

R&D Spirits

Toward High-quality and High-reality Teleconferencing

Yoichi Haneda

Senior Research Engineer, Supervisor
Speech, Acoustics and Language Laboratory
Acoustic Information Processing Group
NTT Cyber Space Laboratories



The rapid growth of IP (Internet protocol) networks has placed new demands on teleconferencing systems. NTT Cyber Space Laboratories has long been researching and developing teleconferencing systems focusing on an integrated type that combines a speaker and microphones in one unit. These efforts have produced a 5th-generation teleconferencing product released at the end of last year. To learn more about the technical features of this product, its acceptance by the market, and future R&D goals, we sat down with senior research engineer and project director Yoichi Haneda.

Development of hands-free multipoint teleconferencing systems

—Dr. Haneda, could you tell us about your current research endeavors?

Our group is currently researching elemental technologies and developing products for hands-free teleconferencing systems. The development of teleconferencing systems that use speakers and microphones did away with the older handset-style of teleconferencing. These systems enabled users to manipulate

their computers, operate a projector, and perform other tasks while carrying on a conversation, and enabled multi-participant meetings to be held. All in all, they made teleconferencing even more convenient. At the same time, replacing telephone receivers with speakers and microphones generated acoustic issues that did not previously have to be dealt with. One of these is ambient noise and its reduction. Ideally, a microphone in a teleconferencing system will pick up only human speech, but in reality, an increasing variety of peripheral sounds will be picked up as microphone performance improves (**Fig. 1**), such as

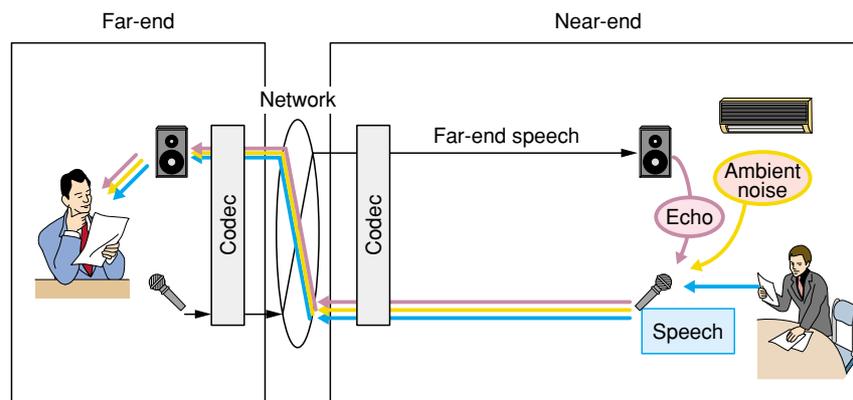


Fig. 1. Problems of echoes, ambient noise, and volume for video and/or teleconferences.

noise produced by fans in laptop computers and projectors or the sounds of air-conditioning equipment. Such sounds might not be noticeable under ordinary circumstances, but they can be bothersome during a teleconference. For this reason, one of our goals has been to remove the noise component from a state characterized by a mixture of speech and noise. Another issue is echoes and their cancellation. During a remote conversation using microphones and speakers, the voice of a person talking at one end is reproduced by the speaker at the other end and in turn picked up by a microphone at that end and returned to the original talker via the network. A device for preventing this is called an echo canceller. We have developed technology for suppressing both noise and echoes simultaneously. The third acoustic issue here concerns volume control. In ordinary signal processing, speech produced near a microphone is reproduced more loudly than that produced far from a microphone. An ideal teleconferencing system, however, would enable the voice of any participant to be heard at a fixed volume level. Another of our goals has therefore been to control the speech volume of each participant in a teleconference.

These are our main acoustic-related issues in our development work. Some of us are also researching and developing speech codecs, and with these researchers included, our group consists of 17 members. As one of these members, my main role is to provide direction for the entire project, but I am also involved in audio signal processing research and the modeling of room transfer functions between microphones and speakers.

—What are the main technical features of this research?

