attack them** (e.g., even the fastest computers take a billion years for breaking some cipher).
- As quantum computers have been actively developed recently, **the security analysis of ciphers against quantum attacks** is rapidly growing in importance.

We provide a **fastest quantum algorithm that finds a multi-collision of a hash function,** an important cryptographic primitive.
➪Our result would contribute to the security analysis of various hash-based cryptosystems against quantum attacks.



Hash Functions

A hash function outputs a short string from which the original string is hard to infer.

Input d

Quantum information science is expected to be an important field in information processing in the near future.

hash funct. f

Output $f(d)$

rbxXx⋯.zp7k MY

Appli-cation

Various situations where tampering detections are required

electronic signature
Web encryption
TLS/SSL

user authenti-cation

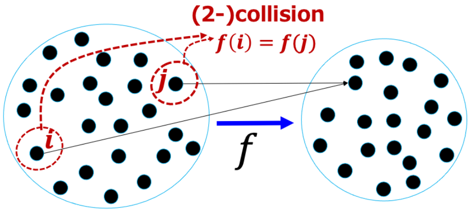
public key crypto. (RSA)

block chain · cryptocurrency

Collision of Hash Functions

A pair of elements is called a **(2-)collision** if they have an identical image via f. Similarly, an **ℓ-collision** is defined as ℓ elements with an identical image via f (e.g., $f(i) = f(j) = f(k)$ for a 3-collision).

(2-)collision



Finding a collision makes it possible to tamper electronic data.
➪For assessing the security, it is necessary to estimate the hardness of (i.e., the time required for) finding collisions.
➪**Such estimation requires algorithms for finding collisions.**

Driving force behind the improvement of security of hash functions has been the discovery of faster collision-finding algorithms

MD5 ['91] ➡ SHA-1 ['95] ➡ SHA-2 ['02] ➡ SHA-3 ['15]

Details of our Algorithm

We provide a theoretical bound on the run-time taken by our quantum algorithm to find a multi-collision for a given random hash function. Then, we illustrate the idea used in our algorithm.

For a given random hash function $f\colon \{1, \cdots, M\} \to \{1, \cdots, N\}$ $(\mathrm{M} \geq N)$,
our quantum algorithm can find an ℓ-collision of f in
$$N^{\frac{1}{2}\left(1-\frac{1}{2^{\ell}-1}\right)} \text{ time.}$$
(assuming that f can be computed quickly).

◎This attains the theoretical time bound (matching with the lower bound [LZ20])

Comparison with Previous Bound [HSX17] on Time Complexity

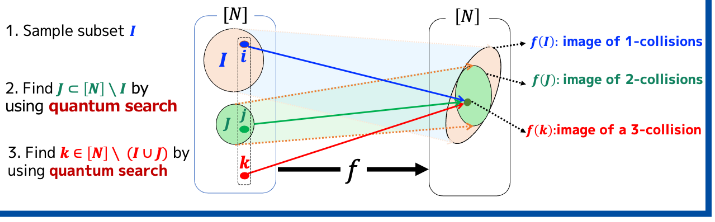
ℓ (multiplicity)	2	3	4	5	· · ·	ℓ
Previous algorithm [HSX17]	$N^{\frac{1}{3}}$	$N^{\frac{4}{9}}$	$N^{\frac{13}{27}}$	$N^{\frac{40}{81}}$	· · ·	$N^{\frac{1}{2}\left(1-\frac{1}{3^{\ell-1}}\right)}$
Our algorithm	$N^{\frac{1}{3}}$	$N^{\frac{3}{7}}$	$N^{\frac{7}{15}}$	$N^{\frac{15}{31}}$	· · ·	$N^{\frac{1}{2}\left(1-\frac{1}{2^{\ell}-1}\right)}$

Ex.) In the case of $\ell = 3$ and $N = 2000$, ours is a **billion times faster** than the previous algorithm.

$$N\text{:2000bit} \rightarrow N^{\frac{4}{9}} : N^{\frac{3}{7}} \approx 1{,}000{,}000{,}000 : 1$$

Outline of Algorithm (3-collision case)

1. Sample subset $I \subset [N]$ and compute the image $f(I)$ of I, where $[N] \equiv \{1, \cdots, N\}$
2. Find a subset $J \subset [N] \setminus I$ that forms 2-collisions with elements in I, and compute $f(J)$.
3. Find an element $k \in [N] \setminus (I \cup J)$ that forms a 3-collision with an element pair in $I \times J$.
4. Output the triplet (i, j, k).


1. Sample subset I

2. Find $J \subset [N] \setminus I$ by using **quantum search**

3. Find $k \in [N] \setminus (I \cup J)$ by using **quantum search**

$[N]$
I
i
J
j
k
f
$[N]$
$f(I)$: image of 1-collisions
$f(J)$: image of 2-collisions
$f(k)$:image of a 3-collision

References

[1] A. Hosoyamada, Y. Sasaki, S. Tani, K. Xagawa, "Improved quantum multicollision-finding algorithm," in *Proc. 10th International Conference on Post-Quantum Cryptography (PQCrypto 2019)*, pp. 350–367, vol. 11505, 2019.

[2] A. Hosoyamada, Y. Sasaki, S. Tani, K. Xagawa, "Quantum algorithm for the multicollision problem," *Theoretical Computer Science*, vol. 842, pp. 100–117, 2020.

Contact

Seiichiro Tani / Computing Theory Research Group, Media Information Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp

13

Finding new arithmetic phenomena via generalized motives

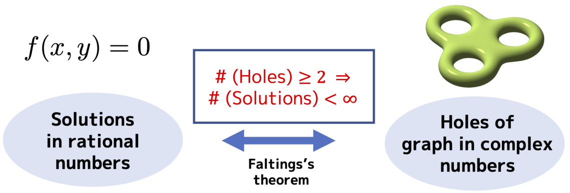
Where does the wonder of numbers come from?

Abstract

In the study of mathematics, we often find mysterious connections between two seemingly unrelated objects and phenomena. The aim of this research is to understand how these mysterious connections appear, by using the theory of generalized motives, which was developed in my previous research. We can study numbers by observing a type of shape called algebraic varieties. The theory of generalized motives enables us to continuously observe algebraic varieties from various points of view. The shapes of algebraic varieties observed from different points of view appear to be different, but they can be connected through this continuous observation. By using the theory of generalized motives, we can systematically connect seemingly different objects without relying on random luck. We anticipate that this study will accelerate the research on number theory, which underpins human activity everywhere.

Mystery in math

There are many mysterious connections in number theory. For example, the number of holes of a shape is related to the number of rational solutions of an algebraic equation.


$f(x,y)=0$
# (Holes) ≥ 2 ⇒
# (Solutions) < ∞
Solutions in rational numbers
Faltings's theorem
Holes of graph in complex numbers

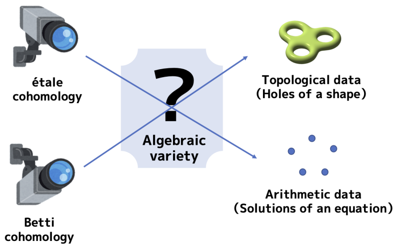
Mathematics is very good at finding a new approach to a difficult problem by connecting two things that seem to be completely different at a glance.

Wonders come from shapes

Such mysterious connections come from a type of shape called algebraic variety. We can see algebraic varieties via mathematical observation devices, i.e., cohomologies.

(Algebraic varieties are an important research subject in pure mathematics, and in applied fields such as cryptography.)

Various information, such as number of solutions and number of holes, can be obtained by observing algebraic varieties via different cohomologies.

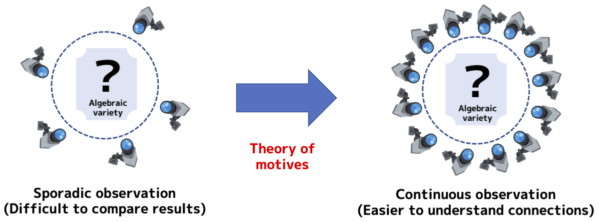


We can "see" algebraic varieties by using cohomologies.

Different cohomologies extract different types of data of algebraic varieties.

How to compare different data

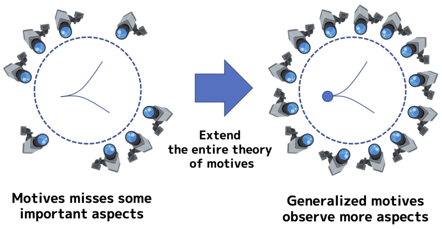
There are many cohomologies to collect different types of data, but it is not easy to find their relations just by looking at them individually. To overcome this difficulty, the theory of motives was developed to continuously observe different aspects of algebraic varieties.



Thanks to the theory of motives, mathematicians could reveal many new and deeper hidden connections.

Towards ultimate observation: generalized motive

However, some important data, such as singularity, cannot be collected by using the theory of motives. In our previous research, we have developed the theory of generalized motives to overcome this disadvantage.



Singular points reflect important properties of a shape.
The theory of motives cannot observe most of the data contained in singular points.

Through high-precision observation using the theory of generalized motives, we will explore further hidden connections in the world of numbers.

References

[1] B. Kahn, H. Miyazaki, S. Saito, T. Yamazaki, "Motives with modulus, III," *Annals of K-theory* (to appear).
[2] B. Kahn, H. Miyazaki, S. Saito, T. Yamazaki, "Motives with modulus, II," *Épijournal de Géométrie Algébrique*, Vol. 5, epiga:7115, 2021.
[3] B. Kahn, H. Miyazaki, S. Saito, T. Yamazaki, "Motives with modulus, I," *Épijournal de Géométrie Algébrique*, Vol. 5, epiga:7114, 2021.
[4] B. Kahn, H. Miyazaki, "Topologies on schemes and modulus pairs," *Nagoya Mathematical Journal*, Vol. 244, pp. 283–313, 2021.
[5] H. Miyazaki, "Nisnevich topology with modulus," *Annals of K-theory*, Vol. 5, pp. 581–604, 2019.

Contact

Hiroyasu Miyazaki / Institute for Fundamental Mathematics
Email: cs-openhouse-ml@hco.ntt.co.jp

14

Robust speech summarization against speech recognition errors

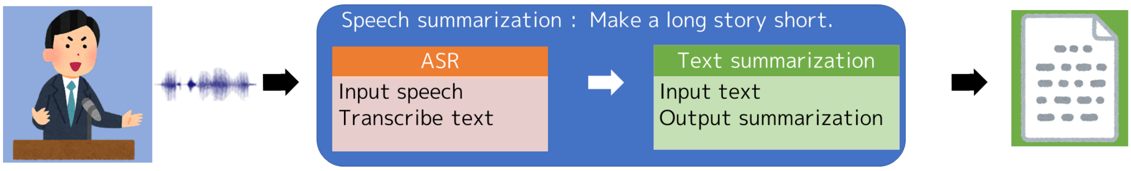
“Huh? What do you mean?” Summarize a long story short

Abstract

Speech summarization aims at creating a summary from a long talk. It is an essential technology if we realize AI systems that can correctly understand human speech. One way to realize speech summarization is cascading automatic speech recognition (ASR) and text summarization. One issue of such approaches is that it is difficult to avoid ASR errors, which degrade the performance of summarization. To alleviate this problem, we propose a robust speech summarization against ASR errors. Our proposed system considers multiple ASR results and looks at the context and relationship between words to generate an accurate summary, even if each ASR result contains errors. The idea we proposed is general and can also be applied to other tasks such as speech translation. This research brings us one step closer to realizing machines that can deeply understand humans, by not only transcribing speech word-by-word but also accessing its meaning and intent.

Mishear but still understand correctly

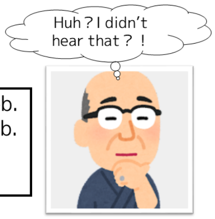
Speech summarization is a technology that summarizes the main points from a long speech, such as a lecture. It can be realized by combining automatic speech recognition (ASR) and text summarization.



Speech summarization： Make a long story short.

ASR
Input speech
Transcribe text

Text summarization
Input text
Output summarization



Problem: **ASR errors seriously affect text summarization!**

Announcements: We will reconsider the content of Salmon lunchbox. The revised result will be announced on the web.
ASR result: We will reconcile the content of Solomon lunchbox. The revised result will be a sold on the web.

