Advanced Research in Speech, Language, and Hearing for Communication of the Future

Eisaku Maeda

Abstract

Research at NTT Communication Science Laboratories draws on both information science and human science with the aim of building a new technical infrastructure that will connect humans and information. These Feature Articles introduce new trends in the fields of speech, language, and hearing, which have a relatively long history of basic research.

Keywords: communication science, basic research, information science

1. Introduction

As we move forward in the 21st century, we are witness to the dramatic changes occurring at a truly amazing pace in the information environment that surrounds us in our daily lives. This phenomenon is clearly reflected by the transition in well-known keywords over the last ten years, for example, from *ubiquitous, grid*, and *sensor network* to *Semantic Web, Web 2.0, cloud,* and *big data*. Similarly, the information devices used to access networks have migrated from mobile phones and desktop computers to smartphones and tablets, and the range of users has expanded to include children and the elderly. These developments have also changed the face of provided services and generated a need for research and development tailored to these changes in the environment.

At NTT Communication Science Laboratories, we seek to build a new technical infrastructure that connects humans with information amid these dramatic changes in the information environment. In contrast to service development, which seeks to meet current needs, basic research aims to bring about technical innovations from a medium- and long-term viewpoint. However, as the pace of change accelerates, the strategies used to advance basic research must also change. NTT Communication Science Laboratories promotes research in a variety of scientific fields in information science and human science. These can be broadly divided into four areas: signal processing, media processing, computer intelligence, and human science (**Fig. 1**). Of particular interest here is that our successes in recent years have almost without exception combined multiple fields and technologies in a synchronized and skillful manner. This outcome can be seen in both scientific fields and service development. It is safe to say that each and every researcher in the upcoming era will need to have a good background in multiple fields.

2. Cultivating trees that bear fruit

Minor paradigm shifts or the encountering of problems often become the seeds for new research, and if those seeds are given water, they will eventually sprout. If the resulting buds are then given fertilizer and exposed to sunlight, they will grow into trees, and on those trees, flowers representing patents, papers, and other achievements will bloom. Although flowers cannot normally be eaten, they give forth fruit that can be picked. However, this harvested fruit cannot always be eaten in its original state. There is hard, unripe fruit, sour fruit, and even poisonous fruit. There is also some fruit that must be prepared and cooked in various ways before being eaten while some fruit can be stored away and preserved for later
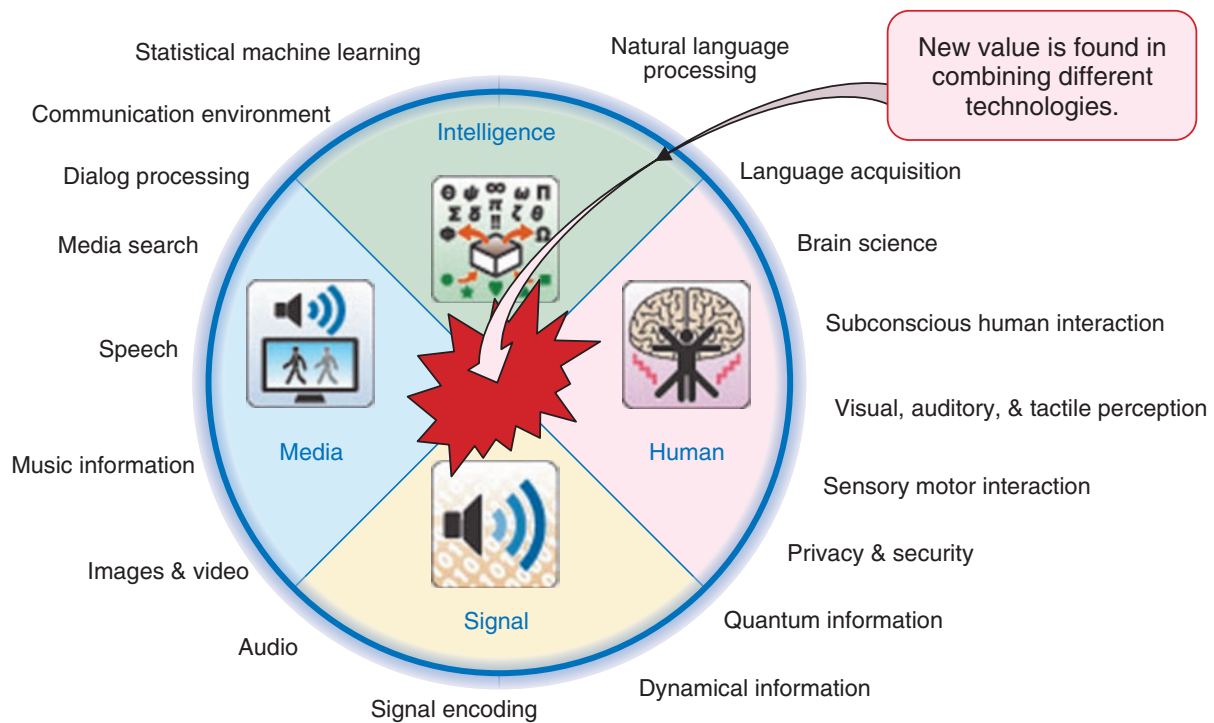
Fig. 1.   Research fields at NTT Communication Science Laboratories.

