High grade IP Telephony Technology

– The Background of Development and Technology of the e-Sound IP Phone –

Shinji Usuba　　Hiromi Aoyagi

As we entered 2000, we started to witness the emergence of common communication carriers offering broadband connections at affordable prices. This triggered a full-scale popularization of the Asymmetric Digital Subscriber Lines (ADSL) and the Fiber To The Home (FTTH, providing communication services with optical fibers), resulting in the assignment of dedicated IP telephone numbers (the “050” numbers) in June 2002. Further, by the end of 2003 the broadband user population exceeded 30 million. Communication charges for broadband in Japan, is the lowest at US$0.09 per 100kbit/sec., which means there is a staggering difference when compared with the United States (US$3.53) or Germany (US$4.42)[1]. This popularization of the broadband infrastructure resulted in an extension of the IP telephone services that was offered to business enterprises, making it available to ordinary homes. As of June 2004, the number of telephone numbers that start with “050”, an area code dedicated to IP telephone subscribers, which were distributed to IP telecommunication service providers, number far in excess of 10 million[2].

A certain standard of quality has been stipulated for the IP telephone services provided for subscribers of the “050” numbers (refer to page 37 for another article on this issue for details). This resulted in a voice quality on a par with that of ordinary “telephones” in recent years. Even before such a quality standard was established, we were consistently providing products with an aim to attain the best voice quality suitable for the IP network (hereinafter referred to as the “network”) environment, with a particular emphasis on the improvement of voice quality for the Voice over Internet Protocol (VoIP) telephone calls (hereinafter referred to as the “VoIP communications”) made over networks.

Voice communications made through such networks have a lot of potential. The voice quality of VoIP communications is evolving rapidly from a “usable voice quality” (level of voice that is comparable with conventional telephones) to a more “desirable voice quality”. This paper will describe the aspects and possibilities of the new communications in the full-IP and full-broadband era, followed by a description of the voice quality assuring technology that supports such communications.

Technology with an emphasis on voice quality

In VoIP communications, voice quality problems occur that are particular to voice data transmitted as packets. It is for this reason that the VoIP communications were initially not recognized as “ordinary telephones”. We felt that in order to popularize the VoIP communications around the world, it was essential to resolve this voice quality issue.

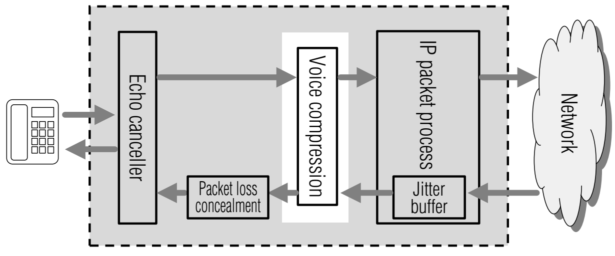

Fig. 1　VoIP functional block diagram

The utmost emphasis was placed on eradicating the problems that create the need to ask a caller to “repeat a sentence”, which occurs due to the loss of the beginning of a conversation or the loss of continuity of the voice transmission, through the faults of voice packets. We have been involved first of all in guaranteeing the continuity of voice transmissions and then for reproducing voice with a quality that is natural and easy to listen to, followed by an improvement to the voice quality. Figure 1 is a block diagram of the basic signal processing (function) of VoIP communications. The standard component (voice compression and then CODEC) secures mutual connectivity by using the standard method. Proprietary developments were made to improve voice quality, where no standard methods were stipulated (shaded segments in the diagram).

Evolution of Voice Quality (e-Sound IP Phone)

We recognized the potential of voice data packet communications over networks, therefore, we have been offering various VoIP products, starting with the release of the “BS1100 - Voice Hub”, which was the very first VoIP-GW (VoIP gateway) product, at the end of 1996. Our many customers used these products in Japan, which helped us nurture the product line to a level where stable quality can be assured in various environments. The greatest issue for the voice quality technology at the very beginning of VoIP communications was to somehow assure a “usable voice quality” in unstable network environments with a narrow transmission band.

