A REVIEW OF TRANSDUCTION TECHNIQUES USED IN ACOUSTICS ECHO CANCELLATION

Shabbir Majeed Chaudhry
Department of Electrical Engineering,
University of Engineering and Technology, Taxila, Pakistan

Farhat Abbas
Thomson Images and beyond,Germany

Yasar Amin
Department of Computer Engineering,
University of Engineering and Technology, Taxila, Pakistan

Alina Majeed Chaudhry
Department of Electronics,
Federal Urdu University of Arts Sciences and Technology, Islamabad, Pakistan

Habibullah Jamal
Department of Electrical Engineering,
University of Engineering and Technology, Taxila, Pakistan

http://www.uettaxila.edu.pk

Abstract: - — In this review we shall describe the most typical and conventional transducers used in the acoustic echo cancellation systems and the effect of the transducer characteristics on achieving echo cancellation. We shall also discuss the typical design issues found in the literature that often lead to failures in achieving echo cancellation. That is, it is not just a question of achieving good echo cancellation, but instead it can be a question of achieving the echo cancellation at all. The transducers used to capture and produce the sounds are a part of the enclosure and therefore form a part of echo cancellation system. We shall also look at some testing issues designers should consider when implementing an echo canceller.

KEY WORDS: Acoustics, Echo Cancellation, transducers, microphones.

1 Introduction

The space separating a sound source from a receiving transducer can be seen as a transmission channel in which the acoustic signal is conveyed and modified, according to the acoustic characteristics of the environment. The *impulse response h(t)* of the acoustic channel between source and sensor in a reverberant room represents all the multiple reflections from the surrounding surfaces that reach the sensor in addition to the direct sound. Reflections from walls and objects produce a variety of paths between the source and the sensor. Since paths have different length, the propagation delay differs from one path to another and several replicas of the radiated signal reach the sensor after the direct wave front. Every time a wavefront hits a surface, it is partially absorbed and partially reflected. All the reflections contribute to create the complicated impulse response typical of an enclosure.The type of microphone, the way in which it is used, and the environmental conditions in which it operates may have a fundamental impact on the performance that can be obtained in applications such as acoustic echo cancellation.An optimal transducer choice can be made depending on microphone mounting conditions, directional response and other specifications that are discussed in the following.

2.General characteristics

A microphone is a device that converts the acoustic energy of sounds into a corresponding electrical

energy. This transduction is generally realized with a diaphragm whose movements are produced by sound pressure and vary the parameters of an electrical system (a variable resistance conductor, a condenser, etc.) inducing a variable voltage that constitutes the microphone output. As other electrical systems, a microphone may be characterized by a *frequency response* and a *signal-to-noise ratio (SNR).* Two other important parameters in microphone performance are *impedance* and *sensitivity.* Another significant parameter that affects the microphone response characteristics is the *directional response* provided by the *polar pattern.*
According to the polar response shape, the microphones can be classified as either omnidirectional or directional. The latter type can be further subdivided into categories that depend on the directionality characteristics and can be roughly classified into cardioid (supercardioid, hypercardioid, shotgun, etc.), and bidirectional (or figure-of-eight).There is also the category of poli-directional microphones, that have a switch for selecting the desired polar pattern[4].

2.1. Microphone mountings

Microphones can be classified along many criteria. An important criterion describes how they are to be mounted. For speech recognition, the most common mounting options are: *hand-held, head-mounted,* and *table-stand* (or desktop). Recently, improvement in local transmitter-receiver technology has led to the flexible solution of head-mounted wireless microphones (with a radio-frequency connection to the sound acquisition system).
Some microphones (e.g. condenser) are very robust to handling vibrations. Moreover, the angle between the axis normal to the microphone diaphragm and the mouth-microphone direction may influence the quality of the acquisition: for hand-held cardioid microphones, an angle of about 45° is generally suggested. The head-mounted microphones have the advantage of reducing the variations in the "talker's mouth to microphone" distance and the effects of the environment (e.g. background noise, reverberation, echoes, etc.) [6]. Table-stand and *Lavalier* microphones belong to the category of microphones that allow the user to have hands-free interaction [7]. However, they are better suited for teleconferencing, film and television applications than for speech recognition. Besides, a non-optimal noise reduction/compensation performed in the device may alter the speech input spectrum[8].

