# Voice Band Widening Technology - Voice Band Enhancer -

Hiromi Aoyagi Atsushi Tashiro

Voice communications (so called telephony) are about to undergo a dramatic change. The infrastructure of voice communications are about to be integrated into packet communication networks, with IP networks represented by such technologies as the Next Generation Network (NGN) and Long Term Evolution (LTE). It is anticipated that this will bring forth a wide variety of new telephone services in the future and the most basic of all, voice transmissions are expected to undergo significant changes. The frequency range for voice calls that could be transmitted with the infrastructure of telephone services in the past had been restricted to 300 Hz to 3.4 kHz, due to system constraints. The frequency range audible to the human ear is usually considered to be between 20 Hz and 20 kHz, which means only a portion of the range had previously been transmitted. These constraints are lifted when the transmission network is changed, facilitating the prospect for voice to be transmitted with a quality that more closely resembles natural voice. The evolution of voice services had thus far provided a narrowband sound range that reached a mere 4 kHz, however, the popularization of a voice service with a wideband voice range that exceeds 4 kHz is expected to accelerate in the future. This paper describes the issues relating to the popularization of such a wideband voice service and the activities undertaken by OKI in response to such issues.

## Wideband voice service

In order to transmit voice with a wideband range, a voice coding function for compressing and decompressing voice with a wideband range must be embedded. In the recent years standardization organizations have been proceeding with the standardization of wideband voice coding, for example ITU-T established G.729.1[1)] and G.711.1.[2)] A wideband voice service can be realized via new transmission networks through the embedding of such a wideband voice coding function in both the sending and receiving terminals.

Such new telephone services cannot, however, be replaced from previous services overnight. During the transition period for such a replacement, a changeover period is required during which the legacy narrowband telephone service networks and the new wideband telephone service networks coexist. During the initial stage of such a period, furthermore, a larger proportion of networks will be comprised of the legacy networks, with the wideband networks gradually and eventually taking over. The likelihood that the wideband terminals on the wideband service networks will not be able to provide adequate effectiveness during such a transition period is a concern, however. This is due to the fact that in order to realize wideband voice communications it is necessary for the wideband voice coding function to be embedded at both the sending and receiving terminals. In other words, any calls made with a narrowband terminal would only provide narrowband voice calls as before, since the party on the narrowband terminal would not be equipped with the wideband voice coding function. During the initial stage of the wideband telephone service, the vast majority of parties making and receiving calls will be expected to use the legacy narrowband terminals, which would halve the value of using wideband terminals (**Fig. 1**).

In order to respond to this issue, OKI is proceeding with the practical implementation of a band widening function for simulating the higher band lost during calls based on narrowband voice signals. Even when a call has a narrowband range, the use of this function enables a receiver using wideband to receive the wideband voice signal, which is obtained internally by the receiving terminal with a wideband capability (**Fig. 2**). Descriptions on the outline of this band widening function and the activities undertaken for its practical implementation, as

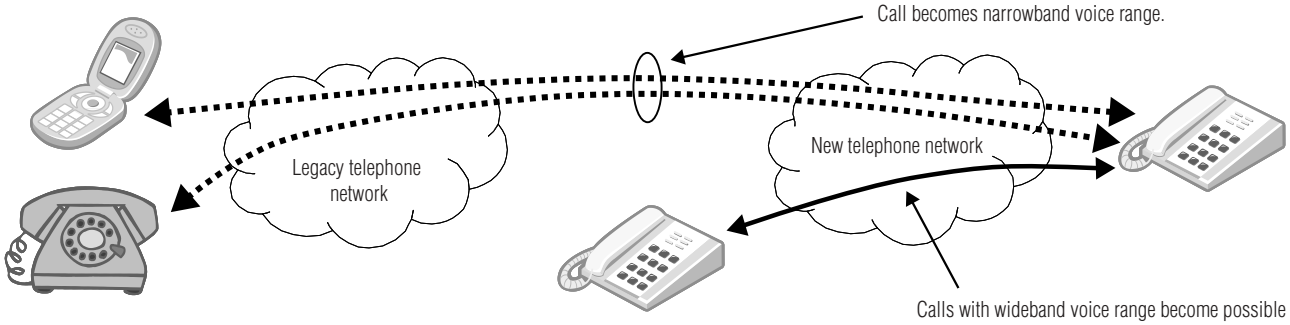


**Fig. 1 Voice communication during service transition period**

well as the evaluation results, are provided herein.

