ECHO CANCELLATION II: DOUBLE TALK DETECTION AND ENVIRONMENTAL NOISE INFLUENCE

Miroslav BRODSKÝ, Pavel SOVKA
Dept. of Circuit Theory
Czech Technical University in Prague
Technická 2, 16627 Praha 6
Czech Republic

Abstract

Two problems arising in the real-life application of echo cancellation systems are analysed. The first, simultaneous activity of both telephone users (double talk) deteriorates the echo suppression. The second, environmental noise is the crucial point in echo cancellation system applications. Experimental evaluation of the influence of both phenomena is given together with possible solution.

Keywords

echo , echo cancellation, hands free, LMS adaptive filter, double talk, crosstalk

1. Introduction and problem definition

For real-life application of the echo canceller two problems must be solved.

As shown in [1], the adaptive algorithms are driven by the mean square error of the output signal e(n). If this signal contains the far end speaker speech r(n) and drivers near end speech s(n) (double talk) the result is the ERLE deterioration.

The driver's speech acts as the disturbance for the adaptive filter adjustment. His speech s(n) penetrates into the directional output signal e(n) and causes the wrong filter adaptation. The greater driver's power of speech causes the worse ERLE. As the criterion for the ERLE evaluation the signal to signal ratio (SSR) is used (see Fig. 1).

To minimise the effect of double talk the robust voice activity detector (VAD) must be used. While the driver's speech is detected the update of adaptive filter must be immediately stopped. In this way the errors in the impulse response cabin car estimation can be minimised.

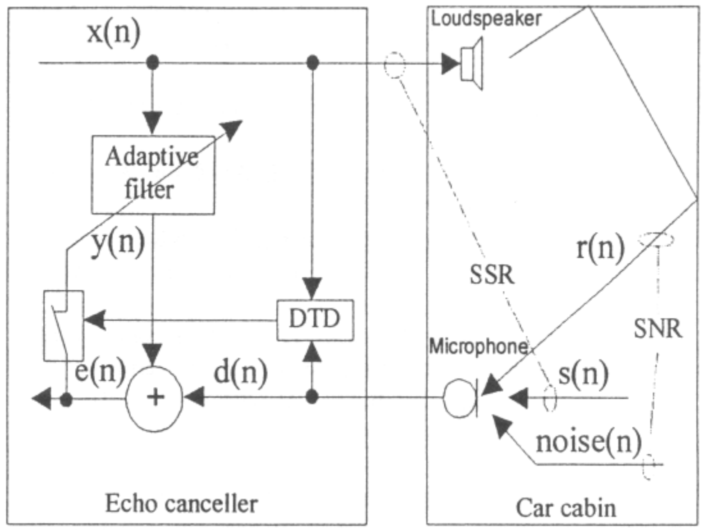

Fig. 1 Block diagram of an echo cancelling system

Legend:

x(n)	... speech of far-end speaker
r(n)	... far-end speaker output
Noise(n)	... environmental car noise
s(t)	... driver's speech
d(n)	... digitised microphone signal
e(n)	... error signal
SNR [dB]	... signal to noise ratio
SSR [dB]	... signal to signal ratio (far/near)
DTD	... double talk detector
+	... summer

The second problem is the environmental noise presence in the car. This noise represents another disturbing factor and therefore it is desired to suppress it. In this case it is not possible to use VAD detector because the car noise is not intermittent like driver's speech. Therefore the noise suppression system must be used.

In this contribution we analyse only the influence of a car noise on the echo canceller behaviour to find proper requirements on the noise suppression system.

2. The double talk influence

Let's study behaviour of MDF and LMS adaptive algorithms, which were published in [1], in the case of near-end speaker speech. The results are shown in Fig. 2.

The order of analysed filters is 256 and for testing is used speech signal with 66.400 samples and sampling rate is 8 kHz. The block length is 256, overlapping 50 % with OLS method. Independent variable is therefore segment of the length 128 samples, which will be called Half Frame (HF). All-important signal changes in time are quantified in HF, samples and also in seconds (1 HF=128/8000=16μs).

Average value of ERLE is calculated only during double talks.

The computer program MATLAB® for Windows, version 4.2c.1. was used.

The red rectangle frame in Fig. 2 denotes the near-end speaker activity. Red or green colours represent MDF and LMS algorithms, respectively.

The MDF algorithm exhibits two transitional events in nonstationary case during one driver's movement.

- First part it is the initial setting up coefficients of filter, its duration is approximately 80 HF (1.28 s).
- Second event occurs under untuning the filter caused by the speech of near-end speaker, its duration is approximately 150 HF (2.4 s).

