Advanced Speech Processing Technologies for Contact Centers

Satoshi Takahashi† *and Yoshiaki Noda*

Abstract

This article introduces several applications that reflect the latest advances in speech processing technologies developed to support operations at contact centers and describes some of the underlying technologies that support these developments.

1. Introduction

The era in which companies uniformly provide common products and services for the majority of people is rapidly coming to an end. A more customer-oriented strategy has become much more important. Companies seek to determine the needs and interests and past buying behavior of individual customers and then sell products that are closely tailored to the preferences of those customers. To accommodate the more diversified needs of consumers, companies have endeavored to differentiate themselves and beat their competitors by diversifying their own products and services, offering a wider array of products and making their products and services more advanced and user friendly. Amid these developments, the contact center[*1] plays a critical role in collecting customer data that provides rapid feedback for developing new products and services and has emerged as a key element of customer-oriented one-to-one marketing strategies.

As products and services become more diversified, questions and requests from customers become correspondingly more complex. In after-sales followup, the ability to respond to these requests with cordial and appropriate answers and new information has also emerged as an important corporate strategy for improving customer satisfaction. In short, "timely provisioning of information wanted by customers and timely identification of customer trends" are indispensable. The contact center is essentially the face of the company. It is the point of contact that consumers have with the company and a site where good communication is critical. To a large extent, the contact center determines a company's image and perceived quality. Consequently, an increasing number of companies are expanding and enhancing their contact centers by improving hardware and software capabilities. Moreover, considering the need for migration to an IP (Internet protocol) infrastructure, the contact center market today is growing at a rate of about 5% per year and is projected to have a market value of close to ¥400 billion in Japan by the year 2008.

2. Challenges for contact centers

With inquiries from customers becoming more diversified as services become more diversified, the task that customer service representatives (CSRs)[*2] face when interacting with customers has become extremely complex. CSRs require both training and skills development. Moreover, they must also acquire basic knowledge and skills in how to comport themselves when dealing with customers. Administrators are responsible for ensuring that the large number of CSRs working for the contact center meet uniformly high standards in the quality of their interactions with customers. This is especially challenging in today's

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

*1 Contact centers: Since call centers often support multiple channels (email and the Web in addition to telephone and fax), they have come to be referred to as contact centers.

*2 CSRs: People responsible for interacting with customers. They are also referred to as operators, communicators, agents, and so on.

environment because more companies have outsourced their contact center operations and the turnover rate among CSRs is growing because of the high workload. This has led to a high rate of new employees and temporary workers at centers, which makes the task of problem solving much more difficult.

The contact center depends heavily on human service workers. Indeed, it is estimated that approximately 60–70% of a center's operating costs are labor costs. How to reduce them is one of the top priorities for administrators. While reducing the number of CSRs or taking other shortsighted cost-cutting measures would reduce costs, such approaches are ineffectual because they always lead to reduced customer satisfaction. Administrators are thus faced with the seemingly impossible challenge of improving customer satisfaction and service quality while simultaneously reducing operating costs.

3. Improving CSR productivity

Labor costs are essentially the sum of CSR salaries multiplied by the number of CSRs employed at each salary working at a contact center. If one tries to reduce labor costs by simply cutting the workforce or changing the balance between higher-paid higher-grade workers and lower-paid lower-grade workers, this only diminishes the quality of service provided by the contact center as a whole, which leads to reduced customer satisfaction. On the other hand, outsourcing to locations with lower salaries has already be fully exploited by many companies and also has its own drawbacks. The only practical solution is to improve the productivity of each individual CSR by reducing the time needed by the agent to deal satisfactorily with each inquiry. In this context, times measured in mere seconds can have a tremendous overall impact when multiplied by the huge number of inquiries being handled. How might such time savings be achieved? Attempting to reduce the call duration by simply accelerating the pace or by reducing the time for data input would only increase the burden on CSRs. It would put stress on the CSRs, causing them to rush the interaction and not focus full attention on the customer's problem, which would very likely lead to reduced customer satisfaction. The only way to truly improve individual productivity is to reduce the workload on employees. This applies not only to the CSRs themselves, but also to the supervisors in charge of overseeing contact center operations.

