SUBJECTIVE ASSESSMENT OF ECHO RETURN LOSS REQUIRED FOR SUBBAND ACOUSTIC ECHO CANCELLERS

Sumitaka Sakauchi, Yoichi Haneda, and Shoji Makino

NTT Human Interface Laboratories
3-9-11, Midori-cho, Musashino-shi, Tokyo, 180 Japan
sakauchi@splab.hil.ntt.co.jp

ABSTRACT

The frequency characteristic of the echo return loss requirement (ERLR$_f$) was investigated using subjective assessments. The ERLR$_f$ is an important factor in the design and performance evaluation of a subband echo canceller (SBEC). The ERLR$_f$ during single-talk was obtained as attenuated band-limited echo levels that subjects did not find objectionable when listening to test speech and its band-limited echo under various transmission conditions. When we investigated the ERLR$_f$ during double-talk, subjects also heard near-end speech. Here, the echo was limited to a 250-Hz bandwidth assuming the use of an SBEC with 32 subbands. The test results showed that: (1) as the transmission delay rose above 100 ms, the ERLR$_f$ at low-frequency bands around 1 kHz increased significantly; (2) when the room reverberation time is relatively long (about 450ms), the ERLR$_f$ increases especially at low-frequency bands around 1 kHz even if the transmission delay is short (28 ms); and (3) the ERLR$_f$ during double-talk is about 5 to 10 dB lower than during single-talk. The obtained ERLR$_f$ will be useful for designing an efficient SBEC.

1. INTRODUCTION

To design an acoustic echo canceller (AEC), it is important to clarify how much returned echo should be suppressed perceptually, because the adaptive filter length and loss insertion level depend on the echo return loss requirement (ERLR). Moreover, the specified ERLR is also used as a basic criterion for the objective performance evaluation of an AEC.

Some subjective test results concerning ERLR for some transmission delays and room reverberation times have been reported [1], [2]. These showed that the ERLR increased significantly when the transmission delay rose above about 50 ms.

The conventional ERLR, however, has been investigated assuming the use of a fullband echo canceller, so the findings are not necessarily applicable to a subband echo canceller (SBEC) [3], [4].

The frequency characteristic of the echo return loss requirement (ERLR$_f$) has been studied theoretically using the known perceptual characteristics to determine optimal adaptive filter tap allocation tables for an SBEC [5]. The investigation did not, however, sufficiently consider the influence of various conditions in hands-free telecommunication, such as transmission delay or room reverberation time. Nor did it take into account whether the speech was single-talk or double-talk, whether the subject was a listener or a talker, or assessmental judgements as to whether the subject could hear the echo or whether the echo was objectionable.

We have used a subjective assessment to empirically investigate the ERLR$_f$ due to transmission delay ($T_{\rm D}$), reverberation time ($T_{\rm R}$) in the echo-path model room, and differences between single-talk and double-talk, which greatly affected the performance of an AEC as determined by preparatory experiments. This paper presents the subjective test results and a theoretical analysis of the ERLR$_f$. Furthermore, we present an example of a desired filter tap profile and discuss the loss insertion levels for an SBEC based on our obtained ERLR$_f$.

2. SUBJECTIVE ASSESSMENT SYSTEM

To investigate the ERLR$_f$, subjective tests were done with various transmission delays and various room reverberation times under single-talk and double-talk conditions. Figure 1 shows the subjective assessment system used to investigate the ERLR$_f$. The subjects heard both the test speech from a mouth simulator and its band-limited echo from a loudspeaker in the assessment room (the far-end). For the double-talk assessment, they also heard the test speech and its band-limited echo with the near-end speech in the echo-path model room added. The subjects suppressed the returned band-limited echo with an attenuator until the echo was not objectionable, and the attenuation level at that point was regarded as an ERLR_f.

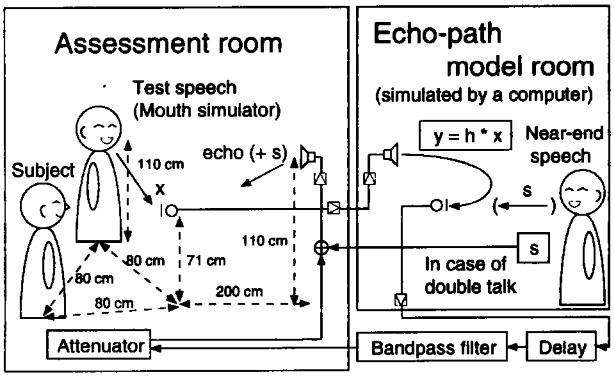

Figure 1: Configuration of the assessment system.

