iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 4 **Issue:** XI **Month of publication:** November 2016

DOI:

Multi microphone Signal Processing Technique to Remove Room Reverberation with Unbiased Coherent-To-Diffuse Power Ratio Estimation

Nikhil S
Cochin college of Engineering and Technology, Valanchery

Abstract—In everyday humans encounter different types of speech communication applications, such as hands-free mobile telephony, voice-controlled systems, hearing aids etc, the recorded microphone signals are corrupted by background noise, room reverberation and far-end echo signals. The signal degradation can lead to total unintelligibility of the original speech signal and degrades the performance of automatic speech recognition systems. In this project the estimation of frequency and time-dependent coherent-to-diffuse power ratio (CDR) from the measured spatial coherence between two omni directional microphones were investigated. From traditional methods for unbiased CDR estimators it is shown that knowledge of either direction of arrival (DOA) of the target source or the coherence of noise field is sufficient for unbiased CDR estimation. A CDR-based dereverberation system is presented and evaluated using signal-based quality measures as well as automatic speech recognition accuracy.

Index Terms---Reverberation, Coherent to diffuse ratio, Recursive averaging, Spatial averaging

I. INTRODUCTION

I N everyday life, humans encounter multiple auditory events either in enclosed spaces or in open-spaces. Auditory events include any audible sound such as speech, music and environmental sounds, and encompass the acoustical identity of the space, where the sounds are reproduced. The human auditory system has an amazing capability in performing a wide range of auditory functions like speech recognition, pitch perception, sound localization, noise suppression, assessment of qualitative characteristics of sounds, auditory scene analysis etc[1]. Humans are able to perform such actions easily under varying environments independently of their specific geometrical or acoustical characteristics. Nevertheless, the neurophysiologic mechanisms that render such functions are still not well understood mainly due to the complexity of the auditory system, which consists of several peripheral subsystems as well as millions of interconnected central neurons. Besides that, when a sound is emitted in an enclosed space, the reflections and diffractions of the surrounding surfaces distort the original source and reverberation becomes a crucial component of the auditory event. Reverberation has influenced the evolution of human culture since the prehistoric times and is one of the mostly-studied concepts in the field of room acoustics [2].

Reverberation, in psychoacoustics and acoustics, is the persistence of sound after a sound is produced. A reverberation, or reverb, is created when a sound or signal is reflected causing a large number of reflections to build up and then decay as the sound is absorbed by the surfaces of objects in the space which could include furniture, people and air. This is most noticeable when the sound source stops but the reflections continue, decreasing in amplitude, until they reach zero amplitude.

Reverberation is frequency dependent: the length of the decay, or reverberation time, receives special consideration in the architectural design of spaces which need to have specific reverberation times to achieve optimum performance for their intended activity. In comparison to a distinct echo that is a minimum of 50 to 100 ms after the initial sound, reverberation is the occurrence of reflections that arrive in less than approximately 50ms. As time passes, the amplitude of the reflections is reduced until it is reduced to zero. Reverberation is not limited to indoor spaces as it exists in forests and other outdoor environments where reflection exists.

Reverberation occurs naturally when a person sings, talks or plays an instrument acoustically in a hall or performance space with sound-reflective surfaces. In the 2010s, the sound of reverberation is often electronically added to the vocals of singers in live sound systems and sound recordings by using effects units or digital delay effects[3].

In various applications such as mobile communications and digital hearing aids, the presence of interfering noise may cause serious deterioration in the perceived quality of speech signals. Thus, there exists considerable interest in developing speech enhancement algorithms that solve the problem of noise reduction in order to make the compensated speech more pleasant to a human listener. The noise reduction problem in single and multiple microphone environments was extensively studied. Single microphone speech enhancement approaches often fail to yield satisfactory performance, in particular when the interfering noise statistics are time-varying. In contrast, multiple microphone systems provide superior performance over the single microphone schemes at the expense of a substantial increase of implementation complexity and computational cost [4].

Dereverberation technology automatically removes the effect of reverberation from audio signals; this is useful in hands-free speech recognition and other speech applications that use distant microphones since room reverberation degrades the quality of speech signals [5].

