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，聲學傳輸函數（Acoustic Transfer Functions）通常優於相對傳輸函數（Relative Transfer Functions），但在沒有源信號的情況下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）的創新盲估計方法。首先，使用加權預測誤差（Weighted Prediction Error）算法和基於到達時間差（Time Difference of Arrival）預測源位置的延遲和加總（Delay and Sum）波束形成器來獲取目標源信號的初始估計。隨後，使用維納濾波器或卡爾曼濾波器計算卷積傳遞函數係數，並通過粒子群優化（Particle Swarm Optimization）優化其參數。為了獲取估計的聲學傳輸函數的房間脈衝響應，將單位脈衝序列的短時傅立葉變換（Short-time Fourier Transform）與卷積傳遞函數係數進行卷積，然後應用逆短時距傅立葉變換（Inverse Short-time Fourier Transform）。數值模擬和在真實房間環境中進行的實驗證明了我們所提出的聲學傳輸函數估計技術的有效性。這些驗證是通過與最先進的自適應多通道時域最小均方（Adaptive Multi-channel Time Domain Least Mean Square）方法的比較分析實現的。此外，本論文還強調了需要精確聲學傳輸函數估計的應用，包括使用多輸入/輸出逆定理（Multiple Input/Output Inverse Theorem）進行去混響、使用吉洪諾夫正則化（Tikhonov regularization）進行聲源分離以及使用最小功率無失真響應 (Minimum Power Distortionless Response) 波束形成器進行聲源增強。

***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***

# 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 real source input signal 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 sound source using Time Difference of Arrival (TDOA) with delays estimated by Generalized Cross Correlation-Phase Transform (GCC-PHAT), by using a distributed array. Next, we apply the Weighted Prediction Error (WPE) algorithm to dereverberate the signals received by a hybrid array, using the Delay and Sum (DAS) 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 with various reverberation times and experiments in a room with a reverberation time of 0.12s using a thirteen-microphone hybrid array demonstrate the efficacy of the proposed technique. The Adaptive Multichannel Time Domain Least Mean Square (MCLMS) method was used as the baseline for validation. In addition, this thesis presents applications, including signal dereverberation, source separation and speech enhancement, by using the ATF estimated by the proposed technique.

***Index Terms — convolutive transfer functions, weighted prediction error, delay and sum beamformer, Wiener filter, Kalman filter, particle swarm optimization***