Double Talk Detection in hands-free mobile communication- A comprehensive survey

S Hannah Pauline, Dhanalakshmi Samiappan*, and R B Kumar
Department of Electronics and Communication Engineering, College of Engineering and Technology, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamil Nadu, India
Email: *dhanalas@srmist.edu.in

Abstract. In modern teleconferencing systems, echo cancellation is a mandatory requirement. Echoes arise in Loudspeaker Enclosure Microphone devices because of coupling between loudspeaker and microphone signals. To remove these unwanted echoes from microphone signals, Acoustic Echo Cancellers are used. Several techniques exist for acoustic echo cancellation. Acoustic echo cancelling an adaptive FIR/IIR filter can predict the path of echo and cancel it. The adaptive algorithms give good performance when only the far-end signal is present. But when the near-end signal is present, the adaptive filter tends to diverge. Such a situation creates a double talk scenario. The filter adaptation must be prevented during the double talk period. Accurate detection of double talk is quite difficult and computationally complex. Hence the algorithms for detecting Double Talk are not easy to implement. Increased detection accuracy is achieved at the cost of computational complexity. Further, it is desired to cancel out the far end echo signal, thus giving a good quality of near-end speech signal as output. Here, we provide a detailed study of the different algorithms existing for double talk detection, and future direction of research in the field of double talk detection is proposed.

Keywords: adaptive filters, dual, distant-end signal, Loudspeaker Enclosure Microphone, close-end signal.

1. Introduction
Remote & hands-free access devices, the quality of speech are the most important criteria. However, acoustic echo degrades the speech signal quality. Acoustic echo is created when a coupling is provided between a loudspeaker and microphone in a LEM device. Acoustic Echo Canceller (AEC) is employed to cancel out these undesired echoes and provide a good quality of speech.

1.1 Acoustic Echo Cancellation
Several researchers have conducted extensive research to cancel out the acoustic echo. Most of these algorithms are based on Least Mean Square (LMS) and Normalised Least Mean Square (NLMS) algorithm, even though other methods have also been used.

The basic function of Acoustic Echo Canceller is:
(i) To identify the echo path, this is not known.
(ii) Generate a replica of the echo signal.
(iii) To subtract the echo signal from the total received signal.

Figure 1: Acoustic Echo Canceller

It uses an adaptive filter that should have good tracking and convergence properties and effectively cancel the echo signal. The adaptive filter converges as long as there is no speaker at the near end. Figure 1 shows an Acoustic Echo Canceller.

1.2 Double Talk Detection
In situations where the far-end signal is present and the near end signals, the adaptive filter instead of converging trends become divergent. Such a scenario where speakers are active on both ends of the conversation is known as Double Talk. During the Double Talk period, the filter should be notified to halt its adaptation. This function can be performed by a Double Talk Detector (DTD). Thus the Double Talk Detector should detect that a near-end speaker is present and notify the adaptive filter to halt its operation. This function should be performed within a few milliseconds failing which the adaptive filter will diverge.

Figure 2 shows the Dual Detector. The DTD output is provided to the adaptive filter that either continues to adapt or halts its operation depending on Double Talk’s absence or presence. The procedure is as follows:

(i) To determine a detection statistic ‘β’ calculated from the input signal, desired output signal, and error signal.
(ii) To compute the decision threshold parameter ‘τ,’ to determine whether double talk has occurred. Therefore, Double Talk Detector’s function is to detect a near-end signal and freeze the adaptation of the filter. Double Talk Detection's various approaches can be widely categorized as time domain, and frequency domain approaches. The most popular are based on energy, correlation, and coherence. Few methods are also based on voice activity detection and extraction of speech features. Apart from the above stated, there are other solutions to detect double talk scenario in Acoustic Echo Cancellation. Section II will provide an extensive review of the different techniques adapted for Double Talk Detection while highlighting each
method's merits and demerits. Further section III will provide the conclusion and future scope for researches in the field of DTD.

**Figure 2: Double Talk Detector**

### 2. Double Talk Detection Methods

#### 2.1 Energy-based DTD

Geigel algorithm [1] is formulated by comparing the amplitude of far end signal and near end signal by using a decision statistic ‘$\beta$’

$$
\beta = \frac{\max |s(n) - s(n-N)|}{|k(n)|}
$$

Where $s(n)$ is the signal from far, and $k(n)$ is the signal near. A double talk scenario can be declared if the maximum signal from the near end exceeds the signal from the far end over a while $L$. This method is simple to compute and has less memory requirement. However, the echo path estimate is not accurate; it does not perform well in a noisy environment. For quick detection of Double Talk and distinguishing varying gain of the channel from Double Talk, two new algorithms are proposed [2] based on the power of a near-end signal (i.e.) on the adaptive gain constant gain methods. These perform better than the conventional Geigel algorithm. An extension to the Geigel algorithm is proposed in [3] using holder inequality. In this method, three decision statistics have been formulated. The variance of the noise is included in computing the decision statistics. The threshold values depend on the impulse response of the path of echo. Choosing the threshold is crucial. This algorithm performs similar to Geigel in a low noise environment, but in a high noise environment where the Geigel algorithm fails to perform, this method performs well. This algorithm is computationally simple and performs well in a noisy environment. However, selecting the appropriate threshold is a challenge. Table 1 provides a comparison between Geigel and holder inequality algorithms.

