

國立交通大學

機械工程學系

博士論文

多聲道音響重現之分析與實現

**Analysis and Implementation of  
Multi-channel Audio Reproduction**



研究生：李志中

指導教授：白明憲 教授


中華民國九十六年九月

多聲道音響重現之分析與實現  
Analysis and Implementation of  
Multi-channel Audio Reproduction

研究生：李志中  
指導教授：白明憲

Student : Chih-Chung Lee  
Advisor : Mingsian R. Bai

國立交通大學  
機械工程學系  
博士論文



A Thesis  
Submitted to Institute of Mechanical Engineering  
College of Engineering  
National Chiao Tung University  
in partial Fulfillment of the Requirements  
for the Degree of  
Doctor of Science  
in  
Mechanical Engineering

September 2007

Hsinchu, Taiwan, Republic of China

中華民國九十六年九月

# 多聲道音響重現之分析與實現

學生：李志中

指導教授：白明憲

國立交通大學機械工程學系（研究所）博士班

## 摘 要

當使用多聲道揚聲器重現立體空間聲場時，交越干擾消除系統扮演了非常重要的角色。然而此技術因為有限的有效區域與大量的計算量，沒有普遍的應用到一般的系統上。在所有的參數中，揚聲器間的夾角是影響分離效果和有效區域最重要的因素。本篇論文有以下幾項重點：首先，針對兩聲道揚聲器規劃了一套完整的研究來探討聆聽角度對於交越干擾消除效果的影響。目的是要找到一個兼具效能與強健性的配置。文中使用兩種有效區域的定義來評估強健性。在數值模擬階段，除了採用理想的點聲源，為了更接近實際的情況，採用頭部相關轉移函數模擬如頭部在高頻的遮蔽效應。針對三個聆聽角度(10、60、120 度)作主觀與客觀的實驗。第二，提出以次頻帶濾波為基礎之限頻寬交越干擾消除器降低運算的負擔。由於人類對於低頻的訊號較敏感，我們將頻帶限制到 6 千赫茲以下。為了驗證此系統，在無響室內實行包含方位測試和音質測試的主觀聆聽實驗。實驗結果使用變異數分析判斷處理前後是否有統計上顯著差異。第三，發展以免感測器之振膜速度估測器為基礎的重低音加強系統，讓整個 3D 立體空間重現系統更加完整。最後將此系統擴展到多聲道反算濾波，應用在汽車音響上。由於小空間造成的反射、揚聲器與聆聽者沒有位於理想位置和環境噪音等問題，造成車內並不是一個好的聆聽環境。有必要發展一套系統讓聲音在這種環境下能夠正確的傳送。本文針對兩聲道與 5.1 聲道輸入，提出四種方法。針對兩聲道輸入，提出兩種方式，一為以空間響應合成器為基礎之聲道擴展技術加上反算濾波。另一為聲道擴

展技術加上聲道縮減技術和權重與延遲。聲道擴展技術是將兩聲道轉換成 5.1 聲道，而聲道縮減技術則相反。反算濾波目的是定位出正確的 5.1 聲道音像。針對 5.1 聲道輸入也提出兩種方式：一為聲道縮減加上反算濾波；另一為聲道縮減加上權重與延遲。用模擬和實驗來驗證這些演算法且實現於一般的轎車上。主觀聆聽實驗用來比較每個方法並且使用多變量分析結果。



# Analysis and Implementation of Multi-channel Audio Reproduction

Student : Chih-Chung Lee

Advisor : Dr. Mingsian R. Bai

Department (Institute) of Mechanical Engineering

National Chiao Tung University

## ABSTRACT

Crosstalk cancellation system (CCS) plays a vital role in spatial sound reproduction using multi-channel loudspeakers. However, this technique is still not of full-blown use in practical applications due to small sweet spot and heavy computation loading. Among the parameters of loudspeaker deployment, span angle is a crucial factor that has a profound impact on the separation performance and sweet spot robustness achievable by the CCS. First, a comprehensive study was conducted to explore the effects of listening angle on crosstalk cancellation in spatial sound reproduction using two-channel stereo systems. The intention is to establish a sustainable configuration of CCS that best reconciles the separation performance and the robustness against lateral head movement. Two kinds of definition of sweet spot are employed for assessment of robustness. In addition to the point source model, HRTF are employed as the plant models in the simulation to emulate more practical localization scenarios such as the high-frequency head shadowing effect. Three span angles including 10 degrees, 60 degrees, and 120 degrees are then compared via objective and subjective experiments. Second, a bandlimited CCS based on subband filtering approach is presented to reduce the computation loading. A pseudo Quadrature Mirror Filter (QMF) bank is employed in the implementation of CCS filters which are

bandlimited to 6 kHz, where human's localization is the most sensitive. To justify the proposed system, subjective listening experiments were undertaken in an anechoic room. The experiments include two parts: the source localization test and the sound quality test. Analysis of Variance (ANOVA) is applied to process the data and assess statistical significance of subjective experiments. Third, a bass enhancement system based on a sensorless cone velocity observer is developed to construct a complete spatial audio reproduction system. At last, this technique is extended to multi-channel inverse filtering for automotive virtual surround audio system. The interior of a car is known as a notorious listening environment due to reflections in a confined space, non-ideal user/loudspeaker positions, and ambient noise, etc. It is then desirable to develop audio systems that are capable of rendering quality spatial sound fields in harsh car environments. Four design approaches are proposed for 2-channel input and 5.1-channel input, respectively. For 2-channel input, a method of reverberation-based upmixing with inverse filtering and another method of up/down mixing with weighting and delay are presented. The upmixing algorithm is used to convert two-channel signals to four-channel signals, while the downmixing algorithm does just the opposite. Inverse filters are employed to position the virtual sound images according to the 5.1 configuration. For 5.1-channel input, a method of downmixing with inverse filtering and another method of downmixing with weighting and delay are presented. These processing algorithms have been practically implemented on a car. Simulations and experiments were conducted for validating the proposed spatial audio systems. Subjective listening tests were also conducted to compare these methods, with the data processed by multivariate analysis of variance (MANOVA).

## 誌 謝

時光飛逝，四年博士班研究生涯轉眼就過去了。首先感謝指導教授白明憲博士的諄諄指導與教誨，使我順利完成學業與論文，在此致上最誠摯的謝意。而老師指導學生時豐富的專業知識，嚴謹的治學態度以及待人處事方面，亦是身為學生的我學習與景仰的典範。更感謝老師帶領我參與兩次國際研討會，增廣我的國際視野，這絕對是我求學生涯最難忘的回憶。

在論文寫作上，感謝中正大學鄭志鈞教授、成功大學蘇文鈺教授、本校成維華教授、謝世福教授和鄭泗東教授在百忙中撥冗閱讀並提出寶貴的意見，使得本文的內容更趨完善與充實，在此本人致上無限的感激。

回顧這四年的日子，承蒙同實驗室的博士班曾平順學長、歐昆應學長、嚴坤龍學長與陳榮亮學長在研究與學業上的適時指點，並有幸與林家鴻同學互相切磋討論，每在烏雲蔽空時，得以撥雲見日，獲益甚多。特別感謝與我相同研究主題的學弟，董志偉、施畊宇和陳暉文。感謝你們在研究與計畫上大力的幫忙，讓我省了不少力氣。還有和91級到95級學弟在生活上的朝夕相處與砥礪磨練和助理劉嫻婷學妹在行政事務上的幫忙，這些都是我得以完成研究的一大助因，在此由衷地感謝他們。

能有此刻，我也要感謝所有在精神上給我鼓舞支持的人，謝謝各位的幫忙與鼓勵。最後僅以此篇論文，獻給我摯愛的雙親李錦興先生、鄭碧蓮女士、哥哥李志遠、妹妹李宜芳、女友黃美菁。今天我能順利取得博士學位，要感謝的人很多，上述名單恐有疏漏，在此也一致上我最深的謝意。

# TABLE OF CONTENTS

摘 要 .....	I
ABSTRACT .....	III
誌 謝 .....	V
TABLE LIST .....	VIII
FIGURE LIST .....	IX
1. INTRODUCTIONS .....	1
2. MULTI-CHANNEL INVERSE FILTERING FOR CCS FROM A MODEL-MATCHING PERSPECTIVE .....	8
3. THE ANALYSIS OF THE LISTENING ANGLE EFFECTS .....	10
3.1 Numerical Simulations.....	10
3.1.1 Free-field point source model.....	10
3.1.2 HRTF model.....	14
3.2 Objective and Subjective Experiments.....	17
3.2.1 Objective experiment.....	17
3.2.2 Subjective experiment.....	18
4. BANDLIMITED IMPLEMENTATION OF THE CROSSTALK CANCELLATION SYSTEM .....	21
4.1 Four-channel Pseudo QMF Bank.....	21
4.2 Subjective Experiments.....	23
5. BASS ENHANCEMENT BASED ON SENSORLESS VELOCITY OBSERVER .....	28
5.1 Modeling of Moving-coil Loudspeakers .....	28



5.2	Implementation of the Cone Velocity Observer.....	29
5.3	Experimental Investigations.....	30
<b>6.</b>	<b>AUTOMOTIVE VIRTUAL SURROUND AUDIO SYSTEMS.....</b>	<b>32</b>
6.1	Theory and Numerical Simulation .....	32
6.1.1	Equivalent Complex Smoothing Techniques .....	32
6.1.2	Free-field Point Source Model.....	33
6.2	Design Strategies of Automotive Audio Spatializer .....	35
6.2.1	Two-channel Inputs.....	36
6.2.2	5.1-channel Inputs.....	38
6.3	Objective and Subjective Experiments.....	39
6.3.1	Objective Experiments.....	40
6.3.2	Subjective Experiments .....	42
<b>7.</b>	<b>CONCLUSIONS .....</b>	<b>49</b>
<b>8.</b>	<b>FUTURE WORK.....</b>	<b>54</b>
	<b>REFERENCES .....</b>	<b>55</b>
	<b>PUBLICATIONS .....</b>	<b>60</b>



**TABLE LIST**

**Table I. The description of five levels of grade for the subjective localization test. ....61**

**Table II. ANOVA results of the first group .....62**

**Table III. ANOVA results of the second group .....63**

**Table IV. Five-grade impairment scale. ....64**

**Table V. The comparison of computation loading of the fullband CCS and the bandlimited CCS  
with direct convolution. ....65**

**Table VI. The comparison of computation loading of the fullband CCS and the bandlimited CCS  
with fast convolution. ....66**

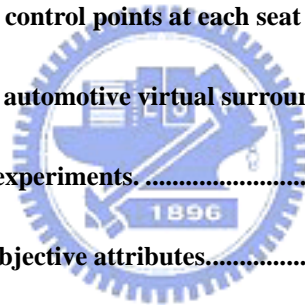
**Table VII. The THD results measured with/without feedback compensation.....67**

**Table VIII. The coordinates of the control points at each seat for three simulation cases. ....68**

**Table IX. The descriptions of four automotive virtual surround processing methods. ....69**

**Table X. The descriptions of four experiments. ....70**

**Table XI. The definitions of the subjective attributes.....71**



**FIGURE LIST**

**Fig. 1** The block diagram of a multi-channel model-matching problem in the CCS design. ....72

**Fig. 2.** The geometry of the free-field point source model.....73

**Fig. 3.** The values of regularization in (a) the free-field point source model and (b) the HRTF model. ....74

**Fig. 4.** The contour plots calculated using the point source model of (a) the condition number of acoustical plant matrix H, (b) the filter gain, and (c) the channel separation.....76

**Fig. 5.** The contour plots of channel separation at the right ear calculated using the point source model. (a) 10-deg span. (b) 60-deg span. (c) 120-deg span.....78

**Fig. 6.** The contour plots of average channel separation at the right ear calculated using the point source model. (a) Bandwidth to 1 kHz. (b) Bandwidth to 6 kHz. (c) Bandwidth to 20 kHz. ....80

**Fig. 7.** Two sweet spot definitions calculated using the point source model for 1, 6, and 10 kHz bandwidths. (a) Relative sweet spot. (b) Absolute sweet spot. ....81

**Fig. 8.** The contour plots calculated using the HRTF model of (a) the condition number of acoustical plant matrix H, (b) the filter gain, (c) the channel separation, and (d) the uncompensated natural channel separation. ....83

**Fig. 9.** The contour plots of channel separation measured at the right ear of the acoustic manikin. (a) 10-deg span. (b) 60-deg span. (c) 120-deg span. ....85

**Fig. 10.** The contour plots of band-average channel separation measured at the right ear of the acoustic manikin. (a) 1 kHz bandwidth. (b) 6 kHz bandwidth. (c) 20 kHz bandwidth. 87

**Fig. 11.** Two sweet spot definitions calculated using the HRTF model for 1, 6, and 10 kHz bandwidths. (a) Relative sweet spot. (b) Absolute sweet spot. ....88

**Fig. 12.** Photo of the experimental arrangement.....89

**Fig. 13.** Channel separations measured at the right ear of the acoustic manikin. The dotted lines, solid lines, and dashed lines represent 10-deg, 60-deg, and 120-deg spans, respectively. (a) In the nominal position ( $x = 0\text{cm}$ ). (b) Rightward 5cm displacement. (c) Rightward 10cm displacement. ....91

<b>Fig. 15. Results of the subjective localization test of azimuth angles with 5-cm head displacement to the right. (a) 10-deg. (b) 60-deg. (c) 120-deg. span. ....</b>	<b>93</b>
<b>Fig. 16. Results of the subjective localization test of azimuth angles with 10-cm head displacement to the right. (a) 10-deg (b) 60-deg (c) 120-deg span. ....</b>	<b>94</b>
<b>Fig. 17. Means and spreads (with 95% confidence intervals) of the grades for three kinds of head displacements. (a) Grades of the 10-deg arrangement (solid line) and the 120-deg arrangement (dotted line). (b) Grades of the 60-deg arrangement (solid line) and the 120-deg arrangement (dotted line). ....</b>	<b>95</b>
<b>Fig. 18 The experimental configuration. ....</b>	<b>96</b>
<b>Fig. 19 The frequency responses of the plants including ipsilateral (solid line) and contralateral paths (dotted line). ....</b>	<b>97</b>
<b>Fig. 20 (a) The frequency responses of <math>Q_{11f}</math> (solid line) and <math>Q_{12f}</math> (dotted line). (b) Natural channel separation (solid line) and compensated channel separation (dotted line). ....</b>	<b>98</b>
<b>Fig. 21 The block diagram of the bandlimited CCS. ....</b>	<b>99</b>
<b>Fig. 22 The magnitude responses of (a) Prototype FIR filter and (b) Analysis bank. ....</b>	<b>100</b>
<b>Fig. 23 (a) The frequency responses of <math>Q_{11b}</math> (solid line) and <math>Q_{12b}</math> (dotted line). (b) Natural channel separation (solid line) and compensated channel separation (dotted line). ....</b>	<b>101</b>
<b>Fig. 24 Results of the subjective localization test of azimuth. (a) Fullband CCS. (b) Bandlimited CCS. ....</b>	<b>102</b>
<b>Fig. 25 Means and spreads (with 95% confidence intervals) of the grades for two kinds of CCS approaches. (a) Grades of the source localization experiment. (b) Grades of the sound quality tests. ....</b>	<b>103</b>
<b>Fig. 26 Shuffler filter structure for 2x2 CCS. ....</b>	<b>104</b>
<b>Fig. 27 Hybrid control architecture. ....</b>	<b>105</b>
<b>Fig. 28 Electromechanical analogous circuit. ....</b>	<b>106</b>
<b>Fig. 29 The experimental arrangement of the cone velocity observer. ....</b>	<b>107</b>
<b>Fig. 31 Block diagram of the down-sampling procedure. ....</b>	<b>109</b>
<b>Fig. 32 Frequency responses of cone velocity: proposed observer (solid line) vs. laser vibrometer</b>	

