My Responsibility Is to Continue Researching Speech/Audio Coding to Provide Highest Quality Services and High Reliability

Takehiro Moriya
NTT Fellow, Head of Moriya Research Laboratory, NTT Communication Science Laboratories

Overview

Performing business tasks and holding meetings using personal computers, the Internet, and phones is accelerating the trend toward more efficient work processes and diverse work styles. It is generally agreed that sound quality can greatly influence the sense of presence in videoconferencing, and the results of NTT research have long been making major contributions on a global scale toward high-quality speech in fixed-line phones and mobile phones as well as in videoconferencing and video calls, as used in teleworking. NTT Fellow Dr. Takehiro Moriya of NTT Communication Science Laboratories has been involved in the research of speech/audio coding for about 40 years. We asked him about his current areas of research and his mindset as a researcher.

Keywords: speech and audio coding, telephone, standardization

Our goal is to meet our customers' needs in the fastest way possible

—Dr. Moriya, please tell us about your research thus far.

I have been involved for about 40 years in the research of speech and audio coding to digitize speech and music and to compress that information efficiently with high reproducibility at playback. For example, the music that we can hear from portable music players or digital broadcasts is not the original signal but rather a compressed signal in which the amount of information has been reduced to about one-tenth its original size. In short, this is research related to methods of compressing and reproducing signals while maintaining sound quality. Speech and audio coding technologies based on digital signal processing have been progressing over the past 50 years (**Fig. 1**). I will omit a detailed description of this history, but I'll say that these technologies were developed through the dedicated efforts of many researchers and engineers throughout the world. From the 1990s onward, these technologies have

AAC: Advanced Audio Coding
ADPCM: adaptive differential pulse code modulation
ALS: Audio Lossless Coding
AMR: Adaptive Multi-Rate
AMR WB: AMR Wideband

APC: Adaptive Predictive Coding
CELP: code excited linear prediction
EVRC: Enhanced Variable Rate Codec
EVS: Enhanced Voice Services
HE-AAC: High-Efficiency AAC
LSP: Line Spectrum Pair

MPEG: Moving Picture Expert Group
PARCOR: partial autocorrelation
PSI: Pitch Synchronous Innovation
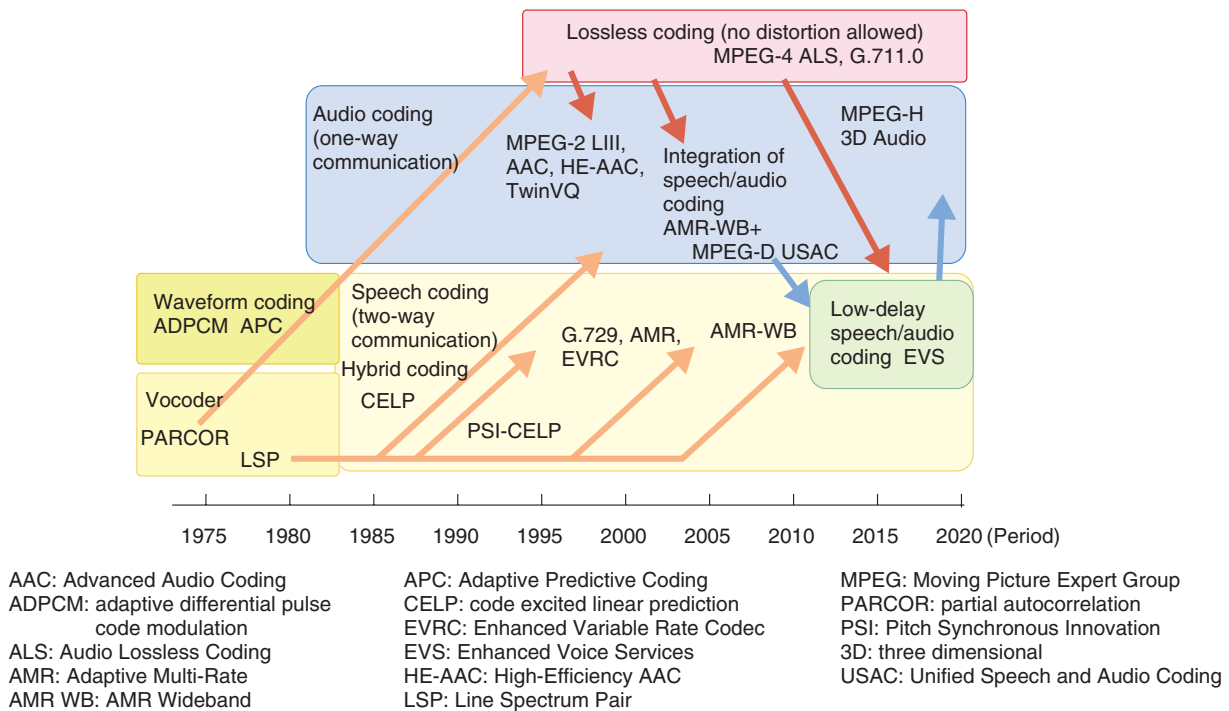3D: three dimensional
USAC: Unified Speech and Audio Coding

Fig. 1. Progress in the development of speech/audio coding.

been making significant contributions to everyday life and business in the form of telephones, broadcasts, etc. Among the many speech and audio technology fields, I believe that data compression/coding technology has made the most significant contribution to commercial services.

