Quality assessment of speech signals under a process of echo cancellation in telecommunication systems

Andrzej Dobrucki, Maurycy Kin, Maciej Walczyński
Wrocław University of Science and Technology, Poland

The echo impression is created when a reflected signal is delayed by at least 100 ms from the direct signal. Then the human ear is able to extract two sounds: direct sound and reflected one. In the case that the delay of the reflected signal is less than 100 ms, we are talking about reverberation and a listener is not able to distinguish the direct and reflected sound. The phenomenon of echo in the telecommunications channel is caused by the reflection of an electrical signal in a long line and by the processing of sound in hybrid telecommunications systems. In order to improve the quality of the transmitted sound, various adaptive filters are used to remove or at least reduce the level of the reflected delayed signal. However, such a process may result in a degradation in the quality of speech, although its intelligibility may not get worse.

The work presents the results of subjective studies on assessing the quality of speech signals under the process of acoustic echo cancellation using different algorithms. The algorithms studied were: Least Mean Squares (LMS), Normalized Least Mean Squares (NLMS) and Affine Projection (AP). The study consisted of assessing the signal quality after applying the echo elimination process using the Degradation Category Rating method. A total of 312 signals were used in the test: 192 male speech and 120 female speech samples. Echo simulation was used using different delay times and levels of echo signal. Both types of speech have signal delay times of 20 ms, 50 ms, 100 ms and 200 ms with echo level values of -6 dB, -12 dB, -18 dB and -24 dB. In addition, for female speech signals, a delay time of 150 ms was introduced. The study involved 14 people aged between 18 and 38, including six women and eight men. All subjects had normal good hearing. Half of those participated hearing tests to assess the quality of music signals so they were experienced in the subjective research. The listeners' opinions were collected on prepared questionnaire.

It was found that the highest ratings were given to the Affine Projection (AP) filter, while the worst ratings were featured the Least Mean Squares (NLMS). It should also be noted that the range between the results obtained for AP and NLMS for female speech is smaller in comparison to male. It is also interesting that the discrepancy in ratings was greatest for a delay time of 100 ms for the AP filter and 200 ms for the LMS filter. It can therefore be concluded from the obtained results that, in the case of acoustic echo cancelation, the Affine Projection (AP) filter introduced the lowest quality degradation while the Least Mean Squares (LMS) achieved slightly worse average ratings when compared to AP filter. The Normalized Least Mean Squares (NLMS) filter characterized by the worst ratings, and in some cases received twice the quality degradation compared to the AP filter.