NTT

NTT Communication Science Laboratories

NTT CS Labs. 30th 30th ANNIVERSARY

OPEN HOUSE 2021

6/3 thu・4 fri

Jun 3rd 12:00– release

On the Web

Access from: https://www.kecl.ntt.co.jp/openhouse/2021/index_en.html

Science of Machine Learning

01. **Learning how to learn from various datasets** Meta-learning from tasks with heterogeneous attribute spaces
02. **Ask me how to make a fair decision for everyone** Learning individually fair classifier based on causality
03. **Ask me anything about network structure** Indexing graph structure with decision diagrams
04. **No labels? Count on me!** Self-supervised Adaptation for Unknown Domains/Classes
05. **Finding groups through connection relationships** Relational data model with infinite flexibility
06. **Finding features in data fast and accurately** Acceleration of feature selection with group regularization
07. **Privacy-aware machine learning** Distributed learning algorithm and medical application
08. **Estimating Risk of Infection in a City** People Flow Reconstruction Based on Anonymous Sensor Data

Science of Communication and Computation

09. **Being greedy makes quantum computers work well** Economic rationality makes cloud quantum computing reliable
10. **Revealing hidden structures behind sentences** Neural rhetorical structure parsing with pseudo-labeled data
11. **Two experts, one result** Fusing two experts for enhancing their specialties
12. **Can a chatbot mediate trust between humans?** Bridging doctor-patient rapport through a chatbot
13. **Recipes for enjoy-talking conversational systems** Development of Transformer-based conversational systems

Science of Media Information

14. **Detecting faint sound by light** Non-contact sound measurement by precision interferometry
15. **Extracting voices out of noise & reverberation** Joint Signal Separation, Dereverberation and Noise Reduction
16. **AI that acquires knowledge just by watching TV** Crossmodal learning for concept acquisition of human movements
17. **Real-time speech emotion contololler using face** Emotional voice conversion via facial expression recognition
18. **Visualizing touched places for Corona prevention** Touched places detection using heat traces by thermography
19. **Telestethoscope: Looking into Body by Listening** Biomedical sound analysis utilizing physical characteristics

Science of Human

20. **Is falling birthrate related to population density?** Prediction and demonstration via life history theory
21. **Moves of magnetic fields, Moves by magnetic fields** Magnetact: A Magnetic force-based Tactile technology
22. **Make hard objects soft, make rough objects smooth** Simple method for modulating tactile texture
23. **Sense of touch connects our hearts beyond distance** Empathetic telecommnunication by vibrotactile transmission
24. **Why do mothers approach to infants' "crying"?** Oxytocin as a neural regulator of maternal implicit approach
25. **Auditory attention that appears in the eye** Relation between microsaccades and auditory spatial attention
26. **Blink pattern of elite car race drivers** Professional drivers always blink at the same time over laps
27. **Essence of "keeping your eye on the ball"** Eye-hand coordination in motor learning
28. **How do you define your dominant hand?** Quantifying motor-skill performance using a smartphone
29. **Brain functions to recognize and hit a fastball** Brain mechanisms for quick judgment and motor control

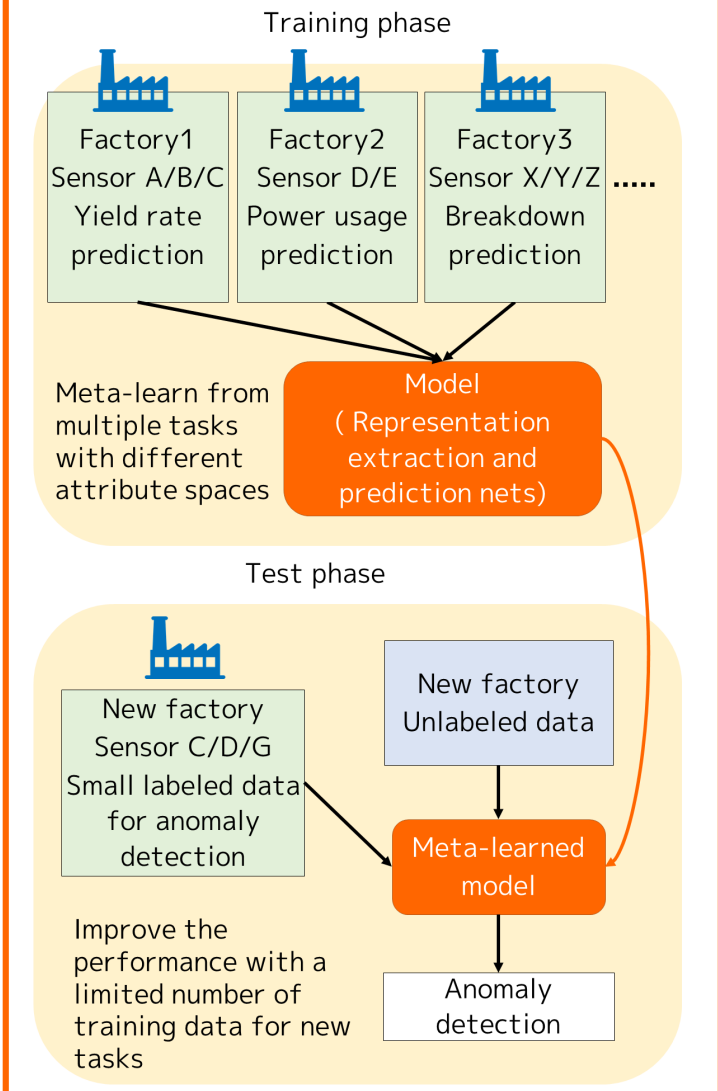
01 Meta-learning from tasks with heterogeneous attribute spaces

Learning how to learn from various datasets

Abstract

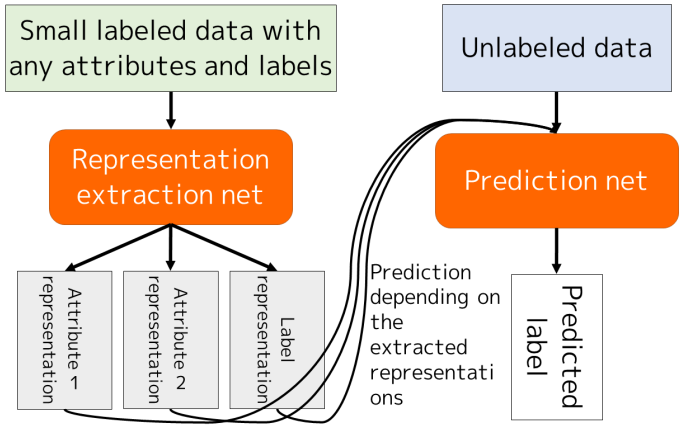
Many training data are required for deep learning. We propose a meta-learning method that can improve the performance with a limited number of training data in the target task by using data in different tasks. For example, the proposed method can be used for learning from data in different factories. Existing meta-learning methods assumes that all tasks have the same attribute space. The proposed method is the first method that can meta-learn from tasks with different attribute spaces. With the proposed method, we design permutation invariant neural networks that can handle data with different attributes and labels. Currently, we cannot apply machine learning methods when sufficient number of training data are unavailable. We want to expand field where machine learning methods are applicable by extending the proposed method.

Meta-learning form different factories



Proposed model

We design two neural networks that can handle different attributes and labels.



Evaluations

- Data：59 Datasets for regression in OpenML (Machine learning data sharing service) with 10-300 instances and 2-30 attributes

Test mean squared errors

Ours	Meta-learning1	Meta-learning2	Linear regression	Nonlinear regression
0.788	0.854	0.845	1.179	0.828

Meta-learning1: Gradient-based adaptation [ICML2017]
Meta-learning2: Neural network-based adaptation [ICML2018]
Linear regression: Ridge regression trained by test data
Nonlinear regression: Kernel ridge regression trained by test data

References

[1] T. Iwata, A. Kumagai, "Meta-learning from Tasks with Heterogeneous Attribute Spaces," in *Proc. Neural Information Processing Systems (NeurIPS)*, 2020.

Contact

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

02

Learning individually fair classifier based on causality

Ask me how to make a fair decision for everyone

Abstract

Machine learning predictions are increasingly used to make critical decisions that severely affect people's lives, including loan approvals, hiring, and recidivism prediction. For this purpose, we developed a novel machine learning technology that makes predictions that are accurate and fair with respect to sensitive features such as gender, race, religion, and sexual orientation. To achieve high prediction accuracy, we utilize prior information about societal demands for each decision-making scenario, e.g., "rejecting applicants based on physical strength is fair if the job requires physical strength." Although existing methods cannot ensure fairness when the data are not generated by a restricted class of functions, our proposed method can use various data to guarantee fairness. Thus, admitting that "what is fair" depends on a particular sense of societal values, we create innovative machine learning technologies that can more flexibly respond to societal demands by bridging the gap between technical limitations and societal needs. In this way, we hope to mold a society that can make automatic decisions while ensuring that nobody will suffer detrimental treatment.

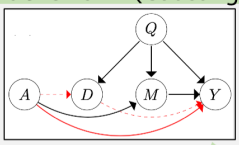
Problem: How can we build an AI that makes fair predictions (decisions) for individuals?

Input — Training data

A Gender	Q Qualification	D Department	M Physical strength	Y Decision
Female	A	Economics	B	Accept
Male	B	Literature	B	Accept
Male	C	Science	C	Reject

Sensitive feature

Prior information about "**What is fair**" (Causal graph)



Arrows provided by experts or estimated from data

Learn

Output: **Classifier (AI) that makes fair predictions**

Examples
- Hiring
- Loan approval
- Child abuse screening
- Recidivism prediction

1. Achieve high prediction accuracy

Using prior knowledge about societal demands for fairness **to learn an AI without imposing unnecessary constraints** and **to achieve high prediction accuracy**

E.g., hiring decisions for physically demanding jobs

1. Making decision Y based on gender A is **unfair** (A→Y)
2. Making decision Y based on department D is also **unfair** (A→D→Y)
3. Making decision Y based on physical strength M (which is necessary) is **fair** (A→M→Y)

Although **3. yields gender difference in rejection rates, it is unnecessary to impose a constraint on it**, which only decreases prediction accuracy.

2. Guarantee individual-level fairness using various data

Building an AI that makes **individually fair** predictions **regardless of what functional model generates data**

Reduce unfair differences in decision outcomes to zero for each individual

Original data (C.V.)

A Gender	Q Qualification	D Department	M Physical strength	Y Decision
Female	A	Economics	B	Accept

Reduce differences in acceptance probabilities to zero

Hypothetical data (Modified C.V.)

A Gender	Q Qualification	D Department	M Physical strength	Y Decision
Male	A	**Science**	**B**	Accept

Change gender to male — **Likely department for the male applicant** — **Leave unchanged since physical strength is irrelevant to discrimination**

Difficulty

To modify C.V., we must express true data-generating processes, which are impossible to approximate if data are not generated from simple functional models

Proposed method

- Propose unfairness measure that can be computed regardless of data-generating processes
- We can make an unfair difference in decision outcomes zero by forcing this unfairness measure to be zero

References

[1] Y. Chikahara, S. Sakaue, A. Fujino, H. Kashima, "Learning Individually Fair Classifier with Path-Specific Causal-Effect Constraint," in *Proc. the 24-th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2021.

Contact

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

03

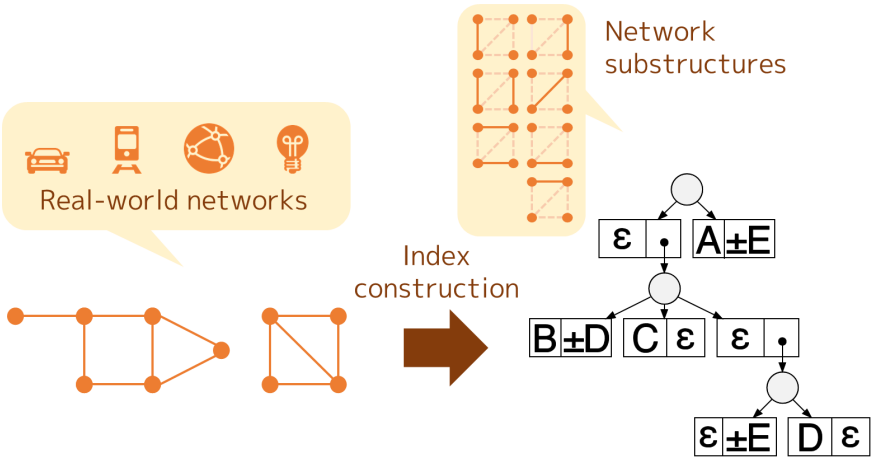
Indexing graph structure with decision diagrams

Ask me anything about network structure

Abstract

We developed a new substructure index called the zero-suppressed sentential decision diagram (ZSDD) and efficient algorithms with which to construct it. A substructure index is a compact representation of network substructures. With it, we can count and find exponentially many substructures in time linear to the index size. Since ZSDD is generally smaller than similar data structures, we can solve network-related problems much faster. For example, we can count more than 100 million network routes within a second using ZSDD. Our algorithm can solve a wide range of problems on real-world networks like communication networks, traffic networks, and power grids. Therefore, it will make society more efficient by finding better solutions to network-related problems.

Substructure index for solving network problems



We can efficiently find and count network substructures with the substructure index.

Applications

- Find best travel route
- Show multiple delivery routes
- Evaluate reliability of communication networks

This is a result of a joint research with Kyoto University.

Key technologies

1. Zero-suppressed sentential decision diagram (ZSDD)

compact index that supports many useful operations

2. ZSDD construction algorithm

Faster than baseline construction algorithm

Network	Index size		Construction time (ms)	
	ZSDD (Proposed)	ZDD (baseline)	ZSDD (Proposed)	ZDD (Baseline)
att48	279,613	387,715	3,494	3,005
berlin52	937,746	3,194,017	11,826	62,706
eil51	838,254	7,178,190	25,828	94,272
ulysses22	3,036	16,762	39	65
grafo10106	1,756	4,057	28	53
grafo10183	224,373	16,414,697	2,866	538,878
grafo10223	1,009,299	7,313,087	48,563	128,097
grafo10248	16,524	47,605	301	672

References

[1] M. Nishino, N. Yasuda, S. Minato, M. Nagata, "Zero-suppressed Sentential Decision Diagrams," in *Proc. 30th AAAI Conference on Artificial Intelligence (AAAI 16)*, 2016.
[2] M. Nishino, N. Yasuda, S. Minato, M. Nagata, "Compiling Graph Substructures into Sentential Decision Diagrams," in *Proc. 31st AAAI Conference on Artificial Intelligence (AAAI 17)*, 2017.
[3] Y. Nakahata, M. Nishino, J. Kawahara, S. Minato, "Enumerating All Subgraphs Under Given Constraints using Zero-suppressed Sentential Decision Diagrams," in *Proc. 18th Symposium on Experimental Algorithms (SEA 2020)*, 2020.

Contact

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

04 Self-supervised adaptation for unknown domains/classes

No labels? Count on me!

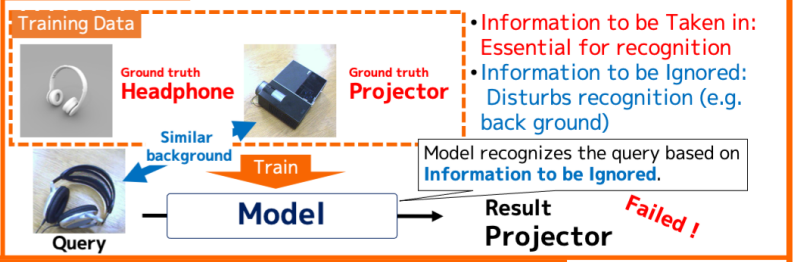
Abstract

We propose a method that enables us to use miscellaneous data collected under various conditions directly for deep learning. When we train a model using miscellaneous data, the model's performance is often degraded by "Information to be Ignored," which disturbs correct recognition. Our proposed method dramatically improves the model's performance by estimating the "Information to be Ignored" and training the model not to be affected by it. Our technique will make it possible to easily utilize miscellaneous data for learning and will contribute to expanding AI services into fields where deploying deep learning has been challenging.

Learning from miscellaneous data

Deep learning using data collected under **various conditions**

Difficulty Information to be Taken in and to be Ignore

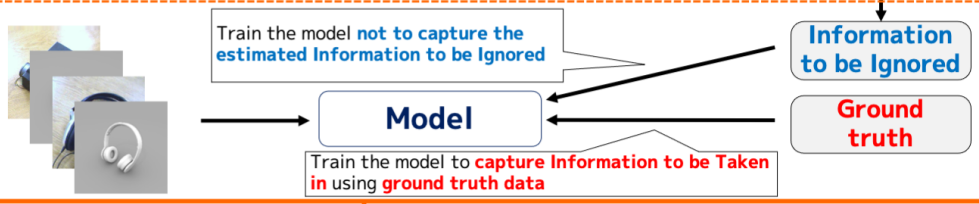


Self-supervised adaptive learning to capture Information to be Taken in

Estimate Information to be Ignored and train the model not to be affected by them.

