
Preliminary Evaluation of Speech to Text Query Application for Parts Database

2020/10/18-22

Yuichiro Aoki email: yuichiro.aoki.jk@hitachi.com

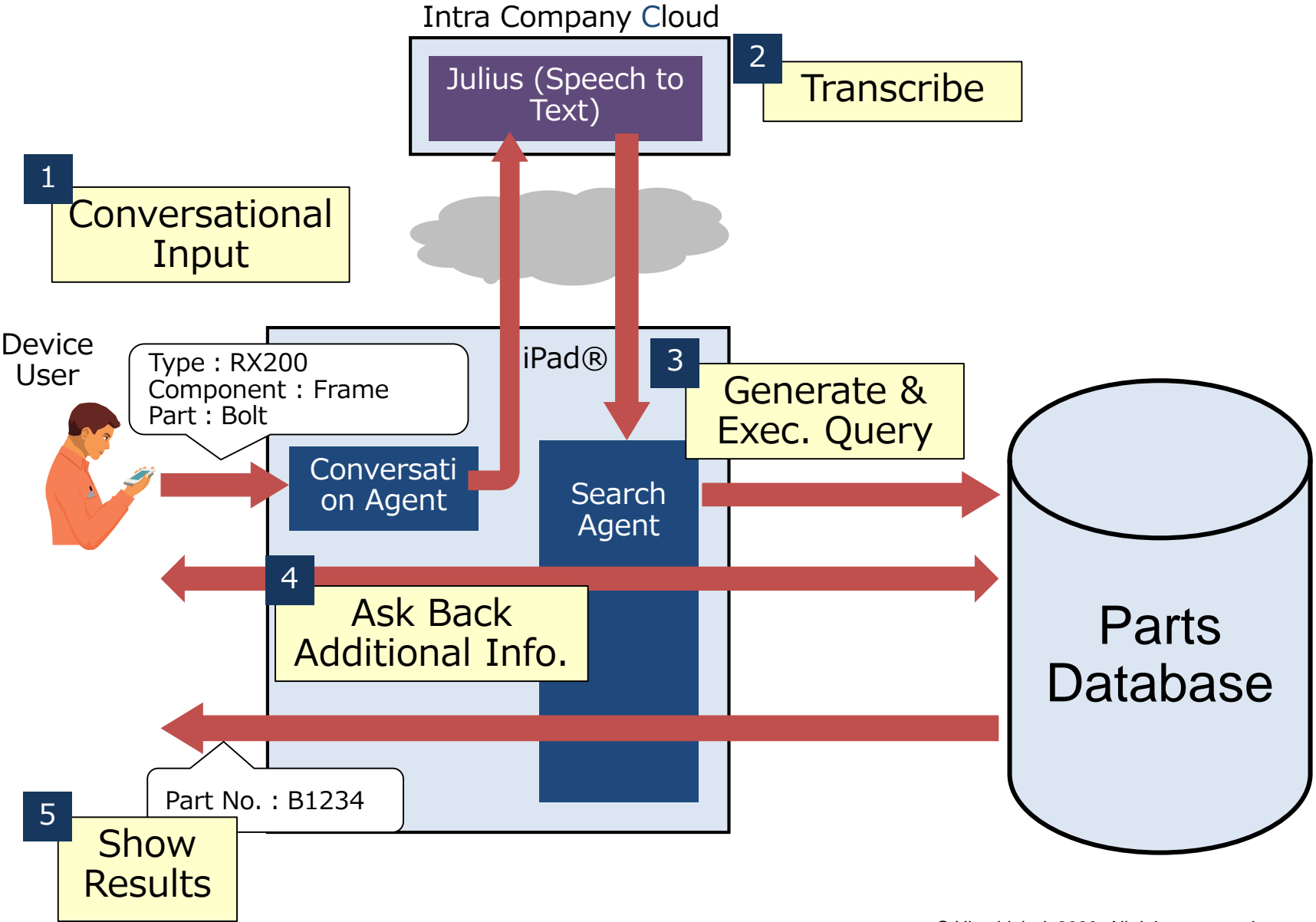
Tadashi Takeuchi email: tadashi.takeuchi.dt@hitachi.com

Research and Development Group, Hitachi, Ltd.

