

EVALUATION OF A HANDS-FREE UNIT DURING DOUBLE-TALK

Kenneth Wiklund^a, Fredric Lindström^a, Ingvar Claesson^b

^aKonftel AB, Research and Development, Box 268, SE-90106, Umeå, Sweden, fli@konftel.com

^bBlekinge Institute of Technology, Department of Signal Processing, SE-37225, Ronneby, Sweden

Abstract- The double-talk performance of a hands-free speech communication system is crucial to the overall quality. This paper presents a method to evaluate double-talk performance of conference phones and speakerphones. The method measures loudness, clipping and the amount of echo during double-talk.

Index Terms- Hands-free, double-talk, performance evaluation, speech quality

1. INTRODUCTION

Hands-free operations are desirable in many different situations and in relation to many products, e.g. car phones and conference phones. The A-side in a hands-free communication refers to the side of the loudspeaking telephone. The B-side is thus the other side. The telephone is said to be in sending mode when the A-side speaker is active, and in receiving mode when the B-side speaker is active.

In hands-free systems the most critical situation is the double-talk situation, i.e. when the A-side and B-side talkers talk simultaneously. Three parameters that are decisive for the performance of the device during a double-talk situation are: the loudness level, the degree of clipping and the amount of echo [1]. The impairments during double-talk normally occur on the send signal due to the manufacturers trying to keep the loudspeaker signal unaffected.

This paper describes a method to evaluate the send signal during double-talk. The method is suitable for optimization of algorithms and comparative performance tests between different hands-free devices.

2. EVALUATION METHOD

The method uses two test signals, $a(k)$ and $b(k)$, which are input signals on the A-side and B-side respectively, see figure 1. Two output signals, $s(k)$ and $m(k)$, are recorded. The physical test arrangement and the test table used are the same as described in [2]. A simplified scheme of the test arrangement is shown in figure 1. Three different parameter

values are calculated, L_{dt} , C_{dt} and E_{dt} . These three parameters corresponds to loudness, degree of clipping and the level of echo of the send signal, $s(k)$, during double-talk.

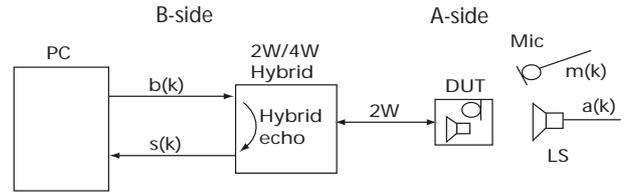


Fig. 1. A simplified schematic of the test arrangement.

The test signals $a(k)$ and $b(k)$ consist of five segments each, see figure 2. The first segment is an initialization segment which can be used to get the device under test (DUT) in a steady state, i.e. in a normal operation condition. The second segment contains white gaussian noise (WGN) on the B-side signal.

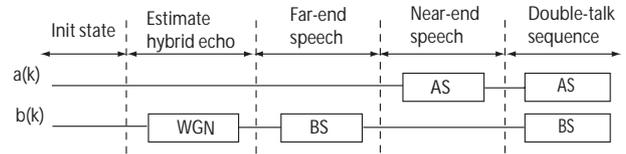


Fig. 2. Signal construction.

The last three segments consist of single B-side speech BS, single A-side speech AS and a double-talk sequence. Speech could be replaced with any test signal, e.g. white noise or a composite source signal (CSS). However, using speech facilitates comparisons between objective test values and subjective assessments.

3. PARAMETER ESTIMATION

The hybrid, see figure 1, echoes a part of the $b(k)$ signal back to the B-side. This echo will mask echoes generated by the DUT and should thus be removed when evaluating the double-talk performance of the DUT.

The LMS algorithm [3] is used with the WGN block in the $b(k)$ and $s(k)$ signal to estimate the impulse response of the hybrid h . The hybrid echo can thus be removed from the

Thanks to the Swedish Knowledge Foundation (KKS) for funding.

$s(k)$ signal, resulting in the signal $r(k)$, i.e. $r(k) = s(k) - b(k) * h$, where $*$ denotes convolution.

The loudness, L , between two signals $x(k)$ and $y(k)$ is calculated as a frequency weighted ratio, i.e.

$$L(x, y) = -\log \sum_{i=0}^N \frac{\sum_k (x(k) * h_i)^2}{\sum_k (y(k) * h_i)^2} \alpha_i. \quad (1)$$

where N is the total number of frequency bands, i is the frequency band index, h_i are frequency band selective filters and α_i the weighting factors. The filters h_i and the weighting factors α_i are basically as those given in [4].