(dotted line).....	110
<b>Fig. 33</b> Frequency responses of sound pressure: uncompensated (solid line) vs. compensated (dotted line).....	111
<b>Fig. 34</b> The geometry of the free-field point source model.....	112
<b>Fig. 35</b> The contour plots of the performance of Case I. (a) Front-right seat (b) Rear-right seat. ....	113
<b>Fig. 36</b> The contour plots of the performance of Case II. (a) Front-right seat (b) Rear-right seat. ....	114
<b>Fig. 37</b> The contour plots of the performance of Case III. (a) Front-right seat (b) Rear-right seat. ....	115
<b>Fig. 38</b> The block diagram of Method I. ....	116
<b>Fig. 39</b> The block diagrams of (a) comb filter and (b) overall reverberator.....	117
<b>Fig. 40</b> The block diagram of Method II.....	118
<b>Fig. 41</b> The block diagram of Method III.....	119
<b>Fig. 42</b> The block diagram of Method IV.....	120
<b>Fig. 43</b> (a) A 2-liter sedan. (b) The photo of the experimental arrangement.....	121
<b>Fig. 44</b> The frequency responses of the plants at the front-left seat with (a) Front loudspeaker and (b) Rear loudspeaker. The dotted lines represent original measurement and solid lines represent smoothed measurement.....	122
<b>Fig. 45</b> The impulse responses of the plants at front-left seat. (a) Original measurement from front loudspeakers. (b) Original measurement from rear loudspeakers. (c) Smoothed measurement from front loudspeakers. (d) Smoothed measurement from rear loudspeakers. ....	124
<b>Fig. 46</b> Frequency responses of the inverse filters at front-left seat. (a) For front sound image. (b) For rear sound image.....	125
<b>Fig. 47</b> Impulse responses of the inverse filters at front-left seat. (a) For front sound image. (b) For rear sound image.....	126
<b>Fig. 48</b> The frequency responses of the virtual sound images. The solid lines represent desired	

	responses M and the dotted lines represent the multi-channel filter-plant product HC. (a) $\pm 30^\circ$ HRTF (b) $\pm 110^\circ$ HRTF.....	127
<b>Fig. 49</b>	The frequency responses of the plants at rear-right seat with (a) Front loudspeaker and (b) Rear loudspeaker. The dotted lines represent original measurement and solid lines represent smoothed measurement. ....	128
<b>Fig. 50</b>	The impulse responses of the plants at rear-right seat. (a) Original measurement from front loudspeakers. (b) Original measurement from rear loudspeakers. (c) Smoothed measurement from front loudspeakers. (d) Smoothed measurement from rear loudspeakers. ....	130
<b>Fig. 51</b>	Frequency responses of the inverse filters at rear-right seat. (a) For front sound image. (b) For rear sound image.....	131
<b>Fig. 52</b>	Impulse responses of the inverse filters at rear-right seat. (a) For front sound image. (b) For rear sound image.....	132
<b>Fig. 53</b>	The frequency responses of the virtual sound images. The solid lines represent desired responses M and the dotted lines represent multi-channel filter-plant product HC. (a) $\pm 30^\circ$ HRTF (b) $\pm 110^\circ$ HRTF.....	133
<b>Fig. 54</b>	The means and spreads (with 95% confidence intervals) of the grades in Exp. I. (a) The first four attributes (b) The last four attributes.....	134
<b>Fig. 56</b>	The means and spreads (with 95% confidence intervals) of the grades in Exp. III. (a) The first four attributes (b) The last four attributes.....	137
<b>Fig. 57</b>	The means and spreads (with 95% confidence intervals) of the grades in Exp. IV. (a) The first four attributes (b) The last four attributes.....	138

## 1. INTRODUCTIONS

---

The central idea of spatial audio reproduction is to synthesize a virtual sound image. The listener perceives as if the signals reproduced at the listener's ears would have been produced by a specific source located at an intended position [1], [2]. This attractive feature of spatial audio lends itself to an emerging audio technology with promising application in mobile phone, personal computer multimedia, video games, home theater, car audio, etc.

The rendering of spatial audio is either by headphones or loudspeakers. Headphone reproduction is straightforward, but suffers from several shortcomings such as in-head localization, front-back reversal, and discomfort to wear. While loudspeakers do not have the same problems as the headphones, another issue adversely affects the performance of spatial audio rendering using loudspeakers. The issue frequently encountered in loudspeaker reproduction is the crosstalk in the contralateral paths from the loudspeakers to the listener's ears, which may obscure source localization. To overcome the problem, crosstalk cancellation systems (CCS) that seek to minimize, if not totally eliminate, the crosstalk have been studied extensively by researchers [3–8]. Various inverse filtering approaches were suggested for designing multi-channel pre-filters for CCS.

Notwithstanding the preliminary success of CCS in academic community, a problem seriously hampers the use of CCS in practical applications. The problem stems from the limited size of the so-called “sweet spot” in which CCS remains effective. The sweet spots are generally so small especially at lateral side that a head movement of a few centimeters would completely destroy the cancellation performance. Two kinds of approach can be used to address this problem – the adaptive design and the robust design. An example of adaptive CCS with head-tracker was presented in the work of Kyiakakis *et al* [9], [10]. This approach

dynamically adjusts the CCS filters by tracking the head position of the listener using optical or acoustical sensors. However, the approach has not been widely used because of the increased hardware and software complexity of the head tracker. On the other hand, instead of dynamically tracking the listener's head, an alternative CCS design using fixed filters can be taken to create a "widen" sweet spot that accommodates larger head movement. Ward and Elko in Bell Labs have conducted a series of insightful analysis of the robustness issue of CCS. In their paper on this topic in 1998, robustness of a two-channel stereo loudspeaker ( $2 \times 2$ ) CCS was investigated using weighted cancellation performance measure at the pass zone and stop zone, respectively [11]. In the other paper by the same authors in 1999, robustness issue of a  $2 \times 2$  CCS was revisited using a different measure, the condition number, which focuses more on numerical stability during matrix inversion, in the presence of noise in data and/or perturbations to system properties [12]. Yet, in another paper by Ward, a joint least squares optimization method is employed to obtain a CCS that is robust to head misalignment [13]. The above-mentioned research winds up with a simple but important conclusion that the optimal loudspeaker spacing should be inversely proportional to the operating frequency. Along the line of robust CCS design, a celebrated "stereo dipole" configuration was suggested by Kirkeby, Nelson, and Hamada [14], [15]. In their arrangement, two loudspeakers are closely spaced with only  $10^\circ$  span. Their analysis of robustness of CCS also focused primarily on numerical stability in relation to the errors in matrix inversion. The consistent finding of these studies was that the optimal loudspeaker spacing is inversely proportional to the operating frequency. Since the optimal spacing is frequency dependent, a multidrive configuration of the optimal source distribution (OSD) system, comprising pairs of loudspeakers with various spacing, was suggested to deal with crosstalk for different frequency bands [16]. Another



multidrive CCS design was also developed by Bai *et al.*, based on the genetic algorithm and array signal processing [17]. Their approach requires no crossover circuits as in the OSD system.

According to Gardner, loudspeakers spaced apart tend to yield a smaller equalization zone than loudspeakers spaced closely [18]. However, the improvement is predominantly along the front-back axis and the equalization zone widens only slightly when the speakers are positioned closely together. One disadvantage of close spacing is the lack of natural high frequency separation due to head shadowing. Another problem is that small head rotation will cause both speakers to fall on the same side so that the panning mechanism fails.

Thus far, there are pros and cons in the closely spaced CCS. The question of which kind of loudspeaker arrangement is the best has been puzzling people for quite some time. It is worth exploring further the underlying physical insights from all possible angles. This motivates the current research to undertake a comprehensive study in a hope to resolve this optimal CCS problem more conclusively. In Gardner's work, the head-related transfer functions (HRTF) were measured in the MIT Media Lab [19], [20] and subjective listening tests were conducted. However, only the crosstalk below 6 kHz was considered to result in a bandlimited CCS design. Furthermore, the robustness of CCS to head misalignment was discussed in depth by Takeuchi and Nelson [15]. In both works, only two listening spans including 10-deg and 60-deg spans were investigated. On the other hand, the emphasis of this work is placed on the analysis of the effects of listening angles on CCS in terms of not only robustness but also performance. There are several special features in this work. First, not only the robustness but also the performance of CCS is examined with the aid of a more comprehensive set of indices. Second, two kinds of definitions of sweet spot are employed for assessment of robustness. Third, the present work

considers the entire audible 20 kHz band in which the listener's head may provide natural separation for certain loudspeaker arrangements. Fourth, apart from the objective physical tests, subjective listening tests are conducted to practically assess the CCS arrangements with different listening angles. The results of subjective tests will be validated by using Analysis of Variance (ANOVA) test. Although the last three points have been investigated in [15] and [18], this study examines the design issues in further detail and in some cases reaches different conclusions than the previous research. The intention is to establish a sustainable configuration of CCS that best reconciles the separation performance and the robustness against lateral head movement, not only in theory but also in practice.

Besides sweet spot issue, another one is the computation loading. It usually needs long-tapped filters to achieve excellent performance, especially in a reverberant room. An efficient method of bandlimited implementation based on the subband approach is presented in Sec. 4. In considering the robustness against uncertainties of HRTFs and head movement and head shadowing effect at high frequencies, the proposed CCS is bandlimited to frequencies below 6 kHz [18]. That is, the CCS only functions at low frequencies and the binaural signals are directly passed through at high frequencies. The bandlimited implementation approach suggested in [18] is more computationally demanding due to its fixed operating rate. In this work, we adopted a subband filtering technique based on a cosine modulated Quadrature Mirror Filter (QMF) bank [21]. In this design, the approximated perfect reconstruction condition is fulfilled and the CCS is operated at low rate. Therefore, it can use more effort at low frequencies for characteristics of human perceptual hearing. In order to verify the proposed CCS, subjective listening experiments were conducted to compare it to the traditional CCS. The results of subjective tests will be validated by using ANOVA. The intention is to develop the CCS with light computation loading that

performs comparably well as the fullband CCS.

In addition, since the subwoofer channel plays an important role in watching DVDs and listening to music, a bass enhancement system based on a sensorless cone velocity observer is proposed in Sec. 5. There are two features in this system. One is that the cone velocity observer requires no sensor. The other is that a hybrid control architecture comprising a feedback controller and a feed-forward controller is employed. Experimental results are discussed.

At last, this technique is extended to multi-channel inverse filtering for automotive virtual surround audio system. In recent years, car electronics has received considerable attention and is regarded as the fourth 'C' industry in addition to the 3C industries (Computer, Communication, and Consumer electronics). Research efforts are currently directed toward new applications in car electronics, including audio/video entertainment, global positioning system (GPS), mobile communication, active safety control, intelligent engine control, and so forth. As opposed to the traditional audio entertaining system comprised of a radio set and a cassette recorder, watching TV or Digital Versatile Disc (DVD), playing video games and even conducting video conference in a car becomes reality nowadays, owing to the rapid advances of the flat panel displays and digital telecommunication technologies.

With the increased proliferation nowadays of automotive audiovisual systems, the interior of a car is also known as a notorious listening environment due to reflections in a confined space, non-ideal user/loudspeaker positions, and ambient noise, etc. This motivates the current research to develop automotive audio spatializers to create a proper listening environment for vehicles. In addition to conventional multi-channel panning techniques [22], there are two advanced methods for spatial audio rendering: binaural audio [1]–[18] and wave field synthesis (WFS)

[23]–[26]. Binaural audio is usually intended for one user using a pair of stereo loudspeakers. This approach, however, suffers from the limited size problem of the so-called “sweet spot” in which the system remains effective [11]–[18]. In the other extreme, the WFS technique is ideally immune from the sweet spot problem and the listeners are free to move in the reproduction area. However, considerable coverage of WFS in academia has not lead to widespread commercial adoption of this technique. The key issue is that large number of loudspeakers, and hence complex processing, is required in the use of this approach, which limits its implementation in practical systems. Pragmatic approaches will be presented in this study as a compromise between binaural audio and WFS.

Although spatial audio reproduction has been studied extensively by researchers, little can be found for automotive applications with regards to this technology. By contrast, there are already some luxury cars in the market place which are equipped with multi-channel surround system. These systems are usually comprised of many high-quality loudspeakers alongside digital audio processors, e.g., Lexicon’s LOGIC 7™ [27], Dolby’s® Prologic II [28], and SRS® Labs’ SRS Automotive™ [29]. Logic 7 and Prologic II are upmixers for extending 2-5.1-channel systems. Bose® AudioPilot® [30], and Bang & Olufsen advanced sound system [31] can automatically adjust the volume according to the background noise. Crockett *et al.* pointed out new trends in automotive audio technology and suggested methods to improve stereo imaging for off-center listeners [32]. Although many commercial systems have emerged, they are mostly based on panning and equalization methods. Few if any have addressed the spatial audio rendering problem for vehicles using more sophisticated and accurate approaches. This paper aims at rendering sound fields in a car environment using various inverse filtering and up/down mixing techniques. These approaches are targeted at less expensive cars in which only limited number of

loudspeakers is available. The proposed system can handle two kinds of audio input: 2-channel content in CD and MP3 format and 5.1-channel content in DVD and Digital Video Broadcasting (DVB) format.

This paper presents several approaches of automotive spatial audio for various passenger sitting modes. Multi-channel inverse filtering in conjunction with up/down mixing is employed to design audio spatializers for 2-channel and 5.1-channel inputs. Sweet spot analysis is conducted using the free-field point source model. Although the simulated conditions are simplified from realistic scenarios, it shows the effects of head movement on rendering performance. The proposed approaches have been implemented on a real car using a fixed-point digital signal processor (DSP) and the loudspeakers installed in the car. Listening tests were conducted for comparing the presented virtual surround systems. The results of subjective tests were processed using multivariate analysis of variance (MANOVA) [33] and design strategies are discussed.



