Turnitin 原創性報告

已處理到: 2024年07月17日 13:56 CST
代碼: 2418117957
字數: 12113
已提交：1

袁安志_碩論.docx 經由 安志 袁

相似度指標	依來源標示相似度	
11%	Internet Sources:	7%
	出版物:	15%
	學生文稿:	0%

3% match (Rui Wang, Kuangrong Hao, Lei Chen, Tong Wang, Chunli Jiang. "A novel hybrid particle swarm optimization using adaptive strategy", Information Sciences, 2021)
Rui Wang, Kuangrong Hao, Lei Chen, Tong Wang, Chunli Jiang. "A novel hybrid particle swarm optimization using adaptive strategy", Information Sciences, 2021

1% match (從 2022年06月22日 來的網絡)
https://coek.info/pdf-adaptive-multi-channel-least-mean-square-and-newton-algorithms-for-blind-channel.html

---

1% match (從 2023年01月18日 來的網絡)
https://israelcohen.com/wp-content/uploads/2018/05/YekutiaelAvargel_PhD_2008.pdf

---

1% match (從 2020年07月19日 來的網絡)
https://webee.technion.ac.il/Sites/People/IsraelCohen/Publications/TASL_2006.pdf

---

1% match (Yiteng Arden Huang, Jacob Benesty. "Adaptive multi-channel least mean square and Newton algorithms for blind channel identification", Signal Processing, 2002)
Yiteng Arden Huang, Jacob Benesty. "Adaptive multi-channel least mean square and Newton algorithms for blind channel identification", Signal Processing, 2002

1% match (Mingsian R. Bai, Shih-Syuan Lan, Jong-Vi Huang. "Time Difference of Arrival (TDOA)-Based Acoustic Source Localization and Signal Extraction for Intelligent Audio Classification", 2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM), 2018)
Mingsian R. Bai, Shih-Syuan Lan, Jong-Vi Huang. "Time Difference of Arrival (TDOA)-Based Acoustic Source Localization and Signal Extraction for Intelligent Audio Classification", 2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM), 2018

1% match (從 2010年04月06日 來的網絡)
http://externe.emt.inrs.ca/users/benesty/papers/sp_aug2002.pdf

---

1% match (Nakatani, Tomohiro, Takuya Yoshioka, Keisuke Kinoshita, Masato Miyoshi, and Biing-Hwang Juang. "Speech Dereverberation Based on Variance-Normalized Delayed Linear Prediction", IEEE Transactions on Audio Speech and Language Processing, 2010.)
Nakatani, Tomohiro, Takuya Yoshioka, Keisuke Kinoshita, Masato Miyoshi, and Biing-Hwang Juang. "Speech Dereverberation Based on Variance-Normalized Delayed Linear Prediction", IEEE Transactions on Audio Speech and Language Processing, 2010.

1% match ("Communications, Signal Processing, and Systems", Springer Science and Business Media LLC, 2020)
"Communications, Signal Processing, and Systems", Springer Science and Business Media LLC, 2020

1% match (從 2020年05月18日 來的網絡)
https://b-ok.org/book/904757/6c5598

國立清華大學 碩士論文 聲學傳遞函數盲估計以應用於去混響、聲源分離以及增強 Blind estimation of acoustic transfer functions with application to signal dereverberation, source separation, and speech enhancement 系級：動力機械工程學系碩士班組別：電機控制組學號姓名：111033537袁安志Anchi Yuan指導教授：白明憲博士(Dr. Mingsian R. Bai) 中華民國一一三年七月

摘要 雖然在陣列信號處理中，聲學傳遞函數（Acoustic Transfer Functions）通常比相對傳遞函數（Relative Transfer Functions）具有更好的性能，但由於源輸入信號通常不可用，獲得可靠的聲學傳遞函數估計具有挑戰性。為了解決這一問題，我們提出了一種基於卷積傳遞函數（Convolutive Transfer Functions）的創新盲聲學傳遞函數估計方法。我們首先使用到達時間差（Time Difference of Arrival）和廣義互相關相位變換（Generalized Cross Correlation-Phase Transform）估計來定位分佈式陣列中的聲源。接著，我們應用加權預測誤差（Weighted Prediction Error）算法對混合緊湊-分佈式陣列接收到的信號進行去混響，並使用延遲和求和波束形成器作為源信號的初步估計。卷積傳遞函數係數可以使用維納濾波器或卡爾曼濾波器計算，並使用粒子群優化（Particle Swarm Optimization）優化其參數。數值模擬和使用十三麥克風混合陣列進行的實驗證明了所提出技術的有效性。最先進的自適應多通道時域最小均方（Adaptive Multi-channel Time Domain Least Mean Square）方法被用作基線。為了進一步驗證，我們將所提出的方法應用於信號去混響、聲源分離和語音增強等應用。 關鍵詞—卷積傳遞函數，加權預測誤差算法，延遲和加總波束成形器，維納濾波器，卡爾曼濾波器，粒子群優化 ii ABSTRACT While Acoustic Transfer Functions (ATFs) generally lead to better performance than Relative Transfer Functions (RTFs) in array signal processing, obtaining reliable ATF estimates is challenging because the source input is usually unavailable. To address this problem, we propose a novel blind ATF estimation approach formulated using Convolutive Transfer Functions (CTFs). We start by locating the source using Time Difference of Arrival (TDOA) estimated by Generalized Cross Correlation-Phase Transform (GCC-PHAT), by using a distributed array. Next, we apply the Weighted Prediction Error (WPE) algorithm to de-reverberate the signals received by a hybrid compact-distributed array, using the Delay and Sum beamformer as an initial estimate of the source signal. The CTF coefficients can be computed using either the Wiener filter or the Kalman filter with the parameters optimized using Particle Swarm Optimization (PSO). Simulations and experiments using a thirteen-microphone hybrid array demonstrate the efficacy of the proposed technique. The state-of-the-art Adaptive Multichannel Time Domain Least Mean Square (MCLMS) method was used as the baseline. For further validation, we applied the proposed technique to applications, including signal dereverberation, source separation, and speech enhancement. Index Terms — convolutive transfer functions, weighted prediction error, delay and iii sum beamformer, Wiener filter, Kalman filter, particle swarm optimization iv

致謝 時光飛逝歲月如梭，轉眼間碩士生涯也即將邁入尾聲，首先感謝指導教授—白明憲教授的諄諄教誨，在研究上給予許多靈感與導正，如今也才能順利進行口試，這些日子也在老師的身上看到了對於研究的不放棄與熱忱，這也勉勵我未來不論是對於工作甚至是生活都應該抱持相同的態度，且實驗室的生活也讓我了解到如何將理論應用於實際情境下，這些累積的經驗對於我來說可謂是收穫滿滿。 非常感謝陳榮順教授及陳科宏教授願意於百忙之中抽空擔任口試委員，且撥冗閱讀並指導學生的口試論文並使其更加完整，於此獻上無限的感激。 求學生涯中特別感謝家人們尤其是父母在我遇到困難與挫折時給予各方面上的支持與鼓勵，也感謝身邊的朋友不離不棄的陪伴，最後感謝同實驗室的陳佑祥學長、許逸誠學長、孔繁傑學長、賴柏儒學長、林邑軒學長、吳仲倫學長、張馨予學姐及陳思涵學姊在學業上的幫助與指導，得以讓我及時解決研究上的問題，且有幸與同學曾大容、鍾沛霖、陳星宇、李鼎珅、于芷萱、范家萱及李可欣互相討論學業，一起成長，承蒙一路上支持與幫助我的貴人，由衷感謝大家。 v CONTENTS

