

- Speech and Image Processing –
**“Media Technologies for Support of
Human Communication”**

Tetsuya Takiguchi
Kobe University

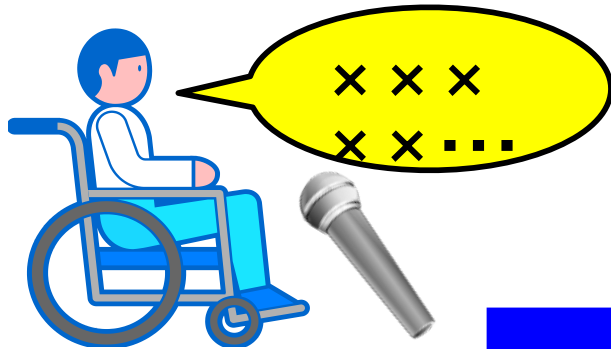
Speech Disorders



- His way of speaking is different
- No intelligence deficiency

Support of Speech Communication

People with speech disorders

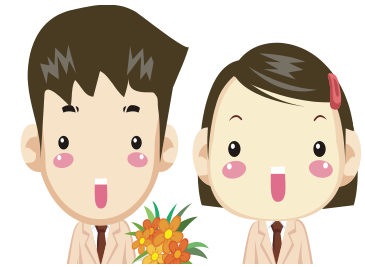
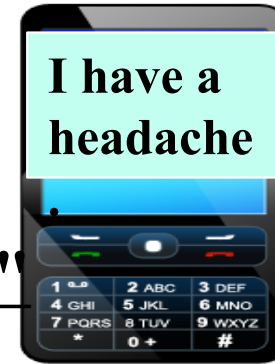


(Difficult for us to understand what they say)



Speech recognition

"Output text"

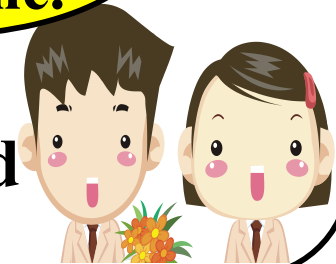


Speech synthesis

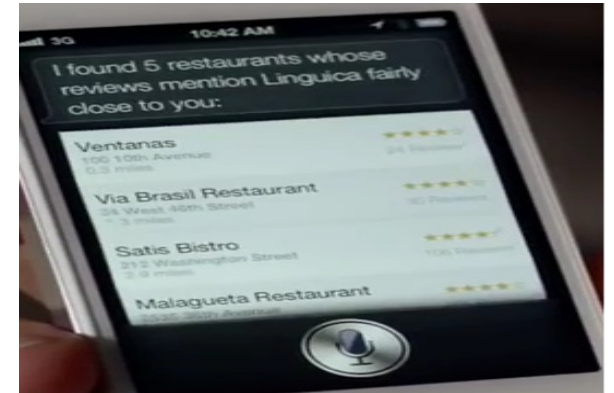
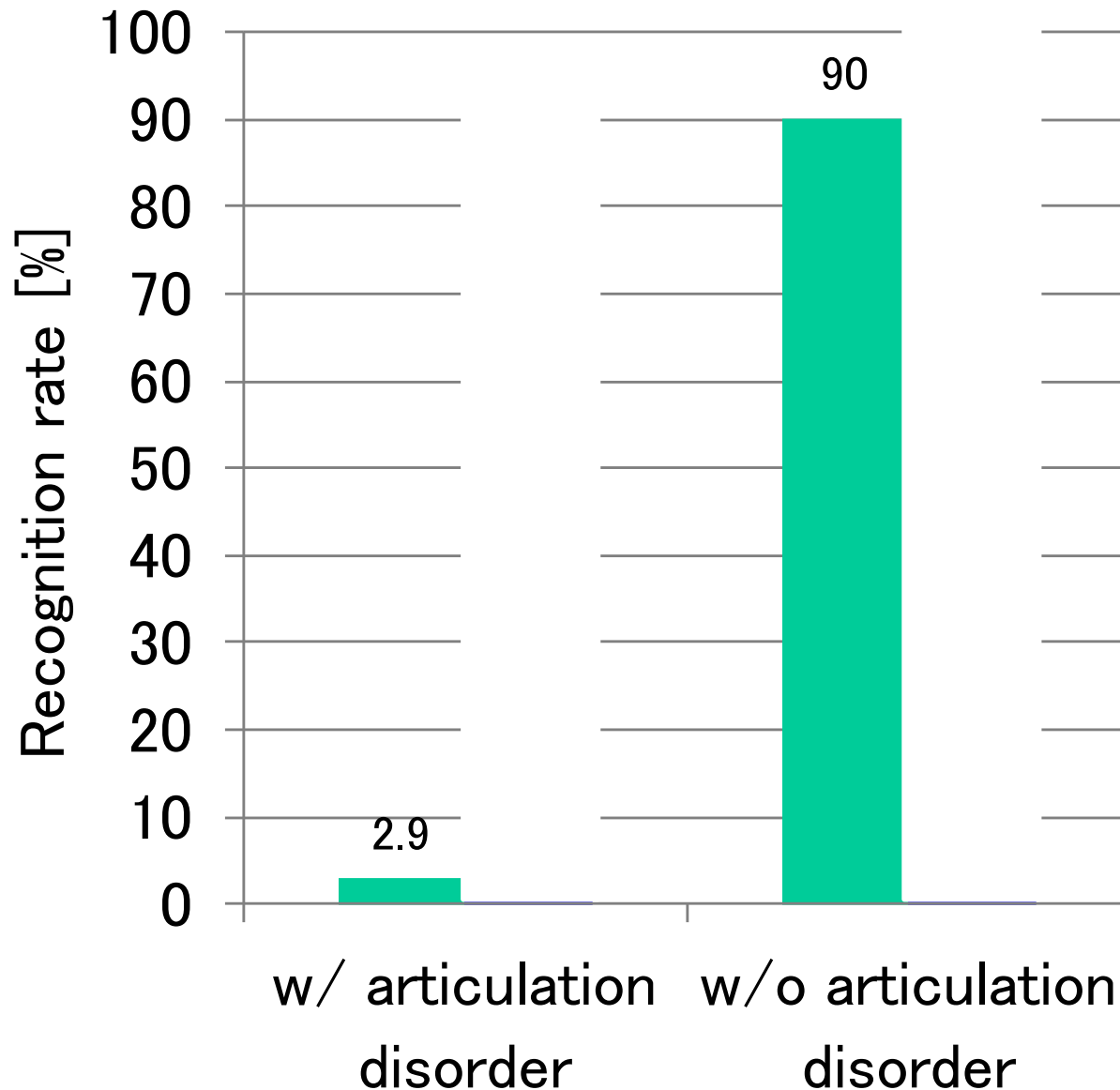
"Output sound"