Summary： **We will sell a Solomon lunchbox.**

Speech summarization robust against ASR errors

Difficult to achieve perfect ASR

- We exploit results from various ASR systems showing different error tendencies, and expect that the correct meaning can be extracted from the multiple ASR results

Summarize text without assuming that ASR is perfect

- We generate an accurate summary by combining the multiple recognition results
- We utilize a state-of-the-art natural language processing model (BERT*) to model word meanings and relationships.

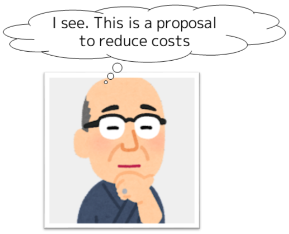
* Bidirectional Encoder Representations from Transformers

ASR A	We will reconcile the content of Solomon lunchbox. The review result will be a sold on the web.
ASR B	We will ________ the content of ______ lunchbox. The review result will be announced on the web.
ASR C	We will reconsider the content of Salmon _____. The review result will be announced on the web.

ASR hypothesis fusion considering context (BERT-based summarization model)

We find out the correct summary by comparing the meaning and content of each word in the ASR results A, B, and C, even if each recognition result individually is wrong

- “Salmon lunchbox” is more natural than “Solomon lunchbox”
- “reconsider” is related to “revised result”
- The “revised result” should be “announced” not “sold.”



Summary： Changed the content of the Salmon lunchbox.

The proposed method **recovers up to 30% of performance degradation due to ASR errors!**

References

[1] T. Kano, A. Ogawa, M. Delcroix, S. Watanabe, “Attention-based multi-hypothesis fusion for speech summarization,” in *Proc. IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pp. 487–494, 2021.

[2] T. Kano, A. Ogawa, M. Delcroix, S. Watanabe, “ASR hypothesis fusion using BERT for speech summarization,” in *Proc. The 2022 Spring Meeting of the Acoustical Society of Japan (ASJ)*, 2022.

[3] T. Kano, A. Ogawa, M. Delcroix, S. Watanabe, “Integrating multiple ASR systems into NLP backend with attention fusion,” in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2022.

Contact

Takatomo Kano / Signal Processing Research Group, Media Information Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp

15

Deep generative model for learning depth and bokeh effects only from natural images

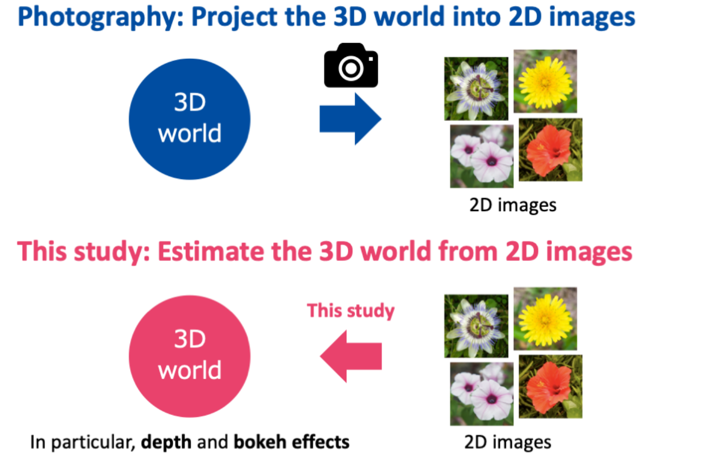
Flexible bokeh renderer based on predicted depth

Abstract

Based on their experience and knowledge, humans can estimate depth and bokeh effects from the corresponding 2D images. However, computers have difficulty in doing so because they lack the necessary experience and knowledge. To overcome this limitation, we propose a novel deep generative model that can control bokeh effects based on predicted depth. If it is possible to collect pairs of 2D images and 3D information, learning a 3D predictor is simple because of direct supervision. However, collecting such data is often difficult or impractical owing to the requirement for specific sensors, such as a depth sensor or stereo camera. To eliminate this requirement, we developed the world's first technology that enables learning depth and bokeh effects only from standard 2D images. Because we live in a 3D world, a human-oriented computer must understand the 3D world. This study addresses this challenge by eliminating an application boundary in terms of data collection cost. We expect that this technology will cultivate a new field of 3D understanding.

❶ Objective: Understand 3D world from 2D images

Solve inverse problem of photography


Photography: Project the 3D world into 2D images

3D world

2D images

This study: Estimate the 3D world from 2D images

3D world

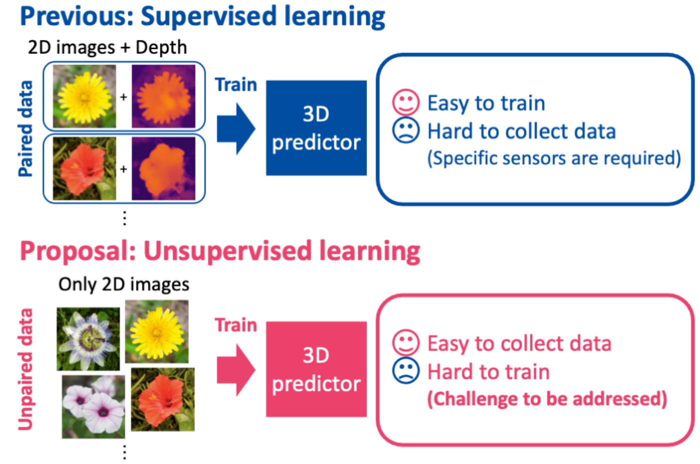
This study

2D images

In particular, **depth** and **bokeh effects**

❷ Approach: Unsupervised learning

Focus on unsupervised learning, where data collection is easy


Previous: Supervised learning

2D images + Depth

Paired data

Train

3D predictor

☺ Easy to train

☹ Hard to collect data

(Specific sensors are required)

Proposal: Unsupervised learning

Only 2D images

Unpaired data

Train

3D predictor

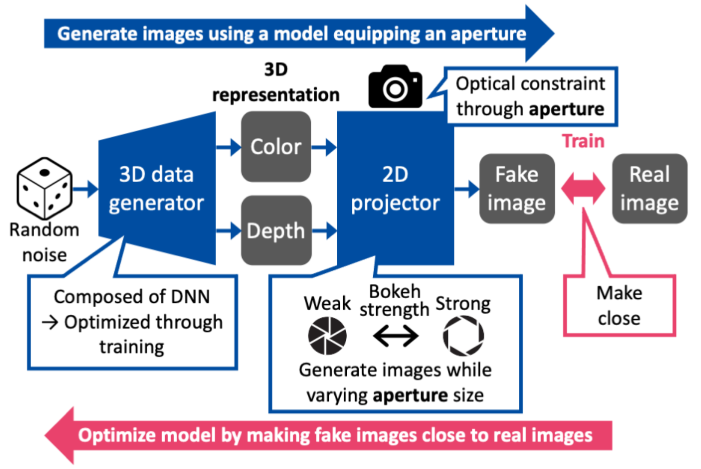
☺ Easy to collect data

☹ Hard to train

(Challenge to be addressed)

❸ Method: Deep generative model equipping aperture

Obtain 3D representation consistent with optical constraint


Generate images using a model equipping an aperture

3D representation

Optical constraint through **aperture**

Random noise

3D data generator

Color

Depth

2D projector

Fake image

Train

Real image

Composed of DNN → Optimized through training

Weak Bokeh strength Strong

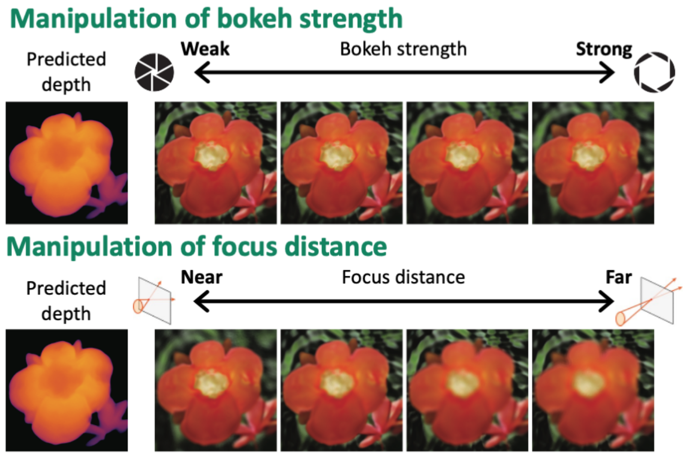
Generate images while varying **aperture** size

Make close

Optimize model by making fake images close to real images

❹ Results: Depth prediction → flexible bokeh control

Able to manipulate bokeh effects based on predicted depth


Manipulation of bokeh strength

Predicted depth

Weak Bokeh strength **Strong**

Manipulation of focus distance

Predicted depth

Near Focus distance **Far**

References

[1] T. Kaneko, "Unsupervised learning of depth and depth-of-field effect from natural images with aperture rendering generative adversarial networks," in *Proc. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR2021)*, pp. 15679–15688, 2021.

[2] T. Kaneko, "AR-NeRF: Unsupervised learning of depth and defocus effects from natural images with aperture rendering neural radiance fields," in *Proc. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR2022)*, 2022 (to appear).

Contact

Takuhiro Kaneko / Recognition Research Group, Media Information Laboratory

Email: cs-openhouse-ml@hco.ntt.co.jp

16

Estimating heart activities from multichannel sounds and ECG signals

Heart health monitoring with sounds and electric signals

Abstract

Early detection of heart problems requires estimation of heart activity based on information that can be easily measured on a daily basis. To this end, we are researching technologies to estimate and visualize the mechanical and electrical activities within the heart based on the non-invasive observations on the surface of the body. Our technique called Physically-Constrained Unsupervised Signal Decomposition (PCUSD) method incorporates a physical heart sound generation model and makes it possible to estimate cardiac vibration components such as opening and closing of valves inside the heart that cannot be directly heard with a conventional stethoscope. In addition, our newly proposed technique called tensor electrocardiography can capture and visualize the action potentials of cardiac muscle cells, and has the potential to detect abnormalities that are not readily apparent in conventional electrocardiograms. Potential applications of these techniques will include a system that allows users to easily assess the condition of their cardiovascular system by themselves which can contribute to early detection of heart diseases such as heart failure, ischemic heart disease, and arrhythmia associated with sudden death. The same system can also be used to support rehabilitation after treatment of heart disease as well as training for healthy people.

Estimation of Activity within the Heart

Tasks: Biometric information ➡ Activity within the heart

Biometric information that can be easily measured non-invasively:

Heart Sounds

Mainly caused by the opening/closing of valves in the heart.

Doctors have used stethoscopes for centuries to listen to heart sounds, a process called auscultation, in order to diagnose the health and condition of the heart.

Electrocardiogram (ECG)

Caused by the action potentials of cardiac muscle cells.

Medical institutions widely use ECG to estimate heart activity.

[Technical hurdles]

Both types of signals are mixtures of signals transferred from multiple internal sources, and therefore, difficult to infer them only from data observed on the surface.

[Approach]

(1) Process biometric data using novel statistical / physical models.

(2) Capture multiple channels of observed data from different locations on the body surface to enable localization of various internal signal sources.

ECG Technique † ⬇ Acoustic Technique † ➡

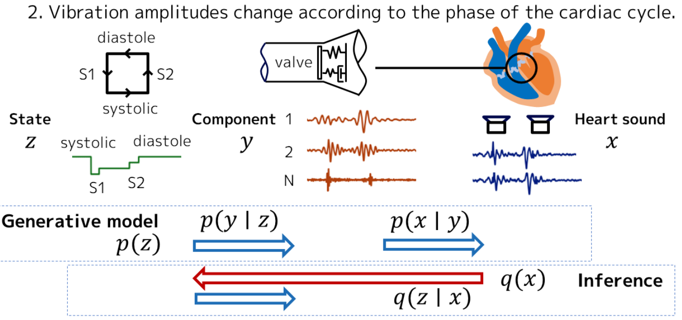
† Investigational (unapproved)

PCUSD (Physically-Constrained Unsupervised Signal Decomposition)

(1) A probabilistic model is defined to describe the physical mechanism of heart sound generation.