 LIST OF ALGORITHMS
LIST OF FIGURES

 Chapter 1. INTRODUCTION Blind System Identification (BSI) refers to system identification techniques used in scenarios where the input signal is unavailable, but only the output signal is available. This is an enormous challenge, despite its importance in many practical applications requiring Acoustic Transfer Functions (ATFs), such as acoustic echo cancellation [1] , dereverberation [2] , blind source separation [3] , and beamforming in reverberant environment [4] . Most BSI techniques have typically been formulated in the time domain [5] or the Short Time Fourier Transform (STFT) domain [6] [7] , where they estimate the time domain convolution by multiplying the source STFT with the room impulse response (RIR) STFT. This approximation, called the multiplicative transfer function (MTF) approximation [8] or narrowband approximation, is theoretically valid only if the length of the RIR is shorter than that of the STFT window. In practice, however, this requirement is rarely met, even in moderately reverberant environments. This is due to the limitations of the STFT window in assuming local stationarity of audio signals. In addition, the use of a long STFT window can lead to increased estimation variance and computational complexity. To address this problem, especially in situations involving extended RIRs, crossband filters (CBFs) for linear system identification were introduced in [9] . These 1 CBFs provide an alternative to the MTF approach. In this alternative, the STFT coefficient output is represented as the sum of different convolutions between the STFT coefficients of the input source signal and the RIRs along the frame axis in different frequency bins. For analytical tractability, an approximation of the CBFs called the convolutive transfer function (CTF) [10] has been proposed. This model proposes that for each frequency, the output STFT coefficient can be represented as a unique convolution between the STFT coefficients of the input source signal and the CTF along the frame axis. This thesis outlines a method for estimating blind ATF by CTF approximation. To start the process, a source localization algorithm is required to obtain an accurate source location in space for the source signal pre-processing procedures of the next stage. In this thesis, we use Generalized Cross Correlation-Phase Transform (GCC-PHAT) [11] to estimate the Time Difference of Arrival (TDOA) [12] of each microphone in a distributed array and then use it for source localization. Once the source location has been obtained, we can proceed to the source signal pre-processing stage, which consists of dereverberation and source signal extraction using Weighted Prediction Error (WPE) [13] and Delay and Sum (DAS) beamforming [14] . Then, CTF coefficients are computed with the extracted source signal using Wiener filters [15] or adaptive filters such as Recursive Least Squares (RLS) [16] and Kalman filters [17] . Furthermore, the 2 parameters of the aforementioned filters are optimized using Particle Swarm Optimization (PSO) [18] and its enhanced version [19] . To obtain ATFs in the time domain, namely Room Impulse Responses (RIRs), from the CTF coefficients, the estimated CTF coefficients are convolved with the filter whose magnitude is constant along the frequency axis. The resulting convolved sequence is then subjected to the inverse STFT to obtain RIRs. The convergence performance is evaluated using the Normalized Root Mean Square Projection Mismatch (NRMSPM) between the ground truth RIR and the RIR estimated by the proposed method and the baseline approach, namely the Adaptive Multichannel Time Domain Least Mean Square (MCLMS) method [20] . The simulations cover a wide range of reverberation times from 0.01 seconds to 1.6 seconds, using a hybrid compact-distributed array of 38 microphones, and the applications including signal dereverberation using the Multiple Input/Output Inverse Theorem (MINT) [21] , source separation using Tikhonov Regularization (TIKR) [22] , and speech enhancement using the Minimum Power Distortionless Response (MPDR) [23] beamformer are also performed in the simulation. The effectiveness of these applications is evaluated using several metrics, including the Perceptual Evaluation of Speech Quality (PESQ) [24] and the Signal-to-Distortion Ratio (SDR) [25] . In addition, experiments are conducted in a realistic room with a reverberation time of 0.128 3 seconds, using a hybrid compact-distributed array of 13 microphones. The simulation and experimental results show that the proposed approach drastically outperforms the MCLMS method. 4 Chapter 2. BASELINE APPROACH This section introduces a state-of-the-art blind system identification algorithm, known as Adaptive Multichannel Time Domain Least Mean Square (MCLMS), which is adopted as a baseline approach for comparison with the proposed ATF estimation method. This approach constructs an error signal based on the cross-relations between different channels in a novel and systematic manner, as detailed below. The i-th observation, denoted as xi(n), is the result of a linear convolution between the source signal, s(n), and the corresponding channel response, hi. This relationship is expressed as follows: xi (n) ? hi ? s(n), i ? 1, 2,..., M , (1) where the * indicates linear convolution with respect to the time index n, and M represents the number of channels. In vector form, the relationship between the input and the observation for the i-th channel can be expressed as follows: xi (n) ? His(n) , (2) where 5 xi (n) ? ? xi (n) xi (n ?1) xi (n ? L ?1)?T ?hi,0 ? hi,L?1 0 ? Hi ? ? ? , (3) ?? 0 h i,0 hi,L?1 ?? ? s(n) ? ?s(n) s(n ?1) s(n ? 2L ? 2)?T and T denotes the transpose of a matrix. The channel parameter matrix Hi has dimensions L × (2L-1) and is constructed from the channel's impulse response, which can be expressed in the following form: hi ? ??hi,0 hi,1 hi,L?1??T , (4) where L represents the length of the longest channel impulse response, as assumed. In the absence of knowledge regarding the input signal, the cross-relation between sensor outputs can be employed to estimate the channel impulse responses. This is based on the assumption that xTi (n)h j ? xTj (n)hi , i, j ? 1, 2,, M , i ? j . (5) Nevertheless, during the process of prediction, the aforementioned cross-relation no longer holds true, resulting in the formulation of an error signal as follows: eij (n) ? ?????x0Ti (n)h j ? xTj (n)hi ii ?? jj i, j ? 1, 2,, M . (6) We therefore have (M-1)M/2 distinct error signals eij(n) , excluding the case where eii(n) = 0 and counting the pair eij(n) = -eji(n) only once. On the assumption that all of the error signals are of equal importance, a cost function is defined as follows: 6 ? (n) ? ? ? ei2j (n) . M ?1 M (7) i?1 j?i?1 Subsequently, the channel impulse responses are defined by minimizing the aforementioned error function. In order to prevent the estimate from becoming trivial and consisting entirely of zeros, a unit-norm constraint is imposed on h = [h1T h2T...hMT]T at all times. This results in the error signal becoming eij (n) ?ij (n) ? h(n) . (8) The corresponding cost function is given by J(n) ? M??1 ? ? i2j(n)= ?(n2) . M i?1 j?i?1 h (9) The optimal solution for h is identified by minimizing the mean value of the cost function J(n), which can be expressed in the following form: hˆ ? arg m in E ?J (n)? . h (10) Direct minimization is a computationally intensive process that may prove intractable in the case of long channel impulse responses and a large number of channels. It is for this reason that an LMS algorithm is proposed as a solution to this minimization problem, offering an efficient solution by hˆ(n ?1) ? hˆ(n) ? ??J (n) h?hˆ(n) , (11) where μ represents a small positive step size, while ∇ denotes the gradient operator. In order to ascertain the gradient in (11), it is necessary to take the derivative of J(n) with respect to h, which can be expressed in the following form: 7 ?J (n) ?J (n) ? ?h ? 1 ???(n) h 2 ?? ?h ? 2J (n)h??? , (12) where ?? (n) ?? ?? (n) ? ? ?? (n) ? T T ?h ? ????? ?h1 ? ? ? ? ?h2 ? ? ? ?? (n) ??T ??T . ? ? ?hM ? ?? (13) We will now proceed to evaluate the partial derivative of χ(n) with respect to the coefficients of the k-th (k = 1,2, …, M) channel impulse response as follows: ???h(nk) ??hk ??M?i??11 j??i? 1ei2j(n)? M ? ? ? ? k ?1 M =?i?12eik(n)xi(n)?j??k?12ekj(n)(?xj(n)), k ?1 M (14) =? 2eik(n)xi(n)? ? 2ejk(n)xj(n) i ?1 j?k?1 =? 2eik (n)xi (n) M i ?1 where the final step follows from the fact that ekk(n) = 0. This equation may be expressed in matrix form concisely as follows: ??(n) ?hk ? 2X(n)ek (n) , (15) =2X(n)?Ck (n) ? Dk (n)? h where we have defined, for the sake of convenience, 8 X(n) ? ?x1(n) x2 (n) ek (n) ? ?e1k (n) e2k (n) xM (n)?L?M e Mk (n)?T = ? x1T (n)hk ? xTk (n)h1 ? ? ? ??? xTM (n)hk ? xTk (n)hM ?? ? ? ? xT2 (n)hk ? xTk (n)h2 ? ? = ?Ck (n) ? Dk (n)?h ?0 ? 0 xT1 (n) 0 0? ? 0 0 xT2 (n) 0 0 ? Ck (n) ? ? ? ? ? (16) ?0 0 xTM (n) 0 0? ? ? ?? (k?1)L ? (M?k)L ??M?ML = ??0M?(k?1)L XT (n) 0M?(M?k)L ??M?ML . ?xT (n) ? k 0 0 ? Dk (n) ? ?? 0 xTk (n) 0 ? ? ? ? ? ?? 0 0 xTk (n)??M?ML ? Subsequently, the two matrix products in (15) are evaluated as follows: X(n)Ck (n) ? ?0L?(k?1)L ? ? X(n)Dk (n) ? ??Rx1xk (n) ?i?1Rxixi (n) 0L?(M?k)L ??L?ML M ? , (17) Rx2xk (n) RxMxk (n)??L?ML where Rxixj (n) ? xi(n)xTj (n), i, j ?1,2,...,M . (18) Substitution of (17) into (15) yields the following result: ??(n) ?hk ? 2??Rx1xk (n) ? ? ?Rxixi (n) ? RxMxk (n)??h . (19) i?k ? Subsequently, the integration of (19) into (12) yields the following result: 9 ??(n) ?hk ? 2R(n)h 1 , (20) ?J (n) ? 2 ??2R(n)h ? 2J (n)h?? h where ? i?1 ?? R xixi (n) R(n) ? ?? ? ?Rx1x2 (n) ? ? ?R ? ? x1xM (n) ?Rx2x1 (n) ? Rxixi (n) i?2 ?Rx2xM (n) ?RxM x1 (n) ? ?Rx M x2 (n) ?? . ? ? ? ? Rxixi (n)? i?M ? ? (21) The updated equation is ultimately obtained by substituting (20) into (11), as follows: hˆ(n?1)?hˆ(n)? hˆ2(n?) 2 ??R(n)hˆ(n)?J(n)hˆ(n)??. (22) Provided that the channel estimate is consistently normalized following each update, the simplified algorithm can be implemented as follows: hˆ(n ? 1) ? hˆ(n) ? 2? ?? R(n)hˆ(n) ? ? (n)hˆ(n)?? . hˆ(n) ? 2? ??R(n)hˆ(n) ? ? (n)hˆ(n)?? (23) Assuming that the independence assumption set out in [26] is valid, it can be demonstrated that the LMS algorithm converges in the mean if the step size satisfies the following constraint: 0 ? ? ? 1 , ?max (24) where the largest eigenvalue of the matrix E{R(n) − J(n)IKKxKK} is denoted by λmax. 10 Chapter 3. CTF SIGNAL MODEL 3.1. Representation of LTI Systems in Crossband Filter This section provides a succinct overview of the manner in which digital signals and LTI systems are represented in the STFT domain. For further details, please consult the following sources: [27] and [28] . Firstly, the interrelationship between the crossband filters in the STFT domain and the impulse response in the time domain is established through the utilization of analysis and synthesis windows. Unless otherwise stated, the summation indexes are defined to range from -∞ to ∞. The STFT representation of a signal x(n) is given by xp,k ? ? x(m)w%*p,k (m) , m (25) where 2? w%p,k (n) @w%(n ? pLs)e N k(n ?pLs) j . (26) An analysis window of length N is denoted by w̃(n). The frame index is denoted by p ∈ [1, P], and k ∈ [0, N−1] represents the frequency-band index. The discrete-time shift is represented by Ls. The complex conjugation is represented by *. The reconstruction of x(n), which is inverse STFT, is achieved by N ?1 x(n) @? ? xp,k wp,k (n) , (27) p k?0 where 11 2? wp,k (n) @w(n ? pLs)e N k(n?pLs) j , (28) and w(n) denotes a synthesis window of length N. This thesis assumes w̃(n) and w(n) are real functions. By substituting (25) into (27), we acquire the completeness condition as follows: ? w(n ? pLs)w%(n ? pLs) ? 1 N for all n . (29) p If the analysis and synthesis windows meet the requirements outlined in (29), the signal x(n) can be reconstructed flawlessly using its STFT coefficients xp,k. However, for Ls≦N and for a given synthesis window w(n), there might be an infinite number of solutions to (29); thus, the choice of the analysis window may not be unique according to [29] and [30] . We will now delve into the STFT representation of LTI systems. Let h(n) denote the impulse response of an LTI system with a length of Q, where the input x(n) and output o(n) of this system are connected through the relation as follows: Q?1 o(n) ? ? h(i)x(n ? i) . (30) i?0 From (25) and (30), the STFT of o(n) can be written as op,k ? ?h(l)x(m ?l)w%*p,k (m) . (31) m,l Substituting (27) into (31), we obtain 12 N ?1 o p,k ? ? m,l h(l)? ? xp?,k?wp?,k?(m ? l)w%*p,k (m) k??0 p? N ?1 , (32) =? ? xp?,k?hp,k,p?,k? k??0 p? where hp,k,p?,k? ? ?h(l)wp?,k?(m?l)w%*p,k(m) (33) m,l may be interpreted as the STFT of h(n) using a composite analysis window Σm wp′,k′(m- l) w̃p,k*(m). Substituting (26) and (28) into (33), we obtain h p,k,p?,k? ? ? h(l)w(m ? l ? p?Ls)e j 2N? k?(m?l?p?Ls) ?w%(m? pLs)e N ?j2?k(m?pLs) m,l =? h(l)? w%(m)e? j 2? km ? w((p ? p?)Ls ? l ? m)e N j 2? k?((p?p?)Ls?l?m) N , (34) l m =?h(n)??k,k?(n)? n?(p?