| Adaptive Filter | Loudspeaker | Acoustic Echo |
|-----------------|-------------|---------------|
| Far end signal  | Microphone  | Near end signal |
| $\sum$         | Double Talk Detector |

**Table 1. Comparison of decision statistics of Geigel algorithm vs Holder inequality based algorithm**
Algorithm | Decision statistic
--- | ---
Geigel | $\eta = T_0|\text{x}(n)|_\infty$

Holder Inequality | $\eta_1 = T_\infty|\text{x}(n)|_1 + \sigma_v$
$\eta_2 = T_1|\text{x}(n)|_\infty + \sigma_v$
$\eta_3 = T_2|\text{x}(n)|_2 + \sigma_v$

where

- $\sigma_v^2$ is the variance of noise
- $h(n)$ is the impulse response of echo path
- $T_\infty$, $T_1$, $T_2$ are thresholds that depend on $h(n)$

Table 2. Comparison of Misalignment of Geigel algorithm vs. Holder inequality based algorithm at ENR=0 dB

| Algorithm          | Time | Misalignment |
|--------------------|------|--------------|
| Geigel             | 2 seconds | -5 to -4 dB |
|                    | 6 seconds | -8 to -9 dB |
| Holder Inequality  | 2 seconds | -10 to -9 dB |
|                    | 6 seconds | -10 to -9 dB |

Table 2 proves that holder inequality-based DTD performs better than the Geigel method. The detection statistic computed in the Geigel algorithm is not static since it is formulated by comparing the samples' absolute values. Instead of the signal energy [4] is used for computing the decision parameter, accurate results can be obtained. The procedure used to detect double talk using signal energy is by first computing the signal envelope by measuring the energy of signals from far, near, and echo estimates. A decision function is then computed using the measured signal envelope, which detects double talk presence compared to the threshold. To improve dynamic accuracy threshold is computed. The accuracy is better than the Giegel algorithm but is computationally complex. Also, accurate estimation of echo is required to reduce the detection errors.

Table 3. Comparison of results of Geigel algorithm vs. signal envelope based algorithm

| Algorithm         | SNR   | ERLE(dB) | PESQ Score |
|-------------------|-------|----------|------------|
| Geigel algorithm  | 44.59 | 8.38     | 3.63       |
| Signal envelope based | 68.12 | 32.26    | 3.71       |

Table 3 proves that signal envelope-based DTD performs better than the Geigel method. To make the DTD compatible with the existing embedded systems, the major requirement is that computational complexity should be reduced. Signal envelope-based Double Talk Detection [5] methods and other easy-to-compute values will result in computationally simpler DTD. The signal envelope rise and fall contain information about the increase in energy level. Hence this measure is sufficient to detect Double Talk. Thus accuracy can be improved with reduced computations.

2.2 Cross Correlation-based DTD

Several cross correlation-based methods exist in the literature. The general method is to compute a cross-correlation function between the signal received from the far and near ends. The double talk occurrence is then predicted by comparing the calculated correlation function and a pre-set threshold. The earliest double talk detector using the cross-correlation function was based on orthogonality [6] theorem (i.e.), the cross-
correlation between input and error signal. Depending on the correlation function, double talk is detected, which then stops the adaptation of the filter. The storage requirements are high due to its computation complexity. In normalized cross correlation-based DTD [7], the decision variable is computed based on Normalized Cross-Correlation (NCC) between input and output signal vectors. In this, the echo path variations do not affect the decision variable. The same method can be extended to derive a cross-correlation matrix.

Further performance improvement can be obtained by combining with the coherence function. The major disadvantage of the above method is computational complexity. Also, it does not perform well in a dynamic environment and the presence of noise. In the frequency domain [8], the test statistic is computed for a frequency domain adaptive algorithm using a Normalized Cross-Correlation vector. A pseudo coherence vector is used to formulate the new decision statistic. Results are compared in terms of ROC and found to be better than the Geigel algorithm. Single-channel Acoustic echo Cancellation DTD algorithms based on NCC can be generalized to a multichannel situation [9] as well. This is possible by merging the data of all microphone signals into a single decision statistic. The same can be formulated in the frequency domain.

