R&D Spirits

Speech Quality Assessment Technology for the IP Era

Akira Takahashi

Group Manager
Communication Traffic & Service Quality Project
NTT Service Integration Laboratories

The research areas of the Communication Traffic & Service Quality Project [1] include the "assessment and management of speech quality," an eternal issue in telephony services. Today, as IP (Internet protocol) telephony services begin to spread throughout the world, research and development (R&D) in this field is again attracting attention. We asked the Group Manager Akira Takahashi, a member of this project, to brief us on current R&D trends and international standardization efforts in this field and to give us his outlook for speech quality assessment in the near future.

Comprehensive speech quality estimation technology toward high-quality speech communication

—Mr. Takahashi, please tell us about your current area of research.

Our group is researching quality assessment in communication services [2]. Of course, "quality" can mean many things, but the focus of my research is the quality of "speech," the most basic aspect of communication services. In the past, the most common method of assessing speech quality was to conduct a subjective assessment in which people decide on their own whether the speech quality of actual calls is good or bad. This method, however, requires many monitors and specialized devices as well as much time and is by no means easy. Researchers therefore wondered whether the speech quality perceived by people could be estimated by taking physical measurements, or in other words, by "objective assessment." Physical measurements are appealing because they are reproducible and enable standard assessment. They can also be used to assess service quality in real time, something that subjective assessment cannot do. For these reasons, interest began to grow at NTT in comprehensive speech quality estimation technology for performing objective assessments. This is exactly what I am researching. Our comprehensive speech quality estimation technology can be broadly divided into two layers, one dealing with physical measurements and the other with human factors. My specific research theme is closer to the latter. That is, I am researching the correspondence between quality factors like delay, distortion, and echo that can be perceived by human beings and physical factors obtained from measurements. My main target, at present, is IP telephony. Factors that degrade the quality of IP-telephony services are known to include delay, coding-related distortion, and packet loss. But when these occur simultaneously and in combination, how will they be perceived by users in terms of overall quality? To answer this question, I am mapping individual factors to relative quality values and modeling that process (**Fig. 1**).

—What are the key technical points of this research?

The way we handle speech quality in our comprehensive speech quality estimation technology is completely different from that adopted by current international standards. To give some background here, conventional quality estimation technology treated quality as something that resulted from the independent actions of individual degradation factors. For example, if delay produced a degradation value of 10 and
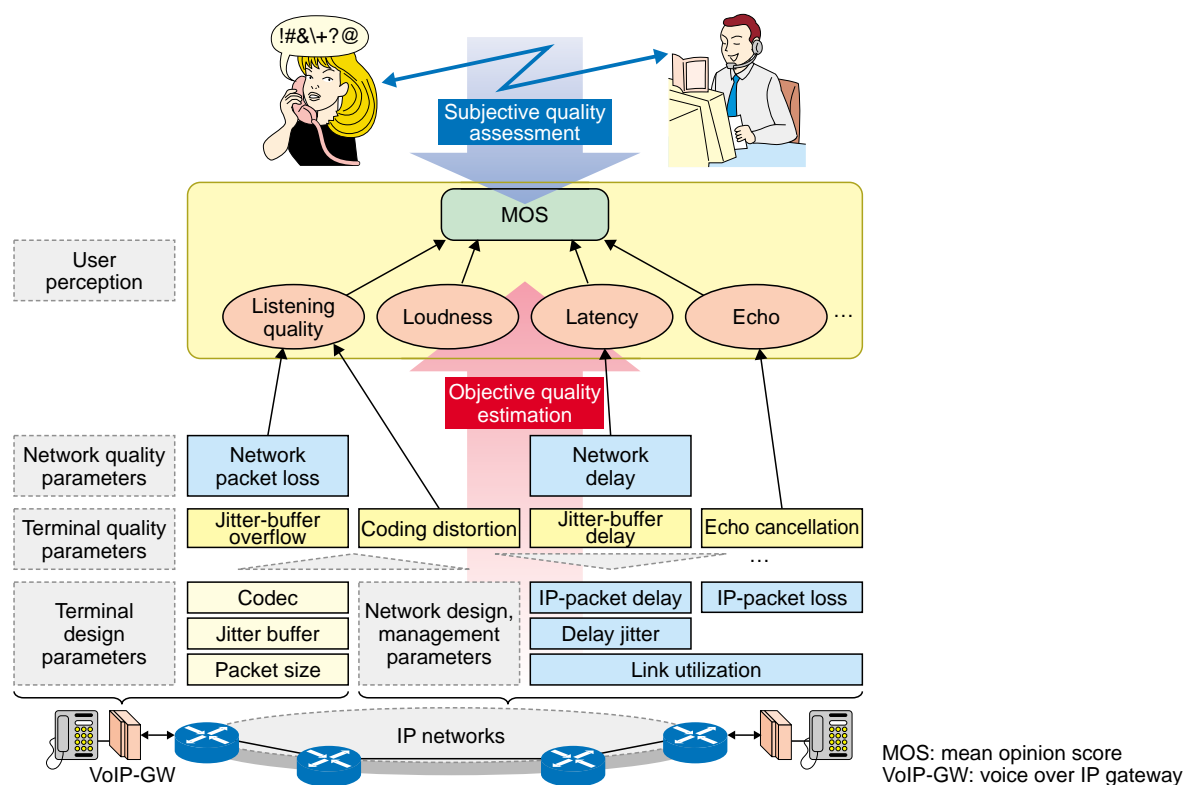
Fig. 1. Quality assessment of IP-telephony services.