p?)Ls @hp?p?,k,k? where * indicates linear convolution with respect to the time index n, and 2? 2? ? ?j k ?n k,k?(n) ? e N ? w%(m)w(n ? m)e? j m(k?k?) N . (35) m From (34), we know that hp,k,p´,k´ depends on (p - p´) rather than on p and p´ separately. Substituting (34) into (32), we obtain N ?1 N ?1 o p,k ? ??x p?,k? p? p?,k,k? h ? ??x p?p?,k? p?,k,k? h . (36) k??0 p? k??0 p? From (34), we also obtain h p,k,k? ? ?h(n) ?? k,k? (n)? n? pLs . (37) From (35), we get 13 ? j2? k?n N ? w%(m)w(n ? m)e ? j2? m(k?k?) ? j2? k?n ? k,k?(n) ? e N ? e N wn,k?k? , (38) m where wn,k is the STFT representation of the synthesis window w(n) calculated with a decimation factor Ls=1. Equation (36) demonstrates that for a particular frequency- band index k, the temporal signal can be acquired by convolving the signal xp,k´ in each frequency band k´(k´= 0,1,…,N-1) with its corresponding filter hp,k,k´ and subsequently adding up all the outputs. Here, the term for k = k´ is referred to as a band-to-band filter, and k ≠ k´ is referred to as a crossband filter, and crossband filters are employed to eliminate the aliasing effects resulting from subsampling. Note that (37) indicates that, in general, for fixed k and k´, the filter hp,k,k´ has ⌈N/ Ls⌈ − 1 non-causal coefficients. Hence, in echo cancellation applications, these coefficients must be taken into consideration. Extra delay of (⌈N/ Ls⌈ − 1) Ls samples is typically introduced into the microphone signal to deal with this problem, as illustrated in [31] . 3.2. Band-to-band Filter as CTF Signal Model In this paragraph, we will derive a CTF signal model for blind ATF estimation using band-to-band filters. In a noise-free and reverberant environment, a speech signal transmits to microphones via the room effect. In the time domain, the received source image y(n) is specified by 14 y(n) ? a(n) ? s(n) , (39) where s(n) and a(n) represent the source signal and the RIR, respectively, with * indicating linear convolution with respect to time index n. The RIR in (39) is often estimated using MTF in the STFT domain, as demonstrated by y p,k ? ak s p,k , (40) where yp,k and sp,k represent the STFTs of their respective signals, while ak denotes the Fourier transform of the RIR a(n). In addition, p ∊ [1, P] refers to the frame index, N indicates the STFT window size, and k ∊ [0, N−1] represents the frequency index as in crossband filter. However, it is important to note that the approximation in (40) is accurate only if the length of the RIR a(n) is shorter than the STFT window size N. In practical applications, a multitude of filter taps must be considered, numbering in the thousands, which ultimately leads to a severely compromised approximation. Consequently, a considerable rise in the level of computational complexity and a gradual reduction in the rate of convergence will be observed. In order to address this issue, the crossband filter model is employed in the present study. From (36) the STFT coefficient yp,k is expressed as the sum of several convolutions between the STFT- domain source signal and the filter across the frequency bins, as follows: N ?1 yp,k ? ? ? sp?p?,k? ap?,k,k? . (41) k??0 p? 15 Assuming Ls is the STFT frame step as stated above, if Ls is less than N, then ap´,k,k´ becomes non-causal, with ⌈N/ Ls⌈ − 1 non-causal coefficients. The number of causal filter coefficients is dependent on the reverberation time. For simpler notation, we assume that the filter index p´ ranges from 0 to L − 1, where L is the length of the filter. This requires shifting the non-causal coefficients to the causal component, which leads to a fixed delay shift of ⌈N/ Ls⌈ − 1 of the frame index for the received microphone signal [31] . From (37) the STFT domain impulse response ap´,k,k´ relates to the time domain impulse response a(n) by ap?,k,k? ? (a(n)??k,k?(n)) n?p?Ls , (42) which indicates the convolution with respect to the time index n evaluated at frame steps using (38). Note that for the remainder of this article, we will continue to refer to the analysis and synthesis windows in the STFT procedure as w̃(n) and w(n), respectively. To streamline the analysis, we employ the so called CTF approximation, which focuses exclusively on the band-to-band filters with k = k´, as described in the following: L? 1 yp,k ? ? sp?p?,k ap?,k ? sp,k ?ap,k . (43) p??0 Based on this, we are considering a version with multiple channels of M microphones as follows: 16 yip,k ? ? L ?1 s p? p?,k aip?,k , (44) p??0 where yip,k and aip,k represent the i-th microphone signal and the corresponding CTF, respectively. Therefore, the source signals can be expressed in matrix form as follows: ? y1p,k ? ? ?a10,k a11,k L a1L?1,k ? ? ? ?? y2p,k ? ? ??a02,k a 2 2 ? ? sp,k 1,k L aL?1,k ? ? s ? p?1,k . ? M ? ? ? M M O M ? ? ? ? ? M ? (45) ???{yMp,k ?? ???14a40M,4k444a41M4,k42 44444a4LM4?14,k4??3 L 1??4s4p4?2(L?414),k43?? ? yd A s Since the proposed algorithm functions on a frequency basis, the frequency index will be omitted henceforth for the sake of brevity. From (45) we can rewrite the matrix form with respect to frame index as follows: y d , p ? As p , (46) where the bold symbols denote vectors or matrices and the subscript d denotes the delayed signal. Up to this point, we have derived the CTF signal model, which corresponds to equation (46). 17 Chapter 4. PROPOSED METHOD This section presents a technique for estimating the ATF in a blind manner.

It should be noted that the data available for analysis is limited to the positions of the microphones, the delayed microphone signal yd,p (by ⌈N/ Ls ⌈ − 1 frames [31]), and the pre-processed source signal sDAS,p, which is obtained from the non-delayed microphone signal yp. The position of the source is determined through the estimation of the TDOA using the GCC-PHAT technique. In particular, the pre-processed source signal sDAS,p is derived through the application of the WPE algorithm and the DAS beamformer. It is worthy of note that three techniques for estimating the CTF coefficients are provided, namely the Wiener filter, the RLS and the Kalman adaptive filter. Moreover, the parameters of all these filters were optimized using PSO and its enhanced version, as detailed in this thesis. 4.1. TDOA-based Source Localization TDOA represents the difference in the arrival times of an emitted signal at a pair of microphones. Upon reception of the signal by the microphones, an estimation of the TDOA can be made by means of GCC-PHAT technique. Subsequently, the distance difference between the source and the two microphones can be obtained by multiplying the known propagation speed. Subsequently, the distance difference can be employed 18 to ascertain the source position via the CLS algorithm. 4.1.1. GCC-PHAT for TDOA Estimation A signal emitted from a distant source is received by two spatially distinct sensors. The signal can be mathematically represented as follows: x1(t) ??(t) x2(t) ??(t ??12) , (47) where the signal received at the first microphone is designated as ?(t), while the TDOAbetweenthefirstandsecondmicrophonesisdesignatedas ?12.Itisnecessary to transform the two time-domain signals into the frequency domain individually. Subsequently,thecrossspectrumcanbeobtainedasfollows: G12 ?k? ? E ??X1*[k]X2[k]?? , (48) where the Fourier transforms of x1(t) and x2(t) , respectively, are represented as X1[k] and X2[k]. The phase transformation weighting scheme, as illustrated in (49), can be employed to achieve unity gain for each frequency component while preserving the phase data, which contains the actual delay information. Gˆ12 ?k? ? GG1122 ??kk?? (49) The aforementioned result is then transformed to the time domain with the objective of obtaining a correlation function as follows: ?12 ?n? ? IFFT ?Gˆ12[k]? ? IFFT(G ?k?) ? IFFT(ej?G12[k]) . G12 ?k? (50) 19 In theory, upon returning Ĝ12?k? to the time domain, we should obtain a unit impulse function. This result is based on the following fact: FT?1(e?j??)? 1 e?j?? ej?td???(t??). ? 2? ?? ? (51) It can thus be concluded that the peak of the correlation function will indicate the delay time. Nevertheless, in order to enhance the precision of the results, it is possible to employ an interpolation method based on the convolution of a Sinc function and a correlation function. ?12 (t) ? ? ?12[n] ? sin[? (t ? nT) / T] ? (t ? nT) / T (52) n??? The delay time can be calculated as follows: ?12 ?argmax?12(t). t (53) 20 4.1.2. TDOA Measurement Model Figure 1 Relative positions of microphones and the sound source in TDOA-based source localization algorithm. Figure 1 depicts the relative positions of microphones and the sound source in a TDOA-based source localization algorithm. For the purposes of this discussion, it is assumed that S represents the source, that mm (m ? 1,..., M) are the microphones, and that the reference microphone is designated as m0 . In accordance with the stipulations of section 4.1.1, the TDOA can be estimated by utilizing the GCC-PHAT algorithm. Consequently, the distance, defined as the difference between the source- to-microphone distance and the source-to-reference microphone distance, can be 21 Commented [B1]: Not at the bottom expressed as follows: dm0 ? rm ? r0 ? ? m0c ? nm0 , m ? 1, 2,..., M , (54) where rm represents the distance between the source and the m-th microphone, while r0 denotes the distance between the source and the reference microphone. The symbol ?mo represents the TDOA between the m-th microphone and reference microphone, while c represents the speed of sound. Finally, nm0 represents Gaussian white measurement noise with a variance of ?2. It is proposed that the source coordinate is (x, y,z) , the m-th microphone coordinate is (xm,ym,zm), m?1,...,M , and the reference microphone is placed at the origin. Therefore, by neglecting the effects of measurement noise, the equation (54) can be rewritten as follows: dm0 ??moc ?(x? xm)2 ?(y ? ym)2 ?(z ? zm)2 ? x2 ? y2 ? z2 , m ?1,...,M . (55) Subsequently, the second item in (55) is defined as R as follows: R ? x2 ? y2 ? z2 . (56) Finally, we can convert the system of linear equations in (55) to matrix form as follows: Aθ = b , (57) where ? x1 A? ?? x2 ? ? ? xM y 1 z 1 d 10 ? ? ? x ? ? x12 ? y12 ? z12 ? d102 ? y2 z2 d20 ? , θ ? ?? y ?? 1 ? x22 ? y22 ? z22 ? d202 ? ? , and b ? 2 ? ? ? ? ? . ? z ? (58) yM zM dM 0 ? ? ??R?? ?? xM 2 ? yM 2 ? zM 2 ? dM 0 2 ? ?? Subsequently, θ may be estimated utilizing the standard least-squares (LS) method, 22 as follows: θ ? arg min(Aθ ? b)T (Aθ ? b) ? (AT A)?1ATb , θ (59) where θ ? [x, y, z, R]T represents a vector of optimization variables. In order to achieve enhanced performance, it is necessary to incorporate considerations of measurement error.