Double talk can be detected by computing a new decision statistic from error signal and microphone signal [10].

Further, a hybrid DTD can be used where the double talk is detected by using 2 detectors. Real-Time Recurrent Learning performs speech detection and discrimination. The changes in the echo path can also be tracked.

| Algorithm                                  | Probability of miss for different NFR |
|--------------------------------------------|---------------------------------------|
|                                            | NFR=-10                               | NFR=+10                               |
| Orthogonality based DTD                    | 0.8                                   | 0.3 to 0.4                             |
| NCC between error and near-end signal based DTD | 0.1 to 0.2                     | 0                                      |

Table 4 proves that NCC between error and near-end signal-based DTD performs better than orthogonality-based DTD. The major drawbacks in DTD based on NCC between input and error signals are that in a noisy environment, errors may occur in detecting Double Talk during the initial convergence period. During periods of single talk, the power of the error signal fluctuates. Therefore it is required to normalize the error signal power [11] and smooth it to reduce error in detecting double talk. In non-linear AEC [12], double talk is detected from the cross-correlation between received and the signal from the far end. The adaptive algorithm employed is Variable Step Size-LMS. Performance in ERLE is better than the Geigel algorithm; to improve this algorithm, the dynamic threshold can be formulated. The most recent research on detecting double talk using cross-correlation is developed using both energies of the signal and an estimate of cross-correlation [13]. For estimation of energy, the envelope of the signals from far and near is used. The occurrence of Double Talk is decided by the information contained in the envelope. If this information fails, the cross-correlation estimate will determine accurate double talk presence. Errors in detecting Double Talk are reduced by using cross-correlation-based techniques, whereas the delay in detection is increased.
2.3 Coherence based DTD

The major problem with double talk detection using the cross-correlation function is that it does not exhibit robustness and high computational complexity. This can be overcome by computing the decision statistic using similarity between the signals from far and near, called the coherence function. The working principle of coherence-based DTD is that when coherence function is almost equal to 1, no double talk is present. In contrast, when the coherence function is almost equal to 0, double talk is detected. In conventional DTD [14], a spectrogram is used to estimate density spectra from signal blocks. Decision static is formulated by taking an average of estimated coherence function calculated in a specific frequency range. Even though signal levels of a wider range can be handled, double talk detection in the presence of high noise and microphone and loudspeaker are placed apart. To improve the precision of the conventional coherence-based double talk detector modified version, the same can be used [15]. In this method, a statistical model is built for the coherence function, and it is smoothed by a classifier and first-order HMM filters. Then, the probability of the presence of double talk per frame and bin is estimated. The results of the above stated conventional coherence based algorithm and modified coherence based algorithm are compared in terms of classification error function defined as

\[
\%E = \frac{\text{FN} + \text{FP}}{\text{TN}} \times 100
\]  

(2)

FN and FP denote the false negatives and positives, and TN denotes the total number of frames.

Table 5. Comparison of classification error of conventional Coherence based algorithm and new coherence based algorithm

| Algorithm                  | %E  | Number of false positives | Number of false negatives | Total frames |
|----------------------------|-----|---------------------------|---------------------------|--------------|
| Conventional coherence based DTD | 2.94 | 575                       | 511                       | 36920        |
| Soft decision coherence based DTD | 1.26 | 251                       | 214                       | 36920        |

Table 5 proves that Soft decision coherence-based DTD has less classification error compared to Conventional coherence-based DTD. Double talk can also be detected using partial coherence [16] in two microphones and one loudspeaker. The partial coherence is zero when the near-end signal is not present; otherwise, it is non-zero. One of the problems that arise with double talk detection using the cross-correlation method is that the set threshold may not discriminate variations in echo path from DT. Thus there will be errors in detecting double talk. This problem can be overcome by using an auxiliary filter [17], which can use the reference input signal and estimate the error signal. The presence of residual echo in the main filter causes filters to converge, whereas the filter does not converge in the presence of signals from the near end along with the error. Further, to track the echo path, even better dual auxiliary filters [18] can be used for DTD. In this case, ERLE is much improved. To estimate the exact transition between single and double talk, a robust double talk detector [19] with the following components can be used, voice detector for near signal and DTD to control the adaptive filter, two auxiliary filters. Divergence of AEC is prevented by saving estimates of the echo signal. ERLE is 30dB. Double Talk Detection can be made simpler and compatible with the available DSP processors and memory chips. A new auxiliary filter [20] structure can be used. This method calculates the number of maxima in speech descriptor and then, using this measure, detects single or double talk. A dynamic threshold can be used in future methods.
2.4 Voice Activity Detection (VAD) based DTD
The problem with cross correlation-based DTD is that if the echo is large, detection error occurs. Similarly, if the echo signal is small, detection error occurs in energy-based DTD methods. To overcome the above two problems, a novel DTD method is proposed that utilizes Voice activity for DTD.

