A Double-Talk Detector Based on Coherence

Tomas Gänsler, Maria Hansson, Carl-Johan Ivarsson, and Göran Salomonsson

Abstract—**In this paper, we address the problem of detecting double-talk in a full duplex transmission line. A new double-talk detector (DTD) based on measuring the similarity between the far- and near-end speech signals is proposed. The detector is block oriented and operates in the frequency domain where the similarity between the signals is measured by means of the coherence function. The coherence is estimated with a short sequence of data by exploiting the multiple window spectrum estimation technique. Theoretical evaluation and examples of its performance are presented. The proposed DTD operates accurately in a wide range of situations, i.e., a difference in speech levels and hybrid attenuations ranging from 0 to 20 dB.**

Index Terms—**Double-talk detector, coherence, multiple-windows.**

I. Introduction

A LONG-DISTANCE telephone network consists of two-wire lines in the local networks, connecting the subscribers to the local offices, and a four-wire line used in the long haul transmission. The transition from the four-wire line to a two-wire line is performed by a hybrid circuit. Since this balancing circuit never is perfectly matched to the connected two-wire circuit, a part of the incident signal will leak over and the transfer function of this leakage will form the echo path. The part will be transmitted back to the talker, who recognizes it as an echo. Today, the most common equipment for suppressing echoes is echo cancellers.

An echo canceller in a long-haul connection is usually implemented by an adaptive finite impulse response (FIR)-filter, the adaptation algorithm which belongs to the least mean square (LMS)-family. The adaptation procedure has to be inhibited during double-talk periods. A double-talk detector (DTD) is used as a controller of the adaptation procedure. Since the DTD is a main function in the echo canceller, a number of papers can be found in the literature concerning this topic [1]–[8].

In this paper, a block-oriented frequency based algorithm for detection of double-talk is proposed. The algorithm compares the frequency properties of the input and output signals of the echo path. The comparison is made by means of the coherence function, which is estimated by using the speech signals from far and near-ends. To obtain this estimate, the power density spectra of the signals as well as their cross density spectrum are estimated. A sliding window forms blocks of the signals and the density spectra are estimated by a multiple window spectrogram [9]. The method allows the blocks to be short in length and yet a low variance estimate is obtained.

Paper approved by S. Dimolitsas, the Editor for Speech Processing of the IEEE Communications Society. Manuscript received August 10, 1995; revised October 24, 1995. This work was supported in part by Telia AB. This paper was presented in part at the 6th International Conference on Signal Processing and Technology, Boston, MA, October 24–26, 1995.

The authors are with Signal Processing Group, Department of Applied Electronics, Lund Institute of Technology (LTH), Lund, Sweden (email: tg@tde.lth.se; mh@tde.lth.se; cji@tde.lth.se; and gs@tde.lth.se).

Publisher Item Identifier S 0090-6778(96)08584-4.

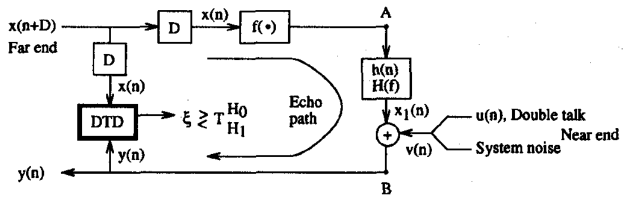

Fig. 1. The block diagram of the echo path.

An average of the estimated coherence function over some frequencies is used as a decision parameter. The frequencies are selected from an interval where the coherence function is sensitive to double-talk. The paper is organized as follows: Basic equations and notations are given in Section II, where requirements on the DTD in terms of probabilities of misses and of false alarm are discussed. A short description of the method for estimating the coherence function is given in Section III together with an example of the estimate with and without double-talk. A statistical analysis of the decision parameter is carried out in Section IV and the performance of the decision rule is given by a receiver operating characteristic (ROC). In Section V a comparison between the proposed DTD and a standard power level detector is presented by means of examples.

II. Basic Equations, Notations

A block diagram of an echo path is shown in Fig. 1. The transmission between the points A and B can either be electric as in the transition from a four-wire line to a two-wire line, or it can be acoustic as in handsfree telephony. The echo path is characterized by three boxes. The first, denoted D, is a model of the round-trip delay (flat delay). The delay D is in the following assumed to be accurately known by an estimation procedure. A nonlinear circuit $f(\cdot)$ is included in the echo path. This nonlinearity may be caused by over-loaded amplifiers or codecs. The circuit $f(\cdot)$, will not be taken into consideration in the following since it is assumed to be small. The linear part of the echo path is characterized by the impulse response $h(n)$ or by the corresponding transfer function $H(f)$.