In noise reduction, the continuous sounds emanating from an air conditioner and computer fan are treated as noise signals when stored, and they are suppressed by subtracting similar signals. An echo canceller, meanwhile, predicts the sounds that would be conveyed from a speaker to a microphone and subtracts the predicted echo replica from the microphone input signal. In principle, this is a relatively simple process, but its implementation in product form requires know-how such as in the appropriate setting of parameters. To this end, we have made good use of research results accumulated over many years. For volume control, our approach is to detect the direction of each participant's voice using multiple microphones and to adjust the volume in each of those

directions using a specially designed algorithm. In this way, we can reproduce a loud nearby voice as is, while making adjustments to only those voices that are relatively far and subdued (**Fig. 2**).

—How is your research progressing?

Our 5th-generation unit went on the market at the end of last year (**Fig. 3**). This new IP (Internet protocol) teleconference unit, called the MB-1000, offers the latest technology in noise reduction, echo cancellation, and volume control (automatic gain control). It also features an improved sound-collection function. While previous units required participants to speak within a radius of one meter from the equipment, the MB-1000 can adjust and convey voices within a range of three to four meters all at a one-meter level of volume. The MB-1000 also incorporates a 7-kHz speech codec, which provides especially clear sounds compared with ordinary narrowband telephone communications. Moreover, MB-1000 enables interconnection between the telephone network and the Internet. Last but not least, it can simultaneously connect four locations, enabling four-point teleconferencing using only these terminals. We are proud to say that the MB-1000 represents a quantum leap in ease of use compared with past teleconferencing systems.

—What future issues do you foresee, and how do you plan to deal with them?

While we have definitely made great strides in improving performance on various levels, there are still a number of problems that must be solved if we are to achieve a teleconferencing system that is closer to ideal. For example, the MB-1000's functions for automatic volume adjustment for small voices and echo suppression are actually at odds with each other from a technical point of view. In truth, part of this system is self-conflicting. This is the first problem that we must solve.

Also, as an example of improving a basic teleconferencing function, we must make it easier for participants to understand the speech of other participants. This could be achieved by bringing the sound-collection component of the system closer to each user such as by using clip-on microphones or headset-type microphones. But considering that users do not normally attach devices to themselves in ordinary meetings, this solution would probably be met with some psychological resistance. For this reason, the approach that we find most desirable is to use micro-