2.2 Basic Transduction Categories

Microphones can be classified according to the type of transduction . Lavalier microphones are small microphones that are suspended from the neck or attached to clothing. perform The first classification is between *passive transducers* and *active transducers.* A second classification is based on the physical quantity that is transduced (i.e. pressure or pressure gradient)[9].
Another possibility of classifying microphones is given by the material and physical principles on which the transduction is based. In the following a classification is given according to the type of transduction employed.

2.2.1.Electromagnetic and electrodynamic microphones

The most common microphones belonging to these two categories are: *ribbon microphones* and *moving-coil microphones.* Both these microphones are widely utilized, reliable for indoor and outdoor use, have extended and smooth frequency response, good transient response, and are available at moderate cost.Generally, ribbon microphones provide a bidirectional response (also called "figure-of-eight pattern"). Their low inertia results in an excellent transient response.The moving-coil microphone induces a much larger voltage than the ribbon microphone, due to the greater length of the coil.

2.2.2 Electrostatic microphones

Electrostatic microphones are based on variation of sensor capacity caused by air-pressure waves. The most common types are *condenser microphones* and *electret microphones.* The output impedance of a condenser microphone is extremely high. So, the amplifier acts as an impedance adaptor. Condenser microphones have excellent frequency response, low distortion and an excellent transient response. They require a power supply (with batteries or phantom power systems) both for the amplifier and for the condenser element [10].
Electret microphones are special types of condenser microphones with a specially designed capacitor that can hold a charge for a very long time. Electret microphones provide uniform frequency response and good transient capability. However, the dynamic range and sensitivity values in electrets are lower than standard values for condenser units.

2.2.3 Piezoresistive and piezoelectric microphones

Piezoresistive and piezoelectric microphones are based on the variation of electric resistance of their sensor induced by the variations of sound pressure level. *Carbon microphones* consist of a small cylinder ("button") packed with tiny granules of carbon. A diaphragm produces pressure against the button containing the carbon granules. Pickup range, frequency range, and transient response of carbon microphones are limited [10].

Crystal microphones and *ceramic microphones* are based on the same principle as carbon microphones. Crystal microphones contain a crystal of Rochelle salt between two metal plates with the upper one free to move. Crystal microphones can be very small and are characterized by low cost and high voltage output. However, the average quality of crystal microphones is not satisfactory for acoustic echo canceller, and the crystal can easily be damaged by high temperatures and humidity[11].

2.3 Specific Microphones

2.3.1. Pressure -zone microphones

The *pressure-zone microphone (PZM)* has an electret placed approximately in the centre of a plate, with a gap of less than 1 mm between the transducer cover and the plate. They are characterized by a hemispherical pickup pattern and are able to transduce sound levels of up to 150 dB without distortion.

2.3.2. Pressure -gradient microphones

Pressure-gradient microphones have a response proportional to the gradient of the pressure wave. They are suitable for directional acquisition, and generally are based either on a single diaphragm, both sides of which are exposed to the direct sound field, or on a more complex device that can acquire sounds from different directions and compensate for relative delay of the wavefront arrival. The most common types of pressure-gradient microphones are the first-order and the second-order gradient, and the toroidal [3], [4].

2.3 3. Noise-canceling microphones

Differential or noise-canceling microphones are useful in noisy environments, where it is important to reduce the effect of background noise. Differential microphones contain two parallel diaphragms, one of which faces the talker's mouth. Sound arriving from lateral directions will produce equal pressures on both sides resulting in signal cancellation.