II. RELATED WORK

It has been observed as early as 1969 that the measured spatial coherence between two microphones allows the discrimination between direct sound and reverberation [6]. A first signal enhancement algorithm based on this observation was proposed by Allen *et al.* in 1977 [7], where the magnitude of the coherence is estimated in the Short-Time Fourier Transform (STFT) domain and used as a gain for reverberation suppression. Other heuristic methods for noise reduction and dereverberation using coherence estimates have since been proposed [8]–[12]. Related methods have also been investigated for noise suppression in connection with beamforming, and post filters which are statistically optimal under certain conditions have been proposed for the suppression of uncorrelated [13] and diffuse [14] noise.

More recently, explicit estimators for the ratio between direct and diffuse signal components, termed the coherent-to-diffuse power ratio (CDR), from short-time coherence estimates have been formulated [15], [16], based on the same assumptions as the earlier optimum post filter derivations [14]. Also, results have since been generalized from omnidirectional microphones to other microphone directivities [17], [18] and spherical microphone arrays [19]. While these estimates can be used for the formulation of post filters for signal enhancement [20], which is the main application considered in this contribution, short-time CDR estimates (or the equivalent "diffuseness" measure) also have applications in parametric coding of spatial audio signals [21] and the extraction of spatial features for automatic speech recognition (ASR) [22].

III. SIGNAL MODEL

We consider the recording of a reverberant or noisy speech signal by two omnidirectional microphones with a spacing d, located in the same horizontal plane. The signal $x_i(t)$ of the i–th microphone is composed of a desired signal component $s_i(t)$ and an undesired component $n_i(t)$ consisting of noise and/or late reverberation, i.e.,

$$x_i(t) = s_i(t) + n_i(t), \qquad i = 1,2 \qquad (1)$$

The microphone, desired and noise signals are represented in the time-frequency (STFT) domain by the corresponding uppercase letters, i.e., $X_i(l,f), S_i(l,f)$ and $N_i(l,f)$,respectively, with the discrete-time frame index l and continuous frequency f, and are assumed to be short-time stationary. Using the representation in the STFT domain, the short-time auto- and cross-power spectra between two signals u(t) and v(t) are defined as:

$$\Phi_{uv}(l,f) = \varepsilon\{U(l,f)V^*(l,f)\}, \qquad (2)$$

where ε is the expectation operator. It is assumed that the auto power spectra of the signal components are the same at both microphones, i.e.

$$\phi_{s1s1}(l,f) = \phi_{s2s2}(l,f) = \phi_s(l,f), \qquad (3)$$

$$\phi_{n1n1}(l,f) = \phi_{n2n2}(l,f) = \phi_n(l,f), \qquad (4)$$

Note that this assumption is generally appropriate for a plane wave as desired signal as well as for noise and late reverberation, but may in practice be impacted by the presence of early reflections causing destructive or constructive interference. The time- and frequency-dependent signal-to-noise ratio (SNR) of the microphone signals can be defined as

$$SNR(l,f) = \frac{\phi_s(l,f)}{\phi_n(l,f)} \quad (5)$$

The complex spatial coherence functions of the desired signal and noise components are given by

$$\Gamma_s(f) = \frac{\phi_{s_1 s_2}(l,f)}{\phi_s(l,f)}, \quad (6)$$

$$\Gamma_n(f) = \frac{\phi_{n1n2}(l,f)}{\phi_n(l,f)}, \quad (7)$$

Respectively, and are assumed to be time-invariant, i.e., dependentonlyonthespatialcharacteristicsofthesignalcomponents.It is furthermore assumed that signal and noise are mutually orthogonal, such that

$$\phi_x(l,f) = \phi_s(l,f) + \phi_n(l,f) \quad (8)$$

The complex spatial coherence of the mixed sound field can then be written as a function of the SNR and the signal and noise coherence functions:

$$\Gamma_x(l,f) = \frac{SNR(l,f)\Gamma_s(f) + \Gamma_n}{SNR(l,f)+1} \quad (9)$$

This relationship is valid for any signal and noise coherence function. For the special case of a fully coherent desired signal component and diffuse noise, the term CDR or direct-to-diffuse ratio (DDR) is often used for the SNR. We will adopt the term CDR in the following. (9) can be rewritten as a parametric line equation in the complex plane, highlighting that Γ_x lies on a straight line connecting Γ_n and Γ_s:

$$\Gamma_x(l,f) = \Gamma_s(f) + \frac{1}{CDR(L,F)+1}(\Gamma_N(f) - \Gamma_s(f)) \quad (10)$$

IV. COHERENT TO DIFFUSE POWER RATIO ESTIMATION

For the application to dereverberation, the CDR is equivalent to the direct-to-reverberation power ratio (DRR), under the assumption that reverberant sound can be modeled as a mixture of a direct component and a perfectly diffuse reverberation component which are mutually uncorrelated, thus neglecting early reflections.