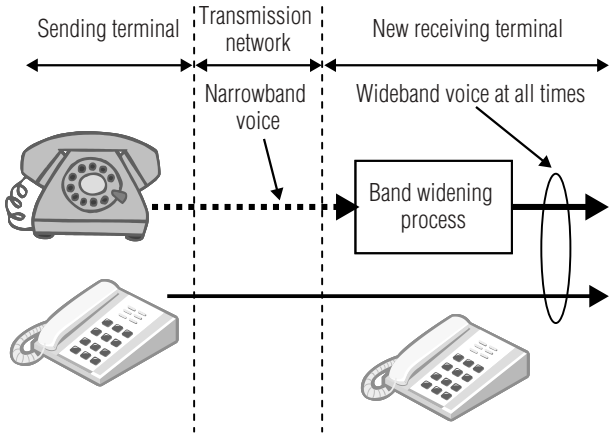


**Fig. 2 Band widening function**

## Band widening function (Voice Band Enhancer)

### (1) Widening method

Two primary methods can be used for the band widening process. The first method entails the processing and shaping of an existing lower band signal (up to 4 kHz) before transferring it to a higher band (4 kHz to 8 kHz). The other method involves analyzing a lower band signal, generating a residual signal and re-synthesizing these signals with consideration for the higher band. In some cases, especially when the latter method is employed, this is used in conjunction with a method that investigates (learns) and correlates the rough spectral shapes of lower band and higher band signal beforehand, in order for the information to be used in a codebook (quantization table). That is, this process obtains at first the rough spectral shape information of the lower band signal by analyzing the lower band signal and, based on the information thus obtained, the correlative rough spectral shape information of the higher band signal is then taken from the codebook for further synthesizing both of lower and higher band signals.

OKI investigated such methods on a variety of aspects. The results indicated that even though reproducibility of the rough spectral shape is somewhat inferior with processing and shaping conducted by the former method, it was found that the noise level of the higher band signals could be suppressed to a relatively low level. On the other hand, even though the noise level of the higher band signals increased somewhat in the analysis and synthesis of the latter method, it was found that the reproducibility of the rough spectral shape was considered relatively favorable. The method involving the use of a codebook, furthermore, proved to have even more favorable reproducibility of the rough spectral shape, however, some ingenuity was required for the design (how to conduct the learning and the size of the codebook) for actual implementation. The dependency (linguistic and transmission network characteristics, etc.) on the input signals or the amount of memory used, for example, must be considered.

OKI realized a method based on the processing and shaping type, with the implementation of further analysis

