July 26, 2005

Development of a High-quality Multi-point Voice Teleconferencing System that Features Freedom in Microphone Placement
-- Application of an automatic gain control using directional discrimination and a wideband speech codec --

The NTT Corporation, headquartered in Chiyoda-ku, Tokyo and headed by President and CEO Norio Wada, have developed directional Automatic Gain Control (AGC[*1]) technology that automatically adjusts the loudness of simultaneously spoken voices to an appropriate level. NTT has also developed a wideband speech codec[*2] for multi-point communication that allows communication in the respective frequency-bandwidths even when wideband enabled terminal and conventional narrowband (telephone quality) terminals are used simultaneously in a conference.
When these technologies are applied to multi-point teleconferencing systems, the users will be free from complicated adjustments required in system installations, and can connect to the other terminals regardless of their wideband or narrowband capabilities.

Development
In conventional voice conferencing systems equipped with microphones and speaker units, there was a problem that only the voices of the talkers close to the microphones are conveyed well but the voices away from the microphone are not. To overcome this problem, there was a need for automatic gain control technology which controls the speech volume to desired level and suppresses ambient noise and acoustic echo.
In addition, multi-point teleconference with wideband speech capabilities has become popular as the result of the rapid growth of the Internet. In the conventional systems, however, all participants had to use the narrowband (telephone quality) speech codec if a single terminal was without wideband capabilities. Thus, there was also a demand for a wideband speech codec which can interoperate with wideband and conventional narrowband speech codec at the same time. Moreover, the decoder needs to suppress the degradation of voice quality due to packet losses.

Technical Features
1. Directional AGC technology for simultaneous adjustment of voice loudness according to direction of speakers in a single conference room ([Appendix 1](#))
In conventional TV conferencing or voice teleconferencing where a number of people gather in a single conference room, a microphone could be assigned to each participant, but it was too costly. In case of sharing a single microphone by multiple participants, conventional AGC technology for automatic volume adjustment could be used, but there was a problem with the volume being too high for participants that were close to the microphones when the AGC adjusts the volume to the participants further away from the microphone.
The directional AGC technique described here is based on microphone array technology[*3] that employs four microphones. When a participant speaks, the system

discriminates the speaker from the others by determining the direction and the loudness of the voice, and adjusts the volume to an appropriate level. For example, when ten persons are sitting at a long table as shown in Appendix 1, this system makes it possible for all voices to be conveyed to the other ends at the same loudness, whereas an ordinary system would only convey the voices of the persons close to the equipment. Also, where the participants would have to speak in a loud voice to be heard properly, by using this technology, communication in a softer and more natural manner is possible.

Furthermore, this AGC technology also includes an acoustic echo canceller[*4] and a line echo canceller. It thus inherits the basic functions of a conventional voice teleconferencing system and is capable of connecting to the telephone network as well.

2.Wideband speech codec that enables multi-point teleconferencing where narrowband and wideband speech terminals are mixed(Appendix 2)

The speech codec digitizes the human voice to transmit it to another party via the Internet or other digital networks. The speech codec generally used for VoIP[*5] (IP phones) provides only telephone quality, narrowband speech. However, teleconferencing systems equipped with microphones and speaker units requires a higher speech quality than the conventional telephony equipped with a handset. For this reason, a wideband speech codec has also come to be used, but because VoIP has a telephone frequency bandwidth, narrowband codecs were still to be used for mutual connections. It is also the case for multi-point teleconferencing, where if a single conventional narrowband VoIP terminal joins the conference, all terminals had to use the narrowband codec. Thus, the clarity in the conferencing was insufficient when both conventional terminals and wideband speech terminals are used in a conference.

This speech codec has a scalable structure[*6] that enables mutual connectivity both with conventional VoIP terminals and wideband speech terminals. The input speech that enters a microphone is divided into conventional narrowband (telephone frequency-bandwidth) speech signal and a high-frequency component. The narrowband signal is then encoded with the conventional codec and the high-frequency component is encoded using a newly developed original codec. Either the narrowband speech component or both narrowband and high-frequency components can be transmitted, depending on the capability of the terminal on the other end. This approach achieves multi-point teleconferencing using both narrowband and wideband capable terminals simultaneously. Furthermore, this codec sends additional data used for packet-loss concealment together with the high-frequency component. Thus, voice degradation caused by packet-loss is minimized and the speech quality is maintained.

Future Development

A voice conferencing system that incorporates this technology will appear in about six months as a commercial product from NTT EAST and NTT WEST. For further improvement for comfortable voice communication, the NTT Laboratories plan to move forward with research and development on the extraction of voice from the midst of the great variety of ambient household sounds such as TV audio and the sounds of housework.

Terminology
*1:AGC (Automatic Gain Control)
    A function that automatically adjusts the microphone level to match the level of the sound. It is used for video camera microphones, etc.
*2:Speech codec
    Codec is an abbreviation of enCOder/DECoder, and encodes the speech signal

from a microphone to encoded digital data and decodes it to a speech signal. Narrowband speech, which is limited to a frequency bandwidth up to 3.4 kHz, is mainly encoded with the ITU-T G.711 in VoIP communication. For wideband speech containing signal components greater than 3.4 kHz, codecs such as the ITU-T G.722, which is capable of frequency bandwidth up to 7 kHz, are standardized.

*3: Microphone array technology

A technique for emphasizing or de-emphasizing the sound from an arbitrary direction by complex processing of the volume and temporal relations of signals output from two or more microphones. Unlike when using a directional microphone, the directionality can be changed freely.

*4: Echo canceller

Echo is the phenomenon that the sound of one's own voice is reproduced from his/her own speaker unit with some delay, and an echo canceller eliminates this. There are two types of echo. In acoustic echo, the voice of another party that is reproduced by a speaker is picked up by a microphone on this side and conveyed back to the other side. Line echo is generated by circuit-loop in the telephone network. Functions for eliminating these two kinds of echo are called acoustic echo cancellers and line echo cancellers.

*5: VoIP (Voice over Internet Protocol network)

The generic name for voice communication over the Internet, in which speech signal is converted to fragmented digital signal data and transported in a stream of packets. The Internet was originally designed for data communication.

*6: Scalable structure

This is a hierarchical data structure that makes it possible to reproduce voice using even a part of the encoded data produced by a codec. The developed codec is structured so that it can reproduce wideband speech when all of the encoded data is available and narrowband speech when only part of the data is available.

- Appendix 1
- Appendix 2

Send inquiries concerning this article to:
NTT Cyber Communication Laboratories
Planning Department, Public Relations: Kouno or Yamashita
Tel: 046-859-2032
E-mail: ckoho@lab.ntt.co.jp

NTT NEWS RELEASE ▶