packet loss a value of 5, a total degradation value of 10 + 5 = 15 could be expected to occur. In contrast, we consider that degradation factors can mutually interact and that total degradation in this example may not necessarily be their simple sum. We have improved estimation accuracy by about two times by incorporating this mutual interaction into our model. This model, by the way, is not limited to conventional telephone-band speech assessment. It can also be used in wideband speech assessment slated for next-generation telephony services. As far as I know, there is no other model in this field that also covers wideband speech.

We have constructed this model by applying the technical data and know-how that NTT has been accumulating in this field for more than 20 years and by using the company's excellent research facilities. The steady accumulation of this experimental knowledge over time gives me the confidence to say that our quality-estimation technology has considerable value (**Fig. 2**).

—What kind of changes do you think this research will bring about?

I believe it can promote the spread of new speech communication services. One example is wideband communications. Compared with existing telephone services operating at a bandwidth of 3.4 kHz, wideband services operate at a bandwidth of 7 kHz or more, the same as AM (amplitude modulation) radio. At this bandwidth, calls can be made with natural sounding speech like that on AM radio programs. Wideband services are scheduled to be launched in about one or two years. On a more long-term basis, I believe our research will help spread and enhance videophone services. Of course, past years have seen many attempts to popularize the use of videophones, but widespread use has yet to be achieved. On the other hand, the appearance of DoCoMo's FOMA (freedom of mobile multimedia access) handsets equipped with a videophone function makes me feel that the time is ripe for this mode of communication. I think that the spread of videophone service with natural-sounding speech will lead to an even larger lineup of video-related services.

As we can see from the evolution of cellular phones and IP telephony, a new service will generally be lacking in quality at the time of its initial appearance even though it may be superior in terms of convenience and cost. Quality engineers like us want to do something about this! We are making steady progress toward the provision of new services that not only are convenient and economical but also stand out in terms of sound quality.