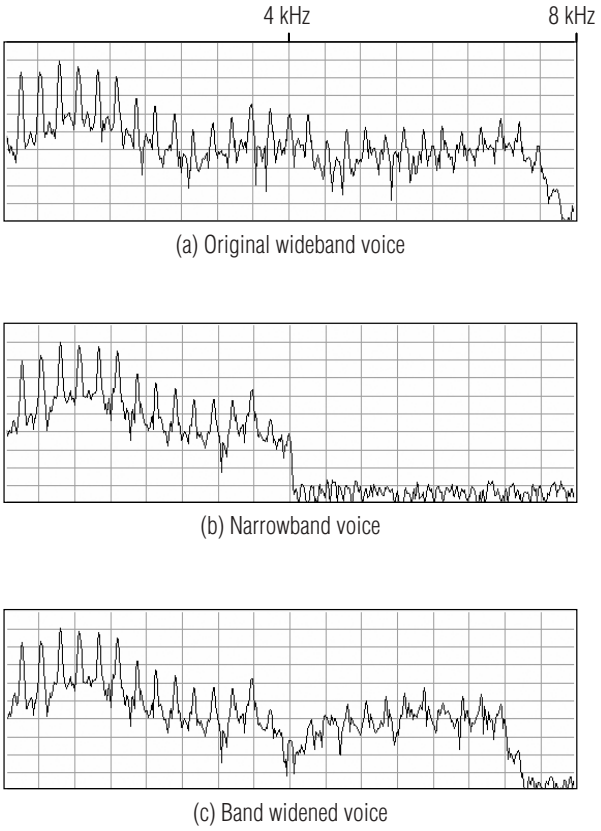


**Fig. 3 Frequency characteristics of VBE processed sound**

before adaptively processing and shaping the obtained information. We named this method for band widening the Voice Band Enhancer (VBE). An example of frequency characteristics for the VBE processed voice is shown in **Fig. 3**.

### (2) Input signal characteristics

The basic widening function is important, in addition two major aspects were considered for practical implementation of VBE. One was the characteristic of the input signal to VBE and the other was the characteristics of the output signal from VBE.

The signal input to VBE is the signal sent by a remote terminal, through a transmission network. The voice coding method used and the characteristics of the transmission network vary depending on the communication system (fixed line network, mobile network, communication carrier) and the extent of background noise also varies depending on the usage condition of the distant terminal. As a result, the input signal to VBE has a wide variety of characteristics (**Fig. 4**). The basic processing of the band widening involves the generation of a frequency portion in the lost higher band (4 kHz to 8 kHz), based on the existing lower band (up to 4 kHz) signal. The characteristics of widening, therefore, have the essential feature of being affected in no small way by the characteristics of existing signal portions. The preprocess, located in the stage prior to the widening process, is a vital in order to take advantage of the widening performance in a stable manner. The preprocess therefore can be considered to have a significant impact on the widening characteristics.

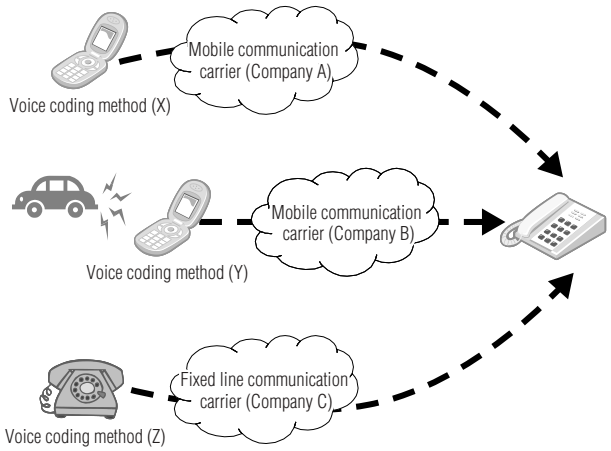


**Fig. 4 Diversity of input signals**

Several communication patterns were measured and anticipated at OKI, in the accumulation of know how for the compensation methods used for the preprocess.

### (3) Output signal characteristics

The output signal from VBE, nevertheless, ends up reaching the human ear via an electroacoustic transducer (e.g. loudspeaker). The uniqueness of the characteristics emitted by the loudspeakers actually vary, including the characteristics of the analog circuit prior to a loudspeaker (**Fig. 5**).