- Replacement of the deteriorated parts is necessary.
- Users have to know part numbers to order and replace them. Many of them call to the support center of the vendors, instead of searching by themselves. Thus, a lot of phone calls bother its staff and prevent them from improving productivity and customer satisfaction.
- We propose an automatic answering system of part numbers using the open-source speech to text engine and the SQL generation from the natural language.

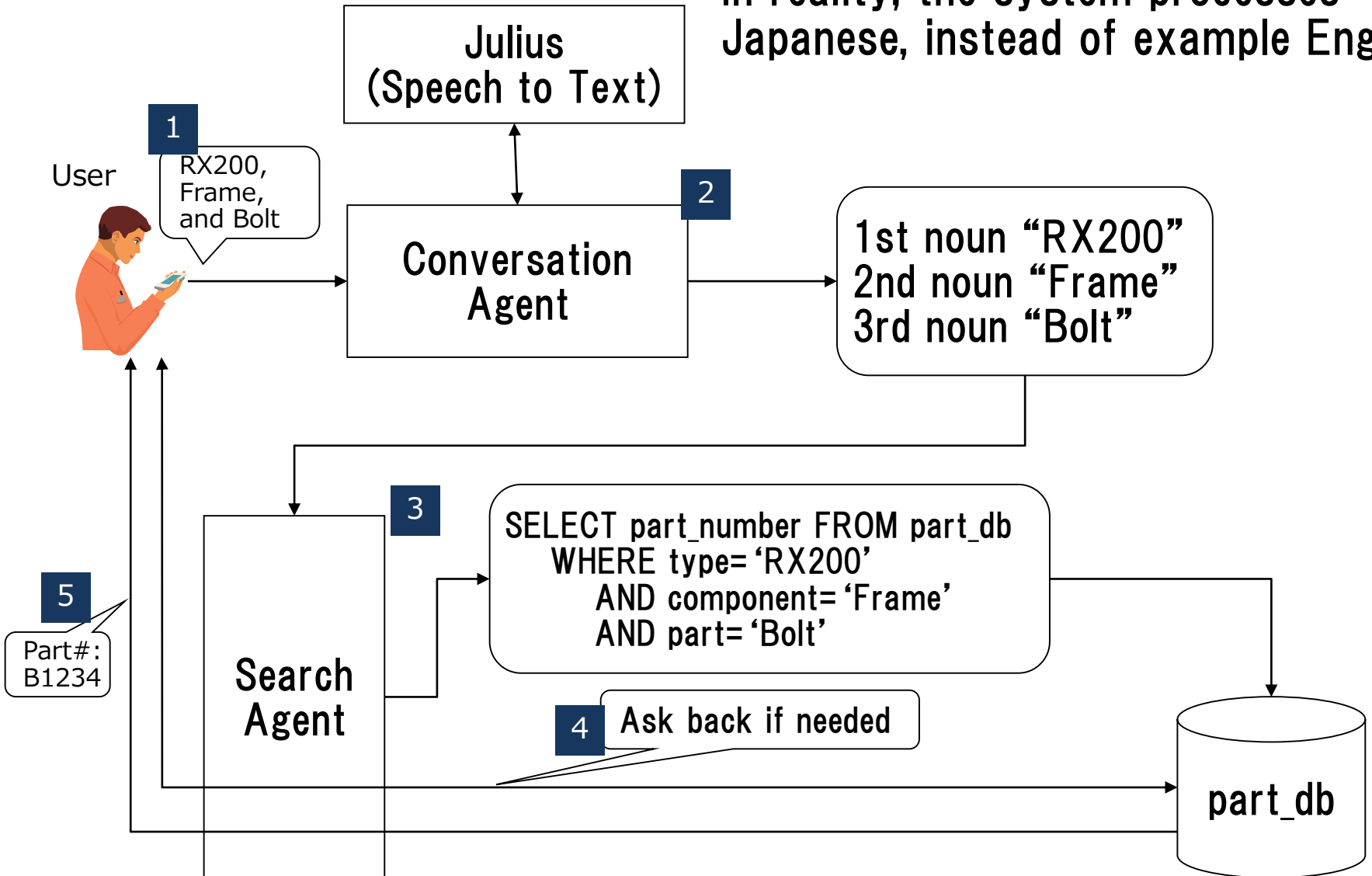
- In our customer company, the small number of the support staffs have to receive the huge number of phone calls from users. Thus, they cannot improve productivity and customer satisfaction of the support center.
- We analyzed the statistics of the phone calls manually and found that the most frequent phone calls is related to the part number inquiries.
- The part numbers are needed when the parts are deteriorated or break down, and they must be replaced. Thus, the number of the part number inquiries cannot be decreased.
- We propose the automatic answering system of part numbers using the open-source speech to text engine and the SQL generation from the natural language.