- Applicable to both telephone-band (3.4-kHz) and wideband (7-kHz) speech communication services
- Comprehensive quality index for directly comparing these services
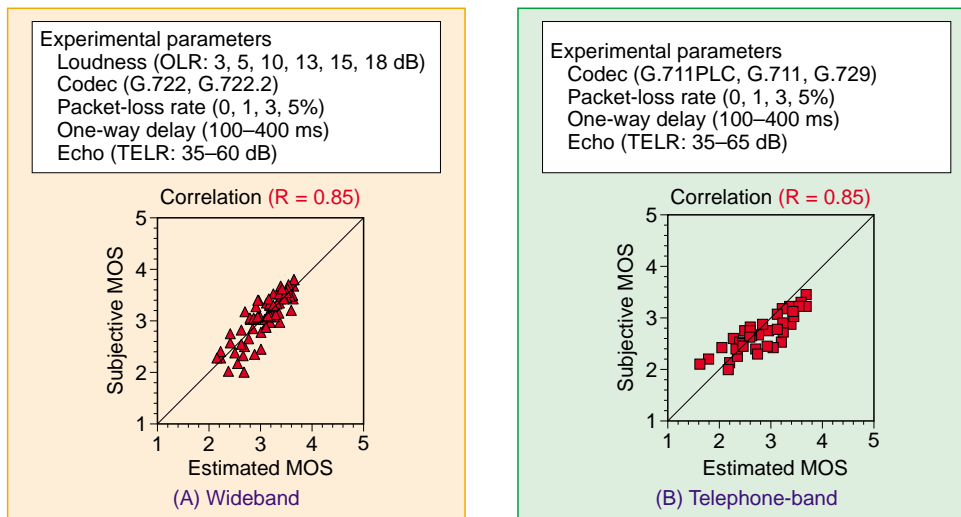- High accuracy in estimating the actual subjective quality (MOS)

Experimental parameters
    Loudness (OLR: 3, 5, 10, 13, 15, 18 dB)
    Codec (G.722, G.722.2)
    Packet-loss rate (0, 1, 3, 5%)
    One-way delay (100–400 ms)
    Echo (TELR: 35–60 dB)

Correlation (R = 0.85)

(A) Wideband

Experimental parameters
    Codec (G.711PLC, G.711, G.729)
    Packet-loss rate (0, 1, 3, 5%)
    One-way delay (100–400 ms)
    Echo (TELR: 35–65 dB)

Correlation (R = 0.85)

(B) Telephone-band

Fig. 2.   Relationship between estimated and subjective MOS values. (OLR: overall loudness rating, TELR: talker echo loudness rating)

—Are you faced with any problems in this research? And if so, how are you dealing with them?

One problem is how to deal with a diverse array of terminals. Until recently, terminals essentially meant handsets, and standard assumptions could be made when making speech quality assessments. Nowadays, users may connect earphones and headphones to their cellular phones and may even make calls from personal computers in a hands-free manner by connecting microphones and speakers. Various calling formats have been considered. Our task here is to construct a quality-estimation model that can accommodate any terminal with the aim of providing services of sufficient quality in any situation. How to go about this is our number one problem. We might consider classifying terminals into a number of patterns and constructing models for each, but this is not a very elegant solution from an engineer's viewpoint. Instead, we are working on a mechanism that enables a single model to cover a variety of terminal types.

A leader in international standardization as a rapporteur and in joint research with universities and enterprises

—Mr. Takahashi, what are the trends in research on speech quality objective assessment?

Flagship carriers like BT, Deutsche Telecom, and France Telecom have taken the lead in this research area. NTT has, of course, been active in this area for some time. In recent years, moreover, IP-related equipment vendors, measurement equipment vendors, and other related companies have been energetically researching quality assessment. It appears that there is a broad range of research in this area. In addition to this background, the global spread of cellular phones and IP telephony is making the quality of voice-based services a subject of great interest.

—What about international standardization activities in this area?

NTT has been contributing to international standardization efforts for a long time. I myself have been involved for some ten years and have been serving as a rapporteur of ITU-T SG12 since 2005 [3]. NTT has, of course, been actively submitting proposals in this area. In terms of comprehensive speech quality estimation technology now on the discussion table, NTT has taken a leadership role not only in technical proposals but also in the overall concept and framework of quality assessment (**Fig. 3**).

—How have NTT's research results been received?