2.3.4. Micromechanical silicon microphones

Silicon microphones have low sensitivity to vibration and electromagnetic interference and can be manufactured at low cost depending on the underlying transduction principle [5]. Its particular type, the optical waveguide microphone is based on the principle of conversion of an acoustic signal into an intensity or phase modulation of a light wave. It has been introduced in 1994 and consists in a two-chip set (a membrane chip and a waveguide chip).

2.4. MICROPHONE ARRAYS

2.4..1. Introduction

The ideal method for capturing an acoustic message is **to** use a transducer placed close to the emitting source. For Acoustic echo cancellation, available modeling techniques are based on the assumption that a single, clear speech message is conveyed by the acquired signal. However, this assumption is clearly not valid in many practical situations where the message to be processed is generated at some distance from the transducer and, therefore, is mixed with other sounds [14].

In these conditions, it could be advantageous to exploit the spatial selectivity of a microphone array to acquire the signal of the desired source. A microphone array consists of a set of acoustic sensors placed at different locations to spatially sample a sound pressure field. Using a microphone array it is possible to selectively pick up a speech message, while avoiding the undesirable effects due to distance, background noise, room reverberation and competitive sound sources. This objective can be accomplished by means of a spatio-temporal filtering approach [15], [16]. The directivity of a microphone array can be electronically controlled, without changing the sensor positions or placing the transducers very close to the speaker [17]. Moreover, detection, location, tracking, and selective acquisition of an active talker can be performed automatically [16] to improve the intelligibility and quality of a selected speech message in applications such as teleconferencing and hands-free communication (e.g. car telephony).

2.4..2. Beamforming

A beamformer exploits the spatial distribution of the elements of a microphone array to perform spatial filtering [18]. The microphone signals are appropriately delayed, filtered and added to constructively combine the components arriving from a selected direction while attenuating those arriving from other directions [20].

Delay-and-sum beamforming is the simplest and most straightforward array signal processing technique as shown in the figure 1.

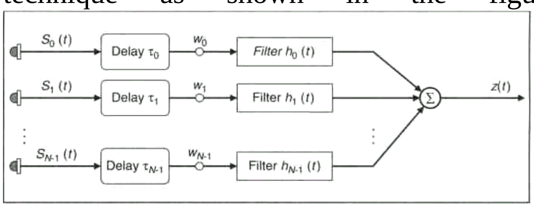

Figure.1. Proposed filter-and-sum beamformer configuration

Figure 1. Moreover, the transfer functions of the filters can be chosen according to the statistical characteristics of the desired signal and interfering noise [19].

2.4..3. Uniform Linear Array

The uniform linear array is the most commonly used sensor configuration in multichannel signal processing. It consists of N transducers located on a straight line and uniformly spaced by a distance d.

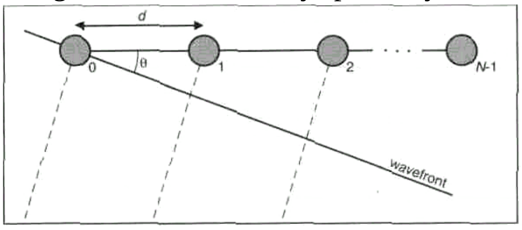

Figure-2. Uniform linear array of N elements and intersensor distance d. A plane wave-front reaches the array from a direction identified by the angle ?.

It becomes apparent that a linear array with equispaced elements is analogous to a digital FIR filter. Therefore, the same design procedures used for FIR filters can be applied to obtain a desired directivity pattern. [21].

Figure 3(a) shows the directivity pattern, at a frequency/ = 2kHz, of a uniform unweighted linear array with 16 elements and an intersensor distance of 10cm, steered to broadside.

