Innovative Speech Applications for Contact Center Operations

Yoshiaki Noda†, Masafumi Tamoto, Tsubasa Shinozaki, Katsutoshi Ohtsuki, Tetsuo Amakasu, and Yoshikazu Yamaguchi

Abstract

This article introduces three speech applications that greatly improve contact center operations—speech recognition-based automatic document retrieval, speech recognition-based support for generating customer interaction reports, and a synthetic speech-based agent for delivering standard content to callers. The effectiveness and potential of these speech applications are also described.

1. Introduction

The chief issues facing contact centers—the point of contact between a company and its customers—are to (1) improve the quality of interactions with customers and (2) improve the efficiency of customer contacts while minimizing any loss of quality and maintaining a high standard of customer satisfaction [1], [2]. The first line of customer support is provided by customer service representatives (CSRs), who deal directly with customers on the telephone. CSRs are charged with receiving calls from customers, maintaining a cheerful positive disposition, quickly figuring out and verifying the gist of the customer's query, and finally providing an appropriate response or solution to the customer's problem. It is common operating procedure once a customer interaction is completed to prepare an *interaction report* that provides a brief description of the interaction, the content of the customer's call, the content of the interaction, any views or opinions expressed by the customer, and so on. Once this sequence of tasks is completed, the CSR then moves on to the next call.

To put the customer's mind at ease and instill trust, skilled CSRs should maintain a cheerful and supportive attitude while talking to the customer and of course must have the knowledge to respond to and solve the customer's problem [2]. The complexity and degree of knowledge that the CSR is expected to have vary greatly depending on the industry and the products and services supported, but the general trend is toward more diversified and functionally sophisticated products and services, and this makes the products themselves as well as the CSR's job more complex [2]. The network services supported by NTT are a case in point. At one time, NTT provided only basic telephone services, but now the company's billing systems are far more diversified, optional services have proliferated, and convergence with IP network services has led to configurations of equipment on customers' premises that are more diversified and functionally advanced. Thus, the range and complexity of operational knowledge that CSRs must be familiar with has vastly increased.

Experienced CSRs are remarkably efficient and adept at multitasking. They are able to retrieve the information they need (or expect to need) and begin writing a customer interaction report even while listening to the customer's inquiry. However, novice CSRs still in training are inefficient at navigating the terminal and lack the operational knowledge to respond to the issues raised. Supervisors charged with overseeing CSRs and ensuring smooth, efficient contact center operations generally have to assist novice CSRs on a one-to-one basis to enable them to provide the right answers and gradually acquire the hands-on operational knowledge that they need to become experienced CSRs.

† NTT Cyber Space Laboratories
Yokosuka-shi, 239-0847 Japan
Contact: https://www.ntt.co.jp/cclab/contact/index.html

2. Improving operations with speech applications

This article gives an overview of three speech applications that exploit recent advances in speech recognition and speech synthesis technologies to assist CSRs and contact center supervisors deal more effectively with customer interaction operations.

Speech recognition technology can be used to convert the dialog between customer and CSR into text data. The text data derived from the conversation can then be used to help generate customer interaction reports much more efficiently and automatically retrieve the exact information needed by the CSR to handle the inquiry. For some time now, we have had the ability to store and play back conversations between a customer and CSR recorded by an ordinary voice recorder, but this new ability to apply speech recognition to the actual content of conversation enables actions supporting CSR operations in all kinds of ways to improve contact center operations as never before. Using NTT's speech recognition engine VoiceRex, which was modified specifically for application to contact centers [3], we have achieved the high level of speech recognition performance required to implement these kinds of advanced contact center operation enhancements.

Moreover, state-of-the-art speech synthesis technology is now capable of generating sufficiently intelligible speech from any string of text. This has the potential of freeing up an enormous amount of time that CSRs spend giving out routine information by letting synthetic speech applications convey standard information and messages. For example, by using NTT's advanced speech synthesis engine Cralinet [4], we have now achieved remarkably natural-sounding synthetic speech that certainly has high enough quality to convey standard information and directions to customers who call contact centers.