Point 1.
Destruct Information to be Take in; then **capture Information to be Ignored** by **self-supervised learning**

Point 2.
Train the model to **capture Information to be Taken in** and **not to capture Information to be Ignored**



Performance

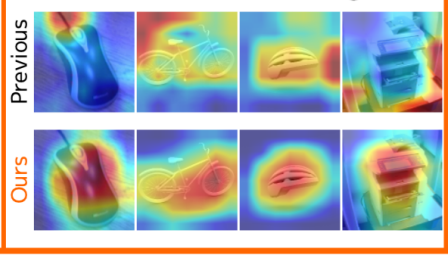
Ours achieves high recognition accuracy.

- Previous method degrades model performance due to **Information to be Ignored.**

Methods	Accuracy[%]
Previous	3.6
Ours	**77.4**

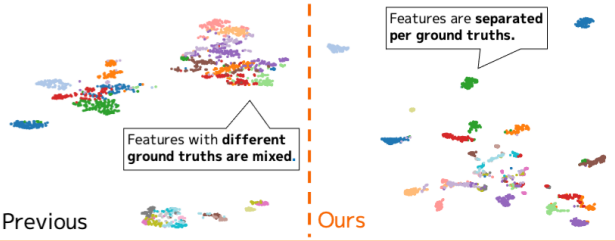
Visual explanations

Our method focuses on **targets.**



Visualization of feature embeddings

Ours forms clusters based on **ground truths.**



References

[1] Y. Mitsuzumi, G. Irie, D. Ikami, T. Shibata, "Generalized Domain Adaptation," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.
[2] R. Tobias, R. Stiefelhagen, "Adaptiope: A Modern Benchmark for Unsupervised Domain Adaptation," in *Proc. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2021.

Contact

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

05 Relational data model with infinite flexibility

Finding groups through connection relationships

Abstract

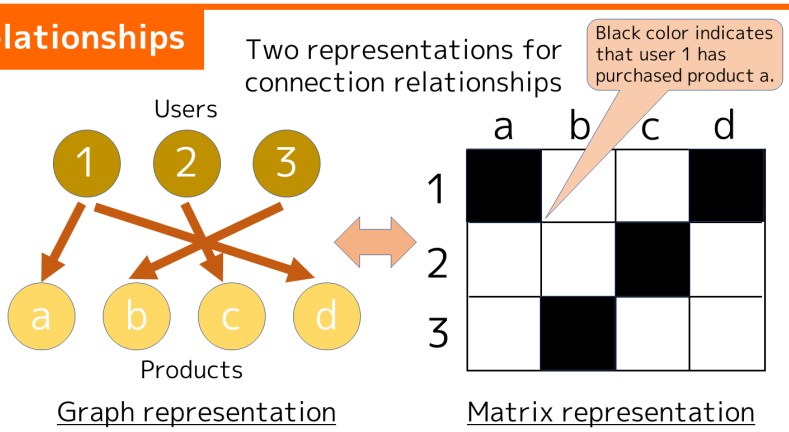
Relational data, including network graphs, such as the connection relationships of users in social networking services and the purchase histories of users for products, appear all around us. In this research, we aim to find the hidden groups in the relational data. In general, when finding groups in relational data, it is difficult to set the number and size of groups appropriately by hand. Therefore, we propose a relational data analysis method that has the ability to automatically adjust the number and size of groups in a data-driven manner according to the size and nature of the input data. By finding appropriate groups in larger relational data, we will contribute to the development of technology for more efficient information storage, search, and retrieval.

Finding groups through connection relationships

- What is connection relationships:
  - Networks that connect people to people,
  - Purchase history of users for products.
- What is groups in connection relationships:
  - Groups of people who are strongly connected each other,
  - Specific group of products that certain people prefer to buy.

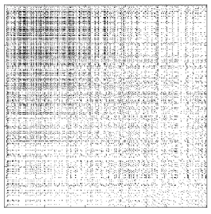
Applications :
- The discovery of potential clusters in a community to prevent the spread of infectious diseases.
- Product recommendation system to present products with high probability of purchase.

Two representations for connection relationships

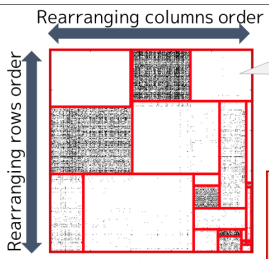


Rectangular cluster extraction from connection matrix:

The size of the input matrix is generally unknown, and it may have an **infinite number of rows and columns**.



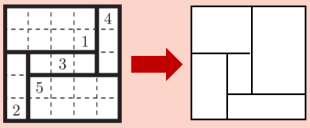
Input data



Analysis result

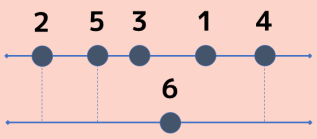
For all possible rectangular partitions **(with infinite number of cases)**, we wish to infer the high degree of fitness to the input data.

Contribution: Modeling the stochastic transformation from floorplan partitions to rectangular partitions



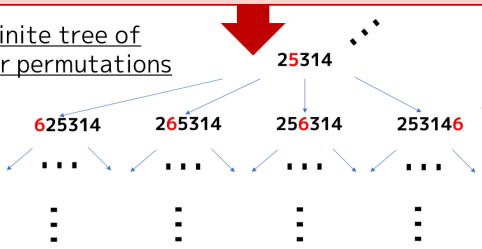
Floorplan partitions (that ignore the size of rooms) Rectangular partitions (that considers the size of rooms)

Contribution: Construction of a generative probabilistic model for the infinite tree of Baxter permutations [Baxter, 1964].



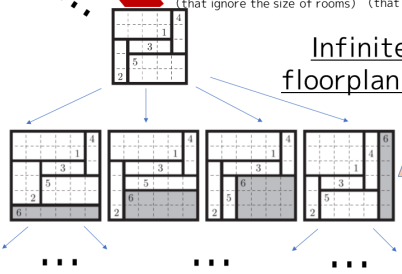
Surjection [Reading,2012]

Infinite tree of Baxter permutations



Bijection [Hong+,2000]

Infinite tree of floorplan partitions



Floorplan partitions have a one-to-one correspondence with Baxter permutations on the diagonal.

References

[1] M. Nakano, A. Kimura, T. Yamada, N. Ueda, "Baxter Permutation Process," in *Proc. Advances in Neural Information Processing Systems 33 (NeurIPS)*, 2020.

Contact

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

06

Acceleration of feature selection with group regularization

Finding features in data fast and accurately

Abstract

Selecting important feature groups from data is one of the most fundamental tasks in data analysis. However, the computation cost is high when the size of the data is large. In this study, we propose a fast method for feature group selection without degrading the accuracy. It safely skips unnecessary computations and intensively optimizes important groups. As a result, our method is up to 35 times faster than the original method without any loss of accuracy. In addition, our method has no additional hyperparameters and no additional tuning costs. It will be possible to create value from complex and large scale data by speeding up the analysis of data with complex structures such as groups. We create value from a wide variety of data and contribute to society.

Feature group selection

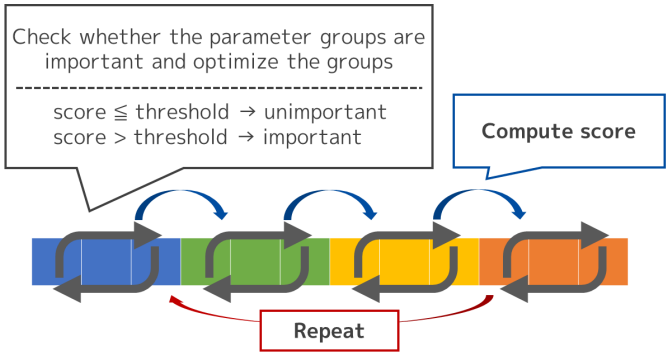
- Selecting important feature groups from data is one of the most fundamental tasks in data analysis.

e.g. finding out which areas are important for Tokyo's weather forecast from the climate data of all areas in Japan.

Which feature groups are important?

Problem: computation cost

- The cost is high when the size of the data is large.

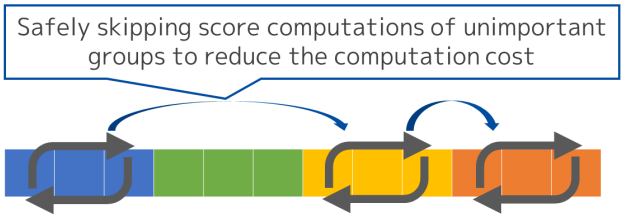


The bottlenecks are "# of score computations" and "# of iterations".

Fast method for feature group selection

- We utilize two ideas:

1. Safely skipping unnecessary computations



We efficiently identify **unimportant groups** by using the **upper bounds** of the scores.

Upper bound of score ≦ threshold → unimportant

2. Intensively optimizing important groups

Intensively optimizing important groups to reduce the value of the objective function

We efficiently identify **important groups** by using the **lower bounds** of the scores.

Lower bound of score > threshold → important

Experiments show that our method is **up to 35 times faster than the original method without any loss of accuracy.**

References

[1] Y. Ida, Y. Fujiwara, H. Kashima, "Fast Sparse Group Lasso," in *Proc. Neural Information Processing Systems (NeurIPS)*, 1700-1708, 2019.

Contact

Yasutoshi Ida / Software Innovation Center
Email: cs-openhouse-ml@hco.ntt.co.jp

07

Distributed learning algorithm and medical application

Privacy-aware machine learning

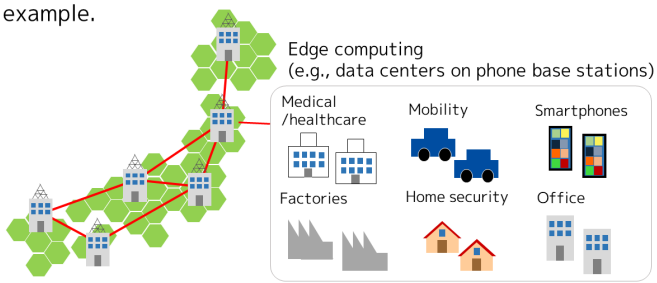
Abstract

While use of massive data benefits on training deep learning models, aggregating all data into one physical location (e.g. a central cloud) may not be possible due to data privacy concerns. For example, according to the EU GDPR, data transmission should be minimized among processing nodes. Our goal is to construct training algorithms to obtain global deep learning models that can be adapted to all data, even when individual nodes only have access to different subsets of the data. We assume that this algorithm is allowed to communicate between nodes in an asynchronous/sparse manner, exchanging such information as model variables or their update differences. However, data are prohibited from being moved from the node on which they reside. We aim to indirectly exploit the overall data across countries and provide high performance services for such industries as the medical/health-care field while protecting privacy.

Goal and application

Background: We are entering an era of distributed data processing (inference/training) due to data volume, privacy-aware issues, and legal regulations, e.g., GDPR.

Goal: To train deep learning models without aggregating data to a central cloud, where asynchronous communication among nodes are allowed to exchange latent variables, for example.



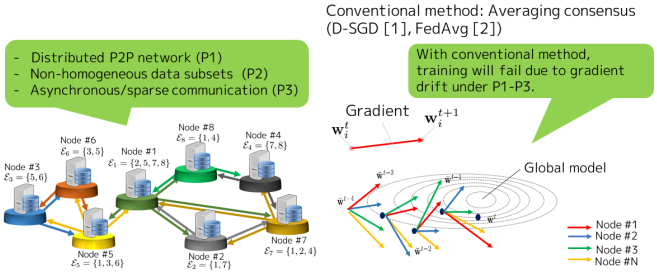
Problems

Aim: Our goal is to obtain a global model, which is equivalent to a model trained at a central cloud using all datasets, even data distributed in N(≧2) nodes.

Problem 1: Network structure is distributed in P2P manner to scale service at any scale.

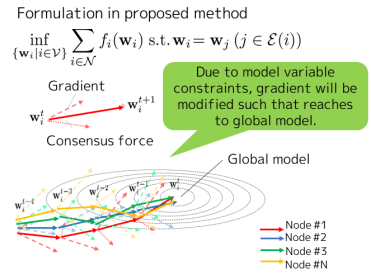
Problem 2: Non-homogeneous data subsets are placed for each node. Training procedure is unstable due to gradient drift.

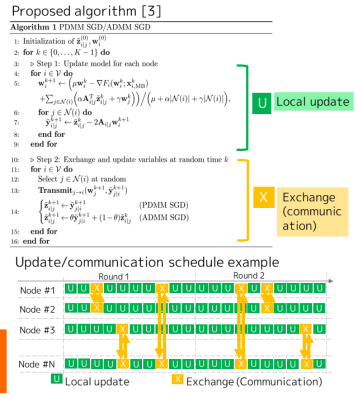
Problem 3: Communication among nodes is in asynchronous/sparse manner due to large-scale network.



Asynchronous decentralized deep learning

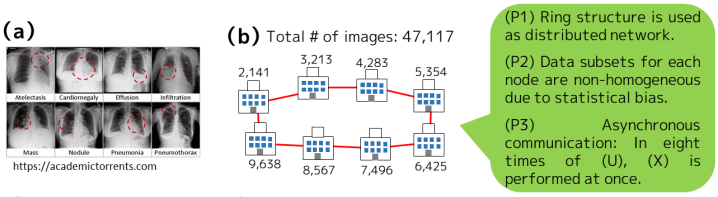
Proposed algorithm: We solved a model variable constrained minimization problem. Training is achieved by alternatingly repeating (U) local node model updates and (X) exchange auxiliary variables for progress w.r.t. making consensus between nodes. This scheme is runnable on arbitrary network structure with asynchronous communication.



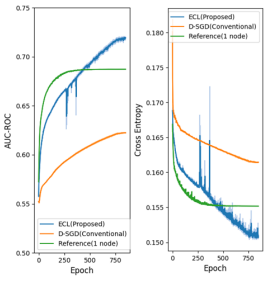


Medical image analysis

We trained 14 different disease detection models using chest X-ray datasets [4]. Image data were not transmitted from eight hospital nodes.



(Experimental results)



- Detection accuracy for some diseases was measured using Area Under curve of Receiver Operating Characteristic (AUC-ROC).
- Conventional method (orange) did not attain global model performance (green) due to gradient drift.
- Proposed method (blue) reached global model performance (green) even though P1-P3 are present.
- Detection accuracy for a part of disease was practical level: AUC-ROC of 0.75 or higher for e.g., emphysema, pneumothorax, cardiomegaly, pleural effusion.

References

[1] J. Chen, A. H. Sayed, "Diffusion adaptation strategies for distributed optimization and learning over networks," *IEEE Transactions on Signal Processing*, Vol. 60, No. 8, pp. 4289–4305, 2012.

[2] B. McMahan, E. Moore, D. Ramage, S. Hampson, B. A. y Arcas, "Communication–efficient learning of deep networks from decentralized data," in *Proc. Artificial Intelligence and Statistics (AISTATS 2017)*, pp. 1273–1282, 2017.

[3] K. Niwa, N. Harada, G. Zhang, W. B. W Kleijn, "Edge-consensus learning: deep learning on P2P networks with nonhomogeneous data," in *Proc. the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD 2020)*, pp. 668–678, 2020.

[4] National Institutes of Health (NIH) clinical center, ChestXray14 data set.

Contact

Kenta Niwa / Learning and Intelligent Systems Research Group, Innovative Communication Laboratory

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

08 People flow reconstruction based on anonymous sensor data

Estimating risk of infection in a city

Abstract

We can measure the risk of infection in a city by knowing the movement routes of infected people, but collecting information of people's movements infringes privacy. We thus propose a new method for estimating multiple routes on the basis of anonymized passage information to estimate the risk of infection while preserving privacy. For a better estimation, we need to choose more appropriate path patterns of people that correctly explain the anonymized passage information. Therefore, we consider a movement model, estimate the transit probability between passage information, and find the most likely set of routes efficiently on the basis of the model. It can improve the accuracy of the estimation. Infectious disease control will be one of functions of smart cities to be realized in the future.By using our work, the risk of infection can be estimated without collecting personal movement information.

Summary

Background

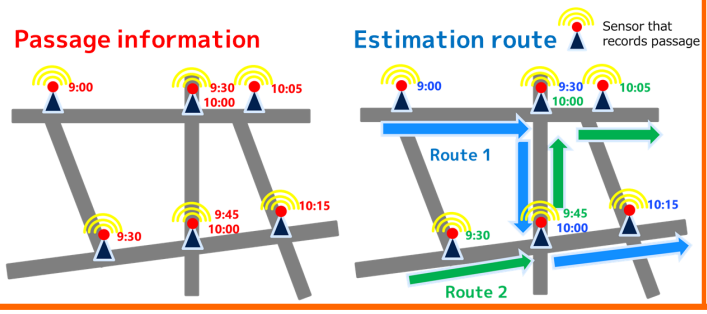
To prevent future outbreaks of infectious disease, we need to measure the risk of infection in urban areas. To do this, we want to know the movement routes of infected people. However, collecting information of personal movements infringes privacy.

Our Research

We estimate **multiple trajectories of infected people** on the basis of **anonymous passage information**.

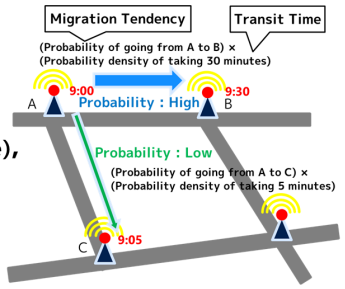
Framework

1. Deploy sensors at each intersection in a city and collect anonymous passage information (time, position, anonymous ID).
   (Anonymous IDs change regularly and people's movements can't be tracked.)
2. Determine if each passage information is from an infected person on the basis of anonymous ID, and obtain the list of infected people's passage information.
   (We don't know who correspond to each passage information.)
3. Estimate **multiple trajectories of infected people** by using the list of **anonymous passage information**.