Assuming a high signal-to-noise ratio (SNR) of the measurement, the squared difference of distance can be expressed as dm20 ? (? m0c ? nm0)2 ? (? m0c)2 ? 2(? m0c)nm0 . (60) It thus follows that the discrepancy between the actual and the measured squared difference of distance is ? m ? dm20 ? (? m0c)2 ? 2(? m0c)nm0 ? 2(? m0c)(nm ? n0) . (61) In vector form, it can be expressed as ε ? [2(? 10c)(n1 ? n0) 2(? 20c)(n2 ? n0) 2(? M 0c)(nM ? n0)]T . (62) The covariance matrix of the disturbance is therefore of the following form: ψ ? E[εεT] ? 4? 2 ?? ? 2(?10c)2 (?10c)(? M 0 c)? ? , ? (63) ??(?10c)(? M 0c) 2(? M 0c)2 ? ? where E?(ni? n0)(nj?n0)(?i0c)(?j0c)? ??E?n02??E?2n0ni??E?2n0nj??E?ninj??(?i0c)(?j0c) . ??2?2(? i0c)2, fori? j (64) ?????2(?i0c)(?j0c), fori? j Finally, the weighting matrix W?ψ?1 can be employed to formulate the weighted 23 least-square localization problem as follows: θW ? arg min(Aθ ? b)T W(Aθ ? b) . θ (65) 4.1.3. Constrained Least Squares (CLS) Method It is important to note that, in (56), the variable R is dependent on the variables x, y, and z. Therefore, it is necessary to apply a constraint in order to satisfy the basic relationship as follows: ? x ?T ? ?1 ? y ? ? ? 0 0 0 ? ? x ? x2 ? y2 ? z2 ? R2 ? ? 0 1 0 0 ? ? y ? ? z ? ?0 ? ? ? ? 0 1 0 ? ? z ? ? θT Pθ ? 0 ?R? ? ? ?0 0 0 ?1? ?R? ? ? ? ? x ? ? θ ? ? y ? ? . ? z ? (66) ?R? ? ? ?1 0 0 ? 0 ? P ? ? 0 1 0 0 ? ? ?0 ? 0 1 0 ? ?0 0 0 ?1? ? As a result, we adopt the method of Lagrange multipliers as a strategy for identifying the local minimum of a function subject to equation constraints. This is achieved by: L(θ, ?) ? (Aθ ? b)T W(Aθ ? b) ? ? (θT Pθ) . (67) The solution of (67) can be readily obtained by applying the partial derivative with respect to θ as follows: ?L(θ,?) ?0 ? θcw ?(ATWA??P)?1(ATWb). (68) ?θ From (68), it can be seen that the sole variable is ? , which allows us to modify CLS to a root-finding problem as follows: 24 Commented [B2]: What is psi? θˆTcw(?)Pθˆcw(?) ? 0 . (69) Nevertheless, the real roots of (69) are likely not singular. In the event that multiple real roots are identified, a number of solutions may be obtained by substituting each root into (68). Subsequently, the optimal solution is selected by minimizing the objective function in (65). To date, we have successfully identified the genuine location of the source. 4.2. Pre-processing In order for the subsequent CTF estimation algorithm to function effectively, it is necessary to have access to a source signal that is free from contamination and echoes. Nevertheless, in practical applications, obtaining a clean source signal is frequently a significant challenge. Accordingly, this thesis utilizes the Weighted Prediction Error (WPE) method for dereverberation, as detailed in [13] . Subsequently, the Delay and Sum (DAS) beamformer is employed with the source location obtained from section 4.1. in order to extract a clean source signal from the WPE outputs of all channels. 4.2.1. WPE In the event that a single speech source is captured by M microphones, it is possible to rewrite (39) as follows: La ?1 y m (n) ? am (k)s(n ? k) , ? (70) k ?0 25 where m and La represent the ordinal numbers of the m-th microphone and the length of the RIR. The reverberant signal ym(n) in (70) is comprised of three distinct components: a direct signal, early reverberation, and late reverberation. It is a common practice to take the first two components as the desired signal, which is denoted by dm(n). Concurrently, the late reverberation is designated as the signal to be eliminated and is denoted by rm(n). The relationship between these signals can be expressed as follows: y m (n) ? d m (n) ? r m (n) , (71) where d m (n) ? ? am (k)s(n ? k) D?1 k?0 La ?1 , (72) rm (n) ? ? am (k)s(n ? k) k?D where D is the sample index that distinguishes the RIR into the early and late reverberation parts. This index is subsequently referred to as the prediction delay. From now on, we assume that there two microphones, namely M = 2, for the sake of simplicity. If the RIRs am(n) in different channels do not share common zeros, the relationship between speech signal and the microphone signals in (70) can be rewritten (stepwise derivation is shown in [32]) as ym (n) ? (cm)T y(n ? D) ? d m (n) , (73) where 26 y(n) ? ??y1(n)T,y2(n)T ??T ?i 2Lc?1 ym(n) ? ?? ym(n), ym(n?1),L , ym(n?Lc ?1)??T ?i Lc?1 cm ? H(HTH)?1almate ?i 2Lc?1 ? a1 (0) ? a1(1) L a1(La ?1) 0 L L 0 ? ? 0 a1(0) a1(1) L a1(La ?1) 0 L 0 ? ? ? (74) O ? H ? ?? a2(0) ? 0 L L L 0 a1(0) a1(1) a1(La ?1)? ? ? a2(1) L a2(La ?1) 0 L L 0 ? ? i 2Lc?LH ? 0 a2(0) a2(1) L 0 L 0 ? ? a2(La ?1) ? ? O ? ? ?? 0 L L 0 a2 (0) a2 (1) L a2(La ?1)?? amlate ? ??am (D), am (D ?1),L , am (La ?1), 0,L , 0?? ? i LH ?1 , where cm and Lc indicate the vector of regression coefficients and the regression order, respectively, and LH equals Lc + La − 1. Using the estimated vector of regression coefficients ĉm and following (73), it is possible to acquire the desired signal as follows: dˆ m (n) ? ym (n) ? (cˆm)T y(n ? D) . (75) Hence, the dereverberation can be achieved by obtaining a suitable estimated vector of regression coefficients ĉm from the microphone signals. Because ĉ1 is completely determined independently of ĉ2, in the following, we disregard the optimization of ĉ2 without loss of generality. The resultant optimization algorithm can be summarized (stepwise derivation is shown in

