Exploring the Essence of Communication to Reach the Heart
Day 4 – Thursday, 9 December 2021
(Time zone - UTC+1)
12:00-12:30
Keynote session
## NTT R&D organization

<table>
<thead>
<tr>
<th>NTT IOWN Integrated Innovation Center</th>
<th>NTT Network Innovation Center</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTT Software Innovation Center</td>
<td>NTT Device Innovation Center</td>
</tr>
<tr>
<td>NTT Human Informatics Laboratories</td>
<td>NTT Social Informatics Laboratories</td>
</tr>
<tr>
<td>NTT Computer and Data Science Laboratories</td>
<td></td>
</tr>
<tr>
<td>NTT Information Network Laboratory Group</td>
<td>NTT Network Service Systems Laboratories</td>
</tr>
<tr>
<td>NTT Access Network Service Systems Laboratories</td>
<td></td>
</tr>
<tr>
<td>NTT Space Environment and Energy Laboratories</td>
<td></td>
</tr>
<tr>
<td>NTT Communication Science Science Laboratories</td>
<td></td>
</tr>
<tr>
<td>NTT Network Innovation Laboratories</td>
<td>NTT Device Technology Laboratories</td>
</tr>
<tr>
<td>NTT Device Technology Laboratories</td>
<td>NTT Basic Research Laboratories</td>
</tr>
</tbody>
</table>

- **2,300 researchers**
- **$1.0 billion invested**
- **18,000+ patents granted**
- **10 consecutive years**
  Received Derwent Top 100 Global Innovators by Clarivate Analytics
- Founded on July 4, 1991
The mission of NTT Communication Science Laboratories

- Human: deep understanding of humans
- Brain: Communication that reaches the heart
- Media: approaching human abilities
- Knowledge:
SpeakerBeam: Computational selective listening based on deep learning

Humans are good at selective listening.

The voice of the target speaker can be extracted simply by teaching the characteristics of the voice.
Voice Conversion: Converting impression and intelligibility of speech

Voice conversion modifies speech signals to one’s desired form of expression without changing the content.

It not only makes communication smoother but also creates new forms of communication that extend human vocal and auditory functions.
Cross-modal voice conversion

What kind of face does the owner of this voice have?

May be this face?

Learn to match these coordinates

Time-independent voice feature

Time-independent face feature

common space

coordinate

Face encoder

Voice encoder
Building common space by multiple encoders for concept acquisition

On a clear day, you can see a windmill on the hill. Image object and word originating from the same ‘thing’ tend to co-occur in many samples.
Tele-stethoscope

A doctor can listen to sounds coming from various parts of the patient’s body.

An examination vest is equipped with multi-channel* acoustic sensors.

*18 acoustic channels and 1 electrocardiogram (ECG) channel
From tele-stethoscope to AI stethoscope

Visualize 3D heart shapes based on the heart states

sound → 3D shape

sound → text

Explain the meaning of the sounds in text

Specificity controls amount of information in the output sentences

Conditional sequence-to-sequence caption generator (sound-to-sentence translator)

Examples

- Low specificity: The heart sounds are abnormal.
- High specificity: 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.
Remote Haptic Communication

Vibrotactile Telephone Lite

Ultra-Future Sensory Public Phone
Implementing perceptual illusions into the real world

変幻灯 Hengento (2015)

HiddenStereo (2017)

浮像 Ukuzo (2018)

Danswing Papers (2019)

Velvet hand illusion (2021)
Language development of infants

Picture books supervised by NTT CSL have achieved 280,000 copies in total.

Do human infants autonomously learn from the co-occurrence of phenomena in the natural world?

mon • key ⇒ monkey

child vocabulary development database

Printed books stimulate five senses
Personalized Picture Book

1. A main character is set to match the child's name and gender
   - Hiro-kun (boy)
   - Aki-chan (girl)

2. The story is customized to match the child’s interests
   - Dog story
   - Elephant story

3. Words with pictures that the child is just about to learn are selected
   - elephant
   - ant
   - ladybug
   - butterfly
   - cat
   - monkey

Target Age: 1-2 years of age

Enter information about a child at: https://ehon.nttprint.com

- child’s name and gender
- favorite objects
- vocabulary checklist

Designed by Akio Kashiwara and NTT Print Corp.

Child Vocabulary Development Database
Conclusion

What it means to be a human? is an ultimate question

The way we interact with other people is changing

Technologies are approaching toward human intelligence

Study illusions provide important clues to understanding the brain functions of humans

Pursue the possibilities of new communication media

Challenge the technical limits toward approaching human intelligence

Reach out and touch someone’s heart
Thank you!