The test speech and the near-end speech were several short Japanese sentences spoken by either a man or a woman. The echo signals were calculated in a computer, where the test speech was convolved with the measured impulse responses corresponding to various reverberation times in the echo-path model room. The presented echo was limited to a 250-Hz bandwidth with a bandpass filter assuming a 64-subband echo canceller. Figure 2 shows the test speech, an example of its band-limited echo whose frequency range is from 625 Hz to 875 Hz (the number of subbands i=3), and the near-end speech, which were presented to the subjects.

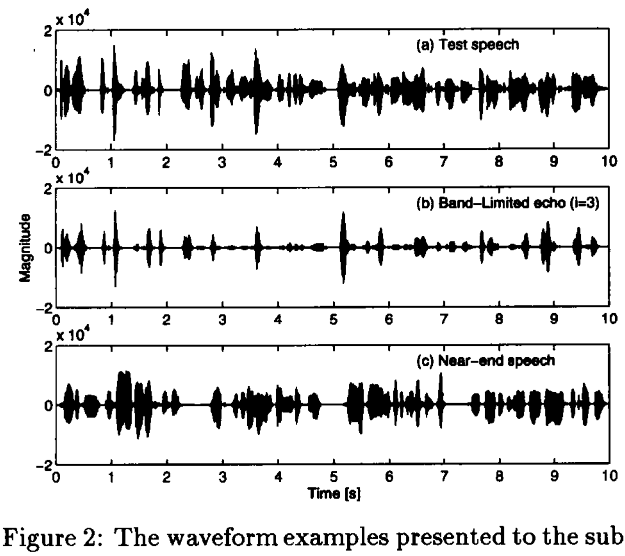

Figure 2: The waveform examples presented to the subjects; (a) test speech, (b) its band-limited echo, and (c) near-end speech.

The system's frequency range was set from 0.1 to 7 kHz, and the sound-pressure levels at the loudspeaker and the microphone were set according to ITU-T recommendation P. 34. The acoustic coupling of the echo-path model room was about -2 dB at all frequencies. The transmission delay was changed by using digital-delay equipment inserted into the echo-send line. The assessment room had a reverberation time of about 270 ms. The background noise levels in the two rooms were 30 dBA or less. There were 15 subjects whose ages were from 25 to 35. Here, all experimental data was evaluated by a t-test where the significance level was 0.01.

3. TEST RESULTS AND DISCUSSION

3.1. Influences of the transmission delay

Figure 3 shows subjective test results on the ERLR_f for transmission delays T_D of 28, 50, 100, and 300 ms. The reverberation time of the echo-path model room was set to 110 ms to minimize the effect of room reverberation.

These results show that when T_D is short (up to 50 ms), the ERLR_f reaches a maximum at 2 to 3 kHz. When T_D is 100 ms or more, the ERLR_f at low-frequency bands around 1 kHz increases significantly, but that of the other bands over 2 kHz is already saturated.

The maximum values of the ERLR_f for several transmission delays are consistent with previous ERLR results for a fullband AEC [1], [2].

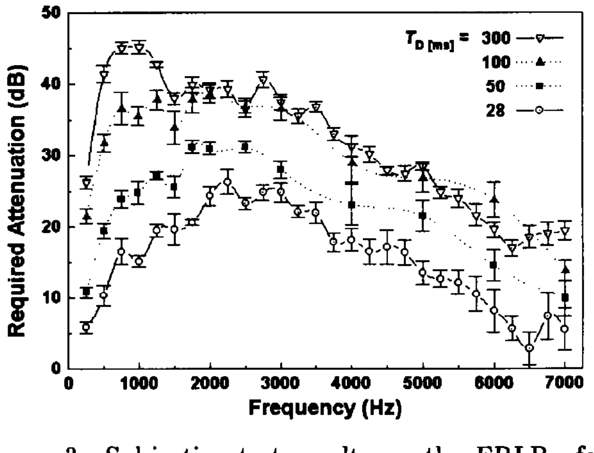

Figure 3: Subjective test results on the ERLR_f for different transmission delays.