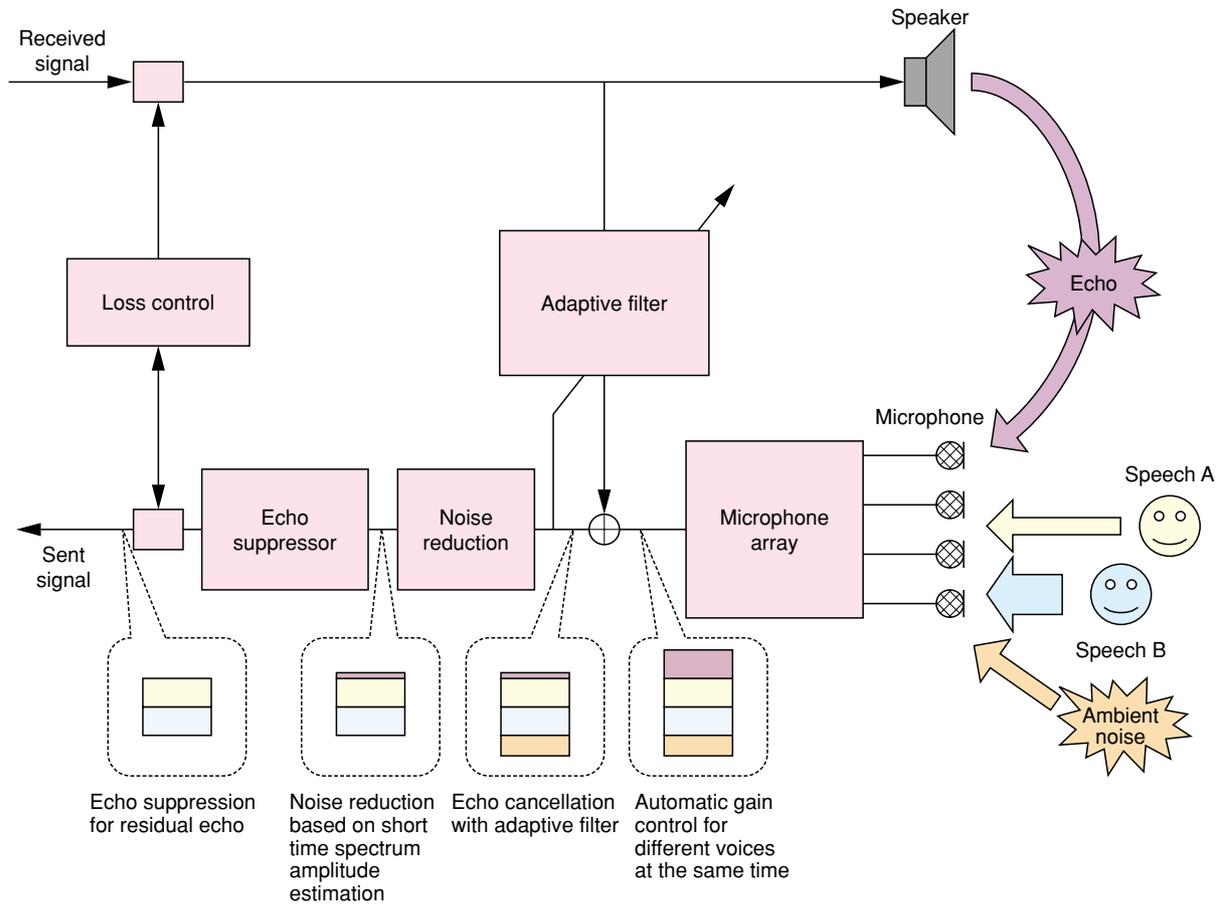


Fig. 2. Block diagram of the latest acoustic echo canceller.



Fig. 3. New IP teleconference unit.

phones that are built into the equipment and to handle any reverberation that might occur by signal processing.

Next, as a long-term research theme, we want to find a way of giving users a high-reality experience when participating in a teleconference. Let me tell you why. Strange as it may sound, trains and airplanes can be regarded as competitors to teleconferencing systems. While people have traditionally

moved from one place to another by trains, airplanes, and cars etc. to hold face-to-face meetings, telecommunication equipment has enabled people to hold meetings without having to travel. As a result, achieving a sense of presence in which participants feel as if they are all present at the same place is a very important function, and a desire for this function has, in fact, been expressed by users. On the other hand, the means of achieving a high-reality environment certainly differs between the bidirectional world as in teleconferencing and the unidirectional world as in music and movies. Most teleconferencing systems in use today use only one speaker, but there are still some doubts as to whether the reproduction of three-dimensional speech simply by using multiple speakers would enable a meeting to proceed smoothly. On the contrary, such a configuration might even disturb a user's concentration. Such attempts at achieving high-reality teleconferencing must be tested and the meaning of high-reality in the context of teleconferencing might also need to be redefined.

Meeting user expectations with the high-performance MB-1000

—*Could you tell us about R&D and market trends in teleconferencing equipment in Japan and overseas?*

Well, in terms of competing equipment, products supplied by U.S. manufacturers have the largest market share at present. These products have become well-known by virtue of their good performance and availability in many countries throughout the world. Here in Japan, a number of major electrical equipment manufacturers have been involved with teleconferencing equipment for some time, and in the last year, two new companies have entered the field.

You might therefore wonder where exactly does our equipment fit among all the products that are being offered. I think it's fair to say that our equipment is the most advanced in terms of basic functions. We are also proud of its excellent cost-performance characteristics. Much to my regret, however, our equipment is hardly as well known as that of U.S. manufacturers. To rectify this, we have been aggressively promoting our products both within and outside the NTT Group by holding exhibitions whenever an opportunity arises. For example, last September we exhibited them at the NTT Collection held by NTT West in Osaka, Fukuoka, and Nagoya, and in December, we displayed them at the NTT Group Communication Expo. At events such as these, it is common to exhibit equipment as company products, but we displayed ours in the technology corner as research achievements of NTT Laboratories. This appeared to attract greater interest in our products.