use.

In any case, just as fruit will eventually nourish people in one shape or another, we can treat the results of research as technology that will eventually serve a useful purpose in society. Here, trees signify research, and *cultivating trees that bear fruit* is the most important role of basic research. It's been more than 20 years since the founding of NTT Communication Science Laboratories, and our technologies that have found a place in society have been increasing slowly but surely. If we look at examples of our successes in recent years in areas like media search, speech recognition, reverberation control, question answering, statistical translation, and *texture* information science, we can see that a period of about ten years is needed after sowing the seeds before any fruit will be ready to consume.

3.   Advanced research in speech, language, and hearing

The five articles concerning speech, language, and hearing in these Feature Articles report on recent research achievements of NTT Communication Science Laboratories. These achievements hold various

positions in the research scenarios based on the fruit analogy described above.

Machine translation technology entitled "Recent Innovations in NTT's Statistical Machine Translation" [1] represents a genuine era of practical application after a long research history spanning more than 30 years. The accumulation of language resources and know-how through years of research as well as recent technical innovations lie behind this innovative development period.

The two articles entitled "Advances in Multi-speaker Conversational Speech Recognition and Understanding" [2] and "Speech Recognition Based on Unified Model of Acoustic and Language Aspects of Speech" [3] introduce the latest trends in speech recognition technology. We are now entering an era of practical speech recognition technology that will enable speech recognition to be used, for example, in preparing the minutes of proceedings in Japan's National Diet. The speech recognition field, however, still has some issues that must be solved depending on the usage environment and application. Here as well, the combination of multiple technologies will give rise to new technologies that have a competitive advantage, and the germination of these technologies

has already started.

"Speaking Rhythm Extraction and Control by Non-negative Temporal Decomposition" [4] is an achievement born of research into the mechanism of human speaking. The human voice is generated as a sound originating in the vibration of speech organs, and we are working on fascinating developments by combining that process with information-science technologies related to speech processing.

"Link between Hearing and Bodily Sensations" [5] introduces research that was the first in the world to unravel the relationship between human bodily sensations and the sense of hearing. The so-called flowers are finally blooming because of this research, and we look forward to seeing what kinds of fruit these flowers will bring forth.

References

[1] M. Nagata, K. Sudoh, J. Suzuki, Y. Akiba, T. Hirao, and H. Tsukada, "Recent Innovations in NTT's Statistical Machine Translation," NTT Technical Review, Vol. 11, No. 12, 2013.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr2013 12fa2.html

[2] T. Hori, S. Araki, T. Nakatani, and A. Nakamura, "Advances in Multi-speaker Conversational Speech Recognition and Understanding," NTT Technical Review, Vol. 11, No. 12, 2013.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr2013 12fa3.html

[3] Y. Kubo, A. Ogawa, T. Hori, and A. Nakamura, "Speech Recognition Based on Unified Model of Acoustic and Language Aspects of Speech," NTT Technical Review, Vol. 11, No. 12, 2013.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr2013 12fa4.html

[4] S. Hiroya, "Speaking Rhythm Extraction and Control by Non-negative Temporal Decomposition," NTT Technical Review, Vol. 11, No. 12, 2013.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr2013 12fa5.html

[5] N. Kitagawa, "Link between Hearing and Bodily Sensations," NTT Technical Review, Vol. 11, No. 12, 2013.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr2013 12fa6.html

Eisaku Maeda
Director, NTT Communication Science Laboratories.
He received the B.E. and M.E. degrees in biological science and the Ph.D. degree in mathematical engineering from the University of Tokyo in 1984, 1986, and 1993, respectively. He joined NTT in 1986. He was a guest researcher at the University of Cambridge, UK, during 1996–1997. His research interests are in statistical machine learning, intelligence integration, and bioinformatics. He is a senior member of IEEE and a member of the Institute of Electronics, Information and Communication Engineers and the Information Processing Society of Japan.