The behaviour of LMS algorithm is maybe surprising. This algorithm is too slow therefore a short sermon can not untune it.

For the evaluation of algorithm dependence on the near-end speaker's activity we use ERLE computed only during the second transitional event. MDF algorithm reached the average ERLE –16,7 dB and LMS -24,5 dB. These results show the better echo suppression of slow algorithm, which is just for its slowness and small precision practically unemployable.

3. Coherence Double Talk Detector

The solution the problem with filter untuning is in using a detector, which allows recognise double talks. For the detection of this situation such method must be used, which can discern the difference between changes of impulse response and changes caused by the double talk. This way allows stopping updating coefficients of adaptive filter and after finishing of double talk the system can continue in the convergence process with a minimal loss of ERLE.

The detector uses for the discerning of the present near-end speaker activity in the d(n) the coherence function [6] between signals d(n) and x(n) on a given frequency k:

$$\vec{\gamma}_{xd}^{\,2}(k)=\frac{\left|\overrightarrow{CSD}_{xd}(k)\right|^2}{\overrightarrow{PSD}_x(k)\otimes\overrightarrow{PSD}_d(k)},\quad \gamma_{xd}^2(k)\in\langle 0,1\rangle\quad [-] \tag{1}$$

Where sign ⊗ represents array multiplication of the vectors PSD. CSD is the cross spectral density between d(n) and x(n), and PSD is the power spectral density of both signals:

$$\overrightarrow{PSD}_x(k)=E\{\left|\vec{x}(k)\right|^2\}=E\{\vec{x}(k)\otimes \mathrm{conj}(\vec{x}(k))\} \tag{2}$$

$$\overrightarrow{PSD}_d(k)=E\{\vec{d}(k)\otimes \mathrm{conj}(\vec{d}(k))\} \tag{3}$$

The cross spectral density of signals x(n) and d(n) is defined:

$$\overrightarrow{CSD}_{xd}(k)=E\{\vec{x}(k)\otimes \mathrm{conj}(\vec{d}(k))\} \tag{4}$$

For practical calculations we have to used estimates:

$$\overrightarrow{P\hat{S}D_x}(k+1)=\beta.\overrightarrow{P\hat{S}D_x}(k)+(1-\beta)\sum\overrightarrow{P_x}(k) \tag{5}$$

$$\overrightarrow{P\hat{S}D_d}(k+1)=\beta.\overrightarrow{P\hat{S}D_d}(k)+(1-\beta)\sum\overrightarrow{P_d}(k) \tag{6}$$

$$\overrightarrow{C\hat{S}D_{xd}}(k+1)=\beta.\overrightarrow{C\hat{S}D_x}(k)+(1-\beta)\sum\overrightarrow{P_{xd}}(k) \tag{7}$$

Where β is the parameter controlling the time constant of integrators.

Using equations (5)(6)(7) we obtain the coherence estimate:

$$\hat{\vec{\gamma}}_{xd}^{\,2}(k)=\frac{\left|\overrightarrow{C\hat{S}D}_{xd}(k)\right|^2}{\overrightarrow{P\hat{S}D}_x(k)\otimes\overrightarrow{P\hat{S}D}_d(k)} \tag{8}$$

The accuracy of estimates of these functions is very important for the proper function of the detector. This accuracy depends mainly on the size of the used DFT (FFT) transformation, which is directly connected with a spectral resolution. From this point of view it is required the large size of transformation. This condition is in antagonism with the condition of fast algorithm convergence. Experiments revealed that the minimal size of transformation, which can assure rightful function of algorithm, is 256 samples.

The MDF algorithm equipped with the coherence detector will be called MDFC in the rest of this article.

4. Experiments and results

MDFC algorithm is tested on real signals and car noises, which were picked up in a driving car. Transfer path between loudspeaker and microphone was approximated with a measured acoustic impulse response of cabin a car [2]. The accuracy of detector searching the area of talking near-end speaker is described by the error estimation of begin and end of this domain. This error is quantified in the HF - see Tab. 1. When the near-end speaker's signal is strong, the detector correctly estimates the beginning of the double talk while the estimation of the end is delayed. And on the other hand (stronger far end speaker's signal) reversibly.

4.1 Double talk problem

Let's study the sensitivity of the DTD to the SSR value first. The testing signals are parts of speech obtained from different speakers with a different ratio SSR [dB] (far/near) and with constant value of SNR=100. All introduced units are used in Tab. 1 for better overview.