This algorithm has the following steps [21]:
I. The high spectrum signals are cancelled using Hamming window.
II. The Voice endpoint of each speech is detected by individual voice activity detection. The power of each frame is calculated and passed through a moving average filter.

The discrimination between speech and noise or silence is done by using a floating threshold. VAD-based DTD method is robust to noise, has low mis-adjustment, and better accuracy. Statistical Voice Activity Detection based DTD [22] suggests that Global Near end Speech Presence Probability (GNSPP) along with Speech activity, recognizing close & distant end signals can be used to identify Double Talk.

To derive the GNSPP statistical model of the speech signal is used. To further improve ERLE and PESQ score, a new DTD is based on probability [23]. In this method, the Speech Presence Probability of near-end signal (NSPP) is estimated by Beta Mixture Model and Bayesian rule. This probability is used to modify the gain of Acoustic Echo suppression (AES). The AES is performed using Short Term Fourier Transform (STFT). The echo path response updating depends on the Double talk decision obtained from NSPP. Echo Return Loss Enhancement (ERLE) is given by

$$ERLE(n) = 10\log_{10}\left(\frac{E(y^2(n))}{E(e^2(n))}\right)$$  \hspace{1cm} (3)

Table 6. Comparison of ERLE and PESQ score of Stereophonic AES vs. Speech Presence Probability-based DTD

| Parameters          | Algorithm                                     | SNR 30dB | 20 dB | 10 dB |
|---------------------|------------------------------------------------|----------|-------|-------|
| ERLE (Single Talk)  | Stereophonic AES                              | 22.04    | 15.97 | 9.8   |
| ERLE (Double Talk)  | Speech Presence Probability-based DTD         | 26.52    | 20.35 | 14.02 |
|                     | Stereophonic AES                              | 6.05     | 5.80  | 5.10  |
|                     | Speech Presence Probability-based DTD         | 5.73     | 5.14  | 4.81  |

The comparison Table 6 proves that acoustic echo cancellation is better by using the above method. The near-end signal is retained well during the double talk period.

2.5 Double Talk Detector based on Mutual Information
The use of cost-effective loudspeakers in a Loudspeaker Enclosure Microphone (LEM) device causes non-linear distortions in the echo path. For a DTD to perform well in non-linearity, mutual information [24] based method is used. The mutual information between the input to the Acoustic Echo path and output from it is used to detect double talk. The decision statistic is computed using the mutual information between the loudspeaker and microphone signal. When these two signals are fully dependent, maximum mutual information value is obtained, and thus there is no Double talk. If the mutual information value is reduced, Double Talk is present. This method is used in monophonic systems. For stereophonic systems with non-linearity, Generalized Mutual Information (GMI) [25] should be computed between the input signal vector at the Acoustic Echo path and the echo path's output vector. This GMI is used to detect Double Talk.
2.6 DTD based on other frequency-domain methods
To enhance the performance of DTD, several frequency-domain methods have been proposed. In the Gaussian Mixture Model (GMM) [26], DTD suppression of acoustic echo in the frequency domain is first applied. The components which are used to compute cross-correlation are analyzed. Gaussian Mixture Model is applied to select the feature vector that discriminates between single and double talk. Its performance in terms of probability of error is good compared to time-domain methods.

Table 7. Comparison of time-domain vs. frequency-domain method for DTD in terms of probability of error (PE) for one echo path

| Noise type     | Method           | SNR(average of 10, 20 and 30) |
|----------------|------------------|-------------------------------|
| White noise    | Time-domain      | 31.88±0.10                    |
|                | Frequency domain | 17.97±0.09                    |
| Car noise      | Time-domain      | 29.66±0.10                    |
|                | Frequency domain | 18.05±0.09                    |
| Street noise   | Time-domain      | 29.96±0.11                    |
|                | Frequency domain | 19.87±0.09                    |

Table 7 proves that GMM based DTD performs better than time-domain methods. To enhance its performance, the feature that is used to detect double talk can still be improved. In almost all Double Talk Detection methods, the unknown echo path has to be estimated. Also, the accuracy of echo path estimation will impact the double-talk decision. In the spectral slit [27] method, the estimation of the echo path is not required. Here the far-end signal is deleted spectrally. This permits the spectral content of only near-end signal to be present during double talk scenario, thus making it easy to detect the near-end signal.