The aim is now to estimate the CDR from an estimate of the short-time spatial coherence, $\hat{\Gamma}_x(l,f)$ exploiting the known coherence functions of the signal and/or noise component, and the relationship of these coherence models and the mixed sound field coherence to the CDR given by (10). Solving (10) for the CDR yields (for brevity, the time- and frequency-dependency is omitted in the following)

$$CDR = \frac{\Gamma_n - \Gamma_x}{\Gamma_x - \Gamma_s} \quad (11)$$

or, reformulated as the diffuseness D,

$$D = \frac{1}{CDR+1} = \frac{\Gamma_x - \Gamma_s}{\Gamma_n - \Gamma_s}, \quad (12)$$

Although $\Gamma_x$andΓ_s may be complex, the CDR and diffuseness are real-valued quantities; however, when inserting a coherence estimate $\hat{\Gamma}_x$ for Γ_x in (11), the resulting values are in general complex-valued, due to mismatch between the coherence models and the actual acoustic conditions, and the variance of the coherence estimate. Estimating the CDR by direct application of (11) is therefore not feasible, which is why a number of different estimator implementations, which yield a positive, real-valued CDR estimate for all possible values of $\hat{\Gamma}_x$, $|\hat{\Gamma}_x| \leq 1$, have been proposed.

A. CDR Estimation for Unknown DOA

However, the instantaneous phase of the mixture is not an unbiased estimate of the phase of the direct signal component, since, for low CDR values, the coherence of the mixture is dominated by the coherence of the diffuse signal component[18], which is real-valued, i.e., has a phase of zero. For $\theta \neq 0^0$, the estimator is therefore biased.

It is possible to derive an unbiased CDR estimator which does not require an estimate of the source DOA, since the knowledge that$|\Gamma_s| = 1$, i.e., that the direct signal is fully coherent is sufficient to solve (11). This can be explained using a geometric interpretation: according to (10),$\Gamma_x, \Gamma_s,$ and$\Gamma_n$all lie on a straight line in the complex plane, and it is furthermore known thatΓ_s lies on the unit circle and on the realaxis.Γ_s can therefore be obtained by the intersection of the linethrough $\Gamma_n$and),$\Gamma_x$with the unit circle, and inserted into (11).An alternative way of obtaining this solution is by solving (10)for Γ_s,and setting the magnitude to 1:

$$|\Gamma_s| = |\Gamma_x - (\Gamma_n - \Gamma_x)CDR^{-1}| \cong 1 \quad (13)$$

which leads to a quadratic equation for the CDR:

$$(|\Gamma_x|^2 - 1)CDR^2 - 2Re\{\Gamma_x(\Gamma_n - \Gamma_x)^*\}CDR + |\Gamma_n - \Gamma_x|^2 = 0 \quad (14)$$

In contrast to the DOA-dependent estimators, where an infinite number of unbiased estimators exist, the DOA-independent estimator is uniquely determined by the requirement of unbiasedness.

V. PROPOSED SYSTEM

In this contribution, the estimation of the CDR from the measured coherence between two omnidirectional microphones, and the application of the CDR estimates to dereverberation, is investigated. First, the signal model for the recording of a noisy or reverberant signal with two omnidirectional microphones is described, the relationship between signal and noise coherence models and the coherence of the mixed signal is given, and coherence models for the application to dereverberation are discussed.

Then, several known CDR estimators are formulated in a common framework, illustrated using a geometric interpretation in the complex plane, and improved unbiased estimators are proposed. It is shown that knowledge of either the target signal direction or the noise coherence is sufficient for an unbiased CDR estimation, and estimators are proposed for the cases of unknown target signal direction and unknown noise coherence. Finally, the CDR estimators are applied in a post filter for reverberation suppression and evaluated by processing reverberant speech and comparing ASR recognition accuracy as well as various signal quality measures.