In the beginning, NTT's research in this area attracted international attention. A lot of technology coming out of NTT was also adopted in international standards. There was a time, however, when NTT was somewhat removed from international standardiza-

SED: speech enhancement device such as acoustic echo canceller (AEC), automatic gain canceller (AGC), or noise canceller.
RTCP-XR: realtime protocol (RTP) control protocol extended report
Ie, eff: effective equipment impairment factor
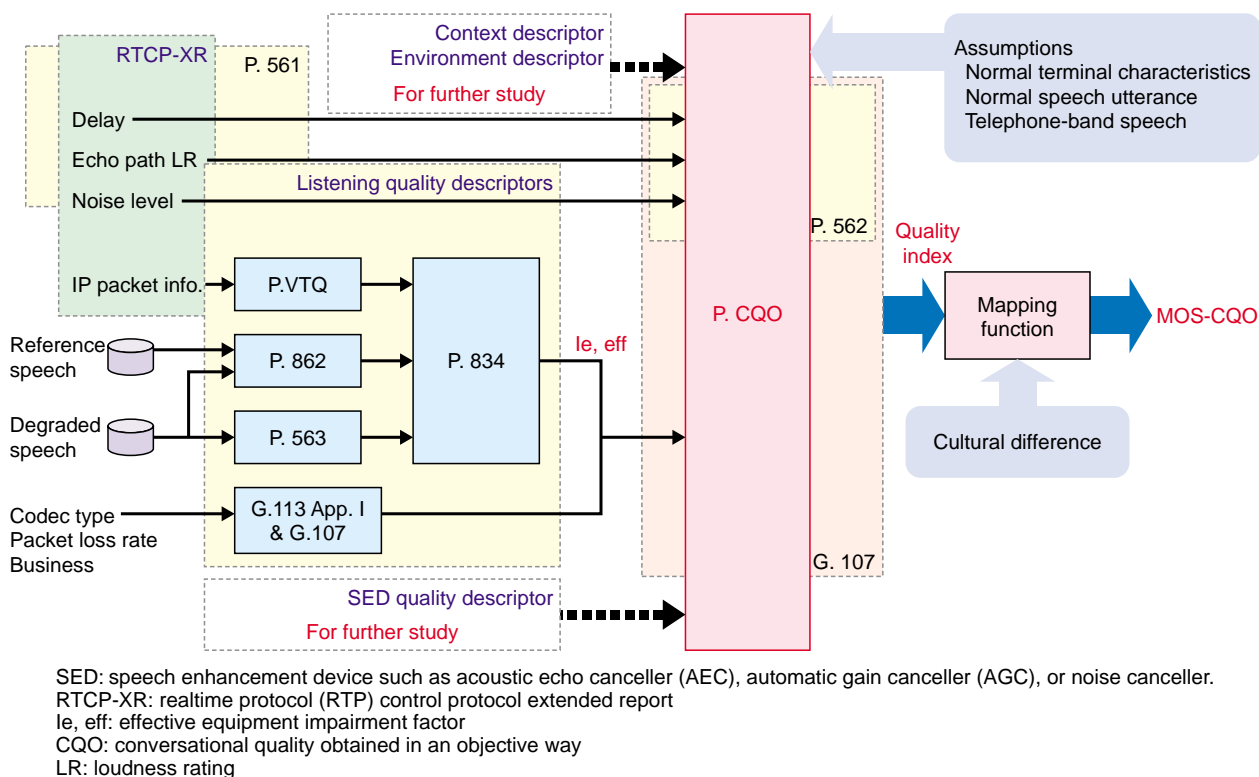CQO: conversational quality obtained in an objective way
LR: loudness rating

Fig. 3.   International framework on conversational quality estimation standard (P.CQO).

tion forums in this area, and several years ago, NTT's impact on this research field began to fade. When I resumed my involvement in international standardization activities several years ago, it took a lot of effort to make NTT's research and proposals attractive to others. Consequently, when we proposed our comprehensive speech quality estimation technology, the initial reaction was overwhelmingly negative. As a result, there was little discussion of our technology. But later, after several years of perseverance in presenting our proposal, the need for such a technology was finally recognized. My present position as a rapporteur of ITU-T SG12 is a sign that NTT is again having an impact in this research area.