The far-end input signal to the echo path is $x(n+D)$. The corresponding output signal is $x_1(n)$. It is corrupted by additive noise $v(n)$, which is the sum of double-talk, near-end signal $u(n)$, and system noise. Finally the observed signal $y(n)$ is the sum of $x_1(n)$ and $v(n)$. The DTD will use the two signals $x(n)$ and $y(n)$ to make the decision whether $y(n)$ contains double-talk or not. The power of $u(n)$ and $x(n)$ are denoted P_u and P_x and the power ratio of the talkers is

$\kappa = 10\log_{10}(\frac{P_u}{P_x})$. The noise level of the system is described by the x-signal to noise ratio (SNR) $= 10\log_{10}(\frac{P_x}{P_n})$, where P_n is the power of the system noise.

A parameter, ξ, has to be formed, with which a decision can be made whether the corrupting noise, $v(n)$, is annoying or not. It is equivalent to the question of making a reliable estimate of the signal, $x_1(n)$, since if this is possible then the near-end disturbance is small. This estimator will in the MSE-sense be the Wiener filter, $F(f)$, given by

$$F(f) = \frac{S_{yx_1}(f)}{S_{yy}(f)} = \frac{S_{xx}(f)|H(f)|^2}{S_{yy}(f)} \tag{1}$$

where $S_{..}(f)$ are the cross- or autopower density spectra of the signals involved. If the estimated filter function is close to one, i.e., $S_{yy}(f) \approx S_{xx}(f)|H(f)|^2$, it is unlikely that a double-talk situation has occurred. The noncausal Wiener filter in (1) is recognized as the coherence function $\gamma_{yx}(f)$ between the input signal $x(n)$ and $y(n)$

$$F(f) = \frac{|S_{yx}(f)|^2}{S_{xx}(f)S_{yy}(f)} = \gamma_{yx}(f). \tag{2}$$

Thus a suitable decision parameter, ξ, will be based on the coherence function, $\gamma_{yx}(f)$, in (2).

In Fig. 2(a), speech and noise spectra are shown where the talkers power ratio $\kappa = -25$ dB and the x-SNR is 25 dB. The spectra are generated by 6th order AR-models fitted to correlation functions estimated from 320 ms long speech segments. The $x(n)$ signal is assumed to pass a hybrid with an attenuation of 20 dB. The coherence functions in this situation are seen in Fig. 2(b) for $\kappa = -\infty, -25, -15$ dB and they are always in the interval zero to one. The example shows that $\gamma_{yx}(f)$ is sensitive to double-talk especially in the frequency interval (300–1800 Hz), where the mean value of the speech power is large. The detection problem can now be formulated by forming a decision parameter of the coherence function. One parameter, which can be used for hypothesis testing is

$$\xi = \frac{1}{I+1}\sum_{i=0}^{I}\gamma_{yx}(f_i). \tag{3}$$

In (3), $f_i = f_1 + i(f_2 - f_1)/I$ and (f_1, f_2) are the endpoints of the frequency interval, in which the coherence function is of interest. The detection problem will be formulated as a hypothesis test with the two hypotheses H_0: no double-talk and H_1: double-talk. The decision rule will be: $\xi \geq T$ then H_0 (double-talk is not present), $\xi < T$ then H_1 (double-talk is present), where T is the threshold level. The parameter ξ, will be a random variable (rv) with a probability density function (pdf) which has to be calculated. Due to the randomness of ξ, the decision rule will give errors. Most important is a low probability of miss, P_M, i.e., a double-talk interval is not detected. The probability of false alarm P_F, i.e., double-talk is erroneously decided, is of minor importance. It influences, however, the adaptation of the canceller and thus, it should be kept small. An advantage of using the coherence as a decision function for double-talk is that it does not only decide whether double-talk is present or not, but also indicates the quality of the echo-path, i.e., if it is worthwile using the samples for estimation of other parameters. This property is desirable when the DTD is operated in an echo canceller system.

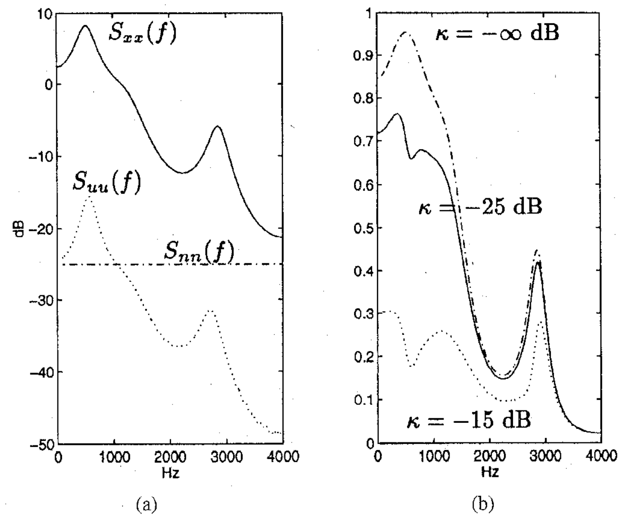

Fig. 2. (a) Spectra of x (solid line), u (dotted line) talkers, and system noise (dashed-dotted line); (b) Coherence function when the u-talker to x-talker ratios are $\kappa = -\infty$ (dashed-dotted line), -25 (solid line), -15 (dotted line) dB.