When T_D is short, the subject hears both the test speech and its echo at almost the same time. As a result, the detectable echo of any frequencies that are over the simultaneous masking levels of the test speech have almost the same power level. Therefore, the ERLR$_f$ has a peak at 2 or 3 kHz where the human ear is perceptually the most sensitive. As $T_{\rm D}$ increases to about 50 ms, the ERLR$_f$ shifts upwards but its frequency characteristics do not change. This is because the temporal masking levels decrease due to the short time lag between the test speech and its echo.

When $T_{\rm D}$ is long, the subjects could fully distinguish between the test speech and its echo. Therefore, the ERLR$_f$ reaches a maximum at low-frequency bands around 1 kHz, since the frequency characteristics of the detectable echo are almost the same as those of the average sound power level of speech.

3.2. Influences of the reverberation time

Figure 4 shows subjective test results on the ERLR$_f$ for the reverberation times in the echo-path model room $T_{\rm R}$ of 110 ms and 450 ms for short (28 ms) and long (300 ms) transmission delays. These results show that when $T_{\rm D}$ is 300 ms, changes in $T_{\rm R}$ do not affect the ERLR$_f$, because the ERLR$_f$ is already saturated.

When $T_{\rm D}$ is 28 ms, the ERLR$_f$ at low-frequency bands for a long $T_{\rm R}$ is higher than that for a short $T_{\rm R}$. This is because even though $T_{\rm D}$ is short, the later reverberations affect the ERLR$_f$ in the same way that the long-transmission-delayed echo does. Here, the ERLR$_f$ of the long $T_{\rm R}$ is lower than that of the long $T_{\rm D}$, because the later reverberations have less energy than direct sound has.

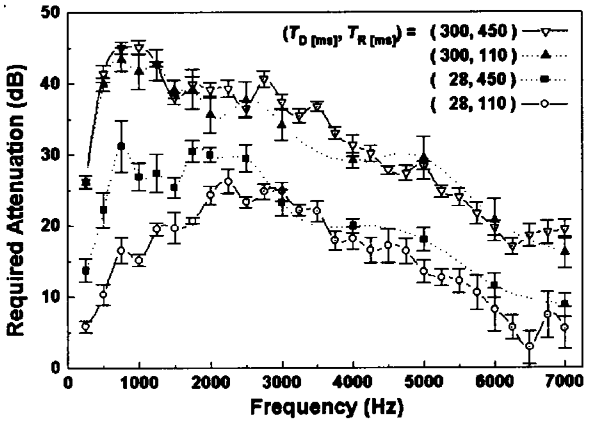

Figure 4: Subjective test results on the ERLR$_f$ for different reverberation times in the echo-path model room.

3.3. ERLR$_f$ during double-talk

Figure 5 shows subjective test results on the ERLR$_f$ for the difference between single-talk (ST) and double-talk (DT). When the transmission delay was short ($T_{\rm D}$ = 28 ms), the ERLR$_f$ during double-talk was about 5 dB lower than that for single-talk. When the transmission delay was long ($T_{\rm D}$ = 300 ms), the ERLR$_f$ during double-talk was about 10 dB lower.

When $T_{\rm D}$ is 28 ms, most of the echo is simultaneously masked by the test speech during single-talk. Therefore, the ERLR$_f$ was slightly lower due to the partial masking of the near-end speech during double-talk. When $T_{\rm D}$ is 300 ms, the echo is partially masked in the frequency domain by both the test speech and the near-end speech, whose influences are almost equal. Therefore, there is a larger difference in the ERLR$_f$ during double-talk than during single-talk.

These experimental results show that the influences on the ERLR$_f$ due to whether it is a case of single-talk or double-talk are as important as those of transmission delay. If the system can detect whether single-talk or double-talk is being used, it should be possible to reduce the loss insertion levels during the double-talk to improve the speech quality.

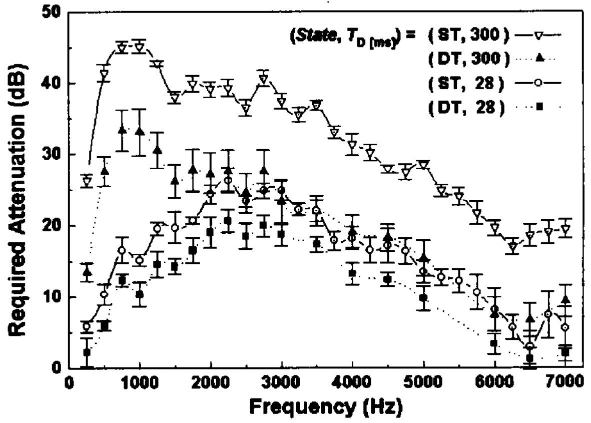

Figure 5: Subjective test results on the ERLR$_f$ for the differences between single-talk and double-talk.