Assumptions:

1. Multiple components vibrate to generate heart sounds based on physical models of valves.
2. Vibration amplitudes change according to the phase of the cardiac cycle.



An example of experimental results

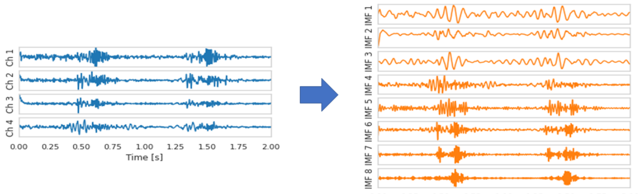
F1	PCUSD (proposed)	Conventional method
S1	**96.1**	86.5
S2	**96.4**	85.7

Improvement of segment estimation accuracy (F1) for S1 and S2 is shown.

Conventional method (right) refers to a decomposition method without a generative model.

(2) Application to multi-channel signals

An example of estimation of 8 vibration components from 4 channels of acoustic signals observed on the body surface.

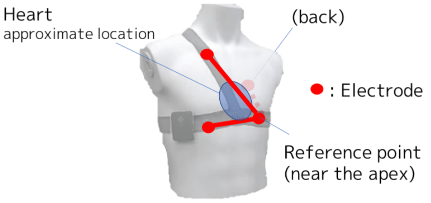


Observed signals (left) and estimated vibration components (right)

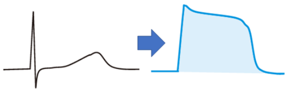
Tensor Electrocardiography

(1) The timing of potential changes (depolarization and repolarization) is statistically modelled with Gaussian distributions. [1]

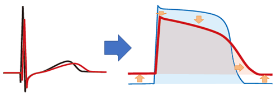
(2) The closest point between the heart and the body surface (near the apex of the heart) is used as the reference point. Electrodes are placed on three nearly orthogonal axes to gather spatial information.



Appearance and electrode arrangement of wearable tensor electrocardiograph



Potential difference observed on the body surface (left) and action potential (right)



Clearer anomaly visualization by tensor electrocardiogram (schematic)

References

[1] S. Tsukada, “Wearable textile electrodes for long-term vector ECG monitoring 'Tensor Cardiography',” in *Proc. ISMICT 2020*, 2020.

[2] R. Shibue, M. Nakano, T. Iwata, K. Kashino, H. Tomoike, “Unsupervised heart sound decomposition and state estimation with generative oscillation models,” in *Proc. EMBC 2021*, pp. 5481–5487, 2021.

Contact

Ryohei Shibue / Basic Research Laboratories
Shingo Tsukada / Basic Research Laboratories
Email: cs-openhouse-ml@hco.ntt.co.jp

17 Crossmodal action unit sequence estimation and image-to-image mapping

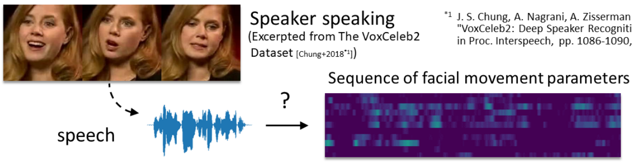
Controlling facial expressions in face image from speech

Abstract

Speech contains not only linguistic information, corresponding to the uttered sentence, but also nonlinguistic information, corresponding to the emotional expression and mood. This information plays an important role in spoken dialogue. This study is the first attempt to estimate the action unit (facial muscle motion parameter) sequence of the speaker from speech alone, assuming that the nonlinguistic information in speech is expressed in the facial expressions of the speaker. Until now, there have been no attempts to estimate action units from speech alone, and how much accuracy could be achieved was not known. This study reveals this for the first time. By combining the action unit sequence estimated from speech with an image-to-image converter, we implemented a system that modifies the facial expression of a still face image in accordance with input speech, making it possible to visualize the expression and mood of speech. Emotional expressions and moods have traditionally been treated symbolically, assigning discrete subjective labels. In contrast, action units are suitable as continuous quantities for expressing emotional expressions and moods, and we have shown that action units can be estimated from speech in this study. In the future, we expect to open up a variety of new applications that simultaneously utilize speech and face images, such as speech synthesis that matches facial expressions and face image generation that matches speech.

Estimating face movement from speech

✓ If face movement can be predicted from speech, ...


Speaker speaking
(Excerpted from The VoxCeleb2 Dataset [Chung+2018*1])
speech
?
Sequence of facial movement parameters

[*1] J. S. Chung, A. Nagrani, A. Zisserman: "VoxCeleb2: Deep Speaker Recognition," in Proc. Interspeech, pp. 1086-1090, 2018.

- it can be used to visualize nonlinguistic information in speech
- it can be used as useful nonlinguistic-information-related feature for speech synthesis and voice conversion applications

✓ Is this task solvable and how difficult is it?

➡ The aim of this study is to answer these questions

Deep learning approach using speaking face-tracks

✓ As quantities that represent facial movements, we focus on the **facial action units (AUs)**[*2]

[*2] **Facial muscular activity units that are related to the contraction or relaxation of specific facial muscles**

✓ Train neural network that predicts AU sequence from speech

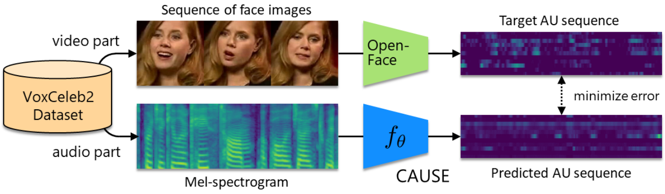
"Crossmodal Action Unit Sequence Estimator (CAUSE)"

Approach

By using many speaking face-tracks, we train CAUSE so that
- AU sequence extracted using OpenFace[*3] from the video part and
- AU sequence predicted by CAUSE from the audio part

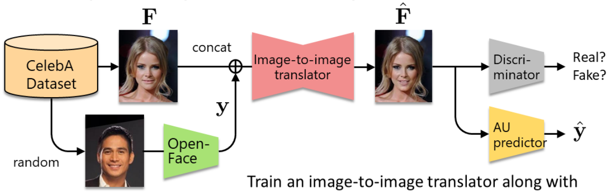
become consistent

[*3] Open-source facial behavior analysis toolkit



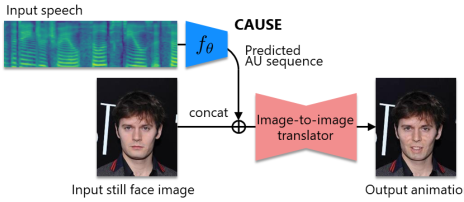
Crossmodal face image control

✓ Train Image-to-image translator using GANimation [Pumarola+2018]



Train an image-to-image translator along with an AU predictor and a discriminator, which scores how realistic a given image is

✓ Convert still face image in accordance with predicted AU sequence



➡ Allows us to control facial expression using speech

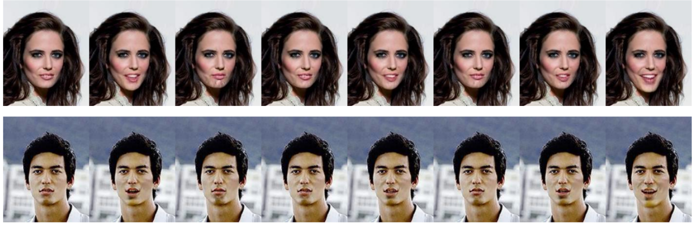
※ All the face images are excerpted from The CelebA Dataset [Liu+2015*4]

[*4] Z. Liu, P. Luo, X. Wang, X. Tang: "Deep Learning Face Attributes in the Wild," in Proc. ICCV, pp. 3730-3738, 2015.

Face image control experiment

Other examples can be found here:

✓ Examples of animations generated from same speech



✓ Generated animations were more natural when controlled by AUs than when controlled by probability vectors of emotional states (neutral, happiness, surprise, sadness, anger, disgust, fear, contempt)

References

[1] H. Kameoka, T. Kaneko, S. Seki, K. Tanaka, "CAUSE: Crossmodal action unit sequence estimation from speech," submitted to The 23rd Annual Conference of the International Speech Communication Association (Interspeech 2022).

Contact

Hirokazu Kameoka / Recognition Research Group, Media Information Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp

Image blending with content-adaptive visibility predictor

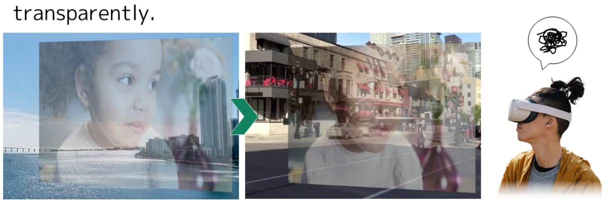
Maintain comfortable visibility anytime, anywhere

Abstract

The visibility of an image semi-transparently overlaid on another image significantly varies depending on the content of images. This makes it difficult to maintain desired visibility when image content changes. To tackle this problem, we developed a perceptual model to predict the visibility of arbitrarily combined blended images. Specifically, we clarified that the influence of each feature on the overall visibility depends on the distribution of features in the presented content, such as fineness and colors. Using the perceptual model that incorporates this effect, we achieved better control on the visibility of blended images than existing techniques. As AR technology matures, there will be more and more situations where information is displayed semi-transparently across our entire visual field. Our technique will make it possible to maintain a comfortable visibility level for such information. It also enables more intuitive control of visibility when blending images with a video editing software.

Visibility of blended images

In media that cover the entire field of view (e.g., AR/VR), image information often needs to be displayed semi-transparently.

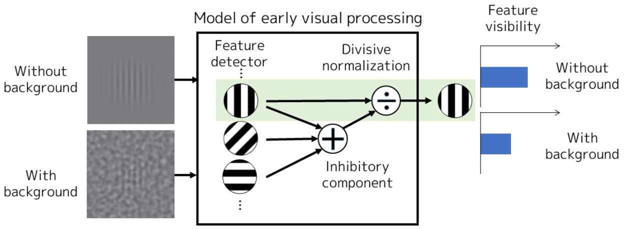


It is difficult to maintain constant visibility in situations where image content and background varies.

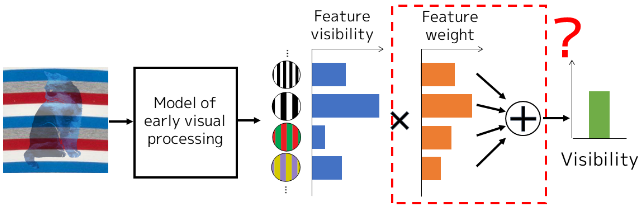
A model that can accurately predict visibility is required.

Visual mechanism related to visibility

The phenomenon in which the visibility of image feature (e.g., fineness or color of patterns) is reduced by the background can be explained by the inhibitory mechanism in the visual system.



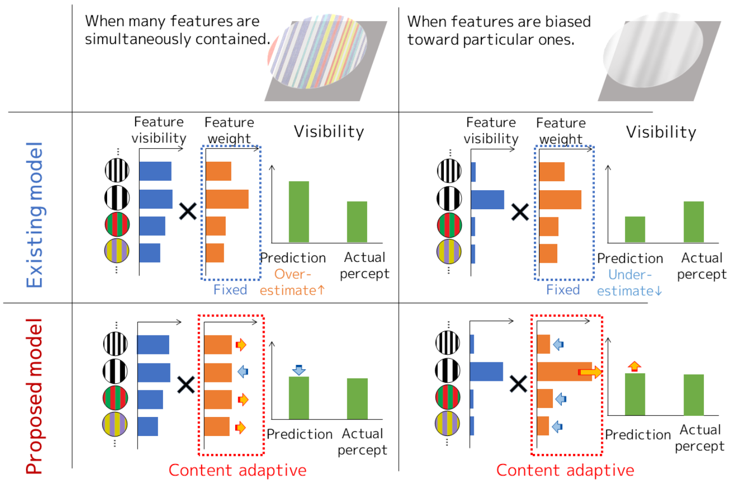
It has not been well understood how the visibility of each feature contributes to the overall visibility when many features are included at the same time, as in natural images



Technical point 1: Content-adaptive feature aggregation

Existing models Weights each feature with a predetermined value.