4. Speech processing technologies supporting more efficient operations

NTT Laboratories has been researching speech processing technologies—speech recognition and speech synthesis in particular—for many years and has a vast amount of expertise in this area. In this Special Feature, the second article introduces some of the latest developments in these technologies for supporting operations at contact centers, focusing on applications that can help to reduce CSR workload and the third and fourth articles describe some of the basic technologies underlying these applications.

There have been instances in the past where speech recognition and speech synthesis have been applied in the area of computer-telephony integration (CTI). The classic example is their application to interactive voice response (IVR). An IVR system with speech recognition capability enables a customer to select an item from a menu without having to click a button by recognizing the category (word) uttered by the customer. Another advantage of speech-synthesis-enabled IVR is that the content of messages can be very easily changed because the response messages are produced by speech synthesis. The effectiveness of this system has now been significantly improved so you can get realtime information about stocks and other information from a voice portal[*3]. Yet we must acknowledge the inherent limitations of using conventional speech recognition and speech synthesis. This is because conventional speech recognition requires you to enunciate very clearly for the machine to understand, and speech synthesis is only capable of generating very mechanical robotic-sounding speech.

Contact centers are based on person-to-person communication between a CSR and a customer. It is interesting that there has recently been a shift away from human-machine interaction to human-human interaction as the target for speech recognition technology. In typical spontaneous speech between people, the speech is often rapid fire and filled with ambiguous phrases that make it difficult for the system to recognize. Yet we can anticipate a vast increase in the situations where this kind of service will be used in the years ahead. Even when 100% recognition accuracy cannot be achieved, the range of situations where speech recognition might be used is

*3 Voice portal: a web site or other service that can be accessed entirely by voice. Voice portals commonly provide information in the form of recorded or synthesized speech to users via the telephone.

extensive. And in the area of speech synthesis as well, we have now been able to achieve very-high-quality synthesized speech, approaching that of a professional announcer, by combining high-speed lookup that concatenates speech segments appropriately using high-speed signal processing on a high-quality, large-scale speech database platform. This corpus-based speech synthesis produces very natural-sounding high-quality speech and is already being used in a range of commercially important applications.

5. Areas of potential application to contact centers

Speech processing technologies have enormous potential for applications that could support contact center operations. Some of the services that could be developed to support next-generation contact centers are shown in **Fig. 1**. The second article in this Special Feature take a closer look at three potential applications.

(1) Automatic document retrieval by CSR speech recognition

While the CSR is dealing with a customer, speech recognition could be used to automatically look up and display documents such as answers to common questions, manuals, and previous interaction history to help the CSR handle the customer's inquiry. This capability would help the CSR stay focused on the customer's inquiry and would be particularly useful for CSRs who are new to the job.

(2) Support for generating interaction reports and initiating interaction logs

This could automatically initiate the process of generating interaction reports and logs by recognizing the conversation between the CSR and the customer. It would significantly reduce the post-processing time required after an interaction with a customer has ended.

(3) Issuing of standardized directions in synthesized speech

This would permit the CSR to select a standardized set of directions in synthesized speech after he or she has determined the nature of the customer's inquiry. It would work in the same way that calls for NTT directory assistance (dial: 104) are handled and would significantly reduce the interaction time.