In other words, our quality target for VoIP communications in the initial stages was the “telephone”. With numerous efforts to improve the voice quality, in order to achieve the target and broadband conversion of infrastructures at the same time, it became possible in recent years to realize a voice quality that is by no means inferior to calls made on public switched telephone networks. VoIP communications became recognized as a “telephone” in this manner, but in reality, its potential vastly surpasses the “telephone” from the very start.

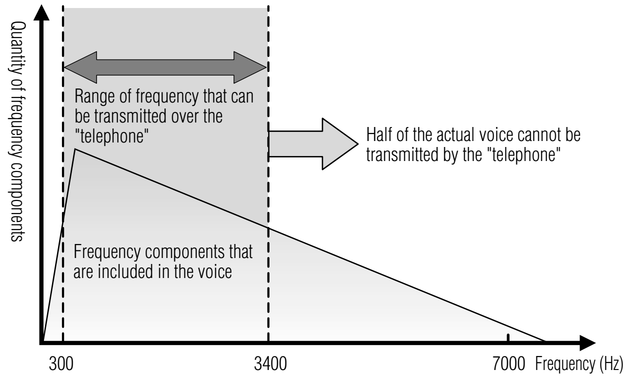

Fig. 2 Voice band and “telephone” band

The frequencies that can be included in voice transmitted in telephone conversations of calls made over the public switched telephone networks are limited. Specifically, it is possible to convey the voice of only up to 3.4kHz (telephone band) for transmissions over analog lines. The frequency of ordinary vocals is said to contain a voice frequency of up to 7kHz. Therefore, only about half of the actual voice frequency is transmitted with the “telephone” (refer to Figure 2). Since there are no such limitations with VoIP communications, on the other hand, it is theoretically possible to transmit the voice almost entirely intact.

Since we attained our target of obtaining a voice quality at a level of the “telephone”, we decided to set the next target to exceed this already attained voice quality level. We call this “real-time bidirectional communications with voice that surpasses the telephone band over an IP network” the “e-Sound” (e-Oto) technology. In order to realize this e-Sound, we adopted the broadband voice coding method, which can transmit voice of up to 7kHz band for our CODEC, to be the base of the voice transmission.

However, it is not possible to improve the voice quality for VoIP communications simply by adopting broadband voice coding. It is necessary to devise a means to manipulate frequencies to draw out the maximum amount of realistic and natural sensations for both male and female speakers, as well as to resolve problems that are inherent in voice communications over networks. The vast amounts of know-how accumulated in the actual use environment are the basis for resolving the issues related to the extension of the voice band. Proprietary devices were added to the echo canceller for the processing of noise components, which tend to increase relatively with the extension of the frequency band. Further, the packet loss concealment function is a new proprietary development. Such technologies were implemented in the e-Sound IP Phone before it was introduced to the world.

Core technologies that support the “e-Sound”

A summary of the proprietary core technologies that have evolved since the beginning of VoIP communications to this day, are provided in this segment.

There are four main factors which contribute to the deterioration of voice quality in VoIP communications.

- Delay (call transmission time lag)
- Echo (a phenomena whereby one's own voice returns after a delay)
- Fluctuation (variation of gaps in the arrival intervals of packets)
- Packet loss (missing voice information)

Our ventures relating to such factors are introduced below.

Core technology 1: Delays and efforts to resolve this problem

Delays of calls in VoIP communications arise from CODEC processes and IP packet processes. Call delays throw the rhythm of a conversation out of synch and deteriorate the call quality, in a similar fashion to those of international calls made a decade or so ago.

The delay arising from CODEC (processing delay) is determined by the particular CODEC method used. The optimum selection of CODEC is being implemented according to the transmission frequency band employed in the actual use environment, in order to maintain a proper balance with the coded voice quality and voice compression rates.