In the first half of the 1980s, the mainstream technologies were high-speed optical fiber in the fixed-line network and analog transmission in the mobile network (mobile phones) and the possibility of digitization was unclear. As a result, technologies for digital compression/coding of speech was losing sight of any uses or applications. However, in the 1990s, the worldwide trend shifted to the digitization of mobile phones with the result that digital compression/coding technologies suddenly became important. During this period, our research achievements in meeting certain conditions such as guaranteeing sound quality at low bit rates even with transmission error codes were adopted in Japan's standard system following a competitive process and used in second-generation (2G) mobile phones. Our elemental technologies were also adopted in 3G mobile phones and Internet protocol (IP) phones, contributing to the improvement of speech quality in mobile telephony throughout the world.

Around 2010, there was a strong desire in the 3rd Generation Partnership Project (3GPP), an international standardization organization targeting mobile communications systems, for establishing new coding protocols toward Voice over Long-Term Evolution (VoLTE), the worldwide 4G mobile communications network. In response to this need, Enhanced Voice Services (EVS) became internationally standardized as a speech/audio coding system through competition and collaboration among many expert parties around the world including the NTT Group.

Up to that point, speech coding systems for mobile phones used Code Excited Linear Prediction (CELP) emulating the human vocalizing mechanism to transmit the human voice with high quality at a low bit rate. The EVS codec combined CELP with newly developed low-delay music-oriented coding that made it possible to transmit speech including background noise and background music or music itself with low delay and high quality that had not been previously possible. During the standardization process, EVS was shown to have a level of quality significantly higher than current systems through large-scale subjective quality evaluation tests conducted by third-party institutions under a variety of conditions, sound sources, and languages. As a result, the EVS
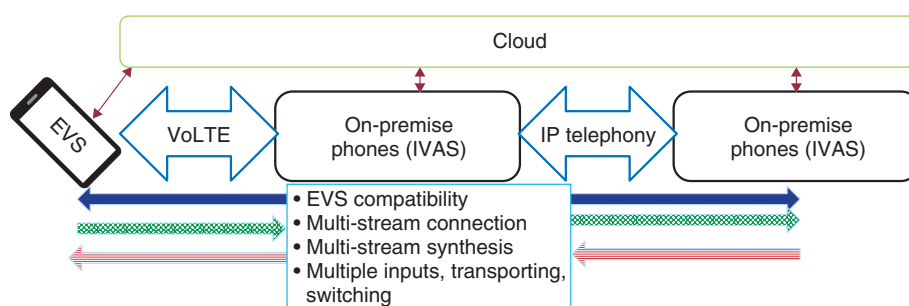
Fig. 2. Example of an extension-phone configuration using IVAS.

codec was adopted in unison by telecom companies, telephone equipment manufacturers, and chip manufacturers worldwide, and in Japan, calls between smartphones that had already come into use achieved a high level of quality with wideband transmission not previously possible in telephones.

—*Those were certainly major achievements.*

For about 40 years, the NTT team together with leading researchers and engineers throughout the world has been working to improve the quality of telephone calls through a process of trial and error. It's been a great joy to see the results of this hard work being used by people around the world.

We are currently working to standardize an extension of EVS called EVS extension for Immersive Voice and Audio Services (IVAS). The aim is to develop telephone services that maintain a sense of immersion in an interactive manner using multiple microphones and speakers such as for virtual reality and videoconferencing applications.

As a researcher at NTT, a global information-communications company with a profound history particularly in the research of telephone speech, my desire is to create services that can be applied to a variety of business scenarios. For example, immersion-related technology that makes participants of teleconferencing feel as if they were all in the same room is desirable, and video may also be necessary, but I would like to place priority on achieving calls with good audio quality irrespective of the medium. Teleconferencing applications using paid-for or free software have recently been gaining in popularity, but the networks used for running such applications are best-effort types with the result that quality degradation due to delay or packet loss cannot be avoided. In addition, using fixed-line phones has been assumed

with teleconferencing systems for business applications, and connecting to such systems from smartphones or IP extension phones can significantly degrade quality. I therefore believe that we should develop IVAS so that EVS codec can be used without degradation in high-reliability networks applicable to voice calls such as Hikari Denwa (IP telephony using optical fibers) and VoLTE networks constructed by network specialists (**Fig. 2**).