## 2. MULTI-CHANNEL INVERSE FILTERING FOR CCS FROM A MODEL-MATCHING PERSPECTIVE

---

The CCS aims to cancel the crosstalk in the contralateral paths from the multi-channel loudspeakers to the listener's ears so that the binaural signals are reproduced at two ears like those reproduced using headphones. This problem can be viewed from a model-matching perspective, as shown in Fig. 1. In the block diagram,  $\mathbf{x}(z)$  is a vector of  $B$  program input signals,  $\mathbf{v}(z)$  is a vector of  $S$  loudspeaker input signals, and  $\mathbf{e}(z)$  is a vector of  $L$  error signals.  $\mathbf{M}(z)$  is an  $L \times B$  matrix of matching model,  $\mathbf{H}(z)$  is an  $L \times S$  plant transfer matrix, and  $\mathbf{C}(z)$  is a  $S \times B$  matrix of the CCS filters. The  $z^{-m}$  term accounts for the modeling delay to ensure causality of the CCS filters. Let us neglect the modeling delay for the moment; it is straightforward to write down the input-output relationship:

$$\mathbf{e}(z) = [\mathbf{M}(z) - \mathbf{H}(z)\mathbf{C}(z)]\mathbf{x}(z) \quad (1)$$

For arbitrary inputs, minimization of the error output is tantamount to the following optimization problem,

$$\min_{\mathbf{C}} \|\mathbf{M} - \mathbf{H}\mathbf{C}\|_F^2 \quad (2)$$

where  $F$  symbolizes the Frobenius norm [34]. For an  $L \times B$  matrix  $\mathbf{A}$ , Frobenius norm is defined as

$$\|\mathbf{A}\|_F^2 = \sum_{b=1}^B \sum_{l=1}^L |a_{lb}|^2 = \sum_{b=1}^B \|\mathbf{a}_b\|_2^2, \quad \mathbf{a}_b \text{ being the } n\text{th column of } \mathbf{A}. \quad (3)$$

Hence, the minimization problem of Frobenius-norm can be converted to the minimization problem of 2-norm by partitioning the matrices into columns. Assume that  $\mathbf{H}$  is of full column rank and there is no coupling between the columns of the resulting matrix  $\mathbf{C}$  which approximates the inverse of  $\mathbf{H}$ , the minimization of the square of the Frobenius norm of the entire matrix  $\mathbf{H}$  is tantamount to minimizing the square of each column independently. Therefore, Eq. (2) can be equal to the

following equation.

$$\min_{\mathbf{c}_b, b=1,2,\dots,B} \sum_{b=1}^B \|\mathbf{H}\mathbf{c}_b - \mathbf{m}_b\|_2^2, \quad (4)$$

where  $\mathbf{c}_b$  and  $\mathbf{m}_b$  are the  $b$ th column of the matrices  $\mathbf{C}$  and  $\mathbf{M}$ , respectively. The optimal solution of  $\mathbf{c}_b$  can be obtained by applying the method of least-squares to each column:

$$\mathbf{c}_b = \mathbf{H}^+ \mathbf{m}_b, \quad b = 1, 2, \dots, B, \quad (5)$$

where  $\mathbf{H}^+$  is the pseudo-inverse of  $\mathbf{H}$  [34]. This optimal solution in the least-square sense can be assembled a more compact matrix form:

$$[\mathbf{c}_1 \quad \mathbf{c}_2 \quad \dots \quad \mathbf{c}_B] = \mathbf{H}^+ [\mathbf{m}_1 \quad \mathbf{m}_2 \quad \dots \quad \mathbf{m}_B] \quad (6a)$$

or

$$\mathbf{C} = \mathbf{H}^+ \mathbf{M}. \quad (6b)$$

For a matrix  $\mathbf{H}$  with full-column rank ( $L \geq S$ ),  $\mathbf{H}^+$  can be calculated according to

$$\mathbf{H}^+ = (\mathbf{H}^H \mathbf{H})^{-1} \mathbf{H}^H \quad (7)$$

Here,  $\mathbf{H}^+$  is also referred to as the left-pseudoinverse of  $\mathbf{H}$  in that  $\mathbf{H}^+ \mathbf{H} = \mathbf{I}$ .

In practice, the number of loudspeakers is usually greater than the number of ears, i.e.,  $L \leq S$ . Regularization can be used to prevent the singularity of  $\mathbf{H}^H \mathbf{H}$  from saturating the filter gains [35], [36].

$$\mathbf{H}^+ = (\mathbf{H}^H \mathbf{H} + \lambda \mathbf{I})^{-1} \mathbf{H}^H \quad (8)$$

The regularization parameter  $\lambda$  can either be constant or frequency-dependent [37]. It is noted that the procedure to obtain the filter  $\mathbf{C}$  in Eq. (6) is essentially a frequency-domain formulation, inverse Fourier transform along with circular shift (hence the modeling delay) are needed to obtain causal FIR filters.

### 3. THE ANALYSIS OF THE LISTENING ANGLE EFFECTS

---

#### 3.1 Numerical Simulations

In this section, numerical simulations are conducted to examine the effects that listening angle has on CCS. The free-field point source model and HRTFs are employed as the plant models in the simulations. Only lateral misalignment is considered because it has been concluded by the previous research that the lateral misalignment has more pronounced effect on CCS than the other types of head movements [15].

##### 3.1.1 Free-field point source model

For the free-field point source model illustrated in Fig. 2, the plant transfer matrix can be shown to be

$$\mathbf{H} = \frac{\rho_0}{4\pi} \begin{bmatrix} e^{-jk_a l_{LL}} / l_{LL} & e^{-jk_a l_{RL}} / l_{RL} \\ e^{-jk_a l_{LR}} / l_{LR} & e^{-jk_a l_{RR}} / l_{RR} \end{bmatrix} k_a = \omega / c_0, \quad (9)$$

where  $k_a$ ,  $\rho_0$ , and  $c_0$  represent the wave number, the density, and sound speed, respectively. In the simulation, we assume that  $c_0 = 343 \text{ m/s}$ ,  $\rho_0 = 1.21 \text{ kg/m}^3$ ,  $l = 1.4 \text{ m}$ , and the spacing between ears  $\Delta\gamma = 0.1449 \text{ m}$  [38]. In Eq. (9), the lengths are calculated as

$$l_{LL} = \left( (l \cos \theta)^2 + \left( l \sin \theta - \frac{\Delta\gamma}{2} + x \right)^2 \right)^{\frac{1}{2}} \quad (10a)$$

$$l_{LR} = \left( (l \cos \theta)^2 + \left( l \sin \theta + \frac{\Delta\gamma}{2} + x \right)^2 \right)^{\frac{1}{2}} \quad (10b)$$

$$l_{RL} = \left( (l \cos \theta)^2 + \left( l \sin \theta + \frac{\Delta\gamma}{2} - x \right)^2 \right)^{\frac{1}{2}} \quad (10c)$$



$$l_{RR} = \left( (l \cos \theta)^2 + \left( l \sin \theta - \frac{\Delta \gamma}{2} - x \right)^2 \right)^{\frac{1}{2}}. \quad (10d)$$

The CCS filters are obtained by using the aforementioned inverse filtering procedure with constant regularization parameters. Overall, 256 frequencies equally spaced from 20 Hz to 20 kHz on a logarithmic frequency scale are selected. The  $k$ th selected frequency can be represented as

$$f(k) = 10^{\log_{10}^{20} + (\log_{10}^{20000} - \log_{10}^{20}) \frac{k}{256}}, \quad k = 0, 1, \dots, 255, \quad (11)$$

where  $\log_{10}^{20}$  and  $\log_{10}^{20000}$  symbolize the logarithm with base 10 for 20 Hz and 20 kHz, respectively. In the simulation, the power of each CCS filter at different span angles is constrained to be equal, which can be achieved by using different regularization values. The  $2 \times 2$  transfer function matrix is assumed to be symmetric.

The power of CCS filters is defined as

$$\frac{1}{P} \sum_{k=0}^{P-1} \left( |C_{11}(k)|^2 + |C_{12}(k)|^2 \right), \quad (12)$$

where  $C_{11}$  and  $C_{12}$  are diagonal and off-diagonal component of the CCS filter,  $P$  is the number of frequency samples and  $k$  represents the frequency index. The regularization values in each span angle are shown in Fig. 3(a).

Let the overall response of the CCS filters cascaded with the acoustic plant be

$$\mathbf{G} = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} = \mathbf{H}\mathbf{C}. \quad (13a)$$

Channel separation, defined as the ratio of the contralateral response and the ipsilateral response compensated by CCS, is employed as a performance index:

$$CHSP_L(k) = G_{12}(k)/G_{11}(k) \quad \text{or} \quad CHSP_R(k) = G_{21}(k)/G_{22}(k). \quad (13b)$$

Figure 4(a) shows the contour plot of the condition number of the plant matrix  $\mathbf{H}$  in the nominal center position ( $x = 0$ ). The  $x$ -axis is the listening angles in degrees and the  $y$ -axis is logarithmic frequency in Hz. Condition number in dB is represented by

gray levels. In addition, the contour plots of the filter gain and the channel separation shown in Figs. 4(b) and (c) are plotted versus the same coordinates as in Fig. 4(a). From the plots, condition number follows similar trend to the filter gain. This reveals that there is indeed a tradeoff between numerical stability and separation performance. Specifically, large condition number leads to high filter gain. This in turn calls for regularization to restrain the filter gain at the compromise of some performance.

Another issue of CCS is concerned with the *ringing frequency* given by [15], [16]

$$f_n = \frac{nc}{2\Delta r \sin \theta}, \quad n = 0, 1, 2, \dots \quad (14)$$

Ringling frequencies appear at high frequency particularly for small span arrangement. Suppose the frequency range of our interest is from 100 Hz to 6 kHz. Although the 10-deg span arrangement is well conditioned at frequencies below the intersection of the 6 kHz line and the first ringling, it suffers from the “corner problem,” where poor conditioning and high gain arise at low frequencies and small spans. This is to be expected because the acoustic plants are almost identical in magnitude and phase when the listening angle becomes exceedingly small.

Figures 5(a) to (c) show the contour plots of channel separation at the right ear for three span angles ( $2\theta$ ), 10, 60, and 120 degrees, respectively. The span of 10 degrees and 60 degrees are selected because they correspond to stereo dipole and ITU standard [39]. The  $x$ -axis is the lateral head displacements in centimeters and the  $y$ -axis is logarithmic frequency in Hz. Channel separation in dB is represented by gray levels. The darker the gray level, the better the separation performance. From the contour plot, it can be seen that the pattern becomes progressively complicated as span angle increases. In the nominal center position, the region of good separation

performance (the dark stripe) extends towards lower frequency limit (near 100 Hz) for the 120-deg span than the frequency limit (above 1 kHz) for the 10 degree span. On the other hand, the region of ringing frequencies (the white stripes for positive head displacements) occurs at lower frequency (600 Hz) for the 120-deg span versus 6 kHz for the 10-deg span. Thus, stereo dipole indeed has the advantage of having a much higher usable frequency limit before hitting the first ringing frequency which could lead to high gain inverse filters. However, it is argued by the authors that stereo dipole also suffers performance problems at low frequencies. These facts also suggest that large span arrangement should be used at low frequency, while small span arrangement should be used at high frequency, as suggested by many previous researchers [11–16].

In order to explore further the effect of listening angle on the separation performance of CCS, an index, average channel separation, is defined as follows:

$$\frac{1}{M_2 - M_1 + 1} \sum_{k=M_1}^{M_2} 20 \times \log_{10} (|CHSP_y(k)|) \quad (\text{dB}) \quad (15)$$

where  $M_1$  and  $M_2$  are the frequency indices of the lower and upper limits, and the subscript  $y$  denotes either  $L$  or  $R$ . In the simulation, the lower frequency limit was selected to be 100 Hz ( $M_1=60$ ) below which the sound is known to be ineffective for localization. The average channel separation in relation to the listening angle and the lateral head displacement is shown with a contour plot in Fig. 6. Figures 6(a) to (c) correspond to the average channel separations for three different frequency upper limits, 1 kHz ( $M_2=145$ ), 6 kHz ( $M_2=211$ ), and 20 kHz ( $M_2=255$ ), respectively. Using small span angle, a wider region of good separation performance (the second darkest stripe) can be attained at the expense of poor performance, especially for extremely small span. For example, Fig. 6(a) shows the 1 kHz-upper-limit average separation, where the lower tip of the second darkest region barely touches the 20-deg span.

The performance of CCS can also be characterized by sweet spot which refers to the region in which the CCS is effective. To be able to better assess the sweet spot quantitatively, two kinds of sweet spot are defined in the work: the absolute sweet spot and the relative sweet spot. The size of absolute sweet spot is defined as two times the maximum leftward displacement that makes the average channel separation go below -12 dB. The size of relative sweet spot is defined with reference to Fig. 6 as two times the maximum leftward displacement for which the average channel separation is degraded by 12 dB as compared to that of the nominal center position ( $x = 0$ ). A value of -12 dB, or 25%, is an empirical value selected from experience. For absolute sweet spot, this value is the minimal requirement for CCS. For relative sweet spot, this value corresponds to the point when the performance drops by 75% from the nominal position. The relative and absolute sweet spots calculated for the point source model are plotted versus span angle in Fig. 7 (a) and (b), respectively. Three curves plotted in each figure correspond to three different bandwidths, 1 kHz, 6 kHz, and 20 kHz. As seen in the Fig. 7(a), the relative sweet spot is increased monotonously as the span is decreased, as predicted by previous researchers. This suggests that small span arrangement is more robust against head misalignment notwithstanding the poor separation performance at the nominal position. However, if the absolute sweet spot is taken as the robustness index, the conclusion is quite different. If this definition of sweet spot is used, the simulation result suggests that the optimal span angle ranges from 80 to 180 degrees.

### 3.1.2 HRTF model

In addition to the point source model, a more sophisticated model based on HRTF is employed in the simulation to better account for the diffraction and shadowing effects due to the head, ears and torso. The HRTF database measured by

MIT Media Lab was employed. In the nominal position, the plant transfer function matrix is written as

$$\mathbf{H} = \begin{bmatrix} H_{\theta}^i & H_{\theta}^c \\ H_{\theta}^c & H_{\theta}^i \end{bmatrix}, \quad (16)$$

where  $\theta$  is the span angle and the superscript i and c refer to ipsilateral and contralateral side, respectively. As the head moves to the right by  $x$  centimeters, the plant matrix is no longer symmetric and should be modified. The azimuth angle should be modified according to

$$\theta_L = \tan^{-1} \frac{l \sin \theta_{L_0} + x}{l \cos \theta_{L_0}} \quad (17a)$$

$$\theta_R = \tan^{-1} \frac{l \sin \theta_{R_0} - x}{l \cos \theta_{R_0}}, \quad (17b)$$

where  $\theta_{L_0}$  and  $\theta_{R_0}$  are the angles in the nominal position, i.e.  $x = 0$ . Linear interpolation is called for when the angle is not a multiple of five-deg interval as the database was originally organized [18]. In addition to angles, the magnitudes and phases are also adjusted to account for attenuation and delay due to distance change.

Thus,

$$\mathbf{H} = \begin{bmatrix} H_{\theta_L}^i & H_{\theta_R}^c \\ H_{\theta_L}^c & H_{\theta_R}^i \end{bmatrix} \begin{bmatrix} \frac{l_{LL_0}}{l_{LL}} e^{\frac{-j\omega(l_{LL} - l_{LL_0})}{c}} & \frac{l_{RL_0}}{l_{RL}} e^{\frac{-j\omega(l_{RL} - l_{RL_0})}{c}} \\ \frac{l_{LR_0}}{l_{LR}} e^{\frac{-j\omega(l_{LR} - l_{LR_0})}{c}} & \frac{l_{RR_0}}{l_{RR}} e^{\frac{-j\omega(l_{RR} - l_{RR_0})}{c}} \end{bmatrix}, \quad (18)$$

where the subscript “0” refers to the nominal position.

The contour plots of the condition number, filter gain, and channel separation are shown in Fig. 8 (a) to (c). The uncompensated natural channel separation is also shown in Fig. 8(d) for reference, where the effect of head shadowing is clearly visible. By and large, the results of point source and HRTF follow similar trend except one

important distinction. Because of head shadowing at high frequencies, ringing does not show up in the HRTF results as pronouncedly as in the point source model except a constant ringing around 8-10 kHz due to concha dip which is almost independent of span. The operation zone of HRTF is thus bounded from above by the concha dip, in contrast to the point source case that is bounded from above by the first ringing. This suggests that large span arrangement seems to provide better numerical stability with larger useful frequency range than the small span arrangement. The separation performance at high frequencies for large spans is also better (reflected by more dark areas) than that of the small span owing to natural separation provided by head shadowing.