3. System Overview



4. Schematic Algorithm

In reality, the system processes Japanese, instead of example English.



■ iPad® Display Images. All operations are processed in Japanese.

Voice Input

User: RX200, Frame, and Bolt

“What part number do you want to search?”

(Webログイン中...)
> 何の部品番号を調べたい?

Send

Record

After voice input, press Record button.

5. GUI Examples



Speech to text transcribes voice input. "RX200, Frame, and Bolt"

5. GUI Examples



5. GUI Examples

Application returns a correct part number "B1234".



■ In case voice input lacks some information (incomplete question)

Voice Input

User: RX200, and Bolt

(Webログイン中...)
> 何の部品番号を調べたい?
> 音声入力内容は?

200のフレームのボルト
品番号：B1234
何の部品番号を調べたい?
> 音声入力内容は?

Send

rx200のボルト	Select
RX200のボルト	Select
(その他)	Select

**Transcribed text.
No component name "Frame".**

Record

5. GUI Examples

Application asks back the user what information is needed (in this case, component name).

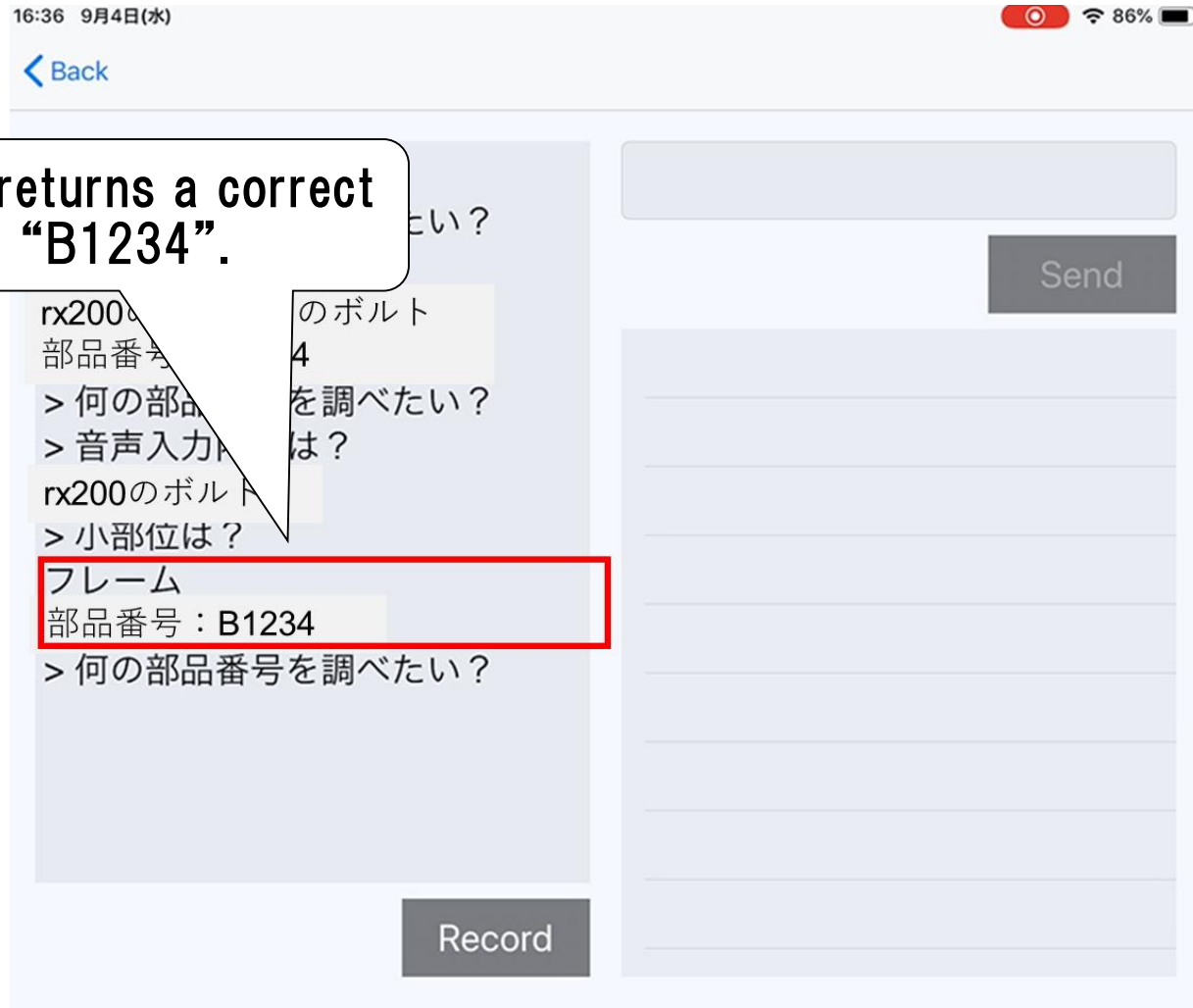
(web口)
> 何の部品を調べたい?
> 音声入力内容は何?
rx200のフレームボルト
部品番号：B
> 何の部品番号を調べたい?
> 音声入力内容は何?
rx200のボルト
> 小部位は?