Table 8. Comparison of probability of miss (P_m) of Geigel algorithm, Signal envelope based algorithm, and NCC based algorithm with spectral slit algorithm.

| Algorithm                | Probability of miss(P_m) |
|--------------------------|--------------------------|
|                          | ENR=30dBi ENR=15dBi ENR=10dB |
|                          | EFR=0dB EFR=0dB EFR=0dB |
| Geigel algorithm         | 0.5 to 0.6 0.6 to 0.7 ≈ 0.7 |
| Signal Envelope based    | 0.3 to 0.4 ≈ 0.5 ≈ 0.5 |
| algorithm                |                          |
| NCC based algorithm      | 0 to 0.1 0.1 ≈0.18 |
| Spectral slit algorithm  | ≈ 0 0 to 0.1 ≈ 0.15 |

Table 8 proves that Spectral slit-based DTD is superior in performance compared to the others. The double talk problem in the multichannel teleconferencing system can be avoided by estimating the fundamental frequency [28]. Using the decorrelation approach, the far-end signal fundamental frequency can be removed. The fundamental frequency of the error signal can then be estimated using second-order adaptive notch filters. If the error signal contains a fundamental frequency, double talk is detected. Results are compared with cross correlation-based DTD in terms of P_m (probability of miss) and EER (Equal Error Rate).
Table 9. Comparison of probability of miss ($P_m$) and EER of Cross correlation-based DTD and fundamental frequency based DTD

| Algorithm                    | Probability of miss ($P_m$) for $P_f = 0.1$ | EER   |
|------------------------------|---------------------------------------------|-------|
| Cross correlation-based DTD  | $\approx 0.015$                             | 2.5%  |
| Fundamental frequency-based DTD | $\approx 0.01$                               | 3.3%  |

From Table 9, it is clear that Fundamental frequency-based DTD performs better than Cross correlation-based DTD in terms of $P_m$ but is less efficient in terms of EER. In frequency domain-based methods, the computational complexity is high.

2.7 Double Talk Detector based on Prediction Error Identification

Double Talk Detection methods work well if the signal from the far end is present for around 20 to 30 percentage of the time. But suppose the signal from the near end is continuously present, such as in noisy teleconferencing systems. In that case, the convergence of the adaptive filter may be affected during a single talk period. This problem may be overcome by using prediction error identification techniques. Conventional Prediction Error Identification [29] methods can be used to calculate echo direction prediction & auto-regressive close-end pulse model. Double talk is detected using the Recursive Error Prediction (RPE) identification algorithm. The above method and 3 other algorithms, namely PEM-AF, PEM-AFROW, and 2ch-AF, are applied to detect Double Talk. Using this method, the convergence is fast. With increased computations to improve the performance and reduce computational complexity, DTD based on Frequency Domain [30] adaptive filtering-Prediction Error Method-Adaptive Filtering using ROW operations is used. The close-end signal is averaged & the Better Linear Unbiased Echo Route Approximation is optimized to improve efficiency. The occurrence of dual talk is observed using the Instantaneous Pseudo Correlation value between the distant end signal & the mic signal.

2.8 Speech Features based DTD

The double Talk scenario can be detected by using the features of the speech signal. To sense the period of double talk, a clear distinction should be made between the far-end and near-end signals. In speech feature-based DTD [31], speech features such as energy and log energy, standard deviation, and maximum value are extracted from the signals from far and near the end. A decision statistic is computed for all iterations to detect the period of double talk. Voice Activity Detection module formulated from the near-end signal is used to compute a dynamic threshold. The presence of Double Talk is declared when the value of decision statistic exceeds the threshold. Otherwise, there is no Double Talk present. ERLE is improved, Misalignment is reduced, and the probability of miss is low.

Further improvements to discriminate between single and double talk can be made by combining the speech’s temporal and frequency features. Another future enhancement includes using voiced and unvoiced speech classification. It is necessary to use Double talk detection in low resource Digital Signal Processors with low power. Therefore the computation should be simple, and the detector should be easy to operate. To meet these conditions, a new DTD based on Zero Crossings [32] can be used. In this method, the Zero Crossings Rate of Acoustic Echo Canceller output is computed and is compared suitably computed threshold. If the Zero Crossing rate exceeds the threshold, Double Talk is absent. Otherwise, Double Talk exists. This method is very simple in computation since only sign comparison is required for computing the zero crossings. However, it is difficult to select the appropriate window size and threshold.
2.9 Double Talk Detector based on transforms