—*What kind of response did you receive from both Japanese and overseas visitors to these exhibits?*

The response was better than we expected! And our improvement of the sound-collection function from one to three-to-four meters was particularly well received. Specifically, the capability of collecting sound from points relatively far away from the equipment without having to connect external microphones was admired, as was the capability of picking up even subdued voices and conveying short discussions among colleagues at one end in a realistic manner. We also received praise for the cost performance of our MB-1000 unit from people familiar with this field. But at the same time, I am sure that additional needs will be expressed once people begin to use the equipment in actual working situations. We plan to

use those needs as feedback in the development of next-generation equipment as a matter of course.

—*Are you involved in any collaboration with other companies or universities?*

Yes, we are. We have a cooperative relationship with NTT Advanced Technology to conduct surveys on market needs and to construct and evaluate prototype equipment. Our people and their people are constantly coming and going, and they have become an important partner for us. We have also been interacting with universities through activities at academic societies. In fact, many of our former colleagues now hold teaching positions at universities providing us with a wide range of contacts on an individual level. Right now, however, we are not involved in any collaborative research related to product development.

—*Could you tell us about any international activities that you might be involved in?*

In terms of individual technologies like echo cancellers and microphone arrays, the researchers in charge of those technologies make presentations at international conferences. As for myself, I am a member of the IEEE Audio Technical Committee.

Taking on diverse acoustic problems by a physical approach

—*Dr. Haneda, what was your university major, and what expectations did you have on entering NTT Laboratories?*

My main field of study was physics, and I researched solid-state crystal structures using lasers. I chose physics as my major from the very start because I believed that a knowledge of basic physical properties would prove valuable on whatever path I might take in the future. Fortunately, a lot of emphasis was being placed on basic research in many fields when I left graduate school in 1989, and physics majors were in great demand. Against this background, I chose NTT as I believed its research environment was far better than that anywhere else. The research of physical properties requires equipment on a large scale, and NTT has equipment and facilities that surpass anything that universities or even other companies have. That was especially appealing to me, so with my strong desire to pursue research on the basic nature of things, I entered the company.

—*What have been your main research themes up to now?*

After entering NTT, I was assigned for a while to the acoustic signal processing section of the Human Interface Laboratories. My job was to research methods of canceling echoes and to model transfer functions that could express the sound path from a speaker to a microphone. Then, in 1998 and 1999, I was in charge of consumer-oriented product development in the Communications Equipment Business Headquarters that later became NTT East. This meant, in essence, the development of ISDN terminal adaptors. At that time, ISDN was spreading rapidly, and with products changing every three months or so, it was a hectic time. Although it was hard work, the experience I gained there broadened my outlook considerably and has come to be immensely useful in my current work. During that time, I also obtained the Ph. D degree for transfer-function modeling. In 2000, I returned to NTT Laboratories, where I again undertook R&D work in the field of acoustics. Last year, I was placed in charge of developing a directional AGC (automatic gain control) echo canceller, MB-1000.

As you can see from my history, I have been involved with echo cancellers for some time, but the fact is, I do not feel entirely comfortable when I am called an expert in echo-canceller research. This is because the most renowned of the several prizes that I have received was for contributions made in directing research. Still, I might consider myself an expert in bringing together various technologies in one physical unit.

—*What has been your goal throughout your R&D activities?*

The acoustic signal processing that we deal with includes three basic elements: physical phenomena, signal processing, and human psychology. Among these, given that I majored in physics at university, I have focused on physical phenomena associated with sound transmission from a speaker to a microphone and have made the modeling of transfer functions between microphones and speakers, which I mentioned earlier, my lifework. At the same time, the world of signal processing is fascinating, and since it is a relatively new technology, it tends to receive lots of attention. Accordingly, in deciding what to pursue in acoustics, I have always thought I should look for undeveloped themes in this field whose solutions would prompt interest.

And, from a different point of view, another of my goals has been to build a foundation for myself as a researcher. To survive in the world of R&D, it is important to decide what one's forte is to be. It is also necessary to take on an area of study that no one can beat you at! I believe that I have pursued transfer-function modeling with the aim of laying a foundation as a researcher even more than out of simple interest.