[13]) as follows: 27 Algorithm 1 WPE Input: y1(n), y(n) 1) Initialize $\sigma(\hat{}n)2$ as ? ˆ(n)2 ?max?? 1 ? n?Lf?1 2 ?Lf n??n?Lf?1 ? y1(n?)2, ???, ? (76) ? ? 2 ? where Lf is length of short time frame and μ > 0 is a certain lower bound to avoid zero division. 2) Repeat the following steps until convergence. a. Update ĉ1 as follows: cˆ1 ? Φˆ? Λˆ , (77) where + denotes the pseudo inverse and Φˆ ? ?n?1 ? y(n ??Dˆ()ny)(2n ? D)T ? (78) Λˆ ? ?n?1 y(n ??ˆ D(n))y21(n) ? , (79) where τ is the largest sample index of the microphone signal. b. Update $d1\hat{}(n)$ as $d1\hat{}(n) = y1(n) - (c1\hat{})T\mathbf{y}(n - D)$. c. Update $\sigma(\hat{}n)2$ as follows: Lf ?ˆ(n)2 ?max????L1f n?n??n??L2?f1?1dˆ1(n) , ???. ? 2 2 ? (80) ? ? 28 It is worth noting that we will refer to the WPE output of the m-th channel $dl(n)$ as $yWlPE(n)$ from this point forward. 4.2.2. DAS Beamformer Once the de-reverberated microphone signal $yWlPE(n)$ from WPE is obtained, clean source signal extraction through DAS beamformer can be executed. First, we convert $yWlPE(n)$ to the STFT domain, resulting in $yWlPE,p$. Here, p still represents the frame index as previously explained. Secondly, we calculate the beamforming weight of the DAS beamformer as follows: 1 ? ?jr?1r1,e?jr?2r2,L ,e?jr?MrM ? , ? T wDAS ? ?e (81) M ?? ? ? where κ denotes the wave number and the distance between the source and the m-th microphone, designated by rm, are computed through the location of the source obtained in section 4.1.. Finally, the inner product is performed between the weight and $yWlPE,p$ to obtain the clean source signal as follows: ? yW1PE , p ? ? 2 ? sDAS,p ?wHDAS ? y ? WPE,p? M ? , (82) ???yWMPE,p?? ? where H denotes Hermitian transpose. 29 4.3. Wiener Filtering Approach For the first technique, the Wiener-based derivations are employed to estimate the matrix of CTF coefficients A. This approach minimizes the mean square error as follows: Aˆ ? arAg?£mM?iLn E ??? y d , p ? As DAS , p 22 ??? , (83) where E[·] denotes the expectation with respect to the frames. Therefore, (83) can be rewritten as Aˆ ? arAg?£mM?iLn tr ? R yy ? AR sy ? R sHy A H ? AR ss A H ? ? arAg?£mM?iLn J , (84) where tr{·} denotes the matrix trace, and the associated covariance matrices is R yy ? E ??y d , p y Hd, p ?? ? M ?M Rss ? E ??sDAS,psHDAS,p ?? ? L?L . (85) Rsy ? E ?? sDAS,pydH,p ?? ? L?M By taking the derivative of (84) with respect to AH , we obtain ?AH J ? ?2Rsy ? 2Rss AH ? 0 . (86) The optimal Wiener solution can be obtained as Aˆ ? RsHyR?ss1 . (87) In practical implementation, instead of the expectation, the recursive averaging is adopted to obtain Rsy and Rss as given by Rss,p ??Rss,p ?(1??)sDAS,psHDAS,p, Rsy,p ??Rsy,p ?(1??)sDAS,pyHd,p (88) where α denotes the forgetting factor for the recursive averaging process. The Wiener 30 filtering approach can be summarized as follows. Algorithm 2 CTF estimation using Wiener filtering Input: yd,p, sDAS,p 1) Initialize forgetting factor α and covariance matrices as Rss,0 ? 0 ? L?L, Rsy,0 ? 0? L?M 2) For each instant of frame, p = 1, 2, …, compute Rss,p ??Rss,p ?(1??)sDAS,psHDAS,p Rsy,p ??Rsy,p ?(1??)sDAS,pydH,p . (89) Aˆp ? RsHy,pR?ss1,p It should be noted that the estimated matrix of CTF coefficients Â changes depending on the processed frame, and the accuracy improves as the number of processed frames increases. 4.4. RLS Approach For the second technique, the CTF coefficients matrix is estimated through the application of the adaptive filter algorithm. It is worth noting that the RLS algorithm optimization process discussed in [16] is currently being conducted in the complex domain. The RLS algorithm aims to minimize the sum of the weighted error norm square as Aˆ ? arg min ? p?i d,i ? AisDAS,i p y 2 (90) A? M?L i?1 ? 2 , 31 where p represents both the adaptation iteration and the frame index, while λ represents the forgetting factor multiplied by the square of the error norm concerning the iteration. Guided by the objective function articulated in (90), the RLS algorithm is employed for the estimation of the CTF coefficients matrix Â . The RLS approach can be succinctly summarized in the subsequent algorithmic routine. Algorithm 3 CTF estimation using RLS Input: yd,p, sDAS,p 1) Initialize RLS forgetting factor λ, weight and inverse of correlation matrix as wRLS,0?0 ? L?M, PRLS,0???1I? L? L,whereεisasmallpositiveconstant 2) For each instant of frame, p = 1, 2, …, compute ep ? yHd,p ?sDAS,pwRLS,p?1 ? ?1P s k p ? 1 ? ? ?1s HDARSL,Sp,Pp?R1LSD,pA? S1,spDAS , p w RLS , p ? w RLS , p?1 ? k pe p . (91) PRLS , p ? ? ?1PRLS , p?1 ? ? ?1k ps HDAS , p PRLS , p?1 Aˆ p ? w HRLS , p It should be noted that the estimated matrix of CTF coefficients Â changes depending on the processed frame, and the accuracy also improves as the number of processed frames increases. 32 4.5. Kalman Filter Approach In the third technique, the CTF coefficient matrix is estimated by applying the Kalman filter. It is worth noting that in this thesis we adapt the Kalman filter as an adaptive filter instead of using it as a state space control filter. Despite this modification, the primary concept remains the same. The process equation of the stationary Kalman adaptive filter of each microphone without process noise is described as wmKalman, p* ? wmKalman, p?1* , (92) where $\mathbf{w}Klallal,p \in \mathbb{C}$Lx1 signifies the optimal weight vector and has a connection with the