Proposed model Adaptively adjusts the weighting of each feature based on the distribution of features in the displayed content.

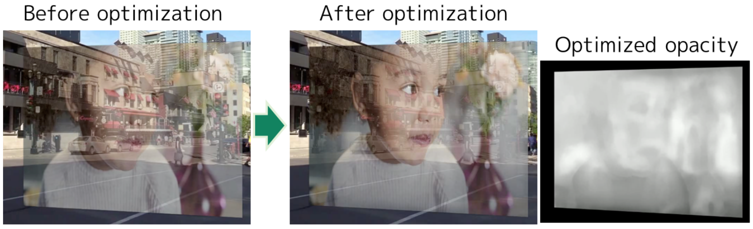


Existing models tend to overestimate visibility for images with many features and underestimate visibility for images with few features.

The proposed model can predict visibility with significantly higher accuracy!

Technical point 2: Visibility-based image blending

Automatically optimizes image opacity to maintain user-specified visibility levels.



This research is the result of a collaborative project with the University of Tokyo

References

[1] T. Fukiage, T. Oishi, "A computational model to predict the visibility of alpha-blended images," *Vision Sciences Society Annual Meeting 2021* (Abstract published in: Journal of Vision, Vol. 21, No. 2493).

[2] T. Fukiage, T. Oishi, "Perception-based image blending based on content-adaptive visibility predictor," in *Proc.* Special Interest Group on Computer Vision and Image Media (CVIM), Vol. 229, No. 45, pp. 1–8, 2022 (in Japanese).

Contact

Taiki Fukiage / Sensory Representation Research Group, Human and Information Science Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp

19

Elucidating effects of social cues on infants' object learning

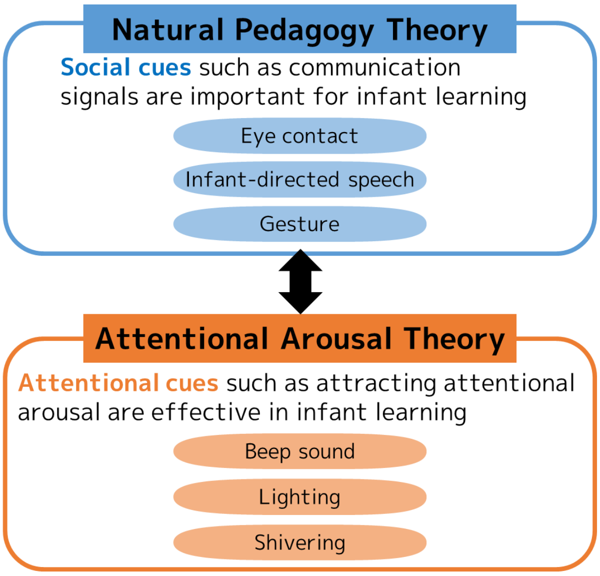
Gazing and talking help infants learn

Abstract

Although infants learn a variety of knowledge from information obtained from the environment, this learning process has not been fully clarified. This study used an experimental psychological approach to determine whether social cues versus attentional cues might affect infants' learning at different levels. By focusing on a new task to clarify the learning process, our experiments showed that although both attentional cues and social cues affected infants' gaze following, only social cues facilitated their object learning. Furthermore, these social cues influenced the infants' vocabulary acquisition. These findings provide evidence that social cues play a distinct role in infant learning and support Natural Pedagogy Theory, which models human learning mechanisms. We believe our study will not only establish theories on how humans acquire language and knowledge but also contribute to practical childcare and education-support methods such as parent training.

Infant learning theory

Two conflicting theories on cues that help infants learn:



Natural Pedagogy Theory

Social cues such as communication signals are important for infant learning

- Eye contact
- Infant-directed speech
- Gesture

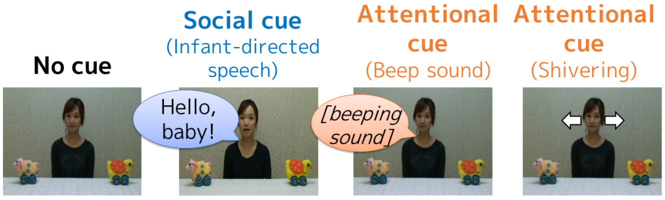
Attentional Arousal Theory

Attentional cues such as attracting attentional arousal are effective in infant learning

- Beep sound
- Lighting
- Shivering

Approach of this study

Effects of social cues and attentional cues on infant learning examined by *experimental psychological approach*

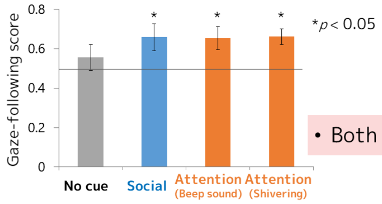


Participants: 140 9-month-old infants

Outcome 1: Clarifying role of social cues

Gaze-following test

Measures whether infants look at objects in the direction of model's gaze

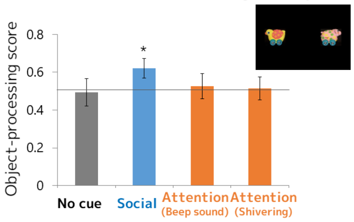


- Both cues facilitate gaze following

Object-learning test

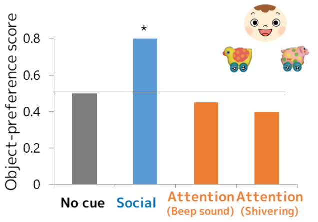
(1) Object-processing test

Measures whether infants recognize objects



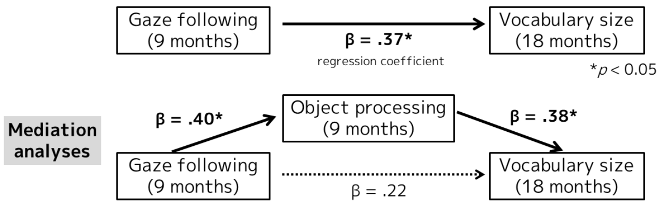
(2) Object-preference test

Evaluation of infants' object choice



- Social cues play a distinct role in infant learning
➡supports Natural Pedagogy Theory

Outcome 2 : Vocabulary-acquisition process


Gaze following (9 months)
β = .37*
regression coefficient
Vocabulary size (18 months)
*$p < 0.05$
Mediation analyses
β = .40*
Object processing (9 months)
β = .38*
Gaze following (9 months)
β = .22
Vocabulary size (18 months)

- Gaze following at 9 months promotes object processing, and it affects vocabulary development at 18 months
➡ Social cues (gaze) and vocabulary acquisition are associated

References

[1] Y. Okumura, Y. Kanakogi, T. Kobayashi, S. Itakura, "Individual differences in object-processing explain the relationship between early gaze-following and later language development," *Cognition*, Vol. 166, pp. 418–424, 2017.
[2] Y. Okumura, Y. Kanakogi, T. Kobayashi, S. Itakura, "Ostension affects infant learning more than attention," *Cognition*, Vol. 195, 104082, 2020.
[3] Y. Okumura, "Social learning in infancy: How and from whom babies learn," *University of Tokyo Press*, 2020.

Contact

Yuko Okumura / Interaction Research Group, Innovative Communication Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp

Social factors influencing adoption of COCOA

Why do people hesitate to use contact tracing apps?

Abstract

Digital contact tracing apps (e.g. COCOA) have been identified as a promising approach to control the spread of viruses, but their usage has been low. Therefore, we investigated people's attitudes about installing and using COCOA, and found that their decisions were shaped by social norms, as well as protecting themselves from financial loss, prejudice, and discrimination. We found that, even if installed, efforts to protect oneself from financial risk and prejudice may cause people not to use the app effectively. Based on this, we identify ways to address people's fears in order to encourage effective use, which is necessary to control the pandemic. The results have implications for the design of future communication technologies that address large collective goals while preserving individual rights. By realizing this, we can help overcome important social problems such as climate change and public health emergencies.

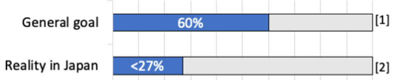
Covid-19 Contact Confirming app (COCOA)

COCOA is a contact tracing app released by the Japanese government.

It uses BlueTooth to detect when people are in close contact and sends an exposure notification to people who have been near an infected person.

The more people use a contact tracing app, the more COVID-19 cases will be reduced.

Worldwide, contact tracing app adoption is much lower than hoped.



→ There is a need to understand why people use or do not use contact tracing apps.

[1] Hinch, R., et al.. “Effective configurations of a digital contact tracing app: A report to NHSX,” 2020.
[2] 厚生労働省, “新型コロナウイルス接触確認アプリ,” 2022. https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/cocoa_00138.html

Background and research design

Past work has found that decisions to install a contact tracing app are influenced by, perceived effectiveness, ease of use, social influence, privacy concerns (surveillance), and other factors.

Once installed, there are two ways to *use* COCOA:

Register to the app if you test positive for COVID-19.

Respond properly if sent an exposure notification.
e.g.: Self-isolate at home, Get a PCR test, Tell family/friends, Tell employer/boss

We extend past work about contact tracing apps by investigating **use after installation** and **fit with daily life.**

Research method: Survey (n=153) & interviews (n=15)

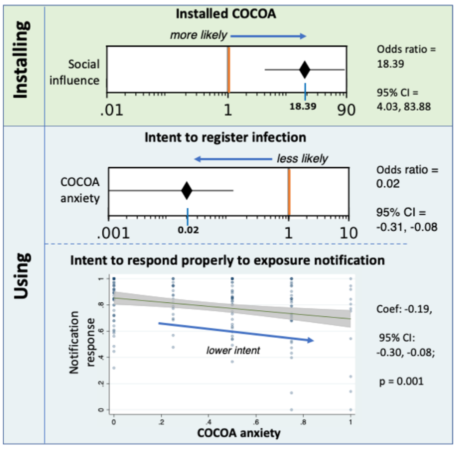
Dependent variables:

- *Installed COCOA* (yes/no)
- *Intent to register if infected* (5-point ordinal)
- *Intent to respond properly to notification* (factor variable)

Key independent variables:

- *Social influence:* Knows at least one app user (yes/no).
- *COCOA anxiety*: Believe COCOA increases anxiety (5-point ordinal).

Factors affecting adoption decisions



Social influence
= More likely to **install**

Believe that using COCOA will increase anxiety
= Less likely to use properly:
a) registering infection.
b) responding to notification

Even if installation increases, proper use may be low.

29% said COCOA would make them **more anxious**. ***Why?***

Fear of stigmatization in community

“Since I live in the countryside, people will immediately identify who I am and the rumors after infection will be very serious.” (P15-S)

Fear of financial loss

“I'm a little worried [about getting a notification] because I see in the news that people will lose their job when they disclose to the workplace.” (P15-S)

Consequences for COCOA use

- Hiding infection information from others
- Not registering to COCOA if infected

Implications and future work

App can create fears of social harm →
Introduce design features to create social rewards.

Design beyond the app→
Collaborate with institutions to reduce stigma (e.g., local governments, workplaces, schools).

Beyond the pandemic →
Next steps: Build on this research to use personal communication technology to address future collective challenges (e.g., climate change, caring for eldery)

References

[1] J. Jamieson, N. Yamashita, D.A. Epstein, Y. Chen, “Deciding if and how to use a COVID-19 contact tracing app: Influences of social factors on individual use in Japan,” in *Proc. ACM Hum.-Comput. Interact. 5, CSCW2, Article 481, (CSCW'21)*, pp. 1–30, 2021.
[2] J. Jamieson, D.A. Epstein, Y. Chen, N. Yamashita, “Unpacking intention and behavior: Explaining contact tracing app adoption and hesitancy in the United States,” in *Proc. the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*, pp. 1–14, 2022.

Contact

Jack Jamieson / Interaction Research Group, Innovative Communication Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp

21

Modifying pitched ball perception by VR

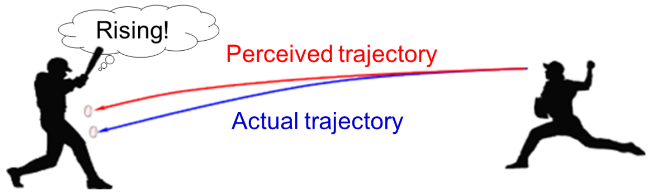
Is the rising fastball a perceptual illusion?