Toward the ultimate high-reality experience focusing on human senses

—*Dr. Haneda, how would you like to expand your personal area of research?*

Well, for one thing, I would like to look a bit further into the psychological aspects of acoustics. I am particularly interested in explaining the sensation of hearing from both the physical and signal-processing perspectives. At present, I am approaching this problem by creating various types of signals and examining how they are heard by people. But I would also like to take this one step further and investigate what types of signals should be created to produce a certain type of hearing sensation. However, my present role is more of a “coach” than a “player,” and I find that I have less and less time to research my personal areas of interest. Nevertheless, if I am to coach others with confidence, and moreover, if I am to leave this role of coach one day, I must endeavor to build up my research expertise and achievements in a manner appropriate for a frontline player.

—*What is your ultimate dream as a researcher?*

Aside from realistic considerations such as what types of technologies should be used, my ultimate dream is to achieve a teleconferencing system that can provide a high-reality experience that makes remote participants feel like they are sitting at the very same table. This, of course, is a theme that has been talked about for some time, and many researchers even at NTT have approached this problem in various ways. My personal position is to approach this task from the viewpoint of solving acoustic problems in bidirectional communications. For this purpose, I believe that a realistic approach is to close the gaps between our ideal system and the present system in a step-by-step manner.

—*What has it been like for you working at NTT Laboratories?*

NTT Laboratories has been like a home to me—a place where I feel protected. Indeed, I think of it as a place where I can demonstrate my abilities to the fullest without restraint. In particular, the Acoustic Information Processing Group that I presently belong to has always been a very comfortable working environment. One reason for this is that each member of this group has expertise in core, elemental technologies that can be combined to create a single piece of equipment. Covering the range from basic research to prototyping, development, and commercialization is something that suits me well. Perhaps this is not such a big job when viewed from NTT Laboratories as a whole, but I strongly feel that it is work worth doing.

—*Dr. Haneda, what would you say to young researchers?*

I would like to ask those who are about to enter the world of R&D to ask themselves whether research is something that they really want to do. I say this for the following reason. There are many students in Japanese universities that, as part of the university educational system, simply enter a laboratory and take on whatever research theme is given to them as opposed to expressing a true desire to become a researcher. Such a state of mind can hardly lead to good research even if they enter a corporate research laboratory. True research is a process in which one finds a problem on one's own and attempts to solve it. Research that follows the whims of someone else or that becomes obsessed with details may have its merits as practice, but it is not something that a researcher should fall into. I would therefore like to see young researchers pursue a research area different from those of others. In this sense, I would say “Don't take

the same approach as someone else” to young researchers.

In addition, I would like young researchers to place particular importance on ideas and approaches. While there has been a tendency in recent years to place emphasis on results, it cannot be said that results are everything in R&D. For example, achieving a two-fold jump in performance simply by using a higher computer processing speed has no particular value from an R&D perspective. What really determines value in the R&D process is insight and approach. The true researcher is a person who holds steadfast to personal strengths and uses those strengths to come up with original ideas and novel approaches. I would love to see all young researchers become just like that.

Interviewee profile

■ Career highlights

He received the B.S. and M.S. degrees in physics and the Ph.D. degree in information sciences from Tohoku University, Miyagi in 1987, 1989, and 1999, respectively.

Since joining NTT in 1989, he has been investigating acoustic signal processing and acoustic echo cancellers. He is now a Senior Research Engineer in NTT Cyber Space Laboratories. He received the President's Award from NTT in 1995, the Outstanding Technological Development Award from the Acoustical Society of Japan (ASJ) in 1995, the Young Engineer Award from the Institute of Electronics, Information and Communication Engineers (IEICE) of Japan in 1996, the Achievement Award of IEICE in 1997, the Kiyoshi-Awaya Incentive Award from ASJ in 1998, the Satoh Paper Award from ASJ in 2002, and the Paper Award from IEICE in 2002. He is a member of IEEE, ASJ, and IEICE.