CTF coefficients matrix as Amp ? wmKalman,pH , (93) where $\mathbf{A}lp$ denotes the m-th row of Ap. The measurement equation of stationary Kalman adaptive filter of each microphone is described as ydm, p ? sTDAS, pwmKalman, p* ? emp , (94) where dpl denotes the measurement noise for each microphone, and E ??empemp* ?? ? Rmp , (95) where Rlp is the covariance of measurement noise. Using the process and measurement equations outlined in (92) and (94), the Kalman gain can be derived by minimizing the error covariance matrix [17] . The subsequent algorithmic routine provides a succinct summary of the stationary Kalman 33 adaptive filter approach. Algorithm 4 CTF estimation using stationary Kalman adaptive filtering Input: ydl,p, sDAS,p 1) Initialize estimated Kalman weight, error covariance matrix, Kalman gain and measurement noise covariance as wmKalman,0 ? 0 ? L?1, PKmalman,0 ? ?I ? L?L , Km0 ? 0 ? L?1, Rm ? ? ? 1?1 , where η and ρ is a small positive constant 2) For each microphone, m = 1, 2, …, For each instant of frame, p = 1, 2, …, compute Kmp ? PKmalman,psDAS,p(sDAS,pHPKmalman,p?1sDAS,p ? Rm)?1 wmKalman,p ? wmKalman,p?1 ?Kmp(ydm,p* ?sDAS,pHwmKalman,p?1) PKmalman,p ? PKmalman,p?1 ?KmpsDAS,pHPKmalman,p?1 . (96) Aˆmp ? wmKalman,pH The non-stationary version of the Kalman filter can be derived by introducing process noise into the process equation. The objective of this process is to render the estimated weights non-deterministic. The utilization of this property enables the effective resolution of the issue of continuously moving sound source positions. The following algorithmic routine provides a concise overview of the non-stationary Kalman adaptive filter approach. Algorithm 5 CTF estimation using non-stationary Kalman adaptive filtering 34 Input: ydl,p, sDAS,p 1) Initialize estimated Kalman weight, error covariance matrix, process noise covariance matrix, Kalman gain and measurement noise covariance as wmKalman,0 ?0? L?1, PKmalman,0 ??I? L?L, Km0 ?0? L?1, QmKalman,0 ??QI? L?L, Rm ??? 1?1 , where η, ηQ and ρ is a small positive constant 2) For each microphone, m = 1, 2, …, For each instant of frame, p = 1, 2, …, compute PKmalman, p ? PKmalman, p?1 ? QmKalman, p?1 Kmp ? PKmalman,p DAS,p(sDAS,pHPKmalman,p?1 DAS,p ?Rm)?1 s s wmKalman,p ? wmKalman,p?1 ?Kmp(ydm,p* ?sDAS,pHwmKalman,p?1) . (97) PKmalman,p ? PKmalman,p?1 ?KmpsDAS,pHPKmalman,p?1 Aˆmp ? wmKalman,pH It is evident that the estimated matrix of CTF coefficients Â exhibits variability depending on the processed frame, both in stationary and non-stationary Kalman filters. Furthermore, the accuracy of the estimation improves as the number of processed frames increases. 4.6. ATF Reconstruction Once the matrix of CTF coefficients $\hat{\mathbf{A}}$ has been estimated from the three approaches aforementioned approaches, the subsequent step is to proceed with the production of the ATFs. The initial step is to generate a unit pulse sequence, which is subject to a delay of (L − 1) Ls points. Subsequently, the sequence is transformed into 35 the STFT domain, resulting in δp,k as follows: 2? STFT??[n?(L?1)Ls]?? ??[n?(L?1)Ls]w%[n?pLs]ejNk(n?pLs) ??p,k. ? (98) n??? It is obvious that the magnitude in different frame index p is a constant along the frequency axis, depending on the analysis window used. Finally, the estimated CTF coefficients are convolved with it to give the following signal: Gˆmp ? ??p?L?p? aˆmp? , L?1 (99) p??0 where p ∊ [0, PATF]. The estimated RIRs $g\hat{l}(n)$, n ∊ [0, NATF], can be obtained by applying the inverse STFT to $G\hat{p}l$. Subsequently, the estimated ATFs, represented by vector $\hat{\mathbf{g}}l$ with each element corresponding to different frequency bins, can be obtained by performing a fast Fourier transform (FFT) on the estimated RIRs $g\hat{l}(n)$. $\hat{\mathbf{g}}l$ is expressed as gˆm ? ??gˆ1m,gˆ2m,L ,gˆmN ? T ATF ? . (100) 4.7. Parameters Optimization Although the algorithms described above demonstrate favorable outcomes in simulations, it is crucial to acknowledge the significant impact that parameters within these algorithms can have on the resulting outcomes. In order to optimize the performance of the aforementioned algorithms, we employ the use of Particle Swarm Optimization (PSO) and its advanced versions [19] in order to optimize the parameters 36 involved. 4.7.1. PSO The PSO algorithm is a swarm intelligent optimization technique inspired by the flocking of birds and schooling of fish [18] . PSO represents each particle's position as a candidate solution during the exploration of a U-dimensional space. At the t-th update iteration, one particle j among the J particles in the population is characterized by its position and velocity as follows: Vj (t) ? ??v1j (t) v2j (t) vUj (t)?? X j (t) ? ??x1j (t) x2j (t) xUj (t)?? . (101) Let the fitness function f : U ? be the one that is required to be minimized. The function accepts a candidate solution in the form of a real vector and produces a real number that represents the fitness value of the given candidate solution. In our case, the candidate solution corresponds to the parameters in our proposed algorithms, and the fitness function can be described as follows: f (Xj(t)) ? ? yd,p ?Aˆ(Xj(t))s DAS,p , p 2 (102) where $\hat{\mathbf{A}}(\boldsymbol{X}j(t))$ represents the estimated CTF coefficients matrix obtained from any one of the three methods when the parameters $Xj(t)$ are specified. After calculating the fitness value of the entire population, pbestj(t) and gbest(t) are updated, which are the personal best position of the j-th particle and the global best position in the 37 population, respectively. pbestj(t) ? ?? pbest1j(t), pbest2j(t), , pbestUj (t)?? gbest(t) ? ??gbest1(t), gbest2(t), , gbestU (t)?? (103) Thevelocityandpositionarethenupdatedusingtheformulasasbelow: Vj(t?1)?win?Vj(t)?c1?r1?(pbestj(t)?Xj(t))?c2?r2?(gbest(t)?Xj(t)), (104) Xj(t?1)?Xj(t)?Vj(t?1) wherewinrepresentstheinertiaweight,r1andr2arerandomvariablesthatfallwithin the interval [0, 1] and c1 and c2 denote two positive acceleration coefficients. It is noteworthythattheupdateprocesswillpersistaslongasthemaximumiterationlimit Tmaxhasnotbeenreached.ThePSOalgorithm'sentireprocessispresentedinFigure2. Figure 2 Block diagram of PSO. 38 4.7.2. ASPSO Although PSO is widely used in the optimization process, it remains limited in its ability to address complicated optimization problems, including premature convergence and insufficient balance between global exploration and local exploitation. To mitigate these challenges, a novel hybrid PSO algorithm using an adaptive strategy (ASPSO) has been developed [19] . It includes four main modifications, namely: inertia weight with chaotic, elite and dimensional learning strategies, adaptive position update strategy and competitive substitution mechanism. These modifications are explained in the following sections. The inertia weight win plays a key role in harmonizing exploration and exploitation within the search progress. Therefore, the choice of the inertia weight is important. While a linear inertia weight is commonly used, the majority of real-world practical scenarios involve complex non-linear systems. Taking advantage of the randomness, ergodicity and sensitivity inherent in chaotic maps, the C-PSO algorithm incorporates a non-linear approach to adjusting the inertia weight [33] . The formula for calculating inertia weight is zt ? Cin ? zt ?1 ? (1 ? zt ?1), zt ? (0, 1) win (t) ? (wmax ? wmin) ? (Tmax ? t) ? wmin ? zt , (105) Tmax where Cin, wmax, wmin and Tmax denote a small positive integer, maximum inertia weight, 39 minimum inertia weight and maximum iteration, respectively. The basic PSO uses personal and global learning strategies to control the velocity and position updates of the particles. Specifically, all particles use their collective best experiences (pbestj(t) and gbest(t)) to accelerate solution progress. However, this approach can lead to trapping in local optimal when dealing with multimodal features. To mitigate this challenge, [19] introduce elite and dimensional learning strategies. In the elite learning strategy, particles learn from exceptional individuals to increase the diversity of the population. Throughout the search, each particle learns from four personal best positions of different particles pbestj(t) randomly selected from the population. Subsequently, the personal best particle j is compared with the above four particles, and the particle with the best fitness value is retained as the new personal best (Fpbestj(t)). The learning strategy is expressed as Cpbest(t) ? arg min? f (pbesta (t)), f (pbestb (t)), , f (pbestd (t))?, a ? b ? c ? d Fbest j (t) ? ???Cpbest(t), f (Cpbest(t)) ? f (pbest j (t)) ?? pbest j (t), f (Cpbest(t)) ? f (pbest j (t)) (106) . An excessive focus on gbest(t) can lead to a rapid diversity in population. To mitigate this potential problem, [19] use the dimensional learning method. By facilitating communication between particles in the dimensional aspect, the mean value provides complementary information, thereby increasing diversity and improving search 40 efficiency. A global particle, denoted Mpbest(t), is defined as J J ? Mpbest(t) ? N 1 ???Jpbest1j(t),?pbest2j(t), ,?pbestUj(t)?. ? ? (107) j ?1 j ?1 j ?1 Finally, the velocity update equation is changed to: V j (t ? 1) ? win (t)V j (t ? 1) ? c1 ? r1 ? (Fpbest j (t) ? X j (t)) ? c2 ? r2 ? (Mpbest (t) ? X j (t)) . (108) Conventional PSO faces the challenge of achieving an effective balance between global exploration and local exploitation during the search process. The position update law induces particles to consistently converge to their previously determined optimal positions, thereby limiting their ability to explore neighborhoods around the known optimal solution. In response to this constraint, a spiral mechanism has been introduced as a local search operator in the vicinity of the known optimal solution region [34] . Building on this inspiration, an adaptive position update strategy that generates particle positions by dynamically orchestrating a balance between local exploitation and global exploration is proposed in [19] . This strategy is articulated by ?j(t)? exp(f(Xj(t))) exp(? f(Xj(t))) 1 J J j?1 X j (t ? 1) ? ????? DX jj (?te)x?p(Vbj ?(lt)?? c1)o,s(2?? lj)(t?) ?gbrest (t), ? j (t) ? r (109) , Dj ? gbest(t)?Xj(t) 2 where Dj represents the distance between the current best position and the j-th particle. The parameter b serves as a constant that determines the shape of the logarithmic spiral, 41 and l is a random number in the range [-1,1]. During each iteration, a ratio βj(t) is calculated by evaluating the fitness value of the current particle in relation to the average fitness value. If βj(t) is small, indicating that the particle is close to the optimal position, there is a need to increase its local exploitation capability. Conversely, if the particle is in a suboptimal position, an update is implemented to increase its global exploration capability, thereby mitigating premature convergence. Finally, a competitive substitution mechanism is introduced to enhance the performance of PSO [19] . In each iteration, the worst-performing particle is identified and replaced, as defined by WX j(t)=argmax?f (X1(t)), f (X2(t)), , f (XJ(t))? NX j (t) ? gbest (t) ? r3 ? (pbeste (t) ? pbest f (t)), e ? f ? j ? [1, 2, , J] , (110) WX j (t) ? ???WX j (t), f (NX j (t)) ? f (WX j (t)) ?? NX j (t), f (NX j (t)) ? f (WX j (t)) where $r3 \in (0,1)$ is a random number. During the search process, all particles in the population acquire knowledge from the global best particle gbest(t). Therefore, gbest(t) significant influences the entire population. In a complex search environment, if gbest(t) becomes trapped in a local optimum, the remaining particles tend to converge towards the suboptimal region, leading to premature convergence. Accordingly, a perturbation strategy is built into ASPSO to facilitate the escape of gbest(t) from local optimal. To minimize the time spent on unfavorable directions, a condition is set to trigger the 42 perturbation strategy if gbest(t) fails to update its value after five iterations. The perturbation strategy is described as follows: Nbest (t) ? r4 ? gbest (t) ? (1 ? r4) ? (gbest (t) ? pbest g (t)), g ? [1, 2, , J] gbest(t) ? ??gbest(t), ?Nbest(t), f (Nbest(t)) ? f (gbest(t)) , (111) f (Nbest(t)) ? f (gbest(t)) where $r4 \in (0,1)$ is a random number. The ASPSO algorithm's entire process is presented in Figure 3. Figure 3 Block diagram of ASPSO. 43 44 4.8. Summary of Proposed Method Table 1 summarizes the method proposed in this thesis, describing the resulting output signals obtained at each stage. Table 1 Flow chart of the proposed method. Flow chart of our proposed method Step 1: Acquire the microphone signal $yl(n)$ and delay it to get $ydl(n)$ Step 2: Do TDOA-based source localization to obtain source location Step 3: Do WPE followed by DAS to $yl(n)$ obtaining sDAS,p Step 4: Estimate the CTF coefficients matrix $\mathbf{A}$ v ̂ia one of the algorithms below a. Algorithm 2 CTF estimation using Wiener filtering b. Algorithm 3 CTF estimation using RLS c. Algorithm 4 CTF estimation using stationary Kalman adaptive filtering d. Algorithm 4 CTF estimation using stationary Kalman adaptive filtering Step 5: Do (98) ~ (100) to obtain estimated RIRs $gl(n)$ or ATFs $\hat{g}l$ Step 6: Filter parameters optimization through PSO or ASPSO. Step 7: Applications: MINT for dereverberation, MPDR beamformer for speech enhancement and TIKR for source separation. 45 Chapter 5. SIMULATIONS A CTF-based blind ATF estimation problem has been proposed and motivated with the aim of achieving fast convergence, adaptivity and low computational complexity. The proposed solution consists of three different approaches developed in the previous sections. For comparison purposes, our approach has also been contrasted with the state-of-the-art BSI method, namely MCLMS. The simulation cases include both fixed and moving sources. This chapter also includes optimizations of filter parameters and applications using estimated ATF or RIR. 5.1. Fixed Source Location Cases Three different room settings were developed to generate RIRs with different reverberation times using the RIR generator [35] , with the specifications of each room listed in Table 2. It is important to note that the system uses a hybrid compact- distributed array of eight microphones in a cuboidal distributed part, with the number of microphones used in the compact part, namely the ULA, at 0.02 m intervals depending on the reverberation time. Speech signals were sampled at 16 kHz and used as sources to generate microphone signals that were convolved with the ground truth RIRs. The layout of each room is shown in Figure 4. Table 2 Specifications of room settings for fixed source location. 