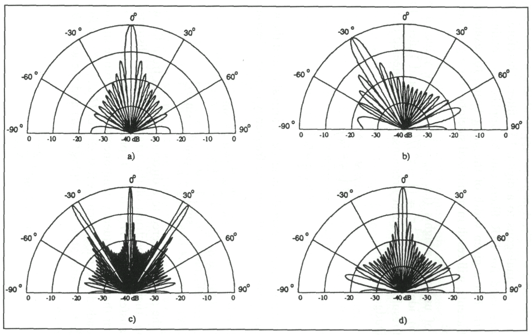

Figure-3. Directivity patterns of uniform linear array

2.4..4. Dereverbe ration

Reverberant speech is the result of a convolution of the original signal with a room impulse response *h(t)*. In principle, reverberation could be removed by an equalizing filter that exactly inverts the room effect. In practice, a room impulse response *h(t)* is generally a non-minimum phase function [22]. Therefore, it is impossible to obtain a direct and exact inverse filter because the inverse function is either unstable or anti-causal (anticipatory)..The exact inverse filtering of the reverberation effect produced by an enclosure has been shown to be possible if more than one sensor is used [20],[24]. This method is based on a principle called multiple-input/output inverse theorem (MINT). Moreover, dereverberation performance is very sensitive to the accuracy of the measured impulse responses and to the variations related to source and receiver positions inside the room as well as to environmental conditions [14]

The computational complexity resulting from long inverse filters can be reduced by the use of a subband MINT technique in which the full speech signal bandwidth is divided into many subbands [23].

Another subband approach to dereverberation [20], [23] considers that the feasibility of inverse filtering a subband signal depends on the z-plane distribution of the zeros of the transfer function in that subband.

2.5. TALKER LOCATION

Acoustic source location is of practical importance for various applications. A possible use for source location may be found in video-conferencing applications, where automatic moving of the camera may be performed based on acoustic source tracking. In hands-free Acoustic Echo Canceller applications, one can envision a target scenario where the microphone array system realizes selective acquisition of a speech message, based on estimated position of the dominant talker, and on a reduction of the captured noise, including background noise produced by other talkers[25].

Acoustic source location can be accomplished following different approaches, which may be grouped into following main categories [26], [27].

(1)Time difference of arrival (TDOA)

(2) Power field scanning

(3) Eigendecomposition

A fourth category that deserves to be mentioned is that known as the *eigen-decomposition-based* techniques. The most representative class of this category is that of the music (multiple signal classification) algorithms[19], [20] generally used for processing narrowband signals and multiple sources. The approach of eigendecomposition techniques should also be investigated to solve the problem of talker location when dealing with multiple speakers in a noisy environment.

Other approaches

Finally, the literature in areas outside talker location (e.g. underwater acoustics) proposes other approaches to source location (see as an example K-rause (1987) for GPS navigation), that are based on the use of high-resolution spectral analysis, maximum-likelihood estimation, ARMA (autoregressive moving average) modelling (Compton, 1988; Bar-Shalom and Fortmann, 1988; Monzingo and Miller, 1980; Wang and Kaveh, 1985; Hassab, 1990; Kay, 1993). These techniques are generally used for narrowband signals processing and require some assumptions about the input signals that are not applicable to speech signals (e.g. the signal is often assumed to be stationary over a long interval).

3 Some Experimental Observations & Testing Tips.

Distortions in transduction hardware is the first thing, which can lead to echo canceller performing very poorly. The very first are the nonlinear distortions in the echo path of the hardware of our device. The echo cancellers perform poorly or don't work at all in systems with the net nonlinear distortions in the echo path higher than -16 dB (typical value). Thus, the smaller the distortion, the better a system will perform. Nonlinear distortions exist everywhere. Certain nonlinearity is inherent in the hybrids, microphones, speakers, amplifiers, and codecs. It is not recommended to make the design with parts, which are highly nonlinear such that the net nonlinear distortions of the device in the echo path are prohibitively high. If there's some preliminary design available, already in a form of a working device, it is a good practice to measure the level of the nonlinear distortions in it [8],[20]. The sooner such measurements are done, the better.