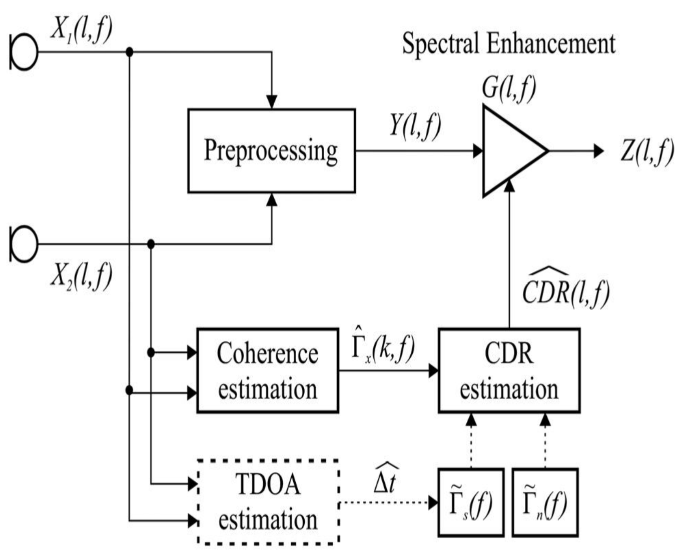


Fig.1. Coherence-based noise and reverberation suppression system consisting of a preprocessor and a CDR-based post filter

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Fig.1 shows the structure of the proposed reverberation or diffuse noise suppression system based on short-time CDR estimates.First, the microphone signals are combined by averaging the squared magnitudes and using the phase from one of the microphone signals:

$$Y(l,f) = \sqrt{\frac{1}{2}(|X_1(l,f)|^2) + (|X_2(l,f)|^2)}.e^{j\arg X_1(l,F)} \quad (14)$$

Spatial magnitude averaging in the STFT domain is typically used to reduce the variance of spectral estimates for the computation of microphone array post filters[14], but has also been used as a preprocessor for signal enhancement[23].It is used here with the purpose of reducing the variations in the transfer function which are caused by constructive and destructive interference of early reflection components with the direct path. For the computation of the coherence based post filter gain, short-time estimates of the spatial coherence are first obtained according to from spectra which have been estimated by recursive averaging. From the coherence, the CDR is estimated based on models for the direct signal and/or reverberation coherence, where the direct signal coherence is derived from a known or estimated TDOA and the reverberation coherence is assumed to be known. A post filter gain is then computed using spectral magnitude subtraction:

$$G(l,f) = \max\left\{G_{min,}, 1 - \sqrt{\frac{\mu}{\widehat{CDR}(l,f)+1}}\right\} \quad (15)$$

with the over subtraction factor and the gain floor. The output signal is computed by applying the post filter gain to the preprocessed signal, i.e., and transformed back into the time domain. Since the preprocessor does not have any spatial filtering effect, the post filter gain can be directly applied to the preprocessor output, and does not require a correction to account for spatial filtering, as it would be the case for a beam former as preprocessor. Note that, when employing a DOA-independent CDR estimator, the proposed signal enhancement system is completely independent of the DOA of the target signal.

Several well-known and some novel CDR estimation methods and their application to dereverberation have been investigated. Using simulated and measured RIRs for different environments, it has been confirmed that the commonly used model of a reverberant speech signal as a plane wave in diffuse noise is sufficiently accurate to justify the application of CDR-based signal enhancement to dereverberation. However, the known CDR estimators were found to be either biased or not robust enough for practical application to signal enhancement.

It has been shown that several variants of unbiased estimators can be derived which improve robustness towards model errors, and that knowledge of either the signal DOA or the noise coherence is sufficient for estimation of the CDR. Employing the improved estimators for dereverberation has been shown to lead to improved dereverberation performance. Using the DOA-independent estimator, the proposed signal enhancement scheme constitutes a completely blind dereverberation system which requires no knowledge or estimation of the signal DOA.

VI. RESULTS AND OBSERVATIONS

In this section we will discuss the obtained results of our experiments. Tool required for the experimentation is MATLAB with Image Processing toolbox.

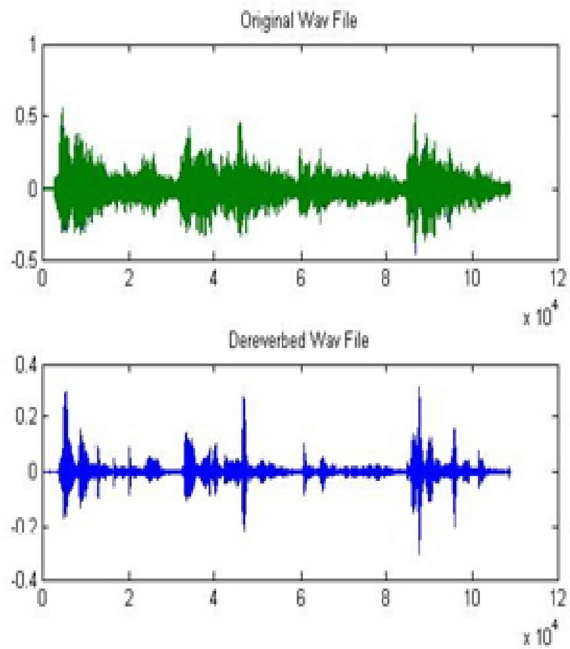


Fig.2 Original and dereverberated signal

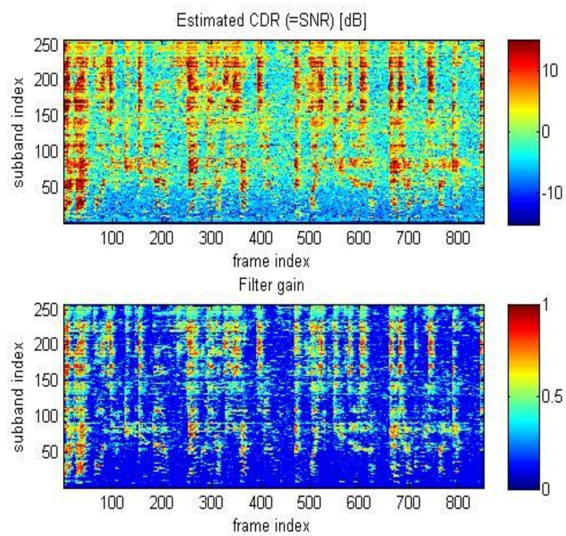


Fig.3.Estimated SNR

VII. CONCLUSION

In many speech communication applications, the recorded microphone signals are corrupted by background noise, room reverberation and far-end echo signals. This signal degradation may lead to total unintelligibility of the speech signal and decreases the performance of automatic speech recognition systems. Hence, high-performance signal enhancement procedures are called for.

The presented algorithm is a multi-microphone signal enhancement algorithm, exploiting both spectral and spatial characteristics. In addition, most algorithms are adaptive, enabling these algorithms to deal with different noise situations and with changing acoustic environments. Generally, we have assumed that the noise sources and the acoustic impulse responses are unknown, requiring 'blind' estimation techniques. Where possible, we have also incorporated robustness against errors in the microphone array characteristics (gain, phase, position) and against other deviations from the assumed signal model (e.g. look direction error, speech detection errors).

However, the known CDR estimators were found to be either biased or not robust enough for practical application to signal enhancement. It has been shown that several variants of unbiased estimators can be derived which improve robustness towards model errors, and that knowledge of either the signal DOA or the noise coherence is sufficient for estimation of the CDR. Employing the improved estimators for dereverberation has been shown to lead to improved dereverberation performance. Using the DOA-independent estimator, the proposed signal enhancement scheme constitutes a completely blind dereverberation system which requires no knowledge or estimation of the signal DOA.

VIII. ACKNOWLEDGMENT

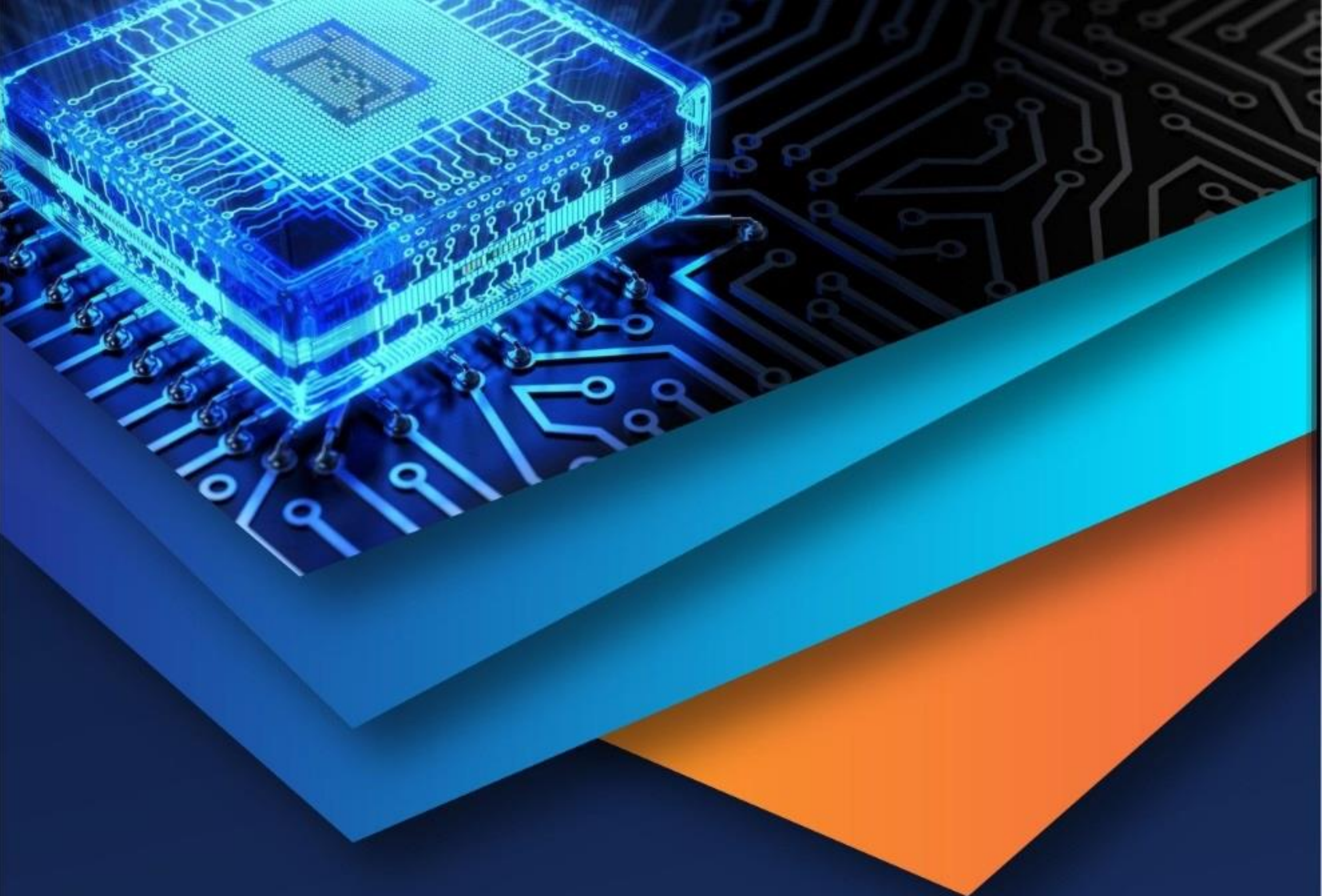
First and foremost we would like to thank **GOD ALMIGHTY** for giving us strength and confidence to see this project through and make it a reality.We are extremely grateful to a lot of people for the successful completion of this project and we feel it our responsibility to thank those who have put in so much time and effort to make our endeavour successful.

REFERENCES

[1] ISO3382. Acoustics measurement of the reverberation time of rooms with reference to other acoustical parameters.
[2] B.Blesser. An Interdisciplinary synthesis of Reverberation viewpoints. J. AudioEng. Soc., 49(10):903.
[3] Reverberation Wikipedia, the free encyclopedia htm
[4] Takuya Yoshioka, Armin Sehr, Marc croix, Keisuke Kinoshita, Roland Maas.Tomohiro Nakataniand Walter Kellermann. "Making machines understand us in reverberant rooms robustness against reverberation for automatic speech recognition," IEEE Signal Processing Magazine, vol. 29. no. 6, pp. 114-126. Nov2012.
[5] Patrick A. Naylor and Nikolay D. Gaubitch "SPEECH DEREVERBERATION"
[6] L. Danilenko, "Binaurales Hören im nichtstationären diffusen Schallfeld," Kybernetik, vol. 6, no. 2, pp. 50–57, Jun. 1969.
[7] J. B. Allen, D. A. Berkley, and J. Blauert, "Multimicrophone signalprocessing technique to remove room reverberation from speech signals," J. Acoust. Soc. Amer., vol. 62, no. 4, pp. 912–915, 1977.
[8] P. Bloom and G. Cain, "Evaluation of two-input speech dereverberation techniques," in Proc. ICASSP, 1982, pp. 164–167.
[9] R. Zelinski, "A microphone array with adaptive post-filtering for noise reduction in reverberant rooms," in Proc. ICASSP, 1988, pp. 2578–2581.
[10] R. Le Bouquin and G. Faucon, "Using the coherence function for noise reduction," IEE Proc. I Commun., Speech, Vis., vol. 139, no. 3, pp. 276–280, 1992.
[11] R. L. Bouquin-Jeannes, A. Azirani, and G. Faucon, "Enhancement of speech degraded by coherent and incoherent noise using a cross-spectral estimator," IEEE Trans. Speech Audio Process., vol. 5, no. 5, pp. 484–487, Sep. 1997.
[12] A. Westermann, J. M. Buchholz, and T. Dau, "Binaural dereverberation based on interaural coherence histograms," J. Acoust. Soc. Amer., vol. 133, no. 5, pp. 2767–2777, 2013.
[13] K. U. Simmer, J. Bitzer, and C. Marro, "Post-filtering techniques," in Microphone Arrays, ser. ser. Digital Signal Process., P. M. Brandstein and D. D. Ward, Eds. Berlin/Heidelberg, Germany: Springer, Jan. 2001, pp. 39–60.
[14] I. McCowan and H. Bourlard, "Microphone array post-filter based on noise field coherence," IEEE Trans. Speech Audio Process., vol. 11, no. 6, pp. 709–716, Nov. 2003.
[15] M. Jeub, C. M. Nelke, C. Beaugeant, and P. Vary, "Blind estimation of the coherent-to-diffuse energy ratio from noisy speech signals," in Proc. EUSIPCO, 2011.
[16] O. Thiergart, G. Del Galdo, and E. A. P. Habets, "Signal-to-reverberant ratio estimation based on the complex spatial coherence between omnidirectional microphones," in Proc. ICASSP, 2012, pp. 309–312.
[17] O. Thiergart, G. Del Galdo, and E. A. P. Habets, "Diffuseness estimation with high temporal resolution via spatial coherence between virtual first-order microphones," in Proc. WASPAA, 2011, pp. 217–220.
[18] O. Thiergart, G. Del Galdo, and E. A. P. Habets, "On the spatial coherence in mixed sound fields and its application to signal-to-diffuse ratio estimation," J. Acoust. Soc. Amer., vol. 132, no. 4, p. 2337, 2012.
[19] D. P. Jarrett, O. Thiergart, E. A. P. Habets, and P. A. Naylor, "Coherence- based diffuseness estimation in the spherical harmonic domain," in Proc. IEEEI, 2012, pp. 1–5.
[20] A. Schwarz and W. Kellermann, "Unbiased coherent-to-diffuse ratio estimation for dereverberation," in Proc. IWAENC, 2014, pp. 6–10.
[21] V. Pulkki, "Spatial sound reproduction with directional audio coding," J. Audio Eng. Soc., vol. 55, no. 6, pp. 503–516, Jun. 2007.

[22] A. Schwarz, C. Huemmer, R. Maas, and W. Kellermann, "Spatial diffuseness features for DNN-based speech recognition in noisy and reverberant environments," in Proc. ICASSP, 2015.

[23] E. A. P. Habets, "Single- and multi-microphone speech dereverberation using spectral enhancement," Ph.D. dissertation, Technische Univ. Eindhoven, Eindhoven, The Netherlands, 2007.



10.22214/IJRASET

INDEX COPERNICUS
45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL
IMPACT FACTOR:
7.429

ISI

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY