TCT/TTA Joint Seminar 2018
Disruptive Technology and 5G supporting Thailand 4.0:
Challenge and Opportunity

Innovative R&D by NTT

NTT's Four AI Directions and Communication Science

February 01, 2018

corevo™: AI Technologies of NTT Group

Press release on 30th May, 2016

- AI technologies accelerating collaborations with variety of players in different fields and creating infinite values
- Human and machine collaboration for revolution

corevo

collaboration + revolution

http://www.ntt.co.jp/index_e.html

NTT

corevo™: Four AI directions set by NTT

Innovative R&D by NTT

Media

Contact Center and office support

Monitors human-generated information and understands human intentions and emotions

Daily activity assistance

Agent-AI

Senior Citizen Support and Care

Healthcare

Driving support

IoT

Analyzes people, things and the environment so that it can instantly make predictions and provide control

Avoid congestions in events

Ambient-AI

Disaster preventions and recovery

corevo

Sports

Heart-Touching-AI

Analyzes people's minds and bodies to understand their deep psyche, their intellect and their instincts

Human

Well-being

Interpersonal Relationship

Network-AI

Failure Predictoin

Connects multiple AIs to optimize the entire social system

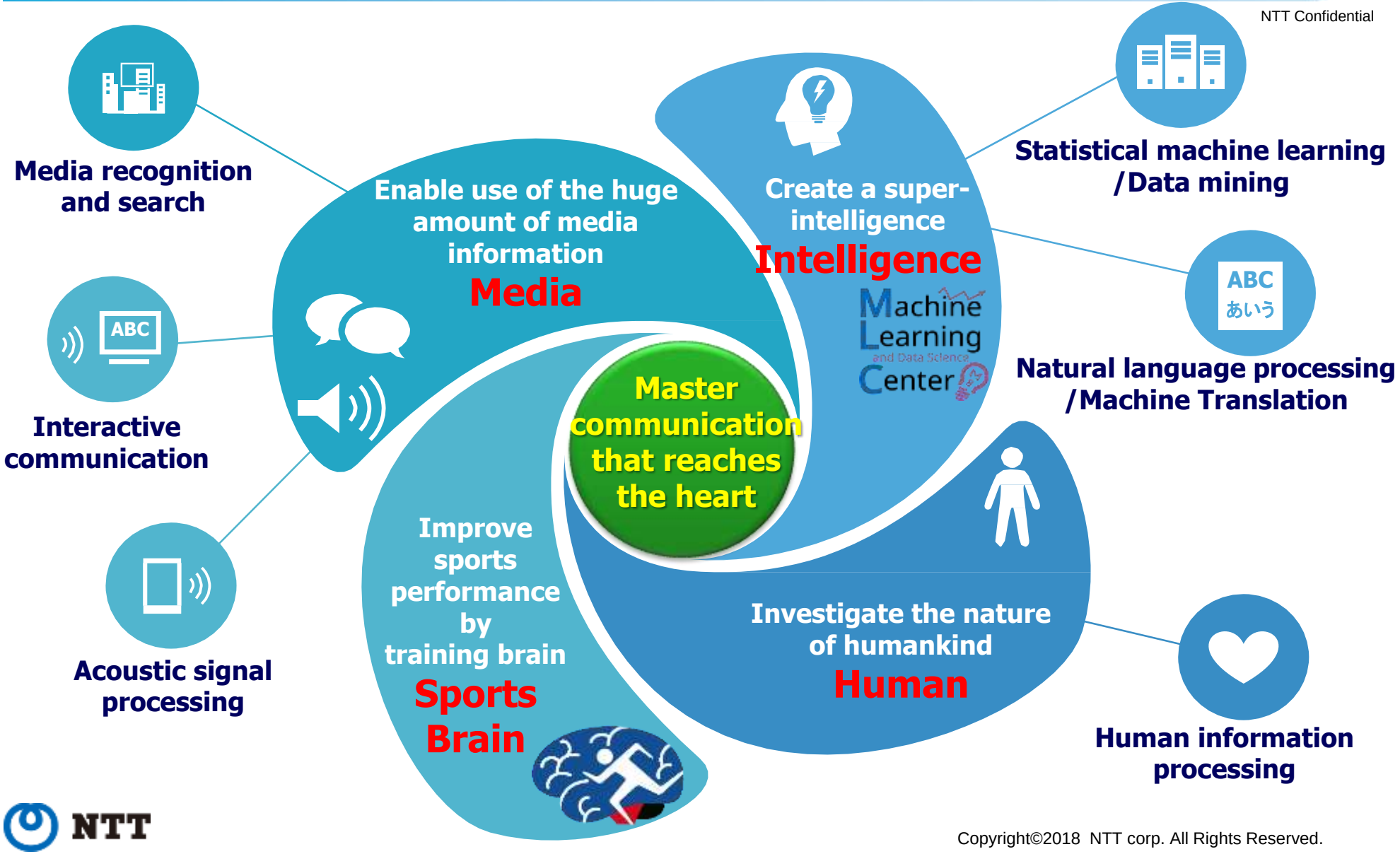
Failure free Network

Network

Global-scale Total Optimization

NTT

NTT CS Labs - Mission and Research domains



Media generation by Deep Learning

Can AI Make a Masterpiece of Painting?

Content

Artistic Style

Vincent van Gogh

Georges Seurat

combination

L. A. Gatys et al., “A Neural Algorithm of Artistic Style,” arXiv 2015
Code: https://github.com/mattya/chainer-gogh

Not Necessarily a Master Piece…

+

=

+

=

+

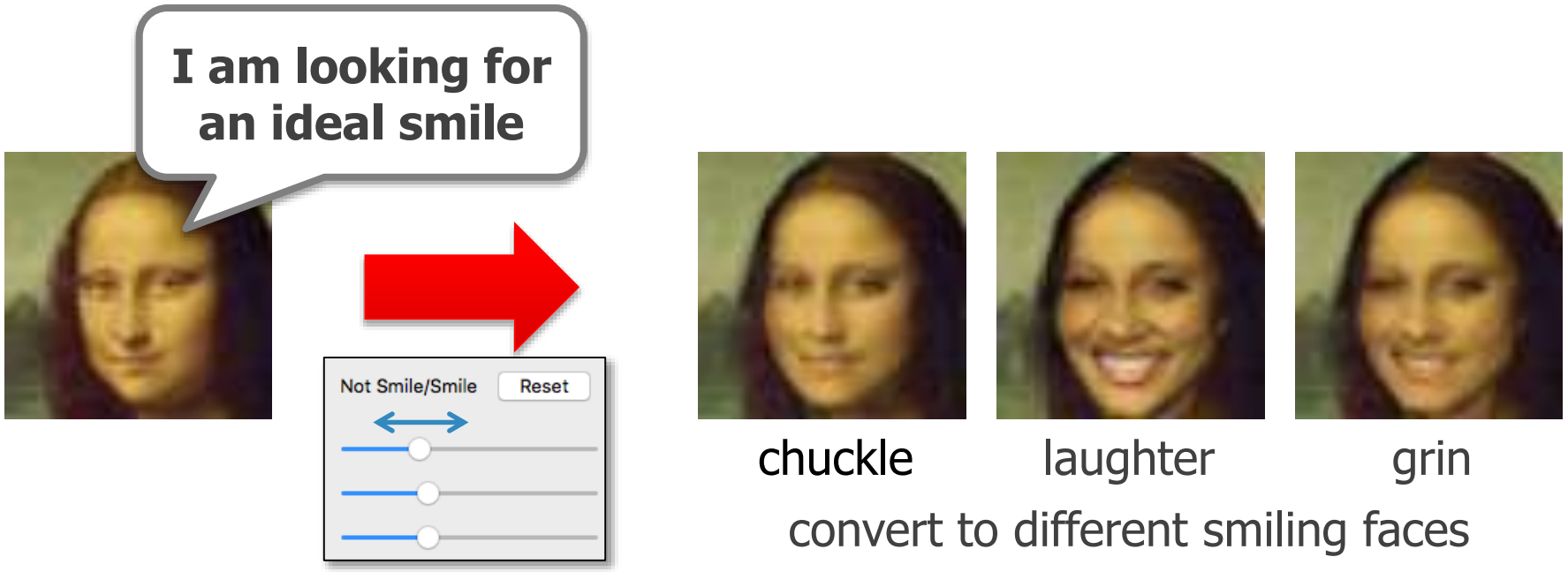
=

In Search of “Ideal Smile”

Deep Attribute Controller

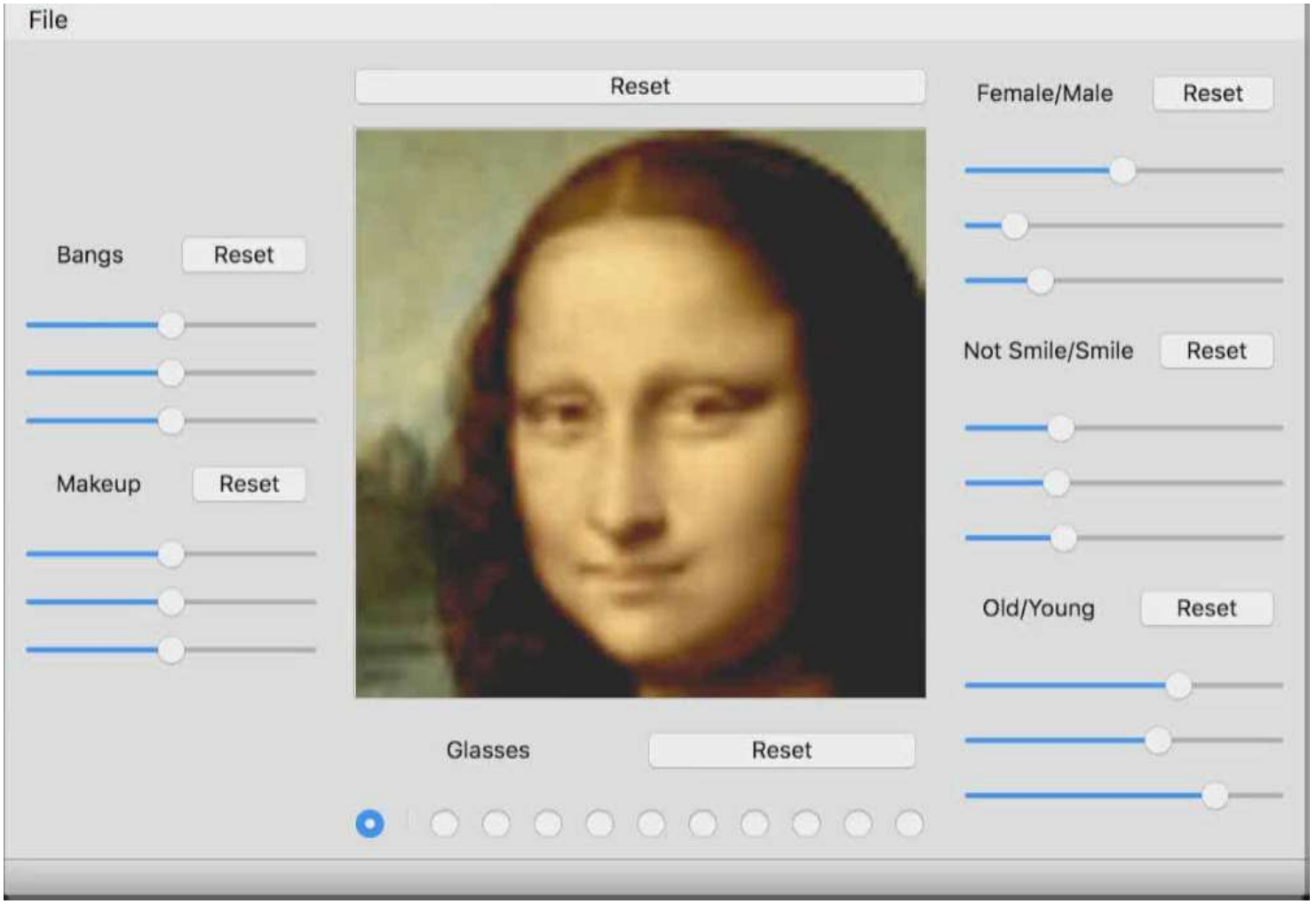
T. Kaneko et al., “Generative Attribute Controller with Conditional Filtered Generative Adversarial Networks,” CVPR 2017

- Multiple attributes can be freely given by interactive operation of image editing sense.
- Smiling faces can be automatically annotated with their styles.
- We can search for a face with a particular smile style in a DB



Deep Attribute Controller

Demonstration of Deep Attribute Controller



http://www.kecl.ntt.co.jp/people/kaneko.takuhiro/projects/gpa/index.html

Ears of AI：Speech Recognition for Agent-AI

Speech recognition has evolved significantly over the past 20 years

2000 2010 2015

technologies

Words/ Commands → **Clear sentences** → **Casual daily conversation**

applications

watch style PHS

TV program subtitles

announcer

weather⇒177
time⇒117

Voice mining at call center

operator customer

natural conversation

Parliament Minutes creation support system

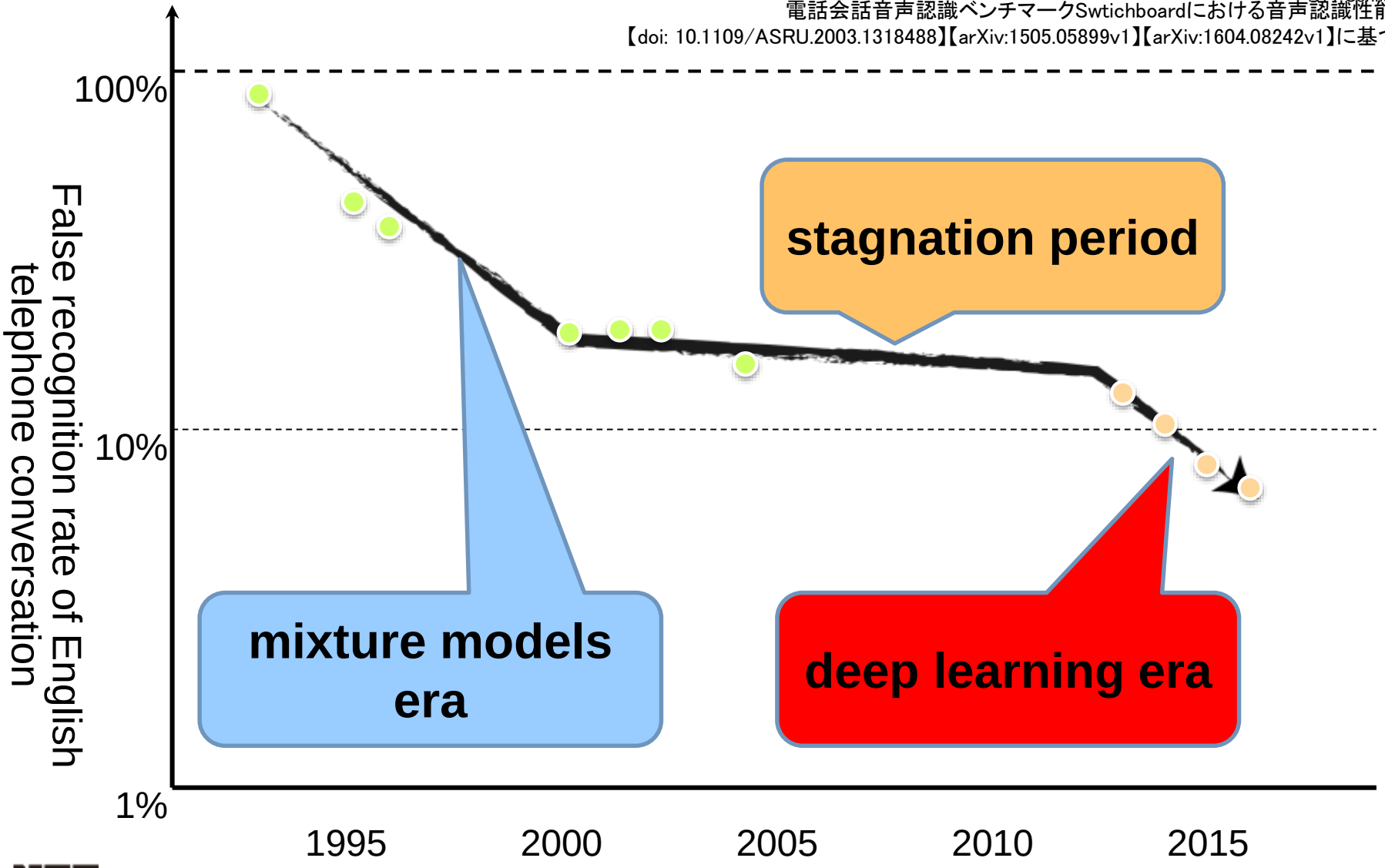
speech looking at notes

conversation with robots

NTT

Evolution of Automatic Speech Recognition

電話会話音声認識ベンチマークSwtichboardにおける音声認識性能の推移
【doi: 10.1109/ASRU.2003.1318488】【arXiv:1505.05899v1】【arXiv:1604.08242v1】に基づき作成



Voice Interface in the Near Future

The scope of speech recognition is being expanded

Current

Near Future

One person speaks after a cue.

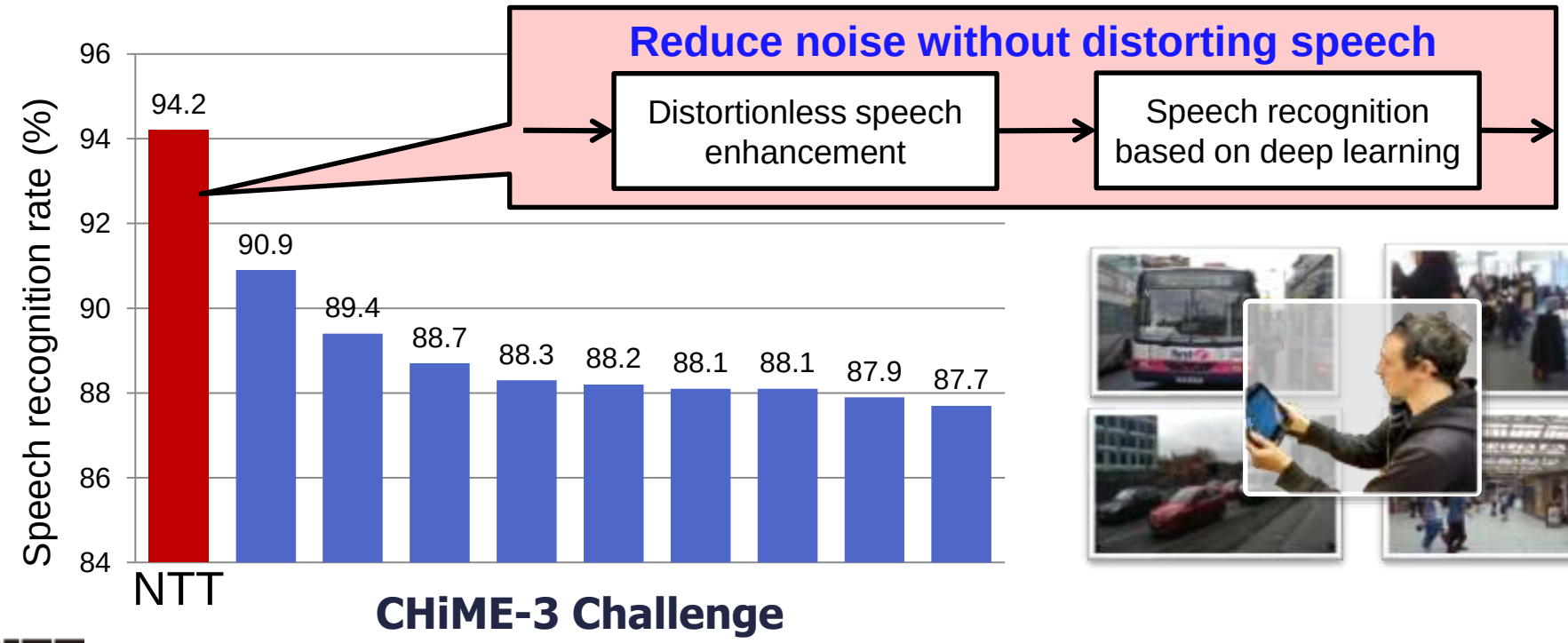
Ex. voice search with a smart phone and an AI speaker.

Multiple people speak freely.

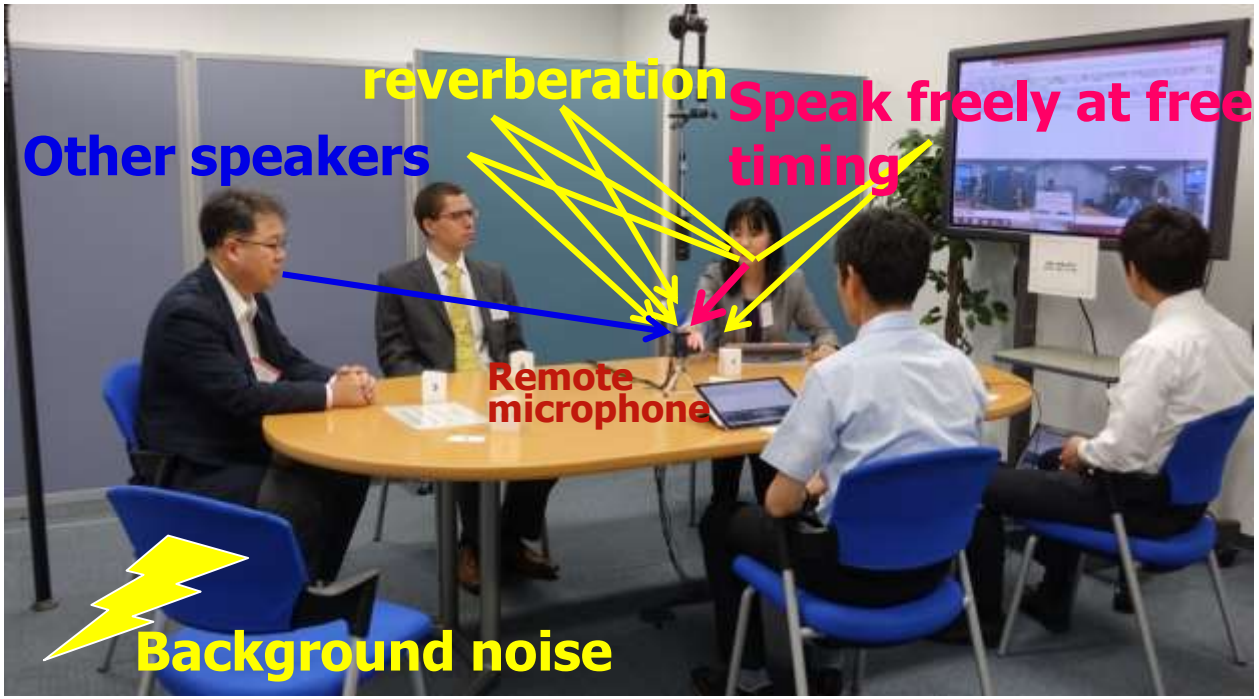
Ex. conversation with robots, group meeting archiving,…

Speech Recognition in Noisy Environments

- **NTT achieved top performance among 25 participating organizations in CHiME-3, an international technical evaluation of speech recognition in various noisy environments: bus, cafe, street, and pedestrian areas)**

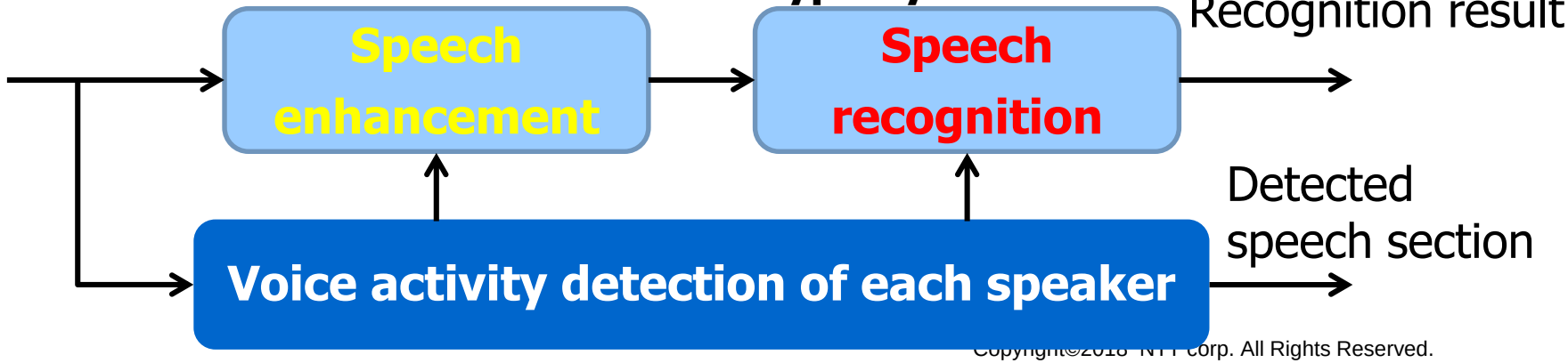


Speech Recognition When Multiple People Speak Freely



Recorded speech (Multi-microphone signal)

Real-time Prototype System



NTT

Demonstration of Meeting Archiving

Demonstration of Meeting Archiving

1：はい[え一]皆さん本日はお集まりいただきましてありがとうございます[えーっと]こちらでは、[えーっと][あの]音声認識システムの援助さしていただいております、で、[えっと]今は、[あの一]ちょうど[ま]ミーティングのようなシチュエーションでの音声認識ということで、やらさしていただいておりまして、で[あの]外から、[あの]かなり、[あの一]、ノイズが聞こえてきて、と思うんですけれども、[ま]こういったをような状況でも[あの一]音声認識ができるようなデモンストレーションになっております、で[えーっと][ま]私が一番の席に座っておりまして、[えー]、2番3番4番5番ちょっと今だけてますけど6番っていう形、、[えー]席仲間なってますので、[ま]、[あの]是非ちょっと皆さん一言ずつでもしゃべっていただければなと思うんですけれども。

6：[えー]。

2：2番です入ってますか。

3：3番ですちょっと質問はあるんですけども、でこれは、話者の数話者の数で喋ってる人の数っていうのは、事前に分かっていて、その下でやってるんですか。

1：[えっと]これは荒木さんじゃあ説明して下さい。

6：[あ]はい分かりました[えーと]これは[あの]最大の人数は今システムはしってるんですけども[えっと一]実際はですにここににたデモシステム実は知りません、[えーっと一]それぞれ、高校使ってるんですけども最初の方向から声が来たらそれをきれいに[えーと]音声強調するということをしています、ちなみにここは6番席でした。

5：4番席ですけれども。

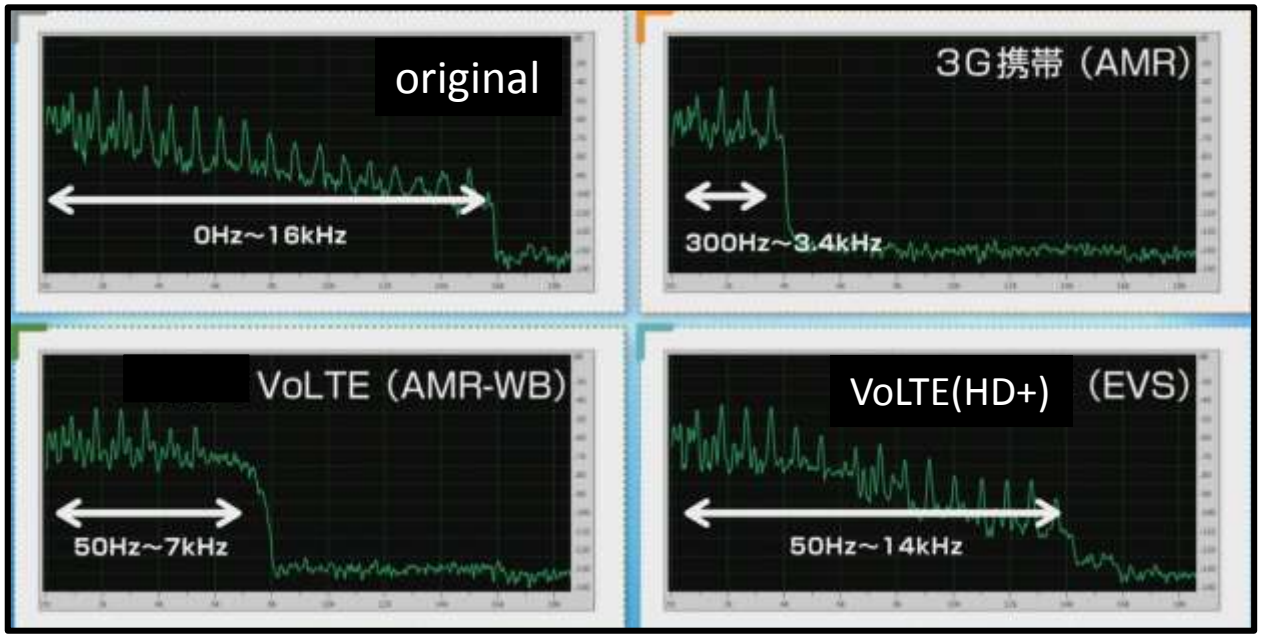
4：これは。

5：これは今マイク[ま]何使ってるんですか。

6：うん一[えっと]マイクは今こちらの[えっと]8つ8個を使っています。

EVS: Enhanced Voice Services Codec for LTE

- Next Gen 3GPP Speech Coding for Improved User Experience in Telephony
- More **natural sounding speech** and **improved music** quality
- Result of global 12 party collaboration including NTT and NTT docomo
- **NTT docomo** launched **VoLTE(HD+)** in May 2016



.....

Analyzing and Converting Speech Prosody

In voice communication, non-verbal information such as intonation and accent is as important as textual information.

- Intonation: pitch changes in a sentence

  「そうですか」Is that so? （納得⇔疑い ↗ ）
  agree doubt

  「これじゃない」（断定調⇔同意を求める ↗ ）
  This is not, isn't it? definitive requesting agreement

- Accent: pitch changes in a word

  「 はし 」と「 はし 」「 おいおい 」と「 おいおい 」


chopstick

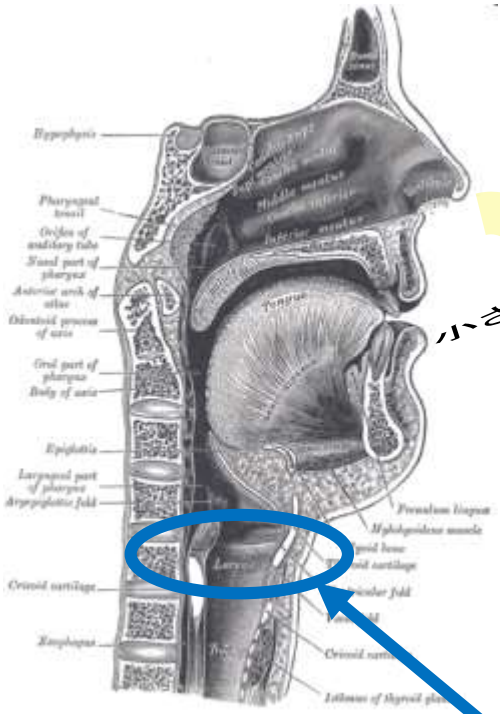

bridge

by and by

Hey, you!

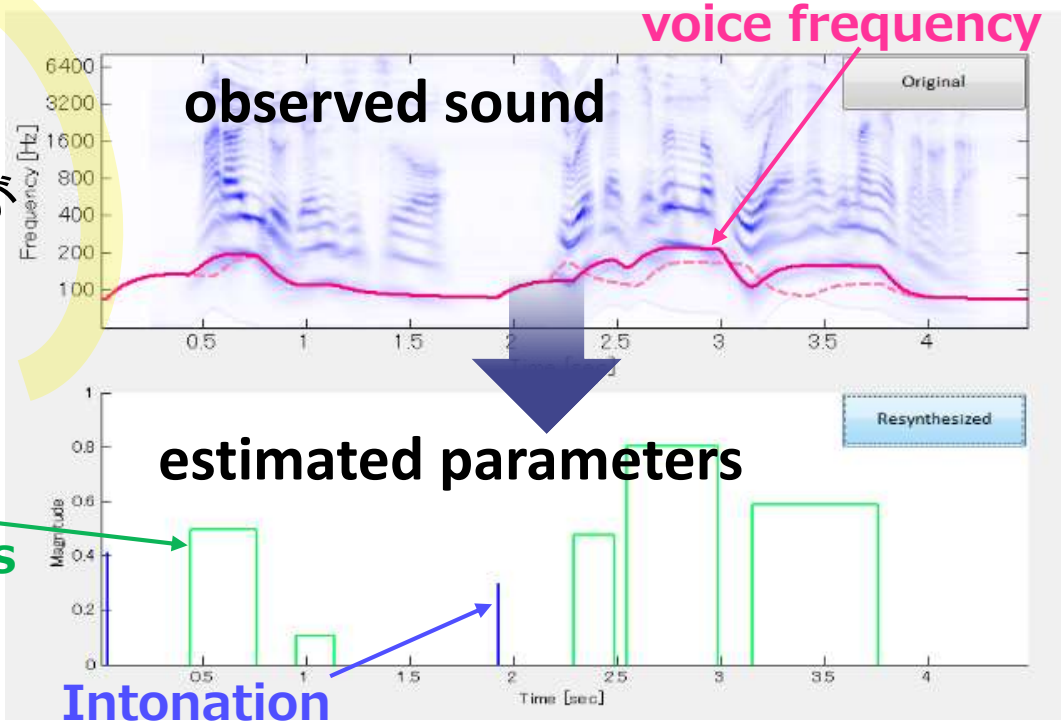
Modeling Voice Frequency Contours

- Voice pitch is controlled by forces pulling the vocal fold.
- The model was known but parameter estimation was hard.
- NTT succeeded in estimating the parameters from the uttered sounds.



voice utterance

ちいさな うなぎ屋に
熱気のようなものが
みなぎる



Accent parameters

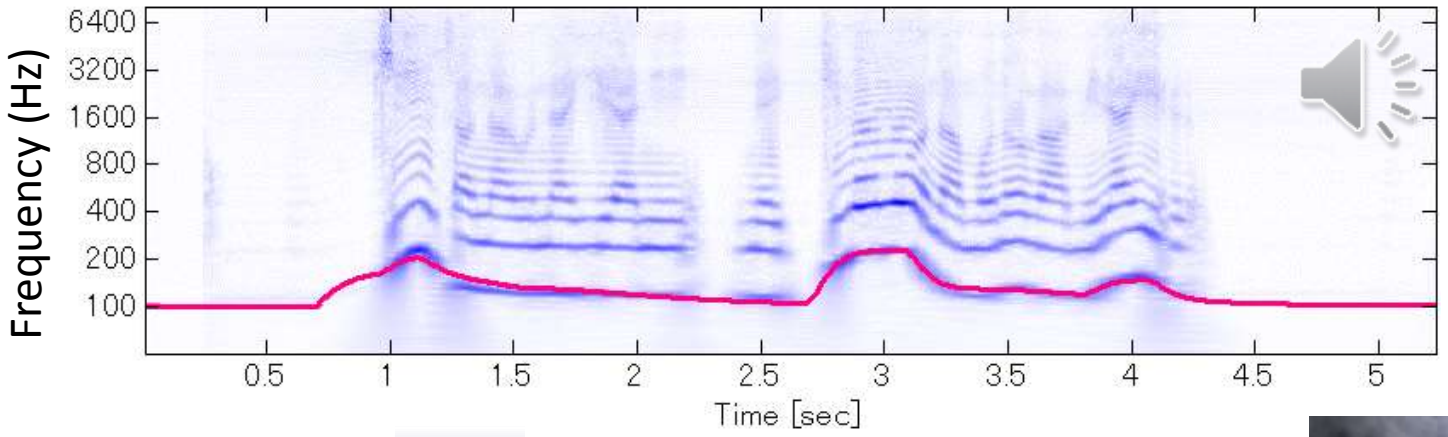
Intonation parameters

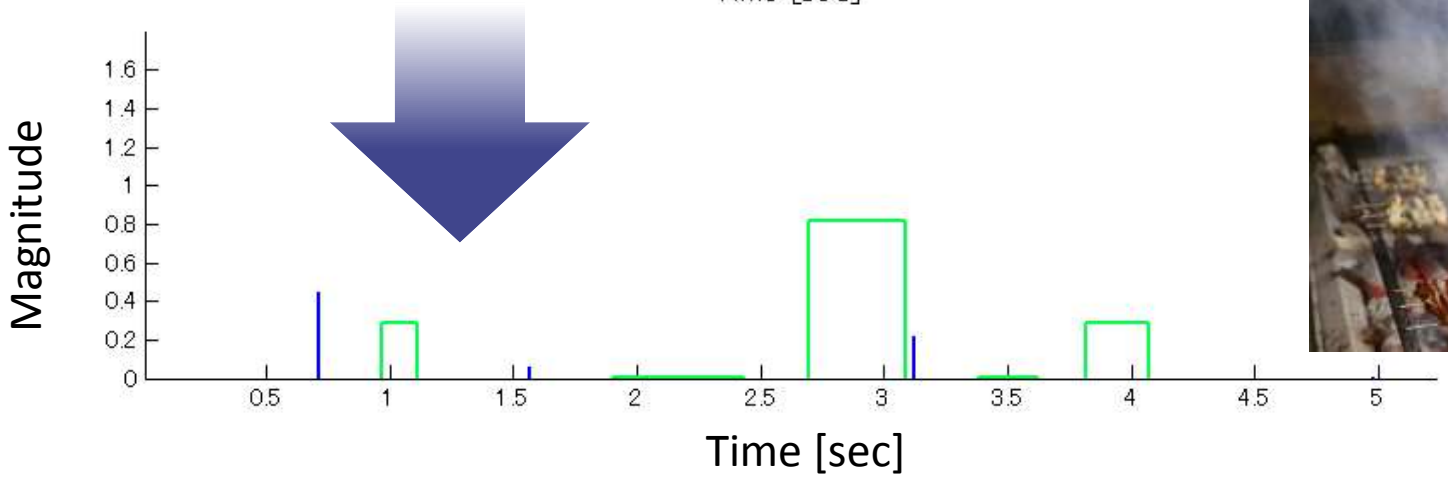
vocal fold

Example (1): Original input sequence

In a small eel restaurant
「小さな　うなぎ屋に

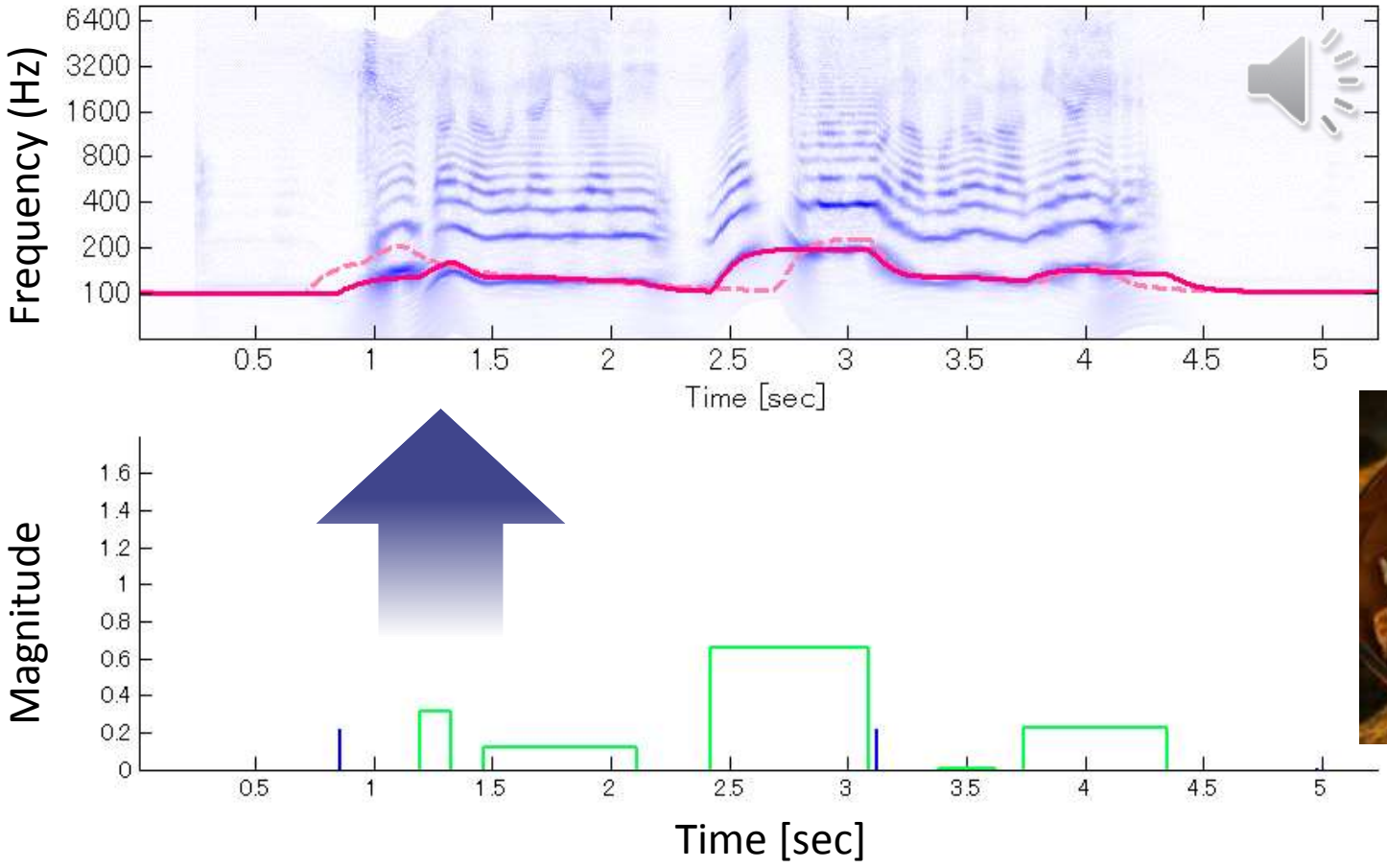
something like fever suffuses
熱気のようなものがみなぎる」





Example (2): Change accent timings

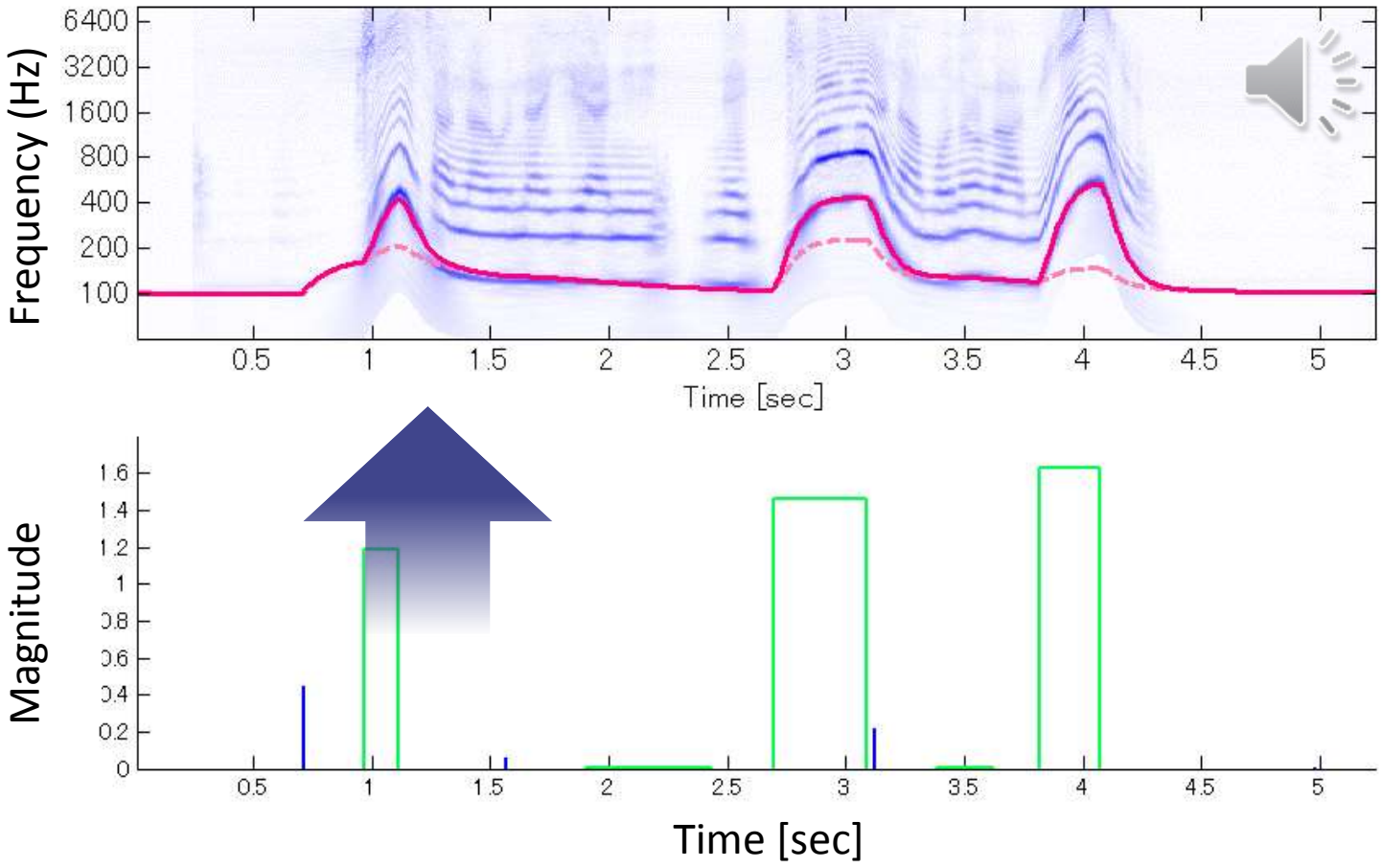
「小さな　うなぎ屋に　　熱気のようなものがみなぎる」



関西風

Example (3): Change accent strengths

「小さな　うなぎ屋に　　熱気のようなものがみなぎる」



Can a Robot Pass a University Entrance Exam?

- NTT joined AI grand challenge project: Can a robot pass a university entrance exam? as English Exam team (The project is hosted by National Institute of Informatics).

問 2

Parker: I hear your father is in hospital.

Brown: Yes, and he has to have an operation next week.

Parker: [19] Let me know if I can do anything.

Brown: Thanks a lot.

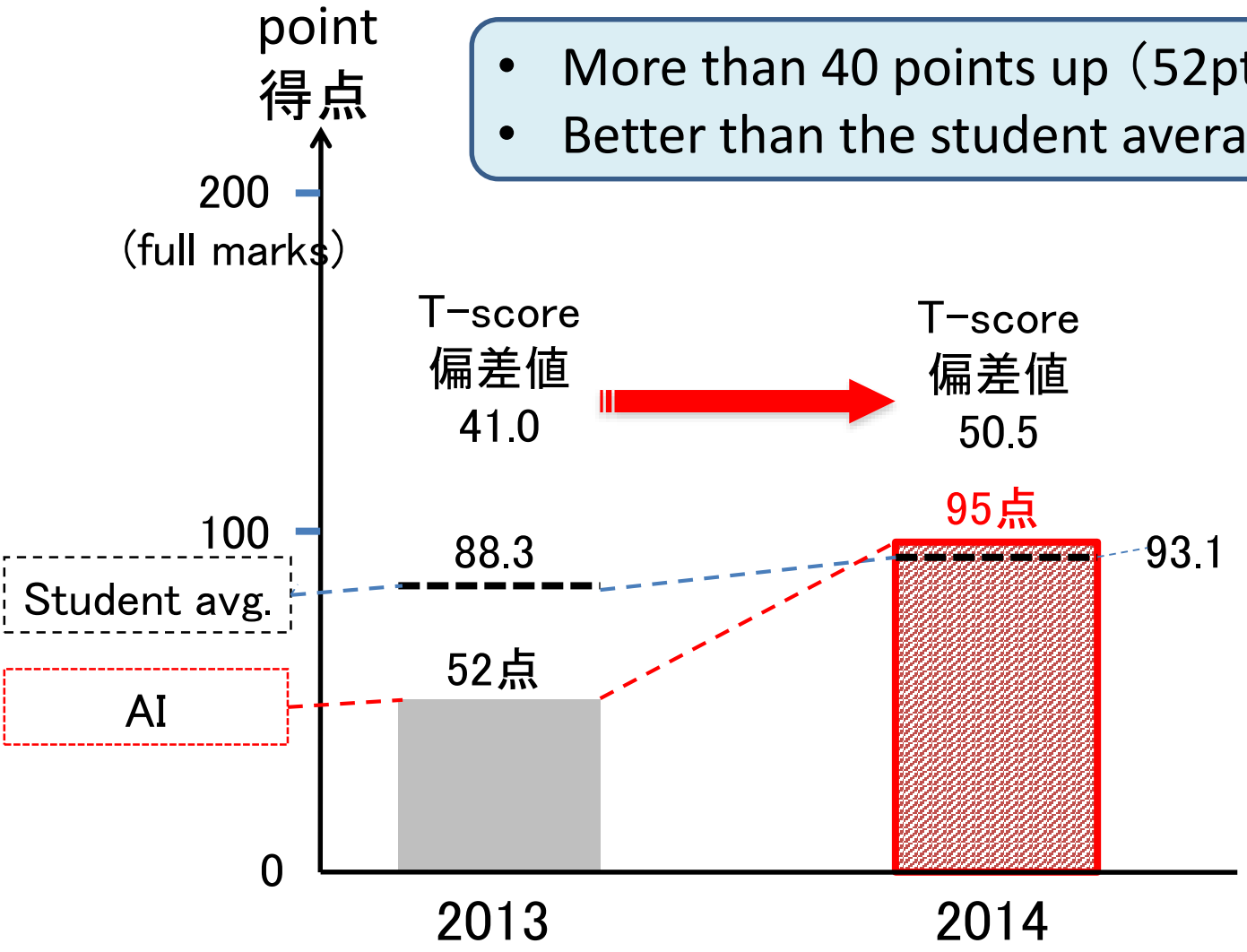
① Exactly, yes.
② No problem.
③ That's a relief.
④ That's too bad.

All are grammatically correct. Commonsense knowledge can select the right answer as 4.

https://www.youtube.com/watch?v=-Vbzp4RAmhY

The Result of English Mock Exam Test in 2014

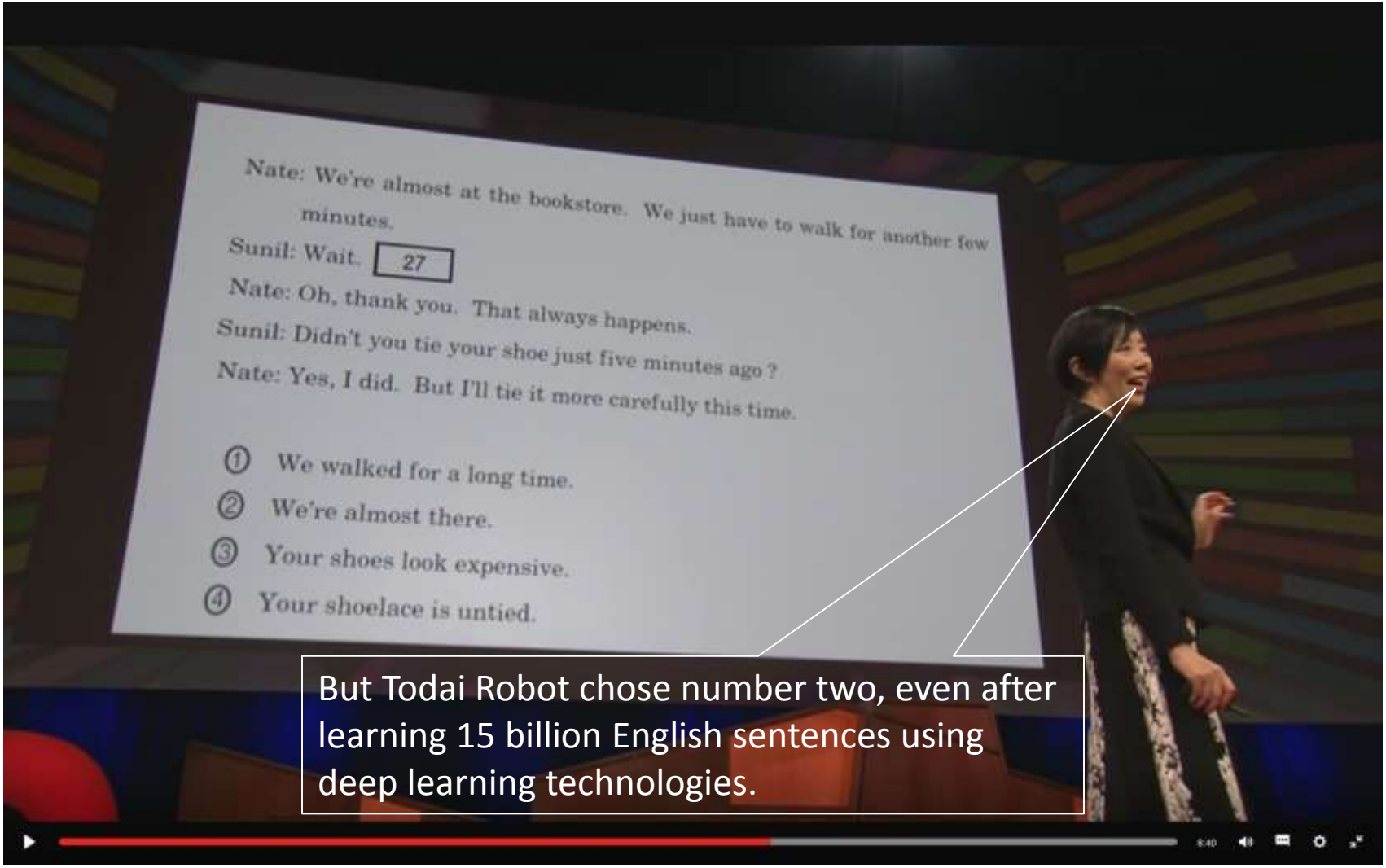
- More than 40 points up（52pt⇒95pt）
- Better than the student average for the first time



- ◆ High score in both years
  - pronunciations
- ◆ Improved over the last year
  - dialogue completion
  - summary understanding
  - word sense induction
  - word order correction
- ◆ Still difficult
  - long reading comprehension
  - explaining photos and pictures

TED: Can a robot pass a university entrance exam? by Noriko Arai

presented at an official TED conference



https://www.ted.com/talks/noriko_arai_can_a_robot_pass_a_university_entrance_exam

Voice of Students

So surprising!

Great and I Feel envious

Feel frustrated, as I work so hard

Many students do not even understand the problem statements.

We must train students.

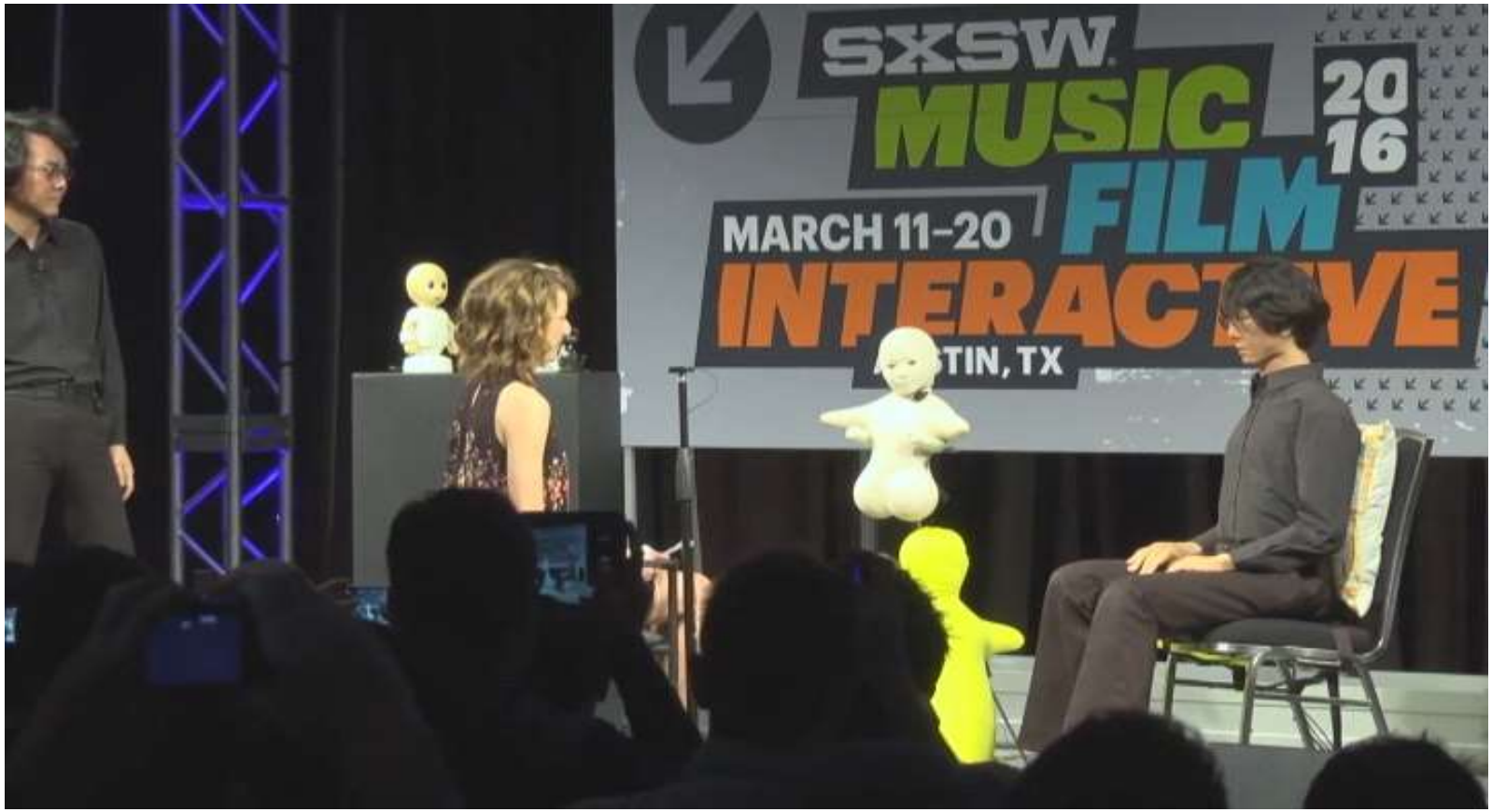
Reading Skill Test

Casual Conversation with Robot

大阪大学 OSAKA UNIVERSITY × NTT

Joint research with Ishiguro Lab. @ Osaka University

SXSW2016



From Casual Conversation to Serious Debate

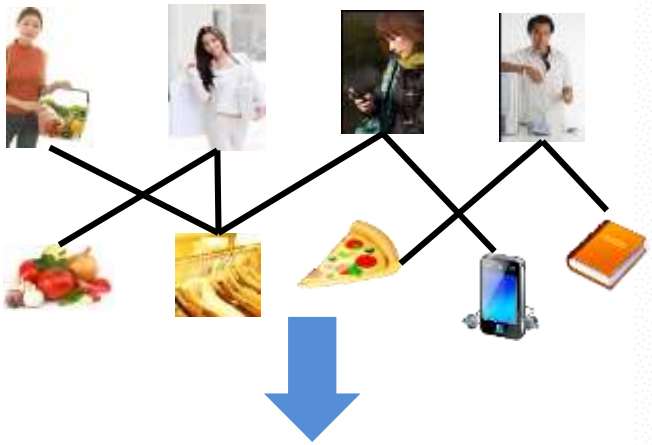
robots

SXSW2017

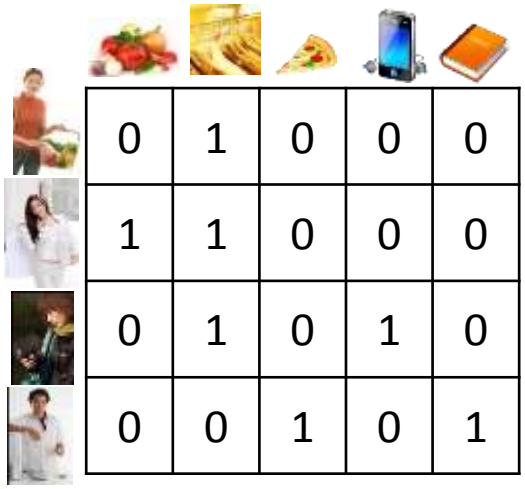
Joint research with Ishiguro Lab. @ Osaka University