Technical Points

Considering **movement model (migration tendency, transit time),** estimate **the transit probability** between passage information.



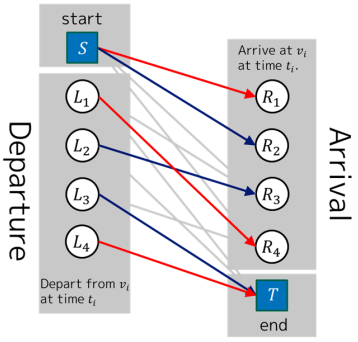
Technical Problem

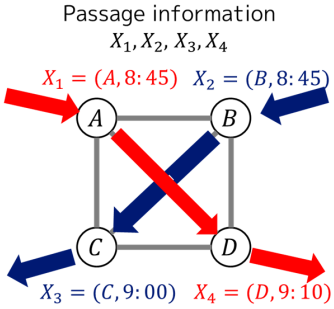
If many people are in the same place at the same time, there exists a huge number of candidates for a set of routes that explain the passage information. It is difficult to list and examine all the candidates.

Solution

The problem of finding **the most likely set of routes** can be efficiently solved by reducing it to **the minimum cost flow problem**.

By "Pass through A" = "Arrive at A"+"Depart from A ", it can be expressed as a problem of associating "Arrival" and "Departure". If we set the cost of associating "Arrival" and "Departure" as − log (transit probability), the problem of minimizing the sum of costs can be formulated as the minimum cost flow problem. The association that minimizes the sum of cost corresponds to the most likely set of routes.





References

[1] K. Matsuda, H. Ikeuchi, Y. Takahashi, T. Toyono, "People Flow Reconstruction Based on Anonymous Sensor Data toward Smart City Infrastructure for Estimating Infection Route," IEICE Technical Committee on Network Systems, 2020.

Contact

Kotaro Matsuda / NTT Network Technology laboratories
Email: cs-openhouse-ml@hco.ntt.co.jp

09

Economic rationality makes cloud quantum computing reliable

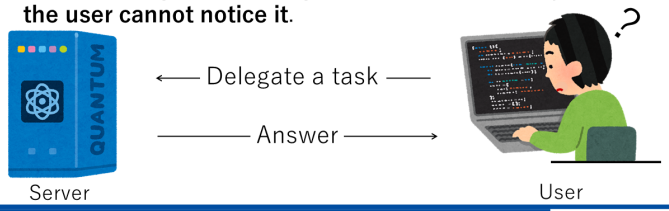
Being greedy makes quantum computers work well

Abstract

We consider cloud quantum computing and have proposed a method guaranteeing the correctness of the answer received from a cloud-quantum-computing server. Since quantum computers are strongly affected by noises, it is indispensable to realize such a method for practical cloud quantum computing. Previous methods work only under the assumption that a verified quantum computer runs in a sufficiently short time or a user can perform quantum communication with the server. By introducing the economic rationality, we have succeeded to remove these assumptions. Furthermore, our method can be applied to a broader class of quantum-computing architectures than that of previous methods. We aim to incorporate large-scale quantum computers into the existing worldwide network by developing our method, which makes it possible to deliver the high computational capability of quantum computers to everyone all over the world.

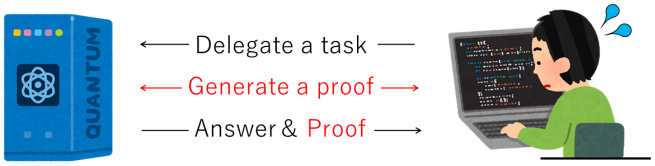
Challenges to realize cloud quantum computing

- Although quantum computers have high computational capability, they are strongly affected by noises. Since **their maintenances require the specialized knowledge**, it is pragmatic to use them **in the cloud service**.
- It is difficult for a user to verify whether the answer given by a server is correct. For example, **even if the server does not use his/her quantum computer to solve the delegated task, the user cannot notice it**.



Limits of existing methods

The server generates a proof guaranteeing that the server runs his/her quantum computer as requested, and **the user checks it**.



In order to make it impossible to forge a proof even with a quantum computer, **a strong assumption or an additional communication is required**.
Existing method 1: it is based on a post-quantum cryptography.
Existing method 2: quantum communication is necessary.
⇒ There are many challenges to be solved for the realization☹

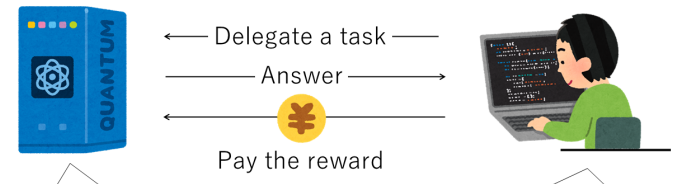
Cloud quantum computing without proofs

Our result: By using economic rationality, we have proposed a guaranteeing method for cloud quantum computing.

Points
- Instead of generating and checking proofs, **the user pays the reward to the server** depending on the received answer.
- **We have constructed a calculation algorithm for the reward** such that the reward is maximized only when the server honestly behaves as instructed by the user.
- Our method can be applied to a broader class of quantum-computing architectures than that of previous ones.

Since the reward is maximized only when the server sends the correct answer, the rational server certainly sends the correct answer!

(※ The maximum of the reward can be adjusted depending on the situation.)

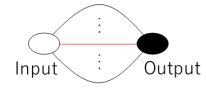


I would like to maximize the received reward.
= Economic rationality

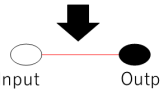
- I can receive the correct answer.
- No additional communication is required.

Overview of our mathematical proof

① We decompose the quantum computation into a huge number of easy calculations.
② The user selects and performs a single calculation uniformly at random. Then he/she uses the result to derive the value of the reward for the server's answer.
③ Since the randomness is included in step 2, there is a possibility that the value of the reward is inappropriate. However, we have shown that the expected value of the reward is maximized only when the server sends the correct answer.



Each line denotes each process in quantum computing.



Although the user cannot efficiently simulate all processes simultaneously, it is possible for a single process (red line).

References

[1] Y. Takeuchi, T. Morimae, S. Tani, "Sumcheck-Based Delegation of Quantum Computing to Rational Server," in *Proc. the 16th International Conference on Theory and Applications of Models of Computation (TAMC)*, 2020.

Contact

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

10

Neural rhetorical structure parsing with pseudo-labeled data

Revealing hidden structures behind sentences

Abstract

This poster presents a method to identify the hidden structures of documents. Each document has a rhetorical structure, which expresses the relations among clauses. Since building a rhetorical structure parser is based on supervised learning, it requires large amounts of manually annotated training data for accurate parsing. However, conventional methods suffer from a lack of training data, resulting in poor performance because manual annotation is quite labor intensive. To tackle this problem, we propose a method that uses silver data: automatically annotated pseudo-labeled data. We pre-trained the parser with silver data and fine-tuned it with gold data: manually annotated data. Our experimental results demonstrated that our method achieved the best performance. The new parser will contribute to various natural language processing applications, such as machine translation and automatic summarization.

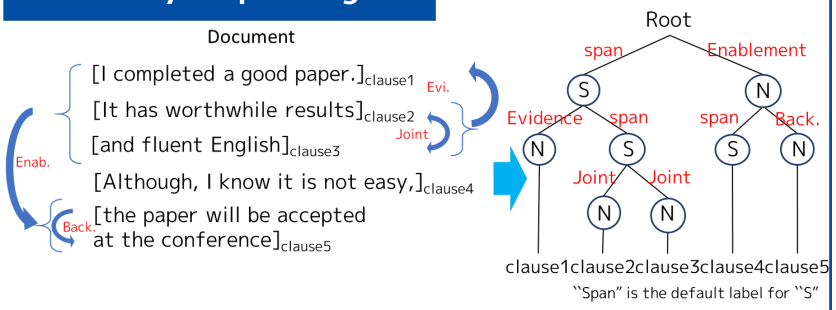
Rhetorical Structure Theory (RST)

RST represents relations among clauses in a document by a labeled binary tree

Leaves of the tree: clauses
Nodes of the tree: span consisting clauses
Labels of the nodes: Role of the span
→Nucleus (N) / Satellite (S)
※An "S" modifies "N"
Labels of the edges: relation label between two spans
→18 classes:
Elaboration, Evidence, etc

Natural language processing applications such as machine translations, automatic summarizations require the trees

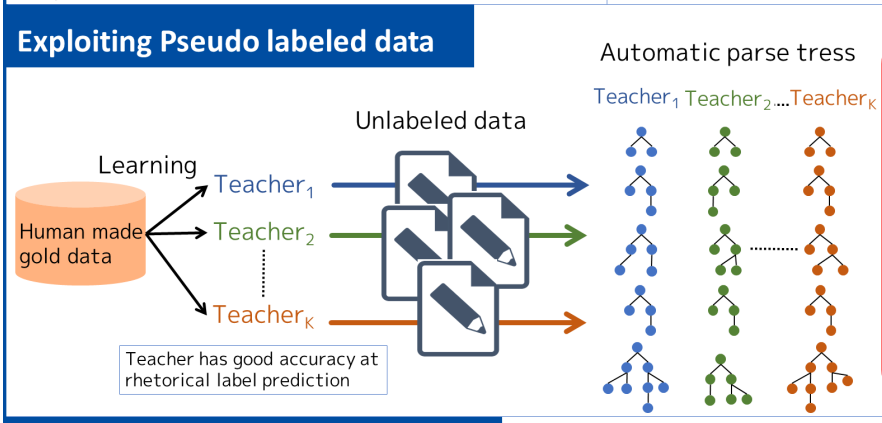
Difficulty of parsing

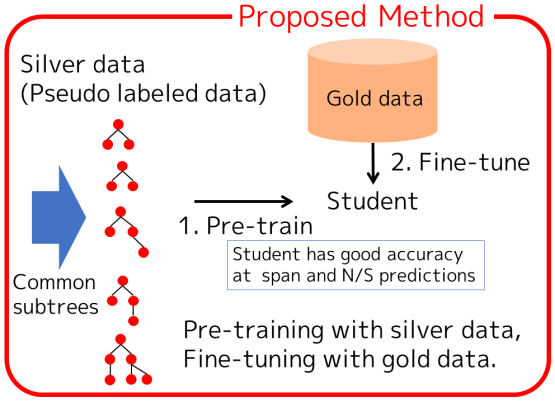


It is hard to annotate large-scale data manually since the annotation is labor-intensive

Only small dataset is available

Exploiting Pseudo labeled data





Effectiveness of our method

F-score (%)

Method＼Metric	Fully-labeled	Span	+N/S label	+Rel. label
Teacher [Wang+17]	58.8	86.0	72.4	59.7
Student [Kobayashi+20]	59.6	**87.1**	74.6	60.0
Proposed method	**62.6** UP!	**87.1**	**75.0**	**63.4** UP!

Fully-labeled score is improved by the gain of relation labeling

This is a collaborative research project between Okumura Lab. at Tokyo Institute of Technology and NTT CS labs.

References

[1] N. Koabayashi, T. Hirao, H. Kamigaito, M. Okumura, M. Nagata, "Improving Neural RST Parsing Model with Silver Agreement Subtrees," in *Proc. 2021 Annual Conference of the Noth American Chapter of the Association for Computational Linguistics*, 2021.

Contact

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

11 Fusing two experts for enhancing their specialties

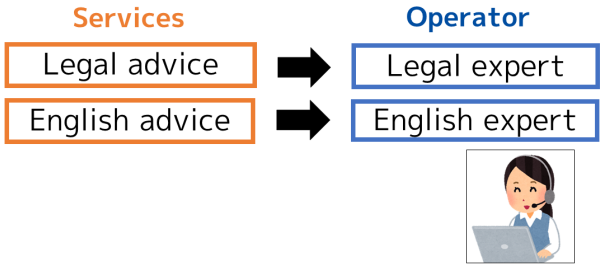
Two experts, one result

Abstract

Although dialogue services sometimes require operators with expertise in multiple fields, such people are rare (e.g., travel guides for foreigners need both travel guide skills and foreign language skills). We focus on a collaboration style in which operators with different fields of expertise work together to solve advanced problems and grow through practice. We focus on a style in which two operators interact with a user as if they were a single operator. Our proposed style enables operators to provide advanced services across multiple fields, to collaborate smoothly, and to learn from each other. By building a dialogue service that allows collaboration among diverse human resources, we will achieve more advanced dialogue services and create a society that supports the employment and growth of many people.

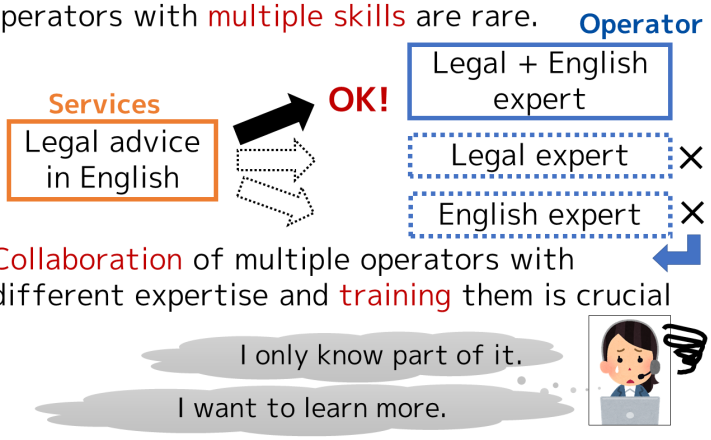
1. Dialogue services with experts

- Dialogue with experts is effective for solving problems.
  (ex. travel guides, customer service)
- We need a variety of experts to meet individual needs.



2. Lack of experts with multiple skills

- Although services require multiple skills, operators with multiple skills are rare.

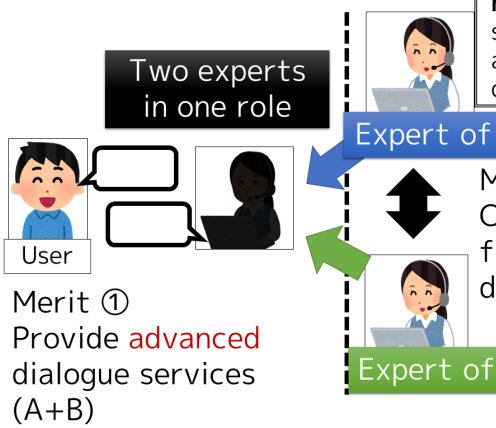


- Collaboration of multiple operators with different expertise and training them is crucial

3. Collaboration style: two experts playing one role

(1) Provide advanced dialogue services across multiple skills and (2) train operators

■Proposed style



Two-person-one-role method: Two operators share an "expert" chat account and take turns depending on topic.

Merit ①
Provide advanced dialogue services (A+B)

Merit ②
Operators learn from each other to develop their skills.

① Achieve travel guides requiring multiple skills
Osaka prefecture expert + Nara prefecture expert
= Osaka + Nara prefectures expert

User
What place do you recommend in Osaka and Nara?
"Expert" (two-person-one-role)
In Osaka, I recommend Osaka Castle Park. **(response by Osaka expert)**
In Nara, Nara Park is a great place to visit. You can see some deer. **(response by Nara expert)**

② Easy for operators to learn different skills
Operator's satisfaction of learning improved more in proposed style than chatting without sharing account (two operators chat as different people).

Acknowledgements : This work was supported by JST-Mirai Program Grant Number JPMJMI18C6, Japan.

References

[1] T. Arimoto, R. Higashinaka, K. Tanaka, T. Kawanishi, H. Sugiyama, H. Sawada, H. Ishiguro, "Collection and analysis of dialogues provided by two speakers acting as one," in *Proc. the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, pp. 323-328, 2020.

Contact

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

12 Bridging doctor-patient rapport through a chatbot

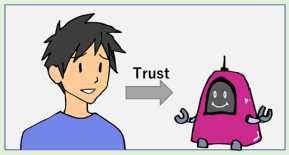
Can a chatbot mediate trust between humans?

Abstract

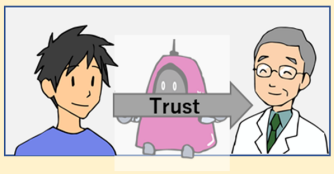
For mental health professionals (MHP) to understand patients' mental health, it is critical that patients engage in deep self-disclosure. However, people tend to avoid revealing their vulnerabilities for fear of being judged by others. Chatbots show great potential in this domain because prior research has shown that people tend to disclose symptoms of depression more truthfully when talking to a chatbot than when talking to a human interviewer. Our work extends this prior work by proposing a novel approach to facilitate people's self-disclosure to MHPs through chatbots. We designed, implemented and evaluated a chatbot that elicits deep self-disclosure and promotes trust-building between users and the MHPs. Results show that people were more willing to share their self-disclosure content with MHPs through the chatbot, which suggests the promise of our approach.