To further enhance Double talk Detectors’ performance in terms of convergence, quality of speech, and Misalignment, transform-based methods are used. In Singular Value Decomposition (SVD) based DTD [33], SVD of the far-end signal is used to detect the presence of double talk. Even though it is computationally more complex than cross-correlation-based methods, the performance is improved in terms of ERLE and misadjustment. For non-stationary signals, the time resolution can be improved by another efficient DTD algorithm based on Stockwell Transform [34], a combination of Short-Time Fourier transform & Wavelet Transform. In this method, the far end and near-end signals are divided into frames, and Hamming window is applied to these frames so that the nearest side lobe is minimized. Then, Stockwell Transform is applied, the mean value is calculated for the far-end signal frame, and the maximum value is calculated for the near-end signal frame. Using these metrics, the presence or absence of Double Talk is declared. The frequency threshold is computed based on the segmentation of speech signals as voiced/unvoiced or silent, where segmentation is performed using the energy of signal or zero crossing rates. With this method, the ERLE is high, Misalignment is less, filter coefficients converge fast, probability of miss is reduced, and more robust to noise.

| Algorithm | ERLE | PESQ | $P_m$ for $P_f=0.3$ and NFR =15 |
|-----------|------|------|-----------------------------|
| SVD based DTD | 6.2105 | 2.331 | 0.2 |
| Stockwell Transform based DTD | 7.3792 | 2.429 | 0.1 |

The Table 10 shows that DTD based on Stockwell Transform performs better than SVD based DTD.

2.10 Echo path variation discrimination and DTD

It is important to find whether there are changes in Double Talk’s echo path or presence [35]. The detection and discrimination can be performed using the likelihood function to compare the global and local channel models. The recursive Least Squares method is used for estimation. [36] Markov Modulated FIR filter can also be employed to discriminate between double talk and echo path changes. The expectation minimization algorithm is used to determine the maximum likelihood estimate of channel characteristics. Two algorithms, one online and one offline, have been formulated. The discrimination between echo path changes and double talk can also be made using an echo path energy level [37]. To improve the accuracy gradient of the energy level is used. The log-likelihood ratio test [38] can also be used to decide whether double talk has occurred or the echo path has changed. A gaussian stochastic model is applied to the input signal, and post-integration is used. It is required to know the power level of double Talk and background noise [39]. A generalized likelihood ratio test is derived and simplified to compute an optimum test statistic to obtain the same. It is found that the detection probability is almost unity.

2.11 Other Double Talk Detection methods

Apart from the ones discussed above, few other methods have also been proposed, such as using a multi-delay filter [40] and Kalman filter [41]. Other methods have been proposed using delay [42]. Watermarking techniques have also been applied to detect double talk [43]. One of the very recent methods is where echo canceller control logic has been formulated as a classification problem. It uses four different hypotheses [44] which will decide the presence or absence of double talk. For each hypothesis, the step size can be adapted and adjusted accurately [45]. Advanced Block Frequency domain-based DTD method is proposed. Further, efficient suppression of echo during a double talk is discussed.
3. Conclusion and Future scope

A comprehensive study of several algorithms for Double talk Detectors has been done in this paper. In general, a DTD performance can be measured in terms of parameters such as ERLE, PESQ score, Probability of Miss, Probability of False alarm, ENR, NFR, Misalignment, convergence speed, computational complexity, and classification error. Energy-based algorithms are the simplest class of DTD algorithms, but the major challenge is selecting thresholds. Also, since the echo path is an unknown accurate estimate of the echo path is difficult. Therefore the performance of these algorithms is not satisfactory in noisy environments. Signal envelope-based DTD has good accuracy, but the accurate estimate of echo is mandatory for the detector to perform well. Cross Correlation-based DTDs perform well compared to energy-based, but they are computationally complex. But their performance is affected by dynamic environment and large echo signal. A combination of energy and Cross Correlation-based DTD can provide a solution but produces more delay in detection. Coherence-based methods can overcome the difficulty in Correlation-based methods, but the background noise affects their performance. Auxiliary filters can improve the performance, but only using a dynamic threshold can give the best results. Voice activity detection and speech feature extraction-based algorithms perform well, but even they fail in noisy environments. Few frequency domain methods offer good solutions for noise but are computationally complex for low-cost Digital Signal Processors. The transform-based methods also face the same problem of computational complexity. Therefore, it is inferred from the above discussion that any of the methods could design a Double Talk Detection algorithm suggested or a combination of these methods to provide a more accurate double talk decision with reduced computation and enhanced speech quality.