The contour plots of channel separation versus displacement and frequency are shown in Figs. 9 (a) to (c), corresponding to span angles 10, 60, and 120 degrees, respectively. The trends of this result are largely the same as that of the point source model. The separation performance at low frequencies is still not good for the 10-deg span (Fig. 9(a)). Figures 10(a) to (c) show the contour plots of average channel separation versus displacement and span angle for frequency bandwidth, 1 kHz, 6 kHz, and 20 kHz, respectively. The trend of the HRTF result is similar to that of the point source result if only a narrow bandwidth, e.g., 1 kHz, is considered (Fig. 6(a) vs. Fig. 10(a)). However, if average separation performance is calculated for a larger bandwidth, e.g. 20 kHz, the results turn out to be quite different. The average performance is poor for extremely small spans. The region of good performance (the darkest strip) is mainly located around the median span area, say, from 100 to 160 degrees. This difference of conclusion with the previous point source model is again due to the fact that the head shadowing effect will come into play at high frequencies.

The relative and absolute sweet spots, as defined previously in the point source simulation, are calculated for the HRTF model in three different bandwidths, 1 kHz, 6

kHz, and 20 kHz, as shown in Figs. 11 (a) and (b). Similar to the point source results, the relative sweet spot is increased monotonously as the span is decreased, which suggests that small span arrangement is relatively robust against head misalignment notwithstanding the poor separation performance at the nominal position. On the other hand, the results of absolute sweet spot suggest that arrangements with listening angle ranging from 120 to 150 degrees (the intersection of bandwidth of 6 kHz and 20 kHz in Fig 11(b)) seem to be good choices.

### **3.2 Objective and Subjective Experiments**

The forgoing simulation results suggest that the optimal listening angle range from 120 to 150 degrees. This observation is further examined in a series of objective and subjective experiments. Three loudspeaker arrangements with 10-deg, 60-deg, and 120-deg spans were compared in the experiments. The 10-deg span represents stereo dipole. The 60-deg span is suggested in the ITU standard of multi-channel stereophonic system [39]. The 120-deg span represents the optimal span previously found in the simulation. All experiments were carried out in an anechoic room, as shown in Fig. 12.

#### **3.2.1 Objective experiment**

This experiment employed a 5.1-channel loudspeaker system, Inspire 5.1 5300 of Creative, and a digital signal processor (DSP), Blackfin-533, of Analog Device. The microphones and the preamplifier used are GRAS 40AC and GRAS 26AM. The plant transfer function matrixes were measured on an acoustical manikin, KEMAR (Knowles Electronics Manikin for Acoustic Research) along with the ear model, DB-065.

The designed CCS filters were implemented on the DSP using 512-tapped FIR

filters. The performance of CCS was evaluated in terms of channel separation. Figure 13(a) shows the right-ear channel separation at the nominal position with three span angles. The  $x$ -axis and  $y$ -axis represent frequency in Hz and magnitude in dB, respectively. The dotted line, the solid line, and the dashed line signify  $10^\circ$ ,  $60^\circ$ , and  $120^\circ$  span angles, respectively. The results of Figs. 13(b) and (c) were obtained for the cases when the manikin was moved to the right by 5 and 10 cm. Notable of these results is that the 10-deg span performed badly at the frequencies below 1 kHz. The separation performance significantly degraded by as much as 15 dB as the head moved to the right by 5 cm irrespective of which span was used. As the head was displaced by 10-cm, CCS failed almost completely, except at high frequencies that the large 120-deg span arrangement still maintained natural separation because of head shadowing.



### 3.2.2 Subjective experiment

For the purpose of comparing the CCS with different span angles, a subjective listening experiment of source localization was undertaken in the anechoic room. Eleven subjects participated in the test. The listeners were instructed to sit at three positions: the nominal position, 5-cm displacement to the right, and 10-cm displacement to the right. In order to ensure that each listener sat at the same designated position, the test subjects were asked to rest their chins on a steel frame. The height of the listener's ear was 120 cm which is the same height as the loudspeaker. A pink noise was used as the test stimulus whose bandwidth ranges from 20 Hz to 20 kHz and the reproduction level was 95 dB. Each stimulus was played 5 times in 25-ms duration with 50-ms silent interval. Virtual sound images at 12 pre-specified directions on the horizontal plane with increment  $30^\circ$  azimuth are rendered by using HRTFs. Listeners were well trained by playing the stimuli of all



angles prior to the test. The listeners were asked to report the perceived direction of source in the range (-180, 180] with 30-deg interval. Experiments were divided into two groups: 10 degrees versus 120 degrees and 60 degrees versus 120 degrees. The experiments were blind tests in that stimuli were played randomly without informing the subjects the source direction. One session of test lasts 15-20 minutes.

The results of localization-test are shown in terms of target angles versus judged angles in Figs. 14-16, corresponding to the cases of nominal position, 5-cm displacement to the right, and 10-cm displacement to the right. In each figure, subplot (a) to (c) refer to the 10°, 60°, and 120° spans, respectively. The size of each circle is proportional to the number of the listeners who localized the same perceived angle. The 45-deg line indicates the perfect localization. It is observed from the results that the subjects tend to localize the sources within  $\pm 30$  degrees about the center line using the 10-deg span arrangement, especially when there is head displacement. On the other hand, the 60-deg span and the 120-deg span were found to be effective in localizing good frontal images and rear images albeit some front-back reversals. Localization error increases with head displacement irrespective of which span arrangement was used. The 10-deg span seemed to have difficulty localizing sources outside the subtending angle because the separation performance in low frequencies is too poor in small span arrangement to maintain proper spatial cues such as interaural time difference (ITD) which works only under 1 kHz. In contrast, the arrangement with large span appear to be more robust than the small span because head shadowing and panning effect help to provide localization effect to certain degree even if CCS breaks down.

To justify the finding, a two-way analysis of variance (ANOVA) on the subjective localization result in relation to span and displacement was conducted. These results were preprocessed into five levels of grade, as described in Table I.

The ANOVA results are summarized in Tables II and III for the first and the second groups, respectively. Figure 17(a) shows the means and spreads (with 95% confidence intervals) of the grades for three kinds of head displacements, where solid line and dotted line represent the 10-deg span and the 120-deg span, respectively. From the means, the 120-deg span consistently outperformed the 10-deg span for all kinds of head displacements. ANOVA output in Table II reveals that the span effect is statistically significant ( $F=80.612$ ,  $p=0.001$ ). The performance degrades progressively as head displacement increases, which is also statistically significant ( $F=9.104$ ,  $p=0.001$ ). No significant interaction was found between span and displacement. Figure 17(b) shows the means and spreads of the grades for three kinds of displacements, where solid line and dotted line represent 60-deg and 120-deg span, respectively. ANOVA output in Table III reveals that performance degrades progressively as head displacement increases ( $F=8968$ ,  $p=0.001$ ). However, the difference of performance of two loudspeaker spans is found statistically insignificant ( $F=0.026$ ,  $p=0.8712$ ). This does not seem to agree with the prediction of the previous simulation that the 120-deg span should perform slightly better than the 60-deg span. It is suspected that the enormous span of 120-deg arrangement is actually quite detrimental to localizing sources at the center position, especially when CCS beaks down. Experience shows overly large angle arrangements seem to have difficulties in positioning images at the center region. In fact, some of the test subjects reported that it sounded like there was an opening of sound field in the front. This offsets somewhat the expected performance gain of CCS using large span arrangement.

## 4. BANDLIMITED IMPLEMENTATION OF THE CROSSTALK CANCELLATION SYSTEM

---

Bandlimited implementation is chosen in this work for several reasons. First, the computation loading is too high to afford a fullband (0~20 kHz) implementation. For the example of the stereo loudspeaker considered herein, the CCS would contain 4 filters. If each filter has 3000 taps, the convolution would require  $1.2 \times 10^4$  multiplications and additions per sample interval. Except for special-purpose DSP engine, real time implementation for a fullband CCS is usually prohibitive for the sampling rate commonly used in audio processing, e.g., 44.1 kHz or 48 kHz. Second, at high frequencies, the wavelength could be much smaller than a head width. Under this circumstance, the CCS would be extremely susceptible to misalignment of the listener's head and uncertainties involved in HRTF modeling. Third, at high frequencies, a listener's head provides natural shadowing for the contralateral paths, which is more robust than direct application of CCS. The CCS in this study is chosen to be bandlimited to 6 kHz (the wavelength at this frequency is approximately 5.6 cm). To accomplish this, a 4-channel pseudo QMF bank is employed to divide the total audible frequency range into subbands for CCS and direct transmission, respectively.

### 4.1 Four-channel Pseudo QMF Bank

The design strategy of subband filter bank employed in this work is the cosine modulated pseudo QMF. In this method, a FIR filter must be selected as the prototype. Using this prototype, an  $M$ -channel maximally decimated filter bank (number of subbands = up/down sampling factor) is generated with the aid of cosine modulation. The maximum attenuation that can be attained by a perfectly reconstructing (PR) cosine modulated filter bank is about 40 dB. Nevertheless, this

PR filter bank would still present an undesirable ringing problem. To alleviate this problem, the PR condition is relaxed in the FIR filter design to gain more stopband attenuation. From our experience, as much as 60 dB attenuation is required for acceptable reproduction.

Based on the method in [21], the following analysis and synthesis filter banks represented by  $g_k(z)$  and  $f_k(z)$ , respectively, are employed to minimize phase distortion and aliasing.

$$g_k(n) = 2p_0(n) \cos\left(\frac{\pi}{M}(k+0.5)\left(n - \frac{N}{2}\right) + \theta_k\right) \quad (19)$$

$$f_k(n) = g_k(N-n), \quad (20)$$

where  $\theta_k = (-1)^k \frac{\pi}{4}$ ,  $0 \leq k \leq M-1$  and  $p_0(n)$ ,  $n=1,2,\dots,N$  are the coefficients of the prototype FIR filter. The remaining problem is how to minimize the amplitude distortion. The distortion function  $T(z)$  for the filter bank is given as [21]:

$$T(z) = \frac{1}{M} \sum_{k=0}^{M-1} F_k(z)G_k(z). \quad (21)$$

Z-transform of Eq. (20) leads to  $F_k(z) = z^{-N} \tilde{G}_k(z)$ , where  $\tilde{G}_k(z)$  is the paraconjugation of  $G_k(z)$ . The distortion function can thus be written in frequency domain as

$$T(e^{j\omega}) = \frac{1}{M} e^{-j\omega N} \sum_{k=0}^{M-1} |G_k(e^{j\omega})|^2. \quad (22)$$

A filter  $P(z)$  is called a Nyquist ( $M$ ) filter if the following condition is met:

$$p(Mn) = \begin{cases} c, & n = 0 \\ 0, & \text{otherwise} \end{cases}, \quad (23)$$

where  $p(n)$  is the impulse response of  $P(z)$  and  $c$  is a constant. In frequency domain

$$\sum_{k=0}^{M-1} P(e^{j(\omega-2\pi k/N)}) = Mc. \quad (24)$$

Equations (22) and (24) indicate that, if  $|G_k(e^{j\omega})|^2$  is a Nyquist ( $M$ ) filter, or equivalently,  $|P_0(e^{j\omega})|^2$  is a Nyquist ( $2M$ ) filter, the magnitude of  $T(z)$  will be flat.

In this QMF design, the Kaiser window is used as the FIR prototype [40]. Given the specifications of transition bandwidth  $\Delta f$  and stopband attenuation  $A_s$ , the parameter  $\beta$  and the filter order  $N$  can be determined according to

$$\beta = \begin{cases} 0.1102(A_s - 8.7) & \text{if } A_s > 50 \\ 0.5842(A_s - 21)^{0.4} + 0.07886(A_s - 21) & \text{if } 21 < A_s < 50 \\ 0 & \text{if } A_s < 21 \end{cases} \quad (25)$$

and

$$N \approx \frac{A_s - 7.95}{14.36\Delta f} \quad (26)$$

An optimization procedure is employed here to make  $P_0(z)\tilde{P}_0(z)$  an approximate Nyquist ( $2M$ ) filter, as posed by the following min-max problem [40]:

$$\min_{\omega_c} \max_{n \neq 0} |p_0(n) * p_0(-n)|_{\downarrow 2M}, \quad (27)$$

where the asterisk  $*$  denotes the convolution operator. Because this is a convex problem, optimal cutoff frequency can always be found [40]. After obtaining the optimal prototype filter, the analysis and synthesis filters are generated according to Eq. (19) and Eq. (20), respectively. The filter bank can be easily implemented with techniques such as polyphase structure or discrete cosine transform (DCT) [21].

## 4.2 Subjective Experiments

In order to compare the performance of the proposed CCS and the fullband CCS, subjective experiments were undertaken in an anechoic room. The experimental arrangement is shown in Fig. 18. This experiment employed a stereophonic two-way loudspeaker system, ELAC BS 103.2. The microphone and the

preamplifier are GRAS 40AC and GRAS 26AM, respectively. The plant transfer function matrices were measured on an acoustical manikin, KEMAR (Knowles Electronics Manikin for Acoustic Research), along with the ear model, DB-065. The frequency responses of the plants are shown in Fig. 19 wherein the solid line and dotted line represent the ipsilateral and the contralateral paths, respectively. Only responses measured on the right ear are shown because of the assumed symmetry. The  $x$ -axis is logarithmic frequency in Hz and the  $y$ -axis is magnitude in dB. The CCS filters with 3000 taps are designed according to the method presented in Sec. 2 with 12 dB threshold. The matrix  $\mathbf{Q}$  is defined as

$$\mathbf{Q} = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} = \mathbf{H}\mathbf{C} \quad (17)$$

This matrix attempts to approximate the model matrix  $\mathbf{M}$  which is set to be an identity matrix here. Figure 20(a) shows the frequency responses of  $Q_{11f}$  and  $Q_{12f}$ , where the subscript  $f$  stands for the fullband method, represented as solid line and dotted line, respectively. After compensation, the ipsilateral magnitude is almost flat from 300 Hz to 8 kHz. Some imperfect match can be seen at low frequencies and at high frequencies because the CCS filter gain is constrained, i.e. large regularization. On the other hand, the contralateral magnitude is degraded to around -40 dB. Channel separation, defined as the ratio of the contralateral response and the ipsilateral response, is employed as a performance index. The channel separation,  $Q_{12f}/Q_{11f}$ , is shown in Fig. 20(b) as the dotted line. The solid line represents the natural channel separation,  $H_{12}/H_{11}$ . As mentioned above, the fullband approach is impractical due to many reasons. The proposed method in this work is bandlimited to 6 kHz with 48 kHz sampling rate. The block diagram of the bandlimited CCS is illustrated in Fig. 21. Through the use of the method presented in Sec. 3, the prototype FIR filter with 120 taps and the analysis bank are plotted in Fig. 22(a) and

22(b), respectively. The CCS only functions at the lowest band and operates at lower sampling rate. The computation loading of an analysis bank or a synthesis bank equals to that of the prototype FIR filter when the polyphase structure is employed. Since CCS operates at low rate, it is able to sample more frequencies at design stage. In the experiment, the tap of the bandlimited CCS is 1500. In other words, the frequency (under 6 kHz) resolution of the bandlimited CCS is twice than that of the fullband CCS. That is, the bandlimited CCS has finer resolution. Figure 23(a) shows the frequency responses of  $Q_{11b}$  and  $Q_{12b}$ , where the subscript t stands for the bandlimited method, represented as solid line and dotted line, respectively. The channel separation,  $Q_{12b}/Q_{11b}$ , is shown in Fig. 23(b) as the dotted line. From Figs. 20(b) and 23(b), we can see that the bandlimited CCS gets better channel separation, especially from 100 Hz to 1 kHz.

Subjective listening experiment includes two parts: the source localization test and the sound quality test. Eleven subjects participated in the test. The listeners were instructed to sit at the position where KEMAR was. In the first part, the test stimulus was a pink noise bandlimited to 20 kHz. Each stimulus was played 5 times in 25-ms duration with 50-ms silent interval. Virtual sound images at 7 pre-specified directions on the right horizontal plane with increment  $30^\circ$  azimuth are rendered by using HRTFs. Listeners were well trained by playing the stimuli of all angles prior to the test. The experiments were blind tests in which stimuli were played randomly without informing the subjects the source direction. The results of localization-test are shown in terms of target angles versus judged angles in Figs. 24(a) and 24(b), corresponding to the cases of fullband CCS and bandlimited CCS. The size of each circle is proportional to the number of the listeners who localized the same perceived angle. The 45-deg line indicates the perfect localization. It is observed from the results that subjects localized well at front (0 degree) and back (180 degrees) no

matter what approach is employed. While the fullband CCS performs well at 30-degree angle, subjects were confused within the range 60°-120°. On the other hand, bandlimited CCS performs slightly better within the range 60°-120°. It is interesting to note that bandlimited CCS exist no back-front reversal problem which means that the subject localizes rear-stimulus to front angle. In addition, a one-way analysis of variance (ANOVA) on the subjective localization result was conducted. These results were preprocessed into five levels of grade, as described in Table I. Figure 25(a) shows the means and spreads (with 95% confidence intervals) of the grades for two kinds of approaches. The mean of the bandlimited CCS is slightly larger than which of the fullband CCS as we observed previously. ANOVA output reveals that two approaches are not statistically significant ( $p=0.2324 > 0.05$ ).

In the second part, the stimulus pre-filtered by the fullband CCS and the bandlimited CCS were treated as the reference and the object, respectively. The “double-blind triple-stimulus with hidden reference” method has been employed in this testing procedure [41]. A listener at a time was involved in three stimuli (“A”, “B”, and “C”) where “A” represented the reference and “B” and/or “C” represented the hidden reference and/or the object. A subject was requested to compare “B” to “A” and “C” to “A” with five-grade impairment scale described in Table IV. The test stimuli contain three types of music including a bass (low frequency), a triangle (high frequency), and a popular song (comprehensive effect). Figure 25(b) shows the means and spreads (with 95% confidence intervals) of the grades for two kinds of approaches. It seems that the fullband CCS earned a slightly higher grade than the subband approach since the fullband CCS was used as the reference. Nevertheless, ANOVA test reveals that the performance difference between two approaches are not statistically significant ( $p = 0.4109 > 0.05$ ).

Here, the proposed method has been validated that it performs comparably well



as the fullband CCS. In Table V, two approaches are compared in terms of computation loading, where MPU and APU represent multiplications and additions per unit time, respectively. The computation loadings are calculated using direct convolution in the time-domain. The computation loading using the proposed subband filtering approach was drastically reduced by approximately eighty percent, as compared to the conventional approach. However, there are still other fast convolution algorithms that can be adopted for efficient implementation. The overlap-add methods of block convolution [42], for example, are compared in the simulation. This method is only used in CCS filters, while the filter bank is still carried out by using direct convolution because of the efficient polyphase implementation. In the procedure of block convolution, the fast Fourier transform is used to realize discrete Fourier transform. Moreover, the number of complex multiplications and additions of the fast Fourier transform is equal to  $N \log_2 N$ , where  $N$  is the number of the transform point. After using block convolution, the results of computation loading are listed in Table VI.

The shuffler method can be applied due to symmetric assumption. The shuffler structure is shown in Fig. 26. It saves around fifty percent of computation [18]. The multi-channel shuffler structure can be found in [43].

## 5. BASS ENHANCEMENT BASED ON SENSORLESS VELOCITY OBSERVER

---

The perception in low frequency (usually below 150 Hz) is very important for not only watching DVDs but also listening to music. It is impossible to reproduce fullband signals, ranging from 20 Hz to 20 kHz, by a single loudspeaker unit perfectly. This is why there is an extra larger loudspeaker in charge of the subwoofer channel in multi-channel audio systems. Due to the importance of subwoofers, a compensation system electronically decreasing the fundamental frequency of the velocity response by 20 Hz is proposed in this section.

The compensation is based on a hybrid control architecture compressing a feedback controller and a feed-forward controller, shown in Fig. 27. The feedback filter and feedforward filter are designed using the quantitative feedback technique (QFT) with an analog circuit and the model matching method with a digital signal processor (DSP), respectively. The feedforward filter is designed according to the cone velocity of loudspeakers that has been long recognized as an important parameter for loudspeaker compensation. In the section, a cone velocity observer that requires no sensor is developed on the basis of state-space estimation. Linear quadratic Gaussian (LQG) theory in conjunction with multirate processing is employed in the design of the observer.

### 5.1 Modeling of Moving-coil Loudspeakers

A moving-coil loudspeaker can be modeled as an equivalent circuit of Fig. 28 using *electro-mechano-acoustical analogy (mobility analogy* in this case). The following definitions are used:

$e_g, R_g$  open-circuit voltage and equivalent resistance of the power amplifier output

$L_E, R_E$  inductance and resistance of the coil measured with the voice coil blocked

$$(u_c=0)$$

- $e, i$  voltage and current of the coil  
 $Bl$  the electromagnetic coupling factor (magnetic flux density  $\times$  coil length)  
 $u_c, f_c$  coil velocity and Lorentz force  
 $M_{MD}$  equivalent mass of the coil and diaphragm  
 $r_{MS}, C_{MS}$  mechanical responsiveness and compliance of the suspension

The sound field radiation loading is neglected since it is small compared to mechanical impedance. With some manipulation, the loudspeaker system can be written in the following state-space form:

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{A}\mathbf{x} + \mathbf{B}u + \mathbf{G}w \\ y &= \mathbf{C}\mathbf{x} + v \end{aligned} \quad (28)$$

where

$$\mathbf{A} = \begin{bmatrix} \frac{-1}{M_{MD}r_{MS}} & \frac{Bl}{M_{MD}} & \frac{-1}{M_{MD}} \\ \frac{-Bl}{L_E} & \frac{R_g + R_E}{L_E} & 0 \\ \frac{1}{C_{MS}} & 0 & 0 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 0 \\ \frac{1}{L_E} \\ 0 \end{bmatrix}^T, \mathbf{C} = [0 \quad 1 \quad 0] \quad (29)$$

$$\mathbf{x} = [u_c \quad i \quad i_{CMS}]^T, y = i, u = e_g, \mathbf{w}(t) \sim (0, \mathbf{Q}), v(t) \sim (0, R)$$

In the equation, it is assumed that  $w(t)$  and  $v(t)$  are zero-mean white noise process with the variances,  $\mathbf{Q} \geq 0, R \geq 0$ .

## 5.2 Implementation of the Cone Velocity Observer

In linear control theory, a state estimator can be constructed if system is observable. It can be verified that the following *observability matrix*

$$\mathbf{Q}_o = [\mathbf{C} \quad \mathbf{C}\mathbf{A} \quad \mathbf{C}\mathbf{A}^2]^T \quad (30)$$

is full rank for the preceding loudspeaker system. Thus the system is observable, which enables the construction of a state observer with the dynamic equation

$$\dot{\hat{\mathbf{x}}} = (\mathbf{A} - \mathbf{LC})\hat{\mathbf{x}} + \mathbf{LCx} + \mathbf{Bu} \quad (31)$$

In LQG design, the observer gain  $\mathbf{L}$  can be obtained by solving the algebraic Riccati equation

$$\mathbf{P} = \mathbf{AP} + \mathbf{PA}^T + \mathbf{GQG}^T - \mathbf{PC}^T R^{-1} \mathbf{CP} \quad (32)$$

$$\mathbf{L} = \mathbf{PC}^T R^{-1}. \quad (33)$$

The observer requires the plant input  $e_g$  and the output  $i$  as two inputs to estimate each state variable. The state equation of the complete observer system can be written as

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{A}'\mathbf{x} + \mathbf{B}'\mathbf{u}' \\ \mathbf{y}' &= \mathbf{C}'\mathbf{x} \end{aligned} \quad (34)$$

where

$$\begin{aligned} \mathbf{A}' &= \mathbf{A} - \mathbf{LC}, \mathbf{B}' = [\mathbf{B} \quad \mathbf{L}], \mathbf{C}' = \mathbf{I}_3 \\ \mathbf{u}' &= [e_g \quad i]^T, \mathbf{y}' = \mathbf{x} \end{aligned} \quad (35)$$

### 5.3 Experimental Investigations

In order to justify the proposed cone velocity observer, experimental investigations were carried out. A 6.5-inch loudspeaker is used in this work and the experimental arrangement is shown in Fig. 29. Figure 30(a) shows the hybrid structure composed of a feedforward  $C_F(z)$  controller and a feedback controller  $C_B(z)$ . Plant  $L(z)$  is our controlled system and the detail is shown in Fig. 30(b) where  $L_S(z)$  and  $O(z)$  is loudspeaker transfer function and the observer controller, respectively. First, the Thiele-Small parameters of the loudspeaker need to be identified using the added mass method. The observer and feedback controller designed by the previously mentioned procedure are implemented in an analog circuit. In addition, the feedforward controller is designed according to the cone velocity measured via the observer and implemented on an ADI BF-533 DSP board. Since our applications are targeted at low frequency range and the sampling rate of the AD

converter is fixed at the rate 48 kHz, we have to reduce the effective sampling rate by using a polyphase technique in multirate signal processing (Fig. 31). With decimation factor 24, the cut-off frequency of the lowpass filter,  $H_L(z)$ , is selected to be 250 Hz. To justify the observer method, the thus obtained cone velocity is compared to that obtained from a laser vibrometer. In Fig. 32, very close agreement between the cone velocities obtained by two methods can clearly be seen. One of the audio applications in which we tested the system was bass enhancement. A feedforward compensation filter was designed to match a frequency-domain velocity template with 6 dB boost and 20 Hz lower fundamental frequency  $f_o$ , relative to the estimated cone velocity. The frequency responses of sound pressure in Fig. 33, where solid line and dotted line represent uncompensated and compensated response respectively, indicate that the target of the observer has been reached quite nicely. The purpose of the feedback control is to increase robustness against plant uncertainties and perturbations. The feedback compensator is designed using QFT and implemented with an analog circuit for fast response. Table VII lists the THD (total harmonic distortion) results measured with/without feedback compensation. It gets a little improvement in THD when the feedback compensation is employed.

## 6. AUTOMOTIVE VIRTUAL SURROUND AUDIO SYSTEMS

---

### 6.1 Theory and Numerical Simulation

#### 6.1.1 Equivalent Complex Smoothing Techniques

It is impractical and not robust to implement the inverse filters based on the measured room response due to its highly complex dynamics and measurement errors associated with it [44]. Some pre-processing should be applied prior to the design of the inverse filters. A simple but elegant way is to smooth the peaks and dips of the acoustic plant using the generalized complex smoothing technique suggested by Hatziantoniou and Mourjopoulos [45]. There are two methods for implementing complex smoothing. The first method, uniform smoothing, is to calculate the impulse response using the inverse FFT of the frequency response. Then, apply a time-domain window to truncate and taper the impulse response, which in effect smoothes out the frequency response. Finally, calculate the ‘smoothed’ frequency response by FFT of the modified impulse response. Alternatively, a nonuniform smoothing method can also be used. This method performs smoothing directly in the frequency domain. The frequency response is circularly convolved with a frequency-dependent window whose bandwidth increases with frequency. This method is based on the notion in psychoacoustics that the spectral resolution of human hearing increases with frequency. The expression of nonuniformly smoothed frequency response is given as [45]

$$H_{\text{ecs}}(k) = H_{\text{cs}}(k) \frac{|H_{\text{ts}}(k)|}{|H_{\text{cs}}(k)|} \quad (36)$$

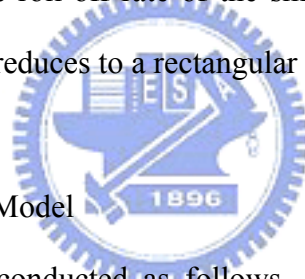
$$H_{\text{cs}}(k) = \sum_{i=k-m(k)}^{k+m(k)} H_{\text{R}}(k) W_{\text{sm}}(i-k+m(k)) + j \sum_{i=k-m(k)}^{k+m(k)} H_{\text{I}}(k) W_{\text{sm}}(i-k+m(k)) \quad (37)$$

$$|H_{\text{ts}}(k)| = \sqrt{\sum_{i=k-m(k)}^{k+m(k)} \left\{ [H_{\text{R}}(k)]^2 + [H_{\text{I}}(k)]^2 \right\} W_{\text{sm}}(i-k+m(k))} \quad (38)$$

where  $k$ ,  $0 \leq k \leq J-1$ , is the frequency index and  $m(k)$  is the smoothing index corresponding to the length of the smoothing window. The smoothing window  $W_{sm}(i)$  is given by

$$W_{sm}(i) = \begin{cases} \frac{1}{2b(m(k)+1)-1}, & i = 0 \\ \frac{b-(b-1)\cos\left[\left(\frac{\pi}{m(k)}\right)(k-J)\right]}{2b(m(k)+1)-1}, & i = 1, \dots, m(k) \\ \frac{b-(b-1)\cos\left[\left(\frac{\pi}{m(k)}\right)k\right]}{2b(m(k)+1)-1}, & i = m(k)+1, \dots, 2m(k) \end{cases} \quad (39)$$

The integer  $m(k)$  can be considered as a bandwidth function by which a fractional octave or any other nonuniform frequency smoothing scheme can be implemented. The variable  $b$  determines the roll-off rate of the smoothing window. As a special case when  $b = 1$ , the window reduces to a rectangular window.



### 6.1.2 Free-field Point Source Model

Sweet-spot analysis is conducted as follows. For simplicity, the reflections from the boundary are neglected and loudspeakers are regarded as point sources. The free-field point source model is illustrated in Fig. 34, where four control points at four seats and the positions of four loudspeakers are indicated. The acoustical plant transfer matrix can be written as

$$\mathbf{H} = \frac{\rho_0}{4\pi} \begin{bmatrix} e^{-jk_a l_{1a}}/l_{1a} & e^{-jk_a l_{1b}}/l_{1b} & e^{-jk_a l_{1c}}/l_{1c} & e^{-jk_a l_{1d}}/l_{1d} \\ e^{-jk_a l_{2a}}/l_{2a} & e^{-jk_a l_{2b}}/l_{2b} & e^{-jk_a l_{2c}}/l_{2c} & e^{-jk_a l_{2d}}/l_{2d} \\ e^{-jk_a l_{3a}}/l_{3a} & e^{-jk_a l_{3b}}/l_{3b} & e^{-jk_a l_{3c}}/l_{3c} & e^{-jk_a l_{3d}}/l_{3d} \\ e^{-jk_a l_{4a}}/l_{4a} & e^{-jk_a l_{4b}}/l_{4b} & e^{-jk_a l_{4c}}/l_{4c} & e^{-jk_a l_{4d}}/l_{4d} \end{bmatrix}, \quad (40)$$

where  $k_a = \omega/c_0$ ,  $\rho_0 = 1.21 \text{ kg/m}^3$ ,  $c_0 = 343 \text{ m/s}$ , and  $l_{pq}$  denote the wave number, the density, sound speed, and the distance between the  $p$ th control point and the  $q$ th source, respectively. The matching model matrix  $\mathbf{M}$  is constructed by calculating the

distances between the control points and the virtual images to be positioned. The inverse filters  $\mathbf{C}$  are obtained by using the preceding inverse filter design procedure. Overall, 256 frequencies equally spaced from 20 Hz to 20 kHz on a logarithmic frequency scale are selected. The  $k$ th selected frequency can be represented as Eq. (11). The errors (measured by 2-norm) between the matching model matrix  $\mathbf{M}$  and the multi-channel filter-plant product  $\mathbf{HC}$  are defined as the performance metric of the sweet spot

$$\mathbf{E}_p(x, y) = \frac{1}{J} \sum_{k=0}^{J-1} \left\| \mathbf{M}_p(k) - \mathbf{H}_p(x, y, k) \mathbf{C}(k) \right\|_2 \quad (41)$$

where  $k$  is the frequency index,  $x$  and  $y$  are physical coordinates,  $J$  is the number of the frequency samples, and subscript  $p$  represents the  $p$ th row of the matrix  $\mathbf{M}$  or  $\mathbf{H}$ . The lower the value of  $E_p(x, y)$ , the better the performance is.

The simulation involves three cases, with the coordinates listed in Table VIII. Case I, one control point is set at the center of each seat, as shown in Fig. 34. The contour plot of the performance metric  $E_p(x, y)$  at the front-right seat is shown in Fig. 35(a). The error between the matching model matrix  $\mathbf{M}$  and the multi-channel filter-plant product  $\mathbf{HC}$  in dB is indicated by gray levels. The circle with 0.2m diameter at the center of the plot represents a human's head. The contour plot of the performance metric  $E_p(x, y)$  at the rear-right seat is shown in Fig. 35(b). It can be observed from these plots, the performance metric increases dramatically away from the control point. The sweet spot takes the shape of an arc due to the influence of the loudspeaker. The sweet spot shown in Fig. 35(a) is long and narrow along the  $-45^\circ$ -line from the vertical, which is unfavorable for the lateral movements. By the same token, Fig. 35(b) reveals that fore-aft movement in the backseat would adversely affects the performance. Figures 36(a) and (b) show the contour plots of the performance metric  $E_p(x, y)$  at the front-right seat and the rear-right seat of Case II,



respectively. The trends of the contour plots resemble those in Figs. 35(a) and (b). The performance metric reaches minimum at two control points around two ears and then increases less abruptly as compared to Case I. The sweet spots defined by the -23 dB contour of Case II are larger than those of Case I. However, it suffers from the same robustness problem of fore-aft movement at the backseat. To overcome the problem, one more control point is added at the position of the nose. Figures 37(a) and (b) represent the results of the front and rear seats, respectively. The results follow similar trends with minimum errors found at the control points. At the back seat, the sweet spot is widened along the  $y$  axis.

In summary, similar trends can be observed among all three cases, which are influenced by the locations of the loudspeakers. Although errors are only negligibly small at the control points in Case I, it suffers from the problem of small sweet spot. Case III can be regarded as a compromise between performance and robustness. Because a car interior is generally full of boundary reflections, it can be anticipated that the size of sweet spot will be further restricted and the robustness of inverse filtering will be further degraded.

## **6.2 Design Strategies of Automotive Audio Spatializer**

In this section, design strategies of automotive audio spatializer will be presented in two categories. One is two-channel content such as CD, MP3, and radio broadcast that can be found in common cars. The other is 5.1-channel content such as DVD and DVB that is becoming popular in recent years. Since an average car is loaded with at least four loudspeakers, the number of transducers is usually not a problem, as compared to home stereo system. The key issue of automotive surround audio is how to render spatial sound field to cope with the reflections in the confined space, with listeners and loudspeakers not in proper positions. By taking

the processing capability of our hardware into account, we have developed several practical processing approaches to address this issue in the context of automotive audio, as will be presented in the sequel. Four processing methods are summarized in Table IX, as detailed in the following presentation.

### 6.2.1 Two-channel Inputs

In traditional automotive audio, the left-input signals are fed to both front-left and rear-left loudspeakers, and the right-input signals are fed to both front-right and rear-right loudspeakers. Balance of the left and right as well as the front and the rear can usually be adjusted. The problem with this approach is that the front and rear channels are too correlated to create natural-sounding surround effects. The paper seeks to develop upmixing algorithms for extending two-channel input to four channels. Upmixing can generally be achieved by two categories of approaches. One approach is decorrelation-based methods, e.g., Prologic II and Logic 7, etc. Another approach is reverberation-based method that is found to be very effective in producing sense of space, especially for small space [46]. In a previous subjective listening test [47], the reverberation-based methods outperformed the decorrelation-based methods in ambient surround effects. Thus, only the reverberation-based upmixing method is adopted in the following discussion.

#### 6.2.1.1 Method I: *Upmixing with Inverse Filtering*

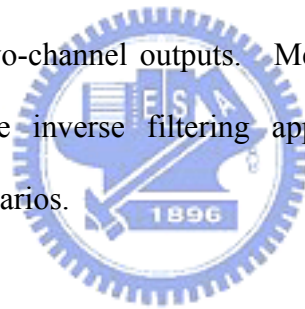
The block diagram of Method I is illustrated in Fig. 38, where two-channel input signals are extended to four channels by the reverberation-based upmixing algorithm and then inverse filtered to produce the outputs. An artificial reverberator is employed to produce the rear surround channels. The artificial reverberator is constructed from 3 parallel comb filters shown in Fig. 39(a) and a 3-layered

nested-allpass filter shown in Fig. 39(b). The parameters in this network are tuned via Genetic Algorithm (GA) [46]. The difference between the left and right input signals is mixed into the rear channel to enhance ambience. The rear-left and rear-right channels are made  $180^\circ$  out of phase.

Prior to the design of inverse filters, the acoustical plants  $\mathbf{H}(z)$  have to be measured. The acoustical plants are the frequency response functions between the loudspeakers and microphones mounted in KEMAR's (Knowles Electronics Manikin for Acoustic Research) ears. The goal of the algorithm is to reproduce four virtual sound images located at  $\pm 30^\circ$  and  $\pm 110^\circ$ , according to the International Telecommunications Union (ITU) standard, ITU-R Rec. BS.775-1 [39]. The HRTF pairs at the corresponding directions are designated as the matching model responses  $\mathbf{M}(z)$ . The HRTF database measured by MIT Media Lab [19], [20] is employed. This problem involves two control points for a single passenger's ears, four loudspeakers, and four input channels. In other words,  $\mathbf{H}(z)$  is a  $2 \times 4$  matrix,  $\mathbf{M}(z)$  is a  $2 \times 4$  matrix, and  $\mathbf{C}(z)$  is a  $4 \times 4$  matrix. For the single passenger mode, however, the design of the inverse filters can be divided into two parts: the front side and the rear side. That is, the frontal loudspeakers are used to reproduce  $\pm 30^\circ$  virtual sound images, the rear ones for  $\pm 110^\circ$ . By doing so, a great saving can be achieved with the number of the inverse filters decreased from sixteen (one  $4 \times 4$  matrix) to eight (two  $2 \times 2$  matrices). This approach will also be employed in the following Method III. Further, some listeners reported in an informal listening test that the sound image width is slightly compromised in applying inverse filtering. To mitigate the problem, the weighted (0.45) and delayed (4 ms) four-channel inputs are mixed into the respective channels.

### 6.2.1.2 Method II: Up/downmixing with Delay and Weighting

The inverse filtering of Method I is intended for single passenger. If the passenger sits in a different seat, the inverse filters have to be redesigned. The sweet spot at the nominal design position is extremely small, as mentioned previously. The computation loading will also increase with increasing number of the passengers. To overcome these problems associated with inverse filtering, the Method II is developed as an alternative solution to the problem of automotive surround audio. Figure 40 shows the block diagram of Method II, in which concatenated upmixing and down mixing processing is required. In the study, weightings (0.65) and delay (20 ms) are used. The upmixing module in Method II is essentially the same as that in Method I. With the upmixed signals, downmixing is only done by standard weighting and summation to produce the two-channel outputs. Method II is less computationally demanding compared to the inverse filtering approach, which lends itself to multi-passenger listening scenarios.



### 6.2.2 5.1-channel Inputs

Another category of automotive surround processors that accepts 5.1 input signals from Dolby Digital or DTS decoder in DVD players will be presented in this section.

#### 6.2.2.1 Method III: Inverse Filtering

The structure of Method III shown in the block diagram of Fig 41 is the same as that of Method I except that it does not require upmix processing. Given the 5.1-channel inputs and four loudspeakers, the center channel has to be attenuated before mixing into the front-left and front-right channels. Next, front two channels and rear two channels are fed to the respective inverse filters. The remaining

channel, LFE, is mixed into each loudspeaker, assuming that the subwoofer is unavailable.

#### *6.2.2.2 Method IV: Downmixing with Delay and Weighting*

Similar to Method II, a down mixing algorithm, Method IV, is developed for inputs in 5.1 format, as depicted in the block diagram of Fig. 42. In the method, the center channel is first mixed into the front two channels and then the ipsilateral channels are summed to produce the two frontal channels. Next, the frontal channels are weighted and delayed to produce the rear channels. Like Method III, the remaining channel, LFE, is mixed into each loudspeaker, assuming that the subwoofer is unavailable. Like Method II, Method IV is less computationally demanding compared to the inverse filtering approach, which lends itself to multi-passenger listening scenarios.



### **6.3 Objective and Subjective Experiments**

In order to evaluate the performance of the four methods mentioned above, a series of objective and subjective experiments were conducted in a 2-liter sedan, as shown in Fig. 43(a). The sedan is equipped with a DVD player, a 7-inch LCD display, a multi-channel audio decoder, and four loudspeakers (two mounted in the lower panel of the front door and two behind the backseat). The experimental arrangement inside the car is shown in Fig. 43(b). The aforementioned audio algorithms are implemented on a fixed-point digital signal processor (DSP), Blackfin-533, of Analog Device semi-conductor. The microphone GRAS 40AC and the preamplifier GRAS 26AM were used for measuring acoustical plants required in the design of inverse filters.

### 6.3.1 Objective Experiments

The acoustical plants were measured before the experiments. Figures 44-48 show the results of the case when single passenger sitting at the front-left seat. Figure 44(a) shows the frequency responses between the frontal loudspeakers and the microphones mounted in the KEMAR sitting in front-left seat. The upper and lower rows of the figures are measured at the left and right ears, respectively. The left and right columns of the figures are measured when the left-side and right-side loudspeakers are enabled, respectively. For example, the upper-left plot is the frequency response measured between the left ear and the front-left loudspeaker. The  $x$ -axis and the  $y$ -axis represent frequency in Hz and magnitude in dB, respectively. The dotted lines and the solid lines signify the original responses and the nonuniformly smoothed responses, respectively. The spiky measured responses have been effectively smoothed out after applying the aforementioned equivalent complex smoothing technique. The frequency responses measured at the rear loudspeakers are shown in Fig. 44(b), where the magnitude above 8 kHz is attenuated dramatically. Comparison of the left column and the right column of Figs. 44(a) and (b) reveals that head shadowing is not significant due to boundary reflections in the small car cabin. Figures 45(a) and (b) show the measured impulse responses for the front-side and rear-side loudspeakers, respectively, whereas Figs. 45(c) and (d) show the impulse responses obtained using frequency smoothing. Clearly visible is that the tails of the impulse responses have been truncated while retaining the main characteristics of the responses.

The smoothed impulse responses are truncated to 512 taps and converted to the frequency domain via FFT. The truncation of the impulse response is tantamount to uniform smoothing in the frequency domain. This ‘double smoothing’ process leads to a much smoothed frequency response. The technique presented in Section 2 is

employed to design the inverse filters. The frequency responses of the inverse filters for the frontal and the rear acoustical plants are shown in Figs. 46(a) and (b), respectively. Figure 46(b) shows that the filter frequency responses above 6 kHz exhibit high gain because of the poor high-frequency response of the rear loudspeakers. In regularization of inverse filters, the gain is always restricted below 6 dB to prevent from overloading the filters. Figures 47(a) and (b) show the impulse responses of the inverse filters corresponding to the frequency responses of Fig. 46. The solid lines in Figs. 48(a) and (b) represent 30° and 110° HRTF pairs, respectively, whereas the dotted lines represent the multi-channel filter-plant product,  $\mathbf{H}(e^{j\omega})\mathbf{C}(e^{j\omega})$ . The agreement between these two sets of responses is generally good below 6 kHz except that notable discrepancies can be observed, especially for the rear-loudspeaker case. This is the inverse filters are gain-limited using regularization at the frequencies where the plants have significant roll-off.

Next, the scenario of single passenger sitting at the rear-right seat is considered. The frequency responses of the plants for the front and the rear loudspeakers are shown in Figs. 49(a) and (b), respectively. The solid lines and the dotted lines represent the smoothed and the measured responses, respectively. The impulse responses of the plants are shown in Figs. 50(a)-(d). The smoothed responses seem to have captured the main characteristics of the measured responses. Figures 51 and 52 show the frequency responses and impulse responses of the inverse filters, respectively. The results are compared in Figs. 53(a) and (b) for the front and rear virtual sources, respectively. It can be seen from Fig. 53(b) that the performance is degraded above 6 kHz. Similar result is obtained for the front seat. This might be due to the poor high-frequency responses of the acoustical plants and/or the non-ideal positions of the rear loudspeakers that are facing upward.

### 6.3.2 Subjective Experiments

Subjective listening experiments in the car were undertaken for comparing the four automotive audio methods proposed in Section 6.2, according to a modified double-blind Multi-Stimulus test with Hidden Reference and a hidden Anchor (MUSHRA) [48]. The experiment cases are described in Table X. In Experiment I, five songs in two-channel stereo format involving various instruments with significant dynamic variations were chosen to be the test materials. In the other experiments, four 5.1-channel movies in Dolby Digital format were used. Both timbre-related and space-related qualities are considered. The loudness of each reproduced signal was adjusted to the same level by measuring the sound pressure level at each seat with a monitoring microphone.

Eight subjective attributes employed in the tests, including *preference*, timbral attributes (*fullness, brightness, artifact*) and spatial attributes (*localization, frontal image, proximity, envelopment*) are summarized in Table XI. Thirty subjects participated in each experiment. The subjects participating in the tests were instructed with definitions of the subjective indices and the procedures before the listening tests. The subjects were asked to respond after listening in a questionnaire, with the aid of a set of subjective indices placed on a scale from  $-3$  to  $3$ . Positive, zero, and negative scores indicate perceptually improvement, no difference, and degradation, respectively, of the signals after processed by the audio spatializers. To prevent the listeners from fatigue, the order of the attributes is randomized except that the index *preference* is always the last question. It took approximately thirty minutes to finish an experiment. In order to access statistical significance, the scores were further processed by using the MANOVA. If the significance level is below 0.05, the differences among all methods are considered statistically significant.

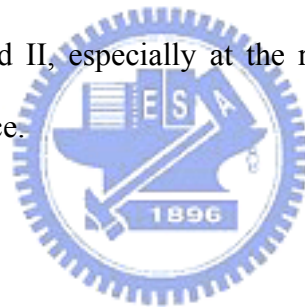


### 6.3.2.1 Experiment I

In this experiment, Method I (MI) and Method II (MII) are compared for the listening positions at the front-left and the rear-right seats. Apart from these two methods, a hidden reference (H. R.) and an anchor (An.) are added into the comparison. The case in which two-channel stereo input signals are fed to the respective front and rear loudspeakers is used as the hidden reference. The signal obtained by summing and lowpass filtering (with 4 kHz cutoff frequency) the two-channel input signals is used as the anchor that is also fed to all loudspeakers.