I have a headache.

(Easy for us to understand what they say)

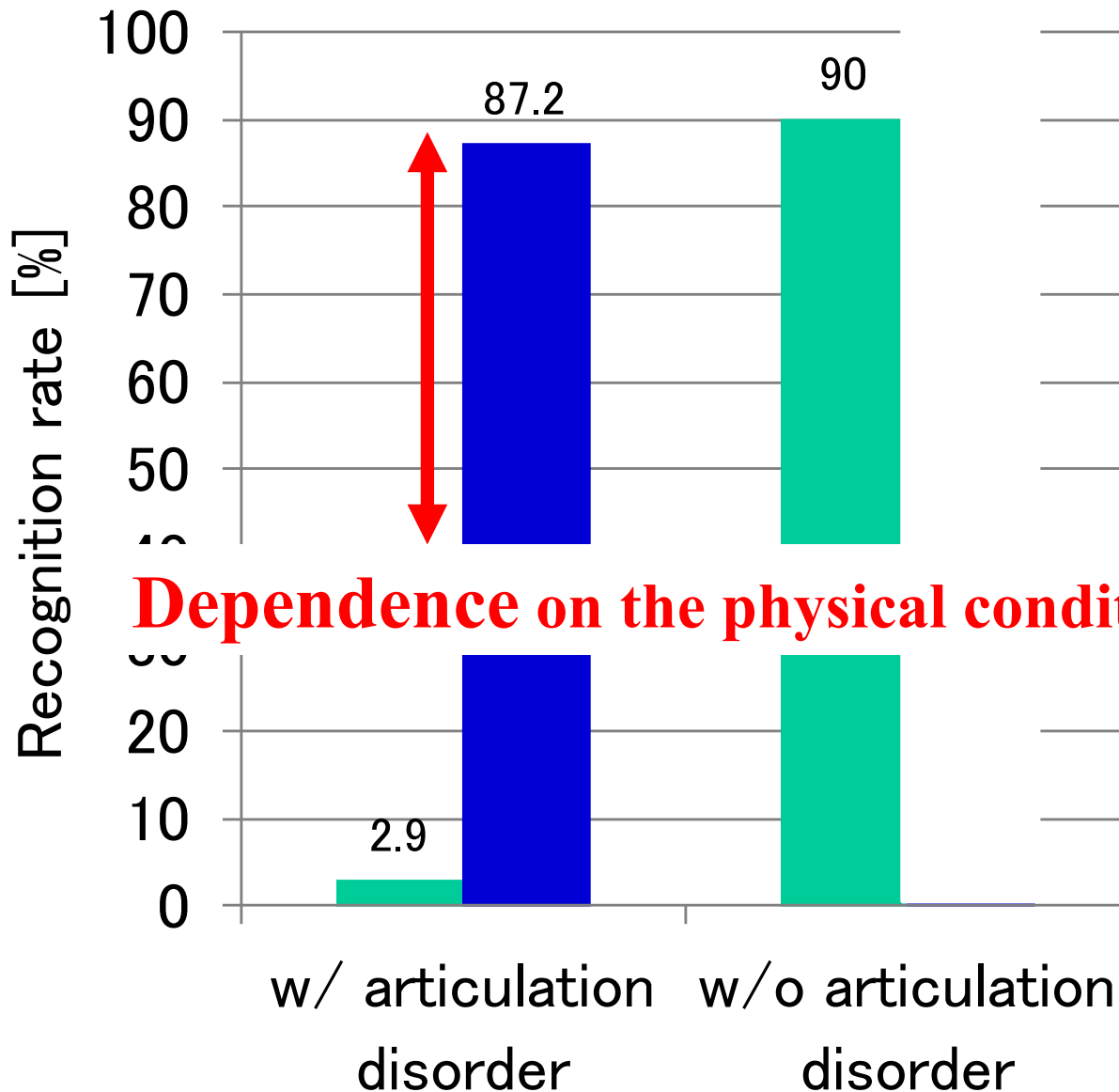


Speech Recognition (210 words)

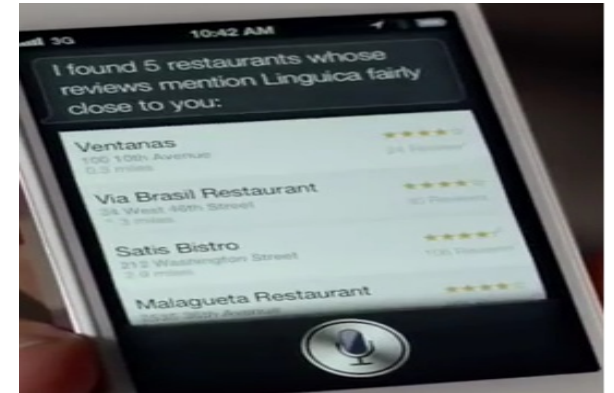


■ Speaker-independent model (w/o disorder) (for unimpaired persons)

Speech Recognition (210 words)