III. Estimation of the Coherence Function

The goal of this section is to derive an estimation algorithm for the coherence function in (2). Available signals are $x(n)$ and $y(n)$, which are both nonstationary. The estimation is made in the frequency-domain and relies on block-processing. A block of input data is processed at a time giving a block of output data. So the signals are segmented by a sliding window of length L. Each new block of data contains typically 5 to 10 new sample values compared to the preceding block, which gives a bound of the time resolution.

In a first step, the power density spectra $S_{xx}(f)$ and $S_{yy}(f)$ and the cross spectrum $S_{yx}(f)$ are estimated for a block. In a second step, an estimate of the coherence function is calculated according to (2). The periodogram is one estimate of spectral functions. Though simple to calculate by the FFT, it has the drawback of having a large variance and the estimated coherence function will always become one when only one block of data is used. An averaged periodogram depends on subdivision of the segment implying a decrease in variance but also a decrease in frequency resolution. In order to obtain an estimate with low variance and an acceptable frequency resolution, a multiple window spectrum estimation method will be used. Then the coherence function can be estimated from one block of data since an averaging occurs using the different windows. The multiple window power spectrum estimation method is based on the Cramér spectral representation of a stationary process

$$x(n) = \int_{-1/2}^{1/2} e^{j2\pi fn}dX(f), \qquad -\infty < n < \infty. \tag{4}$$

The increment process $dX(f)$ has the properties

$$E\{dX(f)\} = 0; \quad E\{|dX(f)|^2\} = S_{xx}(f)df \tag{5}$$

when $x(n)$ does not contain any periodic components. The Fourier transform of the N-point data sequence $x(n),\ 0 \leq n \leq N-1$, is

$$X_N(f) = \sum_{n=0}^{N-1} x(n)e^{-j2\pi fn}. \tag{6}$$

By inserting (4) into (6), $X_N(f)$ can be rewritten as

$$X_N(f) = \int_{-1/2}^{1/2} K_N(f-\nu)dX(\nu) \tag{7}$$

where $K_N(f)$ is the Dirichlet kernel. Equation (7) is a Fredholm integral equation of the first kind [10]. The solution of the integral (7), is equivalent to an eigenvalue problem, where the eigenvalues are denoted by λ_k, and the eigenfunctions by $\Phi_k(f) = \sum_{n=0}^{N-1} \phi_k(n)e^{-j2\pi fn}, 0 \leq k \leq K-1$. The eigenfunctions are local in a frequency interval of width $2W$, which has been predetermined, [9]. The multiple window estimates of the spectrum and cross spectrum are given by

$$\begin{aligned}\hat{S}_{xx}(f) &= \frac{1}{K}\sum_{k=0}^{K-1} c_k|X_k(f)|^2 \\ \hat{S}_{yx}(f) &= \frac{1}{K}\sum_{k=0}^{K-1} c_k Y_k(f)X_k^*(f)\end{aligned} \tag{8}$$

where the weights $c_k = 1/\lambda_k,\ 0 \leq k \leq K-1$, are close to one, and the k:th eigenspectrum, $X_k(f)$, is

$$X_k(f) = \sum_{n=0}^{N-1} x(n)\phi_k(n)e^{-j2\pi fn}. \tag{9}$$

Thus, the estimates are the average of the eigenspectra which have been obtained by using different data windows. These windows are the eigenvectors belonging to the K largest eigenvalues of an eigenvalue problem. The number of eigenspectra is $K \leq 2NW$ where $2W$ is the analysis bandwidth. Only a few eigenspectra are averaged when the length of the data record is short. By inserting the spectra, (8), in (2) the estimated coherence function is found. Equation (3) then gives us the decision parameter ξ.

Examples of the estimated coherence function are presented in Fig. 3. A segment of 128 speech samples and eight windows with a resolution of 640 Hz is used in the multiple window method. Fig. 3(a) shows the result when the signal $x(n)$ is passed through a hybrid circuit with 6 dB attenuation and a system noise (SNR = 25 dB) is present, note the high coherence in the lower frequencies. In Fig. 3(b), a double-talk component is added where κ is 0 dB which results in a significant decrease in the coherence.

IV. Statistical Analysis

In the following the statistics of the decision parameter, ξ are derived. When the statistics are known the detection probability, $P_D = 1 - P_M$, of the DTD can be calculated for various speech and channel situations.

The analysis is made for a specific frequency, f, and the distribution of the estimated coherence function is formed from the distributions of the estimated spectra. One block of speech data is considered to be stationary and fairly well modeled by a gamma distribution [11]. The resulting vector originating from the windowed and Fourier transformed data can be considered normal distributed because of the symmetric shape of the gamma function and the filtering effect of the transform, [12]. In the multiple window spectrum technique, (8), the sum contains squared normal distributions and thus, each term is χ_2^2. The eigenspectra in (9) are approximately orthogonal for a signal with arbitrary spectrum, i.e.,

$$E\{X_k(f)X_j^*(f)\} \approx 0, \quad k \neq j \tag{10}$$

where k and j define different windows. Due to these facts the distributions of the estimated spectra are χ_{2K}^2 [9].