Usually, in the systems, which are not strictly digital (e.g. those involving use of analog circuitry and transmitting analog signals anywhere inside), there's an analog-to-digital or digital-to-analog converter (ADC or DAC) available so the echo canceller implemented on a DSP can work with samples. The path between the DAC output and the ADC input is made up of analog circuitry, which is subject to nonlinearities [28]. If we're talking about using an acoustic echo canceller in some hands-free system, then the analog circuitry in question will include the following: microphone, microphone amplifier, the ADC/DAC itself, the loudspeaker amplifier, and the loudspeaker. This entire echo path must be tested. An easy test for this would be feeding a test signal as samples to the DAC so the speaker would produce it and recording samples from the ADC, e.g. recording what the microphone is picking. The recording should then be analyzed. Note that there can be interference between the digital and analog parts in the device. The interferencemay be in form of additive noise superimposed on the Vcc if the power supply is overloaded or there is no good power supply decoupling. The decoupling capacitors must be placed as close to the power supply pins of the chips as possible.

As already mentioned, the acoustic echo exists between the loudspeaker and the microphone in hands-free phones inside their cases. The echo can be transmitted by both the air inside the case and by the case itself in a form of mechanical waves (vibrations) in the case parts. To reduce this form of echo, there should be a good acoustic decoupling between the loudspeaker and the microphone. To solve this problem, the microphone should be acoustically and mechanically insulated by a soft material, absorbing the case vibrations and sound coming out of the speaker. The microphone should not be directed to the speaker. It can be useful to have a directional microphone, so it can be directed away from the speaker.

Another important thing is the external echo path (e.g. outside the phone's case). The external echo path is actually a number of different echo paths due to the room objects reflecting the speaker's sound back to the microphone. The changes in the echo path impulse response cause an increase in the residual echo error signal.This is the main source of computational errors in acoustic echo cancellation because this forces the Acoustic Echo Canceller to start adapting to the new impulse response and it can even diverge, if the changes are fast or abrupt. In the installed phone, the speaker and microphone should not be directed to the path that is subject to fast changes. It is usually better to direct the speaker and microphone towards the ceiling since this echo path changes rarely[29].

4- CONCLUSIONS

It has been noticed that in order to achieve acoustic echo cancellation or to design an acoustic echo canceller, faithful modeling of room acoustic through experimental setup is the very first and the most important step. Numerical and computational algorithms come next, but they heavily depend upon the offline or online experimental data upon which they work. All this data is collected through the transducers. So it is the most important thing to use the proper transducer for the given circumstances in which acoustic echo cancellation is to be achieved. This in turn requires in-depth study of the transducer types, categories, transducer characteristics in terms of SNR, characteristic impedance, sensitivity, directional response etc. and on special transducer arrangement schemes and finally on the talker-microphone relative positions in teleconferencing and hands free applications. So if the right choice is made in terms of transducers, identification or modeling of acoustic characteristics of ACOUSTIC ECHO CANCELLER systems will lead to good echo cancellation.

5-REFERENCES

[1] Morse,P.M and Ingard,K.U (1986) Theoretical Acoustics . Princeton University Press,Princeton New Jersey, USA.

[2] Davis,D. and Davis,C.(1989) Sound System Engineering. Howard W.Sams &Co.,Second Edition.

[3] Rabiner,L.R. (1994) Applications of voice processing to telecommunications. IEEE Transactions on Speech and Audio Processing,2:199-230.

[4] Clifford,M.(1986) Microphones. TAB BOOKS Inc. Blue Ridge Summit, Pennsylvania, USA, third edition.

[5] Kuttruff, H. (1994) On the acoustics of auditoria. Journal of Building Acoustics. 1(1)

[6] Haykin, S.(ed) (1995) Advances in Spectrum Analysis and Array Processing. Prentice Hall, Englewood Cliffs, New Jersey, USA.

[7]Steeneken ,H.J.M and Houtgast,T.(1980) A physical method for measuring speech transmission quality. Journal of the Acoustical Society of America,67(1).

[8] Gray, R.M., Buzo, A.,Gray,A.H and Matsuyama,Y. (1980) Distortion measures for speech processing. IEEE Transactions on Speech and Audio Processing,28(4).

[9] Sessler,G.M., West,J.E.(1989) Unidirectional,second-order gradient microphone. Journal of the Acoustical Society of America,86.

[10] Flanagan J., Berkley , D., Elko, G., (1991) Autodirective Microphone Systems. Acustica 75

[11] Van Summers,W.,Pisoni,D(1988) Effects of noice on speech production:acoustic and perceptual analyses. Journal of the Acoustical Society of America,84(3).

[12]Sessler,G.M.,(1996)Siliconmicrophones. Journal of the Audio EngineeringSociety,44(1/2).

[13] A. E. Kabir, R. Bashir, J. Bernstein, J. De Santis, R. Mathews, J. O. O'Boyle, C. Bracken,"Very High Sensitivity Acoustic Transducers with ThinP+ Membrane and Gold Back Plate", Sensors and Actuators-A, Vol. 78, issue 2-3, pp.138-142, 17th Dec. 1999.

[14] Berkhout,A.J., de Vries, D. (1980) A new method to acqire Impulse responses in concert halls. Journal of the Acoustical Society of America,68(1). 179-183.

[15] Silverman, H.F.(1997) Some analysis of microphone arrays for speech data acquisition.IEEE Transactions on Acoustics, Speech and Signal Processing (135)

[16] Flanagan J.L.(1987) Three dimensional microphone arrays. Journal of the Acoustical Society of America,82:S.39.

[17] Brandstein, M.S., Adcock,J.E and Silverman, H.F(1996) Microphone array localization error estimation with application to sensor placement. Journal of the Acoustical Society of America,99(6).

[18] Chou,T. (1995) Frequency-independent beamformer with low response error. In proceedings of the IEEE international Conference of Acoustics , Speech and Signal Processing, 2995-2998, Detroit, Michichagan ,USA.

[19] Houtgast, T. and Steeneken, H.J.M.(1985) A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. Journal of the Acoustical Society of America,77(3).

[20] Miyoshi, M. and Kaneda,Y. (1988) Inverse filtering of room acoustics. IEEE Transactions on acoustics, speech and Signal processing36(2)

[21] Sankar,A.,Neumeyer,L.&Weintraub,M.(1996). An experimental study of acoustic adaptation algorithms. . In proceeding for the IEEE international Conference on acoustics,speech and Signal processing. II:713-716. Atlanta,Georgia ,USA.

[22] Peterson, P.M (1986) Simulating the reponse of multiple microphones to a single acoustic source in a reverberant room. Journal of the Acoustical Society of America,80(5).

[23] Wang,H and Itakura,F.(1991) An approach of dereverberation using multi-microphone sub-band envelope estimation. In proceeding for the IEEE international Conference on acoustics,speech and Signal processing,953-956 -Toronto, Canada

[24] Peterson,P.M(1986) Simulating the response of multiple microphones to a single acoustic source in a reverberant room. Journal of the Acoustical Society of America,80(5).

[25] Brandstein, M.S., Adcock,J.E and Silverman, H.F(1997) A closed-form location estimator for use with room environment microphone arrays. IEEE Transactions on Speech and Audio Processing,5(1).

[26] Schroeder, M.R. (1979) Integrated –impulse method measuring sound decay without using impulses. Journal of the Acoustical Society of America,66(2)

[27]Simmer, K.U., Kuczynski,P.(1992). Time delay compensation for adaptive multi channel speech enhancement systems. . In proceeding for the ISSSE,660-663.

[28] R. Jacob Baker.CMOS, Mxed-Signal Circuit Design. July 2002, Wiley-IEEE Press,USA.

[29]ShabbirM.Chaudhry,A.M.Chaudhry. identify-cation of Acoustic Characteristics of Enclosures with resonant second order dynamics. J- PIER 61, 2006.Massachusetts Institute of Technology, Cambridge,USA.