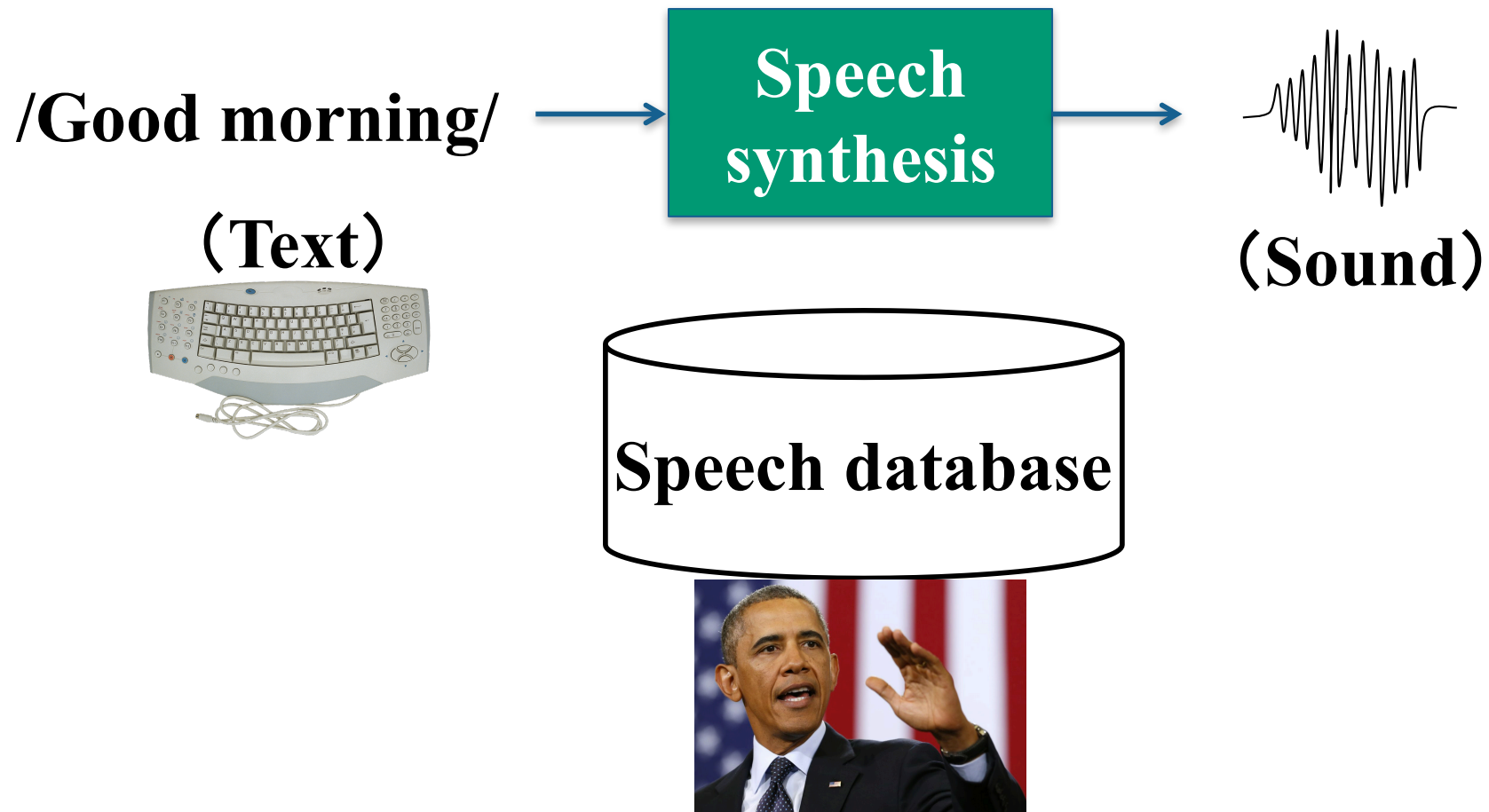


Dependence on the physical condition