Issues of isolation and loneliness

Prevalence of communication technology → increase of **weak relationships,** Isolation, and loneliness



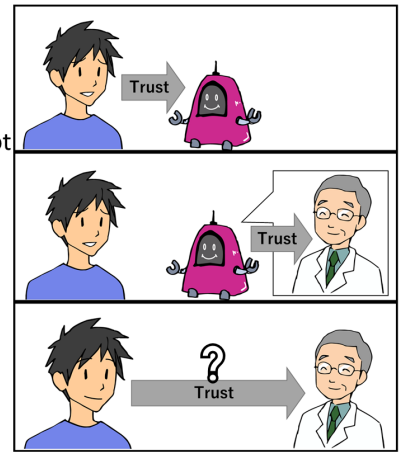
By connecting users with real experts, we aim to create an environment where it is easy to ask for help.



Trust transfer function

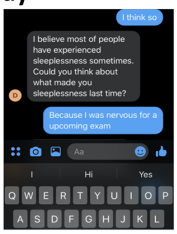
Feature 1
Develop trust with Chatbot
Reciprocal disclosure between users and chatbot

Feature 2
Transfer trust to Doctor
Chatbot talks about **good experiences with the doctor**



User Study and Evaluation

User Study



+

47 participants, 4 weeks
15 minute daily chat

One-hour Online interview

Condition ① : Only **Feature 1**

Condition ② : Both **Feature 1** and **Feature 2**

Trust in chatbot :
Condition ① ≒ Condition ②

Trust in MHP :
Condition ① < Condition ②

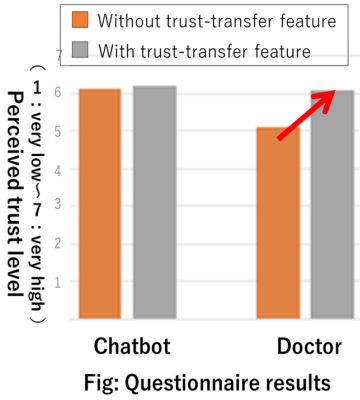
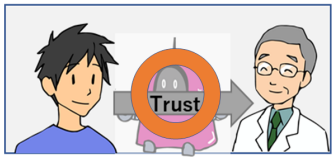


Fig: Questionnaire results

Participants in Condition ② were motivated to share contents of their disclosure with the doctor.

Showed potential for chatbot to transfer trust among humans



References

[1] Y. Lee, N. Yamashita, Y. Huang, "Designing a Chatbot as a Mediator for Promoting Deep Self-Disclosure to a Real Mental Health Professional," in *Proc. ACM Hum.-Comput. Interact. 4, CSCW1, Article 031, (CSCW'20)*, ACM, pp.1-27, 2020.

[2] Y. Lee, N. Yamashita, W. Fu, Y. Huang, ""I Hear You, I Feel You": Encouraging Deep Self-disclosure through a Chatbot," in *Proc. the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*, ACM, pp.1–12, 2020.

Contact

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

13

Development of transformer-based conversational systems

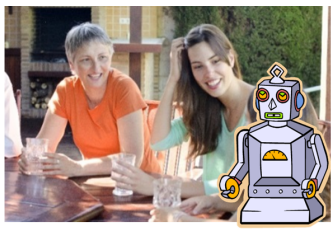
Recipes for enjoy-talking conversational systems

Abstract

We are studying a social dialogue system that satisfies people's desire for dialogue through natural conversation.
We have developed a deep-learning-based Japanese social dialogue system, which is pre-trained with the largest-scale Japanese dialogue data obtained from Twitter and fine-tuned with high-quality dialogue data that NTT has cultivated over many years of research. We also conducted a quantitative analysis of the utterances of the constructed system and identified remaining issues such as contradictions and discontinuous topics.
We believe that the desire to communicate with others is one of our fundamental desires. We aim to realize a social dialogue system as a partner that continuously satisfies this need for dialogue.

Dialogue systems chatting with people

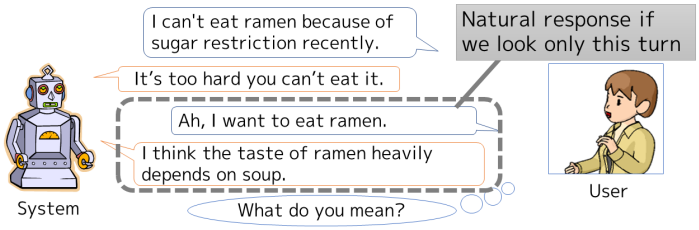
Dialogue systems are recently expected as daily conversational partners.



- **Anytime, anywhere**
  (With no limitation of time and places)
- **Relaxing, ease**
  (No need to be shy nor to worry about the other's evaluation)
- **Deep topics**
  (Easy to talk about private topics and fine-grained favorites)

Issues of conventional systems

- Even a natural utterance as a one-turn response may **become unnatural when considering the context**.
- Too many combinations of utterances complicate considering contexts without appropriately abstracting the meaning of words.

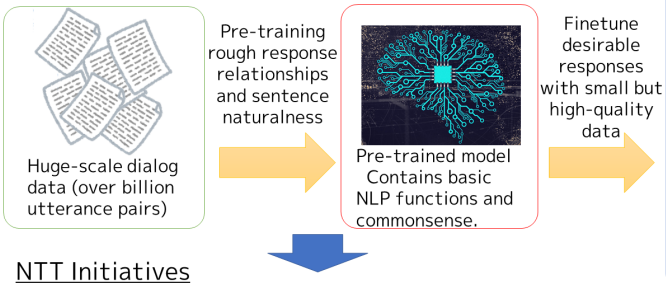


Deep-learning-based dialogue system

Deep-learning-based dialogue systems with huge-scale pre-training* are rapidly developed in English.

*Pre-training: A method to learn sentence naturalness or rough response patterns using huge data in advance. It shows a good performance in machine translation and question answering.

→It dramatically improved dialogue systems' performance by realizing highly natural sentence generation and highly accurate abstraction of the meaning of words.



Finetune with dialogue **about profile**

System	User
Hello. Thank you and best regards.	
	Nice to meet you!
I'm an esthetician. What do you do for a living?	
	I'm an office worker.
You look busy with work. What do you do on your days off?	
	I sleep on my days off.
I see. I love my pet rabbit. Do you have a pet?	
	I don't have one. The rabbit is cute!
Yes, it's soothing. Do you have any hobbies?	

Finetune with dialogue **about favotites**

System	User
Hello. Thank you and best regards, Do you have any hobbies?	
	Hello. My hobby is watching "How about Wednesday".
How about Wednesday? I love it. Which series do you like?	
	I like the confrontational archipelago.
That's a masterpiece! I like Europe revenge. Do you have a favorite episode?	
	I burst out laughing no matter how many times I watch the episode that Mr. Yasuda reverses.
Yasuda is interesting! I also like the combination with Yo Oizumi.	
	Yeah. Their talk is great.
I laughed at the laughter of Yo Oizumi.	

NTT Initiatives

- Development of the largest deep-learning-based Japanese social dialogue system
  - Model Size: 1.6 billion Parameters (3 times larger than the existing Japanese model)
  - Pre-training data: 2.1 billion pairs, 512 GB Twitter reply pairs
  - Finetune data: 200,000 pairs of high-quality dialog data (accumulated by NTT in past studies)
- Error analysis of unnatural utterances
  - Most of the utterances are natural, but about 7% of the utterances are strongly unnatural.
  - Subjective evaluation significantly decreased when contradictions, topic skipping, or utterances that differ from the facts are included.

Awarded the highest prize in the 3rd dialogue system live competition, where social dialogue systems competed.

This study was supported by the Grant-in-Aid for Scientific Research on Innovative Areas "Communicative intelligent systems towards a human-machine symbiotic society" (Issue No. 19 H05693).

References

[1] H. Sugiyama, H. Narimatsu, M. Mizukami, T. Arimoto, Y. Chiba, T. Meguro, H. Nakajima, , "Development of conversational system talking about hobby using Transformer-based encoder-decoder model," in *Proc.* Special Interest Group on Spoken Language Understanding and Dialogue Processing (*SIG-SLUD)*, Vol. B5, No. 02, pp. 104-109, 2020 (in Japanese).
[2] H. Sugiyama, T. Meguro, Y. Yoshikawa, J. Yamato, "Improving Dialogue Continuity using Inter-Robot Interaction," in *Proc.* IEEE International Symposium on Robot and Human Interactive Communication (*RO-MAN)*, pp. 105-112, 2018.

Contact

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

14

Non-contact sound measurement by precision interferometry

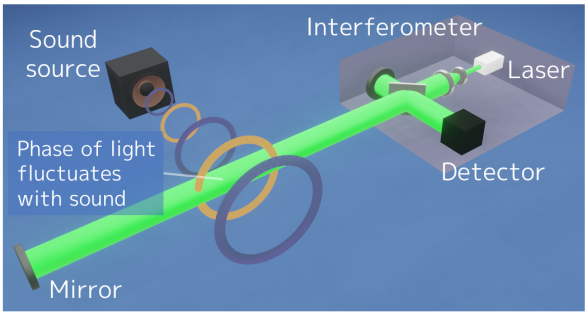
Detecting faint sound by light

Abstract

Optical measurement methods have been attracting attention to realize various sound measurements that are difficult to achieve with microphones. In this research, in collaboration with the Quantum Optical Physics Research Group of the NTT Basic Research Laboratories, we have developed a sound field measurement technique with much lower noise than conventional methods using precision optical measurement technology. Focusing on the difference in physical characteristics between sound and noise components, we proposed a sound measurement technique using a differential midfringe locked interferometer that reduces noise while maximizing sensitivity to sound. By combining this technology with high-stability laser technology, the amplitude of measurement noise has been reduced to 1/30 compared to conventional methods. By applying the precision measurement technology cultivated in optics, we will realize the next-generation sound measurements such as 3D sound-field measurement, remote sensing, and ultra-high precision measurement.

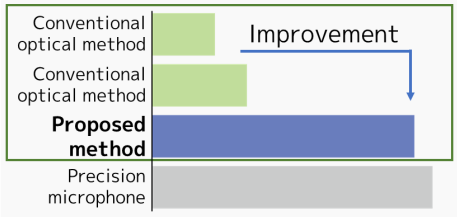
What happens with sound and light?

Optical technology makes it possible to observe sound remotely and without contact. It is expected to be applied to high spatial resolution measurement and remote sensing.



Proposed method

Measurement noise is reduced to **1/30** of conventional optical measurement methods.



Conventional methods are unable to measure small sounds due to noise.

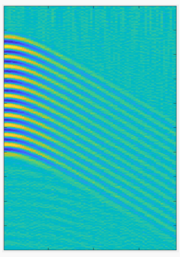
The proposed method enables measurements in various situations.

Key points

The fluctuation of light due to sound is tiny, and it has been challenging to detect faint sounds with light. In this study, we have achieved **the world's lowest noise** level by combining a **differential midfringe locked interferometer with ultra-stable laser technology**.

Future prospects

We realize the next generation of acoustic measurement



Visualization of sound wave

e.g., Non-contact measurement of weak sound

Capable of directly capturing the sound propagation.

Features of optical measurement

Non-contact	Precise	Highly stable
High speed	Controllable	Image sensor

Applications

- High precision measurement
- Sound VR
- Aeroacoustic measurement
- Acoustic Phenomena
- Sound-field imaging
- Remote sensing
- Noise visualization
- Multimodal sensing

References

[1] K. Ishikawa, Y. Shiraki, T. Moriya, A. Ishizawa, K. Hitachi, K. Oguri, "Reducing noise of mid-fringe locked interferometer by optical differential detection," in *Proc. Acoust. Soc. Jpn.*, 2021. (in Japanese)
[2] K. Ishikawa, Y. Shiraki, T. Moriya, A. Ishizawa, K. Hitachi, K. Oguri, "Simple and low-noise optical measurement of sound using mid-fringe locked interferometer," in *Proc. Acoust. Soc. Jpn.*, 2020. (in Japanese)

Contact

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

15 Joint signal separation, dereverberation and noise reduction

Extracting voices out of noise & reverberation

Abstract

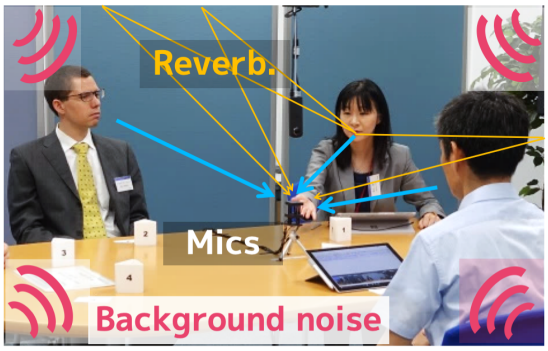
To enable such audio devices as smart speakers to accurately recognize human voices in real-world environments, we must reduce the noise and reverberation from the signals observed by microphones and extract each individual's voice. State-of-the-art (SOTA) technology addresses this problem by sequentially applying the following three techniques: (a) dereverberation, (b) source separation, and (c) denoising. However, SOTA is ineffective in noisy reverberant conditions because all three techniques, (a), (b), and (c), are optimized individually without considering the overall performance. In this exhibit, we introduce a new technology that jointly optimizes (a), (b), and (c) to maximize the quality of the output audio. Our new technology significantly improves the speech recognition performance compared to the SOTA method. It will contribute to a more convenient world where people and computers can interact smoothly in our daily environments, including train stations, streets, and shopping malls.

Speech Extraction from Microphone Recordings

Computers struggle to accurately recognize audio signals recorded by distant microphones

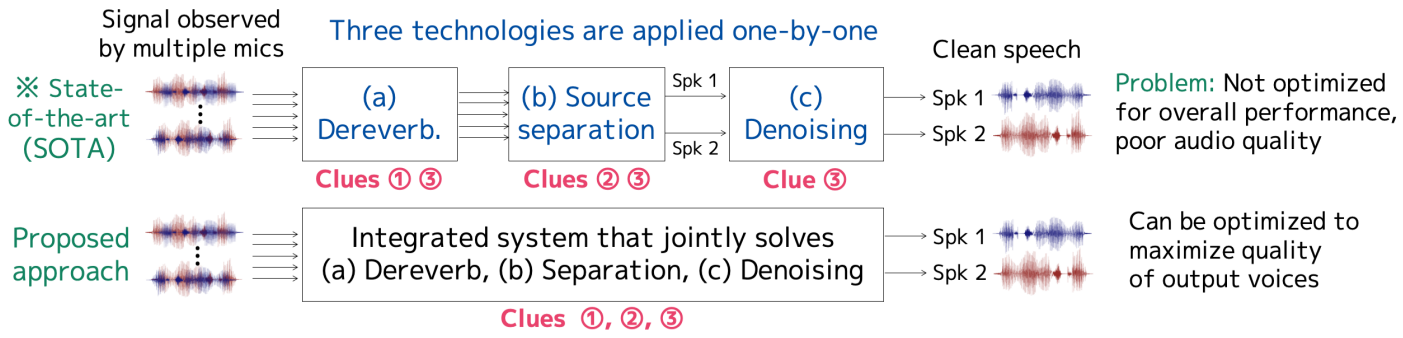
Causes of difficulty	Required technologies
High level of reverberation	(a) Dereverberation
Multi people talking simultaneously	(b) Source separation
High level of background noise	(c) Denoising

⇨ We accurately extract each individual's voice using signal processing technologies



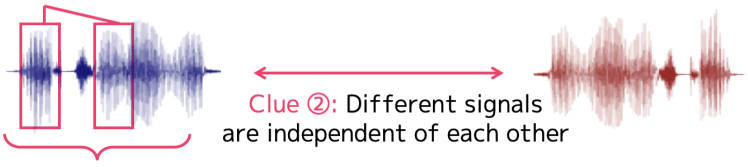
Problem of SOTA technique & our result

※ NTT's CHiME3 technology remains a widely used SOTA technology



Proposed framework unifies (a), (b), and (c)

Clue ① Different time intervals are not similar (independent)



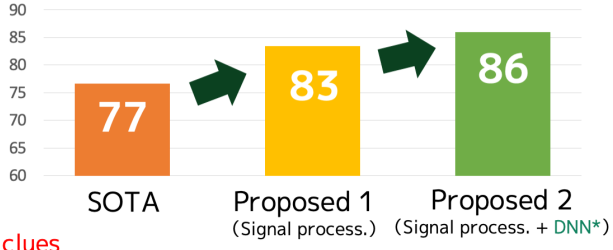
Clue ②: Different signals are independent of each other

Clue ③: Voice intensity is time-varying (nonstationary)

⇨ Integrate (a), (b), and (c) by simultaneously considering all three clues

Speech recognition performance [%]

Speakers	2
Mics	8



*When estimating part of a probabilistic model using DNN

References

[1] T. Nakatani, C. Böddeker, K. Kinoshita, R. Ikeshita, M. Delcroix, R. Haeb-Umbach, "Jointly optimal denoising, dereverberation, and source separation," in *Proc. IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 28, pp. 2267-2282, 2020.
[2] R. Ikeshita, T. Nakatani, S. Araki, "Block coordinate descent algorithms for auxiliary-function-based independent vector extraction," in *Proc. IEEE Trans. Signal Process.*, 2021, to appear.
[3] R. Ikeshita, T. Nakatani, "Independent vector extraction for fast joint blind source separation and dereverberation," in *Proc. IEEE Signal Process. Lett.*, 2021, to appear.