Research laboratories of leading telecom companies throughout the world conventionally possessed a considerable amount of influence, but today, this is no longer the case as the service field and industry players have changed. Amid this trend, NTT and NTT DOCOMO as companies providing fixed-line phone and mobile phone services attach great importance to improving speech-related technologies as a matter of responsibility. This is exactly the domain that I am in.

Imagining the smiling faces of people who use one's research achievements

—*Looking back at more than 40 years of research activities, what lessons have you learned?*

It is important that a researcher think for himself or herself, find a variety of issues, and make his or her ideas into a reality. Half of the motivation here comes from thinking that something would be interesting to research. It's also very important to consider whether the research activities themselves would be interesting. In other words, can those research activities be thought of as exciting and worthwhile? This could not be the case if there were no people that would one day enjoy the results of one's research, so the attitude to take is to devote oneself continually to research while imagining people who use your research results and thinking about ways to make them happy. The point

at which this idea came to me dates back to my student days when I began researching. At that time, I was involved in research completely different from what I'm doing today. I enjoyed thinking about this and that and making something concrete that I would like many people to use. However, no matter how hard I tried, I could not escape from the limits of self-contentment, and when asking other people what they thought, I would only get an indifferent reaction such as "What is the point?" I felt a vague sadness thinking that I could not be useful to anyone or contribute to society. Therefore, I wanted to work at a place in which whatever action I take, the world would respond to it, and I knocked at NTT's gate.

After joining the company, Nippon Telegraph and Telephone Public Corporation was privatized in 1985 as Nippon Telegraph and Telephone Corporation, now commonly known as NTT. At that time, the corporate philosophy was established as "We shall strive to provide highest quality services and high reliability based on technology development with a global vision and contribute to the creation of an affluent life and culture." I was very encouraged by these words as they underscored the importance of being an NTT researcher, and to this day, I still have the paper inscribed with these words on the back of my NTT security pass. These are words that I treasure. They constitute the philosophy that I have always adhered to, and as an employee of a company espousing such a philosophy, I have pursued my research while thinking whether my current work is based on and consistent with this philosophy. These words have provided daily support in my research life, especially when faced with a difficult problem or hardship such as when I could not achieve a consensus on standardization. Perhaps a bit of an exaggeration, but these words truly give worth to my role as a researcher.

It is unbelievable that I researched the same theme continuously for about 40 years. Given that the world changes rapidly, most researchers have to take up a new problem or change research fields after about five years, so sticking with the same theme for a 40-year period is extremely rare. As a researcher, it is important to discover or invent new things and write papers about them so that society will come to use them. For this reason, it frequently happens that a researcher cannot help but change research fields after achieving certain results. Under such conditions, I believe that a researcher in a subject area near practical use, even if at a stage at which final results have not yet been achieved, should discover a different viewpoint or skillfully find another issue to

address. Fortunately, in my case, I have been able to expand upon my research over a 40-year period. For example, in the 1990s, many mobile phone users expressed dissatisfaction with sound quality, and during research on this problem, the idea (new issue) came to me that the same technology for improving sound quality could be used to achieve high-quality playback of music. In this manner, I was able to find new issues one after another based on the same theme of "communications and sound." Of course, 40 years of research activities have also included failures and cases of losing out to the competition.

Pursuing multiple research themes and viewpoints with a balanced approach

—*Can you offer some advice to young researchers on how to lead a better research life?*

My advice would be to work on multiple research themes in parallel. My research team, for example, is currently working on two themes. One is to improve the sound quality and functions of smartphones and other devices as a telecom company and the other is to measure sound and other signals using light. I'm not sure if such light-based measurement would be immediately useful, but I believe that it could contribute to developing services using NTT's high-speed networks in a totally new way by conducting repeated trials. I also would like young researchers to have multiple viewpoints with different dimensions (such as theoretical and experimental and short-term and long-term problems). This is because social trends and conditions are dynamic and if external conditions should change, research results no matter how superior may turn into technology that won't be used. Papers describing one's research results can be written, but it may happen that they will never be used if not matching current conditions or if failing to be competitive in a field with many competitors. It is important to keep an eye on world trends and technological advances and continue researching over a long period while maintaining a balanced approach, and it's better to be prepared to throw out one or another research theme if necessary.

I expect our young researchers to continue in their research efforts over the next 20 years. Research activities have many peaks and valleys, so I think it is good to take on multiple themes having different timeframes and techniques, such as by concentrating on one theme for several years and taking up another over the long term. It is rare to be able to concentrate

on one theme in research or elsewhere and you may spend some time supervising subordinates and dealing with an increasing number of stakeholders. Even so, a researcher should devote his or her efforts to a primary theme while maintaining an interest in peripheral areas because one's primary theme may unfortunately disappear one day.

Promoting the significance of one's research with conviction and winning supporters

—How should research themes be selected?

To begin with, you should avoid research themes that are currently in fashion. Entering an area that many researchers are already pursuing means that you will always be behind someone. When joining in on the competition, it is difficult to promote your own achievements to society. It would be better to select an original theme in which you can integrate your knowledge and skills with social needs.

As a researcher in a company that provides telephone services, I am always thinking about ways of bringing joy to the billions of people on Earth when they use their telephones. I don't think it's a good idea to compete while not knowing who might use the results of one's research. This is the same as running without the finish line in mind. However, each research field has its own characteristics. In the case of basic research, results obtained in the present may become useful in years to come or reflected in textbooks, so the goal here is to improve future society.

Additionally, to drive research forward, a budget is needed for facilities, team formation, etc., but securing that budget is not easy if there is no one who understands the significance of that research. In particular, given the relatively long research span of basic research, people who can understand the importance of that research into the future must be found. It is therefore important to promote the significance of one's research with conviction and winning supporters.

There have been times in which I too have had much difficulty in finding supporters of my research efforts. Around 2000, it was recognized around the world that technology for compressing music signals at a low bit rate was nearly completed and that coding was no longer important. I was convinced, however, that "no matter how large transmission capacity becomes, technology for reducing the amount of information under conditions that strictly prohibit degradation in sound quality will still be useful and

should therefore be standardized." However, this belief was met with some resistance even within the NTT laboratories and was rejected by standards organizations as well. Nevertheless, I pointed out the importance of this technology in collaboration with overseas specialists and gradually won support from both inside NTT and from participants of standards organizations. As a result of this and incorporating requirements from the music industry, we were able at long last to establish an international standard in 2005.

Today, 15 years since establishing that standard, this technology has finally found widespread use in the NTT Group's high-quality music delivery business and other areas. As a researcher, the competitive arena is the world. Representatives of the United States, Europe, and China can be quite assertive and have at times disagreed with my opinions. Yet, technology speaks the truth, and it comes down to competing with technology and having strong convictions about it. In addition, whenever I find myself in troubled times, I return to and follow NTT's corporate philosophy, the base point of my research activity, which also provides me with emotional support.

—Dr. Moriya, please tell us about your future aspirations.

Telephone services including mobile phones have reached a state of saturation, but video services and services based on artificial intelligence and other new technologies are attracting attention not only from the world of information communications but also from platformers such as Google, Amazon, Facebook, and Apple (GAFA) and many other fields. These new fields include many research issues, and NTT as well is actively promoting research in these fields. Within this environment, the number of telephone-related research issues are decreasing, but there are still some solutions that would be useful to society. For example, since different types of telephone networks, such as mobile phones, fixed-line phones, and extension phones, still exist, measures are needed to deal with degraded sound quality in teleconferencing using many and varied terminals. As someone who has been a researcher in this field and an employee of a telecom company for about 40 years, I believe that it's my responsibility to research speech and audio coding through to the end.

Finally, opportunities for giving advice to young researchers have recently been increasing, but many of the experiences that I have accumulated over these

past 40 years are not necessarily helpful at present. Therefore, rather than simply giving out instruction, I speak about the need to take on new challenges that are now appearing in our society.

■ **Interviewee profile**
Takehiro Moriya

NTT Fellow, Head of Moriya Research Laboratory, NTT Communication Science Laboratories.

He received a B.S., M.S., and Ph.D. in mathematical engineering and instrumentation physics from the University of Tokyo in 1978, 1980, and 1989. Since joining the Nippon Telegraph and Telephone Public Corporation (now NTT) in 1980, he has been engaged in research on medium to low bitrate speech and audio coding. In 1989, he worked at AT&T Bell Laboratories, NJ, USA, as a visiting researcher. Since 1990, he has contributed to the standardization of coding schemes for the Japanese Personal Digital Cellular system, ITU-T (International Telecommunication Union - Telecommunication Standardization Sector) G.729, G.711.0, ISO/IEC (International Organization for Standardization/International Electrotechnical Commission) MPEG, MPEG-4 General Audio Coding, MPEG-4 ALS, and 3GPP EVS. He is an IEICE (Institute of Electronics, Information and Communication Engineers) Fellow, an IEEE (Institute of Electrical and Electronics Engineers) Fellow, and a member of the Information Processing Society of Japan and the Acoustical Society of Japan.