Speech applications based on speech recognition and speech synthesis thus have the potential to reduce the burden on CSRs and supervisors who handle customer calls and markedly improve customer contact efficiency. This greater efficiency will not only help reduce the operating costs of contact centers, but also reduce the call duration of each customer contact, which could help shorten the time that a customer has to spend waiting on the phone. These applications also promote a better interactive experience for customers because they deliver accurate information more smoothly. We expect these speech applications to be especially useful for improving contact center operational efficiency when the CSRs are relatively inexperienced and when the range of operational knowledge supported by CSRs is complicated and diverse.

2.1 Automatic document retrieval

Automatic document retrieval uses speech recognition to automatically retrieve and display documents that are needed by the CSR based on the dialog in progress between the CSR and the customer. In particular, product support contact centers or contact centers covering an immense range of different products need CSRs to retrieve the correct documents dealing with specific issues from vast FAQ (frequently asked questions) or product databases. However, novice CSRs have a very difficult time thinking and fumbling with the keyboard trying to locate the right documents while talking to the customer. It is here that automatic document retrieval can improve the efficiency of customer interactions by reducing the burden on CSRs.

Automatic document retrieval works by using speech recognition to convert the dialog between customer and CSR to text, identifying keywords from the text that relate to the customer's problem, and then automatically selecting documents that address this problem. As illustrated in **Fig. 1**, the relevant documents are displayed along with the corresponding keywords in the window of an application developed for the purpose.

A narrowed-down shortlist of search results is automatically retrieved and displayed while the agent is talking to the customer, so the CSR can very easily open the document dealing with the problem at hand by merely selecting it. Even a relatively inexperienced CSR can quickly locate all the documents pertaining to the problem. Giving CSRs the ability to quickly retrieve relevant documents significantly reduces the time that customers are kept on hold while CSRs confer with supervisors and perform other time-consuming activities.

2.2 Support for generating customer interaction reports

This application reduces the time needed to create customer interaction reports by utilizing speech recognition results derived from the dialog between customer and CSR. One can imagine a number of situations where CSRs might need to generate a report such as (1) writing up claims, (2) writing up exemplary customer interactions for CSR training purposes, and (3) writing up customer interaction activities reports at the end of the day.

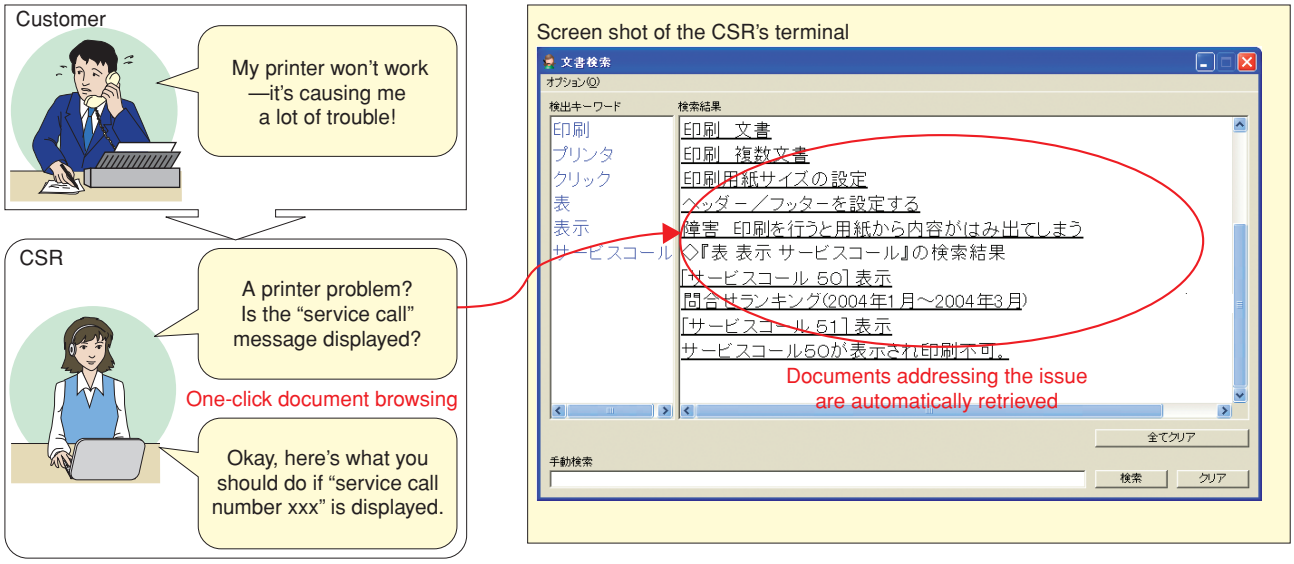

Fig. 1. Automatic document retrieval based on speech recognition.