Contact

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

16

Crossmodal learning for concept acquisition of human movements

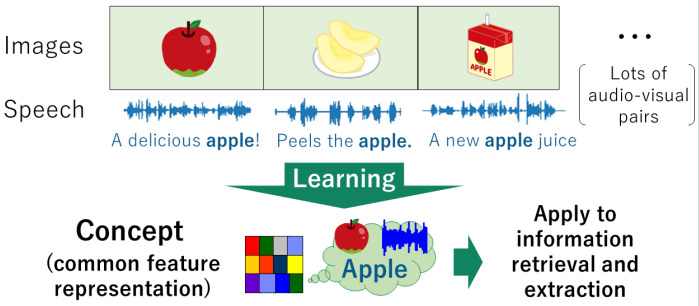
AI that acquires knowledge just by watching TV

Abstract

We developed a crossmodal learning method that can acquire "concepts" corresponding to specific objects and events on unlabeled audio and video signals. Achieving it in an unsupervised way is particularly important, since it is generally difficult to manually label all the objects and events appearing in audio-visual data for supervised learning. Our main idea was identifying concepts by looking at them from different modalities, just like looking at objects from different angles. To efficiently detect and utilize temporal co-occurrences of audio and video information, we employed a guided attention scheme. Experiments using real TV broadcasts of sumo wrestling with live commentaries show that our method can automatically associate specific athlete techniques and its spoken descriptions without any manual annotations. We are aiming for a future in which AI can acquire knowledge autonomously by just watching and listening to everyday scenes, or watching TV.

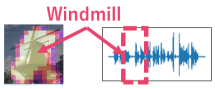
Crossmodal concept acquisition

Acquire “concepts” by learning semantic associations based on co-occurrences across different modalities in unsupervised manner



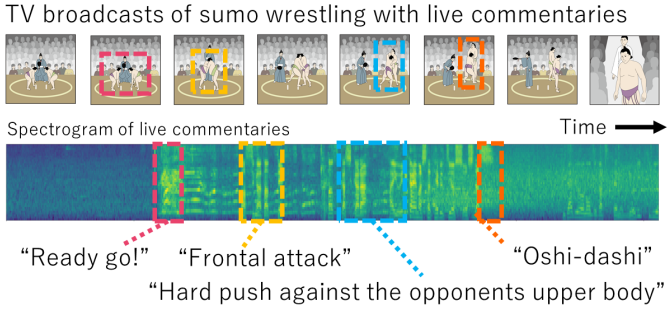
AI that acquires knowledge just by watching TV

Demonstrated feasibility of concept acquisition by associating visual objects with spoken words [1][2]



Concept acquisition of human movements

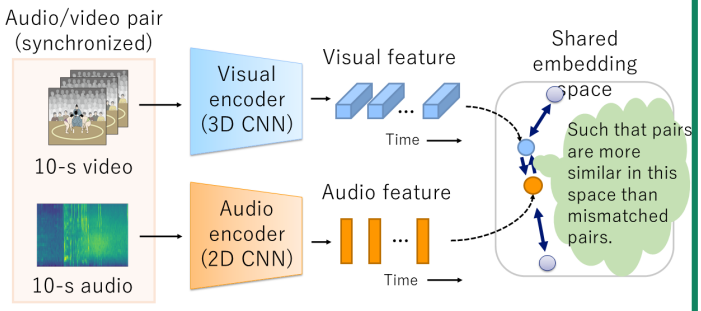
Utilize temporal proximity of spoken words appearing close to human movements in time



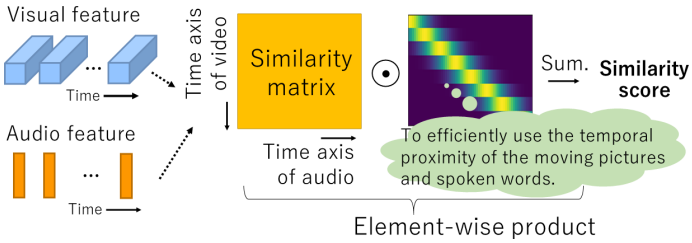
Introduce “guided attention” scheme along time to capture spatio-temporal correspondence [3]

Learning method and results

Method: train parameters of audio/visual encoders to optimize a ranking-based criterion

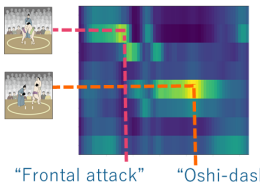


Point: compute similarity matrix between visual and audio features so that matrix becomes nearly diagonal



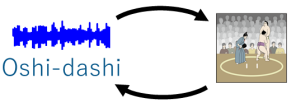
Results: implicitly learned underlying concepts of athlete movements in sumo bouts

Similarity matrix between visual and audio features



Concept retrieval (Recall score)

Audio-to-video retrieval:
Improved from 0.08 to 0.36



Video-to-audio retrieval:
Improved from 0.04 to 0.29

*Experimental conditions:
- 1,218 matches of NHK broadcast video of grand sumo tournaments in which the winners were determined by nine frequent winning techniques (1,128 for training and 90 for validation)

References

[1] Y. Ohishi, A. Kimura, T. Kawanishi, K. Kashino, D. Harwath, J. Glass, “Trilingual Semantic Embeddings of Visually Grounded Speech with Self-attention Mechanisms,” in *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2020*.
[2] Y. Ohishi, A. Kimura, T. Kawanishi, K. Kashino, D. Harwath, J. Glass, “Pair Expansion for Learning Multilingual Semantic Embeddings using Disjoint Visually-grounded Speech Audio Datasets,” in *Proc. Interspeech 2020*.
[3] Y. Ohishi, Y. Tanaka, K. Kashino, “Unsupervised Co-Segmentation for Athlete Movements and Live Commentaries Using Crossmodal Temporal Proximity,” in *Proc. International Conference on Pattern Recognition (ICPR) 2020*.

Contact

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

17 Emotional voice conversion via facial expression recognition

Real-time speech emotion contololler using face

Abstract

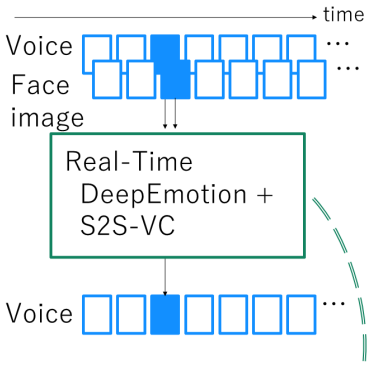
There are many kinds of physical or mental barriers that prevent individuals from smooth verbal communication. One key technique to overcome some of these barriers is voice conversion (VC), a technique to convert para/non-linguistic information contained in a given utterance without changing the linguistic information. Here, we propose a crossmodal voice control system, which offers a way to control the vocal expression of emotion in speech through the facial expression in a face image. The proposed system consists of performing facial expression recognition (FER) followed by VC. For VC, we have developed a method based on sequence-to-sequence (S2S) learning, which is designed to convert the prosodic features as well as the voice characteristics in speech conditioned on the output of the FER system. We believe that this work can provide some insight on what it is like to be able to control our voice through different modalities.

Communication augmentation system

Using voice conversion (VC) technique to help overcome barriers that prevent us from smooth communication

Voice expression control through face

Crossmodal voice control consisting of facial expression recognition and VC



Core techniques

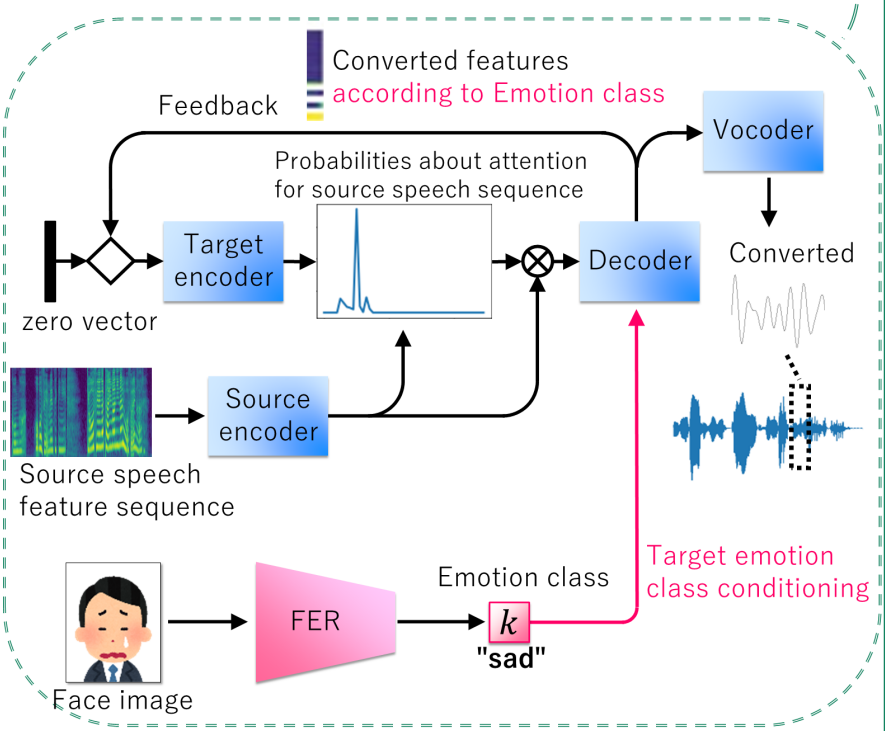
Sequence-to-sequence VC

■Sequence-to-sequence (S2S) learning
- Offers a general framework for transforming one sequence into another variable length sequence
- Encoder/decoder structure and attention mechanism make it possible to learn conversion rules that reflect long-term dependencies in input/output sequences
- Usually requires large-scale parallel corpora

■ VC based on S2S learning (S2S-VC)
- Voice expressions are characterized by prosodic features (e.g., intonation and rhythm)
- S2S-VC is able to convert prosodic features as well as voice characteristics in input speech with limited amount of training data

Facial expression recognition (FER)

■FER using attentional convolutional network
- After prediction, output is passed to VC system



References

[1] H. Kameoka, K. Tanaka, T. Kaneko, N. Hojo, "ConvS2S-VC: Fully convolutional sequence-to-sequence voice conversion," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, pp. 1849-1863, June 2020.
[2] K. Tanaka, H. Kameoka, T. Kaneko, N. Hojo, "AttS2S-VC: Sequence-to-sequence voice conversion with attention and context preservation mechanisms," in *Proc. 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2019)*, pp. 6805-6809, May 2019.
[3] M. Shervin, M. Minaei, and A. Abdolrashidi, "Deep-emotion: Facial expression recognition using attentional convolutional network." *Sensors 21.9:3046*, 2021.

Contact

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

18 Touched places detection using heat traces by thermography

Visualizing touched places for Corona prevention

Abstract

One effective infection prevention method is regularly disinfecting the environmental surfaces of public areas. However, no method currently identifies where and how much others have touched a particular door or shelf. This exhibition introduces a visualization system that identifies touched places by overlay projection. When a person touches an object, her hand's heat remains on its surface. This heat trace can be captured by thermography cameras. By combining Near Infrared (NIR) and thermography cameras, we can detect the touched places with a light algorithm. This technology enables us to identify the objects or places that others have touched, although the virus itself remains invisible to the naked eye. We believe that our system will help relieve anxiety during the COVID-19 pandemic. It will also enable us to gather statistics and data about the places touched by people and improve the efficiency of disinfection activities.

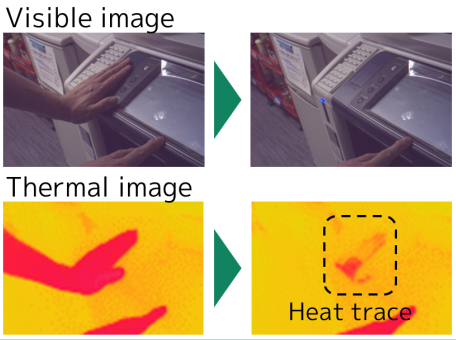
COVID-19 related anxiety

Infection prevention
- Hands → washing and sanitizing
- Air → ventilation
- Environmental surfaces (public areas)
  - Regular disinfection steps
  - Where and how much did others touch?

Visualizes touched places

Touched-places detection using heat traces

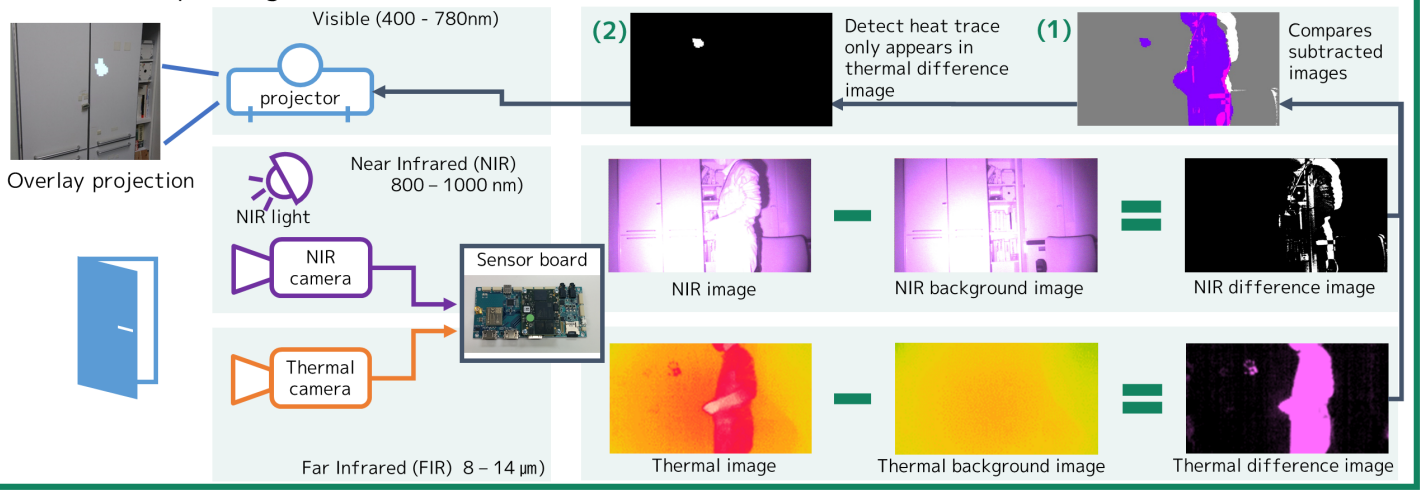
- Heat from hands remains on object surfaces after they are removed.
- Touched places can be detected by heat traces captured by thermography.



System architecture

- Separate wavelengths simplify image processing
  - Visible: information projection
  - NIR: capturing people's movements (no influence from projector)
  - FIR: capturing heat trace

- Light algorithm to detect heat traces by combining thermography and NIR cameras.
  (1) Compares their background images
  (2) Detects heat trace that only appear on thermal images.
- Enable us to identify the touched places by others.



References

[1] Y. Kishino, Y. Shirai, Y. Yanagisawa, K. Ohara, S. Mizutani, T. Suyama, "Identifying Human Contact Points on Environmental Surfaces using Heat Traces to Support Disinfect Activities," *SenSys2020 COVID-19 Pandemic Response*, 2020.

Contact

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

19 Biomedical sound analysis utilizing physical characteristics

Telestethoscope: Looking into Body by Listening

Abstract

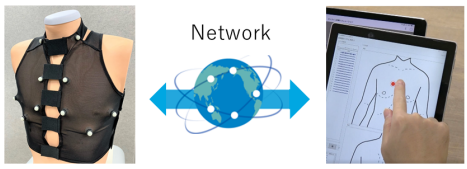
As a basic research toward providing a person with an enhanced sense of well-being, such as early detection of diseases, we developed a wearable acoustic sensor array system that can collect sounds from various parts of the human body and send the signals remotely to a receiver terminal, which is equipped with 18 acoustic sensors inside an examination vest. When the system comes into practical use for medical care, a medical practitioner will be able to listen to sounds from various locations on the patient's body without having to make direct physical contact with the patient or use of a traditional stethoscope, which will be useful in online medical examinations. This system is also expected to play a role, potentially in combination with other information sources, in the research and development of new medical techniques such as the visualization of physical states and direct translation from body sounds into explanatory sentences.

Tele-auscultation / AI auscultation

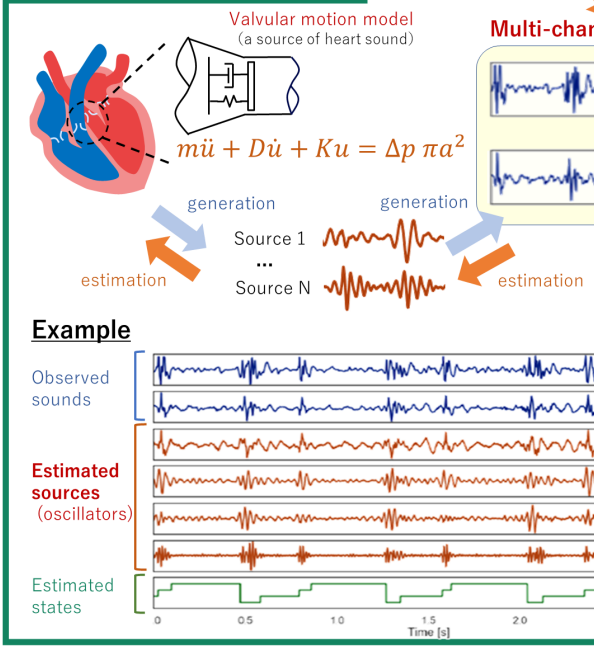
We developed a wearable sensor array system, named telestethoscope, equipped with 18 acoustic sensors that can collect sounds from various parts of the human body and send the signals remotely to a receiver terminal.
We aim for its applications to:

- Tele-auscultation (by medical personnel)
- AI auscultation (by computers, that is, automatic estimation and prediction of physical condition based on the sounds and other sources of information)

Our approach is characterized by attempts to exploit physical properties and laws.

