

Speech



METI Project

Development of Fundamental Speech Recognition Technology

Project Leader: Sadaaki Furui, Tokyo Institute of Technology

Project Sub-leader: Tetsunori Kobayashi, Waseda University

Tokyo Tech, Waseda Univ

OKI, Hitachi, Toshiba, NEC, Asahi-kasei, Mitsubishi



■ **Speech interface for everybody**

- ◆ **Large-scale contents over high-speed network & large storage**
- ◆ **Highly sophisticated functions in diverse services & software**
- ◆ **User-friendly interface and information access technology is indispensable**

Demands for ASR

- ◆ **Improvement of ASR and interface construction technology (interface development framework) is necessary**

Development of fundamental technology for flexible ASR systems

- ➔ **Make the systems easy to use for everybody including beginners and aged people**

Speech

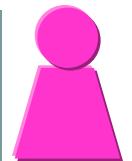
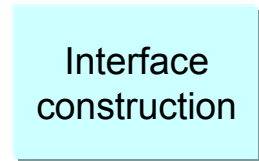
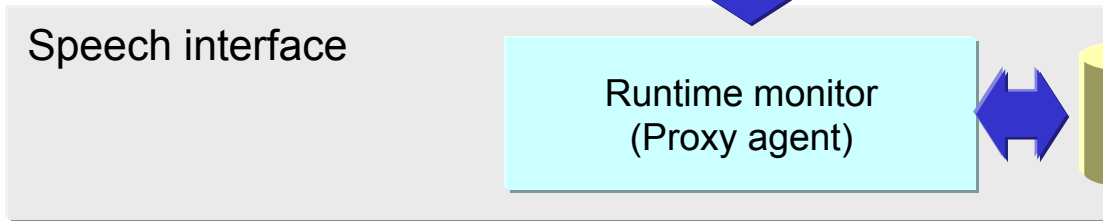
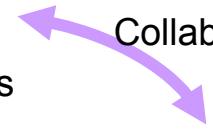
Project overview

⑦ Overall evaluation



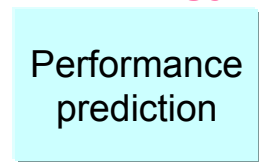
Users

Collaboration

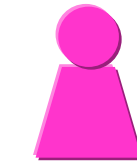
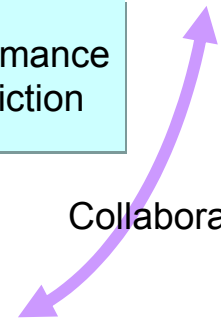


Application developers

⑥ Speech interface technology

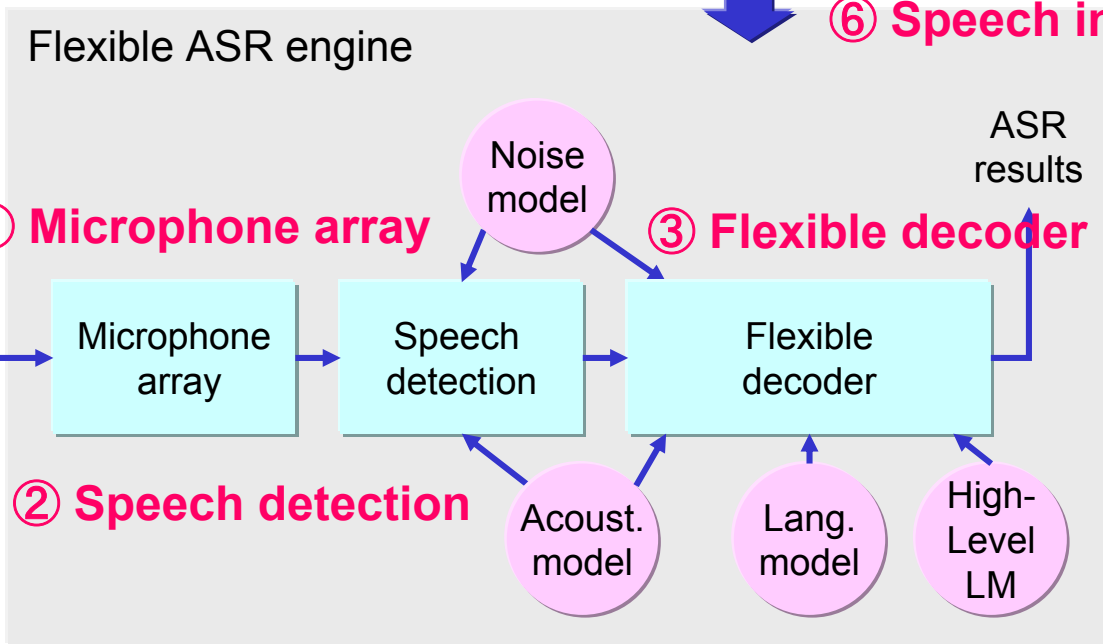


Collaboration



Engine developers

New development framework



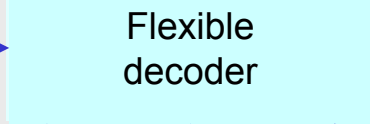
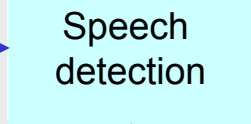
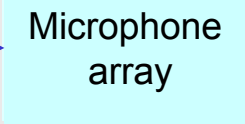
① Microphone array

Noise model

③ Flexible decoder

ASR results

Speech



② Speech detection


Acoust. model

Lang. model

High-Level LM

④ Multi-lingual technology

⑤ High-level LM

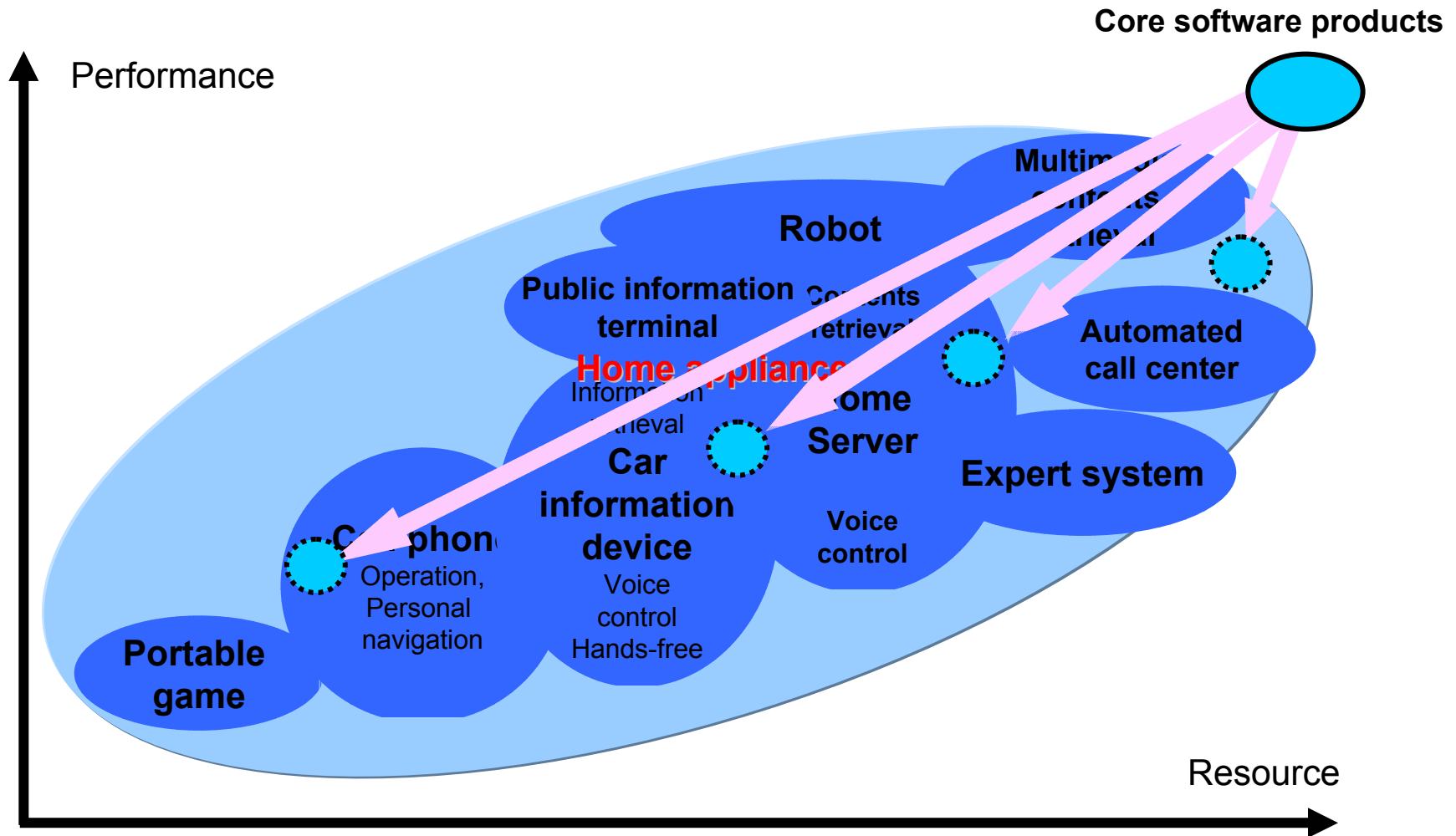


Items	Present	Target
①Microphone array <ul style="list-style-type: none">• Size• Delay• Speech quality	20 cm × 3 cm 1 s MOS value: 2	3 cm × 3 cm 100 ms MOS value: 3
②Speech detection <ul style="list-style-type: none">• Accuracy at ordinary environment	80%	95%
③Flexible decoder <ul style="list-style-type: none">• Recognition accuracy	80~90%	90~95%
④Multi-lingual technology	None	Japanese, Chinese, Thai, etc.
⑤High-level language model	None	Automatic acquisition technology for linguistic information
⑥Speech interface construction technology	None	Speech interface development supporting system

Speech

Applications

■ Various speech applications



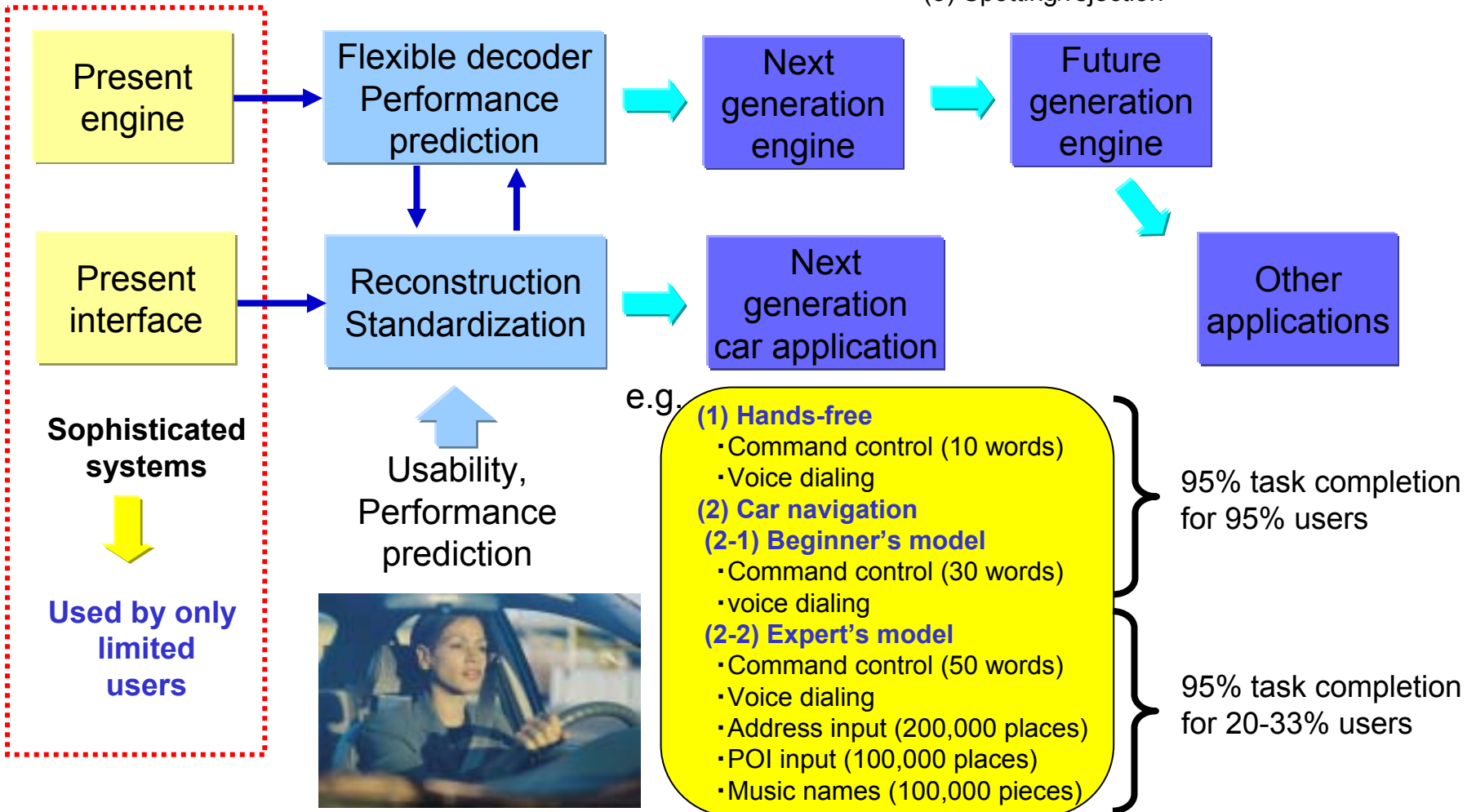
Speech

Car information device

Product image

New functions:

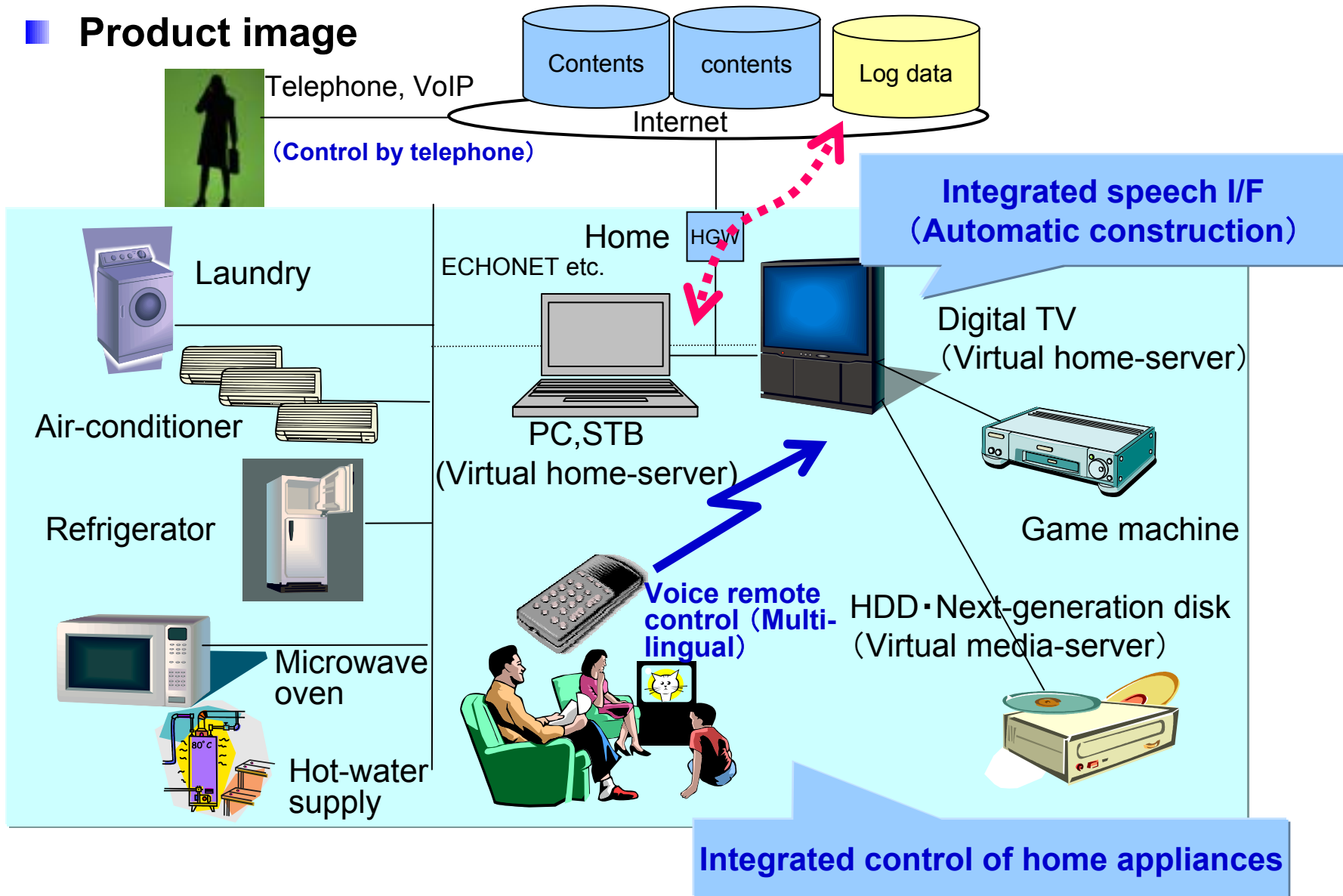
- (1) Performance variation prediction
- (2) Advanced applications
- (3) Spotting/rejection



Speech

Home server

Product image





Related papers presented at ASRU

P. R. Dixon, D. A. Caseiro, T. Ohnishi & S. Furui: “The TITECH large vocabulary WFST speech recognition system,” S4.12

T. Nakano, S. Fujie & T. Kobayashi: “Extensible speech recognition system using Proxy-agent,” S5.19