The multiple window estimate of the cross spectrum is approximately given by

$$\hat{S}_{yx}(f) \approx \hat{S}_{xx}(f)H(f). \tag{11}$$

Then the estimated coherence function in can be rewritten as

$$\hat{\gamma}_{yx}(f) = \left[1 + \frac{\hat{S}_{vv}(f)}{\hat{S}_{xx}(f)|H(f)|^2}\right]^{-1} = [1 + C(f)z(f)]^{-1} \tag{12}$$

where

$$\begin{aligned}C(f) &= \frac{E\{\hat{S}_{vv}(f)\}}{E\{\hat{S}_{xx}(f)|H(f)|^2\}} \\ z(f) &= \frac{\hat{S}_{vv}(f)/E\{\hat{S}_{vv}(f)\}}{\hat{S}_{xx}(f)|H(f)|^2/E\{\hat{S}_{xx}(f)|H(f)|^2\}}.\end{aligned} \tag{13}$$

The quotient, $z(f)$, of the two χ_{2K}^2 distributed spectra is F-distributed and the distribution of the coherence will due to this fact be of transformed F-character. The pdf of $z(f)$ is denoted $p_{z(f)}(z(f))$. The degrees of freedom are equal in the

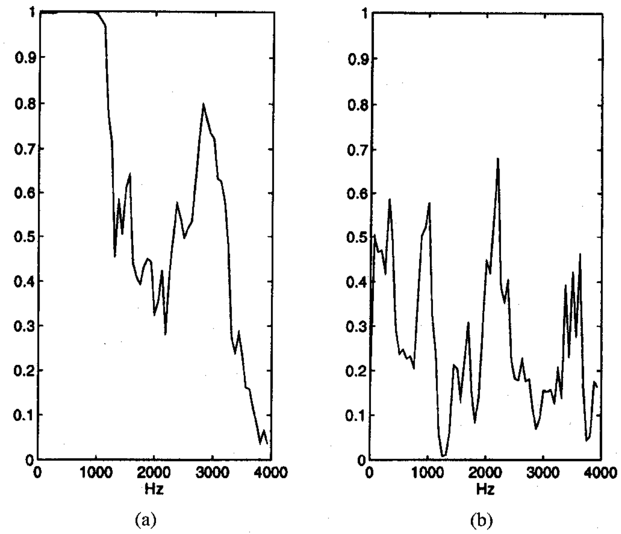

Fig. 3. (a) Estimated coherence function when only x-talker and system noise are present; (b) estimated coherence function during double-talk. Hybrid attenuation is 6 dB, the SNR is 25 dB and the u-talker to x-talker ratio κ = 0 dB.

numerator and denominator of z, so the pdf is (dependence on frequency is from now on omitted)

$$p_z(z) = \frac{\Gamma(2K)}{\Gamma(K)^2} z^{K-1}(1+z)^{-2K}. \tag{14}$$

Since the transform of z is monotonous, $\gamma_{yx}(Cz)$, the probability density of $\hat{\gamma}_{yx}$ is given by

$$\begin{aligned} p_\gamma(\gamma) &= \frac{1}{\gamma^2 C} p_z\left(\frac{1}{C}\left(\frac{1}{\gamma} - 1\right)\right) \\ &= \frac{\Gamma(2K)}{\Gamma(K)^2} C^K (\gamma - \gamma^2)^{K-1}(1 + (C-1)\gamma)^{-2K}. \end{aligned} \tag{15}$$

The decision parameter is the mean value of the coherence over a number of different frequencies according to (3). If the frequency spacing between the coherence samples is sufficiently wide, i.e., $|f_k - f_j| > 2W,\ k \neq j$, the samples are independent and the pdf of ξ is found by convolving the function in (15) $I+1$ times

$$p_\xi(\xi) = p_{\gamma(f_0)} \star p_{\gamma(f_1)} \star \cdots \star p_{\gamma(f_I)} \tag{16}$$

where $\star$ denotes convolution. The binary test for detecting double-talk takes the form

$$\begin{aligned} &H_0\colon \xi \geq T, \text{ no double-talk} \\ &H_1\colon \xi < T, \text{ double-talk.} \end{aligned} \tag{17}$$

The double-talk detection probability and false alarm probability (no double-talk present) are calculated by integration of (16)

$$P_D = \int_0^T p_\xi(\xi \mid H_1)d\xi, \quad P_{FA} = \int_0^T p_\xi(\xi \mid H_0)d\xi. \tag{18}$$

The function $p_\xi(\xi|H_1)$ in (18) is computed for a specific power ratio between the near and far-end signals. When the power ratio κ is large, the result will be a large probability of detection, P_D.