The screenshot shows a mobile application interface. At the top, there is a status bar with a red indicator, a Wi-Fi icon, and 86% battery. Below the status bar is a large grey input field. To the right of the input field is a grey button labeled 'Send'. Below the input field is a list of three items, each with a blue 'Select' button to its right. The items are: 'フレーム', 'ラジエーター', and '(その他)'. A red box highlights this list. Below the list is a large white speech bubble containing the text 'Choices: "Frame", "Radiator", "(Other)". At the bottom of the screen is a grey button labeled 'Record'.

フレーム	Select
ラジエーター	Select
(その他)	Select

Choices:
"Frame"
"Radiator"
"(Other)"

5. GUI Examples



■ Environment

- iPad Pro® (6th Generation)
- Intra-Company Cloud (Julius(Speech to Text), RDB)

■ Test Questions

▪ Functional Evaluation

- + 100 artificial oral questions made by ourselves to test the coverage.

ex.) “RX200, Frame, and Bolt”

- + compare results between by our work and by the correct answers(manually investigated).

▪ Performance Evaluation

- + 11 oral questions from field engineers.
- + our work vs. the conventional keyword search using keyboard.
- + measure the execution time to get the answer.
if it exceeds 300sec., it is assumed to be 300sec..

- Eighty-Three out of 100 oral questions reach the correct part number.

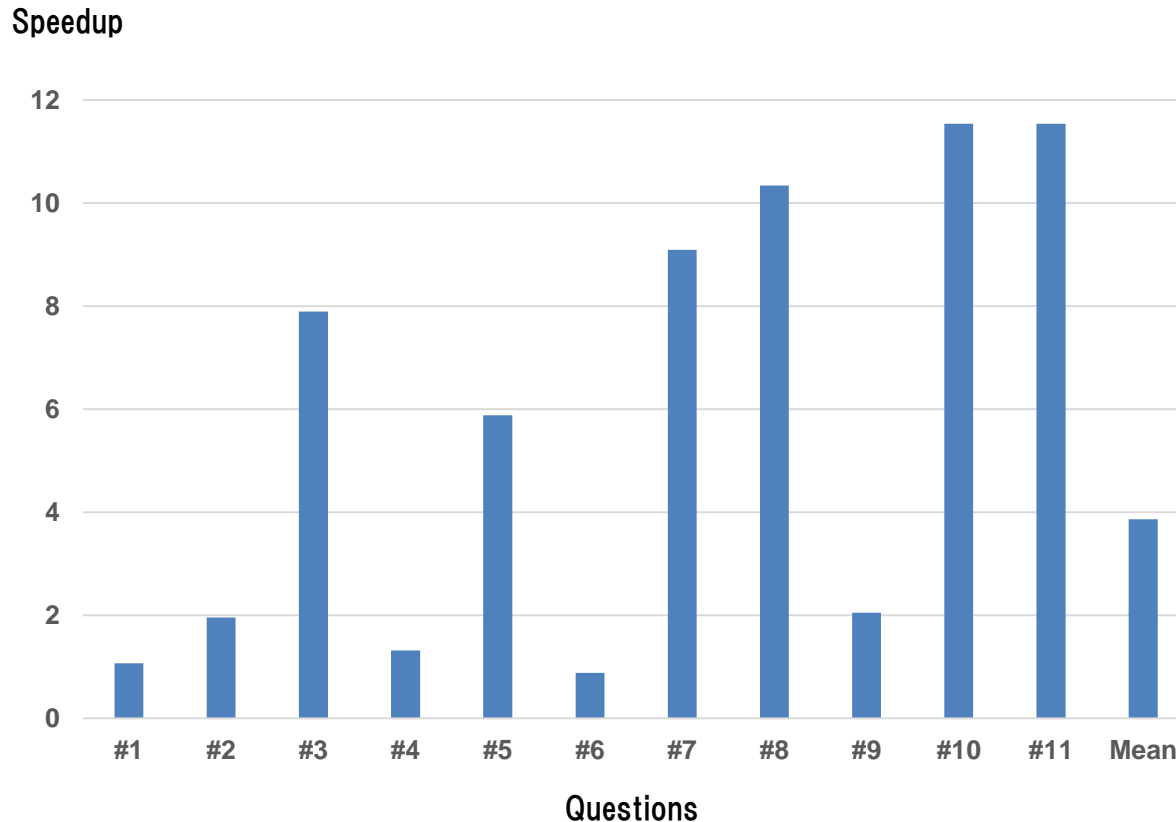
Results	Number of Questions
Correct Part Number	83
Inappropriate Speech Recognition	12
Other Error	5