Abstract

Baseball batters sometimes feel that the pitched fastball rises as it approaches the home plate. While some physical parameters of the pitched ball, such as ball spin rate and axis orientation, can generate the rising perception, we propose that the pitching motion-related information can also cause the "rising" ball effect, since the batters are known to watch pitching motion to predict pitched ball behavior. We used a head-mounted display to evaluate the rising perception of fastballs in elite baseball players. A virtual reality (VR) system was developed that manipulated pitching motion duration with fixed ball behavior. Altering the pitching motion duration changed the rising perception, suggesting that the batters predict ball behavior based on the pitching motion dynamics and the prediction generate the "rising" illusion for fastballs. Our VR system will be useful not only in correcting athletic perception but also enhanced cognitive training in many sports.

Pitched ball perception

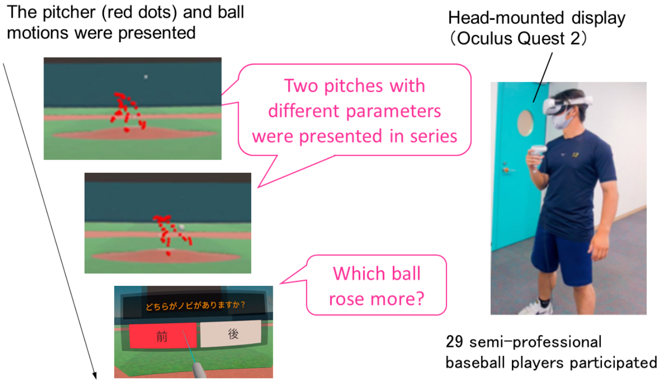
Characteristics of pitched ball depend on more than the ball's physical parameters such as ball speed and trajectory.



Subjective attributes such as the rising perception also important, the perceptual mechanisms remain unclear.

Evaluating the rising perception by VR

Using a virtual reality (VR) system, we manipulated some pitching parameters such as pitching motion duration and ball speed with fixed ball trajectory, and then evaluated the rising perception.



The pitcher (red dots) and ball motions were presented

Two pitches with different parameters were presented in series

Which ball rose more?

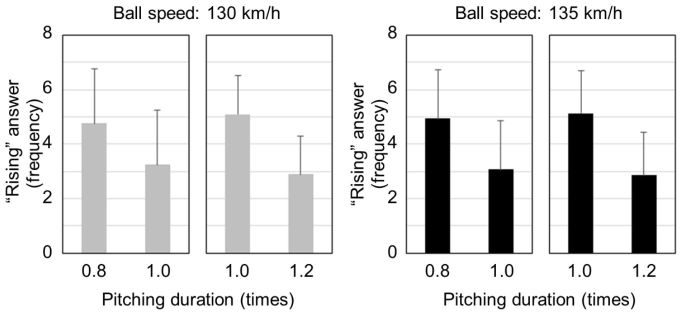
Head-mounted display (Oculus Quest 2)

29 semi-professional baseball players participated

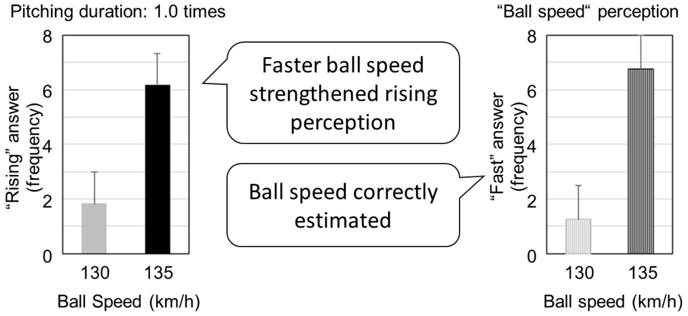
＊This research is in collaboration with Keio University.

Manipulating pitching motion duration changes the rising perception

Manipulating pitching motion duration (0.8, 1.0 or 1.2 times)



Regardless of ball's physical attributes such as ball speed, shorter pitching motion strengthened rising perception



Faster ball speed strengthened rising perception

Ball speed correctly estimated

- **The rising fastball perception was generated not only by ball's physical behavior but also by pitching motion information.**
- **The rising perception includes a perceptual illusion driven by pitching motion-based prediction.**
- **Visual manipulation with VR will be useful not only in correcting athletic perception but also enhanced cognitive training.**

References

[1] T. Fukuda, A. Endo, M. Sugimoto, T. Kimura, "How do elite baseball batters perceive a "rising" fast ball?," in *Proc. North American Society for the Psychology of Sport and Physical Activity 2022 Annual Conference*, 2022.

Contact

Toshitaka Kimura/ Kashino Diverse Brain Research Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp

The relationship between frontal neural oscillation and performance

Mental skills of esports experts revealed by brain measurement

Abstract

In esports, where the outcome is less dependent on physical factors, the importance of mental preparation for the match is considered to be significant. In particular, skilled esports players have superior strategic decision to optimize their behavioral patterns according to their opponents, and emotional control to stay calm under pressure at a critical phase. However, it is not known how the aforementioned abilities affect the outcome of a match. Through EEG measurements during a match and post-match questionnaires, we found that strategic decision is important at the beginning of the match and emotional control is important at the end of the match. In addition, neural oscillations in relation to strategic decision and emotional control were observed at the frontal brain region. By applying these findings, we aim to establish a new training method to bring the mental state of esports players closer to the ideal state for matches.

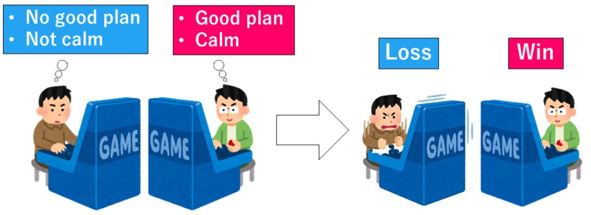
Importance of mental preparation in esports

- **Mental preparation** is important in esports
- Skilled players have superior **strategic decisions** and **emotional control**

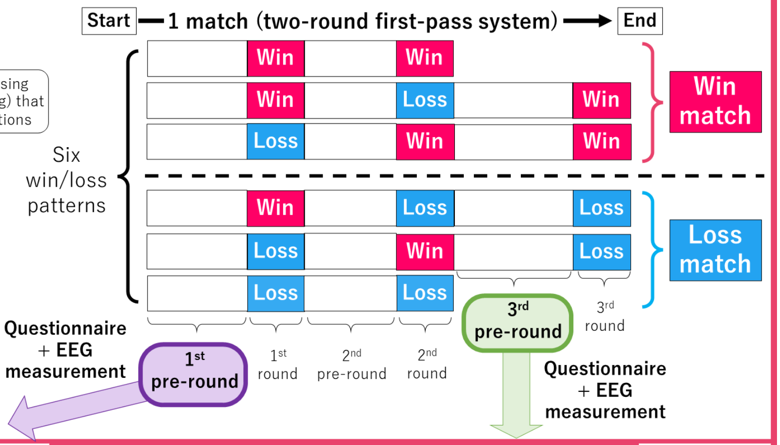
Strategic decision: Inferring in advance the most effective behavioral patterns to take during a round based on the opponent's characteristics

Emotional control: Consciously suppressing the mental agitation (anxiety about losing) that occurs before a round in important situations

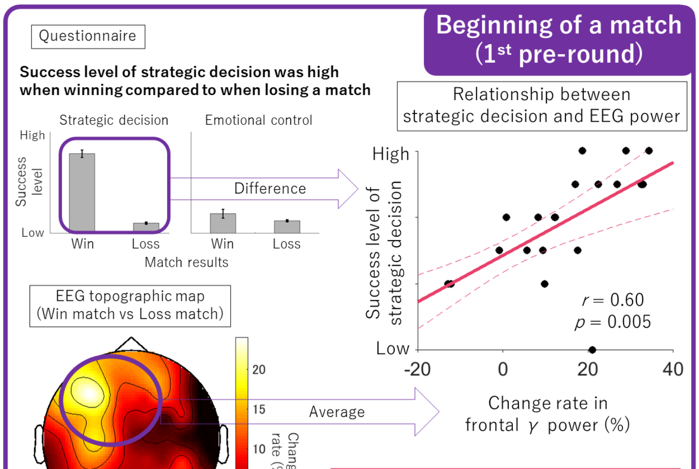
To investigate how these two abilities affect match performance, we focused on **a fighting video game (FVG)**, where the necessary abilities change depending on the situation of a match



Match format and win/loss patterns of a FVG



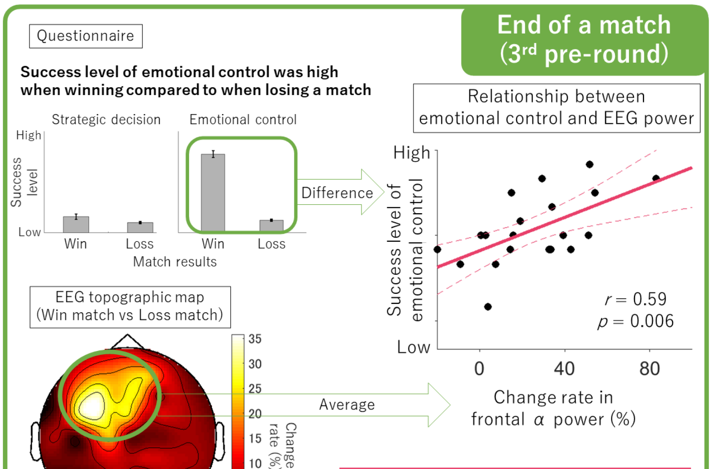
Strategic decision and related neural oscillations at the beginning of a match



Frontal γ power increased when winning compared to when losing a match

- Strategic decision is important at the beginning of a match
- Increased frontal γ power is associated with strategic decision

Emotional control and related neural oscillations at the end of a match



Frontal α power increased when winning compared to when losing a match

- Emotional control is important at the end of a match
- Increased frontal α power is associated with emotional control

References

[1] S. Minami, K. Watanabe, N. Saijo, M. Kashino, "Amplitude of neural oscillations in the parietal area is associated with the results of esports competitions," in *Proc. IEEE Conference on Games (CoG)*, 2021 (in press).

Contact

Sorato Minami / Kashino Diverse Brain Research Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp

23

Approaches to cochlear implant and binaural processing

Unveiling the auditory system with a neural network

Abstract

Although there has been much research on human auditory characteristics, it is difficult to directly address the question of what kinds of input and training lead to the acquisition of these characteristics. In this work, we tackled the clinical and academic aspects of the question by using artificial neural networks (ANNs), and obtained new findings in each case. (1) It is known that people with hearing loss who wear cochlear implants (CIs) have difficulty with pitch perception, but we confirmed that the cochlear implant signal contains a certain amount of pitch information, suggesting that the difficulty in pitch perception is mainly due to physiological factors. (2) By measuring the response of a single unit in an artificial neural network trained to recognize natural sounds, we found out the ANN units (neurons) with the binaural processing characteristics were equivalent to those found in the auditory system of animals. We believe that cochlear implant users may be able to achieve normal pitch perception under a clean environment after an appropriate rehabilitation. We also hope to further develop AI technology and CI devices that behave in a human-like manner by advancing auditory information processing technology that is consistent with the auditory nervous system.

Understanding auditory mechanisms with artificial neural networks (ANNs)

Topic (1): Pitch information transmitted by the cochlear implant (CI)※

※ An artificial organ to partially restore hearing loss caused by damage to the inner ear

Background:

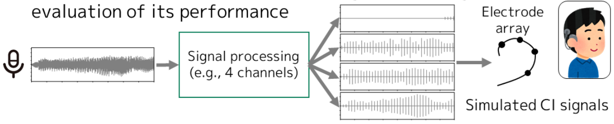
- Cochlear implantation significantly restores speech perception
- Pitch perception is difficult and varies considerably between individuals

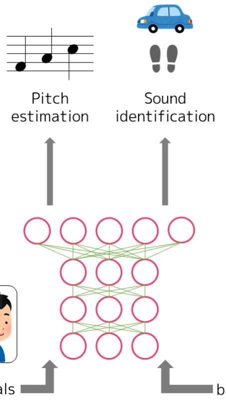
Question:

- Does the signal transmitted by the CI contain the information necessary for pitch estimation?