SSR	begin error			end error		
[dB]	[HF]	[samples]	[μs]	[HF]	[samples]	[μs]
60	12	1536	192	11	1408	176
50	11	1408	176	2	256	32
35	10	1280	160	6	768	96
20	10	1280	160	8	1024	128
15	10	1280	160	9	1152	144
10	10	1280	160	9	1152	144
0	10	1280	160	9	1152	144
-5	10	1280	160	10	1280	160
-10	10	1280	160	10	1280	160
-20	9	1152	144	10	1280	160
-40	6	768	96	13	1664	208
-70	3	384	48	18	2304	288

Tab. 1 The error of DTD algorithm during the double talk

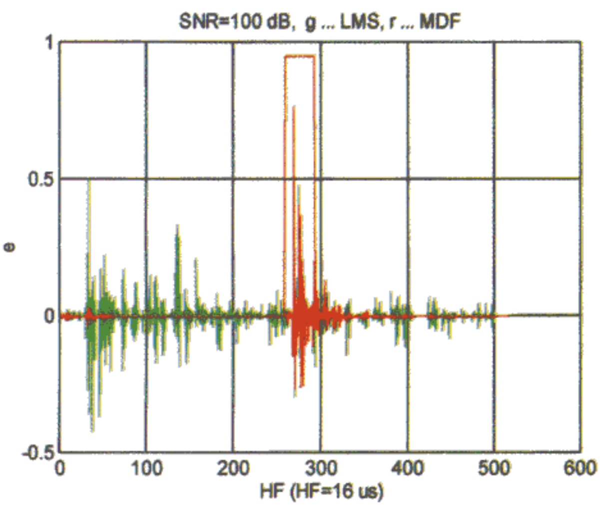

Fig. 2 The influence of double talk on the echo canceller

In Fig. 3 all signals during the whole simulation for SNR=100 dB and SSR = 0 dB are displayed:

- First three graphs represent main signals x(n), d(n), e(n) time development. The black rectangle in x(n) signal graph means voice activity of far end speaker. Red highlighting areas denote occasion of near-end speaker activity.
- Fourth group of curves represents the ERLE factor development, which during double talk increases to 0 dB.
- Fifth group of curves represents the result of coherence detector. As we can see, in this situation the error of double talk beginning is 10 HF (0,16 s) and the error in the double talk end estimation is 11 HF (0.176 s).

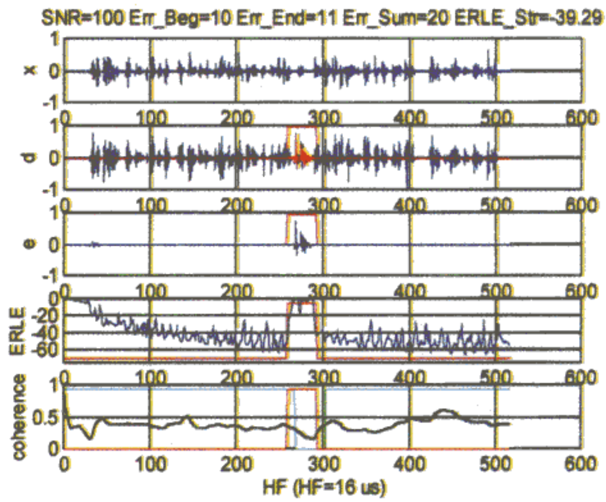

Fig. 3 The echo canceller equipped with DTD

Detailed view to the error signal e(n) development in this case is shown in Fig. 4.

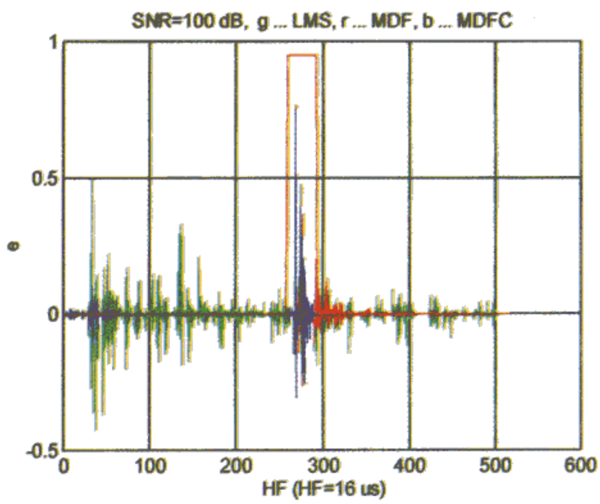

Fig. 4 The impact of DTD function on error signal e(n)

When the near-end speaker finished talking, system with the DTD (blue curve) allowed filter updating and the convergence of LMS algorithm could continue without the loos of previous value.