- Most of them need manual correction of speech recognition results.

Voice Input	Additional Info. Needed	Manual Correction	Number of Questions
✓			0
✓	✓		2
✓		✓	40
✓	✓	✓	41

8. Performance Evaluation

■ Most of our system results except #6 outperforms the conventional keyword search from keyboard. Mean Speedup is 3.86.



■ Functional Evaluation Results

- + improper transcriptions of 12 questions, because of deficiency of the synonym dictionary for technical terms, such as “urea”.
➡ need to expand the synonym dictionary.
- + 81 out of 83 questions which reach the correct answers need the manual correction of the transcribed text, because of loud background noise behind the experiments in the office.
➡ need to tune speech to text OSS.
- + 43 questions need the additional information, because the same part name and component name appear in the same device type.
➡ our system asks back a user what information is needed additionally.

■ Performance Evaluation Results

- + In the conventional keyword search using keyboard, 6 out of 11 questions doesn't reach the part number within 300 seconds, because the user doesn't find the correct part name.
- + On the other hand, our application reaches the correct part number for all the 11 questions, because “ask back” mechanism helps the user to find the correct part name.
- + As a result, our application outperforms the conventional keyword search 3.86 times faster in average.

■ Speech to Text

We choose Julius because of its functionalities and OSS nature.

Name	Japanese	User Dictionary	Sound Model Customization	Open Source Software
Amazon Transcribe		✓		
Google Cloud Speech-to-Text	✓		✓	
IBM Watson Speech to Text	✓	✓	✓	
Microsoft Speech-to-Text	✓		✓	
Julius (This Study)	✓	✓	✓	✓

- Natural Language(NL) questions to SQL queries
 - + Our application translates simple NL questions into SQL queries.
 - + It generates complex SQL queries using WHERE clauses.
 - + It has interactive GUIs for mobile environment using iPad®.
 - + Accuracy is 83%.

	NL level	Interactive	Mobile	Misc.
Precise (2003)	Simple			
Rao et al. (2010)	Simple			User dict. words only
Safari and Patrick (2014)	General			Two step translation
NaLIR (2017)	General	✓		Need user manipulation
Seq2SQL (2017)	General			Low accuracy
This Study	Simple	✓	✓	Generate complex SQL

- We developed the prototype of the automatic part number answering system using open-source speech to text(Julius) engine and SQL generation from natural language on a mobile device such as iPad®.
- Functional evaluation shows that 83% test questions can reach correct part numbers. We suppose that “ask back” mechanism helps users to reach the correct answer. However, noisy environment affects accuracy of voice recognition.
- Performance evaluation shows that our application outperforms the conventional keyword search using keyboard 3.86 times faster in average.
- In future work, we are trying to tune the speech to text against noisy environment.

HITACHI
Inspire the Next 