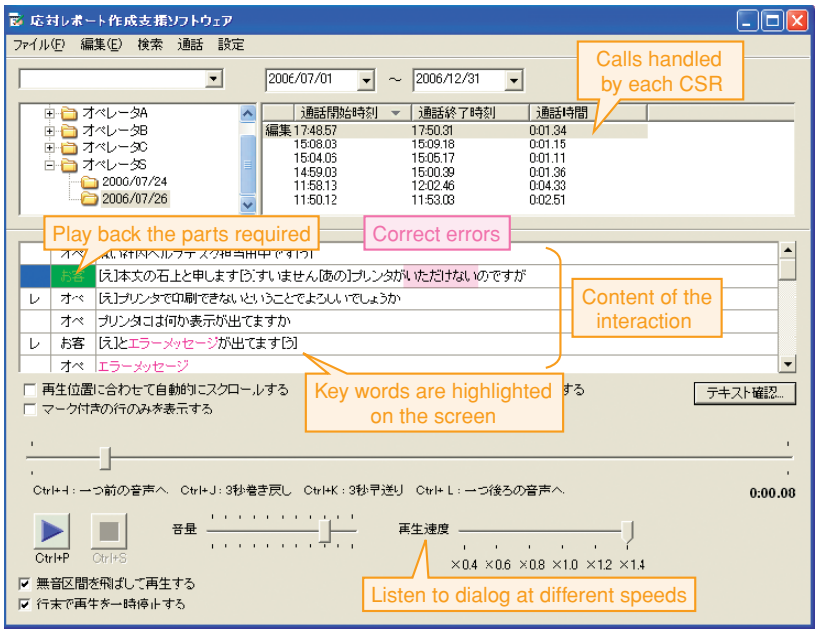

Fig. 2. Speech recognition-based support for generating customer interaction reports.

When writing up claims and training examples, the reports must be fairly detailed in order to convey what actually transpired during the call. Until now these reports have always been prepared starting from zero by replaying the call interaction that was recorded on the voice recorder. The support tool for generating customer interaction reports markedly improves the efficiency of this task.

Through simple trials performed at NTT Laboratories we found that using the new report generation support tool slashed the time needed to prepare the reports by half compared with the old method of first listening to recorded conversation and then writing the reports from scratch. A screen shot of the report generation input window that we developed to speed up the report generation process is shown in **Fig. 2**.

Essentially, the operator selects the parts of the interaction with the customer that need to be included in the report and the speech recognition results for those specific parts are then displayed on the screen.

Sometimes speech recognition results are wrong, but in such cases the dialog corresponding to the text can be played back. So by quickly perusing the entire interaction dialog and playing back just the parts that will be included in the report, the CSR can verify the words and correct them if necessary.

During customer calls, there are long pauses when neither the CSR nor the customer is speaking. Our support tool automatically skips over periods of silence. Moreover, the playback speed of the speech can be varied to enable the CSR to move more quickly through long sections of dialog or slow down for parts that are hard to understand. The system can also highlight special terms that have been registered in advance. The highlighted keywords stand out on the screen, so one can quickly grasp the overall gist of a customer interaction by just perusing these keywords. This function can also be used to identify in-house jargon in the dialog that perhaps should be avoided when talking to customers.

Finally, CSRs are often required to write up customer interaction activities reports detailing the general nature of customer queries and the responses and customer suggestions and opinions. Preparing such a report generally involves going through the day's contacts, picking out the highlights, and editing these disparate details into a brief summary. When CSRs deal with long customer calls or multiple inquires, they can easily omit important things when they rely just on their memories. Our tool can help by serving as notes to jog the memory.

2.3 Synthetic speech-based agent for delivering standard content

Using speech synthesis instead of the operator to deliver standard information and messages would save an enormous amount of time that CSRs currently spend talking to callers. A typical CSR-customer interaction can be divided into two parts: First, the CSR figures out what the customer needs or is asking and then provides the information that answers the customer's question or solves the customer's problem. Grasping the nature of a customer's inquiry can probably be handled more smoothly by a human agent than by a computerized system such as IVR (interactive voice response), but the response provided to the customer is often a fixed message about an event or campaign or some other standard information.

The purpose of this application is to reduce the call duration—the time required for the CSR to deal with a customer inquiry—by using a synthesized speech system to deliver the answer or other information to the customer rather than the human operator, as illustrated in **Fig. 3**. Indeed, this approach has already been implemented in the directory assistance service (dial: 104), where the operator takes the call and the number requested by the customer and then turns the call over to a synthetic voice system to actually say the number. Although the 104 directory assistance service has elementary speech synthesis capabilities for combining a limited set of predetermined words and numbers, the system we introduce here uses far more advanced speech synthesis technology that is capable of articulating a detailed synthetic voice response from any text. One of the big advantages of synthetic speech is that the content of a message can

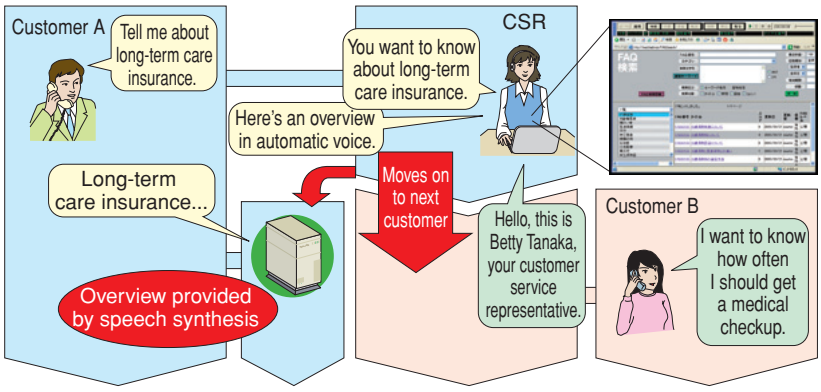

Fig. 3. Flow of synthetic speech-based agent for delivering standard content.

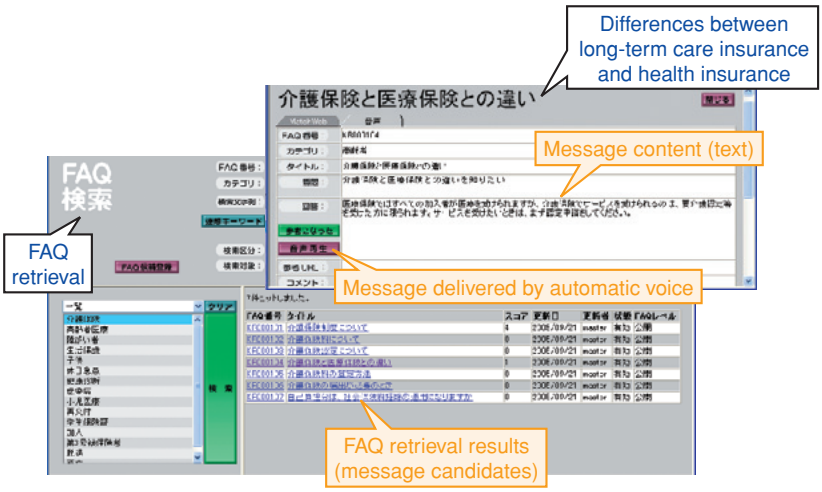

Fig. 4. Screen for operating a synthetic speech-based agent for delivering standard content.

be modified immediately simply by changing the content of the text message. This enables the content of a message to be replaced or changed instantly in the event of an emergency.

The quality of text-based synthesized speech in the past was inadequate for use in contact centers, but we have recently achieved synthesized speech that sounds a lot like a natural voice, a development that now makes this kind of application practical for contact centers. An actual implementation is the event ticket information service provided by Lawson Ticket Inc. based on a joint initiative by NTT Communications and NTT IT. This service has already demonstrated its worth by dramatically reducing the time CSRs spend dealing with customers [5].

Figure 4 shows a screen shot of a new speech synthesis application combining the *MatchContactSolution* system [6] from NTT AT and *ADVICE C3* system [7] from NTT IT. With this system, the CSR uses his or her terminal to retrieve FAQ data managed by the *MatchContactSolution* system. Then, with the customer's consent, one click of a button causes the content to be read to the customer in synthesized speech.

3. Conclusion

The speech applications introduced in this report are currently being evaluated in actual contact centers to quantitatively assess the impact they have on contact center operations. While continuing to assess and collect in-service performance results at contact centers, we also plan to further refine these applications and deploy them much more extensively at NTT Group contact centers.

It is clear that speech applications have tremendous potential to improve contact center operations. With its vast resources and technological expertise in ergonomics, retrieval systems, knowledge processing, and other relevant areas, NTT Laboratories is committed to further enhance these applications and combine them in novel and imaginative ways to obtain even better economical and performance results.

References

[1] C. Hishinuma, "All About "Call Centers," Continued," Ric Telecom, p. 106, 2001 (in Japanese).
[2] "Contact Center Whitepaper 2005," Ric Telecom, pp. 43, 68, 72, 92, 2005 (in Japanese).
[3] H. Masataki, D. Shibata, Y. Nakazawa, S. Kobashikawa, A. Ogawa, and K. Ohtsuki, "VoiceRex—Spontaneous Speech Recognition Technology for Contact-center Conversations," NTT Technical Review, Vol. 5. No. 1, pp. 22-27, 2007 (this issue).
[4] K. Mano, H. Mizuno, H. Nakajima, N. Miyazaki, and A. Yoshida, "Cralinet—Text-To-Speech System Providing Natural Voice Responses to Customers," NTT Technical Review, Vol. 5. No. 1, pp. 28-33, 2007 (this issue).
[5] http://www.ntt.co.jp/news/news04/0405/040531.html (in Japanese).
[6] http://www.matchcontact.jp/ (in Japanese).
[7] http://www.ntt-it.co.jp/product/cat3.html (in Japanese).

Yoshiaki Noda
Senior Research Engineer, Speech, Acoustics and Language Laboratory, NTT Cyber Space Laboratories.
He received the B.E. and M.E. degrees in electrical engineering from Ritsumeikan University, Kyoto, in 1989 and 1991, respectively. Since joining NTT in 1991, he has been working on R&D of automatic speech recognition technologies. He is currently developing speech applications for contact center operations.

Masafumi Tamoto
Senior Research Engineer, Speech, Acoustics and Language Laboratory, NTT Cyber Space Laboratories.
He received the B.E. degree in physical electronics and the M.E. degree in systems science from Tokyo Institute of Technology, Tokyo, in 1991 and 1993, respectively. Since joining NTT in 1993, he has been working on R&D of speech understanding and dialog system technologies. He is a member of the Information Processing Society of Japan and Acoustical Society of Japan (ASJ). He received the Awaya Prize from ASJ in 1999.

Tsubasa Shinozaki
Research Engineer, Speech, Acoustics and Language Laboratory, NTT Cyber Space Laboratories.
He received the B.E. and M.E. degrees in design from Kyushu Institute of Design, Fukuoka, in 1994 and 1996, respectively. Since joining NTT in 1996, he has been engaged in R&D of speech synthesis. He is currently developing applications based on speech synthesis and speech recognition systems. He is a member of ASJ.

Katsutoshi Ohtsuki
Research Engineer, Speech, Acoustics and Language Laboratory, NTT Cyber Space Laboratories.
He received the B.E. and M.E. degrees in electrical engineering from Waseda University, Tokyo, in 1993 and 1996, respectively. Since joining NTT in 1996, he has been working on R&D of automatic speech recognition technologies. From 2004 to 2005, he was a visiting scientist at BBN Technologies, MA, USA, working on statistical language modeling. He is a member of IEEE, the Institute of Electronics, Information and Communication Engineers of Japan, and ASJ. He received the Awaya Prize from ASJ in 2004.

Tetsuo Amakasu
Research Engineer, Speech, Acoustics and Language Laboratory, NTT Cyber Space Laboratories.
He received the B.E. degree in communication engineering and the M.E. degree in computer and mathematical sciences from Tohoku University, Miyagi, in 1997 and 1999, respectively. Since joining NTT in 1999, he has been working on R&D of automatic speech recognition technologies. He is a member of ASJ and the Japanese Society for Artificial Intelligence.

Yoshikazu Yamaguchi
Voice System Division, NTT IT Corporation.
He received the B.E. and M.E. degrees in electrical engineering from Osaka Prefecture University, Osaka, in 1993 and 1995, respectively. He joined NTT Human Interface Laboratories in 1995. Since 2005, he has been working on speech recognition applications for contact centers at NTT IT Corporation. He is a member of ASJ. He was transferred to NTT IT Corporation in June 2005.