If evaluating the factor ERLE during the second transitional event, we obtain the average ERLE – 48,7 dB, which represents improvement 190 % comparing to the system without detector. The second transitional events have duration approximately 40 HF (0,64 s), which is really much better (73 %).

4.2 Car noise influence

Analyse of AEC with the coherence detector was done with SSR=0 dB. The used SNR ratio varied from the 50 dB to 0 dB. The critical SNR level for the proper detector function is 30 dB. Under this level the new parts of signal are assigned as the double talk. The increasing noise level causes these errors. In spite of this fact the determination of the double talk activity remains practically unchanged, and therefore AEC works still correctly. The increasing of ERLE is the result of the slower convergence rate due to often stopping coefficients updating. The average value of ERLE increases with increasing level of noise. The minimal value of ERLE is –35 dB. Subjectively we can determine the break point at SNR= 10 dB, when the value of ERLE is dropping to - 8,4 dB. This situation is described in Fig. 5.

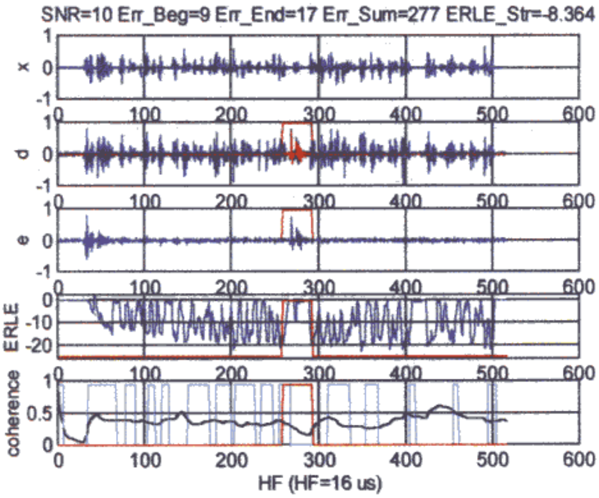

Fig. 5 The influence of car noise on DTD

	error of coherence detector			
SNR [dB]	Total_err [HF]	Begin_err [HF]	End_err [HF]	ERLE [dB]
50	21	10	11	-35
40	25	10	11	-30,3
30	41	10	11	-23,1
20	129	10	13	-15,7
10	277	9	17	-8,4
0	403	8	29	-3,4

Tab. 2 The sensibility of coherence detector to noise

In Tab. 2 the term Total_err means the sum of all errors during the whole echo cancellation process.

5. Conclusions

For real applications the MDFC algorithm with the coherence detector of double talk activity seems to be sufficient. This algorithm can be used without troubleshooting in a room or in a quite car cabin where SNR is greater than 20 dB. This algorithm is able to suppress the echo with the factor ERLE better then – 15 dB.

Under noisy conditions the detector interrupts updating of coefficients of adaptive filter too often and therefore the final value of ERLE is higher (lower suppression of echo).

Further research will be focused on the searching for a optimal combination of noise suppression systems with a echo cancellation system.

References

[1] BRODSKÝ, M. : Echo cancellation I: Algorithms simulation, Radioengineering, Vol.9 No.4, March 2000

[2] TŮMA, M. : Echo cancellation with application in a hand-free radiotelephony, (in Czech) Diploma work, Czech Technical University in Prague, 1998.

[3] BRODSKÝ, M., Sovka, P.: System for echo cancellation of hands-free telephones during double talk, Radioelektronika99, Brno 1999

[4] BRODSKÝ, M. : Echo cancellation of hands-free telephones during double talk, Poster 1999, FEE CTU, Prague 1998

[5] BRODSKÝ, M. : System for echo cancellation for hands free phones in a driving car during double talk, (in Czech) Minimal work, Czech Technical University in Prague, 1999.

[6] UHLÍŘ, J., SOVKA, P. :Digital signal processing, (in Czech) FEE CTU, Prague, 1995

About authors...

Miroslav BRODSKÝ was born in Prague, Czechoslovakia, 12th May 1974. He graduated in 1997 at the Faculty of Electrical Engineering of the Czech Technical University in Prague (FEE CTU). He is Ph.D. student at the same faculty now. His main research is devoted to echo cancellation adaptive systems and algorithms.

Pavel SOVKA – see contribution in vol.8, No.4, Dec 1999