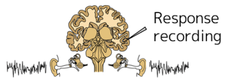
Approach:

- Pitch estimation from simulated CI signals by using ANNs and evaluation of its performance





Topic (2): Characteristics of neuronal response to binaural sound



Background:

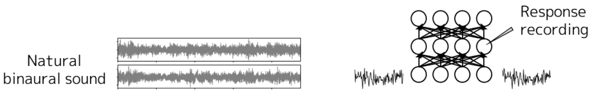
- The time differences between when sound reaches the left and right ear (interaural time difference (ITD)) is important for sound source localization and identification
- Neuronal response characteristics have been investigated

Question:

- How are response characteristics to ITD acquired?

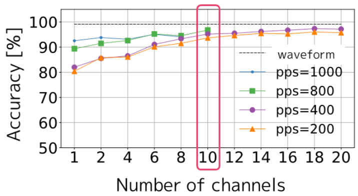
Approach:

- Training ANNs to identify sound sources from natural binaural sounds and measuring single cell responses to stimuli with ITD

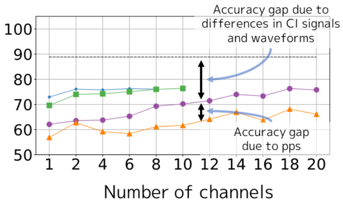


(1) A certain amount of pitch information is contained in the cochlear signal

- When the target sound is presented alone (clean environment (a)), if CI signals have a sufficient number of channels (almost 10), the accuracy is comparable to that of the waveform
  - Containing pitch information in CI signals comparable to that in the waveform under a clean environment
- In the presence of background noise (noisy environment (b)), the accuracy of CI signals is worse than that of the waveform, and improves as pulse per second (pps) increases
  - Pitch perception becomes difficult under noisy environments
  - Finer temporal resolution plays an important role



(a) evaluation = clean environment

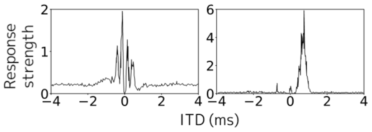


(b) evaluation = noisy environment

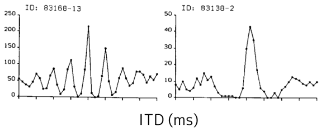
- The difficulty in pitch perception is more likely due to physiological factors than to the signal transmitted by the CI device

(2) Emergence of ITD response characteristics in natural sound identification

- Response strength of ANN neurons varies with stimulus ITD
  - ITD response characteristics qualitatively similar to those of animals are also evident in ANNs



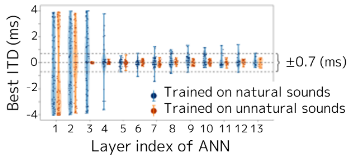
Examples of ITD response characteristics in ANNs

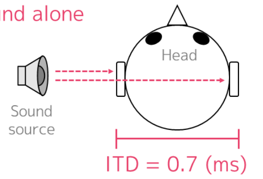


Examples of ITD response characteristics in animals

(Yin et al., 1986, J. Neurophysiol.)

- ANN shows high response strength in the range of ITDs that humans naturally experience
  - No use of human body shape information for ANN training
  - The range becomes narrower when trained on unnatural sounds
  - ANN captures the natural environmental structure for humans from the information contained in sound alone





References

[1] T. Ashihara, S. Furukawa, M. Kashino, “F0 estimation from simulated cochlear-implant signals by using a DNN model,” *Spring Meeting of Acoustic Society of Japan*, 2022.

[2] T. Koumura, H. Terashima, S. Furukawa, “Emergence of ITD selectivity in a deep neural network trained for binaural natural sound detection,” in *Proc. 42nd Association for Research in Otolaryngology (ARO) MidWinter Meeting*, 2019.

[3] TC. Yin, JC. Chan, DR. Irvine, “Effects of interaural time delays of noise stimuli on low-frequency cells in the cat's inferior colliculus. I. Responses to wideband noise,” *Journal of Neurophysiology*, pp. 280–300, 1986.

Contact

Takanori Ashihara / Human Informatics Laboratories
Takuya Koumura / Sensory Representation Research Group, Human and Information Science Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp

24

Autonomic and endocrinological variation by meditation style

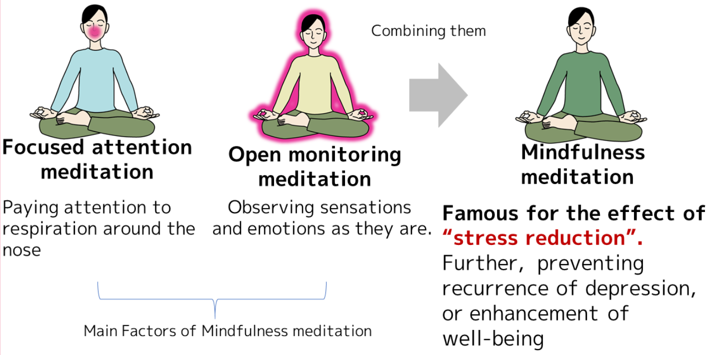
How does mindfulness meditation reduce stress?

Abstract

"Mindfulness meditation" can reduce stress by manipulating our attention. However, the physiological mechanisms have not yet been clarified. In this study, we examined how mindfulness meditation changes the activities of autonomic nerves and secretion of the stress hormone cortisol. Because mindfulness meditation mainly consists of "focused attention" and "open monitoring" meditation, we developed vocal instructions for each. We measured heart rates and took saliva samples to evaluate the strength of autonomic activities and cortisol levels, respectively. We found that focused attention meditation increased parasympathetic activity, while open monitoring meditation increased sympathetic activity with the reduction of cortisol levels. We hope to reveal the physiological, psychological, and neural mechanisms of mindfulness mediation and develop new types of meditation based on our scientific findings. We think we can contribute to people's well-being through social implementation of new types of meditation in the future.

1. Mindfulness meditation consists of “focused attention” and “open monitoring” meditations.

Mindfulness :
The status of monitoring sensation, emotion and thought with an attitude of acceptance; every momentary experience is accepted as it is.



Focused attention meditation

Paying attention to respiration around the nose

Open monitoring meditation

Observing sensations and emotions as they are.

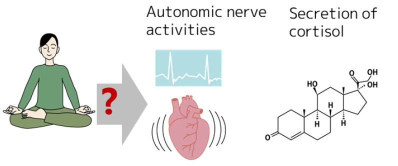
Main Factors of Mindfulness meditation

Mindfulness meditation

Famous for the effect of “stress reduction”.
Further, preventing recurrence of depression, or enhancement of well-being

2. Earlier studies have not revealed the physiological mechanisms of the two meditation styles.

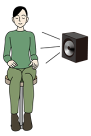
Background:
There has not been consistent thought of the physiological mechanisms, by which mindfulness meditation can reduce stress.



Point to examine:
The physiological mechanism of the effect of focused attention and open monitoring meditation.

No consistency in findings about the effects of mindfulness mediation on stress-related physiological activities, such as autonomic nerve activities and the secretion of stress hormone cortisol.

Solution:
We developed 30-min-long voice instruction for focused attention and open monitoring meditations, respectively.
(Fujino et al., 2019 Japan J. Mindfulness)

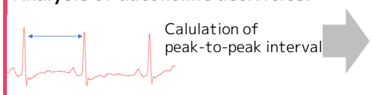


“Now we are starting the exercise for the power of concentration. The power of concentration is … ” the instruction for focused attention meditation explains exercise to help participants maintain attention to respiration around the nose when the mind wanders.

“Now we are starting the exercise for awareness. Awareness is … “ the instruction for open monitoring meditation explains exercise to help participants observe sensations and emotions as they are.

3. We revealed the differential effects of focused attention and open monitoring meditations on autonomic activities and cortisol level.

Analysis of autonomic activities:



Time-domain heart rate variability analysis was performed to calculate
- SDNN
(standard deviation of normal-to-normal interval)
- rMSSD
(root mean square of successive differences)

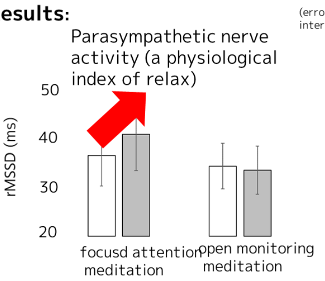
Analysis of stress hormone level:

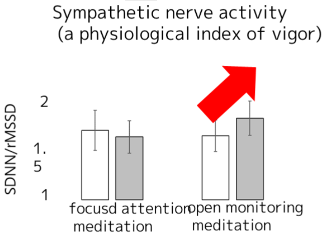
Saliva samples were collected to measure the level of cortisol.

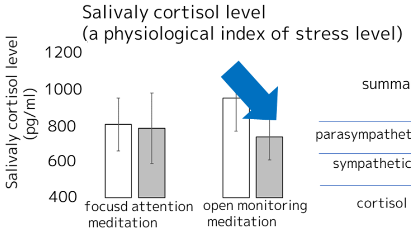
Results:

(error bar: 95% confidence interval)

□ : before meditation
▨ : during meditation







summary	focusd attention meditation	open monitoring meditation
parasympathetic activity	UP	no change
sympathetic activity	no change	UP
cortisol level	no change	DOWN

During focusd attention meditation

By keeping their focus on respiration, they can relax without any disturbance induced by any information other than respiration.

During open monitoring meditation

Their arousal level can be high because they observe several sensations and emotions, and their stress level can be low because they observe them as they are.

References

[1] Y. Ooishi, M. Fujino, V.Inoue, M. Nomura, N.Kitagawa, “Differential effects of focused attention and open monitoring meditation on autonomic cardiac modulation and cortisol secretion,” *Front. Physiol.*, Vol. 12, 675899, 2021.

Contact

Yuuki Ooishi / Sensory Resonance Research Group, Human and Information Science Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp

25

Well-being in terms of mental states, values, and idea of self

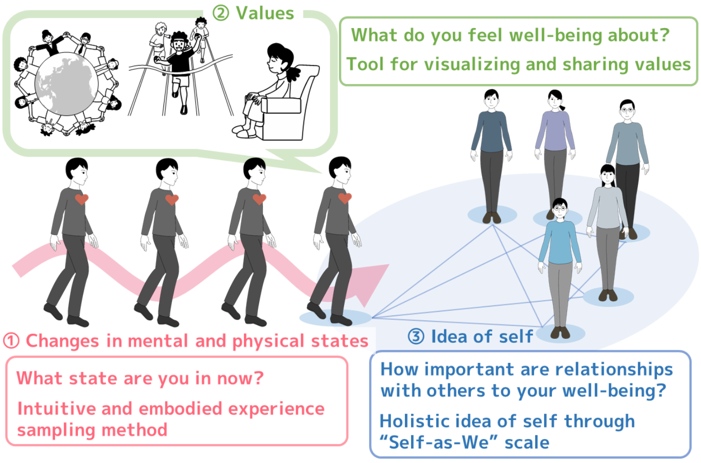
Measuring well-being through diverse aspects

Abstract

When do we feel well-being (a state of physical, mental, and social flourishing)? To find out, it is necessary to comprehensively understand ① our mental and physical state, ② what is important to us, and ③ how we relate to others. In this study, we devised original methods to measure all three. ① We devised a new experience sampling method that uses embodied expressions to intuitively record daily changing states while reflecting physical sensations. ② We devised a tool to visualize the value of each individual's diverse well-being. ③ We developed the "Self-as-We" scale to assess the degree of holistic idea-of-self based on East Asian philosophical traditions. In order for people feel well-being in their daily lives, they need to be aware of and evaluate their own physical and mental states and their values and idea-of-self, and to collaborate with others. We believe our research supports this process from the perspectives of psychology, philosophy, engineering, and design.

Overview

In order to cultivate the well-being of each individual in a society where diverse people interact and support each other, the first step is to comprehensively understand ①②③.



② Values

What do you feel well-being about?

Tool for visualizing and sharing values

① Changes in mental and physical states

What state are you in now?

