NTT DOCOMO

Technical Journal

Vol.16 Apr.2015 | No.4

DOCOMO Today

Technology Reports

NTT docomo

Disruptive Innovation vs. Activity Trap

Nobuyuki Akimoto

Executive Vice President, Chief Operating Officer
NTT DOCOMO Ventures, Inc.

Innovation, open innovation, disruptive innovation. Regardless of whether it is the private sector or the public sector, innovation is in vogue in the world. Why is innovation critical? This is neither the place to debate it nor the place to doubt it. For a large company that has achieved a certain level of success, whether it can realize disruptive innovation for its next move by participating in it itself is extremely important. Even in Japan, strategies related to innovations have been long debated, and many major companies are engaged in pursuing disruptive innovation in their own ways. Unfortunately, at present there are exceedingly few successful cases of disruptive innovation realized by large companies.

Why is that? Why are large companies quite unable to bring about disruptive innovation? It may be because they fall into the well-known “Innovator’s Dilemma.” Even if a company can sustain transformations by extending its current business, for a variety of reasons it is extremely difficult for the company to achieve disruptive innovation that creates new value and brings about major changes to society. For a big picture of the Innovator’s Dilemma, I defer to the presentation made in *The Innovator’s Dilemma* [1] by Professor Clayton Christensen of Harvard Business School. I have recently come to consider “means as an activity trap” as an example of this phenomenon. I would like to introduce this idea here.

The more a large company tries to take measures to advance disruptive innovation, the greater its tendency to fall into an activity trap in which individual policies and systems themselves are implemented and operated. What are some examples of means falling into activity traps? Corporate Venture Capital (CVC)*1 has been growing in number in step with the rising popularity of the term “open innovation.” From its purpose, CVC can be broadly divided into two types. Here, we assume the type of CVC whose primary purpose is to contribute to or create synergy with its parent company and/or affiliated companies in the short, medium, and long-term by investing in a venture to realize strategic return. Because of the large scale of the organization, funding, and other types of arrangements, there are cases where the existence of the CVC and the fund itself has become an activity trap. Synergy is not created, nor is strategic return produced. In actuality, there are several CVC in existence today that makes you wonder, “Wait a minute. What was the original purpose of this company/fund?” They are unfortunate examples of CVC that, without realizing it, have come to hold as the greatest purpose of their existence the sustenance of their organizations and systems instead of switching to a different direction based on new strategic theories from both the parent company and the CVC. What can be achieved by this? Business synergy is not produced, to say nothing of disruptive innovation. Only organizations, arrangements, and systems whose purposes have been lost are maintained.

For NTT DOCOMO and the NTT Group to continue to provide services that customers will keep choosing in the next 10 and 20 years, we must create disruptive innovation and new value. NTT DOCOMO Ventures, a CVC of NTT DOCOMO and the NTT Group, is a means for achieving this goal, along with several funds and programs. We are constantly reminding ourselves that we must contribute to creating disruptive innovation in the NTT Group so as to avoid falling into the activity trap I described above.

Reference

[1] C. M. Christensen: “*The Innovator’s Dilemma: When New Technologies Cause Great Firms to Fail*,” Harvard Business School Press, 1997.

*1 **Corporate Venture Capital:** An undertaking that connects a venture company and a business company to become hopeful partners in the creation of innovation. When the business company invests in the venture company, for example through a fund, the CVC serves as the business company’s subsidiary or related company to carry out a series of tasks, such as operating the fund and advancing cooperative work between the business company and the venture company. CVC can be broadly classified as one of two types from its purposes: (1) CVC that seeks primarily to achieve strategic return from the creation of synergy with the parent or group company, and (2) CVC that primarily seeks to produce capital gains without regard for business cooperation with the parent company or the group company. Note that for the first type of CVC, capital gains are also important.

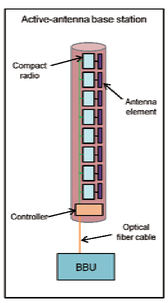
NTT DOCOMO Technical Journal

Vol.16 No.4 Apr. 2015

Contents

DOCOMO Today

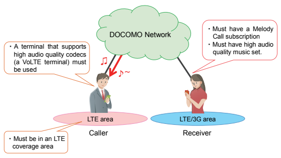
Technology Reports



(P.14)

(P.22)

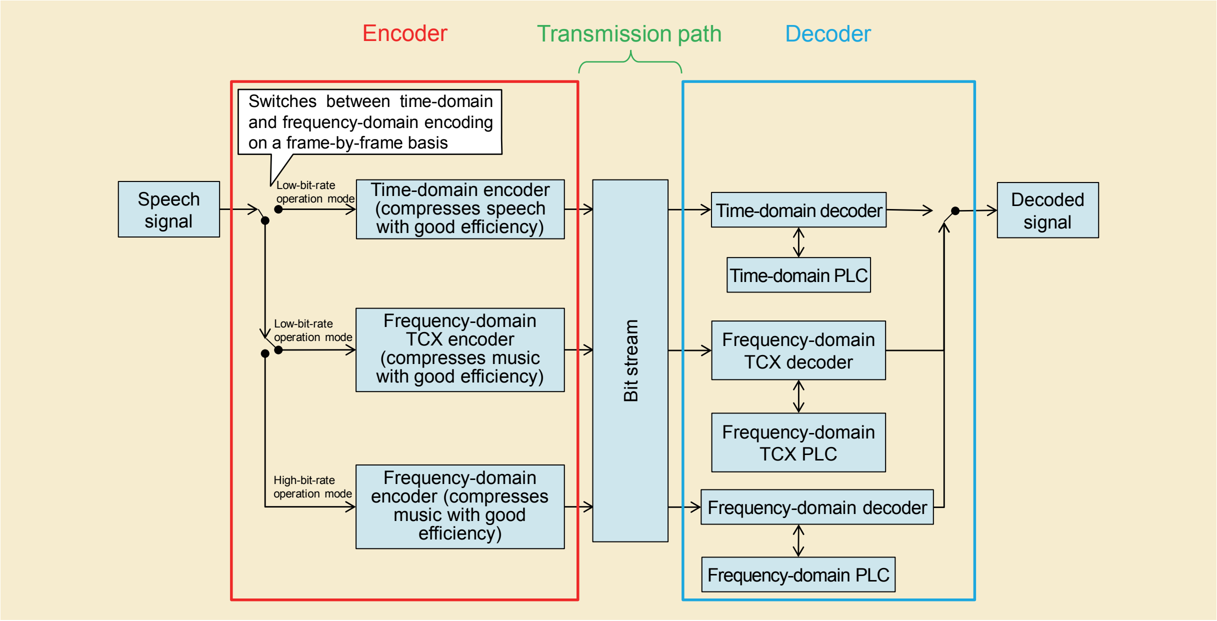
Topics



(P.31)

News

(P.36)



Technology Reports 3GPP EVS Codec for Unrivaled Speech Quality and Future Audio Communication over VoLTE (P.4)

Basic configuration of EVS encoder/decoder

EVS | Speech Coding | Standardization

Technology Reports

3GPP EVS Codec for Unrivaled Speech Quality and Future Audio Communication over VoLTE

Research Laboratories *Kimitaka Tsutsumi*
Kei Kikuiri

NTT DOCOMO has been engaged in the standardization of the 3GPP EVS codec, which is designed specifically for VoLTE to further enhance speech quality, and has contributed to establishing a far-sighted strategy for making the EVS codec cover a variety of future communication services. NTT DOCOMO has also proposed technical solutions that provide speech quality as high as FM radio broadcasts and that achieve both high coding efficiency and high audio quality not possible with any of the state-of-the-art speech codecs. The EVS codec will drive the emergence of a new style of speech communication entertainment that will combine BGM, sound effects, and voice in novel ways for mobile users.

1. Introduction

The launch of Voice over LTE (VoLTE) services and flat-rate voice service has demonstrated the importance of high-quality telephony service to mobile users. In line with this trend, the 3rd Generation Partnership Project (3GPP) completed the standardization of the speech codec for Enhanced Voice Services (EVS) [1] in September 2014.

The speech quality of existing telephony service has been as high as AM-radio quality*1 due to speech codecs such as Adaptive Multi-Rate - Wide-Band (AMR-WB)*2 [2] that is used in NTT DOCOMO's VoLTE and that support wideband speech with a sampling frequency*3 of 16 kHz. In contrast, EVS has been designed to support super-wideband*4 speech with a sampling frequency of 32 kHz thereby achieving speech of FM-radio quality*5. The introduction of EVS in VoLTE can therefore be expected to provide a quantum leap in the quality of telephony services. Additionally, considering the appearance of Moving Picture Experts Group Unified Speech and Audio Coding (MPEG USAC)*6 [3] [4], which in addition to speech can also encode music at high levels of quality and efficiency for non-real-time services, 3GPP experts agreed to adopt high requirements in the EVS codec for music despite its main target of real-time communication. Furthermore, considering that telephony services using AMR-WB are finding widespread use around the world, EVS is required to have a mode compatible with AMR-WB [5].

NTT DOCOMO has been participating in EVS standardization activities since 2010 emphasizing that the goal should be for early and widespread penetration. This would be accomplished

*1 **AM-radio quality:** The capability of expressing a speech band from 100 Hz to 7.5 kHz.

*2 **AMR-WB:** A speech codec used in, for example, telephony services, having a better quality than that of AMR-NB.

*3 **Sampling frequency:** A unit (in Hz) expressing the number of times per second that the acoustic pressure of a speech signal input from a microphone is recorded. A higher frequency enables a higher range of sounds to be recorded.

by establishing EVS design requirements that minimize the changes that would have to be made to the network when introducing EVS in VoLTE. In line with this argument, the EVS bit rate has been set so that data size transported over the radio interface in VoLTE could be the same as that of AMR-WB thereby enabling the design of the VoLTE radio network for AMR-WB to be unchanged. Moreover, envisioning the further evolution of VoLTE due to the future merging of voice and music services, emphasis has been placed on music quality at low bit rates. In the beginning, discussions proceeded in the direction that requirements for music should be set low relative to those for speech. In the end, however, high requirements were set beneficial to music coding specifying that audio quality should be on the same level as a codec having an algorithmic delay*7 longer than that of the EVS codec. In this regard, NTT DOCOMO's Melody Call®*8 (musical ring tone service) has been used widely in existing 3G and VoLTE telephony services, but since it uses the speech-specific AMR codec [6], it cannot necessarily provide high-quality music. It can therefore be expected that EVS satisfying the above high requirements should not only improve the quality of such existing music-content services but also facilitate the appearance of new services that use music as part of telephony services [7].

As a result of the above discussions at 3GPP, EVS has been standardized as a codec having three key features: (1) support of super-wideband speech having FM-radio quality, (2) easy implementation in VoLTE, and (3) high audio quality for both speech and music.

In this article, we first provide an overview of EVS. We then describe the main technologies used for improving quality and the technologies that NTT DOCOMO has contributed. Finally, we present the results of quality evaluation tests for super-wideband speech and music as a part of assessing EVS performance in the characterization phase.

2. EVS Technical Features

2.1 EVS Overview

In addition to wideband and super-wideband audio bandwidths, EVS also supports narrowband speech having a sampling frequency of 8 kHz and full-band speech having a sampling frequency of 48 kHz higher than that of CDs. The EVS codec covers a wide range of bit-rates from 5.9 to 128 kbps.

A high-level overview of the EVS codec is shown in **Figure 1**. In low-bit-rate operational modes, signal classification is performed on a frame-by-frame*9 basis to decide which coding strategy to use: time-domain coding that provides high quality and efficiency for speech or frequency-domain coding for music. In high-bit-rate operational mode in which a substantial amount of information is available, only the frequency-domain coding is used regardless of the type of input signal.

Since it is impossible to avoid packet loss*10 in a packet-switched network such as VoLTE, Packet Loss Concealment (PLC) that reconstructs missing frames and ensures rapid and smooth recovery is important for providing high quality even under erroneous channel conditions. The PLC technique for the EVS codec has also been developed as part of the standard [8]. PLC for the EVS codec switches its strategy between time-domain and frequency-domain based on the coding strategy of the last frame normally decoded before packet loss.

The following provides an overview of time-domain coding, frequency-domain coding, and the PLC technique.

1) Time-domain Coding

An overview of time-domain coding is shown in **Figure 2**. Human hearing is more sensitive to lower-frequency components than higher-frequency components. Therefore, the total amount of information can be efficiently reduced while maintaining speech quality by coding low-frequency and high-frequency components separately with uneven bit allocation, which assigns a lot less bits to high-frequency components.

(1) In the EVS codec, Code Excited Linear Prediction (CELP) is used to encode lower-frequency com-

*4 **Super-wideband:** A speech band with lower and upper frequency limits of 50 Hz and 16 kHz, respectively.

*5 **FM-radio quality:** The capability of expressing a speech band from 50 Hz to 15 kHz.

*6 **MPEG USAC:** MPEG unified speech and audio codec. MPEG is a set of standards specifying coding and transmission systems for digital audio and video. It was formed by a ISO/IEC joint working group.

*7 **Algorithmic delay:** An index indicating the delay in outputting decoded sound with respect to the original sound. It is determined by codec specifications, and in the case of a codec in the frequency domain, a longer delay can generally improve coding efficiency.

*8 **Melody Call®:** A NTT DOCOMO service that enables the user to change the ring tone on the mobile phone to his/her favorite tunes. A registered trademark of NTT DOCOMO, Inc.

*9 **Frame:** The period in which an encoder/decoder operates or a speech signal of a length corresponding to that period. In the EVS codec, the frame length is 20 ms, which means that encoding/decoding is performed once every 20 ms.

*10 **Packet loss:** The failure of a speech packet to be delivered as far as the decoding stage due to congestion or other problems.

ponents to obtain linear prediction*11 coefficients and a linear prediction residual signal*12 after quantization*13. The encoding process for the linear prediction residual signal exploits the property of speech signals that similar waveforms (each being a pitch waveform*14) are repeated along

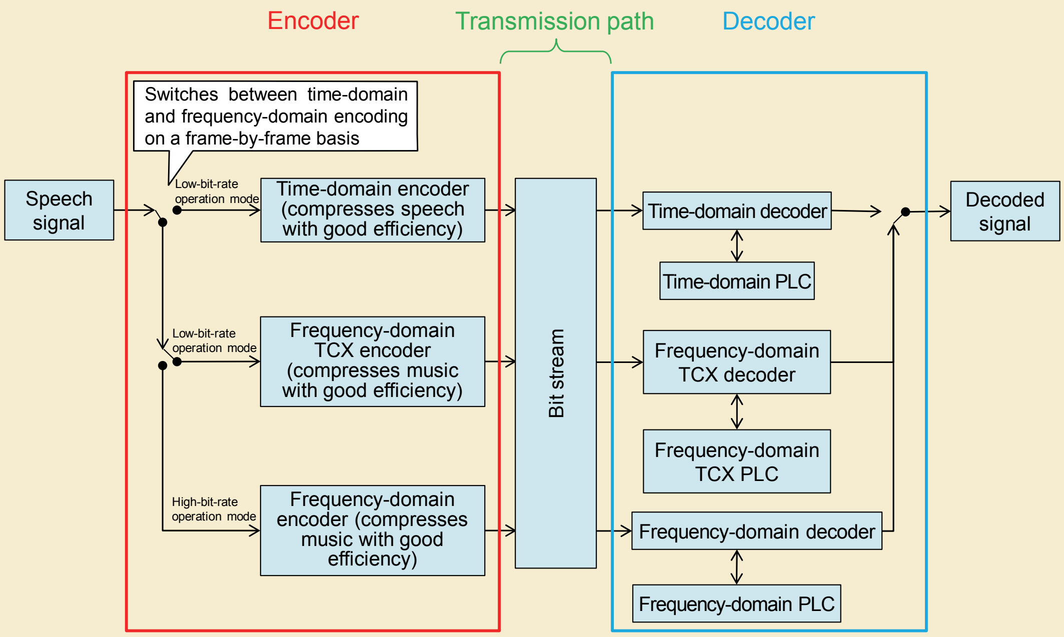


Figure 1 Basic configuration of EVS encoder/decoder

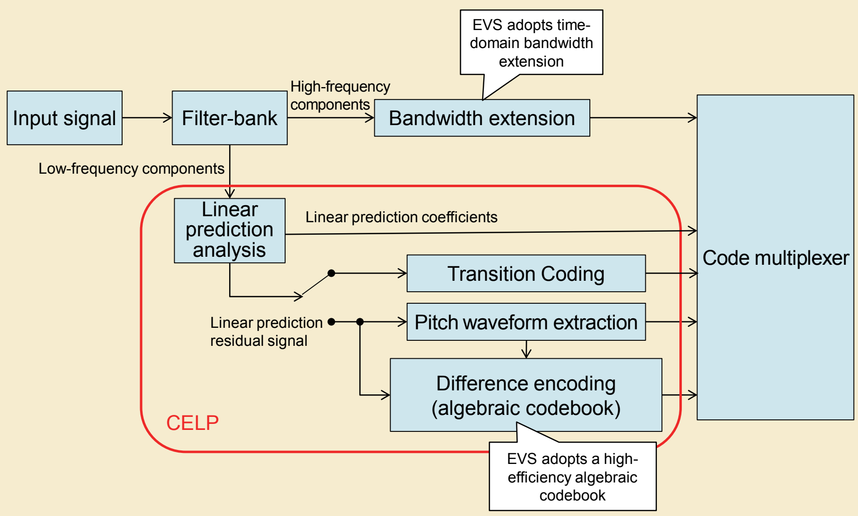


Figure 2 Configuration of time-domain encoding

*11 **Linear prediction:** A technique that approximates the speech signal at a certain point in time by taking a linear sum of previous speech signals.

*12 **Linear prediction residual signal:** The signal representing the prediction error when applying linear prediction to the input signal.

*13 **Quantization:** A process of mapping input values to a smaller set of predetermined discrete values. While resulting in some distortion, quantization can significantly reduce the amount of information.

*14 **Pitch waveform:** One period's worth of similar repeating waveforms that characterize speech.

the time axis. In the encoding, the difference with the immediately preceding pitch waveform together with the length of the pitch waveform is encoded. The difference is encoded by an algebraic codebook*15, which is a technique incorporated in many recent speech codec standards including AMR and AMR-WB. In the EVS codec, coding-efficiency of the algebraic codebook is significantly improved to provide higher sound quality even at low bit rates.

(2) Higher-frequency components encoding is based on a band extension technique that produces higher-frequency components by replicating and shaping lower-frequency components. Only the parameter for the shaping is transmitted at a low bit rate by the encoder, and higher-frequency components are reconstructed based on this parameter by the decoder. Therefore, high-frequency components can be obtained even at low bit rates. In the EVS codec, time-domain band extension encoding is also used to achieve high quality with low delay.

2) Frequency-domain Coding

An overview of frequency-domain coding is shown in **Figure 3**. The input signal is first transformed into a frequency-domain representation using the Modified Discrete Cosine Transform (MDCT)*16 and then encoded.

There are two methods for encoding these MDCT coefficients. The first method divides the MDCT coefficients into sub-bands*17, and obtains the scale factor*18 of each sub-band and the MDCT coefficients normalized by those scale factors. The normalized MDCT coefficients are encoded with vector quantization*19. The second method is Transformed Code Excitation (TCX) that encodes the linear prediction residual signal in the frequency domain.

TCX is also used in MPEG USAC. In the EVS codec, arithmetic coding is improved to obtain enhanced error resistance while maintaining coding efficiency.

In TCX, all MDCT coefficients cannot be encoded due to limited bits for coding and those that cannot are quantized to zero. This results in frequency bands with

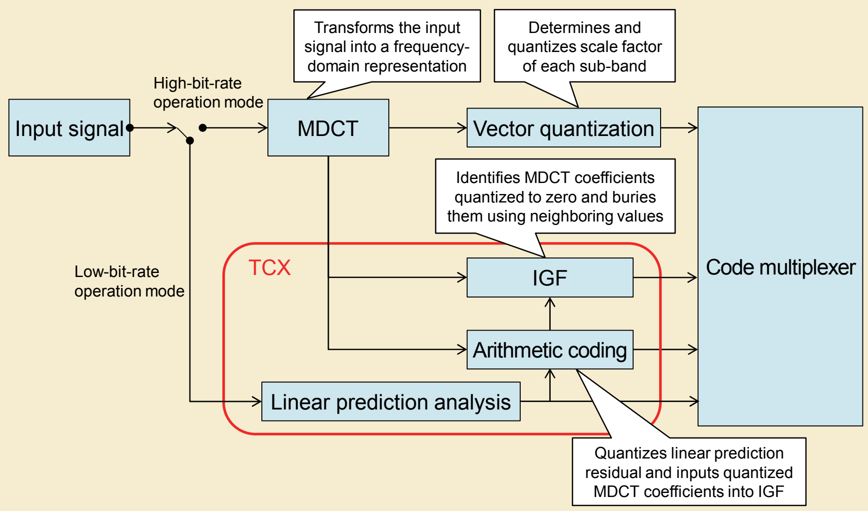


Figure 3 Configuration of frequency-domain encoding

*15 **Codebook:** A set of previously determined candidate vectors for quantizing input vectors.

*16 **MDCT:** A method for converting a time-series signal to its frequency components. It is able to reduce distortion at frame boundaries without losing information by applying an overlapping transform with the preceding and following frames, so it is widely used for audio coding.

*17 **Sub-band:** One of the bands that result from splitting an entire frequency band into multiple parts.

*18 **Scale factor:** A power of a sub-band or its quantized amplitude.

*19 **Vector quantization:** A quantization technique that maps a numerical sequence of length two or more to the closest value of predetermined numerical sequences of the same length.

no signal components and introduces sound quality degradation. To mitigate this degradation, conventional codecs have used a noise filling technique that buries noise in such frequency bands. In EVS, Intelligent Gap Filling (IGF) has been adopted to fill those bands with nearby MDCT coefficients.

3) Packet Loss Concealment

(1) Time-domain PLC

CELP is based on inter-frame prediction*20, which results in error propagation even after receiving packets followed by a packet loss. For smooth and fast recovery of linear prediction coefficients and a linear prediction residual signal, Transition Coding mode*21 (fig. 2) is used in the EVS codec. In this mode, the linear prediction coefficients and the linear prediction residual signal are encoded independently from those in the previous frame thereby improving packet loss resilience.

(2) Frequency-domain PLC

The conventional frequency-domain PLC technique basically copies the MDCT coefficients in the last frame before the packet loss as a substitute for the parameter in the lost frame. However, repetition of the MDCT coefficients in the past frame sometimes introduces discontinuities in the waveform. In the EVS codec, waveform adjustment based on phase control is used for smooth connection between a concealed frame and its adjacent frames.

2.2 Quality-improvement Technologies in EVS

A variety of technologies have been introduced in EVS for speech quality improvement. The following provides an overview of those that are particularly important.

1) Time-domain Bandwidth Extension

Bandwidth extension is a technology that generates higher-frequency components at low bit rates. Higher-frequency components are first generated using lower-frequency components and then shaped so as to have a power distribution indicated by the encoder.

In time-domain bandwidth extension in the EVS encoder, a linear prediction spectrum*22 is calculated based on higher-frequency components from a filter-bank*23 and coded to indicate a power distribution of higher-frequency components. In the decoder, lower-frequency components from the CELP decoder are modified to obtain an excitation signal, and then the excitation signal is fed to a synthesis filter having the decoded linear prediction spectrum (**Figure 4**) to obtain

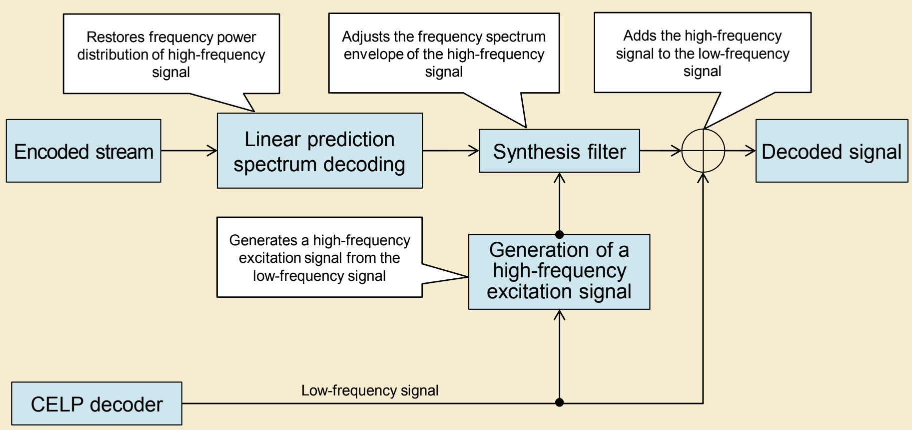


Figure 4 Time-domain bandwidth extension decoding

*20 **Inter-frame prediction:** A technique for improving coding efficiency by quantizing the difference between the values of the current frame and that of the previous frame.

*21 **Transition Coding mode:** A coding mode of Algebraic CELP (ACELP) designed to eliminate inter-frame dependency as much as possible and to control error propagation. The encoder is designed to select this mode in the frame following the frame that includes the onset of speech.

*22 **Linear prediction spectrum:** The frequency spectrum of an IIR filter determined by linear prediction coefficients.

*23 **Filter-bank:** A series of digital filters that split an input signal into multiple frequency bands.

higher-frequency components having a power distribution indicated by the encoder. This scheme enables the encoding of high-frequency components using only a limited amount of information with low computational complexity.

2) Arithmetic Coding

As described in section 2.1 (2), arithmetic coding is used for the quantization of the linear prediction residual in the frequency domain.

As shown in **Figure 5**, a bit plane is first created based on two adjacent MDCT coefficients in binary, and then higher-order bits are encoded with a codebook that is selected based on the immediately preceding results of quantization. Finally, lower-order bits are encoded according to the number of remaining bits available for use.

For higher bit rates in which a sufficient number of linear prediction residual signals can be encoded, the codebook is determined based on the linear prediction residual obtained by decoding in the previous frame. For lower bit rates in which a sufficient number of linear prediction residual signals cannot be obtained, the codebook is determined based on the linear prediction spectrum.

3) IGF

As shown in **Figure 6**, the IGF technique copies adjacent decoded MDCT coefficients to a missing frequency band that could not be encoded. However, if the MDCT coefficients at the copy source have strong peaks, those peaks will also be generated at upper frequencies at the copy destination resulting in degraded sound quality. To suppress such unnecessary strong peaks in the high-frequency band, whitening[*24] is performed in

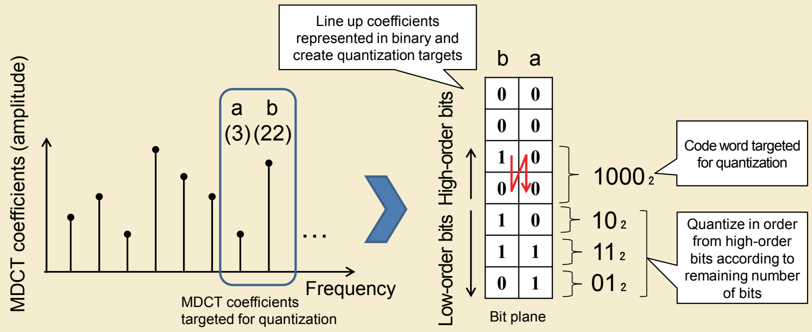


Figure 5 Arithmetic coding

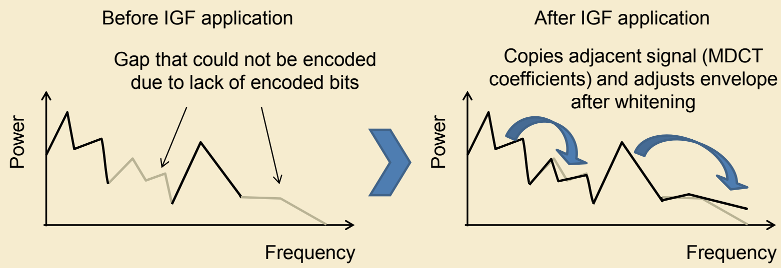


Figure 6 IGF

*24 **Whitening:** Processing for making the frequency distribution of signal power uniform.

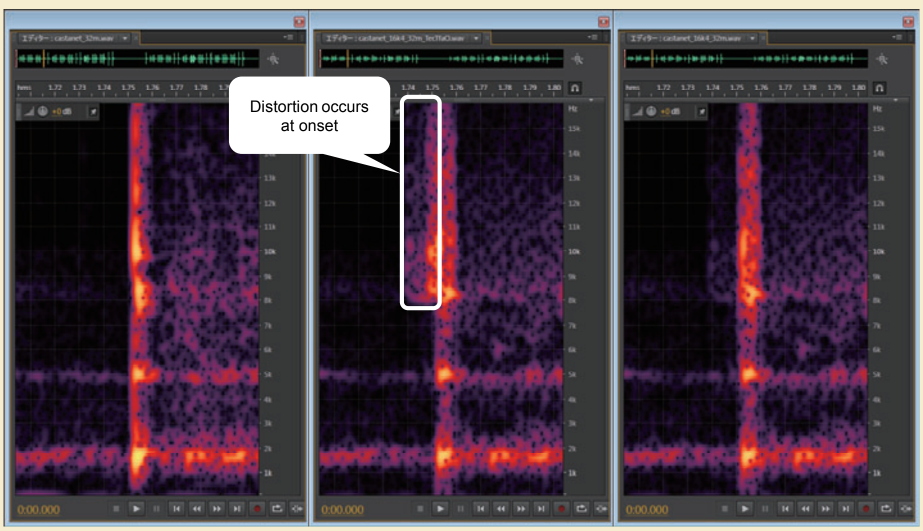
the frequency domain as needed.

On the encoder side, IGF encodes and transmits the rough shape of the frequency spectrum and ON/OFF information for whitening, and on the decoder side, it adjusts the MDCT coefficients to reconstruct the frequency spectrum. In this way, IGF can perform high-quality audio encoding especially for music at low bit rates.

2.3 Technical Contribution from NTT DOCOMO

1) Technology for Improving Higher-frequency Components

The encoding of high-frequency components in EVS is often done by bandwidth extension using low-frequency components for time-domain encoding and by IGF using components of another frequency band for frequency-domain encoding. Thus, a mismatch may occur in the power distribution in the temporal direction between the higher-frequency components of the input signal and the signal components of the frequency band that becomes the basis for encoding high-frequency components, or distortion may occur in the power distribution in the temporal direction when encoding high-frequency components. These issues are addressed by the following method. In the encoder, the method detects whether the power distribution in the temporal direction of the high-frequency components of the input signal is flat and transmits the result, and in the decoder, the method uses that result as a basis for performing applicable processing such as flattening of the high-frequency components. As a result, a match can be achieved with the power distribution of the input signal in the temporal direction. This flattening process is performed either in the time domain or frequency domain according to which coding strategy is selected. If time-domain coding is selected, the encoder transmits the result of detecting a sudden increase in power distribution in the temporal direction, and the decoder adjusts the increase accordingly to reconstruct that power fluctuation. The spectrograms in **Figure 7** (b) and (c) depict the temporal change in a signal's frequency spectrum before and after adjusting the power increase shown in fig. 7 (a) by this technology.



(a) Original sound (b) Before application (c) After application

Figure 7 Effect of technology for improving high-frequency components

In the figure, the vertical and horizontal axes represent frequency and time, respectively, and a change from dark to bright colors represents an increase in power. The results shown in the figure reveal that distortion occurs at high frequencies prior to the signal's sudden power increase before applying this technology and that this distortion is suppressed after applying the technology.

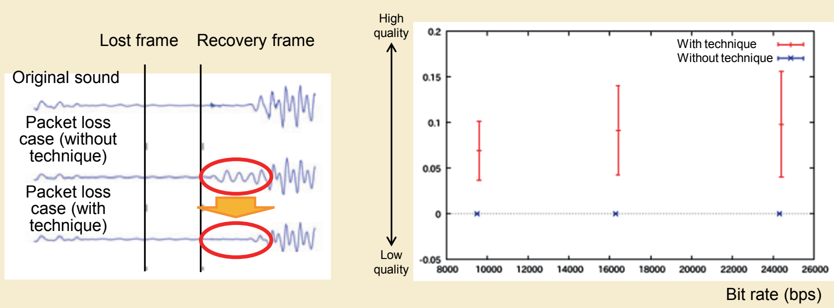
2) PLC Technique

In CELP, pitch lag estimation error caused by packet loss can result in audio discontinuities. To obtain linear prediction coefficients that minimize coding distortion, a portion of the frame following the frame targeted for encoding is exploited for linear predictive analysis as a look-ahead signal. Transmitting the pitch lag calculated for this look-ahead signal provides a pitch estimate for the next frame without adding extra delay and improves the accuracy of pitch-lag estimation under packet loss conditions.

Additionally, since linear prediction coefficients are encoded based on inter-frame prediction in the EVS codec, the linear prediction filter can be unstable at the recovery frame after a packet loss especially at the onset where spectral energy of speech increases from zero. This sometimes causes a large ripple in the decoded signal. To deal with this problem, the linear prediction filter on the decoder side is modified based on auxiliary information from the encoder to prevent it from having excessive gain. In the encoder, frames at which the filter can be unstable are first detected by simulating a linear prediction filter at the lost frame, and the result is transmitted as auxiliary information. In the decoder, the linear prediction filter is modified based on those detection results transmitted as auxiliary information. This technique prevents perceptual quality degradation caused by ripples as shown in **Figure 8** (a). We evaluated the effectiveness of the technique by Perceptual Evaluation of Speech Quality (PESQ)*25 [9]. For this evaluation, we created an error pattern in which the onset of speech is lost. The difference in the scores between with and without the proposed technique is shown in Fig. 8 (b). The error bar*26 represents a 95% confidence interval. These results indicate a significant improvement in sound quality.

3) Technology for Reducing Computational Complexity

When switching between low-bit-rate and high-bit-rate operational modes, linear prediction coefficients need to be recomputed since the internal sampling rate changes. However, if the conventional method is employed for this purpose, a large amount of calculations will be needed. For this reason, linear prediction coefficients are computed here by performing resampling*27 in the frequency domain on the linear prediction spectrum. This approach reduces computa-



(a) Speech waveform on applying technique

(b) Difference in evaluation scores with/without applying the proposed technique

Figure 8　Effect of PLC technique proposed by NTT DOCOMO

*25 **PESQ:** A method of subjective evaluation that estimates speech quality from the difference between a reference signal and the signal being tested.

*26 **Error bar:** A graphical bar that indicates the range of error.

*27 **Resampling:** Performing sampling a second time using a different sampling frequency after returning the digital signal to an analog signal.

tional complexity to around one-third that of conventional values.

3. EVS Performance

3.1 EVS Selection Tests

In the EVS selection phase, subjective quality assessments [10] consisting of a total of 24 tests were conducted at three evaluation institutions with 32 subjects participating in each test. For super-wideband input, a Degradation Category Rating (DCR) test*28 that evaluates degradation from the original sound on five levels was performed multiple times. In the tests, clean speech, noisy speech, music, and clean speech under packet loss conditions were tested [11].

3.2 Evaluation Results

The results of these selection tests are partially shown in **Figure 9**. In the figure, each error bar represents a 95% confidence interval. First, for the tests targeting clean speech, EVS achieved a level of quality equivalent to or greater than reference codec G.722.1 Annex C*29 [12] even at half the bit rate.

For the tests targeting noisy speech, which simulates actual use case in real life, EVS at 13.2 kbps achieved a level of quality significantly higher than AMR-WB at its maximum bit rate of 23.85 kbps. This result indicates that sound quality significantly improves when migrating from AMR-WB to EVS. Similarly, test results revealed an improvement in performance when using EVS for music and under packet-loss conditions. In this regard, AMR-WB+*30 is a codec having a long delay of 80 ms, but EVS could nevertheless achieve an equivalent level of quality with a relatively short delay of 32 ms.

4. Conclusion

This article provided an overview of EVS, described the major technologies making it possible, and presented the results of subjective quality assessment performed in the EVS selection phase. EVS achieves a level of quality as high as FM-radio in telephony and is easy to implement in VoLTE. It is a codec that can achieve a high level of quality for both speech and music at low bit rates. These features can, of course, be leveraged to raise the sound quality of telephony and existing services that make use of music content as in Melody Call®, but they are also expected to give

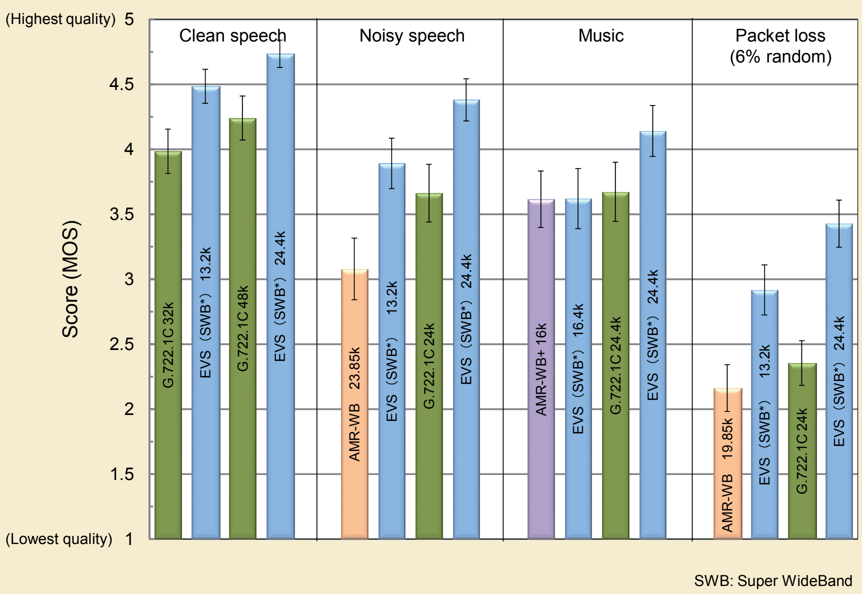


Figure 9 Results of subjective evaluation tests

*28 **DCR test:** A method of subjective evaluation that measures the extent to which the target signal is degraded with respect to a reference signal that represents base quality. A subject listens to both the reference signal and the target signal. This method is specified in ITU-T P.800.

*29 **G.722.1 Annex C:** A speech codec supporting super-wideband signals standardized by ITU-T and used in voice conferencing equipment from Polycom, Inc.

*30 **AMR-WB+:** An extended coding scheme of AMR-WB, the speech coding scheme standardized by 3GPP, which enables it to be used for general audio signals such as music.

birth to a new style of mobile audio communications.

References

[1] 3GPP TS26.441 V12.0.0: "Speech codec speech processing functions; Enhanced Voice Service (EVS) speech codec; General description," 2014.

[2] 3GPP TS26.171 V12.0.0: "Speech codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; General description," 2014.

[3] ISO/IEC 23003-3: "Information technology - MPEG audio technologies - Part 3: Unified speech and audio coding," 2012.

[4] K. Kikuiri et al.: "MPEG Unified Speech and Audio Coding Enabling Efficient Coding of both Speech and Music," NTT DOCOMO Technical Journal, Vol.13, No. 3, pp. 17-22, Dec. 2011.

[5] 3GPP TR22.813 V10.0.0: "Study of Use Cases and Requirements for Enhanced Voice Codecs for the Evolved Packet System (EPS)," 2010.

[6] 3GPP TS26.071 V12.0.0: "Mandatory speech CODEC speech processing functions; AMR speech Codec; General description," 2014.

[7] Kaneko et al.: "VoLTE-compatible High-quality Melody Call," NTT DOCOMO Technical Journal, Vol. 22, No. 4, pp. 29-33, Jan. 2014. (in Japanese)

[8] 3GPP TS26.447 V12.0.0: "Codec for Enhanced Voice Services (EVS); Error Concealment of Lost Packets," 2014.

[9] ITU-T P.862.2: "Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs," 2007.

[10] ITU-T P.800: "Methods for subjective determination of transmission quality," 1996.

[11] 3GPP AHEVS-311: "EVS Permanent Document EVS-8b: Test plans for selection phase including lab task specification," 2014.

[12] ITU-T G.722.1: "Low complexity coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss," 2005.

Base Station | Active Antenna | Standardization

Technology Reports

Active Antenna for More Advanced and Economical Radio Base Stations — Connection with LTE Base Station and Evaluation of Service Area Quality by Field Experiment —

Active antennas that integrate radio transceiver functions in the antenna unit have been attracting attention as an approach to furthering the evolution of radio base stations. Compared to existing base stations, a base station using an active antenna can provide a higher-quality service area, reduce installation space, and improve power efficiency. This article describes the features of a base station using an active antenna, presents an overview and the results of a field experiment conducted by NTT DOCOMO, and outlines standardization trends at 3GPP.

Radio Access Network Development Department

Sho Yoshida
Teruo Kawamura
Teruaki Toeda
Huiling Jiang †

1. Introduction

Radio base stations continue to evolve to provide a better wireless communication environment to customers as mobile communication systems continue to evolve to meet the growing demand for mobile traffic. In recent years, attention has come to be focused on active antennas that integrate radio transceiver functions in the antenna unit as one approach to furthering the evolution of base stations. Standardization work for specifying radio characteristics of a base station using an active antenna is now in progress at the 3rd Generation Partnership Project (3GPP). A base station using an active antenna features a higher-quality service area, a smaller installation space, and improved power efficiency compared to existing base stations, and these features make active antennas a promising technology for the future.

At NTT DOCOMO, we have successfully connected an active antenna to a commercial LTE base station equipment via a standard interface and conducted for the first time in Japan a field experiment using an active-antenna base station [1] [2].

In this article, we describe the technical features of a base station using an active antenna, provide an overview of a field experiment conducted by NTT DOCOMO, and present experimental results. We also outline standardization trends at 3GPP.

2. Features of an Active-antenna Base Station

2.1 Basic Configuration of Base Station using an Active Antenna

The basic configuration and features of an active-antenna base station are shown in **Figure 1**. An active antenna consists of multiple antenna elements and corresponding compact radios as well as a controller for these radios. An

† Currently DOCOMO Beijing Communications Laboratories Co., Ltd.

antenna element acts as an outlet/inlet for radio waves and a compact radio performs transmit/receive signal processing such as digital-to-analog conversion, frequency conversion, and power amplification. The controller, meanwhile, performs digital control of excitation coefficients*1 given to each antenna element and combines/divides digital signals from/to each antenna element. In addition, a BaseBand Unit (BBU)*2 performs digital signal processing of transmit/receive information during communications with a mobile terminal and an optical fiber cable connects the active antenna to the BBU to transmit digital signals.

When changing the antenna beam tilt*3 to adjust the service area radius covered by the base station, a conventional base-station antenna generally uses an analog variable phase shifter, which is a device used to change the relative excitation phase difference*4 between antenna elements. An active antenna, on the other hand, enables the excitation coefficients of each antenna element to be separately controlled by equipping each antenna element with a compact radio. This means a much higher degree of freedom in controlling excitation coefficients compared to an analog variable phase shifter, and this, in turn, means that antenna directivity*5 can be controlled with more flexibility enabling the design of high-quality service areas. For example, referring again to Fig. 1, the tilt range can be expanded, different tilts can be set for the downlink transmission and uplink reception of radio signals at the base station, and tilts can be separately set for different Radio Access Technologies (RAT) such as LTE and W-CDMA, all without having to mount multiple analog variable phase shifters.

To obtain the desired antenna beam pattern, the amplitudes and phases of the antenna elements must be appropriately adjusted. Since an active antenna incorporates multiple radios, noise caused by the power amplification circuits of each radio generates time-varying excitation errors among antenna elements. An active antenna may therefore have a calibration function for correcting these excitation errors. This calibration helps

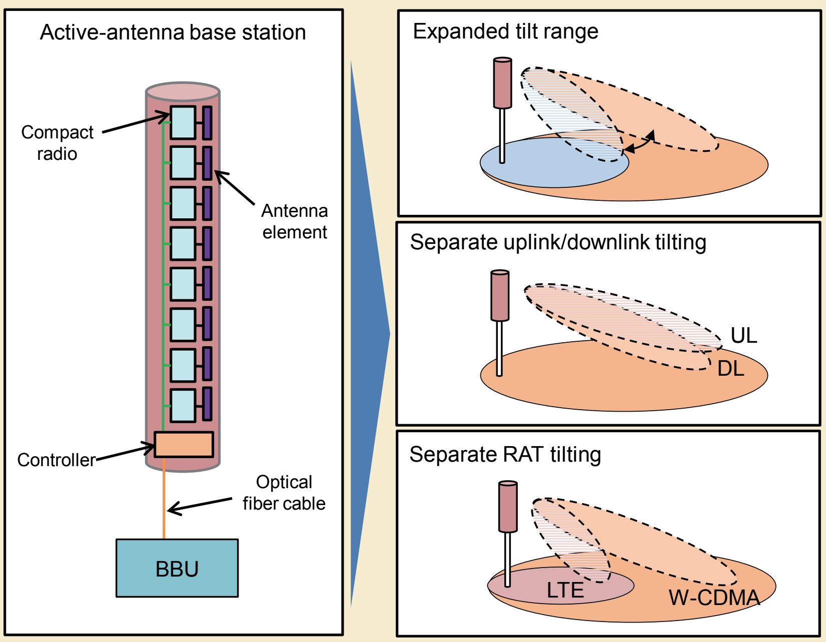


Figure 1 Configuration and features of active-antenna base station

*1 **Excitation coefficients:** Phase and amplitude information given to each antenna element.

*2 **BBU:** One component of base station equipment performing digital signal processing of transmit/receive information when communicating with a mobile terminal.

*3 **Tilt:** Inclination of antenna beam in the vertical plane. If the horizontal direction is designated as 0°, increasing or decreasing the tilt angle changes the communication area.

*4 **Excitation phase difference:** Phase difference between signals that antenna elements radiate or receive.

*5 **Antenna directivity:** The directional characteristics of the radiated or received strength of the antenna.

to achieve stable antenna directivity.

2.2 Advantages of Active Antennas in a Remote-type Base Station Configuration

One deployment configuration of a base station is the remote-installation type (optical-fiber-connected base station) consisting of a master station and multiple slave stations as shown in **Figure 2**. In a conventional base station of this type, a slave station consists of a Remote Radio Head (RRH)*6 performing transmit/receive signal processing and installed apart from the master station (BBU), and an antenna installed near the RRH. The BBU and RRH exchange digital signals through an optical fiber cable and an RRH and its associated antenna exchange radio signals through a Radio Frequency (RF)*7 coaxial cable. In contrast, since a base station using active antennas integrates the RRH function of conventional base stations in the antenna itself, the slave station can consist of one active antenna. In the case of an active antenna mounting multiple radios as shown in Fig. 1, the maximum transmit power required by a single radio is small, so each radio can generally be made small. This means that the total volume of the active antenna can likewise be made small. Furthermore, as there is no need for installing RRH, the entire base station size can be reduced compared to a conventional base station. This is advantageous for installing base stations at locations having limited space for installing equipment as in an urban area. Moreover, as an active antenna can be directly connected to the BBU via an optical fiber cable to transmit digital signals and as the excitation coefficients for each antenna element can be digitally controlled, electrical loss associated with RF coaxial cables and analog variable phase shifters in a conventional antenna configuration can be reduced thereby enhancing power efficiency. As a consequence, the service area covered by a single base station can be expanded and area quality improved while operating costs can be reduced through power savings.

3. Connection with BBU via a Standard Interface

NTT DOCOMO has conducted performance evaluations of an active-antenna base station using a prototype active antenna. To connect to the BBU using optical fiber cable, this active antenna supports a global standard interface based on the Open Radio equipment Interface (ORI) whose specifications are being established by the European Telecommunications Standards Institute (ETSI)*8. Referring to **Figure 3**, if the

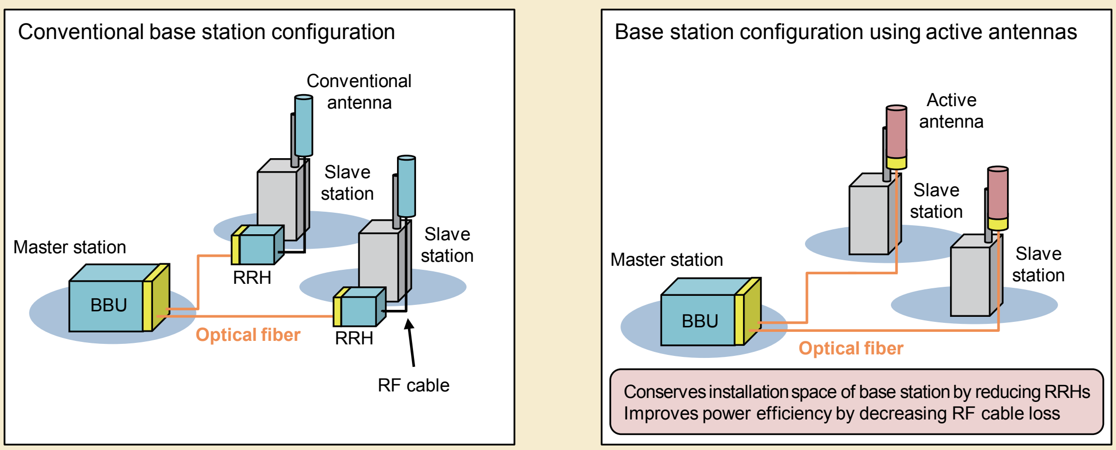


Figure 2 Conventional configuration and active-antenna configuration of a remote-type base station

*6 **RRH:** One component of base station equipment installed at a distance from the BBU using optical fiber or other means. It serves as radio equipment for transmitting/receiving radio signals.

*7 **RF:** The frequency range used in radio communications.

*8 **ETSI:** The standardization organization concerned with telecommunications technology in Europe.

active antenna that will be used to replace the slave station in a conventional remote-type base station has an equipment-specific interface, the BBU (master station) will also have to be replaced with equipment supporting that active antenna. However, if a standard interface is supported, compatibility can be achieved between the active antenna and a BBU from a different vendor making it unnecessary to replace the existing BBU and making it possible to inexpensively and quickly deploy active antennas. NTT DOCOMO has successfully connected a prototype active antenna to LTE base-station equipment used in its commercial network via an ORI-standard interface.

4. Field Experiment

4.1 Purpose of Experiment

At NTT DOCOMO, we conducted a field experiment using an experimental station consisting of LTE base-station equipment and a prototype active antenna. Our purpose here was to test the power-efficiency improvement effect achieved by a reduction in electrical loss, which is one of the key features of an active-antenna base station. Specifically, to clarify the amount of improvement achieved in comparison with a conventional base station, we also installed a conventional antenna (hereinafter referred to as "passive antenna" in contrast to "active antenna") designed so that basic antenna specifications such as antenna gain*9 and half-power beam width*10 were equal to those of the active antenna. We were then able to compare communications quality and service-area range in the downlink between the two types of antennas.

4.2 Configuration of Experimental Station

The major specifications and equipment configuration of the experimental

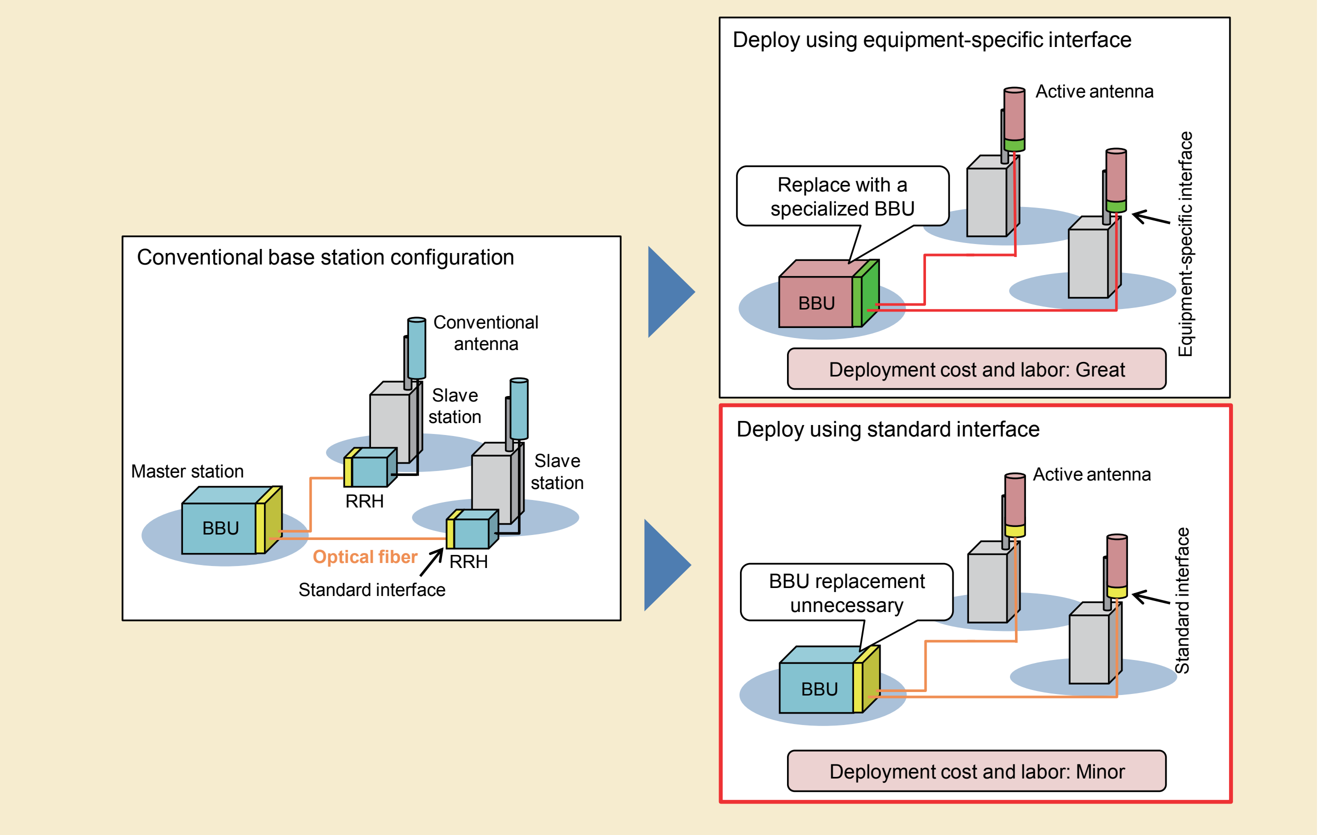


Figure 3 Efficient deployment of active antennas through a standard interface

*9 **Antenna gain:** Radiated power in the direction of maximum radiation usually expressed as the ratio of radiated power to that of an isotropic antenna.

*10 **Half-power beam width:** The angle at which radiated power of the antenna is half that in the direction of maximum radiation.

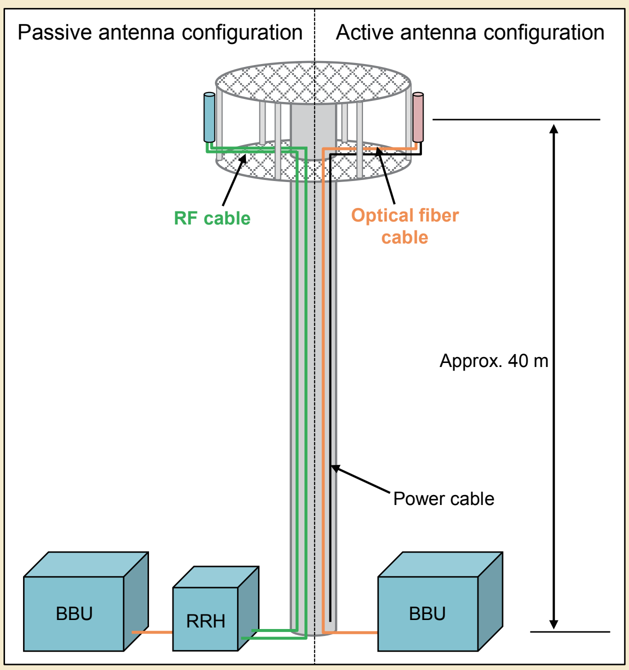
station are given in **Table 1** and **Figure 4**, respectively. The active antenna has an orthogonal polarization configuration[*11] made up of vertical and horizontal polarization and consists of eight antenna elements and eight corresponding compact radios for each polarization. The passive antenna, meanwhile, has the same configuration as the active antenna with respect to antenna elements, and it also mounts an analog variable phase shifter to control tilt angle. The RRH used in the passive-antenna configuration is installed at the foot of an antenna tower. Both antennas are installed at a height of approximately 40 m from the ground and are set to a tilt angle of 8°. The transmit power per polarization, which is the total output power of eight compact radios for the active antenna configuration and the output power of the RRH for the passive antenna configuration, was set to 10 W. The RAT used here was LTE, the radio frequency band was 800 MHz, and the bandwidth was 10 MHz.

Table 1 Major specifications of experimental station

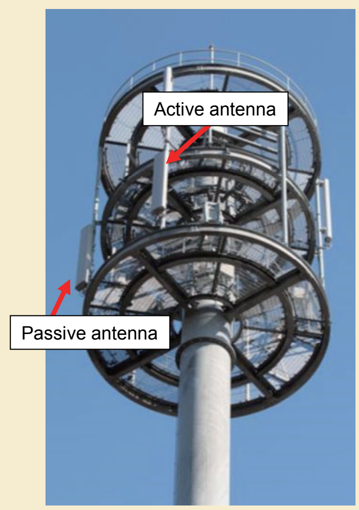
Communications system	LTE
Radio frequency band	800 MHz
Bandwidth	10 MHz
Total transmit power per polarization	10 W
Antenna height	Approx. 40 m
Tilt angle	8°

4.3 Measurement Environment

This field experiment was conducted in a suburb of Chiba City in Chiba Prefecture, Japan. The area surrounding the experimental station was a relatively open environment with few tall buildings. As shown in **Figure 5**, measurements were performed along measurement courses within a short-range area and long-range area at a distance of 200–700 m and 1–3 km, respectively,



Configuration of experimental station



Actual antenna installation

Figure 4 Equipment configuration of experimental station

***11 Orthogonal polarization configuration:** An antenna configuration that can perform transmitting and receiving equivalent to two antennas from a single antenna enclosure by using orthogonal polarization in the vertical/horizontal or ±45° directions.

from the base station. The short-range area was roughly within the range of the main beam's half-power beam width (vertical and horizontal planes). In this area, we evaluated the Reference Signal Received Power (RSRP)*12 and user throughput*13 in the downlink. In the long-range area, we evaluated the range in which communications could be performed.

4.4 Results of Experiment

(1) Short-range area

Measurement results for RSRP and user throughput in the short-range area are shown in **Figure 6**. Measurement values were obtained by taking the average of values measured within a 10-meter-square cell, and these graphs show median values of measurement results obtained within the area. These results show that the median values for RSRP and user throughput improved by approximately 4 dB and 10%, respectively, when using the active-antenna configuration compared with the passive-antenna configuration. In short, for a comparison made within the same area, these results demonstrate that an active-antenna configuration can improve communications quality compared to a passive-antenna configuration.

(2) Long-range area

For the long-range area, the active-antenna configuration and the passive-antenna configuration were compared in terms of the range within which communications could actually be performed. It was found that the range of communications when using a passive-antenna configuration was no greater than 2.5 km from the base station along a straight line and that when using an active-antenna configuration was at least 3 km. These results show that the active-antenna configuration can expand the cell radius covered by a single base station by 1.2 times or more compared with the passive-antenna configuration.

5. 3GPP RAN4 Standardization Trends

The 3GPP Radio Access Network working group 4 (RAN4) is in charge of standardizing RF aspects of Universal Terrestrial Radio Access Network (UTRAN)*14 and Evolved UTRAN (E-UTRAN)*15 in 3GPP. Thus, specifications related to radio characteristics of a base station using the active antenna, which is called Active Antenna System (AAS) in 3GPP, also fall within the scope of this group and Study Item (SI)*16 discussions on those specifica-

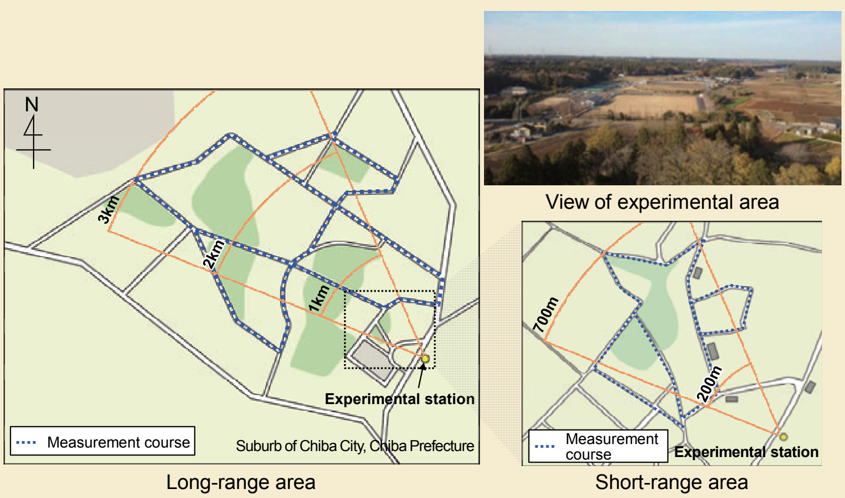


Figure 5 Experimental area

*12 **RSRP:** The received power of a signal measured by a mobile terminal in LTE. Used as an indicator of the receiver sensitivity of a mobile terminal.

*13 **Throughput:** The amount of data transferred through a system without error per unit time.

*14 **UTRAN:** A 3GPP radio access network using the W-CDMA system.

*15 **E-UTRAN:** A 3GPP radio access network using the LTE system.

*16 **SI:** The work of studying an issue in the creation of specifications.

tions began in September 2011. These discussions, which were completed in March 2013, examined differences in transmit/receive signals between existing base stations and AAS. At present, discussions on specifying AAS radio characteristics and measurement methods continue as a Work Item (WI)*17 that began in March 2013. The plan is to complete these WI discussions in early 2015.

5.1 SI Discussions

In the SI, the structure of AAS was first discussed as a basis for future discussions. As shown in **Figure 7**, a consensus was reached on defining the AAS structure as one consisting of a Transceiver Unit Array*18 with K Transceiver Unit(s)*19, an Antenna Array*20 with L antenna element(s), and a Radio Distribution Network (RDN)*21 that divides/combines the signals from Transceiver Unit Array to the Antenna Array and vice versa in a K:L format.

Next, discussions were held on unwanted emissions from the transmitters. The effects of unwanted emissions in AAS were evaluated by simulation, and it was found that AAS causes the same radio quality degradation as existing base stations if the Adjacent Channel Leakage Ratio (ACLR)*22 were specified as 45 dB per transmitter (which is the same value as the requirement for existing base stations). On the receiver side, in-band blocking*23 was also discussed and it was found that the level of interference from other base stations to the AAS would be the same as that from other base stations to existing base stations. Details on the consensus reached in these discussions can be found in a 3GPP Technical Report [3].

5.2 WI discussions

Adding to the SI discussions, simulations on unwanted emissions from the transmitters were performed for a variety of scenarios envisioned for actual propagation environments, and it was agreed that the ACLR requirement for AAS is to be specified as 45 dB.

The requirements for existing base stations are specified at the "antenna connector," which connects the antenna and the radio equipment. An AAS, however, enables antenna directivity and effective radiation gain to be dynamically varied by integrating the antenna and

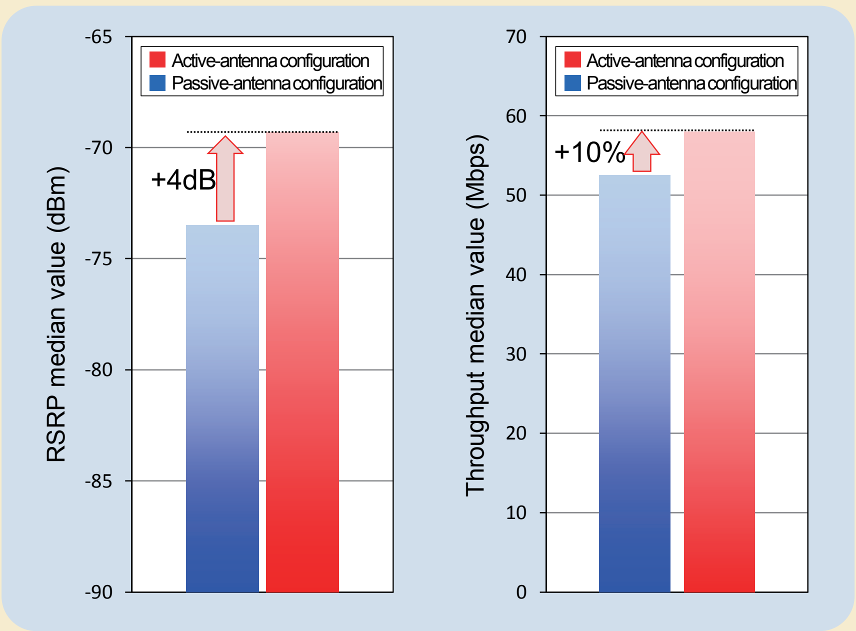


Figure 6 Comparison of communications quality in the downlink (Short-range area)

*17 **WI:** The work of prescribing specifications.
*18 **Transceiver Unit Array:** An array of radios.
*19 **Transceiver Unit:** Equipment that integrates a transmitter and receiver. A radio.
*20 **Antenna Array:** An array of antenna elements.
*21 **RDN:** A logical node situated between and interconnecting the Transceiver Unit Array and Antenna Array.
*22 **ACLR:** Ratio of the wanted signal power to the unwanted emission power in an adjacent channel.
*23 **In-band blocking:** Receiver ability to receive a wanted signal in the presence of an unwanted interferer within the receive bandwidth.

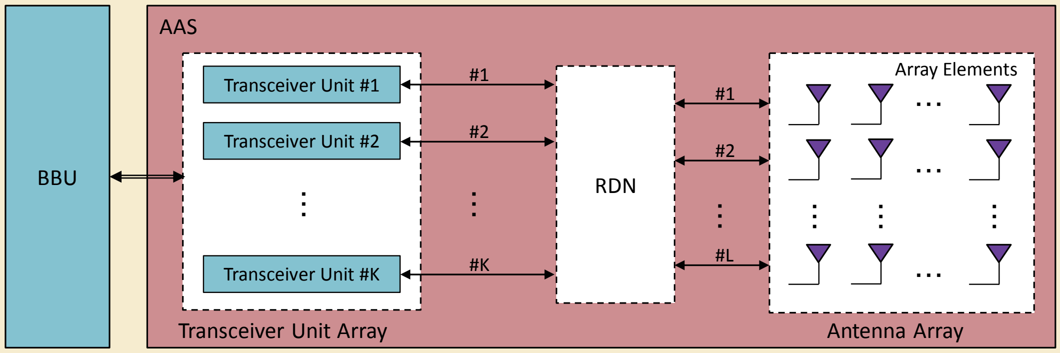


Figure 7 AAS structure agreed upon in RAN4

radio equipment. Discussions are therefore being held on the need for specifying Over The Air (OTA) characteristics such as radiated transmit power requirements and OTA sensitivity requirements in addition to the requirements specified at the antenna connector. Details on the consensus reached up to November 2014 can be found in a 3GPP Technical Report [4].

6. Conclusion

This article described the technical features of an active antenna for furthering the evolution of base stations by integrating radio transceiver functions in the antenna unit. It also presented an overview and results of a field experiment conducted by NTT DOCOMO and outlined standardization trends at 3GPP.

Going forward, we will continue our studies towards introducing an active-antenna base station into commercial networks to further improve the overall quality of mobile communications.

References

[1] M. Murakami, N. Miyadai, S. Yoshida, T. Kawamura and T. Ihara: "Field Trial of Active Antenna System in LTE—Architecture of Active Antenna System in 800 MHz Band—," Proceedings of the 2014 IEICE Society Conference, B-1-150, 2014. (in Japanese)

[2] S. Yoshida, T. Kawamura, T. Ihara, M. Murakami and N. Miyadai: "Field Trial of Active Antenna System in LTE—Configurations of Field Trial Base Station and Downlink Performance Evaluations—," Proceedings of the 2014 IEICE Society Conference, B-1-151, 2014. (in Japanese)

[3] 3GPP TR37.840 V12.1.0: "Study of Radio Frequency (RF) and Electromagnetic Compatibility (EMC) requirements for Active Antenna Array System (AAS) base station," Jan. 2014.

[4] 3GPP TR37.842 V1.2.0: "E-UTRA and UTRA; Radio Frequency (RF) requirement background for Active Antenna System (AAS) Base Station (BS)," Oct. 2014.

SIM BLE Authentication

Technology Reports

Portable SIM: Empowering the User in the IoT Era

The recent widespread use of smartphones is driving the rapid development of new services that use smartphone applications to enhance the user experience. At the same time, the development of devices towards the post-smartphone is accelerating with systems and services that use those devices. NTT DOCOMO has developed a novel SIM-based authentication mini device called Portable SIM that physically separates the SIM function from the smartphone. In this article, we explain the basic configuration and operation of Portable SIM and describe the new ecosystem that Portable SIM will help to create.

Communication Device Development Department *Akira Shibutani*
Kazuma Nachi
Yuta Higuchi
Takashi Okada

1. Introduction

Smartphones are becoming increasingly popular and the development of next-generation (post-smartphone) devices is already underway. Amid this trend, wearable devices such as watch-type, glasses-type, and healthcare-related smart devices are already becoming popular and most of them can be linked to smartphones. If this trend continues, multi-device usage in which an individual uses a full range of multiple devices including wearable devices will become a reality.

In addition, the concept of seamlessly connecting various kinds of devices such as sensors, smart meters, and automobiles on the Internet, i.e., the Internet of Things (IoT), has been proposed [1], and many kinds of IoT devices are now being developed. We can expect an explosive increase in communication devices of diverse types.

On the other hand, the market for mobile IT services such as social networking services (SNS), online shopping, electronic payment services, and content viewing has been growing in much the same way as the use of smartphones. In these services, the ID and password has been used as an essential means of identifying the user. However, the risk of leaking such information is becoming greater as these services come to handle highly confidential information such as payments, and the wide variety of services available to users means that they have to manage a large number of IDs. A more secure and straightforward ID management method is therefore needed.

Against this background, we reevaluated the concept of the post-smartphone. What we found was that the recent evolution of mobile phones has been focused on an “all in one” concept in which a large number of functions can be sequentially added as needed to a smartphone to encourage the user to use a variety of services. To realize the post-smartphone era in which various types of devices are used, we concluded that enabling the user to connect to diverse services regardless of device type (in a multi-device environment) will be of more importance.

Based on this conclusion, we reassess the functions of current smartphones. First, for example, the User Interface (UI) and the cellular-network access function, which are the core functions of the smartphone, are not always needed for the post-smartphone era. In other words, the user identification function as can be achieved by a Subscriber Identity Module (SIM)*1 is essential and all other functions can be activated within a device as needed by the user. In short, this means an authentication mini device with SIM is one solution for multi-device usage in the post-smartphone era.

Based on the above idea, NTT DOCOMO has developed a SIM-based authentication mini device called Portable SIM that physically separates the SIM function from smartphone (**Figure 1**). In this article, we present the hardware configuration of Portable SIM and the software configuration on the smartphone side and describe the basic operation of this novel device. We also introduce the new ecosystem that Portable SIM will help to create.

2. Configuration of Portable SIM

An external view and main specifications of a Portable SIM prototype are shown in **Photo 1** and **Table 1**, respectively.

To realize the Portable SIM, it was necessary to add wireless communication functions since a SIM card itself does not have them. Furthermore, to achieve a pocketsize mini device with a smaller battery, low power consumption was essential. Ease of connection with a variety of communication devices also had to be considered. With these requirements in mind, we selected Bluetooth®*2 for connecting Portable SIM with diverse communication devices, and in particular, Bluetooth v4.0 (commonly known as Bluetooth Low Energy = BLE), which enables low power consumption. We also mounted Near Field Communication (NFC)*3 functions in Portable SIM to simplify ID management and

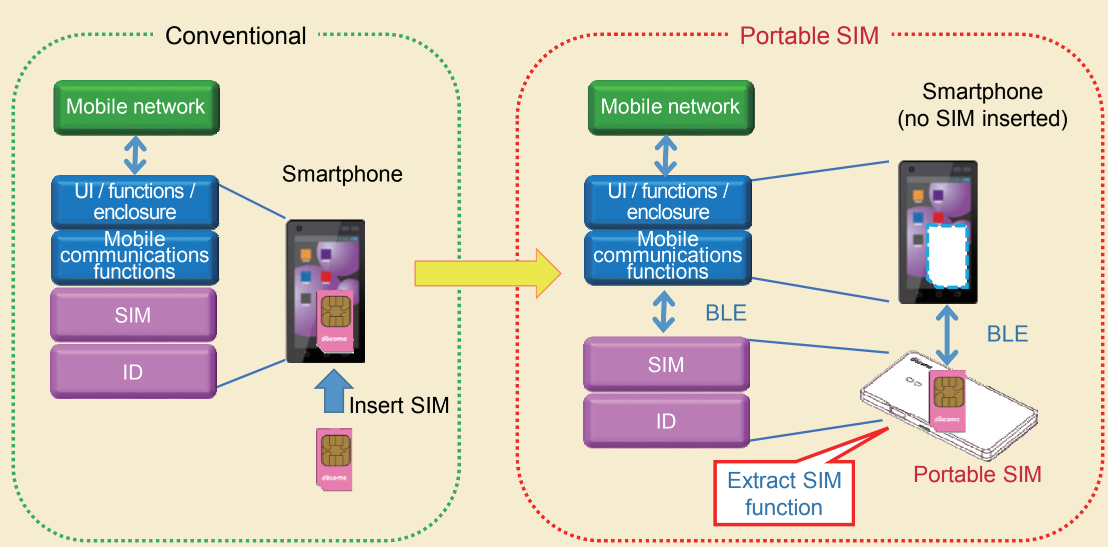


Figure 1 Concept of Portable SIM

---

*1 **SIM:** A smart card that stores subscriber information associated with a mobile phone.

*2 **Bluetooth®:** A short-range wireless communication standard for interconnecting mobile terminals such as mobile phones and notebook computers. A registered trademark of Bluetooth SIG Inc. in the United States.

*3 **NFC:** A short-range wireless communications standard using the 13.56 MHz band and initiated by NXP Semiconductors Inc. and Sony Corp. Provides unified support for FeliCa, Mifare, Type A/B (ISO14443), and IC tags (ISO/IEC 15693).

the establishment of a BLE connection (**Figure 2**).

When connecting SIM with an external device, a high-reliability protocol is essential to exchange SIM information in a secure manner. Bluetooth features SIM Access Profile (SAP) as a protocol that meets this requirement [2]. However, differences between the Bluetooth and BLE specifications prevent SAP from being directly applied to BLE, so we developed a new profile*4 to apply to BLE (SAP on BLE) following SAP policy.

SAP on BLE is essentially SAP constructed on BLE, which involved the one-to-one redefinition of Bluetooth commands as BLE commands. Moreover, as BLE itself is not capable of transmitting a response after the server has completed processing of a request received from the client, response processing was achieved by allocating the appropriate commands on BLE. In addition, while the encryption method used by SAP is Triple Data Encryption Standard (3DES), SAP on BLE achieves encrypted communications through Advanced Encryption Standard (AES)*5 defined on BLE.

The software configuration on the smartphone side to support Portable SIM is shown in **Figure 3**. On the smartphone side, it must be possible to exchange SIM information to the modem via BLE. The need for ensuring the security of this information given the presence of other applications on the smartphone and for facilitating the addition of functions for creating new services must also be considered. In light of the above, this software configuration places the SAP on BLE smartphone-side software (Application for Portable SIM) on the Java layer and connects with the modem via daemon software on the native layer (Daemon for Modem Comm.). As a result, those sections dependent on the modem and OS can be absorbed by daemon software, which means that support can be easily provided for a variety of communication devices without having to make extensive changes to SAP on BLE. Finally, SIM has been removed from the smartphone and connection from the outside has been achieved by

Photo 1 External view of Portable SIM prototype

Table 1 Main specifications of Portable SIM prototype

Dimensions (height × width × thickness: mm)	approx. 80 × 40 × 5.6
Weight	approx. 20 g
Communications method	NFC/Bluetooth (4.0)
Power method	USB charging

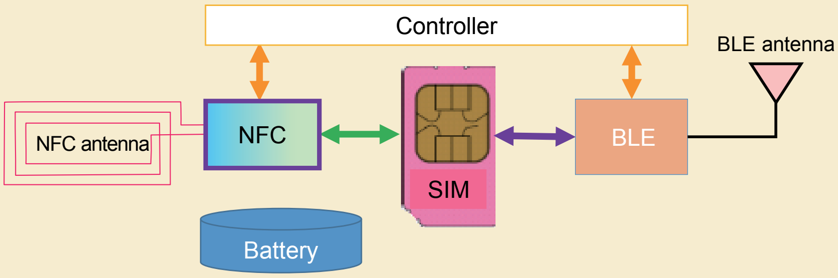


Figure 2 Hardware configuration of Portable SIM

*4 **Profile:** Inter-device protocol formulated on a service-by-service basis for use in communications by Bluetooth and Bluetooth Low Energy.

*5 **AES:** A symmetric key encryption method that has been adopted as a new encryption standard by the U.S.A. One of the cryptosystems used in 3GPP.

assigning Portable SIM the peripheral*6 role and the smartphone the central*7 role in SAP on BLE.

3. Basic Operation of Portable SIM

Basic operation of Portable SIM when using a service in conjunction with a communication device such as a smartphone is described below (**Figure 4**).

(1) Establish BLE link

The Portable SIM and the communication device exchange the information needed to establish a BLE link and then proceed to establish the link. This prototype uses NFC technology to exchange the information needed for a BLE link thereby enabling a connection to be established by a simple and intuitive "wave" operation.

(2) Connection by SAP on BLE

After the BLE link has been established, the SIM in the Portable SIM and the communication device connect using SAP on BLE. At this time, Portable SIM transmits the information for connecting a mobile network to the communication device (modem). The communication device then performs the same processing as if SIM was a built-in component.

(3) Use of SE area

The information stored in the SIM's Secure Element (SE)*8 can be exchanged through an NFC-based touch operation. Once the BLE link is established, the information is exchanged using the SAP on BLE profile.

The above basic operations of Portable SIM can be combined to enhance the user experience when using multiple devices, sharing devices, and performing ID authentication. We describe these three usage scenarios below (**Figure 5**).

- Multi-device

Connecting Portable SIM to a tablet enables mobile-phone functions to be activated with the mobile phone number of that Portable SIM. Then, after the same Portable SIM connects to the user's smartphone, the functions of the smartphone can be activated with the same number that the user used with the tablet. At this time, the tablet becomes deactivated.

In other words, the user can easily select a tablet with a large screen while at home and a smartphone that is easy to carry around while com-

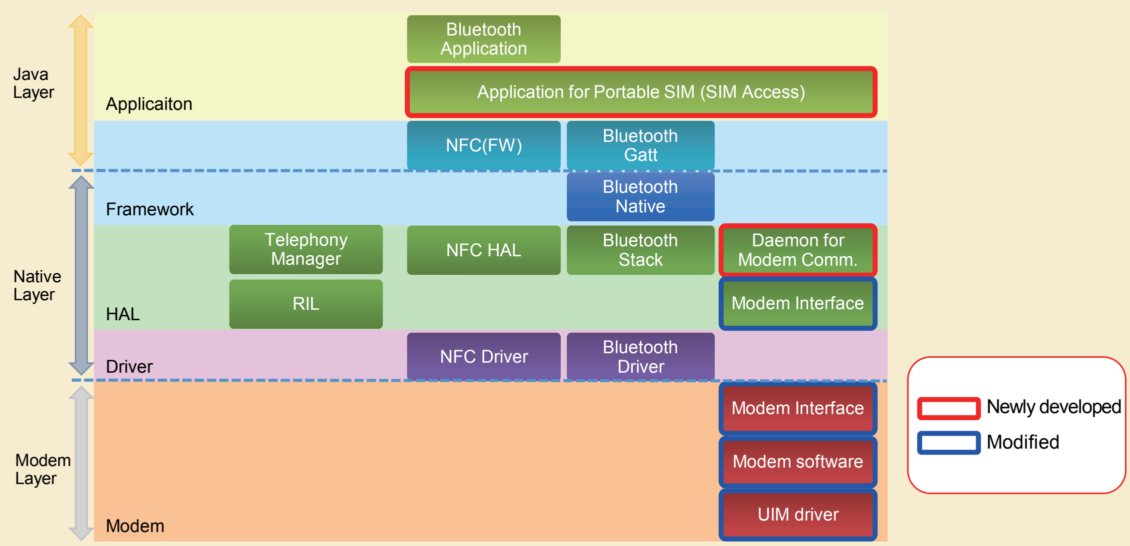


Figure 3 Software configuration of smartphone supporting Portable SIM

*6 **Peripheral:** The role of a device in Bluetooth Low Energy communication, which divides roles into "central" (see *7) and "peripheral." The central device detects and controls the peripheral device.

*7 **Central:** The role of a device in Bluetooth Low Energy communication, which divides roles into "central" and "peripheral" (see *6). The central device detects and controls the peripheral device.

*8 **SE:** An area for securely storing encrypted keys and other types of confidential information.

muting.

- Device sharing

A user can possess multiple Portable SIM devices for use with a single communication device and can easily use one or the other depending on the use case by simply switching connections.

For example, a user could possess two Portable SIM units—one for private use and one for business use—and then use the same smartphone

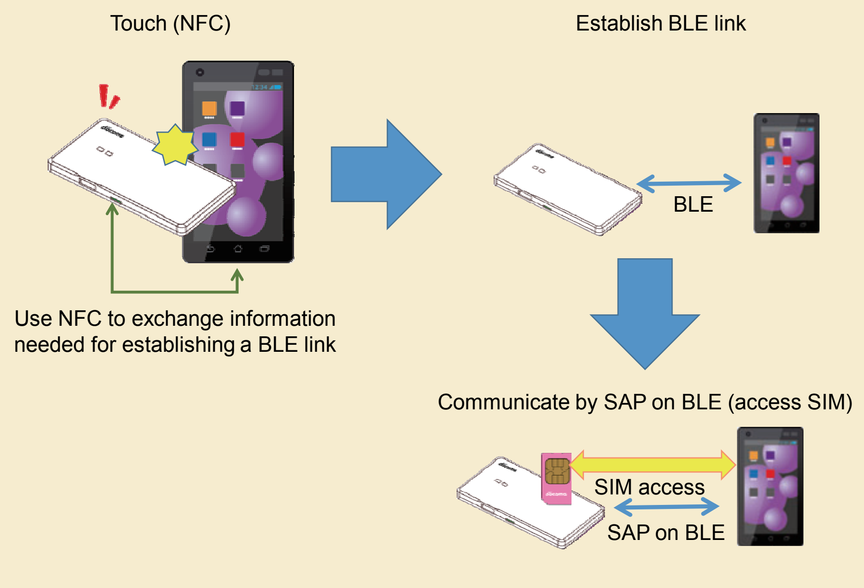


Figure 4 Portable SIM usage scheme

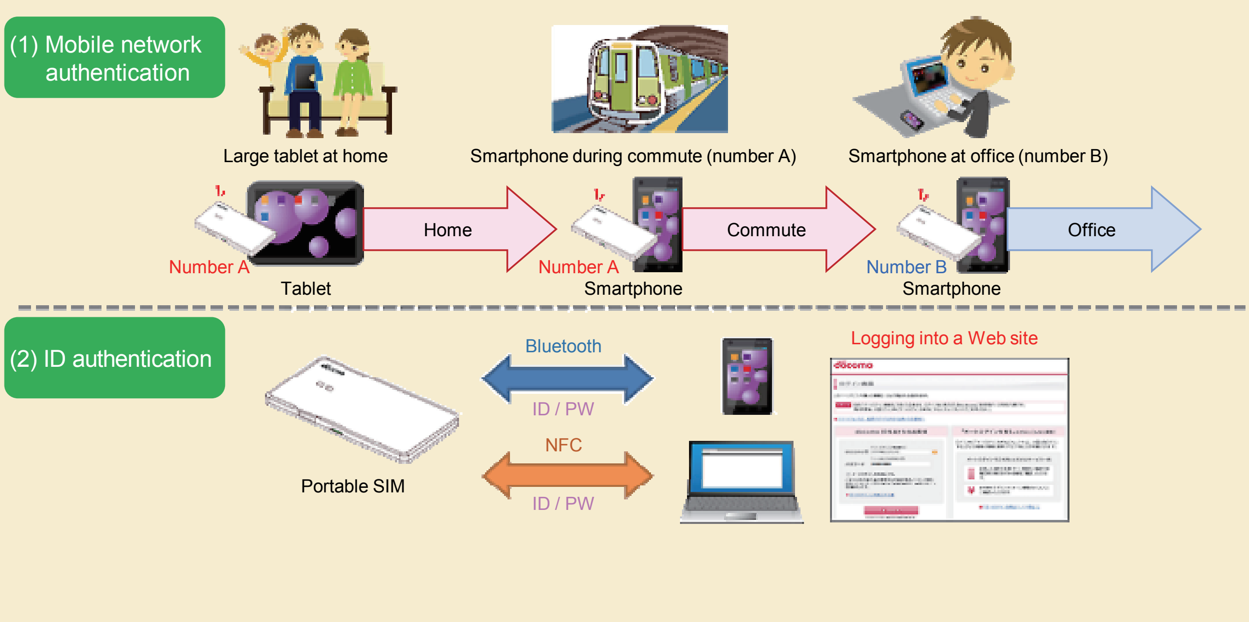


Figure 5 Portable SIM usage scenarios

for either private or business matters. The same idea could be applied to the case in which a family shares a single tablet: different members of the family could use separate Portable SIM units enabling the tablet to be used under settings tailored to each user.

In other words, if Mobile Device Management (MDM) functions were to be linked to phone numbers, it would be possible, for example, to suppress the execution of the camera function or apps when using the Portable SIM for business purposes or to activate parental controls*9 in the case of a family tablet when using the Portable SIM set for children.

- ID authentication

Storing regularly used service authentication information (Web site address, ID/password, etc.) in the SE area of an SIM card simplifies the use of the same service authentication information when going back and forth among various communication devices in a multi-device environment or when using IoT devices.

4. New Ecosystem Enabled by Portable SIM

With Portable SIM, the user can easily switch phone numbers and settings while keeping service authentication information with them. Thus, Portable SIM enables the creation of an environment in which users can carry around their phone numbers and service authentication information. The following describes a new ecosystem enabled by Portable SIM (**Figure 6**).

1) Flexible Combinations of People and devices

A direct value provided by Portable SIM is the ability to “carry around one’s (user-specific) phone number and ID.” However, looking forward to the IoT world that is predicted to grow in the years to come, there will be many occasions when the user will want to connect

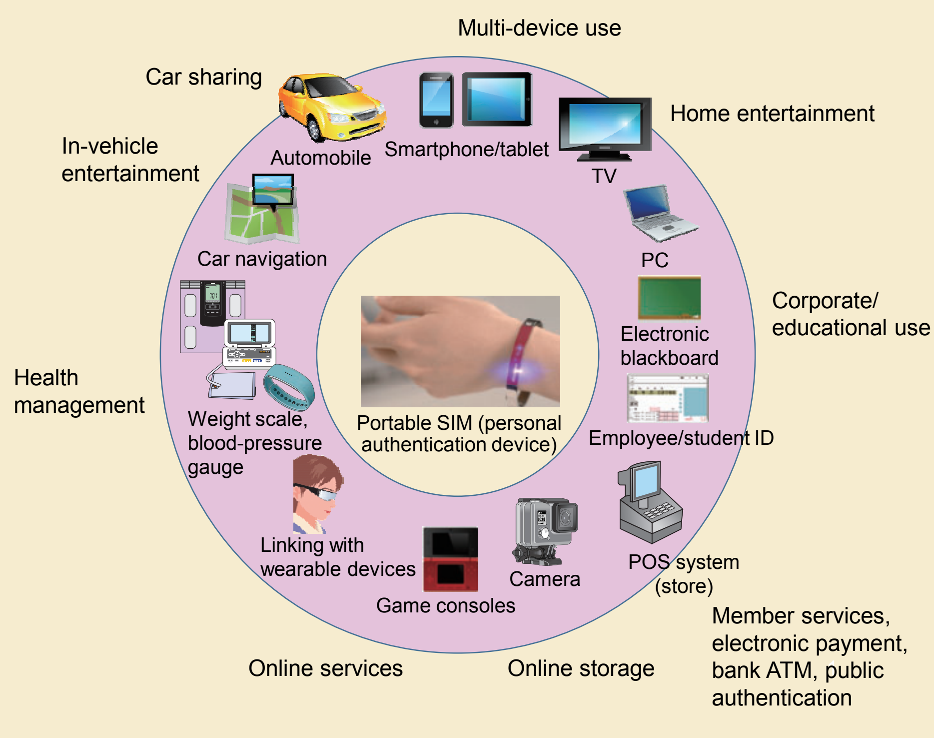


Figure 6　New ecosystem enabled by Portable SIM

*9 **Parental controls:** Functions that enable parents to control how their children use devices such as smartphones and PCs.

to a network (Internet, cellular net, etc.) through devices (such as smartphones and other devices having communication functions), and at these occasions, identifying the user to provide desired services will be extremely important. If we therefore consider what value Portable SIM can provide in an IoT world, we can say, first, from the viewpoint of devices, that it will be able to make all kinds of devices into a network entry point by separating devices and the mobile network (billing contracts) thereby eliminating the need to be concerned about the latter. At the same time, from the viewpoint of users, Portable SIM will make it easy to access desired services from network entry points that will exist at many and varied locations throughout this ecosystem. In short, as shown in **Figure 7**, Portable SIM will make it easy to identify the service user, and this, in turn, will facilitate the creation of a new ecosystem with flexible combinations of user and devices without the user having to own those devices. In contrast, a mobile network (billing contract) and user have conventionally been fixed to a device. This flexibility should lead to the creation of new value that we can call "from ownership to use."

Taking, for example, a user of health services, we can envision a service that enables the user to obtain data from nearby health-related devices through intuitive operations made possible by Portable SIM, even if using those devices for the first time. This service could also guide the user toward appropriate health services based on the data obtained. Additionally, in the case of automobiles, Portable SIM will make it possible to quickly identify the user entering the vehicle so that services tailored to the user can be provided in a seamless manner. For example, travel plans (routes, desired destinations, etc.) that have been previously prepared by the user on the

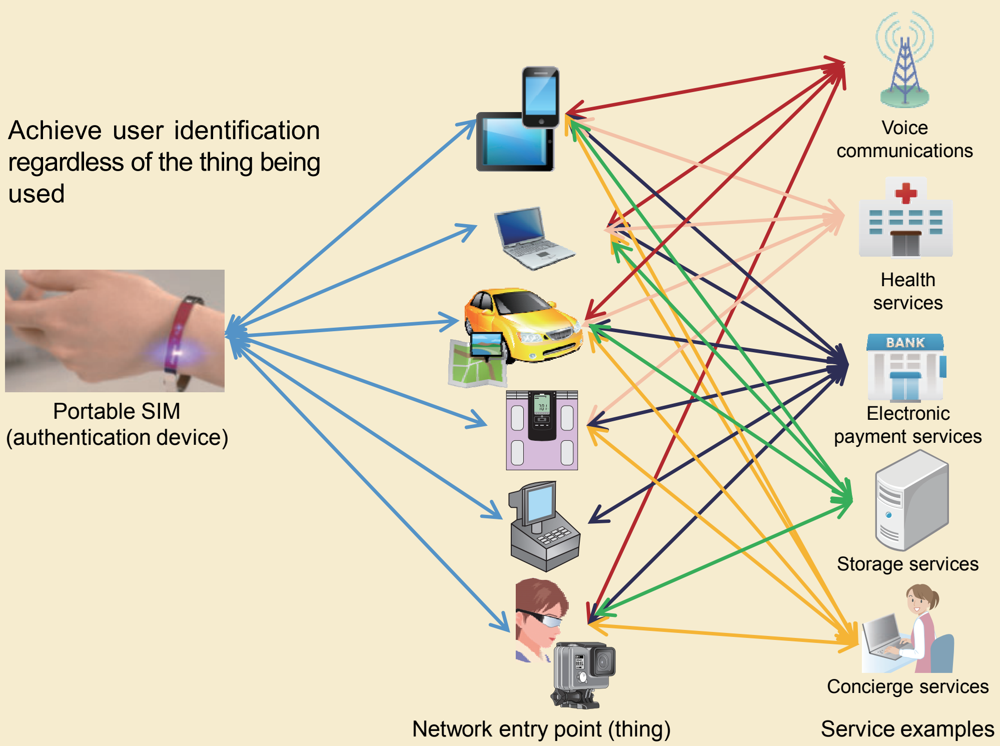


Figure 7 Flexible combinations of people and things

cloud could be set in the car navigation system and in-vehicle entertainment devices could be linked to the user's favorite music, content, etc.

2) Provision of Secure and Open Authentication functions

Up to now, SIM has been used as a component built into a smartphone. However, with the development of this new functional device called Portable SIM, the functions that SIM has conventionally been equipped with can now be used through Portable SIM alone. This is the second direct value that Portable SIM can provide, as shown in **Figure 8**.

SIM has PIN authentication and a remote lock [4] function for use when a handset is lost. Authentication functions can similarly be disabled when a Portable SIM (device) is lost likewise through remote locking.

In addition, Portable SIM has a built-in mechanism for adding an applet*10 [4] to SE from a service issuing/management system [3] called Trusted Service Manager (TSM). An applet stored in SE can be programmed based on JavaCard™*11 specifications [4], making it easy to add ID/password management functions and authentication functions for member ID cards, electronic locks, basic resident registration cards, etc. In this way, applets stored in Portable SIM can be used to perform personal authentication for connecting to a variety of services. This mechanism is advantageous for service providers since it enables them to construct a proprietary authentication platform for specific services without having to use devices such as cards that have been traditionally distributed for personal authentication. It is also advantageous for users since it consolidates authentication functions for diverse services while ensuring safety and eliminates the complexity of ID/password management. In addition, these advantages can be integrated to fortify

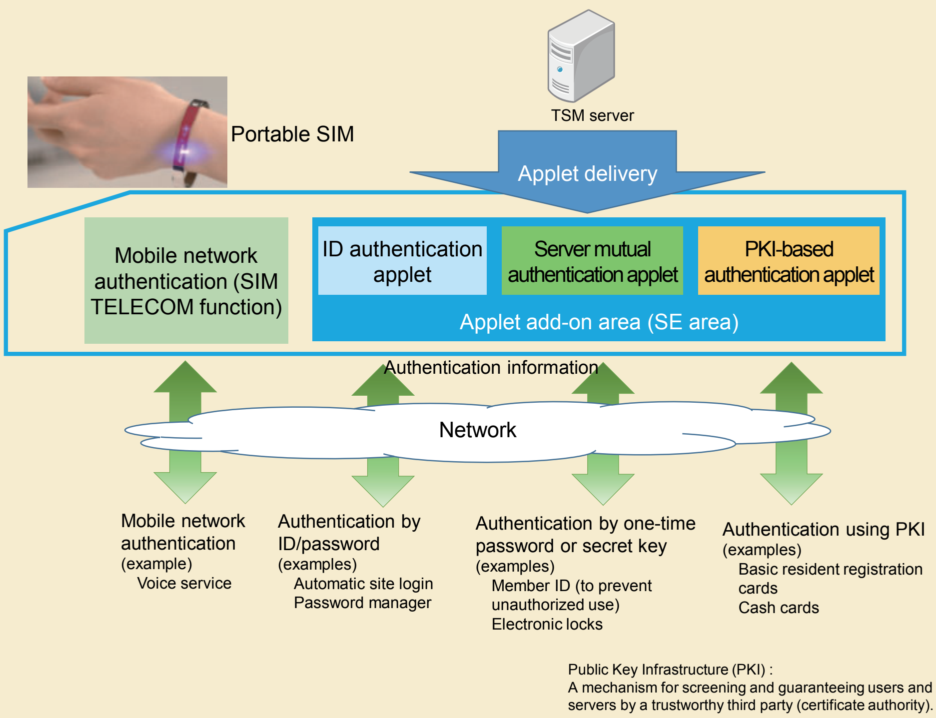


Figure 8 Authentication using Portable SIM

*10 **Applet:** A JavaCard™ (see *11) application running on SIM.

*11 **JavaCard™:** A Java execution environment operating on a device having limited memory and processing ability such as smart cards including SIM. Java is a registered trademark of Oracle Corporation and/or its subsidiaries and affiliates in the United States and other countries.

coordination between services, which assumes a secure authentication platform. We expect such secure inter-service linking to enable the creation of new services.

For example, such authentication functions could be combined with mobile network authentication to automate applet management (add, delete, etc.) so that passwords and signature information held by individual authentication functions can be automatically updated without bothering the user. Additionally, the consolidation of authentication functions should simplify the linking of content payment functions between different service providers and facilitate the mutual use of points and priviledge. Furthermore, in the health and medical care fields, Portable SIM authentication functions could be linked with personal information such as medical records that needs to be securely managed between hospitals. We expect such linking to enhance the user experience in using services.

In the above ways, Portable SIM can increase the number of network entry points, provide flexible combinations of users, devices, and services, and achieve diverse authentication functions tailored to different types of services. Integrating these capabilities in the form of Portable SIM can enhance the value of cloud services and provide new value to users.

5. Conclusion

In this article, we presented a prototype of the Portable SIM authentication mini device with the aim of creating new services. We explained the background leading up to its development, the technology needed to realize it, and its basic operation. We also described a new ecosystem enabled by Portable SIM through the creation of novel and attractive services.

Looking to the future, our plan is to miniaturize the Portable SIM device with an enhanced user experience toward commercialization and to accelerate system development toward the provision of diverse services enabled by Portable SIM.

References

[1] Kevin Ashton: "That 'Internet of Things' Thing," RFID Journal, Jun. 2009. http://www.rfidjournal.com/articles/view?4986

[2] Bluetooth SIG, Inc.: "SIM ACCESS PROFILE Interoperability Specification v1.1," Dec. 2008.

[3] T. Sugano et al.: "Advances with Osaifu-Keitai – Starting Services Supporting NFC (Type A/B) on NTT DOCOMO UIM Cards–," NTT DOCOMO Technical Journal, Vol. 15, No. 1, pp. 22-28, Jul. 2013.

[4] T. Akiyama et al.: "Technologies for Further Evolution of Osaifu-Keitai Service–NFC-enabled Mobile Terminals and NTT DOCOMO UIM Card–," NTT DOCOMO Technical Journal, Vol. 15, No. 1, pp. 13-21, Jul. 2013.

High Audio Quality Melody Call for VoLTE

Core Network Development Department
Mana Kaneko, Yu Kojo†
Corporate Strategy & Planning Department
Hiroaki Nakajima
Online Marketplace Department
Susumu Ogata

1. Introduction

In June 2014, NTT DOCOMO began the first voice over LTE (VoLTE) service in Japan, enabling users to call each other with high quality audio. Because users must have a VoLTE-capable terminal to enjoy high audio quality communications with each other, the number of users who have VoLTE-capable terminals and opportunities for users to enjoy high audio quality calling need to be increased. Thus, to promote VoLTE terminal usage, we have developed a high audio quality DOCOMO Melody Call®*1 service, which became available in October of 2014. This service enables VoLTE terminal users to experience high audio quality regardless of the type of terminal the other party is using. Melody Call is a service that plays music or voice selected by the receiver for the caller, in place of a Ring Back Tone (RBT)*2. NTT DOCOMO began providing this service in September of 2003.

This article discusses the background to the development of high audio quality Melody Call, an overview of the service, and describes how this service was achieved and the steps taken during development.

2. Development Background

For the codec*3, in place of the Adaptive Multi-Rate NarrowBand codec (AMR-NB)*4 [1], VoLTE enables high audio quality calling using the wide band Adaptive Multi-Rate WideBand codec (AMR-WB)*5 [2]. There are two objectives for raising the quality of Melody Call audio. The first is to make VoLTE more appealing, while the second is to raise the added value of the Melody Call service for future service developments.

- Appealing to VoLTE

  The high audio quality of VoLTE is only available when both the caller and receiver are using VoLTE terminals. For this reason, even if users purchase a VoLTE terminal, the opportunities to experience high audio quality calling will be limited to calling other users who are also using VoLTE terminals. To counter this situation and increase the opportunities for users to experience high audio quality even if the person they are calling does not have a VoLTE terminal, we developed the high audio quality Melody Call as a mechanism for raising awareness of the benefits of the high audio quality of VoLTE.

† Currently Solution Business Department

*1 **Melody Call®:** An NTT DOCOMO service that lets users replace the mobile phone ring back tone with their favorite music. It is a registered trademark of NTT DOCOMO.

*2 **RBT:** A ring back tone is played via the network to the caller while the receiver is being called (ring ring, ring ring etc).

- Increasing the added value of Melody Call

  Melody Call [3] has been available since the feature phone era, and is a service currently used by many users. By developing the Melody Call service for high audio quality available with VoLTE, this development aims to raise the appeal of VoLTE with improved audio quality (music audio quality), and thus further promote usage of the service among existing users while acquiring new users.

3. Service Overview

High audio quality Melody Call is available under the following conditions:

- Usage terminals and usage areas

  To enjoy high audio quality Melody Call, callers must use a VoLTE terminal and be within an LTE coverage area. Call receivers must be subscribed to Melody Call, and have high audio quality music set for it (**Table 1**, **Figure 1**).
- Rates

  High audio quality does not incur any changes to rates - existing users can use high audio quality Melody Call at their current rates.

---

*3 **Codec:** Technology for coding and decoding data such as audio signals.
*4 **AMR-NB:** A 3GPP standard conventional audio encoding method.
*5 **AMR-WB:** A 3GPP standard wideband audio encoding method.

Table 1 Conditions for providing high audio quality Melody Call

Caller	Receiver	Melody Call subscription	Music settings	Ring back tone	(Reference) Call
VoLTE terminal user (in LTE area)	VoLTE terminal user (in LTE area)	Yes	High audio quality music	High audio quality Melody Call	High audio quality calling
			Existing music settings	Conventional audio quality Melody Call	
		None	–	RBT (ring ring etc)	
	VoLTE terminal user (in 3G area), or non-VoLTE terminal user (in 3G/LTE area)	Yes	High audio quality music	High audio quality Melody Call	Conventional audio quality calling
			Existing music settings	Conventional audio quality Melody Call	
		None	–	RBT (ring ring etc)	

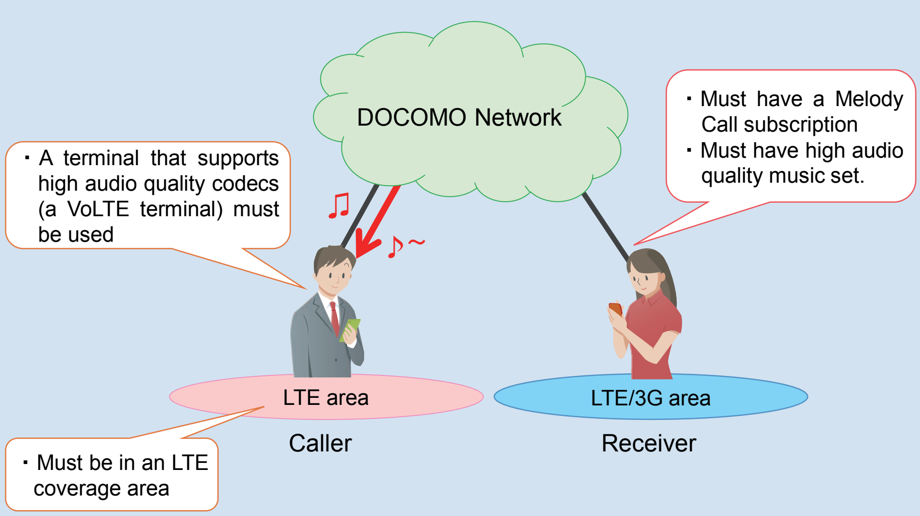


Figure 1 Providing high audio quality Melody Call

- Music settings

  Future plans include high audio quality music to be offered by Content Providers*6 for Melody Call as standard, and a gradual switch to automatic provision of high audio quality sources of the music currently set by users.

- Trial listening

  Trial listening to Melody Call sounds will also be available in high audio quality. This will enable VoLTE terminal users to listen to samples of high audio quality music on Content Provider sites or the Melody Call setting site.

- Trial dialing

  VoLTE terminal users in LTE areas will be able to try high audio quality Melody Call by dialing the 157005 VoLTE trial number, or high audio quality calling by dialing 157001 (As of January 2015).

4. Implementing Method

4.1 Basic Call Processing (High Audio Quality Codec Selection Logic)

The following describes the network architecture and process flow for high audio quality Melody Call (**Figure 2**).

(1) The network receives a call request (SIP_INVITE) from the caller, and then delivers the call request (SIP_INVITE) to the receiving terminal.

(2) The terminal that received the call request pages the receiver, and sends a paging complete response (SIP_180 Ringing) to the network.

(3) The Application Servers Node (ASN) that receives the paging complete response (SIP_180 Ringing) judges whether the receiver is subscribed to Melody Call from receiver subscription data, and then sends a Melody Call output request (SIP_INVITE) to the Media Processing Node (MPN)*7 that is operating as the Melody Call Application Server (AS)*8.

(4) The MPN that receives the Melody Call output request (SIP_INVITE) judges whether the calling terminal is requesting a high audio quality call from the SDP*9 data[5] included in the signal.

(5) The MPN acquires the sound source set for the receiver from storage.

(6) The MPN encodes the acquired sound source with

*6 **Content Provider:** A business that offers Melody Call sound sources.

*7 **MPN:** A node of the NTT DOCOMO core network. It currently provides various media services including voice answering, Melody Call and other voice media services; video media services such as videophone answering; and SMS.

*8 **AS:** A server that executes an application to provide a service.

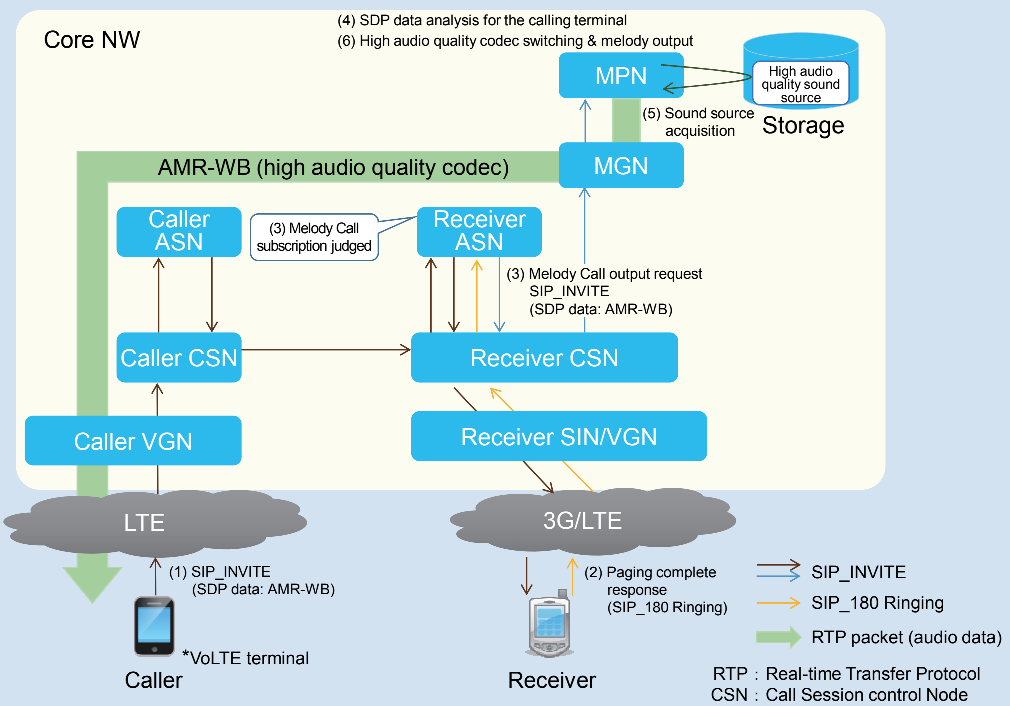


Figure 2 NW architecture and process flow

a high audio quality codec (AMR-WB), and outputs it to the calling terminal so that the high audio quality Melody Call is replayed on the calling terminal. Acquiring sound source from storage and switching to high audio quality codec is performed upon reception. Also, if the calling terminal does not support AMR-WB, codec conversion is performed to match the calling terminal.

This processing is basically the same for currently offered Melody Call[3]. The main difference with existing processing is step (4), judging SDP data in the caller signal, and step (6) outputting Melody Call with the high audio quality codec.

4.2 LTE to 3G Handover Actions

When moving into a 3G area during VoLTE voice calling, calls are continued using Single Radio Voice Call Continuity (SRVCC) technology. NTT DOCOMO has been providing SRVCC since implementing VoLTE [8] so that the caller can continue to hear the sound source even when moving into a 3G area while listening to the RBT or Melody Call. However, with handover to 3G through SRVCC, the codec for the music being played must be switched from the high audio quality Melody Call codec to the conventional audio quality AMR-NB codec.

The following describes SRVCC processing during Melody Call output (**Figure 3**).

(1) When SRVCC is executed, the VoLTE Gateway Node (VGN) sends notification (SIP_UPDATE) to switch from LTE to 3G to the Call session Control Node (CSN). CSN transfers SIP_UPDATE to ASN.
(2) The ASN that receives SIP_UPDATE sends notification (SIP_ReINVITE) to switch from LTE to 3G to the Media Gateway Node (MGN).
(3) The MGN that received SIP_ReINVITE analyses ReINVITE, and converts the AMR-WB received from MPN to AMR-NB and sends it.

4.3 Adoption of High Audio Quality Sound Sources with an Eye to Next-generation Audio Codecs

Sound sources currently in storage are compressed with a quality lower than AMR-WB, therefore, sound sources must at least be upgraded to AMR-WB or equivalent to enable callers to hear high audio quality Melody Call. Sound source upgrade is also described in Chapter 3, and there are plans for content providers to gradually replace sound sources with high audio quality sound sources in future (**Figure 4**).

*9 **SDP:** A protocol to describe information such as IP addresses necessary for initiating sessions in the IMS. It is also used to describe session information relating to SIP, a call control protocol.

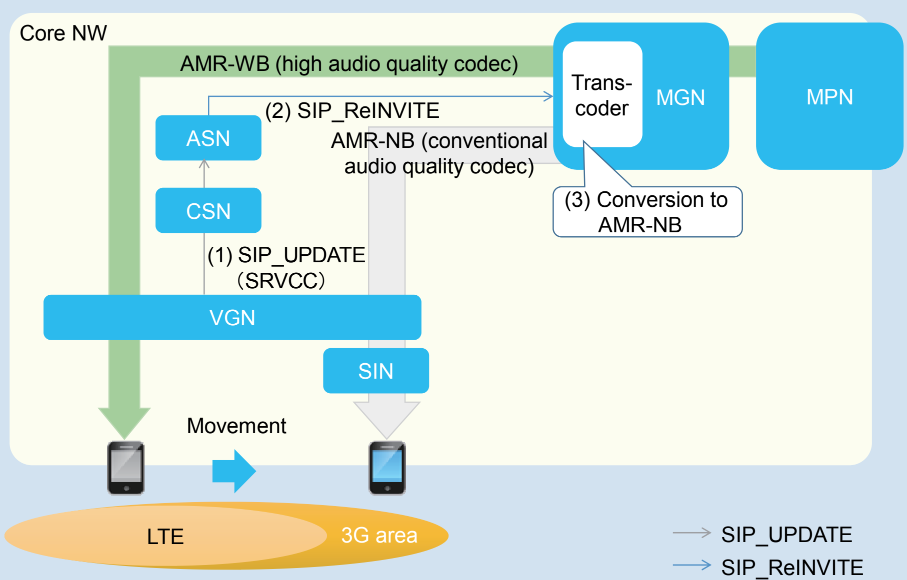


Figure 3 SRVCC processing during Melody Call output

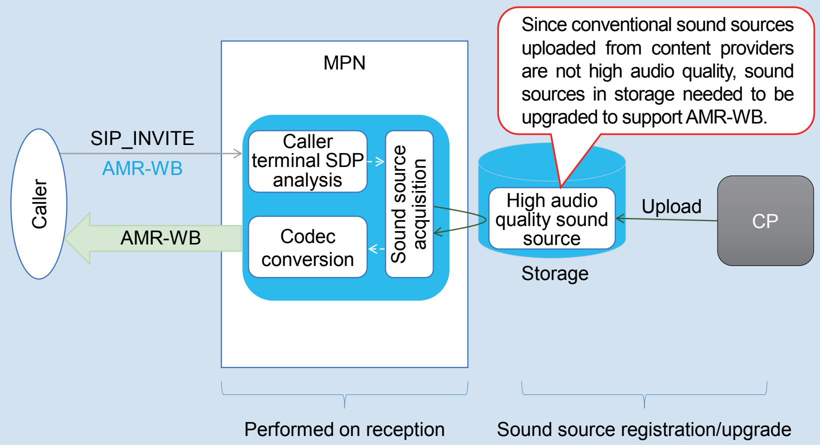


Figure 4　Sound source registration

Currently, Enhanced Voice Services (EVS)*10 [6] [7] is gaining attention as the next-generation audio codec, thus, with an eye to further increases in audio codec performance and in consideration of bit rate and audio range, high audio quality sound sources compatible with EVS have been adopted. These high audio quality sound sources will be compatible with next-generation Melody Call codecs and will not require upgrade (**Table 2**).

5. Conclusion

This article has described the high audio quality Melody Call development accompanying the implementation of VoLTE, and an overview of the service and methods of delivery. Into the future, we aim to continue to make DOCOMO services even more convenient.

Table 2　Sound source quality and codec performance

Audio codec	Conventional audio quality Melody Call sound source	High audio quality Melody Call sound source
AMR-NB (Bandwidth: 300 Hz to 3.4 kHz/bit rate: 4.75 to 12.2 kbps)	Y	Y (Conventional audio quality)
AMR-WB (Bandwidth: 50 Hz to 7 kHz/bit rate: 6.6 to 23.85 kbps)	–	Y
EVS (Bandwidth: 20 Hz to 16 kHz/bit rate: 5.9 to 128 kbps)	–	Y

References

[1] 3GPP TS26.071: "Mandatory speech CODEC speech processing functions; AMR speech Codec; General description."

[2] 3GPP TS26.190: "Speech Codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; Transcoding functions."

[3] H. Hayasaka et. al: "Development of Melody Call Service," NTT DOCOMO Technical Journal, Vol. 12, No. 1, pp. 64-69, Apr. 2004 (Japanese only).

[4] Y. Shimada et. al: "IP-based FOMA Voice Network toward Enhanced Services and Improved Efficiencies," NTT DOCOMO Technical Journal, Vol. 12, No. 1, pp. 4-14, Jun. 2010.

[5] 3GPP TS24.229: "IP multimedia call control protocol based on Session Initiation Protocol (SIP) and Session Description Protocol (SDP); stage3."

[6] 3GPP TS26.114: "IP Multimedia Subsystem (IMS); Multimedia telephony; Media handling and interaction."

[7] K. Tsutsumi et. al: "3GPP EVS Codec for Unrivaled Speech Quality and Future Audio Communication over VoLTE," NTT DOCOMO Technical Journal, Vol. 16, No. 4, pp. 6-13, Apr. 2015.

[8] K. Tokunaga et. al: "VoLTE for Enhancing Voice Services," NTT DOCOMO Technical Journal, Vol. 16, No. 2, pp. 4-22, Oct. 2014.

*10 **EVS:** A 3GPP standardized next-generation audio encoding method.

ETSI NFV Excellence Award

Tetsuya Nakamura[†] of NTT DOCOMO Research Laboratories, who has led NFV activities, received the ETSI NFV Excellence Award at the 8th plenary session of the Network Functions Virtualization Industry Specification Group (NFV-ISG) of the European Telecommunications Standards Institute (ETSI) held in Arizona in the United States from November 17-21, 2014.

NFV-ISG was established within ETSI in October of 2012 to study the virtualization of network functions, and has been examining technologies to achieve telecommunication carrier network virtualization. In that period, 238 companies including 37 telecommunication carriers and providers have become members (as of November 2014). There are more than 1,200 names on the organization's mailing list of people central to NFV studies. The prize was awarded two years after the establishment of NFV-ISG, when concept examinations of NFV and the initial study on architecture was completed, and was awarded to four persons including Mr. Nakamura as well as Joan Triay of DOCOMO Communications Laboratories Europe GmbH for their excellent contributions to the NFV studies.

Modern communications services are achieved through individual communications functions in specialized hardware in dedicated equipment. For this reason, when commencing new services or expanding the size of services to respond to demand increases, there are large costs associated with developing these specialized devices, procurement, construction works and installation associated with implementation, and the maintenance systems that must be set up for these systems. Moreover, in the modern world where new services are unceasingly being brought into existence, shortening the time it takes to implement services is a major issue facing telecommunication carriers. To solve these issues, NFV enables virtualization technologies widely used in IT service data centers to be used in telecommunications services, and enables telecommunications functions through software running on general purpose equipment.

NFV-ISG does not concern itself with creating standard specifications, but seeks to reach consensus within the industry about architecture for network functions virtualization, its requirements and how it can be achieved. The results of the organization's activities are shared with related standards institutions, forums and through open source activities, while the organization aims to create specifications by appealing to relevant associations if standardization is required.

Including Mr. Nakamura's involvement, NTT DOCOMO has made a range of technical proposals for NFV including architecture, interface and service requirements, has taken the lead in discussions, and has provided guidance for the dissemination of NFV concepts as an overall leader of NFV-ISG activities. As a result, many related associations such as 3GPP and IETF have begun examining applications of NFV, and have established open source projects.

† Currently NTT DATA Corp.

NTT DOCOMO
Technical Journal Vol.16 No.4

Editorship and Publication
NTT DOCOMO Technical Journal is a quarterly journal edited by NTT DOCOMO, INC. and published by The Telecommunications Association.

Editorial Correspondence
NTT DOCOMO Technical Journal Editorial Office
R&D Strategy Department
NTT DOCOMO, INC.
Sanno Park Tower
2-11-1, Nagata-cho, Chiyoda-ku, Tokyo 100-6150, Japan
e-mail: dtj@nttdocomo.com