The spectra of the two speech signals used in the following example are shown in Fig. 2(a). Fig. 4 shows the pdf of ξ, $I = 2$, $W = 0.04$, and $K = 8$, with (solid line) and without (dotted line) double-talk. The hybrid attenuation is 20 dB, $\kappa = -25$ dB, and the SNR is 25 dB. The receiver operating characteristic is shown in Fig. 5(a). It indicates that if a false alarm probability of 0.01 is allowed, we detect double-talk with a probability of almost one in the specified situation in Fig. 2(a).

If the threshold, T, is chosen so that the detector has a false alarm probability of 0.01 the detection probability versus κ is shown in Fig. 5(b). The analyses in this section show the detector's theoretical performance, which is somewhat degraded in practice, but give a good indication of how the decision parameter should be chosen. The impairments can, among other things, be attributed to nonzero crossterms between $x(n)$ and $u(n)$ in the cross spectrum estimate and deviations from the distribution assumptions.

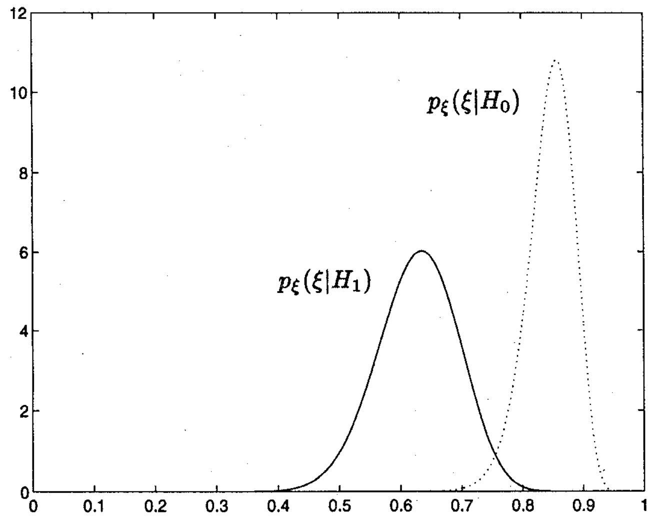

Fig. 4. PDF's of ξ under H_0 (dotted line) and H_1 (solid line).

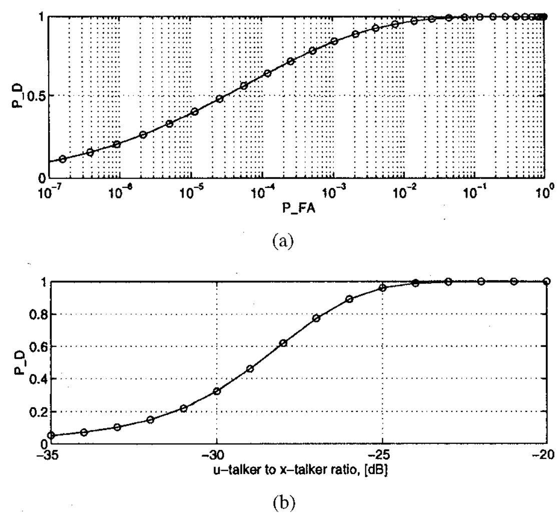

Fig. 5. (a) Theoretical ROC of the coherence DTD; (b) detection probability versus κ when the false alarm probability is 0.01.

V. Simulations and Examples Using Speech Data

In this section, examples of the performance of the proposed DTD are shown. As a reference, a standard power comparing DTD, [8], is used. Table I presents the situations in which we have tested the two DTD's. Three different hybrid attenuations are used, $\alpha = 0, 6, 20$ dB, the u-talker to x-talker ratios are $\kappa = 0, -6, -20$ dB, and the SNR is 40 dB. Cases 1, 2, 3a and b in Table I are situations described more thoroughly in this section. The remaining cases are commented by using the following grades: "Good"—practically no false alarm and misses, correct on-set time decision. "Fair"—a few false alarms and/or misses, somewhat delayed on-set decision. "Poor"—many false alarms and/or misses.

To calculate the spectrum estimates, we use eight windows with $W = 0.04$ and the number of FFT-samples is 128. The decision parameter is calculated from the FFT bins 6, 16, and 26, i.e., $f = 312, 937$, and 1562 Hz so that the

TABLE I
THE HYBRID ATTENUATION (α) AND u-TALKER TO x-TALKER RATIOS (κ) TESTED FOR THE TWO DTD'S. CASES 1, 2, 3A, B ARE THOROUGHLY DISCUSSED IN THE TEXT, AND THE OTHER ARE GRADED IN THREE LEVELS; SEE TEXT FOR THE INTERPRETATION OF "GOOD" ETC. THE GRADE FOR THE COHERENCE DETECTOR IS SHOWN TO THE LEFT AND THE STANDARD DTD TO THE RIGHT IN EACH TEST CASE

α [dB] / κ	0	6	20
0	Case 1	Good/Fair	Case 2
-6	Fair/Poor	Good/Poor	Good/Poor
-20	Fair/Poor	Fair/Poor	Cases 3a, b

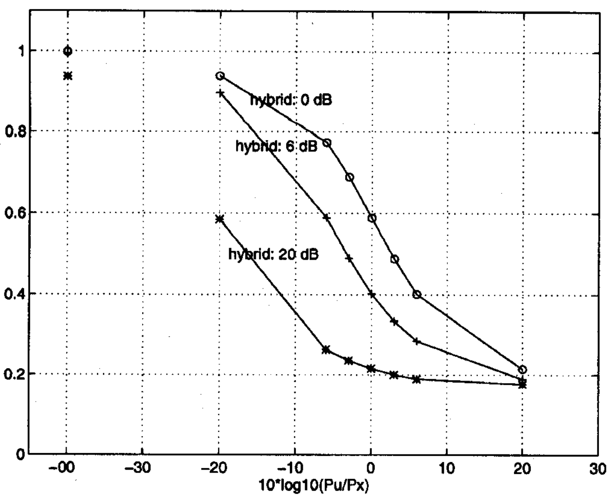

Fig. 6. The mean value of the decision parameter for 20 speech realizations versus different hybrid attenuations and u-talker to x-talker ratios. The SNR is 40 dB.

coherence estimates at these frequencies can be assumed independent.

In Fig. 6, the behavior of the decision parameter in (3) is examplified by a simulation with different hybrid attenuations and different κ. To see the performance, the mean of the decision parameter of 20 realizations is plotted. The left-most point shows the mean value when no double-talk is present and the curves show how the decision parameter changes for larger κ. The parameter without double-talk is sensitive to worse SNR when the attenuation in the hybrid is high (20 dB). For large κ (20 dB), the decision parameter should theoretically be smaller than in Fig. 6. The reason for the fairly large value is the nonzero crossterms between x and u in the crosspectrum estimate. These crossterms do not interfere so much for smaller κ.

The standard DTD assumes a hybrid attenuation, $\alpha = 6$ dB and is described by, [13], [14]

$$\begin{aligned} y_p(n+1) &= (1-\tau)y_p(n) + \tau|y(n)| \\ x_p(n+1) &= (1-\tau)x_p(n) + \tau|x(n)| \end{aligned} \tag{19}$$

where the filtergain τ is 2^{-5} and N is 128. Double-talk is declared if

$$y_p(n) \geq \frac{1}{2}\max\{x_p(n),\, x_p(n-1),\ldots,x_p(n-N)\}. \tag{20}$$

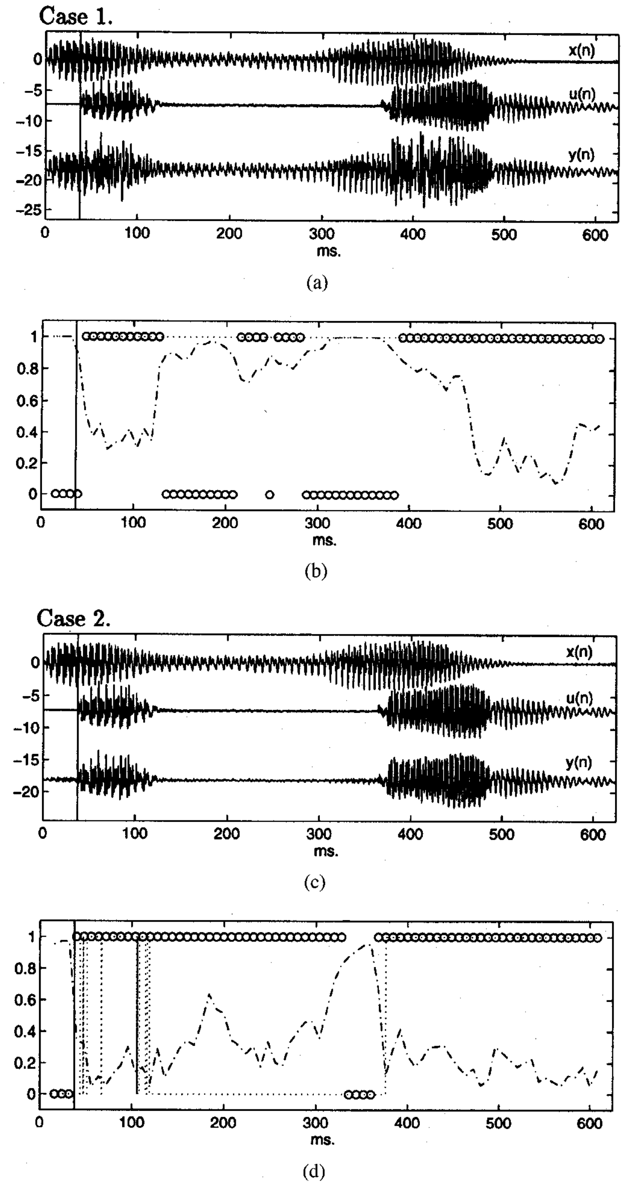

Fig. 7. (a) Far- and near-end speech signals, $\kappa = 0$ dB, $\alpha = 0$ dB; (b) results from detection. The vertical solid line = beginning of double-talk, Dashed-dotted line = value of decision parameter of coherence DTD. Circles: Decision of coherence DTD, "1" = double-talk, "0" = no double-talk. Dotted line = Decision of standard DTD; (c) far and near-end speech signals, $\kappa = 0$ dB, $\alpha = 20$ dB; (d) results from detection of the situation in (c).

Examples used for presenting the detectors are the two names, "Bertil" and "Sune," pronounced in Swedish. These words are normalized in power and merged together with different κ ratios and hybrid attenuations to generate the test situations. In Fig. 7(a) the far-end signal, $x(n)$, "Sune" and the near-end signal, $u(n)$, "Bertil" are shown. Below these two, the measured $y(n)$ signal is shown, containing a hybrid attenuated version of $x(n)$ added to $u(n)$ according to Case 1 ($\kappa = 0$ dB and $\alpha = 0$ dB) in Table I. An appropriate level of white Gaussian noise is added giving an x-SNR of 40 dB. Fig. 7(b) shows the decisions made by our DTD, marked as circles, and the standard DTD, dotted line, where the double-talk

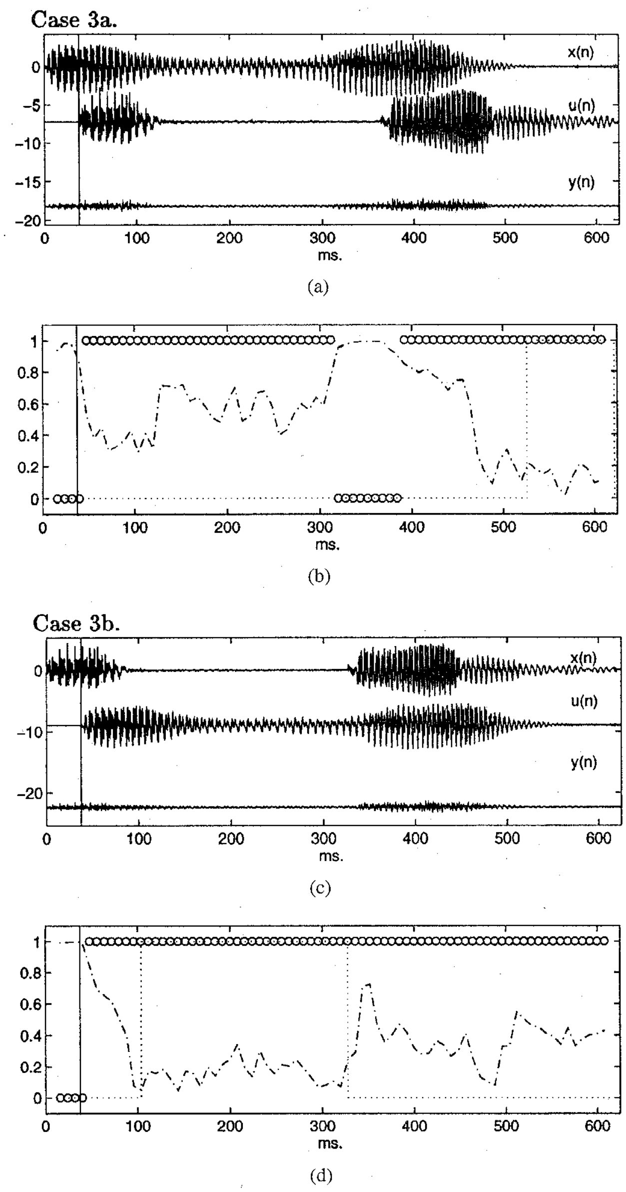

Fig. 8. (a) Far and near-end speech signals, $\kappa = -20$ dB, $\alpha = 20$ dB; (b) results from detection. The vertical solid line = beginning of double-talk. Dashed-dotted line = value of decision parameter of coherence DTD. Circles: Decision of coherence DTD, "1" = double-talk, "0" = no double-talk. Dotted line = Decision of standard DTD; (c) far and near-end speech signals, $\kappa = -20$ dB, $\alpha = 20$ dB. Note that $x(n)$ and $u(n)$ of (a) are exchanged; (d) results from detection of the situation in (c).

decision is represented by "1." The actual value of the decision parameter of our DTD is also shown as a dashed-dotted line and the detection threshold is $T = 0.86$. A new decision is made every 64 samples in the coherence detector and every sample in the standard detector. Since the hybrid attenuation is 0 dB, the standard DTD will continuously indicate double-talk. The coherence based DTD indicates double-talk very close to the start of the near-end word "Bertil." As long as there is a significant level of near-end signal, we detect double-talk but in the low level region of $u(n)$ (300–350 ms) the connection is declared double-talk free.

Fig. 7(c) and (d) show the results for Case 2 where there is a high attenuation in the hybrid, $\alpha = 20$ dB and $\kappa = 0$ dB. The starting point of double-talk is efficiently detected by the coherence detector but also by the standard DTD, although somewhat delayed. In the low level region of $u(n)$, the standard DTD does not detect double-talk while the coherence detector does. In this situation, one normally allows adaptation of an echo canceller but careful examination of $u(n)$ and $x_1(n)$ in this region shows that the echo and near-end talk are of equal amplitude and therefore, divergence of the canceller would appear. The described situation shows that the coherence detector is powerful for deciding the quality of the signals.

The next example, Case 3a in Fig. 8(a) and (b), shows the performance when the levels of the speakers are different. The u-talker to x-talker ratio, κ, is -20 dB and the hybrid attenuation $\alpha = 20$ dB. The standard detector completely fails to detect double-talk when the far-end speech is large while the coherence detector works without any degraded performance.

In Fig. 8(c) and (d), the situation is the same as in Fig. 8(a) and (b) but the signals are exchanged, i.e., "Bertil" is used as $x(n)$ and "Sune" is used as $u(n)$. The standard DTD is only able to detect double-talk in the low level region of "Bertil" (100–320 ms). The coherence detector makes correct decisions in all levels of "Bertil" which is important if divergence of the echo canceller should be avoided.

VI. Conclusion

We have presented a new method for detecting double-talk in a full duplex transmission line. The performance of the double-talk detector has been evaluated theoretically and examples of measured speech situations have been presented. The use of the estimated coherence function results in a more powerful DTD than level comparing detectors. The examples shown in Section V indicate the new DTD's ability to classify the quality of the speech situation, i.e., when the disturbance in the near-end is low compared to the echo. This property allows a more accurate control of the adaptation of an echo canceller compared to relying on a decision based on the level between far-end and near-end signals. Simulations have shown that the proposed DTD can handle wider ranges of signal levels and hybrid attenuations since it does not depend on the absolute values of the levels of signals but on the correlation between the signals before and after the hybrid.

Acknowledgment

The authors would like to thank B. Wide of Telia Network Services AB for his initiation and encouragement of the project that led to the presented paper.

References

[1] S. M. Kuo and Z. Pan, "Distributed acoustic echo cancellation system with double-talk detector," *J. Acoust. Soc. Am.*, vol. 94, pp. 3057–3060, Dec. 1993.

[2] ——, "An acoustic echo canceller adaptable during double-talk periods using two microphones," *Acoustics Lett.*, vol. 15, no. 9, pp. 175–179, 1992.

[3] J. Prado and E. Moulines, "Frequency-domain adaptive filtering with applications to acoustic echo cancellation," *Ann. Télécomun.*, vol. 49, nos. 7–8, pp. 1–15, 1994.
[4] H. Ye and B. X. Wu, "A new double-talk detection algorithm based on the orthogonality theorem," *IEEE Trans. Comm.*, vol. 39, no. 11, pp. 1542–1545, Nov 1991.
[5] S. Kawabe, J. Chao, and S. Tsujii, "A new iir adaptive echo canceller: Give," in *Proc. IEEE Int. Conf. Syst. Eng. IEEE*, 1992, pp. 547–551.
[6] T. Huhn and H.-J. Jentschel, "Kombination von gerauschreduction und echokompensation beim freispechen," *Nachrichtentech. Elektron.*, vol. 43, no. 6, pp. 274–280, 1993.
[7] C. R. Johnson Jr., Z. Ding, and W. A. Sethares, "Frequency-dependent bursting in adaptive echo cancellation and its prevention using double-talk detectors," *Int. J. Adapt. Cont. Signal Processing*, vol. 4, pp. 219–236, 1990.
[8] C. R. Johnson Jr., "On the interaction of adaptive filtering, identification, and control," *IEEE Signal Processing Mag.*, vol. 12, no. 2, pp. 22–37, 1995.
[9] D. J. Thomson, "An overview of multiple-window and quadratic-inverse spectrum estimation methods," in *Proc. of ICASSP IEEE*, 1994, pp. VI-185–VI-194.
[10] G. M. Wing, *A Primer on Integral Equations of the First Kind.* SIAM, 1991.
[11] L. Richards, "Statistical properties of speech signals," *Proc. IEE*, vol. 111, no. 5, pp. 941–949, May 1964.
[12] C. L. Mallows, "Linear processes are nearly Gaussian," *J. Appl. Prob.*, vol. 4, pp. 417–454, 1967.
[13] D. L. Duttweiler, "A twelve-channel digital echo canceler," *IEEE Trans. Commun.*, vol. 26, no. 5, pp. 647–653, May 1978.
[14] M. M. Sondhi, "An adaptive echo canceller," *The Bell Syst. Tech. J.*, vol. 46, no. 3, pp. 497–510, Mar. 1967.