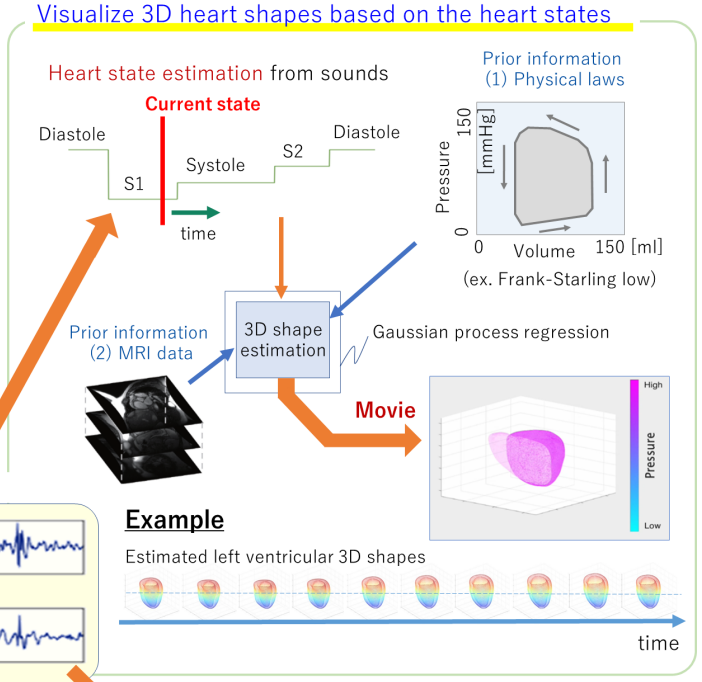


Sound source estimation

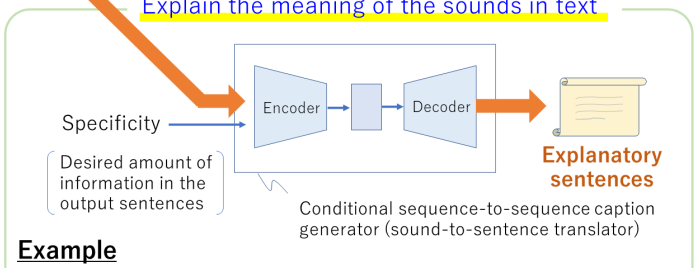


Sound-based visualization and explanation

Visualize 3D heart shapes based on the heart states



Explain the meaning of the sounds in text



Example

Specificity	Generated sentence
Low:	The heart sounds are abnormal.
High:	The heart sounds are abnormal. There may be a problem with one of the heart valves. The 1st sound is normal, and the 2nd sound is split.

References

[1] S. Ikawa, K. Kashino, "Neural audio captioning based on conditional sequence-to-sequence model," in *Proc. Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE)*, 2019.
[2] M. Nakano, R. Shibue, K. Kashino, S. Tsukada, H. Tomoike, "Gaussian process with physical laws for 3D cardiac modeling," in *Proc. European Signal Processing Conference (EUSIPCO), pp. 1452-1456*, 2020.

Contact

Kunio Kashino / Media Information Laboratory and Biomedical Informatics Research Center
Email: cs-openhouse-ml@hco.ntt.co.jp

20

Prediction and demonstration via life history theory

Is falling birthrate related to population density?

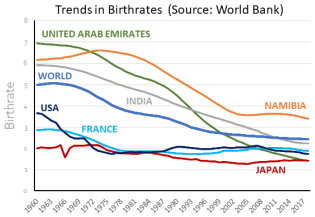
Abstract

Low birthrates and increased longevity are critical social problems. This study approaches the issue from a biological perspective. By analyzing such public statistical data as census information, we found that in Japan, some areas, which have high population densities, also tend to have low fertility rates and extended longevity. This idea may reflect the fact that in highly competitive situations, the number of children will be reduced to allow greater investment for child to increase the chance for success. Declining birthrates are often discussed in terms of childbirth and childcare systems, employment, and economic conditions. Perhaps this phenomenon should be studied from multifaceted perspectives to gain basic knowledge for planning essential solutions.

Low Birthrate and Longevity

Low birthrates and greater longevity are occurring worldwide.

Since this happens even in such different social situations as politics, economics, and culture, we assume some factors are shared by all humans.



This research focuses on human beings as living organisms.

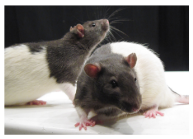
Since low birthrates and greater longevity are complex social issues, they must be examined from a variety of perspectives and their solutions must be widely discussed. This research focuses on biological perspectives as basic knowledge.

Life History Strategies: Fast and Slow

Life history strategy (LHS):

an evolutionary biological theory that studies how organisms allocate their time and resources in reproduction and growth.

Fast LHS



- High birth and death rates
- Less nurturing
- Early maturation
- Small body
- Short life

These species reproduce often and quickly, although available resources for each baby are reduced. Unfortunately, fewer will succeed.

Slow LHS

- Low birth and death rates
- More nurturing
- Slow maturation
- Big body
- Long life

These species limit their offspring and give more resources to each one.
Higher chance of success per offspring.

Low birth rates and longer longevity

Variance exists between people who have fast/slow life history strategies.

A greater survival risk suggests a faster LHS. This idea can be interpreted as a faster LHS, since child mortality was higher when medical care was more primitive.

Low birthrates and extended longevity may be caused by people changing to a slower life history.

Dense Prefectures have SLOW LHS

Parents must consider the potential competition faced by their children.

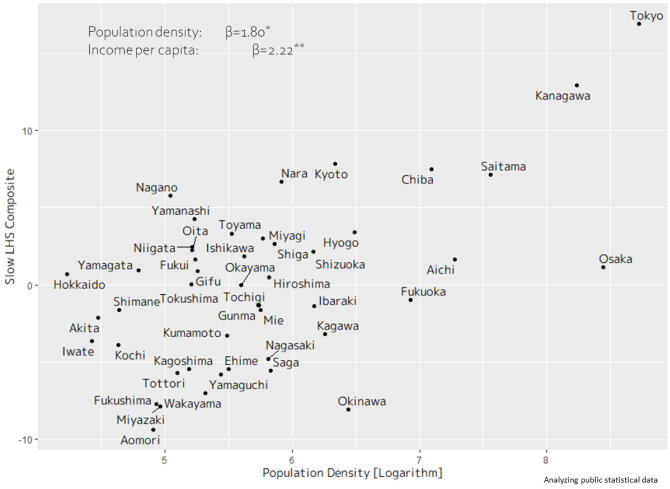
Where competition is intense, a greater nurturing investment is required to ensure a child's success. The number of children will decrease in response to increased investment per child.

Intensity of competition can be approximated by population density.

The more people in a region, the more competition they will experience.

The more densely populated the prefecture, the slower its life history.

This idea is consistent with Sng et al. (2017) who analyzed countries and regions around the world and the 50 U.S. states. Effect of population density remained--even after adjusting for income per capita.



Toward the Future

Today, declining birthrates are often attributed to such social economic situations as childbirth/childcare assistance, employment, and economic conditions. Perhaps these are not the only causes. Future human societies must investigate essential causes and devise solutions through multidimensional analysis.

References

[1] M. Matsuda, T. Kobayasi, "Do population density and economic situation affect residents' life history strategies? -- An analysis via web survey --," IEICE Technical Report, Vol. 119, No. 394, HCS2019-69, pp. 89-92.
[2] M. Matuda, "Low birthrate and longevity in Japan -- Comparison between prefecture level and municipal level --," IEICE Technical Report, Vol. 118, No. 437, HCS2018-66, pp. 107-111.

Contact

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

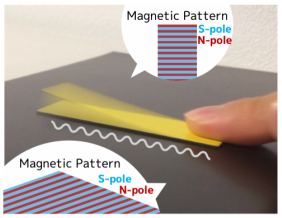
Magnetact: A Magnetic force-based Tactile technology

Moves of magnetic fields, Moves by magnetic fields

Abstract

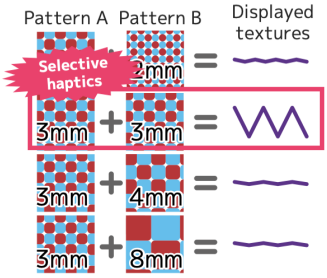
Magnetact is a magnetic tactile technology that does not require any power supply or wiring to present a tactile sensation. The specific patterned magnetic rubber sheets present a unique tactile texture when they are rubbed together. However, it takes time to magnetize the kind of detailed lattice pattern of S/N poles that can provide a selective tactile presentation. Therefore, we invented a new method for generating complex geometric magnetic patterns by layering multiple magnetic rubber sheets with simple magnetic patterns. Moreover, the superposed magnetic lattice patterns can be dynamically changed by rotating the layered magnetic sheets. This method resolves the tradeoff between the complexity of the magnetized pattern and the time required for magnetization.
The Magnetact technology is expected to be applied to a variety of tactile experiences, including in online tactile workshops, portable tactile picture books, and low-cost tactile VR experiences.

Magnetact: A Magnetic Tactile Technology



Magnetact is a magnetic force-based tactile technology. By magnetizing specific magnetic patterns on a magnetic rubber sheet, it presents unique tactile textures when the sheets are rubbed together.

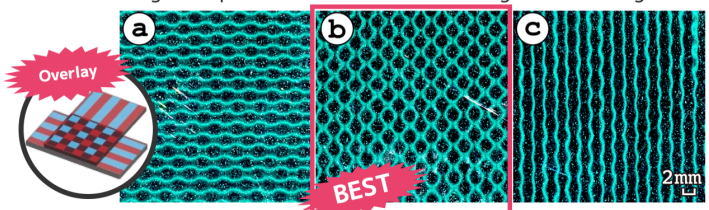
Pattern Superposition by Overlaying Magnets



It is relatively easy to magnetize a simple stripe pattern of S/N poles on the sheet, but it takes time to magnetize the kind of detailed lattice pattern that can provide a selective tactile presentation.

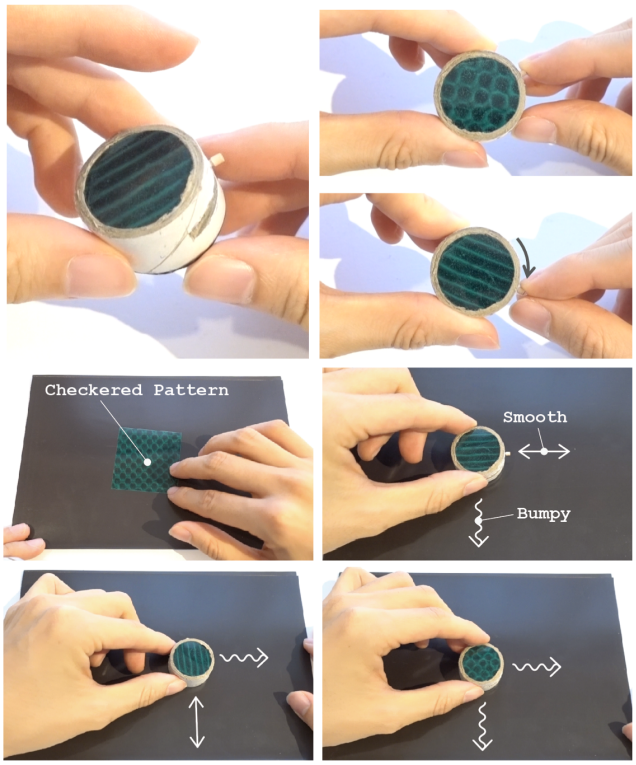
To solve this problem, we have established a method to generate complex geometric magnetic patterns by layering multiple magnetic rubber sheets of appropriate thicknesses. This method shortens the construction time of complex magnetic patterns to about 1/7 that of the machine magnetizing method.

Magnetic patterns visualized with magnetic viewing film



	a	b	c
(upper t_1)	0.31 mm	0.51 mm	0.82 mm
(lower t_2)	1.52 mm	1.52 mm	1.52 mm

Dynamic Change of Magnetic Fields



Moreover, the superposed magnetic patterns can be changed by rotating the layered sheets.

This evolution of Magnetact resolves the tradeoff between the complexity of the magnetized pattern and the time required for magnetization. It is expected to enable online tactile workshops, portable tactile picture books, and low-cost tactile VR experiences.

References

[1] K. Yasu, "Magnetic plotter: a macrotexture design method using magnetic rubber sheets," in *Proc. the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*, 2017.
[2] K. Yasu, "MagneLayer: force field fabrication by layered magnetic sheets," in *Proc. the 2019 CHI Conference on Human Factors in Computing Systems (CHI '20)*, 2020.
[3] K. Yasu, M. Ishikawa, "Magnetact Animals: a simple kinetic toy kit for a creative online workshop for children," in *Proc. CHI EA '21*, Article. No. 198, pp. 1–4, 2021.

Contact

Kentaro Yasu / Sensory representation research Group, Human and Information Science Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp

22

Simple method for modulating tactile texture

Make hard objects soft, make rough objects smooth

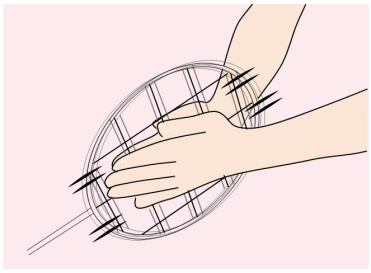
Abstract

To artificially present rich tactile textures, a tactile display that can modulate tactile textures such as hardness and roughness of various daily objects is necessary. In this study, we focused on a tactile illusion, named the 'velvet hand illusion' (VHI), which could be used to modulate tactile textures in a simple setup. We evaluated tactile textures during the VHI qualitatively and quantitatively and found that the VHI is a phenomenon wherein the tactile texture of metallic wires being touched was perceived to be softer than it actually is, with the tactile texture of cloth. Based on the findings, we developed the rotating-frame method wherein tactile textures of arbitrary objects can be modulated to seem softer and smoother than they actually are. Since the proposed method can be applied to the tactile textures of daily objects, it is expected that the method could be used in various application scenes such as product design and in-store demonstrations.

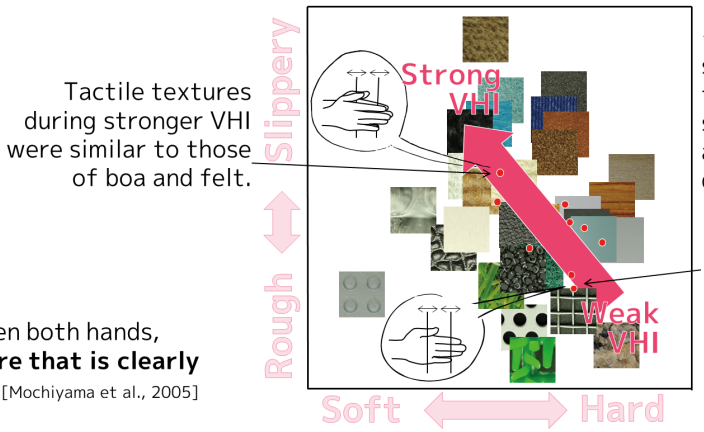
Investigating tactile perception during the velvet hand illusion

-We evaluated tactile textures during VHI by comparing them with tactile textures for various daily materials.
-When feeling stronger VHI, the texture of wires between both hands was perceived to be softer and warmer as if it was cloth.

Velvet Hand Illusion (VHI)



When sandwiching moving wires between both hands, one feels **an unexpected tactile texture that is clearly different from that of moving wires.** [Mochiyama et al., 2005]



Tactile textures during stronger VHI were similar to those of boa and felt.

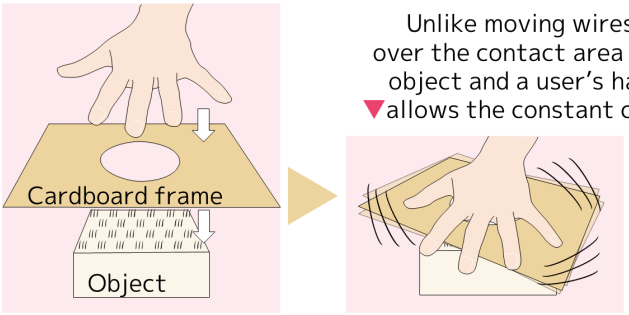
◀Results of the similarities of tactile textures sensed during VHI and various actual daily materials

Tactile textures during the weakest VHI were similar to those of wires.

Extending VHI-like modulation from wire to arbitrary objects

-We developed a method of inducing VHI-like modulation of tactile textures for various objects.
-The method can modulate tactile textures for various objects to seem softer and smoother than they actually are.

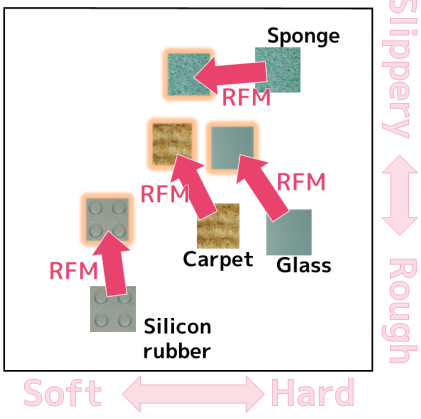
Rotating-frame method (RFM)



The tactile texture of an arbitrary object can be modulated when a rotating frame is sandwiched between an object and a user's hand.

Unlike moving wires which cross over the contact area between the object and a user's hand, the RFM ▼allows the constant contact area.

Results visualizing▶ the RFM's modulation of tactile textures for various objects



References

[1] T. Yokosaka, S. Kuroki, S. Nishida, "Describing the sensation of the 'velvet hand illusion' in terms of common materials," *IEEE Transactions on Haptics*, DOI:10. 1109/TOH, 2020. 3046376. Online ahead of print.
[2] T. Yokosaka, Y. Suzuishi, S. Kuroki, "Feel illusory texture through a hole: Rotating stimulus modulates tactile sensation for touched object's surface," in *Proc. EuroHaptics2020 (WIP137)*, 2020.

Contact

Takumi Yokosaka / Sensory representation research Group, Human and Information Science Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp

23

Empathetic telecommnunication by vibrotactile transmission

Sense of touch connects our hearts beyond distance

Abstract

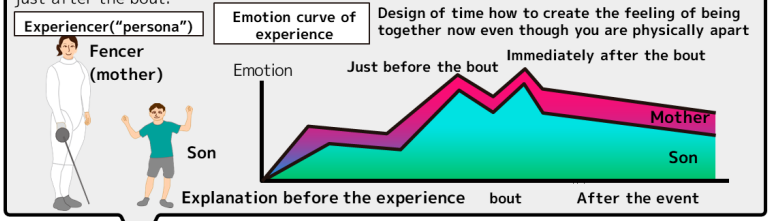
Reconstructing empathic communication that allows people to share feelings even in the remote age, by designing 3 frameworks as a solution. We have created a system called “remote high-five” through which families in remote areas can exchange “high fives” during sports events. As a framework for realizing empathic communication in society, we arranged three elements—namely, design of the story (including the people who experience it and the flow of the experience), technology that supports tactile transmission, and context that inclusively involves all parties—and were able to show the effectiveness of the “remote high-five” system with an example. We aim to create a society in which anyone can communicate with each other and feel that they are together even if the people are far apart.

What is “empathic communication”?

Owing to the novel coronavirus pandemic, conventional physical communication between people is being lost. In such a remote-communication era, it is important to reconstruct **empathic communication** that shares people’s feelings. For that purpose, it is essential to design three frameworks: **experience, technology, and collaboration**.

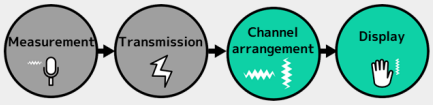
Designing three frameworks to reconstruct empathic communication

It is important to design those frameworks in consideration of who the experiencer is and the timing of the peaks of the experience. During this fencing bout, family members in separate remote locations at the time of the bout could share the momentary feelings that they most wanted to convey to each other at the two “peak” times, namely, just before and just after the bout.

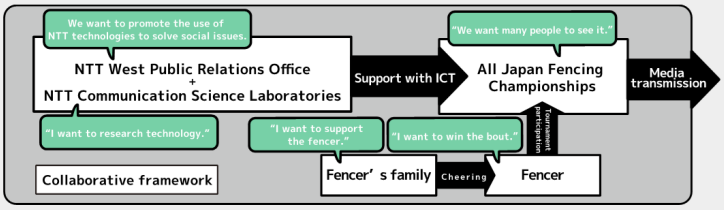


1. Experience framework — Story Design
2. Technology framework — Haptic transmission
3. Collaborative framework — Contextual Design

To realize this experience, we need the technology to support it. So we are working on standardizing the transmission of tactile-vibration signals in conjunction with audio. Specifically, we contributed to the multimedia extension of IEC 60958-5 (published in February 2021), an international standard for digital audio interfaces, in regard to specifying the transmission of tactile-vibration signals and the correspondence between channel number and vibration position.

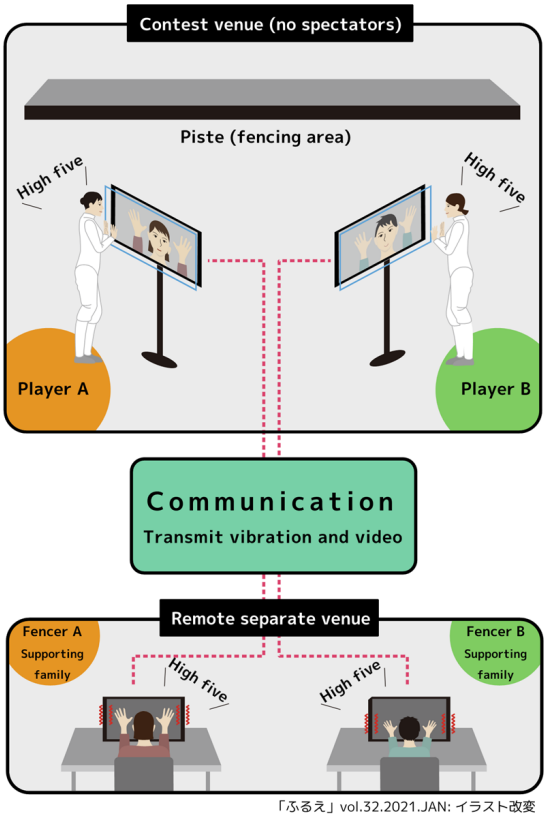


It is important to realize experience on the basis of (i) the background and purpose of each of the people involved and (ii) have a unified awareness of the overall direction. The framework for collaboration is established when all parties involved are empathetic to the story and involved in the design of the experience.



An example of empathic communication

At the All Japan Fencing Championships, we created a remote “high-five experience” (called “remote high-five”) through which people share joy and support even when they are far away, by sending vibrations and images via communication and sharing emotions.



「ふるえ」vol.32.2021.JAN: イラスト改変

Actually used device “remote high-five”

References

[1] H. Hayakawa, R. Owaki, T. Ishikawa, K. Minamizawa, Y. Tanaka, K. Komazaki, Y. Kamamoto, J. Watanabe, “Proposal of Interactive Visuo-Audio-Tactile Media “Public Booth for Vibrotactile Communication” for Telecommunication with High Presence The virtual reality society of japan, Vol. 25, No. 4, 2020.
[2] J. Watanabe, “The Institute of Electrical Engineers of Japan,” Vol. 141, No. 2, 2021.

Contact

Kakagu Komazaki, Junji Watanabe / Sensory Resonance Research Group, Human Information Science Laboratory
Yutaka Kamamoto / Computing Theory Research Group, Media Information Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp

24

Oxytocin as a neural regulator of maternal implicit approach

Why do mothers approach to infants' "crying"?

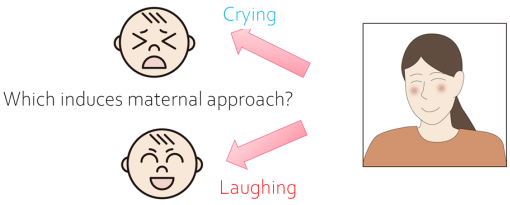
Abstract

It is well known that infant vocalizations, crying in particular, enhance maternal motivation to provide care and induces maternal approach. However, the mechanism has not been clarified. In this study, we examined which infant vocalization induces maternal approach and what neural factor regulates it. We measured the position of the center of pressure of mothers to evaluate their natural and implicit movement when they listened to infant voice stimuli. We took saliva samples from them to measure levels of oxytocin, one of the neural hormones that represent mental states. We found that mothers approached infants' "crying" because it raised a strong sense of "urgency." The approach distance for crying was negatively correlated with their oxytocin levels. We think our study will help establish home-scale well-being, and finally, social well-being.

1. Implicit maternal approach to infant voices

Purpose of this study

Infant vocalizations affect maternal emotion and behaviors.

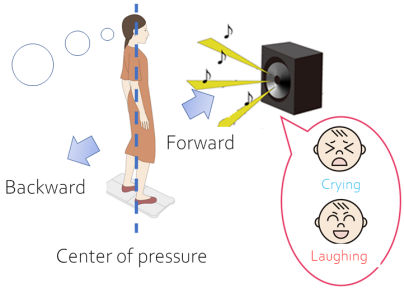


What vocalization induces maternal approach?
What neural factor regulates maternal approach?

2. Procedure

Self-reported emotion

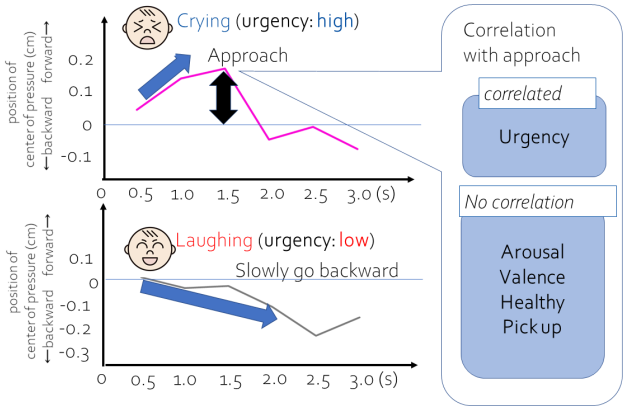
Arousal (aroused-not aroused)
Urgency (urgent-not urgent)
Valence (pleased-displeased)
Healthy (healthy-sick)
Pick up (desire to pick up the baby-no desire to do so)



Analysis of behavior
- Presentation of infants' voice (crying or laughing)
- Evaluation of maternal implicit movement (center of pressure)

Seeking the mechanisms
- Voice stimuli-induced emotion (e.g., urgency, valence)
- Individual difference in neural factor of mental state (oxytocin)

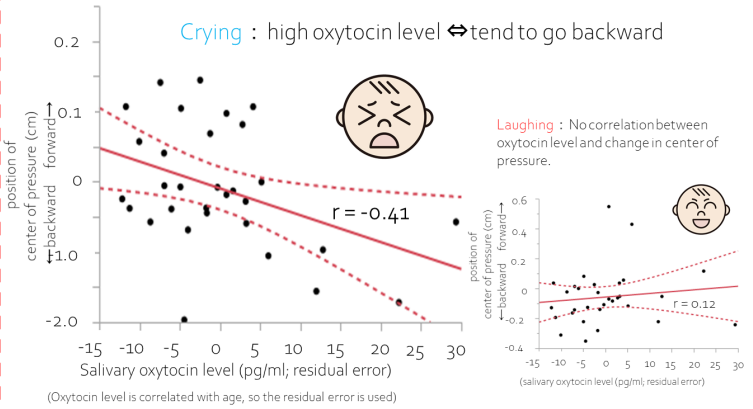
3. Mothers implicitly approach to "crying".



Mothers implicitly approach "crying"; it induces maternal emotion of "urgency" (wanting immediately to comfort a baby)

(Hiraoka, Ooishi, Mugitani, and Nomura, 2019 Frontiers in Psychology)

4. Oxytocin regulates implicit maternal approaching



(Oxytocin level is correlated with age, so the residual error is used)

Neural factor of mental states, oxytocin, regulates implicit maternal approach behavior.

(Ooishi, Hiraoka, Mugitani, and Nomura, 2020 Comprehensive Psychoneuroendocrinology)

References

[1] D. Hiraoka, Y. Ooishi, R. Mugitani, M. Nomura. "Differential effects of infant vocalizations on approach-avoidance postural movements in mothers." *Front. Psychol.*, Vol. 10, 1378, 2019.
[2] Y. Ooishi, D. Hiraoka, R. Mugitani, M. Nomura, "Relationship between oxytocin and maternal approach behaviors to infants' vocalizations," *Compr. Psychoneuroendocrinol.*, Vol. 4, 100010, 2020.

Contact

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

25

Relation between microsaccades and auditory spatial attention

Auditory attention that appears in the eye

Abstract

The eyes reflect various states of mind. It is known that gaze directed to an object of interest, but also that microsaccades, involuntary tiny eye movements, reflect the visual covert attention. Here, we investigated the relationship between microsaccades and spatial auditory attention. We found that microsaccades reflect the direction of auditory attention during a dichotic selective attention task and are also associated with task performance. Although many previous studies have already shown a link between microsaccades and visual attention, this study showed that they are also linked to auditory attention processes. We believe that this finding will lead to the development of technology for estimating attention states that vary spontaneously and instantaneously (e.g., estimating information such as the voice to which a person is paying attention at the party). We also hope that this result encourages future studies to elucidate the mechanisms of how our auditory system coordinates spatial attention.

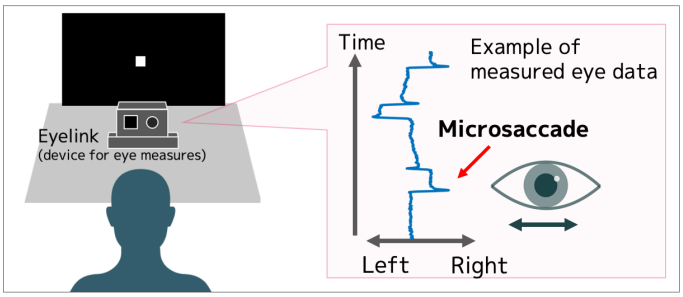
Significance

▶ Tiny saccadic eye movements, called microsaccades, reflect the direction of auditory attention.

This study showed the possibility of tracking listener's attention states by analyzing the properties of microsaccades.

What is a "microsaccade"?

A tiny involuntary saccadic eye movement that occurs even when fixating on one point

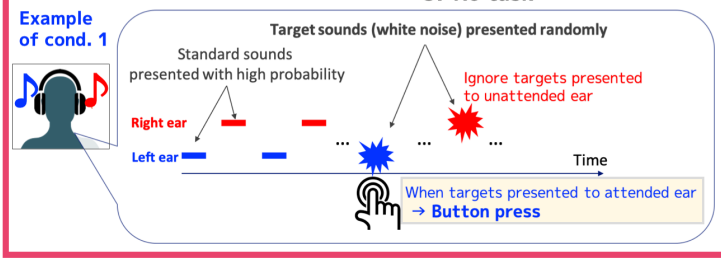


Dichotic selective attention task

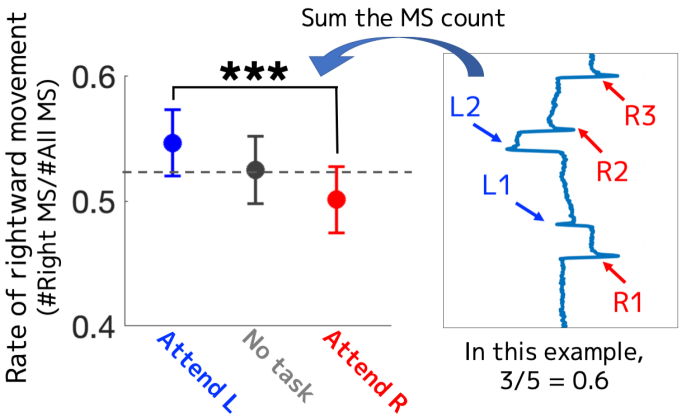
Q1: Do microsaccade reflect auditory attention direction?
Q2: Do microsaccade reflect attention level (task performance)?

Sound sequence presented dichotically → **Three conditions:**

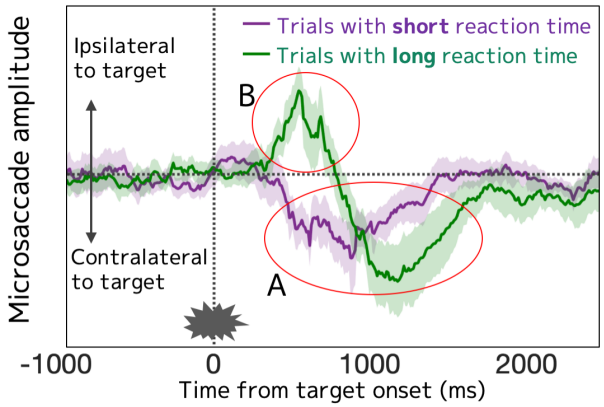
1. **Attend to left sound**
2. **Attend to right sound**
3. **No task**



Microsaccade bias contralateral to attention



Relationship between microsaccades and task performance (reaction time)



A) For both **short** and **long** RT trials, microsaccades were biased opposite to the target.
B) For **long RT** trials, microsaccade bias ipsilateral to the target was observed.

References

[1] S. Yamagishi, S. Furukawa, "Simultaneous measures of auditory brainstem frequency following response, pupillary response, and microsaccade during auditory selective attention task," in *Proc. 42nd Association for Research in Otolaryngology (ARO) Midwinter Meeting*, 2020.
[2] S. Yamagishi, S. Furukawa, "Relationship between auditory selective attention and microsaccades," in *Proc. The Auditory Research Meeting*, 2020.

Contact

Shimpei Yamagishi / Sensory representation research Group, Human and Information Science Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp

26

Professional drivers always blink at the same time over laps

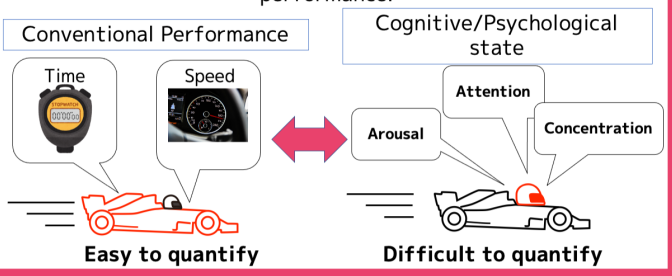
Blink pattern of elite car race drivers

Abstract

How do formula car drivers process the stream of information from the external world during ultra-high-speed driving? In this study, we focused on spontaneous blinking and captured the dynamic changes in the psychological and cognitive states that top Japanese drivers exhibit when driving around a circuit. As the lap speed increased, the drivers blinked at particular locations on the track and these locations were common between the drivers. We revealed for the first time that racers adjust their psychological and cognitive states to control their cars in distinct areas on the track with increasing lap speed. In future, we aim to decode these states that even professionals themselves do not understand through the unconscious physiological phenomena. By clarifying the implicit information processing occurring in the brain that leads to superior performance, we pursue a society where people can easily improve their desired skills.

Quantification of cognitive/psychological metrics in sports

It is difficult to examine athletes' psychological and cognitive states based on the measurement of conventional metrics of performance.

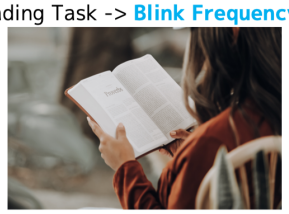


Spontaneous blinks reflect psychological and cognitive states

The frequency of **spontaneous blinking** is known to vary with emotion and task. For example,

Anger -> **Blink Frequency+**

Reading Task -> **Blink Frequency-**

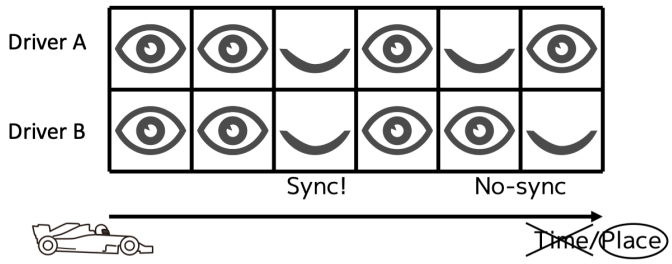


Blinks may capture psychological and cognitive states.

Blink patterns during race car (formula car) driving

Assume that **similar blink behaviors occur when drivers are in similar psychological and cognitive states.**

Q: Do they blink at similar locations during lap iterations? (**synchronization**)

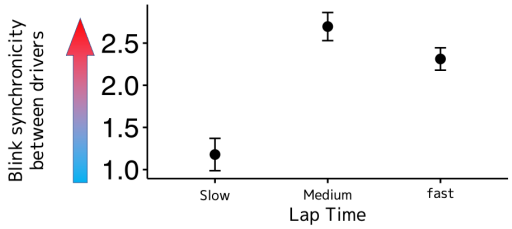


Blinks were synchronized at specific positions on the course.
→ The drivers were not always in the same psychological and cognitive state. Instead, they fine-tuned their state depending on their position on the course.

Both of the drivers blink at similar positions on the course, although the number of blinks in a lap greatly varies between them.
→ This biased pattern may be a trait of elite-level drivers.

▶ It may be possible to visualize and quantify fine attentional control by examining blink patterns.

Q: What is the relationship between driving performance and blink synchronicity?



The degree of synchronization correlated with the speed of the lap.
→ It may reflect psychological and cognitive load associated with the difficulty of the drive.

▶ The strength of synchronization between athletes may be a measure of the psychological and cognitive load imposed by the task.

Summary and future research

✔A course position- and speed-dependent blink pattern was found during race car driving.
✔Blink patterns may provide a quantitative representation of psychological and cognitive states.
✔Future studies should examine differences at various competition levels and applicability to other disciplines.

References

[1] R. Nishizono, N. Saijo, M. Kashino, "Synchronization of Spontaneous Eyeblink during Formula Car Driving", in *Proc. The 13th ACM Symposium on Eye Tracking Research and Applications (ETRA)*, 2021.

Contact

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

27

Eye-hand coordination in motor learning

Essence of “keeping your eye on the ball”

Abstract

Motor learning is indispensable to realize skillful behaviors. Under a dynamically changing environment like in tennis, we need to acquire and execute new skilled reaching movements while our gaze is directed toward the ball (foveal vision) and possibly the opponent’s location too (peripheral vision). While previous studies have taken gaze information into consideration for motor learning, conventional theories were restricted to emphasizing the superiority of the reaching to a foveated target over the reaching to a peripheral target. Our current study elucidated that the eye-hand spatial coordination for both foveal and peripheral reaching movements during learning is inherently linked with the internal model of learned reaching skills. Our results highlighted a novel interaction of gaze information with motor learning. By understanding the brain mechanism of this interaction, we will be able to design a novel training method that utilizes different gaze states to enhance sports training and rehabilitation.

Eye-hand coordination

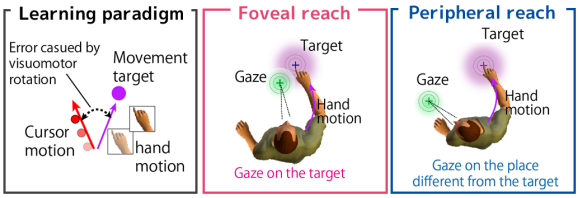
Why should gaze be considered for acquisition and execution of motor skills ?

Conventional theories have focused on the superiority of reaching movements with central vision over peripheral vision.

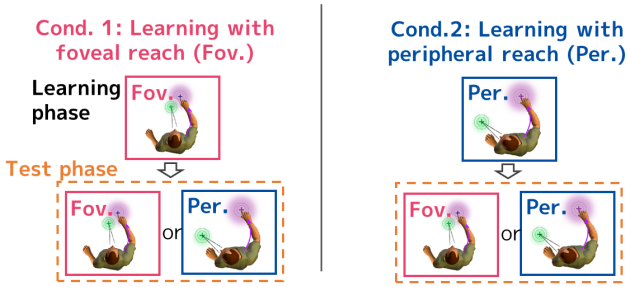
Our study elucidated that the eye-hand spatial relationship, including both foveal and peripheral vision, was inherently related with the motor learning of reaching movements.

Experimental method

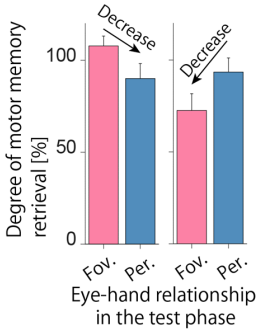
Participants move the cursor to the target. The cursor motion is rotated from hand motion. After repeating trials, hand motion is gradually changed to compensate for visuomotor rotation.



Exp.1: Does the eye-hand relationship have an impact on motor learning?



[Results] In both conditions, the retrieval of motor memory was degraded to about 80 % when the eye-hand relationship for the test phase differed from that of the learning phase.



For better retrieval of motor memory, same eye-hand relationship should be used in both learning and test phases.

Foveal and peripheral reach would be processed differently by the brain.

⇒Is it possible to acquire different motor skills simultaneously by using such distinct representations?

Exp.2: Does foveal and peripheral reach enable us to learn different reaching skills simultaneously?

Learning (480 trials) : Rotational directions (CW & CCW) changed across trials, but were unequally associated with eye-hand relationship.*1

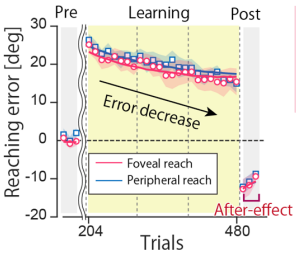
	Trial 1	Trial 2	Trial 3	-----	Trial 480
Eye-hand relationship	Foveal reach	Peripheral reach	Peripheral reach	-----	Foveal reach
Rotational direction	CW	CCW	CCW	-----	CW

An example of learning phase. Two rotational directions appeared with the same frequency and in no particular order.

[Results]

During learning: Error decrease for both rotational directions.
After learning: Clear “after-effect”*2 was observed.

*1 Simultaneous learning of different motor skills is known to be difficult.
*2 After-effect: Estimated in the non-rotational trials after learning. Larger negative value indicates better retrieval of motor memory.



Changing eye-hand relationship enables us to learn different reaching skills simultaneously

Reaching skill consists of gaze status as well as the novel pattern of hand movements

References

[1] N. Abekawa, S. Ito, H. Gomi, “Different learning and generalization for reaching movements in foveal and peripheral vision,” in *Proc. Adv. Mot. Learn. Mot. Control*, 2019.
[2] N. Abekawa, S. Ito, H. Gomi, “Foveal and peripheral vision separate motor memories for reaching movement,” JNNS2020, 2020.

Contact

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

28

Quantifying motor-skill performance using a smartphone

How do you define your dominant hand?

Abstract

The skilled hand in sports is usually determined by a handedness questionnaire [1], but it is blind to forced right-handed correction and does not measure motor performance. We invented a new method to easily evaluate each hand's motor skill by quantifying the variability of fast repetitive motion. Unlike other measures of motor performance that require customized equipment, our method is practical as it uses a smartphone's accelerometer to determine motor skill, making it easier to use in the field and at home. Our method could detect individuals who were forcibly corrected as their left hand was more skilled than the left hand of natural right-handers. In the near future, we plan to use our methodology to quantify sports training aimed at specific movements, and to motivate physical rehabilitation via daily feedback. It may also serve as a tool in understanding how and why the brain's control of the left and right arms is different.

Quantifying motor skill

Preferred and skilled hand could be different

Q1. When forced to be right-handed, which hand is the skillful one?

Q2. How to quantify skill?

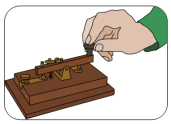
Major hurdles

Problem 1. Unbiased motor skill determination

Forced right-handers may mistakenly use the wrong hand for sports.

Trainers cannot easily determine the skilled hand.

Problem 2. Need for custom equipment



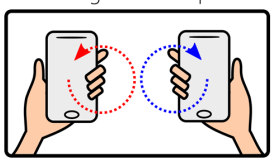
Tapping rate (30s)

Peg-in-hole task (20-40s)

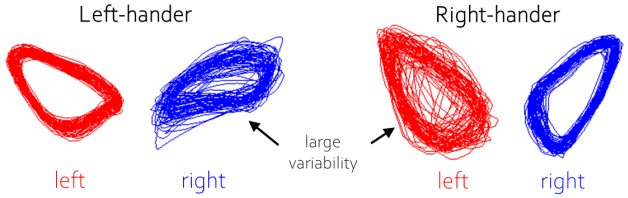
Past methods use custom equipment to determine skill, making them impractical for everyday-use.

Measuring motor skill with a smartphone

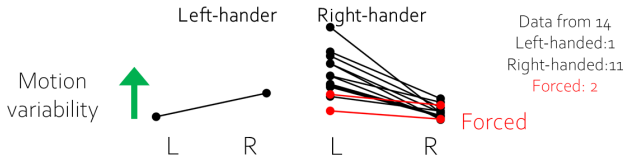
Motor skill was defined as motion with low variability, and was measured using the smartphone's accelerometer.



Rotate quickly for 15s per hand (total 30s)



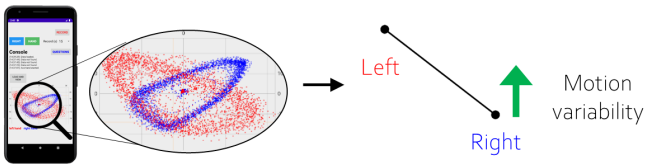
Answer 1. Forced right-handers are more skilled with the right hand.



Answer 2. Motor skill is quantifiable with motion variability.

Summary

- Effect of forced correction apparent by quantifying motor skill as low motion variability.
- Easy to measure, has potential to quantify sports training and left-right skill balance.



References

[1] A. Takagi, S. Maxwell, A. Melendez-Calderon, E. Burdet, "The dominant limb preferentially stabilizes posture in a bimanual task with physical coupling," *Journal of neurophysiology*, Vol. 123, No. 6, pp. 2154-2160, 2020.
[2] R. C. Oldfield, "The assessment and analysis of handedness: the Edinburgh inventory," *Neuropsychologia*, 1971.

Contact

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

Brain functions to recognize and hit a fastball

Abstract

When hitting a 95-mph blazing fastball in baseball, the batter must judge the ball's path and control the bat in about 0.4 seconds. However, it takes longer to achieve accurate judgment and movement. Though many studies have examined judgment and movement processes in the brain separately, they are closely related. We investigated how the brain establishes quick judgment and movement processing under strict time constraints. To this end, we conducted a baseball-like hitting experiment and clarified that the Strike/Ball judgment had less effect on hitting performance as the time constraint became stricter, but changing the movement strategy restrained the decrease. Our goal is to provide novel methods to evaluate and improve the brain functions of athletes to enhance cognitive-motor control in support of conventional approaches to physical fitness testing and training. We believe that this work will elucidate now hidden mental processes and find application in other research fields.

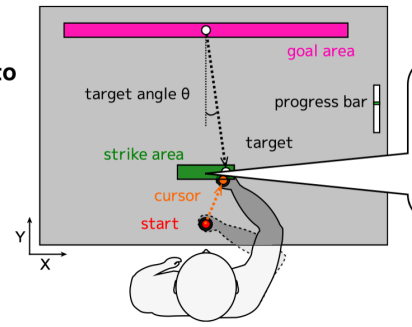
Quick judgment and movement

Though it takes more time for more accurate judgment and movement, how does our brain achieve quick judgment and movement processing under strict time constraints, like a blazing fastball?

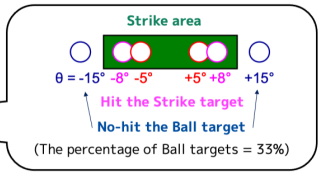
Our designed task is judging Strike/Ball target and hitting Strike targets.

Time from ball release to ball hitting
(time-to-contact: TTC)
- **0.4 s ≒ 95 mph**
- **0.5 s** ≒ 76 mph
- **0.6 s** ≒ 63 mph

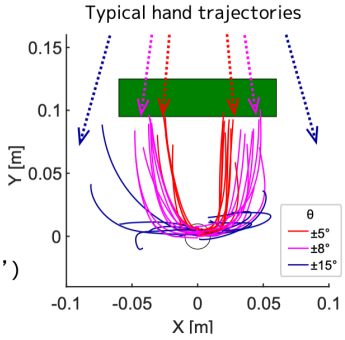
Another task to hit both targets without judgment was also conducted.



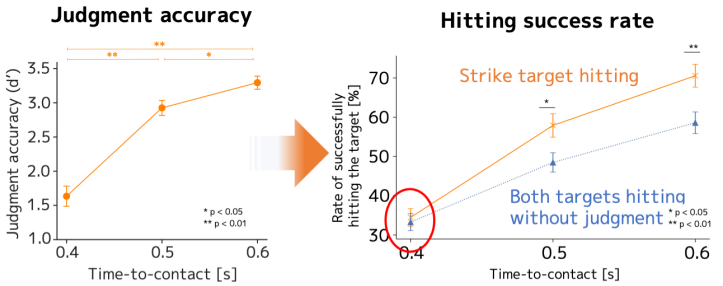
30 healthy adults participated.



Judgment accuracy: sensitivity (d')
(calculated by the signal detection theory)



Judgment and hitting success in a fastball?

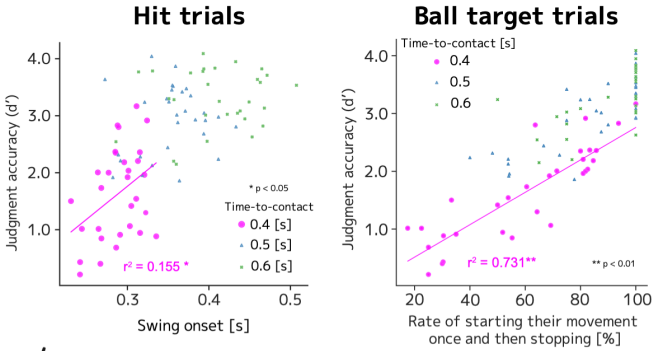


Judgment accuracy greatly impacted hitting success.

Strike/Ball judgement was effective at 0.5s and more TTC, but not at 0.4 TTC.

As the time-to-contact became shorter, the judgment accuracy got worse, resulting in a loss of judgment efficacy.

Swing strategy in a fastball?



'Wait before move as possible'

'Move, then stop when judging Ball'

The ability to change the swing strategy, slowing the swing onset and stopping the swing for Ball targets, is critical for improving judgment success.

References

[1] A. Kobayashi, T. Kimura, "Effects of cognitive strategy on hitting tasks," IEICE Technical Report, Vol. 118, No. 470, pp. 37-42, 2019.
[2] A. Kobayashi, T. Kimura, "Motor redundancy affects decision-making behavior," SICE Division of Life Engineering Symposium 2020, 2020.
[3] A. Kobayashi, T. Kimura, "Go/No-go decision making under severe time constraints interferes with hitting task performance," in *Proc. The Society for Neuroscience 49th Annual Meeting*, 2019.

Contact

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