4. AN EXAMPLE OF EFFICIENT SBEC DESIGN

These subjective test results can be applied to design an SBEC. Figure 6 shows a desired SBEC filter tap profile that was calculated based on our obtained ERLR$_f$ for $T_{\rm D}$ of 300 ms and $T_{\rm R}$ of 450 ms during single-talk in accordance with ITU P.167. The filter taps of each subband are given by

$$L_i = \frac{1}{60} \cdot \frac{T_{Ri}}{T_S \cdot M} \cdot \mathrm{ERLR}_{fi} \qquad (1)$$

where L_i denotes the number of filter taps, T_{Ri} denotes the reverberation times for the ith-subband, T_S denotes the sampling period, and M denotes downsampling rate.

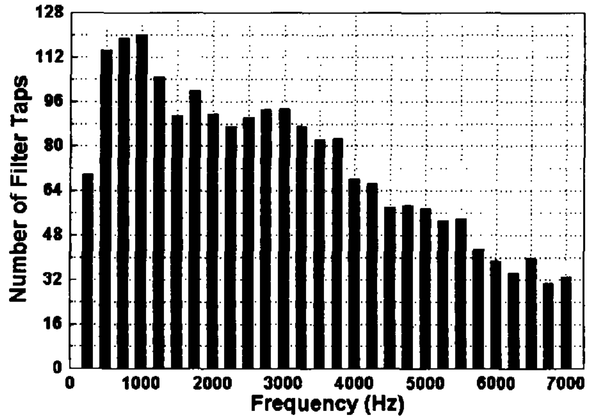

Figure 6: Desired SBEC filter tap profile.

Figure 3 shows, however, the frequency characteristics of the ERLR$_f$ differ depending on whether there is a short or long transmission delay. That makes it possible to optimize the SBEC's hardware design to make it much smaller, because optimal tap profiles can be selected corresponding to the various transmission conditions that an SBEC is used under.

It is well known that appropriate echo reduction cannot be achieved by using only the adaptive filter when, for example, the acoustic path between a loudspeaker and a microphone changes rapidly or in a noisy room. Our obtained ERLR$_f$ can also be used to determine optimal and robust loss insertions for subbands that keep the decline in the quality of the speech to a minimum.

5. CONCLUSIONS

We have investigated the ERLR$_f$ for the transmission delay, the reverberation time in the echo-path model room, and differences between single-talk and double-talk through subjective assessments. Subjective test results show that the ERLR$_f$ at low-frequency bands around 1 kHz increases significantly when the transmission delay is 100 ms or more. The effect of reverberation time on the ERLR$_f$ is weaker than that of the transmission delay. When the reverberation time is long, however, the ERLR$_f$ at low-frequency bands increases, even if the transmission delay is short. The ERLR$_f$ during double-talk is lower than that during single-talk, especially when the transmission delay is long. Using these results from our experiments on the ERLR$_f$, we have obtained an outline for the efficient design of an SBEC.

ACKNOWLEDGMENTS

We would like to thank Dr. N. Kitawaki, and Dr. Y. Kaneda for their discussions and many suggestions.

REFERENCES

[1] H. Yasukawa, M. Ogawa and M. Nishino,"Echo return loss required for acoustic echo controller based on subjective assessment," *IEICE Transactions*, Vol. E-74, No. 4, pp. 692-705, April 1991.

[2] N. Kishimoto, K. Ishimaru and K. Takahashi, "Transmission quality of hand-free audio teleconference services," *IEEE ICC'88*, No. 8.4, June 1988.

[3] A. Gilloire, "Experiments with sub-band acoustic echo cancellers for teleconferencing," *Proc. ICASSP87*, pp. 2141-2144, 1987.

[4] W. Kellermann, "Analysis and design of multirate systems for cancellation of acoustical echoes," *Proc. ICASSP88*, pp. 2570-2573, Apr. 1988.

[5] E. J. Diethorn, "Perceptually optimum adaptive filter tap profiles for subband acoustic echo cancellers," *Proc. IEEE Workshop Applic. Sig. Proc. Aud. Acoust.*, Oct. 15-18, 1995, New Paltz, NY.