46 Specifications Settings Room1 Room2 Room3 Range of T60 (sec) 0.01 0.1 0.2~1.6 Dimensions of the room (m) 0.3 × 0.4 ×0.3 3 × 3×2.5 5× 6×2.5 Number of microphones of ULA 30 10 10 First sensor location of ULA (m) 0.1, 0.1, 0.2 1.1, 1, 1 2.1, 2, 1 Dimensions and first sensor location of distributed array (m) 0.1×0.1×0.1 0.15, 0.25, 0.15 1 × 1 ×1 1, 1, 1 1 × 1 ×1 1, 1, 1 Source location (m) 0.2, 0.3, 0.2 1.7, 1.8, 1.3 2.1, 2.15, 1.1 47 (a) (b) 48 (c) Figure 4 Configurations of the room for fixed source location. (a) Room 1. (b) Room 2. (c) Room 3. In Chapter 5, the values of the free parameters α, λ, ε, η and ρ were consistently set to 0.999, 0.99, 0.01, 0.5 and 0.001 respectively, as they were found to be appropriate for all conditions. The magnitude and phase of the estimated ATF for all frequency bins and the amplitude of the estimated RIR with several reverberation times chosen are compared with their ground truth values and are shown in Figure 5. However, when the reverberation time exceeds 0.1 seconds, it becomes difficult for MCLMS to converge due to the prolonged RIR. Therefore, MCLMS simulations are only performed with 49 reverberation times below 0.1. In addition, it is important to note that in the case of blind estimation, there is an inevitable equalization problem where there will be a scale gap between the estimated RIR and the ground truth RIR. To address this issue, the estimated RIR is rescaled using the ratio calculated as the maximum absolute magnitude of the ground-truth RIR divided by the maximum absolute magnitude of the estimated RIR. Figure 5 shows extremely small absolute magnitude and phase errors across all frequency bins and a remarkable correspondence between the amplitude of the estimated RIR and its ground-truth counterparts, which is the desired result. Table 3 and Figure 6 illustrate the normalized root mean square projection misalignment (NRMSPM) of the estimated ATFs for all algorithms. Again, it is important to note that MCLMS simulations are only performed with reverberation times below 0.1. Consequently, the NRMSPM values for these reverberation times are set to zero. In addition, the algorithms with the lowest NRMSPM for each reverberation time are shown in red. The NRMSPM is defined as follows: NRMSPM ? 20log10 ? ? 1 ?? g N ?i? 1 ψ(i) 1 N 2 ? ? , ? (112) ? where N represents the number of Monte Carlo runs, g denotes a long vector connected by the ground-truth RIR of each channel, (•)(i) denotes a value obtained from the i-th run, and the projection misalignment vector ψ is represented as follows : 50 ψ ? g ? ggˆTT ggˆˆ gˆ, (113) where ĝ denotes a long vector linked by the estimated RIR of each channel. (a) Commented [B3]: Don't different types of figs into one. 51 (b) 52 (c) 53 (d) 54 (e) (f) Figure 5 Magnitude and phase of the estimated ATF and amplitude of the estimated RIR of all algorithms with several chosen T60. (a) ATF with T60= 0.01s. (b) RIR with T60= 0.01s. (c) ATF with T60= 0.5s. (d) RIR with T60= 0.5s. (e) ATF with T60= 1.6s. (f) RIR with T60= 1.6s. Table 3 NRMSPM of RIRs estimated using all algorithms under different T60. method T60 (s) Wiener RLS Kalman MCLMS (baseline) 55 0.01 -9.3472 -9.0147 -9.3561 -7.4119 0.1 -7.1719 -6.1848 -7.1623 -2.98E-07 0.2 -14.2432 -14.9637 -13.5157 0 0.3 -14.3913 -14.8531 -13.8433 0 0.4 -14.7163 -14.793 -14.3051 0 0.5 -14.3667 -14.5819 -13.8722 0 0.6 -14.3403 -14.7172 -13.8901 0 0.7 -14.2824 -14.4914 -13.9392 0 0.8 -14.1173 -14.4959 -13.7197 0 0.9 -13.7846 -13.9609 -13.4715 0 1.0 -13.8204 -13.9497 -13.4665 0 1.1 -13.6486 -13.4545 -13.3825 0 1.2 -13.4506 -13.6967 -13.0737 0 1.3 -13.0662 -13.0599 -12.7338 0 1.4 -12.6869 -12.7366 -12.3377 0 1.5 -12.952 -12.5737 -12.8581 0 1.6 -12.767 -12.2751 -12.7297 0 56 Figure 6 NRMSPM of the estimated RIRs for all algorithms at different T60. 5.2. Fixed Source Location with Parameters Optimization As mentioned above, the parameters of the three filters used to estimate the CTF coefficients can be optimized by applying PSO or ASPSO. Consequently, these optimization algorithms are used to improve the performance of the system and to facilitate a comparison with the non-optimized version. In Section 5.2, we adopt the specifications of Room 3 and focus only on the optimization results of the 38-th microphone for the sake of simplicity. Table 4 presents the NRMSPM of the estimated RIR of the 38-th microphone using the Kalman stationary filter with and without 57 optimization when T60 is 0.2 seconds. The parameters of the Kalman stationary filter to be optimized are η and ρ. The parameters of the PSO, namely U, J, Tmax, win, c1 and c2, are set to 2, 50, 100, 0.6, 2 and 2, respectively, and the parameters of the ASPSO, namely U, J, Tmax, z1, Cin, wmax, wmin, b, c1 and c2, are set to 2, 50, 100, 0.4, 4, 0.9, 0.4, 0.3, 2 and 2, respectively. Table 4 demonstrates that when the filter parameters are optimized using either PSO or ASPSO, the NRMSPM can be reduced to a lower value, which is a more favorable outcome. Table 4 NRMSPM of RIR estimated for the 38-th microphone with and without optimization at T60 = 0.2s. Kalman filter without parameters optimization -13.2238 Kalman filter with PSO -13.5870 Kalman filter with ASPSO -13.5873 5.3. Applications of Estimated ATFs and RIRs In this section, we use estimated ATF or RIR for a variety of acoustic applications, including signal dereverberation using the Multiple Input/Output Inverse Theorem (MINT), source separation using Tikhonov Regularization (TIKR), and speech enhancement using the Minimum Power Distortionless Response (MPDR) beamformer. The primary goal of these applications is to obtain clean source signals. A comparison is made between the unprocessed microphone signal and the signal processed by the above algorithms with the ground truth source signal, using Perceptual Evaluation of 58 Speech Quality (PESQ) and Signal-to-Distortion Ratio (SDR) as evaluation criteria. The room specifications used for MINT are identical to those used for Room 3 in Section 5.1. However, TIKR and MPDR require the presence of an additional source or interferer in the room. Consequently, based on the conditions observed in Room 3, an additional source or interferer is introduced into the room at a position of (2.9m, 2.9m, 1.9m) as shown in Figure 7. Figure 7 Configurations of the room for TIKR and MPDR. Figure 8 shows the PESQ and SDR values calculated between the source signal obtained after MINT dereverberation and the ground truth source signal. The RIR has been estimated from T60 in the range of 0.2 to 1.6 seconds with 0.1 second intervals and are derived from three different approaches: the Winer approach, the RLS approach and 59 the Kalman stationary approach. The unprocessed microphone signal is also evaluated against the ground truth source signal for comparison in terms of PESQ and SDR. The results show that the signals processed by the MINT dereverberation have higher scores than the unprocessed signals. (a) 60 (b) Figure 8 (a) PESQ and (b) SDR of MINT de-reverberated signal using RIR estimated by all algorithms and unprocessed signal. Table 5 and Table 6 present the PESQ and SDR values calculated from the signals obtained after TIKR source separation using the ATF of two sources estimated by the Kalman stationary approach, compared to their respective ground truth signals. This evaluation was conducted under conditions where T60 ranges from 0.2 to 0.6 seconds with intervals of 0.2 seconds. The unprocessed microphone signals are also included in the evaluation for comparison purposes. This comparison demonstrates the effectiveness of TIKR source separation using the estimated ATF in improving the quality and fidelity of the separated signals compared to the unprocessed microphone 61 signal. Table 5 PESQ and SDR of the signal from source 1, separated by the TIKR, and the unprocessed signal. T60 (s) 0.2 0.4 0.6 PESQ of source 1 3.8347 3.6683 3.5515 PESQ of unprocessed signal 1.7486 1.5969 1.5046 SDR of source 1 (dB) 26.2406 23.4508 22.8719 SDR of unprocessed signal (dB) 4.7966 2.4419 0.5752 Table 6 PESQ and SDR of the signal from source 2, separated by the TIKR, and the unprocessed signal. T60 (s) 0.2 0.4 0.6 PESQ of source 2 2.976 2.7389 2.7366 PESQ of unprocessed signal 1.2467 1.2444 1.219 SDR of source 2 (dB) 10.5134 8.0219 9.8436 SDR of unprocessed signal (dB) -5.3411 -5.277 -5.8869 62 For MPDR speech enhancement, the T60 used ranges from 0.2 to 0.6 seconds with intervals of 0.2 seconds. The estimated ATF of the speech target, derived from the Kalman stationary approach, is used for MPDR beamforming to produce the enhanced speech signal. The PESQ and SDR are then calculated between the enhanced speech signal and the ground truth speech signal. The unprocessed microphone signals are also included in the evaluation for comparison. The results of these evaluations are presented in Table 7, and show that the use of MPDR speech enhancement techniques can facilitate the improvement of speech quality. Table 7 PESQ and SDR of the MPDR enhanced speech signal and the unprocessed signal. T60 (s) 0.2 0.4 0.6 PESQ of enhanced speech signal 1.8859 1.7416 1.6923 PESQ of unprocessed signal 1.7486 1.5969 1.5046 SDR of enhanced speech signal (dB) 4.3434 3.6316 3.2715 SDR of unprocessed signal (dB) 4.7966 2.4419 0.5752 5.4. Moving Source Location Cases In real-world scenarios, the position of the sound source may change over time, 63 which poses a challenge to the estimation of the ATF. However, as discussed earlier, the non-stationary Kalman filter is an effective solution to this problem as it introduces process noise into the process equation, which allows for better multiple convergence of the estimated CTF coefficients. In this section, we continue to use the specifications of Room 3 for the simulation. The sound source moves along the x-axis at 0.1 and 0.3 meters, three times every 23 seconds, with the room T60 fixed at 0.4 seconds. Table 8 shows the NRMSPM obtained at three locations using both the stationary and non- stationary versions of the Kalman filter. Figure 9 also shows the ATF and RIR estimated using both the stationary and non-stationary versions of the Kalman filter when the sound source is moved by 0.3 meters. Table 8 NRMSPM of RIRs estimated for three positions using the stationary and non- stationary Kalman filters. Filter type Movement distance NRMSPM1 NRMSPM2 NRMSPM3 stationary 0.1000 -14.3064 -0.5501 -0.0847 Non-stationary 0.1000 -14.0583 -14.4474 -13.8963 stationary 0.3000 -14.307 -0.115 -0.0094 Non-stationary 0.3000 -13.9327 -14.1249 -13.4396 64 (a) 65 (b) (c) 66 (d) Figure 9 Magnitude and phase of the estimated ATF and amplitude of the estimated RIR when the sound source is displaced by 0.3m. (a) ATF of Stationary Kalman filter. (b) RIR of stationary Kalman filter. (c) ATF of Non-stationary Kalman filter. (d) RIR of Non-stationary Kalman filter. The results presented in Table 8 and Figure 9 serve to illustrate the improved effectiveness of the non-stationary Kalman filter over its stationary counterpart in moving source scenarios. 67 Chapter 6. EXPERIMENTS 6.1. Experimental Settings and Parameters To demonstrate the effectiveness of the proposed ATF blind estimation algorithm in a real reverberant room, an experiment was carried out in a room measuring 4 × 4 × 2.5 meters. After measurement, the T60 of the room was determined to be 0.128 seconds. The hybrid compact-distributed array was positioned near a corner of the room. The distributed part consisted of eight microphones mounted on an iron rod frame with a side length of 0.8 meters, forming a cube. The compact part consisted of a five- microphone ULA, with 0.07 meters between each microphone, placed in the lower left corner of the distributed part. The y-axis and z-axis coordinates of the ULA corresponded to the lower left microphone in the distributed part. A loudspeaker was used as the source and positioned within the frame, playing 30 seconds of white noise as the source signal. The ground truth ATF was generated using the signal received by a reference microphone placed directly in front of the loudspeaker. This involved the cross power spectral density divided by auto power spectral density operations on the signals from each microphone in the hybrid compact-distributed array. Figure 10 shows photographs of the experimental setup. 68 Figure 10 Picture of the experimental setup. In Chapter 6, the values of the free parameters η and ρ were consistently fixed at 0.5 and 0.001, respectively, as they were found to be appropriate for all conditions. 6.2. Experimental Results and Discussions Figure 11 shows the magnitude and phase of the ATF, as well as the RIR, estimated using the stationary Kalman approach and the baseline MCLMS approach. Table 9 69 presents the NRMSPM between the RIR estimated by these two approaches and the ground-truth RIR. (a) 70 (b) 71 (c) (d) Figure 11 Magnitude and phase of the estimated ATF and amplitude of the estimated RIR obtained from the experiment. (a) ATF of stationary Kalman filter. (b) RIR of stationary Kalman filter. (c) ATF of MCLMS. (d) RIR of MCLMS. Table 9 NRMSPM of the estimated RIRs obtained from the experiment using the Kalman stationary filter and MCLMS. Kalman stationary filter -2.9109 MCLMS -0.0013 It is clear from these figures and tables that the proposed methods continue to achieve 72 lower NRMSPM than the baseline MCLMS approach, which is a convincing result. 73 Chapter 7. CONCLUSIONS AND FUTURE WORK 7.1. Conclusions This paper presents a blind estimation method for the ATF based on the CTF model. Three techniques are developed to estimate the CTF coefficient matrices using the Wiener filter, the RLS algorithm and the Kalman filter, respectively. The magnitude and phase of the estimated ATF are compared with those of the ground truth ATF, and the NRMSPM of the estimated RIRs is also calculated using its ground-truth counterpart. It can be concluded that the proposed method provides more accurate ATF estimation than the baseline MCLMS approach in both simulation and experimental settings. Furthermore, the results of the dereverberation signal produced by MINT, the separated source signal produced by TIKR and the enhanced speech signal produced by MPDR are compared with the ground truth source signal using PESQ and SDR. The results show a significant improvement of the processed signal compared to the unprocessed microphone signal. Finally, by optimizing the parameters used in the three proposed techniques, it is possible to achieve a reduction in the NRMSPM of the estimated RIRs. 7.2. Future Work Although this thesis enhances the signal using MINT, TIKR and MPDR, the operations of these algorithms are based on the estimated ATF or RIR. Consequently, it 74 is necessary to transform the CTF coefficients back to the ATF or RIR. However, if all these operations can be performed in the CTF domain, it will significantly reduce the processing time of the transformation and increase the quality of the processed signal. The development of such an algorithm is therefore essential. Finally, although several state-of-the-art BSI techniques [5] [6] [7] have been shown to be useful in small reverberation time scenarios, they are still unable to tackle long reverberation time scenarios. Therefore, efforts will be made to modify these techniques using the CTF signal model. 75 REFERENCES [1] J. Benesty, T. Gänsler, D. R. Morgan, M. M. Sondhi, and S. L. Gay, "Advances in Network and Acoustic Echo Cancellation," New York: Springer., 2001. [2] Y. Huang, J. Benesty, and J. Chen, "A blind channel identification-based two- stage approach to separation and dereverberation of speech signals in a reverberant environment," IEEE Trans. Speech Audio Process., vol. 13, no. 5, pp. 882–895, 2005. [3] S. Araki, R. Mukai, S. Makino, T. Nishikawa, and H. Saruwatari, "The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech," IEEE Trans. Audio, Speech, Lang. Process., vol. 14, no. 3, pp. 774–784, May 2006. [4] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," IEEE Trans. Audio, Speech, Lang. Process., vol. 49, no. 8, pp. 1614–1626, Aug. 2001. [5] Y. Huang and J. Benesty, "Adaptive multi-channel least mean square and Newton algorithms for blind channel identification," Signal Processing., vol. 82, no. 8, pp. 1127–1138, Aug. 2002. 76 [6] Y. Huang and J. Benesty, "A class of frequency-domain adaptive approaches to blind multichannel identification," IEEE Trans. Signal Processing., vol. 51, no. 1, pp. 11–24, Jan. 2003. [7] M. K. Hasan, J. Benesty, P. A. Naylor, and D. B. Ward, "Improving robustness of blind adaptive multichannel identification algorithms using constraints," Proc. European Signal Processing Conference (EUSIPCO)., 2005. [8] Y. Avargel and I. Cohen, "On multiplicative transfer function approximation in the short-time Fourier transform domain," IEEE Signal Processing Letters., vol. 14, no. 5, pp. 337–340, 2007. [9] Y. Avargel and I. Cohen, "System identification in the short-time Fourier transform domain with crossband filtering," IEEE Trans. Audio, Speech, Lang.

Process., vol. 15, no. 4, pp. 1305–1319, 2007. [10] R. Talmon, I. Cohen and S. Gannot, "Relative transfer function identification using convolutive transfer function approximation," IEEE Trans. Audio, Speech, Lang. Process., vol. 17, no. 4, pp. 546–555, 2009. [11] B. Van Den Broeck, A. Bertrand, P. Karsmakers, B. Vanrumste, H. Van hamme and M. Moonen, "Time-domain generalized cross correlation phase transform sound source localization for small microphone arrays," 2012 5th European DSP Education and Research Conference (EDERC)., pp.76-80, 2012. 77 [12] T. Li, Z. Deng, G. Wang and J. Yan, "Time Difference of Arrival Location Method Based on Improved Snake Optimization Algorithm, "2022 IEEE 8th International Conference on Computer and Communications (ICCC)., pp. 482- 486, 2022. [13] T. Nakatani, T. Yoshioka, K. Kinoshita, M. Miyoshi and B. -H. Juang, "Speech Dereverberation Based on Variance-Normalized Delayed Linear Prediction," IEEE Trans. Audio, Speech, Lang. Process., vol. 18, no. 7, pp. 1717– 1731, 2010. [14] Harry L. Van Trees, "Optimum array processing: Part Iv of detection, estimation, and modulation theory," New York: Wiley., 2002. [15] J. Benesty, S. Makino, J. Chen, Y. Huang, and S. Doclo, "Study of the Wiener filter for noise reduction," Speech enhancement., pp. 9-41, 2005. [16] S. A. U. Islam and D. S. Bernstein, "Recursive Least Squares for Real-Time Implementation," IEEE Control Systems Magazine., vol. 39, no. 3, pp. 82-85, June 2019. [17] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," Journal of Basic Engineering., 1960. 78 [18] J. Kennedy, and R. Eberhart, "Particle swarm optimization," Proceedings of ICNN'95-International Conference on Neural Networks, IEEE., pp. 1942-1948, 1995. [19] Rui ang, Kuangrong Hao, Lei Chen, Tong Wang, and Chunli Jiang, " A novel hybrid particle swarm optimization using adaptive strategy," Information Sciences., vol. 579, pp. 231-250, 2021. [20] Y. Huang and J. Benesty, "Adaptive blind channel identification: Multi- channel least mean square and Newton algorithms," 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing., pp. II-1637-II-1640, 2002. [21] M. Miyoshi, and Y. Kaneda, "Inverse filtering of room acoustics," IEEE Trans. Audio, Speech, Lang. Process., vol. 36, no. 2, pp. 145-152, Feb. 1988. [22] R. Marxer and J. Janer, "A Tikhonov regularization method for spectrum decomposition in low latency audio source separation," 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)., pp. 277-280, 2012. [23] Q. Trong The and N. Thi Thu Dung, "A Speech Enhancement in Diffuse Noise Field using Minimum Variance Distortionless Response Filter," 2022 79 Commented [B4]: Inconsistent format Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus)., pp. 1395-1398, 2022. [24] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs," IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings., vol. 2, pp. 749-752, 2001. [25] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," IEEE Trans. Audio, Speech, Lang. Process., vol. 14, no. 4, pp. 1462–1469, 2006 [26] S. Haykin, "Adaptive Filter Theory," Englewood Cliffs, NJ: Prentice-Hall,, 1996. [27] M. R. Portnoff, "Time-frequency representation of digital signals and systems based on short-time Fourier analysis," IEEE Trans. Signal Process., vol. ASSP-28, no. 1, pp. 55–69, Feb. 1980. [28] S. Farkash and S. Raz, "Linear systems in Gabor time-frequency space," IEEE Trans. Signal Process., vol. 42, no. 3, pp. 611–617, Jan. 1998. [29] J. Wexler and S. Raz, "Discrete gabor expansions," Signal Process., vol. 21, pp. 207–220, Nov. 1990. 80 [30] S. Qian and D. Chen, "Discrete Gabor transform," IEEE Trans. Signal Process., vol. 41, no. 7, pp. 2429–2438, Jul. 1993. [31] W. Kellermann, "Analysis and design of multirate systems for cancellation of acoustical echoes, " International Conference on Acoustics, Speech, and Signal Processing., vol. 5, pp. 2570-2573, 1988. [32] D. Gesbert, and P. Duhamel, "Robust blind channel identification and equalization based on multi-step predictors," International Conference on Acoustics, Speech, and Signal Processing., pp. 3621–3624, 1997. [33] H. A. Hefny, and S. S. Azab, "Chaotic particle swarm optimization," The 7th International Conference on Informatics and Systems (INFOS)., pp. 1-8, 2010. [34] Seyedali Mirjalili, and Andrew Lewis, "The whale optimization algorithm," Advances in Engineering Software., vol. 95, pp.51-67, 2016. [35] Emanuël Habets, "Room Impulse Response Generator," Internal Report., pp. 1-17, 2006. 81