1. Introduction

Recently, the dissemination of remote conferences in enterprises, VoIP calls on PC, video phone on cellular phones and the prohibition of the use of cellular phone during driving have increased the demand for a handsfree capability without using a headset or handset. The handsfree devices for this purpose currently use a linear echo canceller based on a linear adaptive filter in order to improve the communication quality. However, this technology still poses a problem, which is the incapability of canceling the nonlinear echo that is produced due to the nonlinear characteristics of a compact loudspeaker and the vibrations propagated through the cabinet from the loudspeaker to the microphone.

Techniques to cancel the nonlinear echo using a nonlinear adaptive filter or neural network have been studied, but these devices do not offer a realistic solution because they require a huge amount of computations, more than 10 times that of the linear echo canceller. Therefore, there is a strong need to cancel the nonlinear echo with small computations.

The nonlinear echo canceller proposed in this paper has overcome the above problem and enables the cancellation of nonlinear echo with a small amount of computation. The proposed method allows compact equipment such as a notebook PC or cellular phone to easily perform a quality two-way communication at a similar level to the high-quality TV conference system.3,4)

2. Principles of Nonlinear Echo Generation

Let us assume that you are in Tokyo as shown in Fig. 1(a) and that you are holding a handsfree talk with Osaka, and see how the voice is heard. First, voice “Hello” spoken at Osaka is sent through the communication line and reproduced as “Hel-lo” to the loudspeaker in Tokyo. Since the loudspeaker reproduced volume is high in handsfree talk, “Hello” output from the loudspeaker is fed back to the microphone and returned to Osaka with a delay. As a result, the person at Osaka hears his own voice like an echo. The echo not only makes the pronunciation of voices difficult but may also in the worst case
produce a howling tone, in which case a satisfactory talk would no longer be possible.

The problem of echo is old and has usually been solved with the echo canceller technology. As shown in Fig. 1(b), the echo canceller regards the feedback path from the loudspeaker to the microphone as a filter. It generates a replica signal that resembles the signal fed back to the microphone, using a technology called the linear adaptive filter and uses it to cancel the signal from the microphone. The linear adaptive filter generates the echo replica by modeling the feedback path by applying a linear computation (the delay, addition and multiplication with a constant) in order to reproduce the signal sent to the loudspeaker. The linear echo canceller is used widely as an indispensable technology for the videoconferencing system, etc.

The problem of echo is more serious with compact equipment such as a notebook PC or cellular phone. Fig. 2 shows the principles of nonlinear echo generation. When the voice from the other party is reproduced at a high volume through the loudspeaker on a compact cabinet, the loudspeaker and other parts in the cabinet vibrate, generating distorted sounds like a buzzer. This echo including the distortion is called the nonlinear echo in contrast to the linear echo that can be defined as a linear computation applied to the loudspeaker signal. A linear echo canceller cannot suppress the nonlinear echo completely, so the residual echo degrades quality of conversation. This problem is noticeable especially when a cheap loudspeaker or compact equipment is used.

Many notebook PCs, cellular phones and car-use handsfree kits equipped with a handsfree talk function use a technique called the voice switch to cancel the nonlinear echo. This technique is based on the fact that in most of a conversation, only one of the parties is speaking. It reduces the reproduced volume from one party while the other party is speaking. However, when both parties speak at the same time or the surrounding noise is significantly loud, the voices tend to be intermittent and a high-quality two-way communication becomes impossible.

In order to suppress the nonlinear echo, nonlinear adaptive filters such as a Volterra filter and a neural network are used, which are theoretically capable of expressing any nonlinearity. However, they need more than ten times more computations than the linear echo canceller and still cannot offer satisfactory echo cancellation even with the huge computations.

Fig. 3 shows the structure of the nonlinear echo canceller proposed in this paper. The nonlinear echo canceller first uses a linear echo canceller to suppress linear echo, it then applies FFT (Fast Fourier Transform) to convert the signal into a frequency spectrum, and then applies IFFT (Inverse Fast Fourier Transform) to re-synthesize the audio waveform, which is transmitted to the other party of the communication. The frequency spectrum processing is based on the newly discovered “correlation in spectrum amplitude between the nonlinear echo and the linear echo replica.”

Fig. 4 shows an example of spectrum amplitude correlation. The linear echo and nonlinear echo have completely different waveforms. However, as a result of surveys of quite a few types of compact equipment, it was found that their spectrum ampli-
Fig. 5 shows the results of evaluation with a cellular phone mockup, which is the most compact equipment with a small loudspeaker. Fig. 5(a) is the audio signal from the other party that is reproduced at the loudspeaker, and Fig. 5(b) is the signal waveform of the echo captured by the microphone. In Fig. 5(c), the echo is canceled using only the linear echo canceller. However, the nonlinear echo is still remaining and the sound of the spoken voice is still audible. On the other hand, with the output waveform of the proposed method, shown in Fig. 5(d), any echo including the nonlinear echo is cancelled to an inaudible level.

A new nonlinear echo suppression technology is presented, which enables compact equipment such as a notebook PC or cellular phone to support high-quality handsfree talk. We believe that the proposed method contributes to high-quality handsfree communications using a variety of equipment including PCs, softphones, VoIP handsets and cellular phones. We are planning to utilize the technology in commercial products in the near future.
Nonlinear Echo Suppression Technology Enabling Quality Handsfree Talk for Compact Equipment

References


Author’s Profile

HOSHUYAMA Osamu
Principal Researcher,
Common Platform Software Research Laboratories,
NEC Corporation
Member of IEEE, IEICE (Institute of Electronics, Information and Communication Engineers) and ASJ (Acoustic Society of Japan).