—*Are you collaborating with any overseas enterprises or universities?*

At present, we have several partners with whom we aim to make joint proposals for standardization. The reason for this is interesting. To begin with, there are actually few societies in quality-related research fields specializing in speech. But in general, scientific societies are forums for carrying on technical discussions while standardization conferences are forums for exchanging opinions with an eye to cooperation. As a result, our area of research has come to involve both standardization and technology exchange, and there are many aspects in which joint research and standardization activities are linked. This is how the collaboration that we are involved in was born.

—*What do you aim to accomplish through these standardization activities?*

Since it will not do to talk about quality and assessment technology using different measures or scales, standardization is very important. Though it is true that standardization includes leadership issues between countries and companies, the standardization of something as important as quality must first consider what will be truly useful for society on the whole. I therefore intend to devote my energies to spreading NTT technology in the world while having a strong sense of mission in contributing to society.

Returning to the theme researched as a new employee after 18 years

—*Mr. Takahashi, please tell us about your major at school and how it became your technical foundation.*

I studied probability in the mathematics department

at my university. My character at the time was at odds with an ambiguous world in which answers depend on the manner of interpretation, and I had liked the clear-cut answers and elegant nature of mathematics ever since I was a teenager. But during my undergraduate studies, my interests began to shift toward applied research because I wanted to know how mathematics could actually be put to use. At that time, I also developed an interest in problems related to natural resources and the environment, and I began to think that communications might one day become an important part of preserving resources and the environment. Looking back, this was just a student's hasty conclusion, but nevertheless I entered NTT with the intention of applying mathematics to the field of communications. I had actually received some short-term training at NTT's Yokosuka R&D Center during my third year at university, and I remember feeling quite comfortable at NTT for some reason.

—*What kind of work have you been involved with up to now?*

Well, the research theme given to me just after I entered the company was the same as the one I am working on now: quality assessment of wideband speech communications. This was followed by research on assessing CD-quality music at an even wider band of 20 kHz (CD: compact disc). Then, after nearly two years of working on this, I was given the opportunity to study abroad at the California Institute of Technology. I was not there as a guest researcher, but rather as a first-year student in a master's course, so I was quite removed from research during this stay. It seems as if this overseas study was also meant to introduce me to the real world, because I was soon put in charge of work dealing with international standardization after returning to Japan. This was in 1993, which was just before the launch of PHS (personal handy-phone system) services. As it turned out, I also became engaged in the research of PHS service quality in parallel with my standardization activities. I still have a strong attachment to PHS despite the fact that it has mostly dropped out of sight. Then, in 1998, I was assigned to development work on telephone equipment in the Customer Equipment Division of NTT West. I returned from this assignment in 2001, the year that it was decided that the "050" prefix would be used for IP telephony. I was kept very busy around this time drafting a standard on measuring quality for IP telephony at the Telecommunication Technology Committee.

My current research theme is again quality assess-

ment of wideband speech communications, but my approach is completely different from last time. Whereas subjective assessment was my method of choice in the past, it is now objective assessment. The technology has changed quite a bit as well. But knowing that wideband speech communications is finally nearing deployment after 18 years and that I again have a chance to work on it gives me great pleasure—it's really unbelievable to me.

—*What has become your objective in your R&D activities to date?*

In corporate R&D, the main mission is to contribute to business. I too hope to contribute in some way to the business of the NTT Group. But in parallel with this realistic aspect of R&D, what really motivates me in my R&D work is working on an answer to the question: "What are human beings?" Or more specifically, "how do human beings react to sound?" I find this deeply interesting, and this is why I find research on quality assessment so appealing. I wonder if investigation of human factors in research might be an ongoing theme. The more one researches, the more one realizes that there is more to research than just technology—human sensations and feelings must also be taken into account. As one who is attracted to the conciseness and elegance of mathematics, I was even taken by surprise by this unexpected encounter with a fuzzy and delicate world.

Working to spread quality assessment/management technology and achieve higher levels of speech quality

—*How would you like to expand your research in the years to come?*

There are several themes I would like to pursue. One is research on quality-selection technology with the aim of providing quality in accordance with user attributes. For example, young people and elderly people tend to differ in the way they perceive speech quality in telephone calls, but current telephone services are not really equipped to deal with that difference. Although terminals that can slow down speech to make listening easier for the elderly have recently appeared, this issue has yet to be dealt with from an overall viewpoint. I think that the idea that quality should fit each individual user should be taken more seriously from now on. That is why I am interested in quality-selection technology.

Another direction I'd like to take is research on high-fidelity speech quality assessment technology.

Speech quality at present is closely associated with telephone calls and is therefore on a level where users can clearly judge it to be either good or bad simply by listening. However, when talking about broadband, high-reality communications, we have to start thinking about high levels of quality that cannot be easily distinguished by humans. Researching ways of assessing quality in such an environment is a challenge that I would like to take on.

—What kind of contribution would you like to make to society as a researcher?

I would like to help spread the use of speech quality assessment and management technology in the world. This has traditionally been an area requiring a very high level of specialization because subjective assessment involving human speech has been a world of skilled persons where experience and know-how carry a lot of weight. In the past, even engineers working at NTT business companies would come to us with problems about speech quality and would implement only the measures that we recommended. On the other hand, objective assessment that can estimate quality on the basis of physical measurements, though still a specialized R&D field, is not so difficult to use. And as more measurement tools are made available, I would like to introduce this technology to many more engineers. I would also like to establish a system that can respond immediately to a quality-related problem on site. In short, I would like to be involved with speech quality throughout the NTT Group. If, through these efforts, people start to give more thought to quality in telephone services and speech communication services in general, that would be even better. That, I believe, would be a great contribution for a quality engineer like me.

—What is it like working at NTT Laboratories?

At the risk of being trite, I think NTT Laboratories is a treasure house of talented people. Many of my seniors there set very good examples for me, and I have a very strong sense of being nurtured in this environment. It is exactly for this reason that I think the exchange of personnel between business companies and research laboratories should be carried out on an even larger scale than at present. NTT Laboratories is simultaneously a treasure house of talented people and a place for cultivating great talent. I think that making more use of such a personnel-development function throughout the NTT Group would be a very worthwhile endeavor.

—Mr. Takahashi, could you leave us with a message for young researchers?

Something that I often say in my position is that a researcher should get out into the world as much as possible. Why is this? Well, in the field of research, it often happens that real discussion and debate cannot take place within one's own country. Take, for example, my area of research—speech quality. Up until recently, there was no comparable research being performed outside NTT in Japan. Consequently, there were no grounds for discussion, and I came to be regarded as an authority regardless of my true ability or knowledge. But once I stepped outside Japan, I encountered much controversy. Frank debate was possible in all kinds of environments—not only at formal gatherings such as academic societies, but also at universities, corporate research laboratories, and standardization conferences. Going out to places like these enables you to question the true value of your research for the first time. I want young researchers to occasionally have experiences that bring them down to earth. And I want them to grow to be world-class researchers.

References

[1] http://www.ntt.co.jp/qos/
[2] A. Takahashi, A. Kurashima, and H. Aoki, "Quality Assessment of Wideband Speech Communication Services," NTT Technical Review, Vol. 4, No. 4, pp. 47-51, 2006.
[3] H. Aoki, H. Yamamoto, K. Matsumoto, and A. Takahashi, "Recent QoS Standardization Activities in ITU-T SG12," NTT Technical Review, Vol. 4, No. 4, pp. 52-54, 2006.

Interviewee profile
■ Career highlights
He received the B.S. degree in mathematics from Hokkaido University, Hokkaido, in 1988 and the M.S. degree in electrical engineering from the California Institute of Technology, USA, in 1993. He joined NTT Laboratories in 1988 and has been engaged in research into the quality assessment of speech and audio telecommunications. Currently, he is primarily working on the quality assessment of speech over IP networks. He has been contributing to ITU-T SG12 since 1994. He is a Co-Rapporteur for Question 13/12, which is responsible for multimedia quality assessment and network planning. He has also been engaged in the standardization of quality assessment methodologies for IP-telephony services in the Telecommunication Technology Committee (TTC) since 2002. He received NTT West President's Award in 2001. He received the TTC Award in Japan in 2004 and the ITU-AJ Award in Japan in 2005.