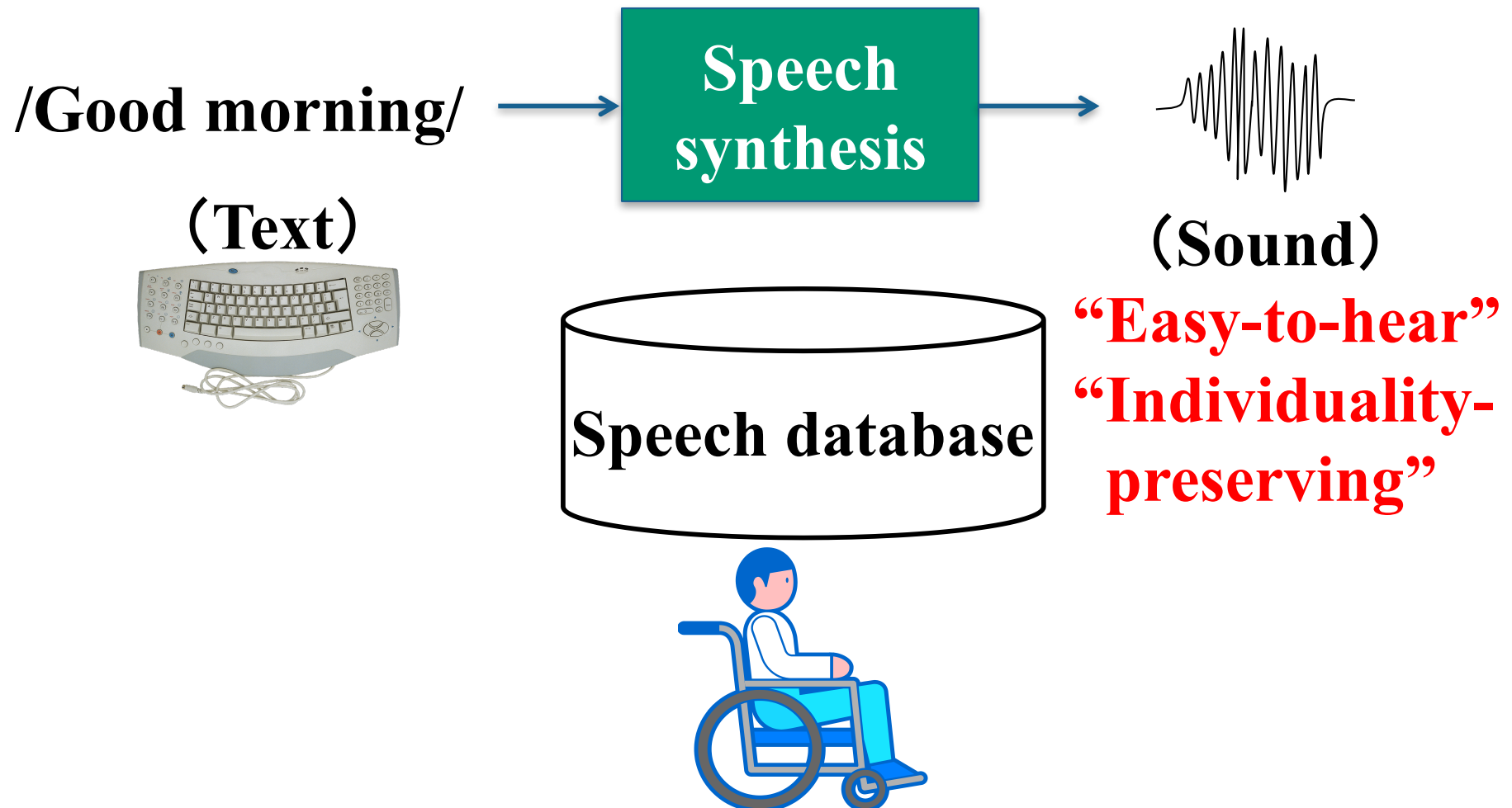
- Speaker-independent model (w/o disorder)
- Speaker-dependent model

Speech synthesis using deep learning



Speech synthesis using deep learning

Individuality-preserving speech synthesis

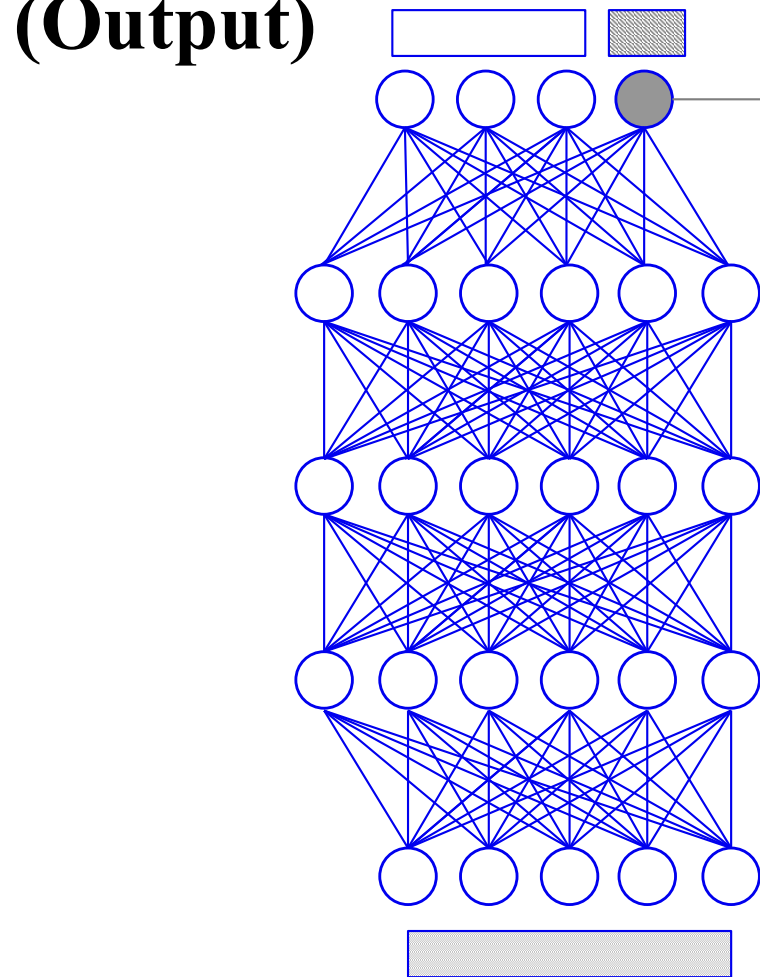


Hybrid deep neural networks

DNN for unimpaired person

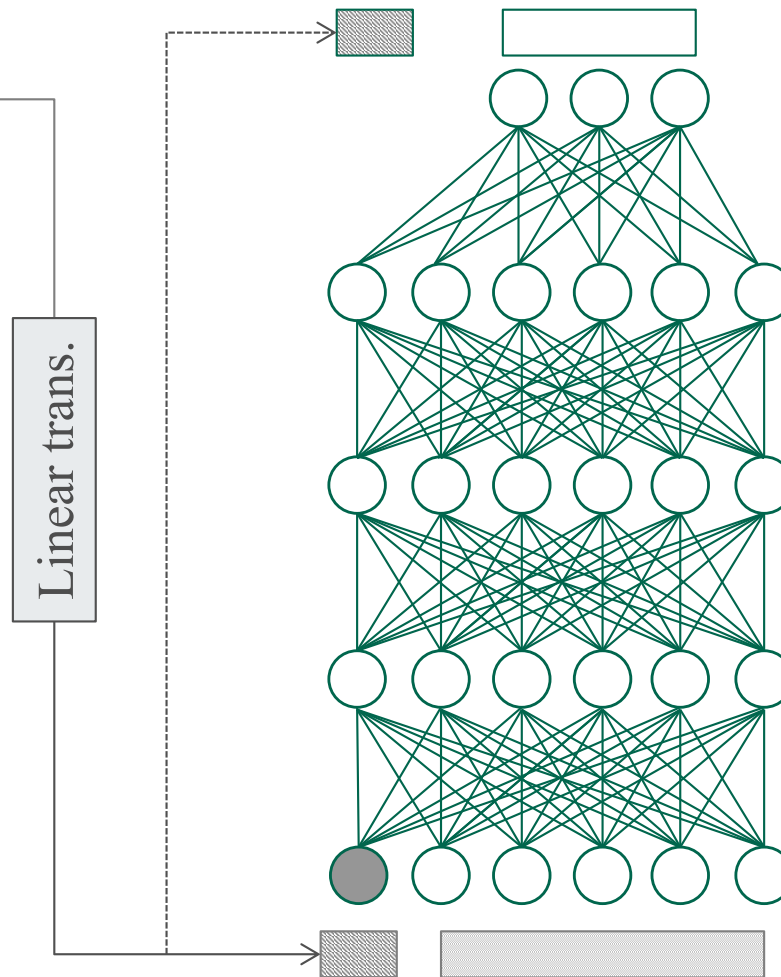
DNN for speech disorders

Acoustic features
(Output)



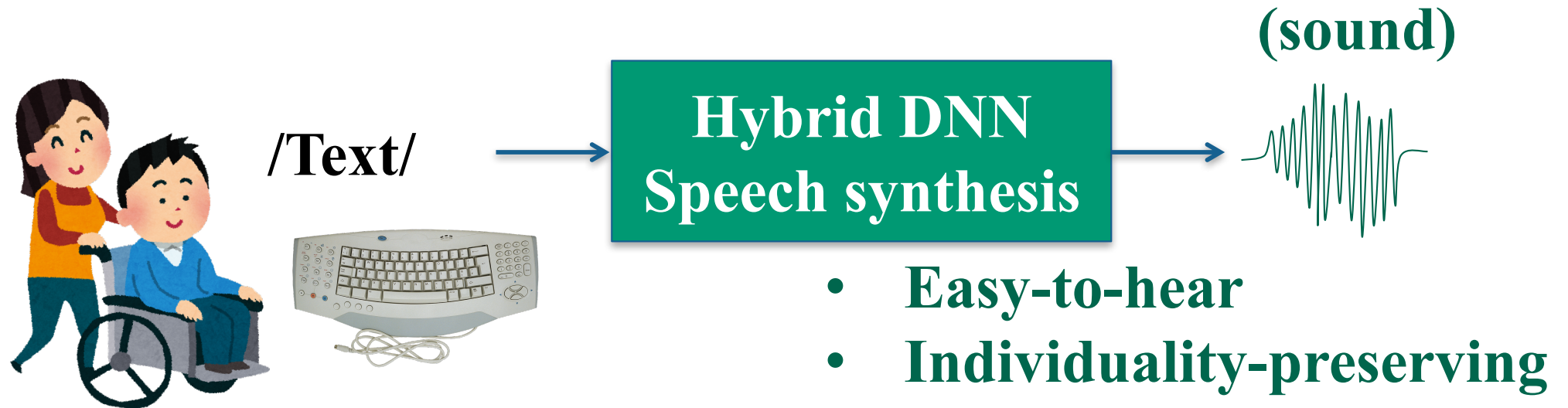
(Input) **/Hello/**

Acoustic features



/Hello/

Individuality-preserving speech synthesis



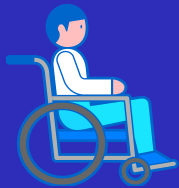
Original sound



Synthesized sound

Augmentative Speech Communication

Communication Support for
People with Articulation
Disorders

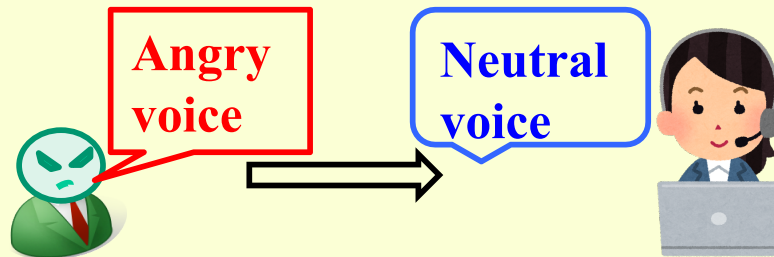


Visual Sound Recovery
from Minute Vibrations

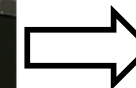


Good
afternoon!

Emotional Voice Conversion

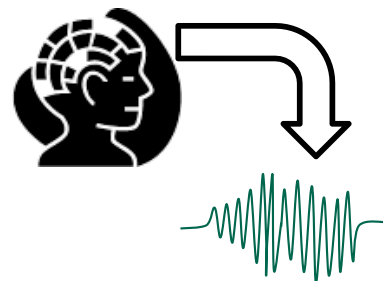


Semantic Segmentation for
Disaster classification



{Bldg.
Water
...}

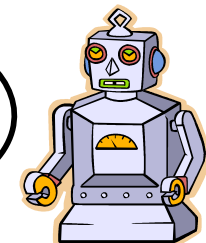
Brain-to-Voice Conversion



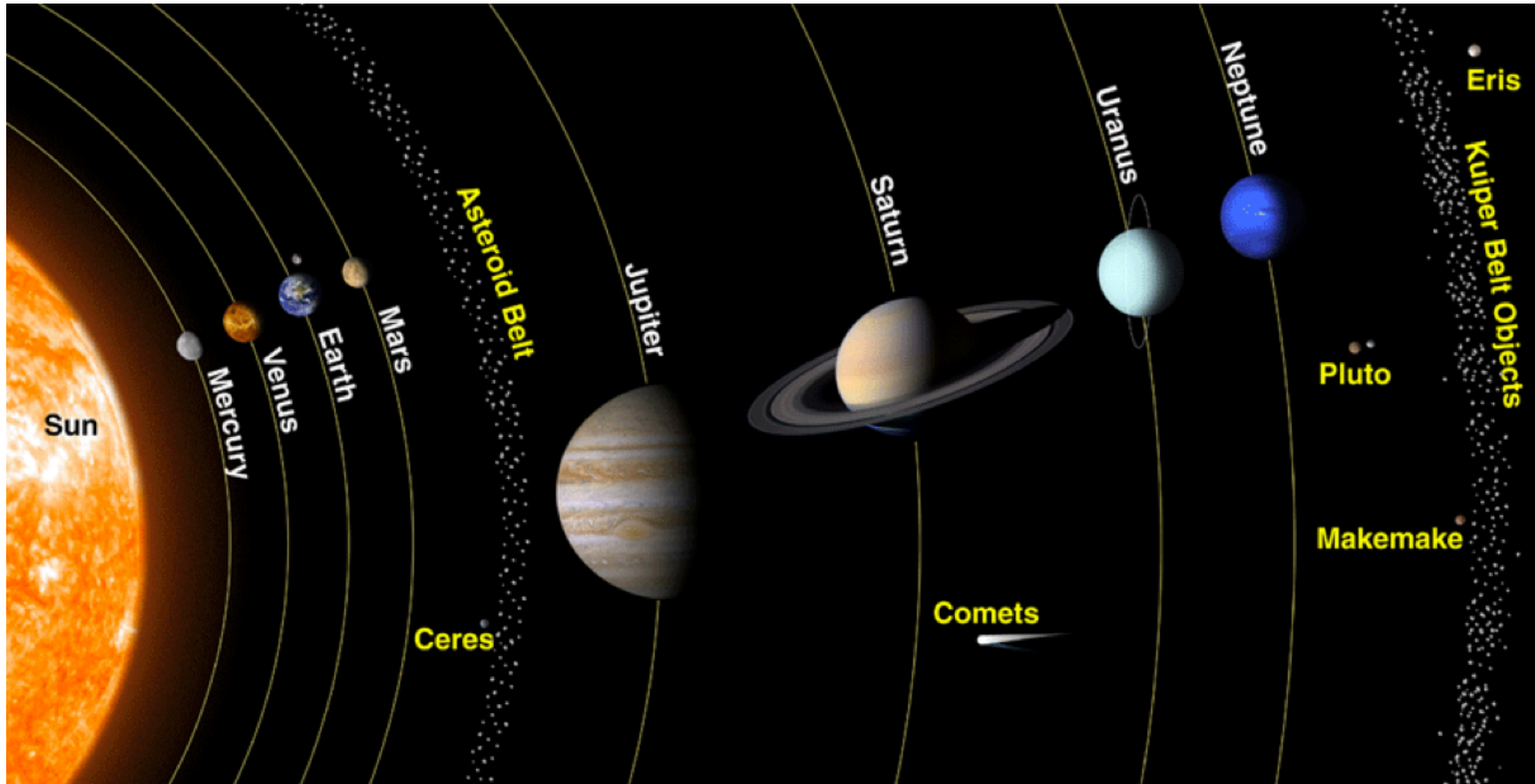
Spoken Dialogue System
-chat dialogue-



What
a bore.

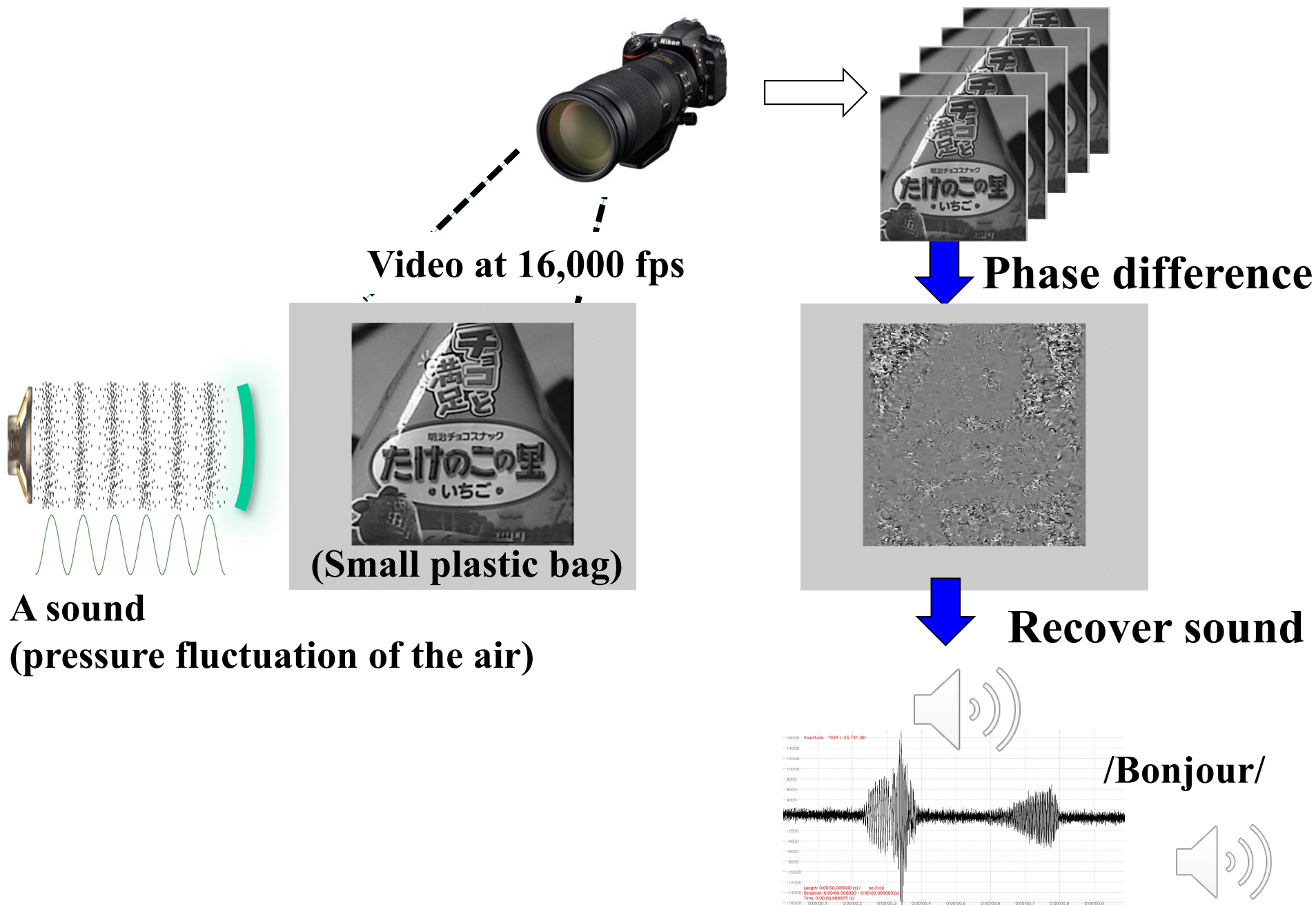


From how far away can you capture the sound?



- “hear” sounds from object images.

Visual sound recovery using a high-speed camera



Thank you for your kind attention.