Intuitive and embodied experience sampling method

③ Idea of self

How important are relationships with others to your well-being?

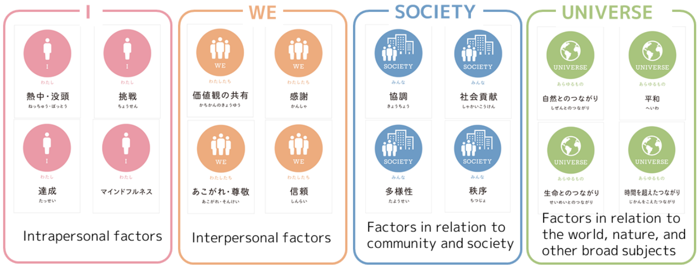
Holistic idea of self through "Self-as-We" scale

② Visualization of values

We have devised a tool to easily visualize what you and others value and promote awareness.

Choose three cards that represent "what are important to you" and share them with others.

➡ Promote awareness of one's own values and diversity of values.



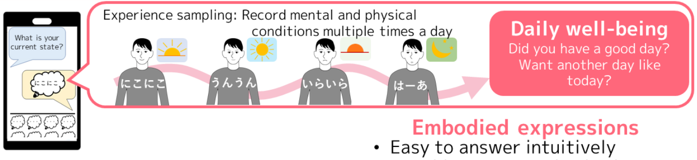
Excerpts from some of the card sets

- We made cards representing diverse factors of well-being

- Approx. 1,300 people were asked to list three "things that are important to your well-being," and categorized approx. 3,900 factors of well-being.

① Measurement of changes in mental and physical states

We have devised a new experience sampling method that can intuitively record the mental and physical states by using embodied expression.


What is your current state?
にこにこ
Experience sampling: Record mental and physical conditions multiple times a day
にこにこ
うんうん
いらいら
はーあ
Daily well-being
Did you have a good day? Want another day like today?

Embodied expressions

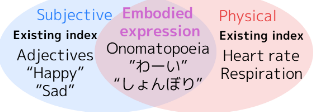
- Easy to answer intuitively
- Capable of expressing both subjective and physical sensations

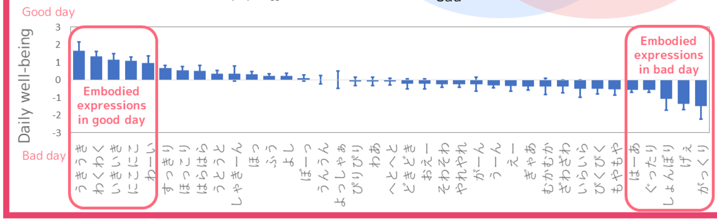
- We developed a list of embodied expressions (emotional onomatopoeia) based on the results of a large-scale survey.[*1]

[*1] approx. 14,000 respondents

- We conducted a survey on experience sampling[*2] using the list of embodied expressions.

[*2] 16 people logged for 4 weeks.



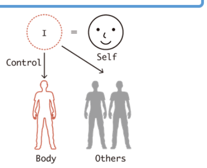


③ Evaluation of idea of self

We have developed a scale to measure how we perceive ourselves in relation to others (i.e., idea of self).

Conventional idea of Self

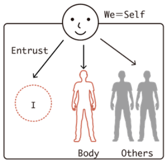
"I" = Self



- "I" am an autonomous being and can live without any relationship with others.
- "I" am in control of my body and my interactions with others by my free will.

Self-as-We

"We" = Self



- "I" alone cannot accomplish any act, and "I" can only live among "We."
- "I", my body, and others are equally entities entrusted by "We."

We developed the Self-as-We scale using psychological methods to assess the degree of "Self-as-We[*3]", a holistic idea of self based on East Asian philosophy.

3* This concept was proposed by Professor Yasuo Deguchi of Kyoto University.

Example of Self-as-We scale items

✓ "Any results that are achieved by the team belong to the team and cannot be attributed to a specific member."

✓ "When I participate in the team's activities, I feel that I am able to take initiative for my actions proactively in addition to passively following the team's requests."

Contact

Aiko Murata / Sensory Resonance Research Group, Human and Information Science Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp

26

Vision-based speedometer regulates human walking

Faster walking by moving the wall forward

Abstract

The visual scene on the eyes expands outward during walking. Such visual information is not only used to detect obstacles on the pathway, but is actually used to control walking in real time. Here we show that our automatic regulator of walking speed based on vision, which estimates and maintains the speed, is robust to changes in the depths. The robustness was not explained by temporal-frequency-based speed coding previously suggested to underlie depth-invariant object-motion perception. On the other hand, it broke down, not only when interocular distance was virtually manipulated, but also when monocular depth cues were deceptive. These observations suggest that our visuomotor system embeds a speedometer that calculates self-motion speed from vision by integrating monocular/binocular depth and motion cues. Elucidating these implicit visuomotor control mechanisms will help us for refining the technology and safety design of virtual reality devices.

How is walking speed controlled?

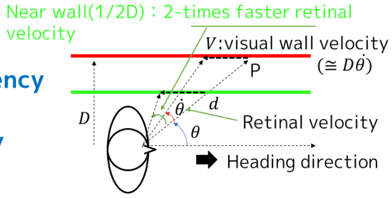
- Human can walk in a constant speed by moving legs.
- To do so, the brain uses sensory information monitoring muscle and limb states and head motion. Visual information is also indispensable to avoid obstacles on the load.
- Additional visual function is known to be used for walking speed regulation.

What information is coded in the brain for walking speed regulation?

Hypo1: **Retinal velocity**

Hypo2: **Temporal frequency of retinal image**

Hypo3: **Walking velocity**

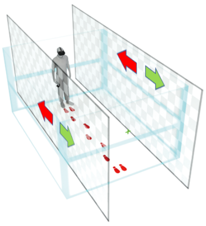


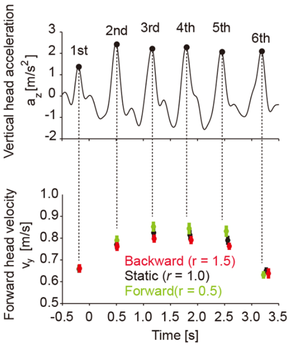
Wall motion impacts on walking speed

The head-mounted display (HMD) shows a passageway with virtual walls, and a person is instructed to walk through it.

Walking speed increases when the wall is moved forward during walking, and vice versa (automatic gait speed regulation).

⇒ Index: Walking-velocity-change



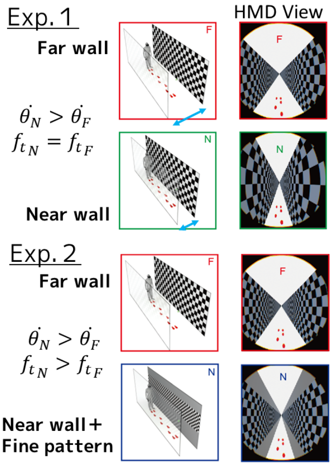


Vision based walking speed regulation

Q1: Does the retinal velocity or visual temporal frequency regulate walking speed?

Three hypotheses were examined by walking experiments with different wall-distances and roughness of wall patterns.

Experimental data showed that walking-velocity-changes were not different under two conditions of Exp1 nor of Exp2, supporting Hypo3.

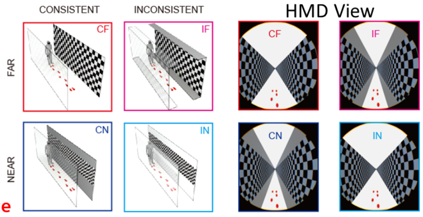


Q2: Does the monocular depth cue contribute to the wall distance estimation needed to calculate walking speed?

Walk under some conditions where the wall distance tends to be misestimated by binocular and monocular depth cues.

Far walls with colored eaves and borders (IF) look similar to the near walls of CN.

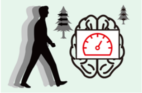
Near walls with narrow wall-pattern (IN) look similar to the far walls of CF.



Misestimation of wall distance

⇒Experimental data showed that misestimation of wall distance results in alteration of walking-velocity-changes.

The brain automatically estimates walking speed for gait control using visual information.



References

[1] S. Takamuku, H. Gomi, "Vision-based speedometer regulates human walking," *iScience 24:103390*, 2021. https://doi.org/10.1016/j.isci.2021.103390.

Contact

Hiroaki Gomi / Sensory and Motor Research Group, Human and Information Science Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp

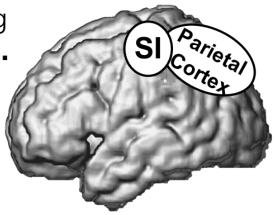
Fingertip illusions direct the mind

Abstract

Feeling directional tactile pulls is important for everyday life, allowing us to feel the weight of an object or be guided by our partner during a dance. We wanted to know what type of brain activity gives rise to the pulling sensation, specifically if it was generated in the primary somatosensory cortex (SI; area responsible for early processing of touch) or parietal cortex (area responsible for spatial and orientation processing). We generated pulling sensations via asymmetric vibration from a hand held device and recorded brain activity with electroencephalography (EEG; a technique for recording the brain's electrical activity from the scalp). We found that the pulling sensation is associated with brain activity 280ms post-stimulus in the parietal lobe. These results may benefit people with sensory impairments (e.g. blindness) or paralysis by helping researchers use vibration feedback for navigation and the control of prosthetic limbs.

How do we feel a pulling sensation?

You perceive a lot via feeling of **directional force on skin.**



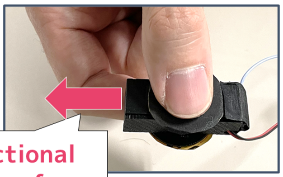
Which part of the brain generates this sensation?

Hypothesis

Parietal cortex generates pulling sensations.

Brain activity relating to pull

We used "Illusory pulling sensation from asymmetric vibration".



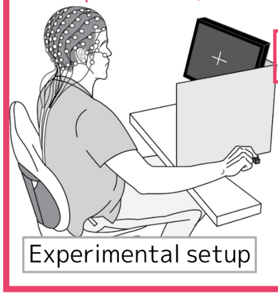
Three types of stimuli:
Left, Right & Neutral pulls

directional pulling force

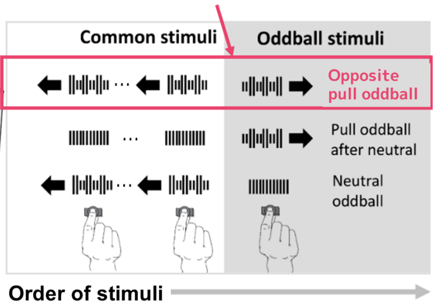
EEG was measured in oddball task*1 with pulls.

*1: response to rare stimulus in stream of commons

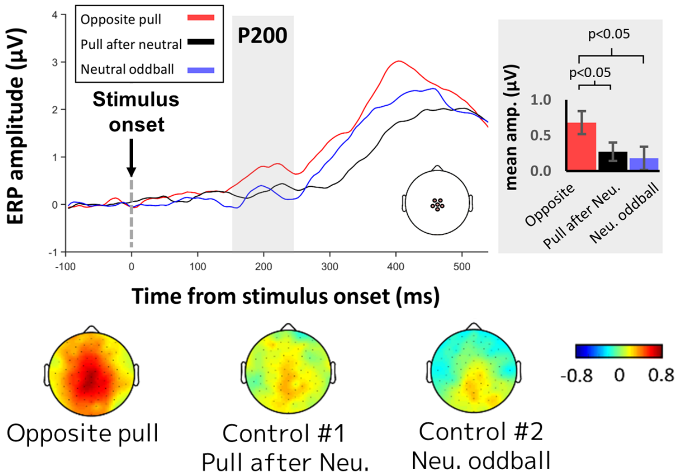
Critical condition = **'Opposite pull'** (pull direction opposite to expectation)



Experimental setup

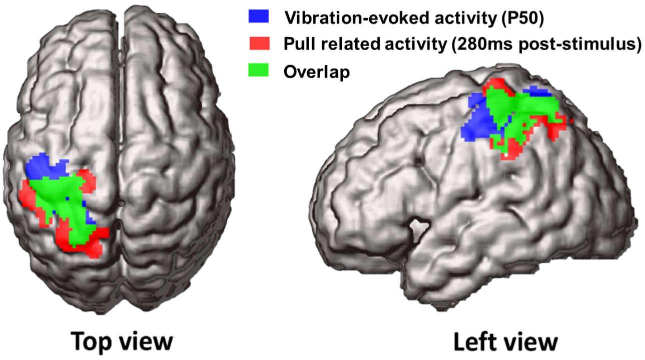


Results: parietal cortex generates pull



EEG **P200** (for orientation and spatial processing) was larger for Opposite pulls than control.