Finding Patterns from Human Behavior Data

Ex) Purchase Log Data

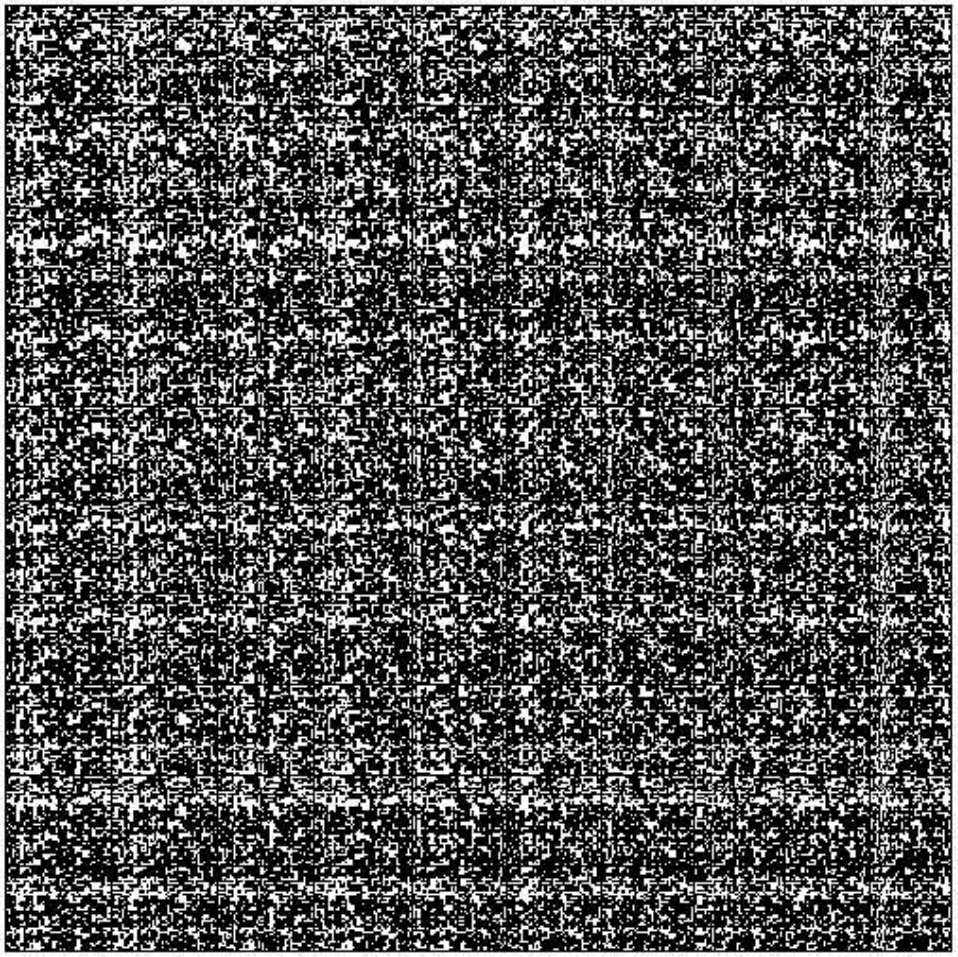


matrix representation



0	1	0	0	0
1	1	0	0	0
0	1	0	1	0
0	0	1	0	1

patterns



Discovering Patterns by Matrix Factorization

- Human behavior data exhibit certain tendencies and patterns.
- NTT developed NMTF* that can efficiently extract characteristic and (possibly) intersectional patterns from such complicated relational data.

Input data

User location data:
time, latitude, longitude,…

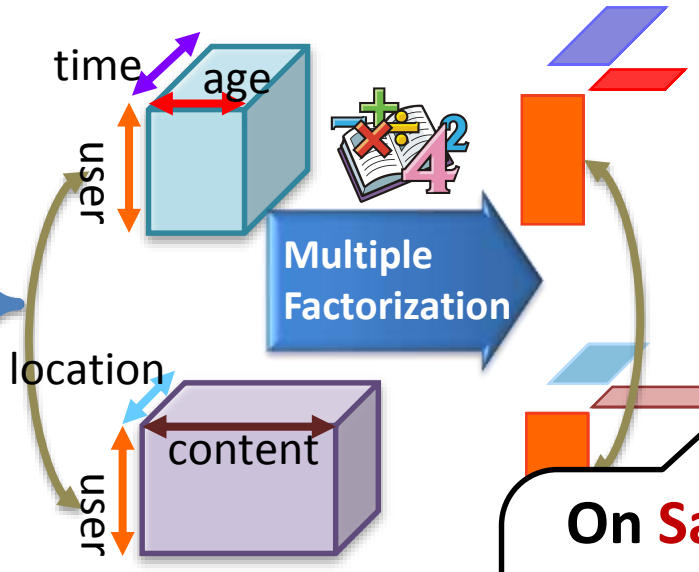
Shop data:
shop location and category,…

User attribute data:
sex, age, living area…

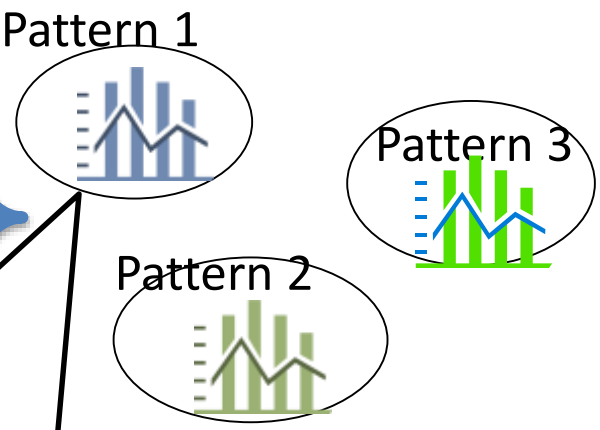
Content browsing history

Climate data:
weather, temperature, humidity, …

Apply NMTF



Discovered patterns



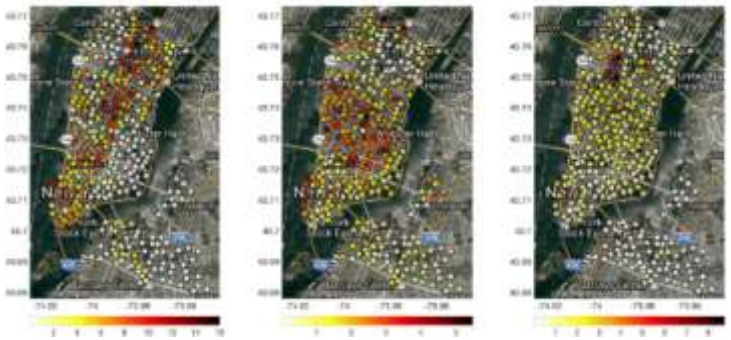
On Saturday and Sunday during the day, many women in their thirties living in the east district visit cafe in the west district

*NMTF=Non-negative Multiple Tensor Factorization

Extensions to Matrix Factorization Analysis

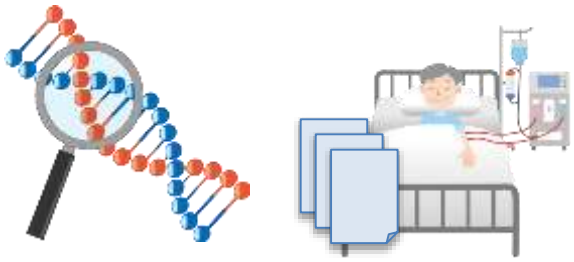
Spatio-temporal extension

Graph-regularized tensor factorization

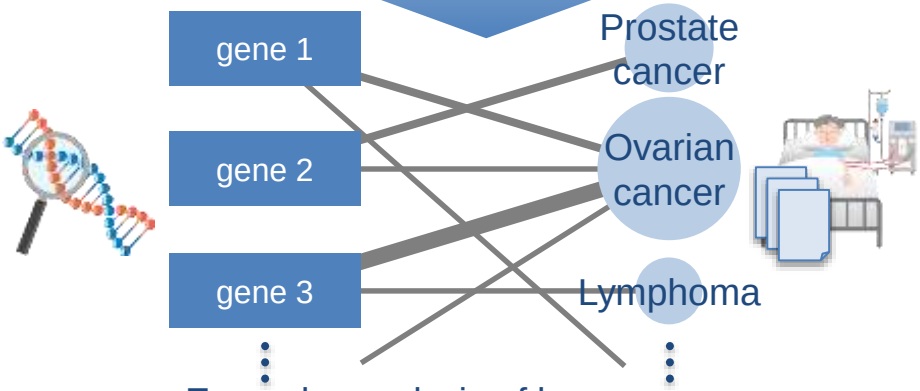


Example: prediction of rental bicycle use in New York

higher-order extension



Higher-order factorization machines



Example: analysis of how genes are related to diseases

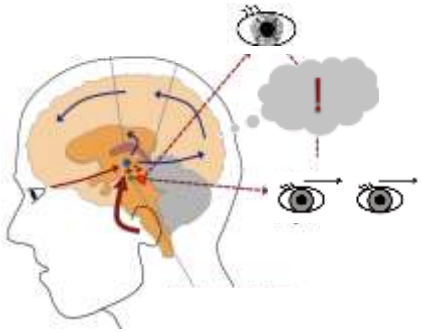
From Agent-AI to Heart-Touching-AI

◆**Agent-AI**

understands your intentions and emotions and behave like a human

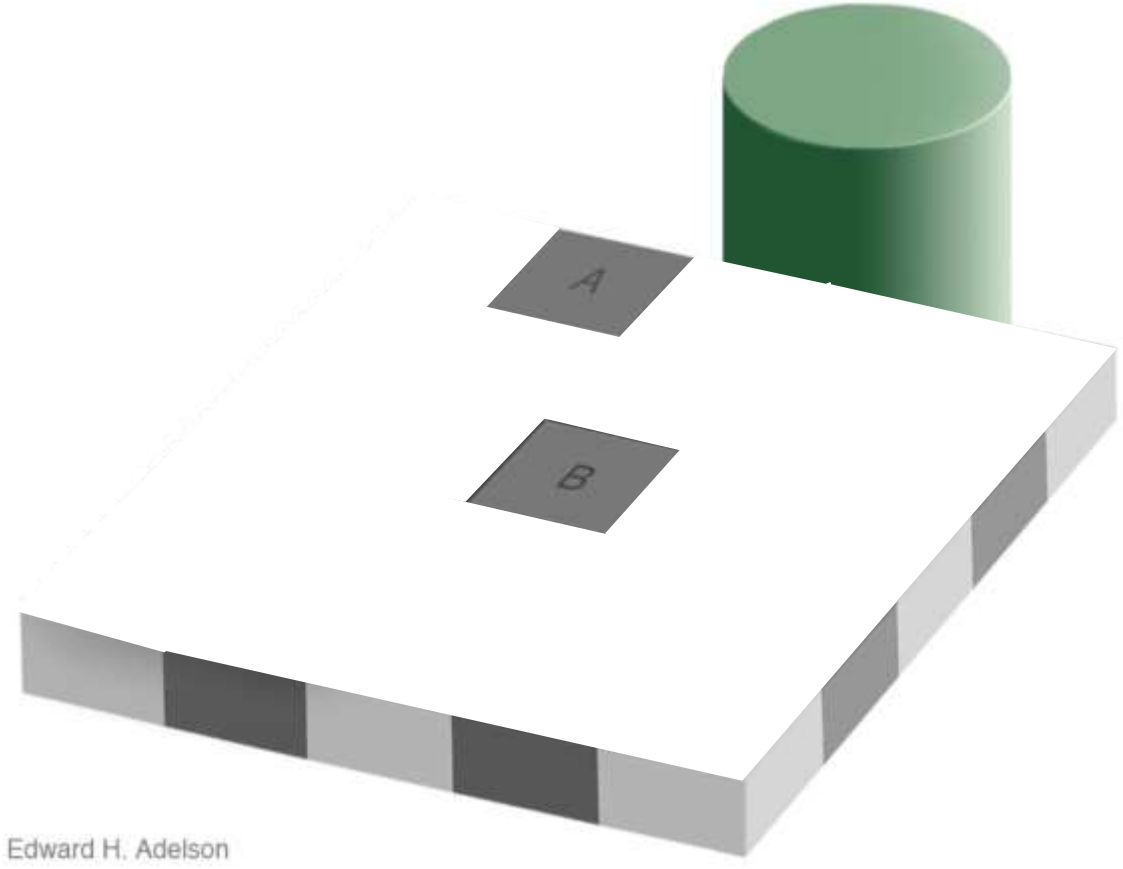
◆**Heart-Touching-AI**

understands your psychological, subconscious and instinctual states and appeals directly to your heart



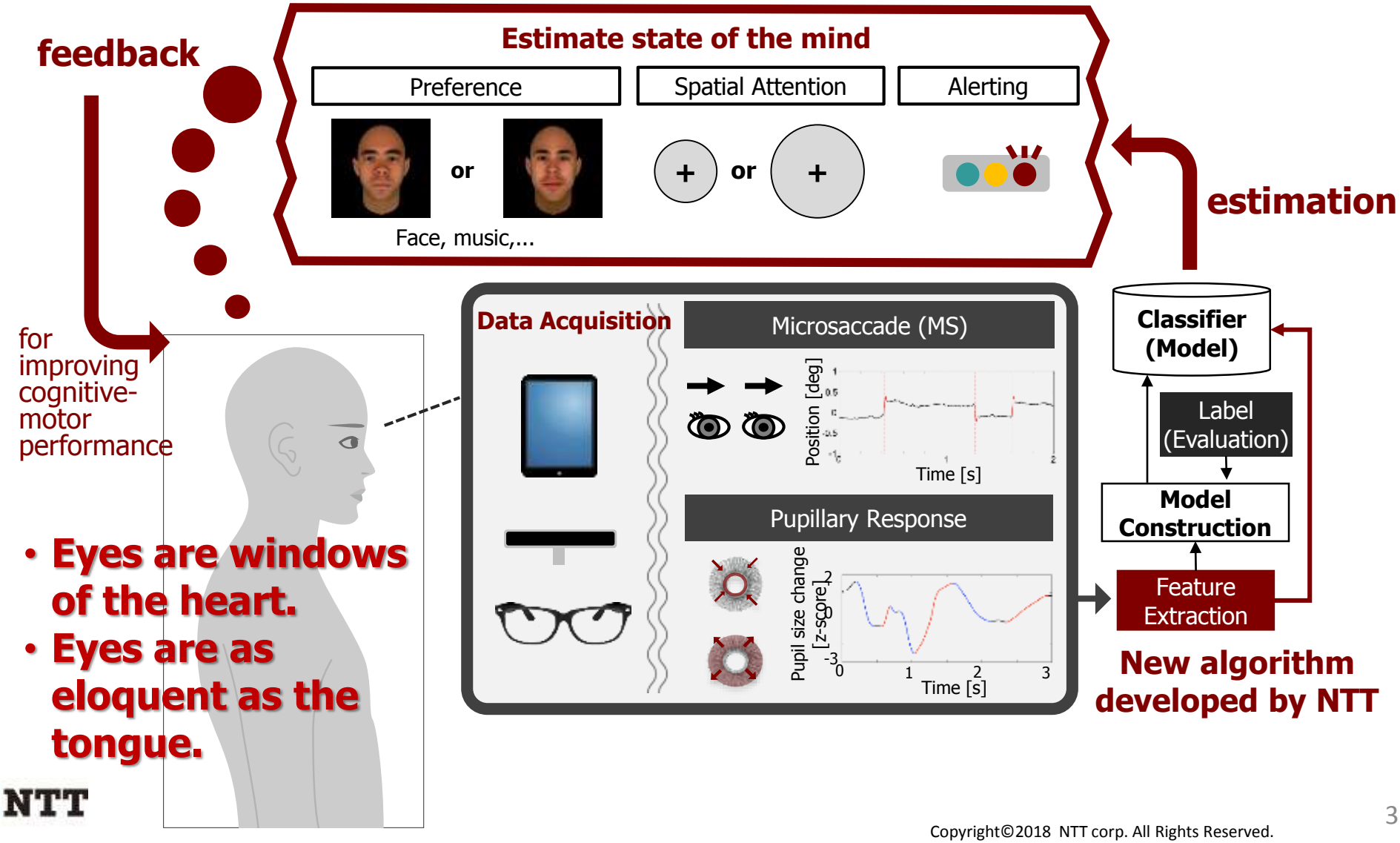
NTT

What you perceive is not what it is



This effect demonstrates the success rather than the failure of the visual system.

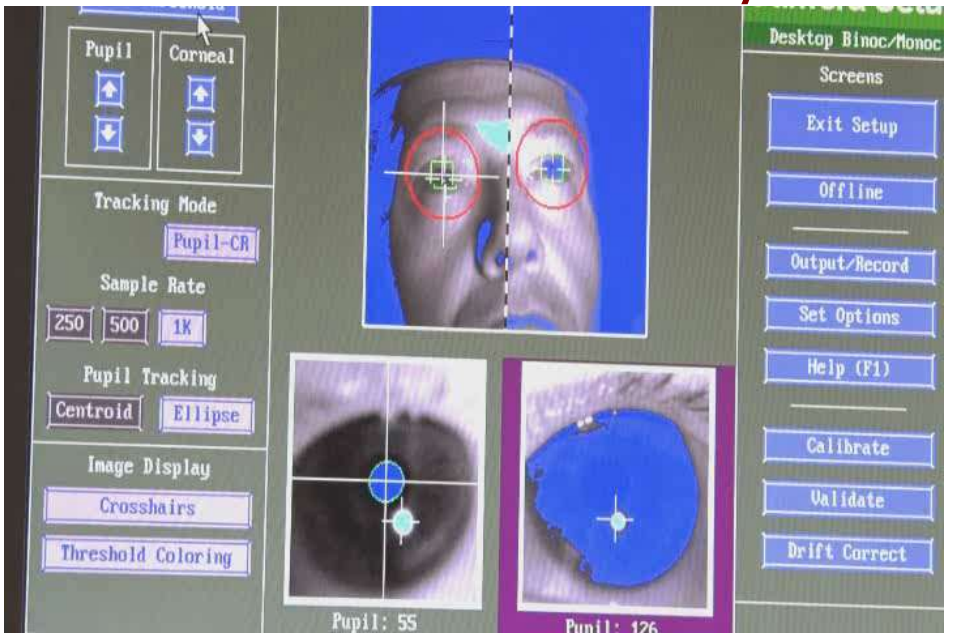
Reading Mind from Unconscious Eye Movements



- **Eyes are windows of the heart.**
- **Eyes are as eloquent as the tongue.**

Mind Reading Overview

81 features are analyzed



estimation accuracy：88.9%

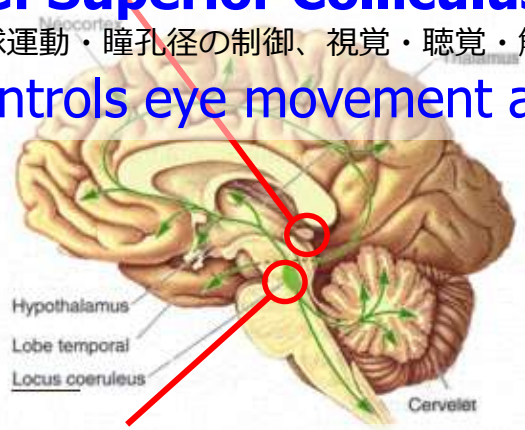
Mechanism of Mind Reading

Aston-Jones, G., & Cohen, J. D. (2005). Annu Rev Neurosci.

SC: Superior Colliculus (上丘)

眼球運動・瞳孔径の制御、視覚・聴覚・触覚入力

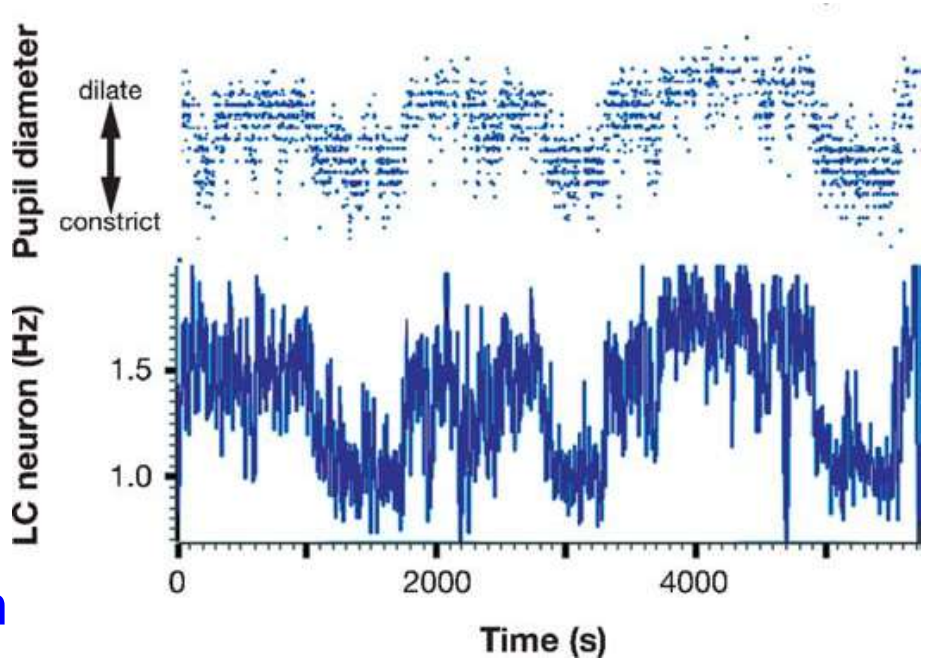
controls eye movement and pupil diameter

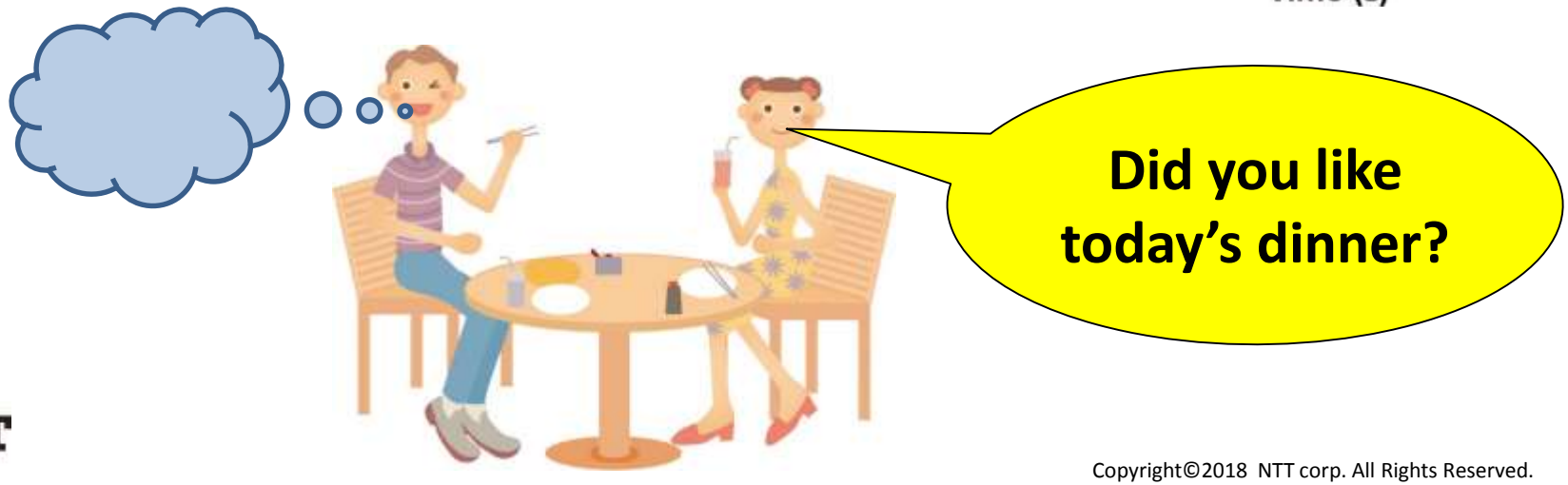


LC: Locus Coeruleus (青斑核)

覚醒レベルの制御、選択的注意などに関与

controls arousal level and selective attention





Tactile/haptic Sense Opens New Information Channels

Palpable Intelligence (触知性):
tactile/haptic sense conveys deep information rooted in the body

Intelligence of tactile sense that generates information

69th Mainichi Bunka Award
（Natural Science Section）
November, 2015

Junji Watanabe

Buru-Navi: Gives You a Feeling of Being Pulled

- The device held in fingers creates the sense of being pulled.
- It makes use of the nonlinear characteristics of human perception and asymmetrically oscillating stimuli.

Buru-Navi 1 2007

10 years

Buru-Navi 4 2017

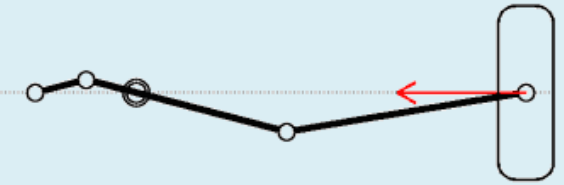
shell case type

cubic type

fingertip driven type

Sensory-motor mechanism in human:

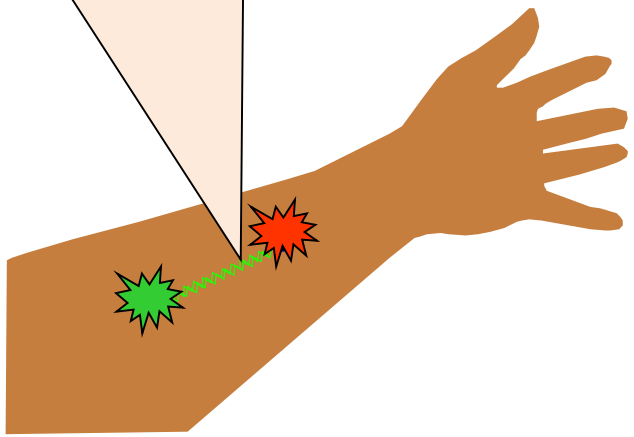
- strong & short stimulus is easy to feel
- weak & long stimulus is hard to feel

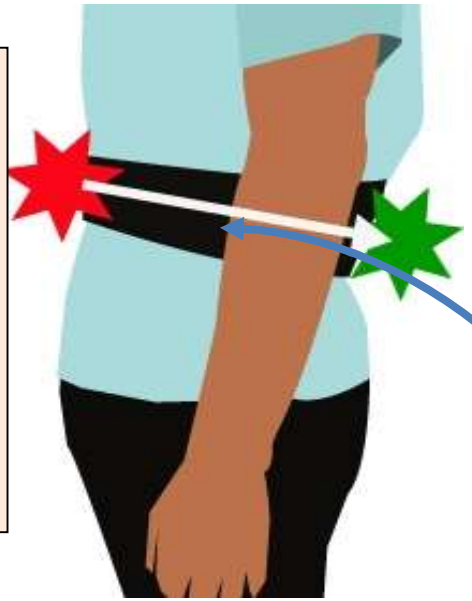


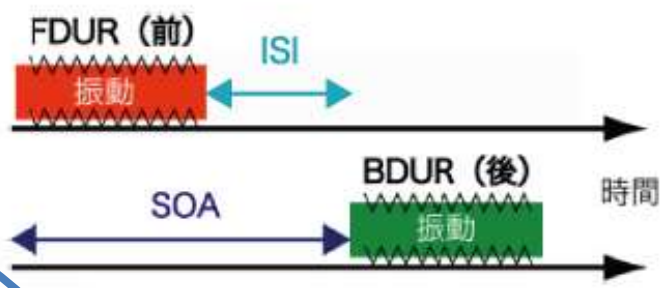
Sensory Illusion caused by Buru-Navi

“Through-and-Through” Sensation Interface

When two separate tactile stimuli with a certain time difference are presented, the impression of continuous movement between the two points is perceived.







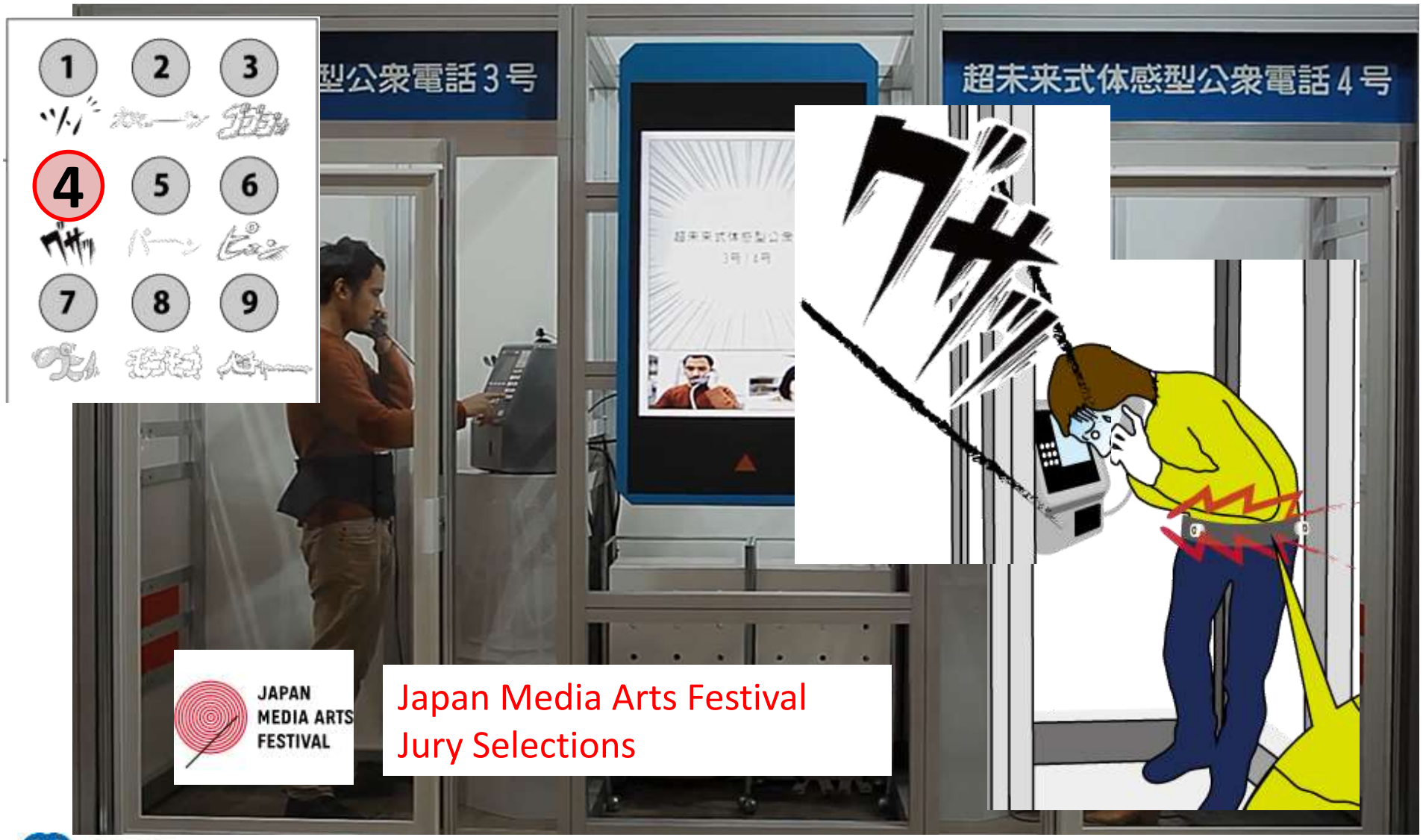
two speakers generate vibrations

A through-and-through sensation is perceived between stomach and back.

Demonstration

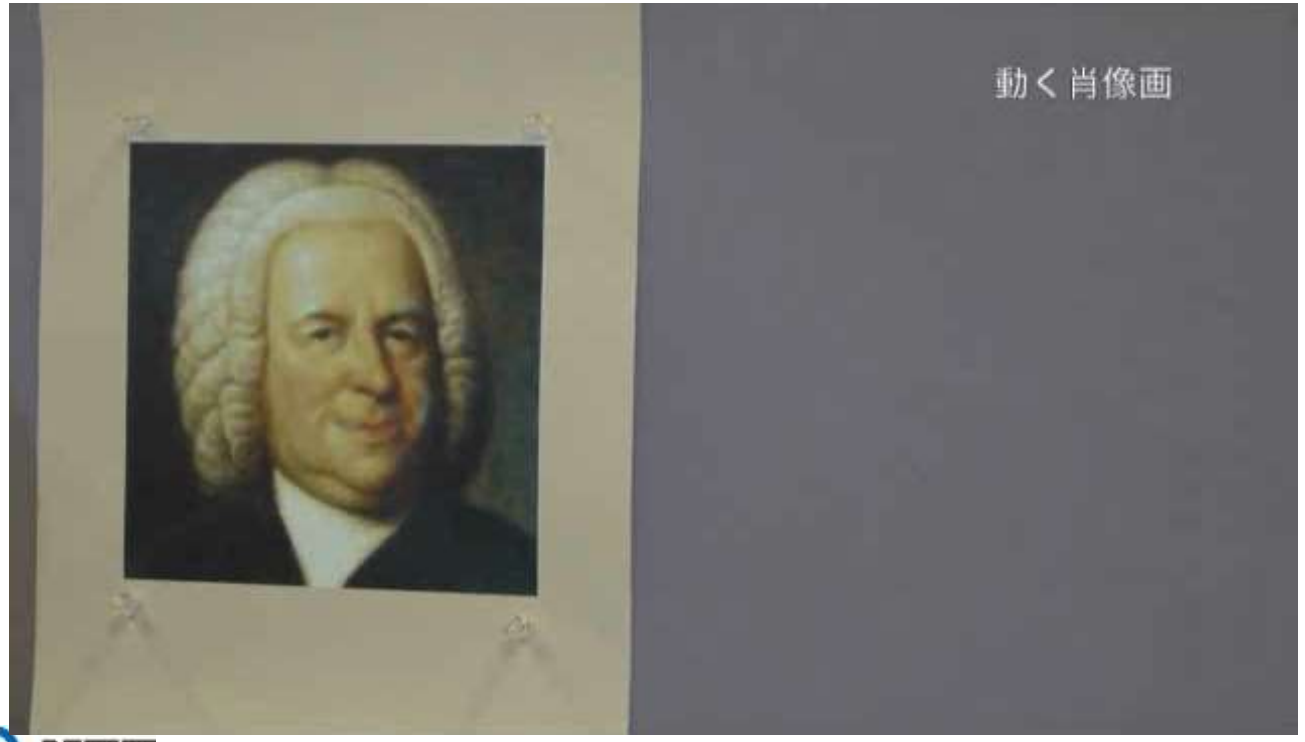
TBSがっちりマンデー！！
2015.07.19

Ultra-Future Experiential Public Phone



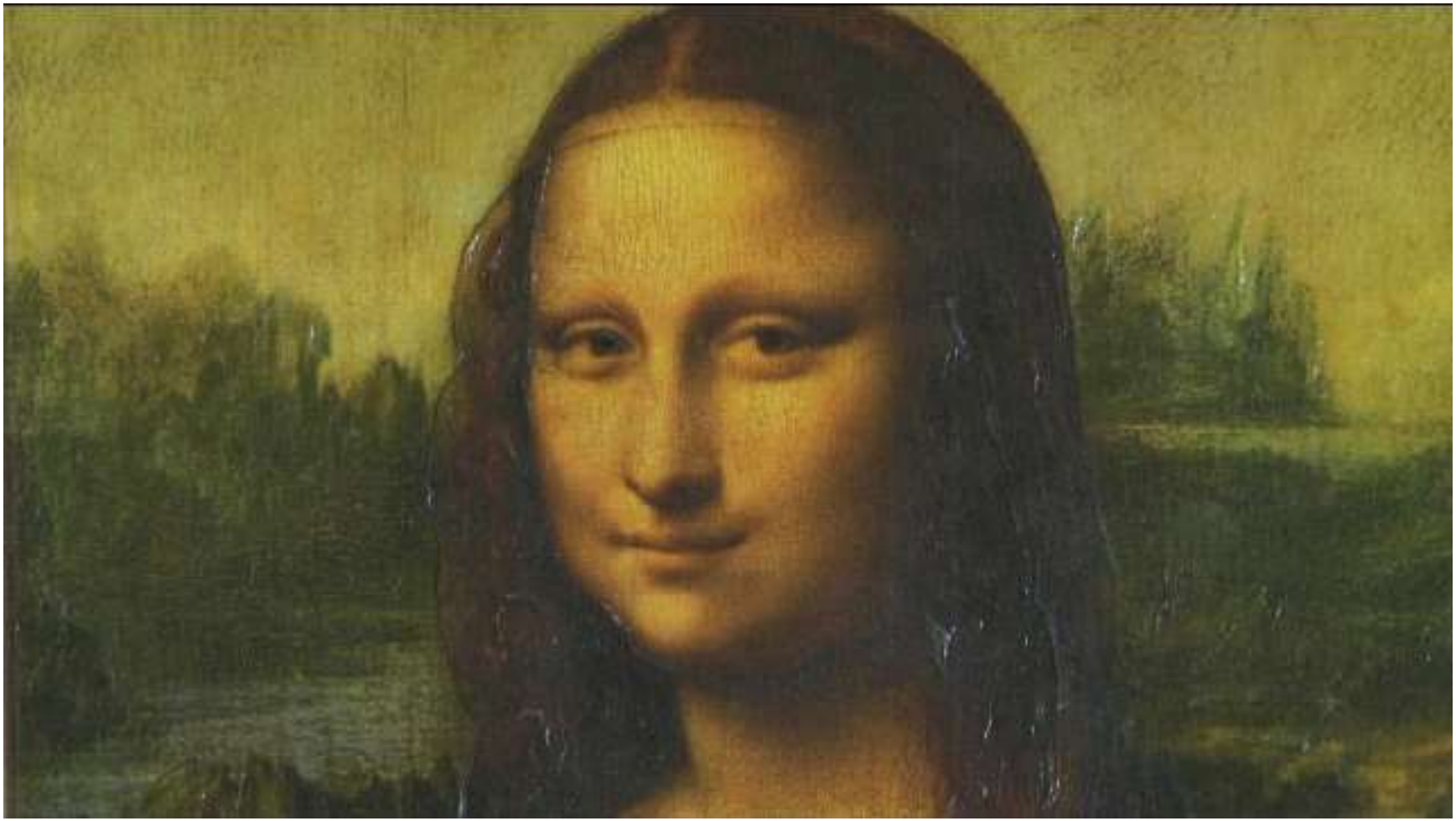
Hengento Projection

- A light projection technology that adds a variety of illusory, yet realistic motions to a static object
- It is applied to the Next-generation POP (point-of-purchase) signage expressing sizzling feelings (collaborated with DNP)



collaboration with DNP

変幻灯: Mona Lisa



変幻灯: Moving Brick Wall

Hacking Human Visual System

Color movie

color →
form → integrated → perceived
motion →

processed separately

inconsistencies are resolved in the brain

Hengento projection

projecting motion patterns

color →
form → integrated → perceived
motion →

perceived just like a normal color movie

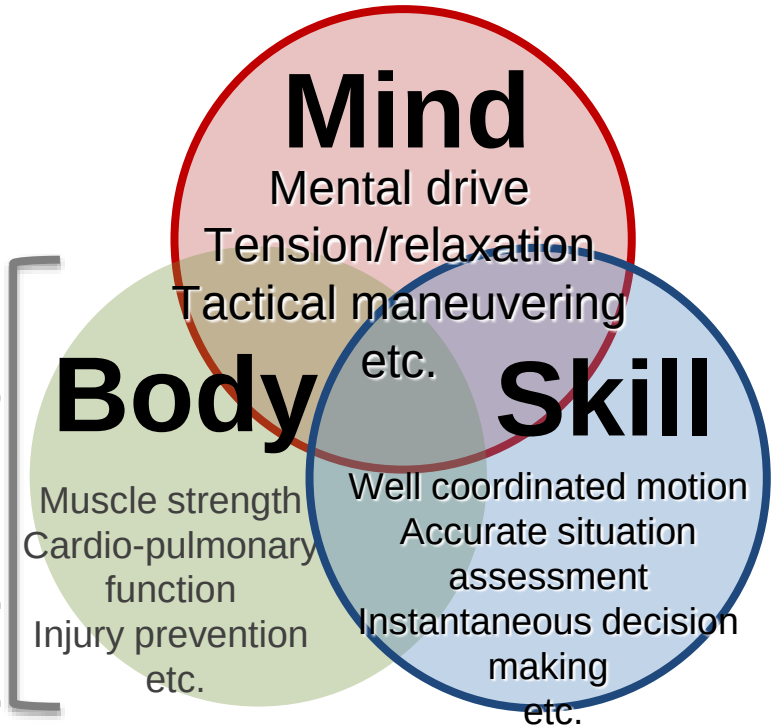
NTT

Curve ball is Created by the Eyes and the Brain

Shapiro et al. 2010

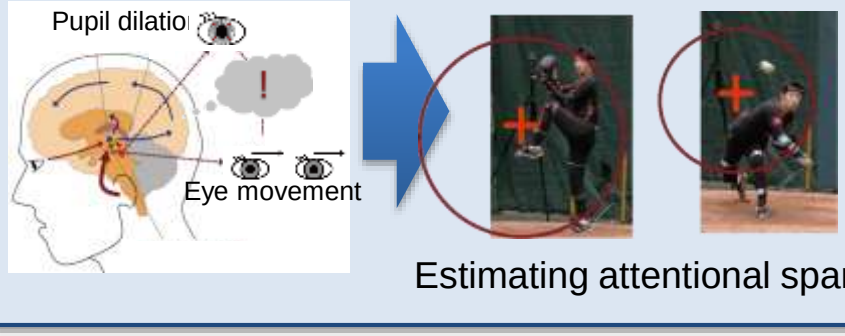
Spors Brain Science Project

SBS
SPORTS BRAIN SCIENCE PROJECT



Measurement using wearable sensing and virtual reality

Estimation of psychological state using measurement of eye movements

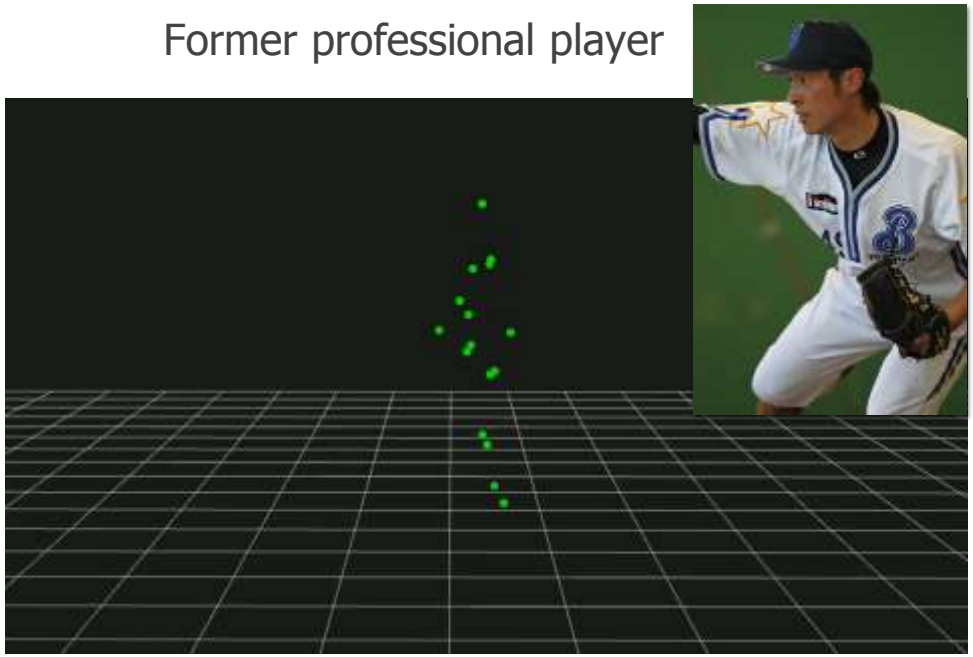


"Split seconds matter - the brain and sport"
http://sports-brain.ilab.ntt.co.jp/document/20170907_nature.pdf

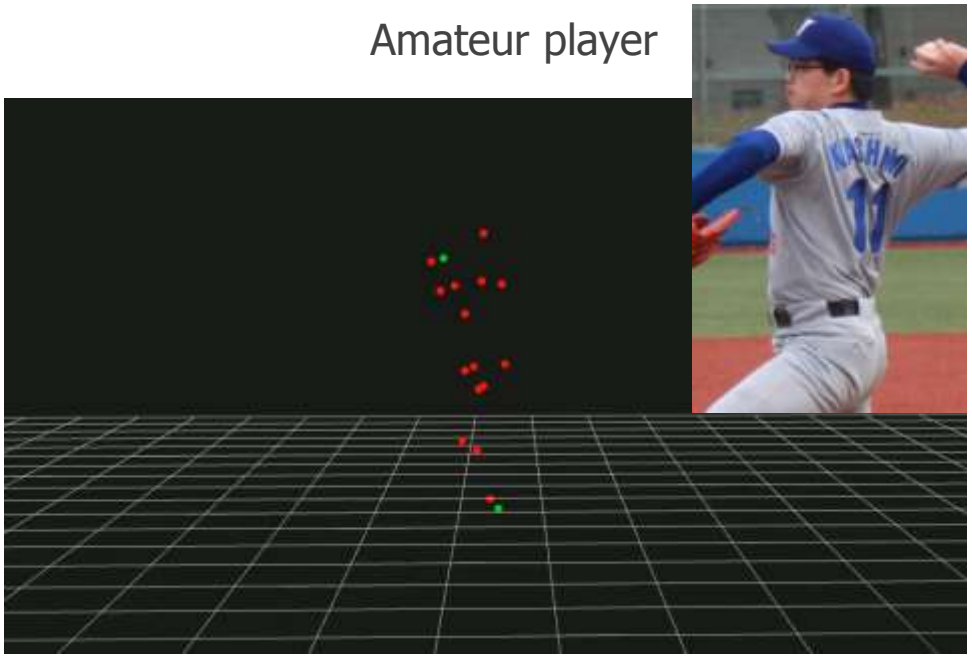
NTT

Which is the Former Professional Player?

Former professional player



Amateur player

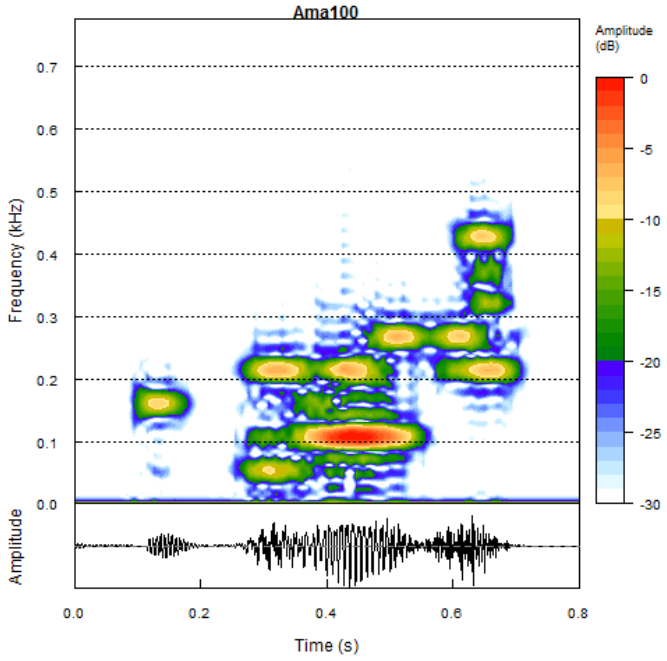


ball speed is the same (100 km/h)

“Sonification” of Muscle Activities

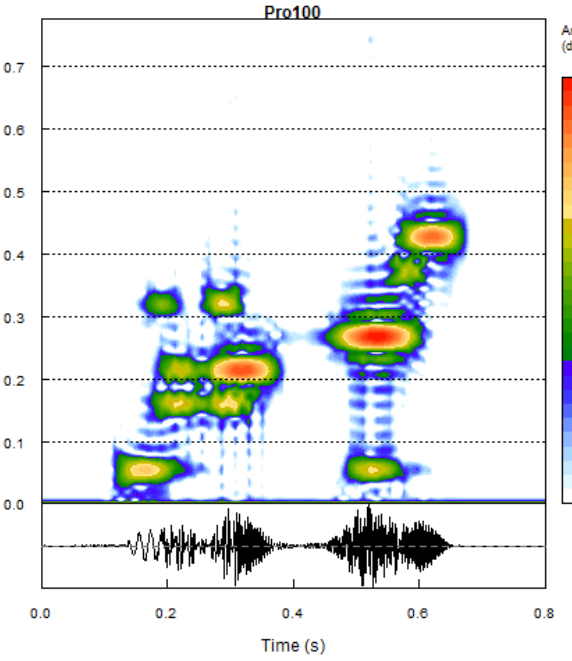
Amateur player

100 km/h

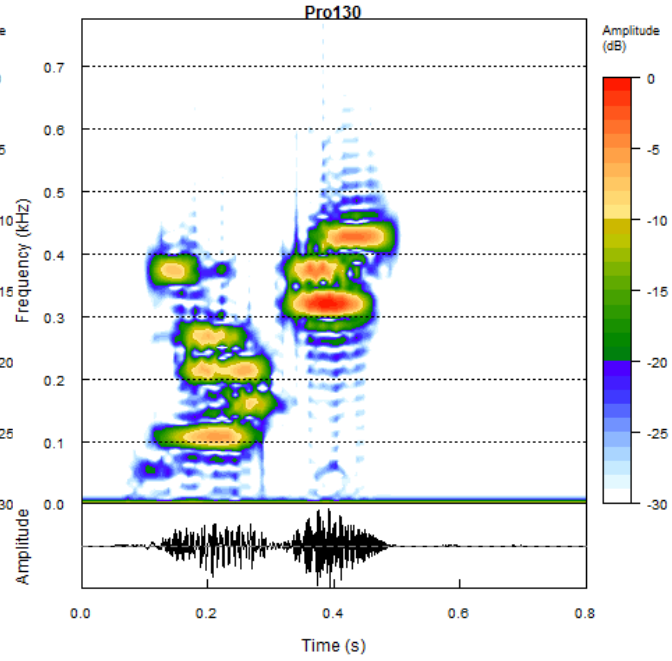


Former professional player

100 km/h



130 km/h



Click each graph to play respective sound

The Split Seconds Matter

make unpredictable

予測する/させない
0.1 sec. をめぐる攻防

battle for 0.1 sec.

predict

0 sec.

0.5 sec.

predict
(conscious level)
予測(自覚的)

predict
(subconscious level)
予測(無自覚的)

plan
運動計画

Go or NoGo

do
運動実行

adjust online
オンライン調整

conscious
意識

Implicit「潜在的」

Yu Darvish Pitch Overlay
https://www.youtube.com/watch?v=jUbAAurrnwU

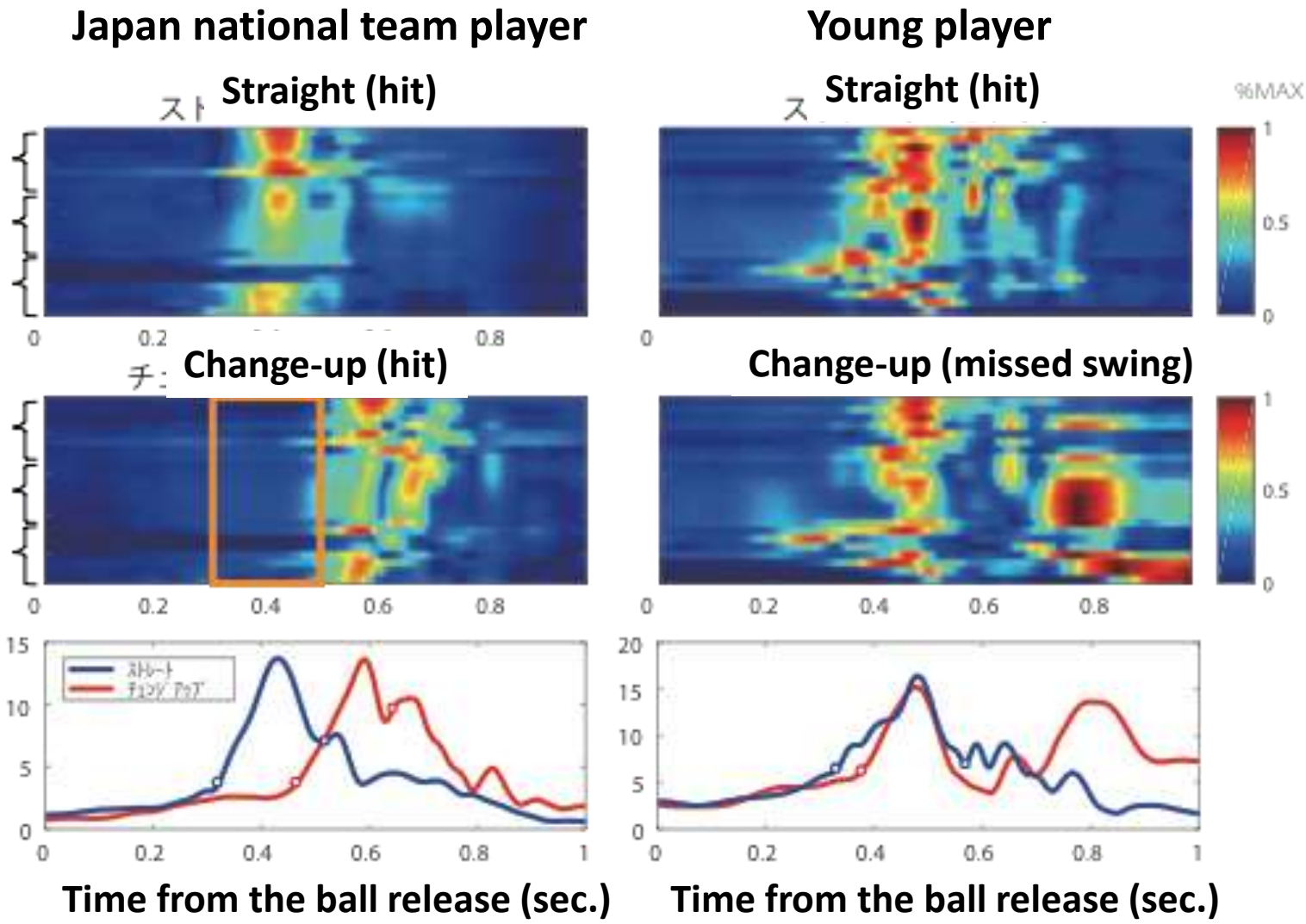
Smart Bullpen

http://sports-brain.ilab.ntt.co.jp/

- Reveal implicit brain functions underlying the outstanding performance of top athletes.
- Develop effective training methods that help athletes to raise their game.

A Good Batter can Hold and Make Time

打てる打者は「タメ」を作れる



Thank you for your time and attention!

corevo

Human and machine collaboration for revolution

Deep Learning is Everywhere

Deep learning ≈ artificial intelligence

It is widely used in many practical applications such as: machine translation, image recognition, anomaly detection, games, robots, automatic driving cars.

NTT

Real-time Prototype System

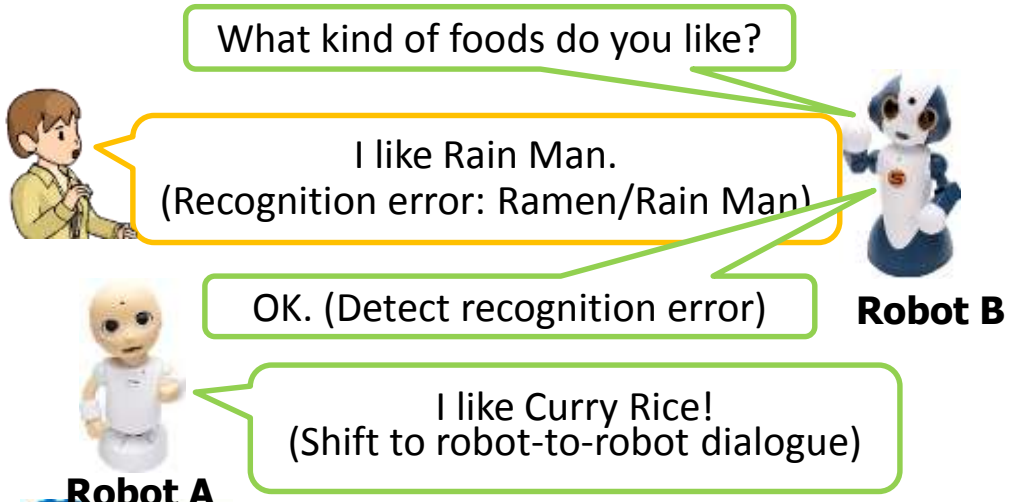
リアルタイム
複数人会話音声認識
Real-time multi-persons speech recognition

発話区間検出、音声強調、音声認識
Voice activity detection, speech enhancement and speech recognition

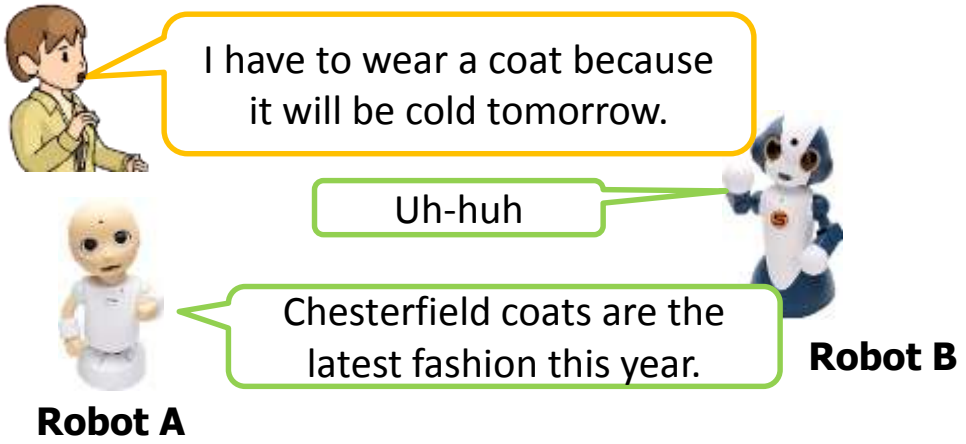
Coordinated Dialogue Control with Multiple Robots

- Coordinated multiple robots can improve dialogue even under some speech recognition/generation errors.
- Impression is greatly improved by switching robots appropriately considering human cognitive characteristics.

Detect recognition error
↓
Switch to robot-to-robot dialogue to avoid breakdown



Detect generation error
↓
Switch from one robot to the other to reduce discomfort



NTT

Joint research with Ishiguro Lab. @ Osaka University

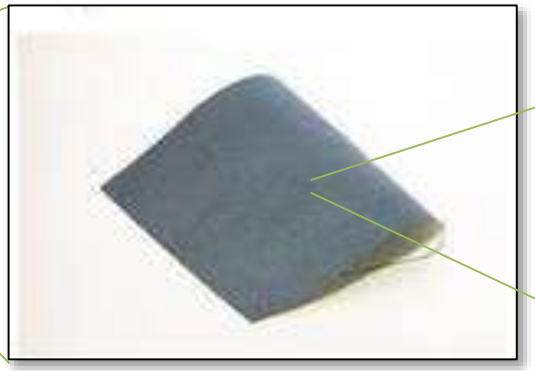
Demonstration (in Japanese)

Functional materials: hitoe®

hitoe®

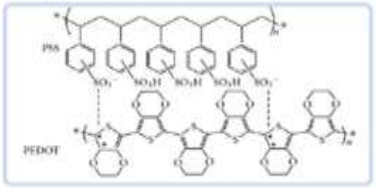
- Just wearing "hitoe" medical wear enables to detect medical-quality ECG (electrocardiography) signals and heart rates.
- It is developed and commercialized jointly by Toray and NTT.

NTT & TORAY Innovation by Chemistry

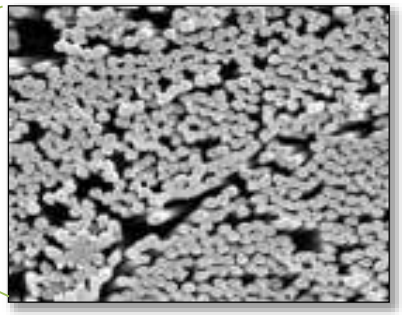


hitoe®

PEDOT-PSS



+



nanofibers

NTT

Home monitoring of a heart disease patient

A simple Holter ECG monitoring system with hitoe will reduce patient burden and improve examination efficiency in health screening and home medical care.


Patient at home

conventional Holter ECG monitoring

is replaced by

with

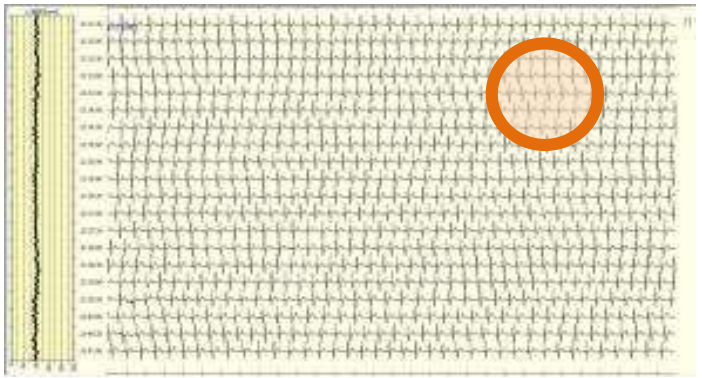

Medical wear

Long-term electrocardiogram waveform obtained

The doctor checks the status and contact patient's home


Hospital

Abnormal values are immediately reported to the hospital



*hitoe medical electrodes: pharmaceutical machine license approved and clinical research in progress