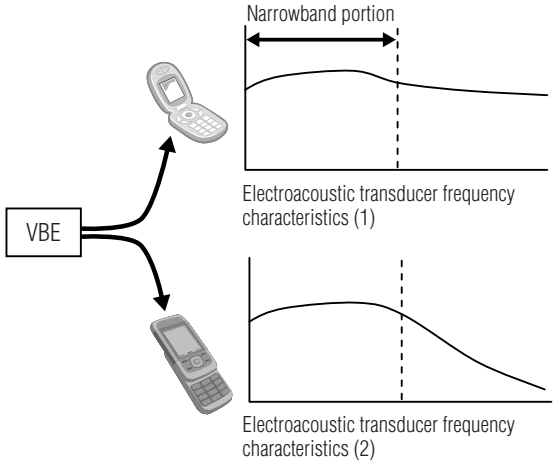


**Fig. 5 Diversity of electroacoustic transducer characteristics**

To ensure that sound reaches the human ear with a stable quality, post processing that compensates for the quality is also a vital responsibility. Loudspeakers (or receivers) built into existing telephone terminals are ordinarily directed to outputting narrowband signals (up to 4 kHz) and thus the output characteristics are not guaranteed for wideband signals (from 4 kHz up). In other words, some are capable of outputting sound in a relatively tidy manner, whereas the sound is greatly attenuated in others. Loudspeakers designed for wideband use provide relatively stable characteristics, but they are still not common and the costs are high also. OKI has been accumulating know how on the compensation methods used in the post process in order to obtain stable widening characteristics, even when ordinary narrowband receivers are used.

Incidentally, companies that manufacture voice communication terminals always employ persons who are referred to as "Sound Masters" or "Sound Gurus". These people make final sound adjustments and the decisions for the products. Accordingly, an important aspect for practical implementation is to prepare and contrive an interface for these people in order to make their sound making efforts easier.

## Subjective evaluations

The evaluation results for the performance of the band widening process, which is the foundation of VBE, are introduced in this section. A third party organization was called upon to evaluate the opinions according to the ITU-T recommendation P.800.1[3] in order to elicit an evaluation on the voice quality of the band widened voice, as perceived by humans. The evaluation voice preparation system and the test system are shown in **Fig. 6**, whereas the test conditions are listed on **Table 1**.

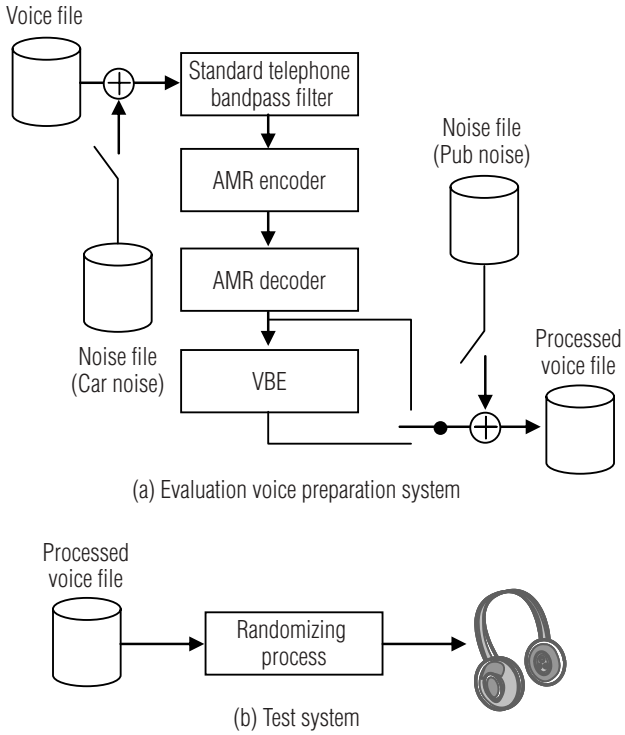


**Fig. 6 Evaluation voice preparation system and test system**

Evaluation voice preparation system operates on the computer and sequentially processes the prepared voice files (two sentences per file, no overlay of noise) with a standard telephone bandpass filter, AMR[4] encoder, AMR decoder, as well as the VBE process (ON/OFF) and records the results in a file. The noise overlay on the sender side was implemented through the addition of a prepared noise file prior to the standard telephone bandwidth filter process. Car noise (noise recorded in a car while driving) with a signal to noise ratio of 15 dB, was used as the noise for the overlay on the sender side.

**Table 1 Test conditions**

| | |
|---|---|
| Speech samples | English, five sentences by pair of male and female speakers |
| Noise conditions | (1) Clean voice (no overlay noise)<br>(2) Noise overlay on receiver side (Pub noise, S/N = Approx. 20 dB)<br>(3) Noise overlay on sender side (Car noise, S/N = 15 dB) |
| Voice processing types | AMR, AMR+VBE |
| Test subjects | Ten persons (ordinary native English speakers) |

The noise overlay on the receiver side was implemented through the addition of a prepared noise file following the VBE process. Pub noise (noise recorded in a dining and drinking establishment) was used as noise for the overlay on the receiver side. Since noise on the receiver side was something that could be heard directly with the ear, the frequency component included noise up to about 20 kHz and in a stereo format (binaural). Because the intended sound was heard by one ear but noise was heard with both ears, a stipulation of the signal to noise ratio was difficult and thus the signal to noise ratio was set for 20 dB on one ear (intended sound + noise).

The test system randomly reproduced recorded processed voice and test subjects listened to the sound using headphones (on both ears). Except for subjects on the receiver side of the noise overlay condition, the sound was reproduced in one ear only, with the test subject making a selection as to which ear the sound was heard. The test subjects evaluated each file for voice quality by rating them in five levels (five points sequentially to one point, with a high to low evaluation). The final evaluation result was obtained by calculating the mean value (MOS value) of the evaluation score given by all test subjects. The results are shown in **Fig. 7**.

As depicted by **Fig. 7**, significant results were obtained for the clean voice, with the MOS value rated higher by 0.5 point. When the background noise was present on the side of the receiver and sender, improved effects brought about by VBE were also observed. The use of VBE therefore can be considered as a means to obtain band-widened voice with improved voice quality in comparison with conventional narrowband voice.

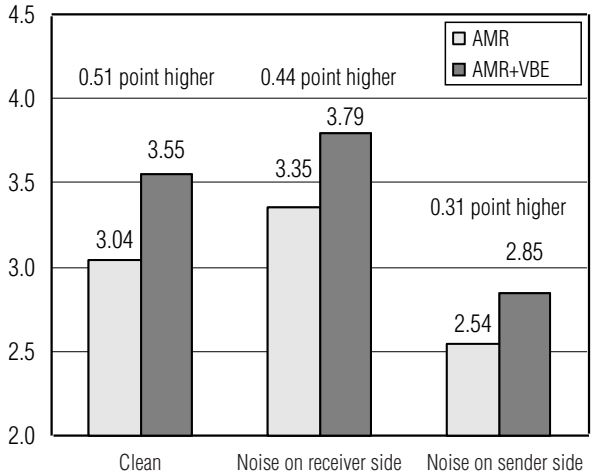


**Fig. 7 Evaluation results**

## Conclusion

This paper introduced issues surrounding the popularization of wideband voice services and activities undertaken by OKI to resolve the related issues. At OKI we intend to develop the practicality of voice quality by continuing to improve the methods used in the future. At this point in time the voice quality available is a step behind the original wideband voice. The issues that need to be resolved, however, are already clear and at OKI we are zealously proceeding with our deliberations to resolve these concerns. We are hoping to obtain band-widened voice quality levels that will not be inferior to the original voice in the near future. We are hopeful that VBE from OKI will play a part in popularizing and accelerating the wideband voice service.

## References

1) ITU-T Recommendation G.729.1 (2006): G.729 based Embedded Variable bit-rate coder: An 8-32 kbit/s scalable wideband coder bitstream interoperable with G.729.
2) ITU-T Recommendation G0.711.1 (2008): Wideband embedded extension for G.711 pulse code modulation.
3) ITU-T Recommendation P.800.1 (2006): Mean Opinion Score (MOS) terminology.
4) 3GPP TS 26.071 V6.0.0 (2004): Mandatory speech CODEC speech processing functions; AMR speech CODEC; General description (Release 6).

## Authors

Hiromi Aoyagi: Telecom Systems Div., eSound Business Dept.

Atsushi Tashiro: Telecom Systems Div., eSound Business Dept.