→ **Involvement of P200 in the pulling sensation.**


Vibration-evoked activity (P50)
Pull related activity (280ms post-stimulus)
Overlap
Top view
Left view

Most pulling activity (red) was **in parietal cortex**, posterior to SI (blue).

References

[1] J. De Havas, S. Ito, S. Bestmann, H. Gomi, "Neural dynamics of illusory tactile pulling sensations," *bioRxiv*, 2021. https://doi.org/10.1101/2021.10.12.464029

Contact

Jack De Havas / Sensory and Motor Research Group, Human and Information Science Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp

28

Understanding of desire to touch using large-scale Twitter data

What do we want to touch?

Abstract

What do people want to touch in their daily lives? We clarified the desire for touch in their daily lives by collecting and analyzing a huge amount of text data that people tweeted "want to touch" on Twitter. We revealed the relationship between the body part that the people want to touch and the way they want to touch it in their daily lives. Also, we revealed the effects of the COVID-19 pandemic on touching desires. Specifically, we observed the "skin hunger", or touch desire for animate' warm skin, and variation of touch avoidance toward inanimate targets such as doorknobs. It is expected that our findings can contribute to problems in broad areas such as elucidating the mechanism of touch desire in their daily lives, designing products that consumers really want to touch, and monitoring the impact of actual social problems such as the spread of infection on people's awareness.

Understanding desire to touch

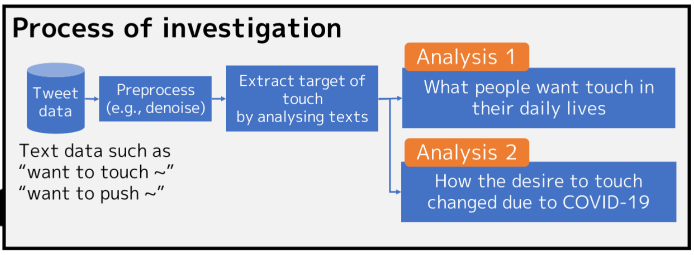
Question What do people want to touch in their daily lives?

Previous study

- focused on experiment-specific object
- did not address the desire to touch in daily lives

This study

- analyzed large-scale Twitter text data representing "want to touch“
- understood the desire to touch in daily lives



Process of investigation

Tweet data → Preprocess (e.g., denoise) → Extract target of touch by analysing texts

Text data such as
“want to touch ~”
“want to push ~”

Analysis 1
What people want touch in their daily lives

Analysis 2
How the desire to touch changed due to COVID-19

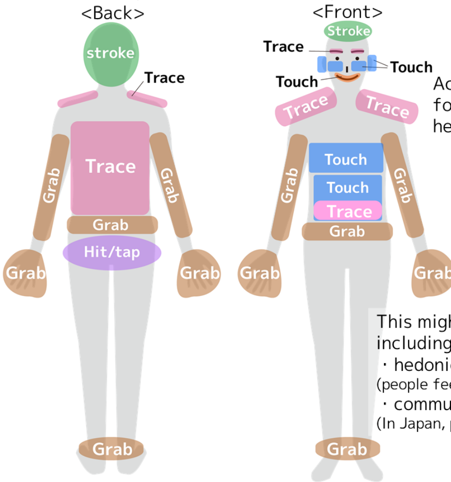
Analysis 1: Daily lives × Desire to touch

We clarified the relationship between targets of touch desire and touch gesture

Touch gesture	1st	2nd	3rd	4th
Touch	Breast	Hair	buttock	Cat
Statically contact	You	People	Skin	Cat
Stroke	Head	Cat	Dog	Abdomen
Grab	Waist	Hand	Buttock	Tail
Push	Button	Stamp	Cart	Abdomen
Hit/tap	Drum	Buttock	Keyboard	Head
Trace	Line	Abdominal muscle	Eyebrow	Muscle

Visualization of body parts and touch gesture

Mapping body part and touch gesture

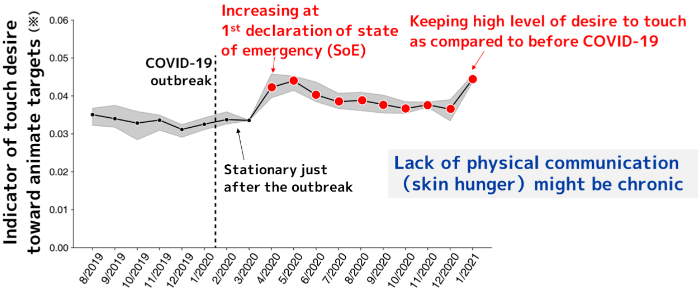


According to this visualization, for example, we see that head is the target of stroking

This might be due to multiple factors including,
・hedonic aspect
(people feel pleasantness by stroking hair)
・communicative and cultural aspects
(In Japan, people stroke the child’s head)

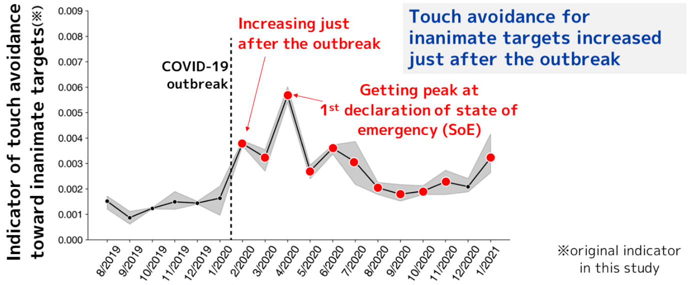
Analysis 2: COVID-19 × Desire to touch

After the outbreak of COVID-19, we investigated how the desire to touch toward animate targets (humans or animals) changed



Lack of physical communication（skin hunger）might be chronic

We also investigated how the touch avoidance toward inanimate targets (e.g., doorknobs) changed



Touch avoidance for inanimate targets increased just after the outbreak

※original indicator in this study

Timing characteristics

Skin hunger did not appear just after the outbreak and appeared at the declaration of SoE. In contrast, touch avoidance for inanimate targets appeared just after the outbreak and increased at the declaration of SoE.

References

[1] Y. Ujitoko, Y. Ban, T. Yokosaka, “Getting insights from Twitter: What people want to touch in daily life,” *IEEE Transactions on Haptics*, Vol. 15, No. 1, pp. 142–153, 2022.

[2] Y. Ujitoko, T. Yokosaka, Y. Ban, H. Ho, “Tracking changes in touch desire and touch avoidance before and after the COVID-19 outbreak,” *PsyArXiv*, 2021.

Contact

Yusuke Ujitoko / Sensory Representation Research Group, Human and Information Science Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp

29 Pupil size tracks subjective perceptual changes

Eyes as a window of our mind

Abstract

Pupil size is indexed to changes in neural activities, which have been shown to reflect a broad range of cognitive processes. We investigated the temporal aspects of pupil size on perceptual bistability. Pupil size increased with an increasing number of perceptual alternations. Furthermore, pupil size was related to the frequency of perceptual alternation at least 35 s before the behavioral report of perceptual alternations. The overall results suggest that variability of pupil size reflects the stochastic dynamics of arousal fluctuation in the brain related to bistable perception. In future work, we plan to use pupil size to predict the representation of brain network shift across modality and task.

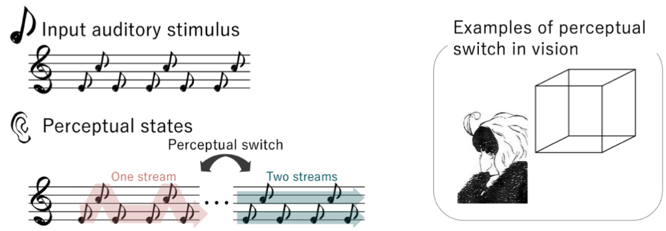
Pupil size tracks subjective perceptual changes

When you listen a certain auditory sound, the perception is spontaneously and temporally changed in multiple ways.

The pupil size may reflect a timing of the switch before we experience alternating percepts.

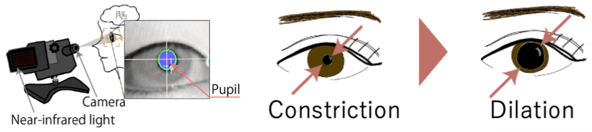
Bistable perception and perceptual switch

The moment-to-moment changes in our perception on a constant sensory input.

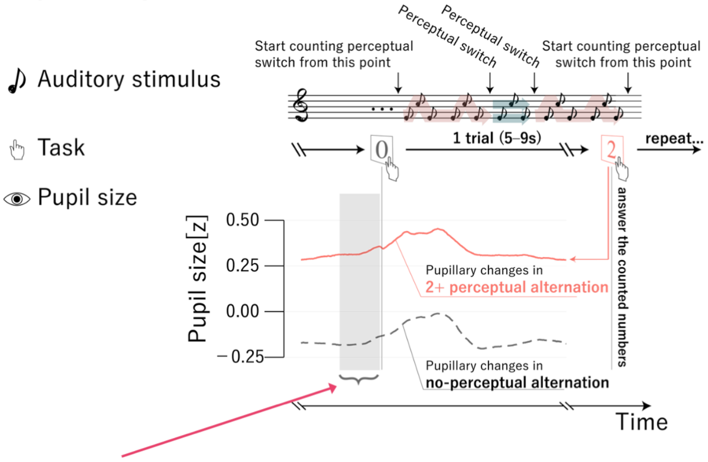


Pupillometry :

Pupil size is related to the autonomic system(norepinephrine) and indexed as an arousal level.

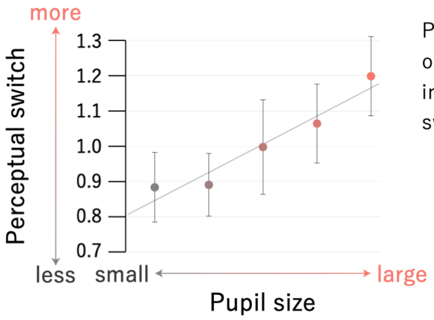


Pupillometry and task


Auditory stimulus
Task
Pupil size
Start counting perceptual switch from this point
Perceptual switch
Perceptual switch
Start counting perceptual switch from this point
…
0
1 trial (5–9s)
2
repeat...
answer the counted numbers
Pupil size[z]
0.50
0.25
0.00
−0.25
Pupillary changes in
2+ perceptual alternation
Pupillary changes in
no-perceptual alternation
Time

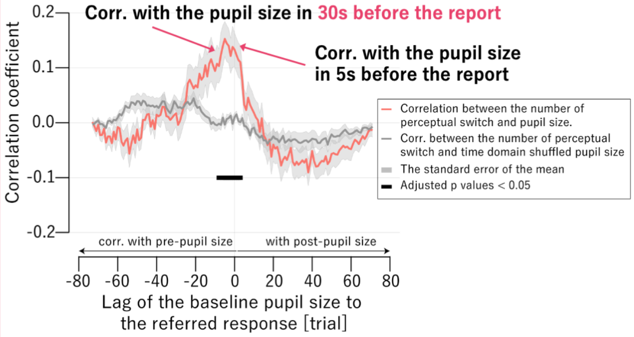
The analysis of the pupil size before the counting task corresponding to the answer.

Norepinephrine level might be related to the stochastic frequency of bistable perception



Pupil size before the assignment of the task increased with increasing number of perceptual switch.

The correlation lasts tens of seconds



References

[1] Y. Suzuki, H. Liao, S. Furukawa, "Temporal dynamics of auditory bistable perception correlated with fluctuation of baseline pupil size," *Psychophysiology*, 2022. doi:10.1111/psyp.14028

Contact

Yuta Suzuki / Sensory Representation Research Group, Human and Information Science Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp