Evaluation of test signals for measuring residual echo characteristics of echo cancellers

Nobuhiko Kitawaki,* Mari Saito,* Futoshi Asano,** and Junko Sasaki***

* University of Tsukuba,

1-1-1, Tennoudai, Tsukuba, 305-8573 Japan

** ETL,

1-1-4, Umezono, Tsukuba, 305-8568 Japan

*** NTT-AT,

3-9-11, Midori-cho, Musashino, 180-8585 Japan

(Received 11 April 2000)

Keywords: Echo canceller, Residual echo characteristics, Artificial voices, ITU, P.50

PACS number: 43.58.-e

1. Introduction

Acoustical echo canceller has been developed to remove a room echo signal through the loudspeaker to the microphone in hands-free telephone network system. Adaptation algorithms such as LMS (Least Mean Square), ES (Exponentially weighted Step-size), and projection methods are used in the echo canceller. Performance of the echo canceller is evaluated by residual echo characteristics expressed in ERLE (Echo Return Loss Enhancement). The ERLE can be conventionally measured by putting white Gaussian noise into the echo canceller system.

However, white Gaussian noise is not adequate as the test signal for measuring the performance of the echo canceller, since the performance may depend on the characteristics of input test signal and the characteristics of the white Gaussian noise differ from those of real voice. If a real speech signal is used for such purpose, sufficient number of real voices is required because of talker dependency. In practice, only insufficient number of samples is used for the measurement because of time saving. This causes ambiguity for the objective quality measurement of echo cancellers.

This paper describes evaluation of test signals with human voice characteristics other than conventional white Gaussian noise for measuring residual echo characteristics of echo cancellers.

2. Test signals for measuring residual echo characteristics

Figure 1 shows schematic diagram for measuring residual echo characteristics. Test signal is input to the telephone network system equipped with echo canceller, and emitted to the room from loudspeaker. Acoustic echo and ambient noise are input to the microphone at the receiving room. Adaptive filter of the echo canceller generates echo replica by estimating impulse response in the acoustic echo path and cancels a room acoustic echo. Residual echo characteristics are expressed in ERLE.

Test signals studied in this paper are: (1) real voice,

(2) male and (3) female artificial voices specified by ITU-T Recommendation P.50 used for measuring speech coding performances,¹⁾ (4) composite source signal (CSS) specified by ITU-T Recommendation P.501 proposed for measuring telephony device performances,²⁾ (5) frequency weighted Gaussian noise used for measuring Loudness Rating (LR) in ITU-T,³⁾ and (6) conventional white Gaussian noise.

Figure 2 shows an example of real voice sequence. Real voices composed of Japanese short sentences spoken by 6 males and 5 females were thought as reference for performance evaluation. Figure 3 shows an example of male artificial voice sequence. Artificial voice P.50 reflects the average characteristics of the human voice such as long-term average spectra, instantaneous amplitude distribution, level distribution of segmental power, spectral distribution of segmental power, voiced/unvoiced structure of speech waveform, and short-term spectral characteristics.⁴⁾ Figure 4 shows an example of CSS sequence. CSS is composed of artificial voice P.50 sequence in voiced intervals, white Gaussian noise and silent sequences in which are mixed in random. Frequency weighted Gaussian noise reflects the long-term average spectra of the human voice as same as P.50.

3. Measurement of residual echo characteristics

ERLE, which expresses residual echo characteristics, is defined as following equation. Where, $y(k)$ denotes acoustic echo at the receiving room, $\hat{y}(k)$ denotes echo replica generated by adaptive filter in the echo canceller, and $n(k)$ denotes ambient noise in the receiving room.

$$ERLE = 10 \log_{10} \frac{p_y}{p_e} \text{ (dB)}$$

Where,

$$p_y = E[y(k)^2]$$

$$p_e = E[\tilde{e}(k)^2] = E[(y(k) - \hat{y}(k))^2]$$

$$\tilde{e}(k) = e(k) - n(k)$$

$$E[\cdot]: \text{Expectation}$$

Fig. 1 Measurement of residual echo characteristics.

Fig. 2 An example of real voice.

Fig. 3 An example of male artificial voice P.50.

The loudspeaker and microphone arrangement shown in Fig. 5 was set up conforming to ITU-T Recommendation P.340 which specifies acoustic measurement of hands-free telephone system.⁵⁾ Echo canceller used in this experiment was "Natural Talk PRO" produced by NTT (Nippon Telegraph and Telephone Corporation).^{6,7)} Sound level from the loudspeaker was 73 dB, and ambient noise level was 45 dB.

ERLE were measured by 6 test signals. Figure 6 shows ERLE characteristics for real voices, and Figs. 7

Fig. 4 An example of CSS.

Fig. 5 Loudspeaker and microphone arrangement conforming to ITU-T Rec. P.340.

Fig. 6 ERLE characteristics of real voices.

to 11 show those for various test signals. In the Figs. 7 to 11, the light colored curves are for real voices as the reference, and the dark colored curves are for each test signal other than real voices. Smooth curves approximate each ERLE characteristics.

Fig. 7 ERLE characteristics for male artificial voice P.50.

Fig. 8 ERLE characteristics for female artificial voice P.50.

Fig. 9 ERLE characteristics for CSS.

ERLE characteristics measured by male and female artificial voices according to P.50 are almost equivalent to those of real voices from Figs. 7 and 8. However, ERLE characteristics measured by CSS, white Gaussian noise, and frequency weighted Gaussian noise differ from those of real voices, and reduce very rapidly

Fig. 10 ERLE characteristics for white Gaussian noise.

Fig. 11 ERLE characteristics for frequency weighted white Gaussian noise.

compared with those of real voices from Figs. 9 to 11. This may cause over-estimation for the echo canceller system if CSS, white Gaussian noise, and frequency weighted Gaussian noise are used as a test signal.

4. Conclusion

This paper described evaluation of test signals with human voice characteristics other than conventional white Gaussian noise for measuring residual echo characteristics of echo cancellers. It is concluded that artificial voice P.50 recommended by ITU-T for evaluation of speech coding can be applied to objective measurement for residual echo characteristics of echo cancellers. Conventional white Gaussian noise may cause over-estimation for the echo cancellers.

References

- 1) "Artificial voices," ITU-T Recommendation P.50 (1993).
- 2) "Test signals for use in telephony on CD-ROM," ITU-T Recommendation P.501 (1996).
- 3) "Determination of sensitivity/frequency characteris-

- tics of local telephone systems," ITU-T Recommendation P.64 (1997).
- 4) K. Itoh, N. Kitawaki, H. Irii, and H. Nagabuchi, "A new artificial speech signal for objective quality evaluation of speech coding systems," *IEEE Trans. Commun.* **42**(2-4), 664-672 (1994).
 - 5) ITU-T Recommendation P.340: Transmission characteristics of hands-free telephones, 1996.
 - 6) S. Makino, Y. Kaneda, and N. Koizumi, "Exponentially weighted step-size NLMS adaptive filter based on the statistics of a room impulse response," *IEEE Trans. Speech Audio* **1**(1), 101-108 (1993).
 - 7) S. Makino and Y. Kaneda, "Exponentially weighted step-size projection algorithm for acoustic echo cancellers," *Trans. IEICE Jpn.*, **E75-A**, 1500-1508 (1992).