The delay arising from the conversion of voice into IP packets is determined by the amount of voice information contained in each individual IP packet. From the aspect of shortening call delays, it is better to break up and convert voice information as much as possible into smaller increments of packets. However, since the header information is included in each IP packet, aside from the voice information, and as the increments are made smaller, more bandwidth bands will be consumed. Because this is a tradeoff relationship, optimized adjustments of packet sizes are being made in accordance with the transmisson bandwidth in the actual use environment.

Core technology 2: Echoes and efforts to resolve this problem

Echoes can be quite prominent in VoIP communications, as previously described call delays occur. Echoes are almost negligible with telephone communications through existing public switched telephone network, therefore, the elimination of echoes is an important issue for securing the voice quality of VoIP communications.

There are two broad categories of technologies for eliminating echoes and these are the echo suppressor and the echo canceller. Although the structure of an echo suppressor is simple, the voice quality it is able to offer cannot be considered to be adequate, as the voice quality level of the other party is often reduced as well. We proceeded with modification and development of our proprietary technologies by using the echo canceller (Figure 3), which is capable of eliminating the echo alone.

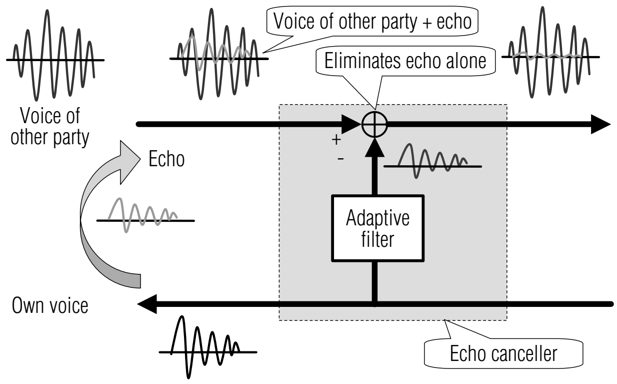

Fig. 3 Echo canceller

Standards and specifications that can be used as reference for the performance of echo cancellers, such as G.165 and G.168 of ITU-T, do exist. Compliance with such standards and specifications, however, does not mean that the performance will be adequate for use in an actual environment. This is because echo cancellers operate by considering both one's own voice and that of the other party. In an actual conversation, an unlimited number of operating conditions exist based on who the parties are, what the context of their conversation is and the pace in which the conversation progresses. It is not possible to conduct an investigative study of all such conditions simply on paper. We have accumulated know-how concerning its use in an actual environment with our early introduction of products in the market. Our proprietary technology for determining the operating conditions (status of conversation) in particular, has been evolving.

Core technology 3: Fluctuations and efforts to resolve this problem

Real-time characteristics are not guaranteed for the packet transmission of VoIP communications. For this reason, IP packets that have been transmitted from the transmitting terminal at a fixed interval do not necessarily arrive at the receiving terminal at the same fixed interval. The continuity of the voice signal is disrupted at the receiving terminal because of this fluctuation of the reception interval. Communication quality deteriorates as voice interruptions and jumps occur.

A buffer that absorbs fluctuations was installed at the receiving terminal as a countermeasure against such effects. Large fluctuations can be dealt with by increasing the capacity of the buffer. Doing so, however, creates a delay, since more data is accumulated in the buffer. A proprietary delay recovery process was developed to resolve this problem.

Firstly, a buffer of a certain capacity was made available to eliminate the voice interruptions. The amount of buffering was then set to change according to the fluctuation time. Finally, a proprietary device was implemented to reduce unnecessary delays to ensure that the natural sensation of a conversation was not spoiled, even if a large fluctuation occurred.

Core technology 4: Packet losses and efforts to resolve this problem

In general, UDP/IP is used for the packet transmission of VoIP communications. To accurately convey information, it is beneficial to use the TCP/IP, which comes with a retransmission control. Contrary to this, it presents a major problem for real-time characteristics. Real-time characteristics are an important factor for conversations, which are bidirectional communications. In the case of a UDP/IP, packet losses that occur during transmission will lead directly to the loss of voice at the receiving terminal. In general, precautions to prevent packet losses are conducted at the receiving terminal, and are usually integrated as a component of the CODEX process. The level of this ability to implement remedies varies significantly depending on the CODEX, thus this became an important issue and was considered to be a criteria for the selection of CODEX.

Where "e-Sound" is headed

Full-IP and full-broadband networks are rapidly becoming popular. In the future, we can expect to see waves of newly created modes of communication that surpass existing telephones. We believe that it will be possible to offer a brand new value, which can appeal to the emotions and sensitivity by extending the transmission of voice from the vocalization range of humans (up to 7kHz), realized by the e-Sound IP phone, to an audible range (up to 20kHz). We are, furthermore, confident that new values will be offered through the expansion of communications with realistic sounds that include not only voice, but also music and ambient

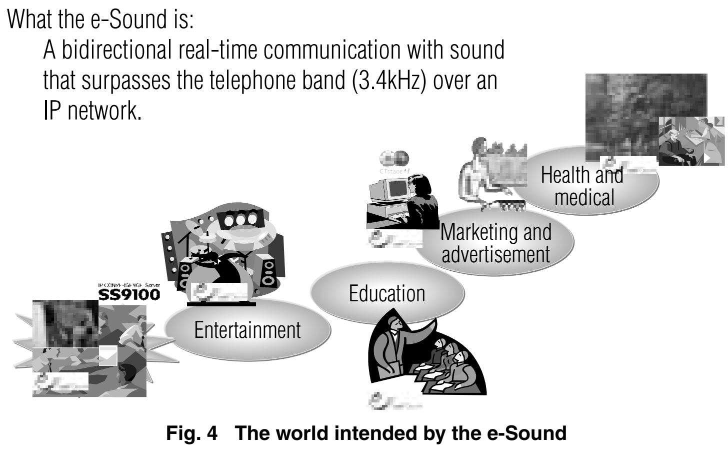

Fig. 4 The world intended by the e-Sound

sounds (Fig. 4).

VoIP communications brought down costs and led the way to pioneering communication infrastructures that are extremely economical. With the establishment of the broadband environment as a complementing infrastructure, it will be possible to offer much more captivating communications. This signifies a breach of the wall of limitations that was inherent with existing telephones, which served only for the conveyance of messages. This means that we can now offer heart-to-heart communications, in which people can truly understand each other and share their experiences. Further, when a maker provides goods and the buyer of those goods engage in deeper communications, it will be possible to elicit the potential needs buried deep inside the hearts of buyers, resulting in the realization of new consumption activities that can serve to make both the maker and buyer mutually satisfied. This is a vision for the world we intend to realize.

References

1) http://www.johotsusintokei.soumu.go.jp/whitepaper/ja/h16/pdf/index.html
2) http://www.johotsusintokei.soumu.go.jp/tel_number/ip_number.htm
3) Ministry of Public Management, Home Affairs, Posts and Telecommunications: Report on the “Study Group for the IP Network Technologies”, http://www.soumu.go.jp/s-news/2002/020222_3.html
4) Ministry of Public Management, Home Affairs, Posts and Telecommunications: On partial amendment of the IP Telephone Service Related Regulations, http://www.soumu.go.jp/s-news/2002/020614_4.html

Authors

Shinji Usuba : IP Solutions Company, Solution Development Div., Incubation Coordination Dept.