Another potential application would be to convert call speech logs to text by using speech recognition, which would greatly facilitate later lookup. Currently when a problem occurs between the CSR and customer, the only proof is a recording of the call speech log, so a text version of the call speech log would be extremely useful for identifying and analyzing inherently valuable customer trends that could be used to create new value. And by analyzing the call speech

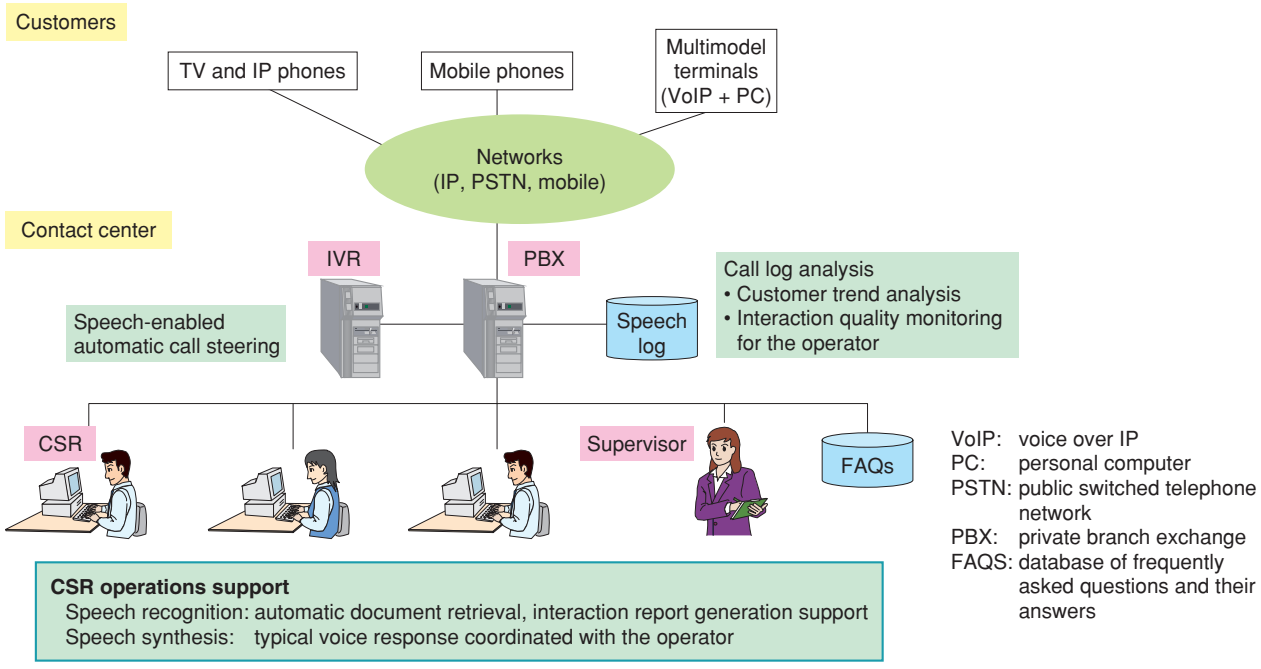

Fig. 1. Next-generation contact center support services.

logs, administrators would be able to better grasp the status and quality of CSR interactions. Enormous potential is seen in this kind of speech mining. Further out on the horizon, all sorts of self-contained services are envisioned such as spontaneous conversational IVR far surpassing conventional speech-recognition-based IVR that would automatically establish a connection to the appropriate department as soon as the caller starts speaking, and intelligent software agents that would assist customers locate information and solve problems on their own.

6. Conclusion

Performance data and quantitative assessment of actual deployment are essential, so we are currently moving ahead with field trials. In the years ahead, we expect to see even closer cooperation among commercial companies, development companies, and research laboratories leveraging their mutual product and problem-solving strengths, and this will lead to widespread deployments of applications based on the technologies outlined in this article.

Satoshi Takahashi
Senior Research Engineer, Supervisor, Speech, Acoustics and Language Laboratory, NTT Cyber Space Laboratories.
He received the B.E., M.E., and Ph.D. degrees in information science from Waseda University, Tokyo, in 1987, 1989, and 2002, respectively. Since joining NTT in 1989, he has been engaged in speech recognition and pattern recognition. He is a member of the Acoustical Society of Japan and the Institute of Electronics, Information and Communication Engineers of Japan.

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.