The sending loudness during double-talk S_{dt} is thus given by

$$S_{dt} = L(r_{dt}, a_{dt}), \quad (2)$$

where $a_{dt}(k)$ is the $a(k)$ signal during double-talk and $r_{dt}(k)$ is the aligned $r(k)$ signal during double-talk, i.e. the signal $r_{dt}(k)$ is shifted so that the largest cross correlation coefficient between $r_{dt}(k)$ and $a_{dt}(k)$ is for lag 0. If the S_{dt} value is low, the B-side listener will not hear any A-side speech, i.e. a poor double-talk performance. However, a high value of S_{dt} does not automatically imply good double-talk performance. The $r_{dt}(k)$ could consist of a lot of clipping and/or echo in the A-side signal.

To measure the amount of clipping during double-talk the A-side input signal during double-talk $a_{dt}(k)$ is compared blockwise with $r_{dt}(k)$. The signal $a_{dt}(k)$ is divided into blocks $a_{dt}^l(n)$ of length L_B , i.e. $0 > n > L_B - 1$, where l denotes the block index. Likewise, $r_{dt}(k)$ is divided into blocks $r_{dt}^l(n)$. The clipping C_{dt} is defined as the ratio between n_{tot} and n_L , where n_{tot} is the number of active blocks, and n_L the number of active blocks for which the loudness level is below a certain threshold T_L , i.e.

$$C_{dt} = \frac{n_L}{n_{tot}} = \frac{\sum_l B_L(l)}{\sum_l B_{tot}(l)}, \quad (3)$$

where

$$B_{tot}(l) = \begin{cases} 1 & \text{if } \sum_n a_{dt}^l(n)^2 \geq T_{tot} \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

$$B_L(l) = \begin{cases} 1 & \text{if } L(r_{dt}^l, a_{dt}^l) < T_L \text{ and } B_{tot}(l) = 1 \\ 0 & \text{otherwise,} \end{cases} \quad (5)$$

where T_{tot} is a threshold defining active speech blocks. A high C_{dt} thus indicates a high degree of clipping.

The echo during double-talk E_{dt} is calculated using a weighted coherence function, i.e.

$$E_{dt} = \sum_{j=0}^M C(j) \beta_j, \quad (6)$$

where $C(j)$ is the coherence function [5] between $b_{dt}(k)$ and $r_{dt}(k)$, j is the frequency bin index, M the number of

frequency bins and β_j is a frequency selective weight factor. β_j is set in accordance with the discussion on the average spectral distribution of a speech signal in [6].

A high value in the coherence estimate $C(j)$ indicates a high correspondence between $b_{dt}(k)$ and $r_{dt}(k)$. Thus, a high value of E_{dt} indicates a high level of echo present in $r_{dt}(k)$.

4. USAGE

The test method can be used in optimizing an algorithm. The optimal algorithm parameter settings are those that maximize S_{dt} under the constraints $C_{dt} < T_C$ and $E_{dt} < T_E$, where T_C and T_E are thresholds. The final parameter setting should be verified using subjective testings.

When the method is used to compare different devices, the worst performance equals a totally interrupted transmission, i.e. a low value of S_{dt} . A lot of clipping, i.e. a high value of C_{dt} , is generally considered to be worse than echo, i.e. high values of E_{dt} . To obtain correct test results it is important that the receive loudness, i.e. $L(b, m)$, of the different devices are the same. The loudness $L(b, m)$ is preferably evaluated during the B-side speech section, see figure 2. The test method presented in this paper has been successfully used by the authors in the optimization of algorithms as well as in comparisons between different hands-free sets.

5. CONCLUSION

This paper describes a method to compare double-talk performance of hands-free sets. The method calculates three different parameters which are used to evaluate the performance.

The method can be used in algorithm optimizations as well as for comparisons of the performance of different devices.

6. REFERENCES

- [1] ITU-T Recommendation P.502, "Objective test methods for speech communication systems using complex test signals", 05/2000.
- [2] ITU-T Recommendation P.340, "Transmission characteristics of hands-free telephones", 08/1996.
- [3] S. Haykin, *Adaptive filter theory*, Prentice-Hall, 4th edition, 2002.
- [4] ITU-T Recommendation P.79, "Calculation of loudness rating for telephone sets", 09/1999.
- [5] Proakis, *Digital Signal Processing*, Prentice-Hall, 3rd edition, 2002.
- [6] T. Gänslar, M. Hansson, C-J Ivarsson, G. Salomonsson, "A double-talk detector based on coherence", *IEEE Trans. on Comm.* Vol 44, No. 11, Nov. 1996.