References

[1]. S. Haykin, Adaptive Filter Theory, Prentice Hall Inc, New York, 1996.
[2]. H. Haaghshenas, M. MaalmirChegini, ‘A innovative dual detector based on an optimal estimate of close-end capacity, Proceedings of the 2007 IEEE International Conference Telecommunications and Malaysia International Conference on Communications (14-17 May 2007).
[3]. C. Paleologu, J. Benesty, T. Gensler, S. Cochin, ‘A class of double-talk detectors based on the holder inequality,’ Proc. ICASSP (2011) 425428.
[4]. G. Szwaoch, A. Czyawski, M. Kulesza, ‘A computationally efficient dual-talk detector based on a signal envelope, Elsevier, Signal Processing Vol.88(11) (2008) 28562862.
[5]. M. Shamim Hussain, M. Ajoad Hasan, A. B. M. Harun-Ur-Rashid, ‘Quick dual detection better signal boxes for the implementations of acoustic echo cancelation in embedded devices,’ Proceedings of the 4th International Conference on Advances in Electrical Engineering (ICAEE) (2017) 199204.
[6]. A. Ye, B. X. Wu, ‘A new double-talk detection algorithm based on the orthogonality theorem,’ IEEE Trans. Comm. 39 (1991) 15421545.
[7]. J. Benesty, D. R. Morgan, J. H. Chao, ‘A new generation of double-talk detectors focused on cross-correlation,’ IEEE Transactions on Speech Audio Processing 8 (2000) 168172.
[8]. T. Gensler, J. Benesty, ‘Wavelength dual-sensor via a structured inter vector, data Processing 82 (2002) 17821786.
[9]. J. Baenesty, T. Gansler, ‘Multichannel adaptive echo canceler double-talk sensor based on a structured inter matrix,, European transactions on telecommunications 13 (2002) 951101.
[10]. M. A. Iqbal, J. W. Stokes, S. L. Grant, ‘Normalized doubletalk detection based on microphone and AEC error cross-correlation,’ Proceedings of the IEEE International Conference on Multimedia and Expo (2007).
[11]. W. Heo, T. Kim, K. Bae, ‘Robust double-talk detection in the acoustic echo canceller using normalized error signal power,’ 9th International Symposium on Signal Processing and Its Applications (2007).

[12]. V. Das, A. Kar, M. Chandra, ‘A new cross correlation-based double talk detection algorithm for non-linear acoustic echo cancellation, TENCON IEEE Region 10 conference (2014).

[13]. A. M. Muzahid, K. M. R. Inggrid, S. I. M. M. RatoanMondaol, Y. Zhou, ‘Extensive dual identification algorithm focused on traditional signal power & inter estimation,’ 8th IEEE International Conference on Communication Software and Networks (2016) 303{306.

[14]. T. Gansler, C. Hansson, J. Ivarsson, G. Salomonson, ‘Dual detector dependent on coherence, IEEE Transactions on Communications 44 (1996) 1421{1427.

[15]. J. Taashev, ‘Dual talk sensor dependent coherence with soft judgment, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2012) 165{168.

[16]. R. Le BouquinJeans, G. Faucon, ‘A double talk detector based on the partial coherence function’, 9th European Signal Processing Conference (EUSIPCO 1998) (1998).

[17]. SihoKim, K. Ba., ‘Applying an auxiliary filter in the adaptive echo canceller for performance improvement of double-talk detection’, Proceedings of 2002 IEEE 10th Digital Signal Processing Workshop, 2002 and the 2nd Signal Processing Education Workshop. (2002).

[18]. ChanfeiZhuang, FeifeiJia, P. Wang, JingyuHua, ‘A double-talk detection based on dual-auxiliary filters of different taps,’ International Conference on Wireless Communications and Signal Processing (2009).

[19]. Lee, H. C. Huang, ‘A robust double-talk detector for acoustic echo cancellation, Proc. International Multiconference of Engineers and Computer Scientist II (2010).

[20]. M. Hamidi, A. Amrouache, ‘A new south wales for echo cancelation in a dual-speak situation utilizing an aux filter, 14th International Workshop on Acoustic Signal Enhancement (2014) 253{257.

[21]. H. Baoao, Y. Yaang, J. Liu, Q. Y. Baoe, ‘A powerful dual algorithm based on the recognition of speech behavior,’ International Conference on Audio Language and ImageProcessing (ICALIP) (2010) 12{15.

[22]. Y.-S. Park, J.-H. Song, S.-I. Kang, W. Lee, J.-H. Chang, ‘A statistical model-based double-talk detection is incorporating soft decision,’ Proceedings of ICASSP (2010) 5082{5085.

[23]. Y. Tong, YapingGu, ‘Acoustic echo suppression based on speech presence probability, IEEE International Conference on Digital Signal Processing(DSP) (2016).

[24]. Shi, X. Ma, G. T. Zhou, ‘A mutual information-based double talk detector for non-linear systems’, Proc. IEEE Conference on InformationSciences and Systems (2008).

[25]. Shi, X. Ma, G. T. Zhou, ‘A double-talk detector based on generalized mutual information for stereophonic acoustic echo cancellation systems with non-linearity, 42nd Asilomar Conference on Signals, Systems and Computers (2008).

[26]. K. H. Lee, J. H. Chang, S. Kimm, N. S. Kang, Y. Kimm, ‘Domain bandwidth dual identification using Gaussian model, IEEE SignalProcess. Lett. 17 (2010) 453{456.

[27]. S. Y. Low, S. Venkatesh, S. Nordholm, ‘A spectral slit approach to doubletalk detection,’ IEEE Transactions on Audio, Speech, and LanguageProcessing (2012) 1074{1080.

[28]. S. Cecchi, L. Rommoli, F. Pizza, ‘Multichannel dual-sensor focused on a fundamental approximation of the frequency,’ IEEE Signal Processing Letters(2016) 94{97.

[29]. T. van Waterschoot, G. Rombouts, P. Verhoeve, M. Moonen, ‘Double-talk robust

[30]. Prediction error identification algorithms for acoustic echo cancellation, IEEE Trans. Signal Process. 55 (2007) 846858.
[31]. J. Gil-Cacho, T. van Waterschoot, M. Moonen, S. H. Jensen, ‘A frequency-domain adaptive filter (FDAF) prediction error method (PEM) framework for double-talk-robust acoustic echo cancellation,’ IEEE/ACM Transactions On Audio, Speech, And Language Processing 22 (2014) 2074–2086.

[32]. Hammidia, A. Amrouache, ‘Dual detector focused on voice abstraction for auditory echo cancelation, 22nd International Conference on Software, Telecommunications & Computer Networks, SoftCOM, IEEE (2014) 393–397.

[33]. M. Z. Ikram, ‘Double-talk detection in acoustic echo cancellers using zero-crossings rate,’ IEEE International Conference on Acoustics, Speech and Signal Processing, (ICASSP), IEEE (2015) 1121–1125.

[34]. M. Hammidia, A. Amrouche, ‘Double-talk detection using the singular value decomposition for acoustic echo cancellation, IEEE International Conference on Communications, ICC (2013) 4745–4749.

[35]. M. Hammidia, A. Amrouche, ‘A new robust double-talk detector based on the Stockwell transform for acoustic echo cancellation, Digit. Signal Process. (2016).

[36]. C. Carlemalm, F. Gustafsson, Wahlberg, ‘On the Problem of detection and discrimination of double talk and change in the echo path,’ Proc. Int. Conf. Acoust., Speech Signal Process. (ICASSP), Atlanta, GA (1996) 2742–2745.

[37]. C. Carlemalm, A. Logothetis, ‘On detection of double talk and changes in the echo path using a Markov modulated channel model,’ Proc. Int. Conf. Acoust., Speech Signal Process. (ICASSP) (1997) 3869–3872.

[38]. H. K. Jung, N. S. Kim, T. Kim, ‘A new double-talk detector using echo path estimation,’ Int. Conf. Acoust., Speech, and Signal Proc. (2002), 123–126.

[39]. J. Bershad, J. Y. Tourneret, ‘Echo cancellation: A likelihood ratio test for double talk versus channel change,’ IEEE Trans. Signal Process. 54 (2006) 4572–4581.

[40]. J. Y. Tourneret, N. J. Bershad, J. C. M. Bermudez, ‘Echo cancellation - the generalized likelihood ratio test for double-talk versus channel change, IEEE Trans. Signal Processing 57 (2009) 916–926.

[41]. H. Buchner, J. Benesty, T. Gansler, W. Kellermann, ‘Robust extended multi delay filter and double-talk detector for acoustic echo cancellation, IEEE Trans. Audio, Speech, and Language Processing 14 (2006) 1633–1643.

[42]. Scheldt., F. Lindstrom, Claesson. I, ‘Dual detector dependent on delay, IEEE Trans on speech and audio processing 20 (2012) 1725–1733.

[43]. G. Szwaoch, A. Czyzewski, A. Ciaarkowski, ‘Dual detection system with audio watermarking,’ Journal of the Audio Engineering Society 57 (2009) 916–926.

[44]. Ruxue Guo and Li Zhao ‘An Improved Double-talk Detection Algorithm for Echo Cancellation in Teleconference System,’ IOP Conf. Series: Materials Science and Engineering 466 (2018) 012040.

[45]. Hyungchan Song and Jong Won Shin, ‘Residual Echo Suppression Considering Harmonic Distortion and Temporal Correlation,’ Appl. Sci. 2020, 10(15), 5291