Hiromi Aoyagi : Corporate Research and Development Center, Human Interface Laboratory

TIPS [Trends in Quality Standards]

An organization, charged with the responsibility for investigating the issues for realizing telecommunications over IP networks, was set up at the Ministry of Public Management, Home Affairs, Posts and Telecommunications around the time when the broadband infrastructure began to spread rapidly. The “Study Group for the IP Network Technologies” disclosed the findings of their investigation as a report[3] in February 2002. This report indicated that the “quality of IP telephony must be considered an end-to-end quality including terminals at both ends, to make it possible for users to understand the quality and to be able to make appropriate selections of the services”. The specific classes of quality are categorized as shown in Table 1.

Table 1 Quality class categories for IP telephones (an excerpt from the report by the Ministry of Public Management, Home Affairs, Posts and Telecommunications)

	Class A (Same level of quality as fixed line telephones*Note)	Class B (Same level of quality as mobile telephones*Note)	Class C
General voice transmission quality rate (R)	> 80	> 70	> 50
End-to-end delay	< 100ms	< 150ms	< 400ms
Loss probability (connection quality) [Reference value]	≤ 0.15	≤ 0.15	≤ 0.15

* The values for the R value and delay in the table are to be satisfied by the probability of 95%.
Note: The "same level of quality as fixed line telephones" and "same level of quality as mobile telephones" are meant to represent the figures that are useful for the general voice transmission quality (R) of the overall call qualities. It is, therefore, not to be understood as requirements to attain the same quality levels as the fixed line telephones or mobile telephones for end-to-end delays and other functions.

Further, for the telephone numbers dedicated to IP telephones, it has been decided that “it is appropriate to use specific numbers of the “0A0” format, from an aspect of support for the provision of services not location specific, routing (distributions) of calls to IP networks, ease of service provider identification, ensuring capacity numbers and convenience of users”. This resulted in the establishment of the 11-digit number assignments that start with “050” as telephone numbers specifically allocated for IP telephones (June 2002[4]), according to the Telecommunication Numbering Regulation. At the same time the Commercial Telecommunication Facility Regulation stipulated a General Quality as a level of quality that needed to be sustained by IP telephony service providers (June 2002[4]). The General Quality referred to here is equivalent to Class C stipulated in the report mentioned previously.

In order for a service provider to use telephone numbers that start with “050”, which are specifically allocated for IP telephones for their IP telephony service businesses, it is necessary for them to file an application with the Ministry of Public Management, Home Affairs, Posts and Telecommunications with the measurement results and proof of measurements indicating that the General Quality (Class C) has been satisfied, and that the quality at such levels can be sustained. As for the method for evaluating the quality, the aforementioned report indicates that the “domestic standardization organizations needed to take a leading role in the study of the issue, while maintaining a close working relationship with international standardization organizations”. This resulted in the “Evaluation Method for IP Telephone Call Quality: JJ-201.01”, which was established (April 2003) by the IP Telephone Call Quality Evaluation SWG of the Network Management Committee at the TTC (Telecommunication Technology Committee). This standard describes a method for measuring the R value, which is used as an index for the quantitative representation of the call quality (further, as the leader of this SWG Oki Electric is making contributions to the establishment of this standard).

The R value is obtained as an output value (general voice transmission quality) of the calculation model (E-model) stipulated by the recommendation G.107 of the ITU-T. The E-model assumes that the quality factors, which affect call quality, have additive effects in terms of psychological measure. The R value is defined, based on the evaluation value for these quality factors, as shown in Figure 5.

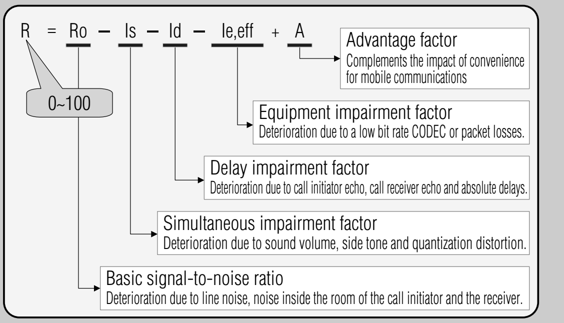

Fig. 5 Definition of R value