Prior to the data analysis, three MANOVA assumptions have been checked that they are not violated. Two-way MANOVA indicates that there is only insignificant difference among all listening positions ( $F = 0.921, p = 0.504$ ), but there is significant difference among all methods ( $F = 22.822, p < 0.001$ ), and there are interactions ( $F = 2.961, p < 0.001$ ). Since there is interaction between the position and the method, the analysis is separated into two parts: the front-left seat and the rear-right seat. Figure 54(a) and (b) show the means and spreads (with 95% confidence intervals) of the grades of the first four (*preference* and *timbral*) attributes and the last four (*spatial*) attributes, respectively. The left column represents the front-left (FL) position, while the right column represents the rear-right (RR) position. The  $x$ -axis and  $y$ -axis represent the method and grade, respectively. Only the index *fullness* showed no significant difference among all methods. Three attributes, *preference*, *localization*, and *envelopment*, showed similar results regardless of the positions. In terms of *preference*, Method I is a preferred choice over Method II, even though Method II performed significantly better than the hidden reference. In terms of *localization*, Method I performs significantly better than Method II, while there is no significant difference between Method II and the hidden reference. In terms of *envelopment*, there is no significant difference between Method I and Method II, albeit they both

perform significantly better than the hidden reference. On the other hand, the results of the other four attributes are quite different for two positions. There is no significant difference among all methods in *artifact* at the front-left seat. However, some clicking artifacts were heard in testing Method I for the rear-right seat. In terms of *brightness*, *frontal image*, and *proximity*, there is no significant difference between Method I and Method II, notwithstanding we obtained significantly higher grades than the hidden reference for the front-left seat. At the rear-right seat, there exists significant difference between Method I and Method II. Further, there exists significant difference between Method II and the hidden reference in terms of *frontal image* and *proximity*. Method I received the highest score in *brightness*, which differs significantly from Method II and the hidden reference. Overall, Method I performed better than Method II, especially at the rear seat, albeit both performed better than the hidden reference.



#### 6.3.2.2 Experiment II

Methods III, IV, and the unprocessed 5.1-channel reproduction are compared in this experiment. Because only four loudspeakers are available in this car, the center channel of the 5.1-channel input is attenuated by -3 dB and mixed into the front channels to serve as the hidden reference. In addition, the four-channel signals are summed and lowpass filtered (with 4 kHz cutoff frequency) is used as the anchor. Fifteen listeners participated in the test for the front left and rear right seats.

The statistical assumptions of MANOVA have been verified. The results of two-way MONOVA indicate that there is no significant interaction between the positions and the methods ( $F = 1.296$ ,  $p = 0.206$ ). However, there are significant intra-group differences in the positions ( $F = 3.165$ ,  $p = 0.003$ ) and in the methods ( $F = 12.046$ ,  $p < 0.001$ ). Since there is no significant interaction between factors, the

following analysis will focus on each factor separately. Figures 55(a) and (b) show the means and spreads (with 95% confidence intervals) of the grades of the first four and last four attributes for all methods. A *post hoc* technique [33] was used to determine which variable contributed most to the overall multivariate significance. The result indicated that the grades of Method III in *preference*, *brightness*, *frontal image*, and *proximity* were significantly higher than those of Method IV. In *fullness* and *envelopment*, Methods III and IV are not significantly different, albeit both grades are significantly higher than the grades of the hidden reference. Further, there is no significant difference among methods in the attribute *artifact*. Note that Method IV received the lowest grade in *localization*, while Method III received the highest grade. This might be caused by the downmix processing in Method IV, where the front channels are mixed into the rear channels. On the other hand, the means and spreads (with 95% confidence intervals) of the grades of the first four and last four attributes are shown in Figs. 55(c) and (d), respectively, where the *x*-axis represents positions. Significant differences exist among the positions for three attributes: *brightness*, *frontal image*, and *proximity*. All grades received are higher for the rear seat than for the front seat. There is no significant difference among the positions in the other attributes.

Overall, the proposed processing methods have significantly improved the audio rendering in the car in all attributes but *localization*, as compared to the hidden reference. In particular, Method III received the highest grades in most attributes, especially in spatial attributes. In addition, Method III performed better at the rear seat than at the front seat in *frontal image* and *proximity*.

### 6.3.2.3 Experiment III

In this experiment, the two-speaker approach is compared with the four-speaker

approach for single passenger. In the two-speaker approach, the inverse filters are divided into the frontal group and the rear group. Each group is responsible for producing respective pair of two virtual sound images. In other words, the number of inputs  $N = 2$ , the number of loudspeakers  $M = 2$ , and the number of control points  $L = 2$ , for each group. In the four-speaker approach, the number of inputs  $N = 4$ , the number of loudspeakers  $M = 4$ , and the number of control points  $L = 2$ . Apparently, the two-speaker approach (two  $2 \times 2$  inverse matrices) is less computationally demanding than the four-speaker approach (one  $4 \times 4$  inverse matrix). The MANOVA test indicates that there is significant differences among all methods ( $p < 0.001$ ). Figures 56(a) and (b) show the means and spreads (with 95% confidence intervals) of the grades of the first four and the last four attributes, respectively. Contrary to our expectation that more inverse filters should mean better performance, the *post hoc* procedure indicates that the two-speaker approach earned higher grades than the four-speaker approach in almost all attributes but *fullness* and *artifact*. No significant differences were found among the methods in terms of *fullness* and *artifact*. This interesting result may be due to the fact that the four-speaker approach remixes the front and the rear channels, which might cause front-back reversal problem when the plants are mismatched. The two-speaker approach does not suffer from this problem because of separate design for the frontal and rear processing. Moreover, it needs only half computation loading as compared to four-speaker approach. Nevertheless, the four-speaker approach still performed significantly better than the hidden reference.

#### 6.3.2.4 Experiment IV

The inverse filters employed in the above experiments are intended for single passenger sitting mode. In this experiment, the inverse filters are designed for two

passengers sitting in the front seats. Method III with the two-speaker approach was first applied using four control points (four ears). However, an informal listening test indicated this was not a viable approach for the non-square nature of inverse filter design (the acoustical plant  $\mathbf{H}$  is a  $4 \times 2$  matrix). For this reason, the design of the inverse filters in Method III is modified to a problem with four loudspeakers and four control points, where the plant  $\mathbf{H}$  is a  $4 \times 4$  matrix and hence this approach is renamed as Method III-2. In this case, the listening test is basically similar to Experiment II except some modified steps borrowed from Method III. Figures 57(a) and (b) show the means and spreads (with 95% confidence intervals) of the grades of the first four and the last four attributes, respectively. The *post hoc* procedure indicates that there is no significant difference between Methods III-2 and IV, while both are significantly different from the hidden reference in overall *preference*, *frontal image*, and *envelopment*. In terms of *proximity*, Method III-2 performed better than Method IV that is also significantly better than the hidden reference. Method IV earned higher grades in *fullness* than Method III-2 and the hidden reference. The grade of *artifact* obtained using Method III-2 is very low with the mean within -0.5, implying that some artifacts are perceptible but not disturbing. Some subjects reported during the test that glitches can be heard in certain sections when applying this method. There is no significant difference among all methods in *brightness* and *localization*. To conclude, although inverse filtering did not perform as well as it did for the single passenger mode, both processing methods significantly outperformed the hidden reference. In terms of computation complexity and rendering performance, Method IV is the adequate approach for the two-passenger mode.

#### 6.3.2.5 Multiple Regressions Analysis of the Attributes

In order to examine the correlation between the overall *preference* (dependent

variable) and the other attributes (independent variables), multiple linear regression analysis is conducted. The following regression model is obtained:

$$\begin{aligned} Preference = & 0.283 + 0.064 \times Fullness + 0.327 \times Brightness + 0.106 \times Artifact \\ & + 0.117 \times Localization + 0.135 \times Frontal\ image \\ & + 0.078 \times Proximity + 0.440 \times Envelopment. \end{aligned} \quad (42)$$

The regression model is considered statistically significant with all  $p$  values below 0.05. The squared correlation coefficient ( $R^2$ ) reaches 0.966, indicating this regression model can interpret almost 97% of the variance of *preference*. As we can see from the model coefficients, *envelopment* and *brightness* are two dominant attributes in overall *preference*, whereas the influence of *fullness* and *proximity* is the least. It should be noted that, however, the high frequency response of the filters should be carefully handled lest it is mistakenly perceived as artifacts.



## 7. CONCLUSIONS

---

A comprehensive study has been conducted to explore the effects of listening angle on crosstalk cancellation in spatial sound reproduction using two-channel stereo systems in Sec. 3. The intention is to establish a sustainable configuration of CCS that best reconciles the separation performance and the robustness against lateral head movement, not only in theory but also in practice. Similar to the previous research which focuses mainly on numerical stability, the present work arrives at the conclusion that inversion of ill-conditioned systems results in high gain filters, loss of dynamic range and hence separation performance. Regularization is required to compromise between numerical stability and separation performance. However, findings different from the previous study had also been reached because this work employed a comprehensive approach. First, it is found from the HRTF results that the problem of high frequency ringing is not as critical as in the point source model owing to head shadowing. In addition, poor conditioning, high gain, and low performance problems at low frequencies may arise for extremely small span arrangements, whereas there is broader useful frequency range with performance and numerical stability if wide span arrangement can be used. The effects of listening angle were also examined in the context of sweet spot. Two kinds of sweet spot definitions are employed in the simulation. The relative sweet spot suggests that robustness is excellent with the use of small span arrangement notwithstanding the poor performance in the nominal position, which is in agreement with the previous research. However, it is not very useful in practical application if the average channel separation in the sweet spot is very poor even though it is relatively robust. Therefore, in addition to the conventional relative definition, we suggest another definition, the absolute sweet spot, to make the evaluation more complete. In an absolute sweet spot, the performance is guaranteed in complement to the relative

robustness, which is desirable in practical use of the CCS. The results of absolute sweet spot reveal that arrangements with listening angle ranging from 120 to 150 degrees are optimal choices.

To justify the conjectures above, objective and subjective experiments were undertaken in an anechoic room for three loudspeaker arrangements, including the stereo dipole (10-deg), standard span (60-deg), and proposed span (120-deg). The results post-processed by the ANOVA test indicate that the 120-deg configuration performs comparably well as the standard 60-deg configuration, but is better than the 10-deg configuration. Small span arrangement produces large relative sweet spot because head displacement would cause minimal change of time-of-arrival differences between two loudspeakers using closely spaced loudspeakers. This configuration is well suited to applications that must be spatially compact, e.g., mobile phones and other portable devices. Nevertheless, the benefit of small span arrangement comes at the price of poor conditioning, high gain, and limited performance problems at low frequencies. Apart from this, due to the lack of natural high frequency separation provided by head shadowing, the small span arrangement is not able to position “out-of-range” source when CCS breaks down at high frequencies, where the phantom source is incorrectly panned within a narrow span. The arrangement with large span appears to be more effective than the small span because head shadowing and panning effect help to provide localization effect to certain degree even if CCS breaks down. While it may seem from this report that large-span configuration is predominantly favored, problems inherent to large span prevent the span to grow indefinitely, e.g., sound image stability will become an issue for wide apart loudspeakers. A practical recommendation is perhaps the conventional 60-deg configuration which is a reasonable compromise between the two extremes (10 and 120 degrees) to achieve both robustness and performance. It was also found that the



120-deg arrangement did not perform as well as the 60-deg arrangement in positioning frontal images. If an additional center loudspeaker is available, the 3/0 format with 120-deg span would be an ideal choice.

A bandlimited CCS based on subband filtering has been developed in Sec. 4. The intention is to establish a computationally efficient CCS without penalty on cancellation performance. The CCS is a bandlimited design which is effective up to the frequency 6 kHz. To achieve the bandlimited implementation, a pseudo cosine modulated QMF is employed, allowing the CCS to operate at low rate within an approximate PR structure. As a result of this, spatial audio processing can concentrate more on the low frequency range to better suit human perceptual hearing.

To compare the proposed CCS to traditional systems, subjective listening experiments were conducted in an anechoic room. The experiments include two parts: source localization test and sound quality test. By means of the techniques presented in Sec. 4.1, the fullband CCS operated at the sampling rate of 48 kHz requires four 3000-tapped FIR filters. On the other hand, the bandlimited CCS operated at the sampling rate of 12 kHz requires only four 1500-tapped FIR filters. The prototype FIR filter has 120 taps. The analysis bank and the synthesis bank are generated from the prototype and implemented via polyphase representation. The results of subjective tests processed by ANOVA indicate that the bandlimited CCS performs comparably well as the fullband CCS not only in localization but also in sound quality. From Table V, the computation loading using the proposed subband filtering approach was drastically reduced by approximately eighty percent, as compared to the conventional approach. After employing fast convolution algorithm, the difference between two methods is reduced. Even though the block convolution is very efficient, it requires more memory to store temporary data. In conclusion, which method is better is dependent upon which one you concern about, speed or

memory. The bandlimited CCS with direct convolution and shuffler method is an acceptable choice.

The cone velocity observer that requires no sensor has been developed and implemented in an analog circuit in Sec. 5. Excellent estimation of cone velocity has been achieved using the suggested system. A hybrid control employing a feedforward filter and a feedback compensator is proposed. A feedforward filter is synthesized on the basis of the velocity observer. A feedback compensator is designed by using QFT. With the aid of such system, the bass response of a loudspeaker has better low-frequency extension with significant level enhancement.

A comprehensive study has been conducted to explore promising but practical approaches for the automotive virtual surround audio systems via simulations and experiments. The simulation using the free-field point source model reveals that setting three control points at each seat position creates the largest sweet spot, but the performance at each control point is compromised. Four processing methods have been presented: the first two methods are intended for two-channel inputs and the other two methods are intended for 5.1-channel inputs. A reverberation-based upmix processing is used to convert two-channel inputs to four-channel signals. In addition, the inverse filters in Method I and Method III are exploited to correct the car responses and then render a spatial listening environment. Methods II and IV are practical approaches in the sense of computation complexity and audio performance. Conclusions can be drawn from the listening tests as follows. First, for two-channel inputs, Method I outperformed Method II, especially for the rear seat, while both performed the hidden reference. Second, for 5.1-channel inputs, Method III received the highest grades in most attributes, especially in spatial attributes. In addition, Method III performed better at the rear seat than it did at the front seat in *frontal image* and *proximity*. Third, for the single passenger mode, the two-speaker

approach is a preferred choice over the four-speaker approach in considering rendering performance and computation complexity. Fourth, inverse filtering did not perform as well for the two passenger mode as it did for the single passenger mode. Further, the number of inverse filters increases drastically with number of passengers, rendering this scheme impractical. Fifth, overall *preference* is dominated by *brightness* and *envelopment*, as indicated by the multiple-regression analysis. It is concluded from the discoveries above that a simple design strategy can be formulated according to the number of passengers, using a hybrid approach. Methods I and III are employed for one passenger, while Methods II and IV are employed for more than one passenger.



## 8. FUTURE WORK

---

A number of topics are planned for future research. Efficient inverse filtering methods applicable to multi-passenger scenarios should be developed. Integration of present surround system to the other audio techniques such as equalizers, superbass systems, dynamic range control, Karaoke machines, acoustical echo and noise control, etc., should be investigated. In addition, self-identification and compensation systems should be developed in the future study for practical application of the automotive audio system.



## REFERENCES

- [1] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA, 1997).
- [2] D. R. Begault, *3-D Sound for Virtual Reality and Multimedia* (AP Professional, Cambridge, MA, 1994).
- [3] R. Schroeder and B. S. Atal, "Computer simulation of sound transmission in rooms," IEEE International Convention Record **7**, 150-155 (1963).
- [4] P. Damaske and V. Mellert, "A procedure for generating directionally accurate sound images in the upper- half space using two loudspeakers," *Acoustica* **22**, 154-162 (1969).
- [5] D. H. Cooper, "Calculator program for head-related transfer functions," *J. Audio Eng. Soc.* **30**, 34-38 (1982).
- [6] W. G. Gardner, "Transaural 3D audio," MIT Media Laboratory Tech. Report 342, (1995).
- [7] D. H. Cooper and J. L. Bauck, "Prospects for transaural recording," *J. Audio Eng. Soc.* **37**, 3-19 (1989).
- [8] J. L. Bauck and D. H. Cooper, "Generalized transaural stereo and applications," *J. Audio Eng. Soc.* **44**, 683-705 (1996).
- [9] C. Kyriakakis, T. Holman, J. S. Lim, H. Homg, and H. Neven, "Signal processing, acoustics, and psychoacoustics for high-quality desktop audio," *J. Vis. Comm. and Im. Rep.* **9**, 51-61 (1997).
- [10] C. Kyriakakis, "Fundamental and technological limitations of immersive audio systems," *IEEE Processing* **86**, 941-951 (1998).
- [11] D. B. Ward and G. W. Elko, "Optimal loudspeaker spacing for robust crosstalk cancellation," *Proc. ICASSP 98, IEEE*, 3541-3544 (1998).
- [12] D. B. Ward and G. W. Elko, "Effect of loudspeaker position on the robustness of

- acoustic crosstalk cancellation,” IEEE Signal Process. Lett. **6(5)**, 106-108 (1999).
- [13] D. B. Ward, “Joint squares optimization for robust acoustic crosstalk cancellation,” IEEE Trans. on Speech and Audio Processing **8(2)**, 211-215 (2000).
- [14] O. Kirkeby, P. A. Nelson, and H. Hamada, “The “stereo dipole” a virtual source imaging system using two closely spaced loudspeakers,” J. Audio Eng. Soc. **46**, 387-395 (1998).
- [15] T. Takeuchi and P. A. Nelson, “Robustness to head misalignment of virtual sound imaging systems,” J. Audio Eng. Soc. **109**, 958-971 (2001).
- [16] T. Takeuchi and P. A. Nelson, “Optimal source distribution for binaural synthesis over loudspeakers, J. Acoust. Soc. Am. **112**, 2786-2797 (2002).
- [17] M. R. Bai, C. W. Tung, and C. C. Lee, “Optimal design of loudspeaker arrays for robust cross-talk cancellation using the Taguchi method and the genetic algorithm,” J. Acoust. Soc. Am. **117**, 2802-2813 (2005).
- [18] W. G. Gardner, *3-D Audio Using Loudspeakers* (Kluwer Academic, Dordrecht, 1998).
- [19] W. G. Gardner and K. D. Martin, “HRTF measurements of a KEMAR,” J. Acoust. Soc. Am. **97**, 3907-3908 (1995).
- [20] W. G. Gardner and K. D. Martin, *KEMAR HRTF measurements* (MIT’s Media Lab, <http://sound.media.mit.edu/KEMAR.html>, 1994).
- [21] P. P. Vaidyanathan, *Multirate Systems and Filter Banks* (Prentice-Hall, Englewood Cliffs, NJ, 1993).
- [22] F. Rumsey, *Spatial Audio* (Focal Press, Oxford, Boston, 2001).
- [23] P. A. Gauthier, A. Berry and W. Woszczyk, “Sound-field reproduction in-room using optimal control techniques: simulations in the frequency domain,” J.

- Acoust. Soc. Am. **117**, 662–678 (2005).
- [24] T. Betlehem and T. D. Abhayapala, “Theory and design of sound field reproduction in reverberant rooms,” J. Acoust. Soc. Am. **117**, 2100–2111 (2005).
- [25] G. Theile and H. Wittek, “Wave field synthesis: a promising spatial audio rendering concept,” Acoust. Sci. and Tech. **25**, 393–399 (2004).
- [26] S. Spors, A. Kuntz, and R. Rabenstein, “An approach to listening room compensation with wave field synthesis,” AES 24<sup>th</sup> International conference on multichannel audio, 1–13, (AES, Canada, 2003).
- [27] Lexicon, LOGIC 7, <http://www.lexicon.com/logic7/index.asp>
- [28] Dolby, Prologic II, [http://www.dolby.com/professional/popup\\_PLII/](http://www.dolby.com/professional/popup_PLII/)
- [29] SRS Labs, SRS Automotive,  
<http://www.srslabs.com/ae-srsautomotivetech826.asp>
- [30] Bose, AudioPilot,  
[http://www.bose.com/controller?event=VIEW\\_STATIC\\_PAGE\\_EVENT&url=/automotive/innovations/audiopilot.jsp](http://www.bose.com/controller?event=VIEW_STATIC_PAGE_EVENT&url=/automotive/innovations/audiopilot.jsp)
- [31] Bang & Olufsen, Advanced sound system,  
<http://www.bang-olufsen.com/page.asp?id=321>
- [32] B. Crockett, M. Smithers, and E. Benjamin, “Next Generation Automotive Research and Technologies,” *AES 120<sup>th</sup> Convention* (Paris, France, 2006).
- [33] S. Sharma, *Applied Multivariate Techniques* (John Wiley, New York, 1996).
- [34] B. Noble, *Applied Linear Algebra* (Prentice-Hall, 1988).
- [35] O. Kirkeby, P. A. Nelson, and H. Hamada, “Fast deconvolution of multichannel systems using regularization,” IEEE Trans. Speech Audio Processing **6**, 189-194 (1998).
- [36] A. Schuhmacher, and J. Hald, “Sound source reconstruction using inverse boundary element calculations,” J. Acoust. Soc. Am. **113**, 114-127 (2003).

- [37] M. R. Bai and C. C. Lee, "Development and implementation of cross-talk cancellation system in spatial audio reproduction based on the subband filtering," *J. Sound Vib.* **290**, 1269-1289 (2006).
- [38] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (New York, 2001).
- [39] ITU-R Rec. BS.775-1, "Multi-channel stereophonic sound system with or without accompanying picture," International Telecommunications Union, Geneva, Switzerland (1992–1994).
- [40] Y. P. Lin, and P. P. Vaidyanathan, "A Kaiser window approach to the design of prototype filters of cosine modulated filter banks," *IEEE Signal Processing Letters*, 132-134, (June 1998).
- [41] Rec. ITU-R BS.1116-1, "Method for the subjective assessment of small impairments in audio systems including multichannel sound systems," International Telecommunications Union, Geneva, Switzerland (1992–1994).
- [42] A. V. Oppenheim, R. W. Schaffer, and J. R. Buck, *Discrete-Time Signal Processing*, 2<sup>nd</sup> ed., (Prentice-Hall, 1999).
- [43] J. Yang, W. S. Gan, and S. E. Tan, "Development of virtual sound imaging system using triple elevated speakers," *IEEE Transaction on Consumer Electronics*, **50**(3), 916-922 (Aug. 2004).
- [44] P. D. Hatziantoniou and J. N. Mourjopoulos, "Errors in real-time room acoustics dereverberation," *J. Audio. Eng. Soc.* **52**, 883–899 (2004).
- [45] P. D. Hatziantoniou and J. N. Mourjopoulos, "Generalized fractional-octave smoothing of audio and acoustic responses," *J. Audio. Eng. Soc.* **48**, 259–280 (2000).
- [46] M. R. Bai and G. Bai, "Optimal design and synthesis of reverberators with a



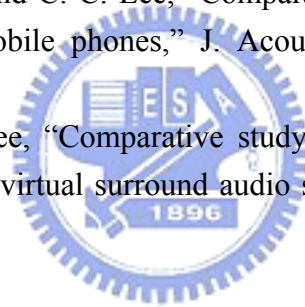
- fuzzy user interface for spatial audio,” J. Audio. Eng. Soc. **59**, 812–825 (2005).
- [47] Mingsian R. Bai and Geng-Yu Shih, “Upmixing and downmixing two-channel stereo audio for consumer electronics,” IEEE Trans. Consumer Electron. (2007) (Accepted).
- [48] ITU-R BS.1534-1, “Method for the Subjective Assessment of Intermediate Sound Quality (MUSHRA)”, International Telecommunications Union, Geneva, Switzerland (2001).



## PUBLICATIONS

### Journal Paper

1. M. R. Bai and C. C. Lee, "Development and implementation of cross-talk cancellation system in spatial audio reproduction based on the subband filtering," *J. Sound Vib.*, 290(3-5), 1269-1289 (2005).
2. M. R. Bai, C. W. Tung, and C. C. Lee, "Optimal design of loudspeaker arrays for robust cross-talk cancellation using the Taguchi method and the genetic algorithm," *J. Acoust. Soc. Am.*, 117(5), 2802-2813 (2005).
3. M. R. Bai and C. C. Lee, "Objective and subjective analysis of effects of loudspeaker span on crosstalk cancellation in spatial sound reproduction," *J. Acoust. Soc. Am.*, 120(4), 1976-1989 (2006).
4. M. R. Bai and C. C. Lee, "Subband approach to bandlimited crosstalk cancellation system in spatial sound reproduction," *EURASIP Journal on Applied Signal Processing*, 1-9 (2007).
5. M. R. Bai, G. Y. Shih, and C. C. Lee, "Comparative study of audio spatializers for dual-loudspeaker mobile phones," *J. Acoust. Soc. Am.*, 121(1), 298-309 (2007).
6. M. R. Bai and C. C. Lee, "Comparative study of design and implementation strategies of automotive virtual surround audio systems," submitted to *J. Audio Eng. Soc.* (2007).



### Conference Paper

1. M. R. Bai and C. C. Lee, "DSP-based sensorless velocity observer with audio applications in loudspeaker compensation," *The 118th AES convention*, Barcelona, Spain (2005).
2. M. R. Bai and C. C. Lee, "Comprehensive analysis of loudspeaker span effects on crosstalk cancellation in spatial sound reproduction," *The 120th AES convention*, Paris, France (2006).

### Patent

1. 白明憲、董志偉、李志中，中華民國專利（發明第 I 249361 號）：多聲道之交越失真消除系統

Table I. The description of five levels of grade for the subjective localization test.

Grade	Description
5	The judged angle is the same as the target angle
4	30° difference between the judged angle and the target angle
3	Front-back reversal of the judged angle identical to the target angle
2	30° difference between front-back reversal of the judged angle and the target angle
1	Otherwise



Table II. ANOVA results of the first group

	SS	df	MS	<i>F</i>	<i>p</i>
Intercept	8359.001	1	8359.001	4812.701	0.0001
Span	140.011	1	140.011	80.612	0.0001
Displacement	31.624	2	15.812	9.104	0.0001
Span * Displacement	3.189	2	1.595	0.918	0.3996



Table III. ANOVA results of the second group

	SS	df	MS	<i>F</i>	<i>p</i>
Intercept	10367.42	1	10367.42	8595.508	0.0001
Span	0.03	1	0.03	0.026	0.8712
Displacement	21.63	2	10.82	8.968	0.0001
Span * Displacement	1.52	2	0.76	0.631	0.5325



Table IV. Five-grade impairment scale.

Impairment	Grade
Imperceptible	5.0
Perceptible, but not annoying	4.0
Slightly annoying	3.0
Annoying	2.0
Very annoying	1.0



Table V. The comparison of computation loading of the fullband CCS and the bandlimited CCS with direct convolution.

	Fullband	Bandlimited
MPU	12,000	1,980
APU	11,998	1,976



Table VI. The comparison of computation loading of the fullband CCS and the bandlimited CCS with fast convolution.

	Fullband	Bandlimited
MPU	1,464	815
APU	1,462	808





Table VII. The THD results measured with/without feedback compensation.

	Without Feedback	With Feedback
25 Hz	3.954 %	2.777 %
30 Hz	2.917 %	2.495 %
35 Hz	1.143 %	1.044 %
40 Hz	0.914 %	0.762 %



Table VIII. The coordinates of the control points at each seat for three simulation cases.

	Front left	Front right	Rear left	Rear right
Case I	(-0.3,0.3)	(0.3,0.3)	(-0.3,-0.3)	(0.3,-0.3)
Case II	(-0.4,0.3)	(0.2,0.3)	(-0.4,-0.3)	(0.2,-0.3)
	(-0.2,0.3)	(0.4,0.3)	(-0.2,-0.3)	(0.4,-0.3)
Case III	(-0.4,0.3)	(0.2,0.3)	(-0.4,-0.3)	(0.2,-0.3)
	(-0.2,0.3)	(0.4,0.3)	(-0.2,-0.3)	(0.4,-0.3)
	(-0.3,0.4)	(0.3,0.4)	(-0.3,-0.2)	(0.3,-0.2)



Table IX. The descriptions of four automotive virtual surround processing methods.

Method	Input content	Design strategy
I	2-channel	Upmixing + Inverse filtering
II	2-channel	Up/downmixing + Weighting & delay
III	5.1-channel	Inverse filtering
IV	5.1-channel	Downmixing + Weighting & delay



Table X. The descriptions of four experiments.

Experiment	I	II	III	IV
Input content	2-channel	5.1-channel	5.1-channel	5.1-channel
Passenger no.	1	1	1	2
Method compared	Method I Method II	Method III Method IV	Two-speaker Four-speaker	Method III-2 Method IV
Reference	$L_{in} \rightarrow FL_{out}$ $R_{in} \rightarrow FR_{out}$ $0.7 \times L_{in} \rightarrow RL_{out}$ $0.7 \times R_{in} \rightarrow RR_{out}$		$FL_{in} + 0.7 \times C_{in} \rightarrow FL_{out}$ $FR_{in} + 0.7 \times C_{in} \rightarrow FR_{out}$ $RL_{in} \rightarrow RL_{out}$ $RR_{in} \rightarrow RR_{out}$	
Anchor	Summation of all lowpass filtered inputs $\rightarrow$ All outputs			



Table XI. The definitions of the subjective attributes.

Attribute	Description
Preference	Over all preference in considering timbre-related and space-related attributes
Fullness	Dominance of low-frequency sound
Brightness	Dominance of high-frequency sound
Artifacts	Any extraneous disturbances to the signal
Localization	Determination by a subject of the apparent direction of a sound source
Frontal image	The clarity of the frontal image or phantom center
Proximity	The sound is dominated by the loudspeaker closest to the subject
Envelopment	Perceived quality of listening within a reverberant environment



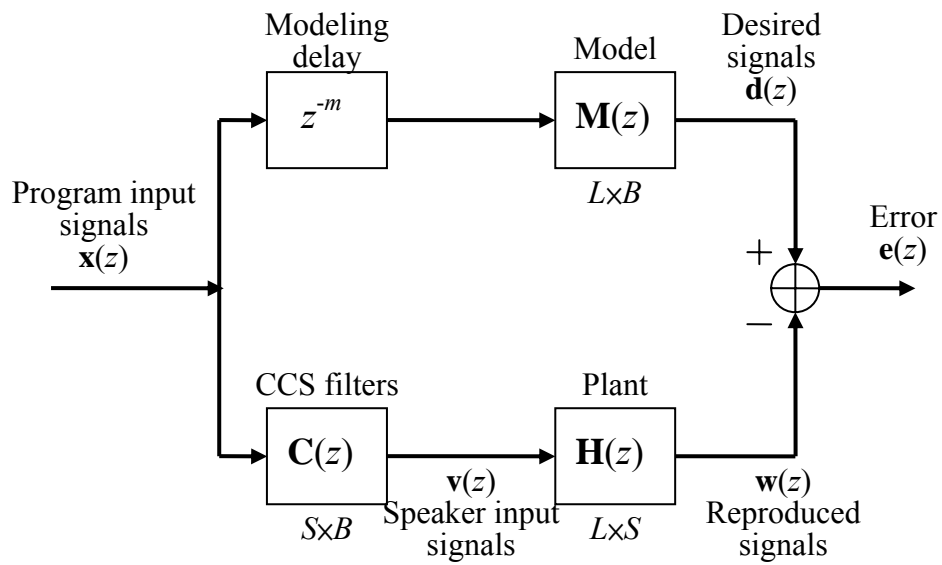
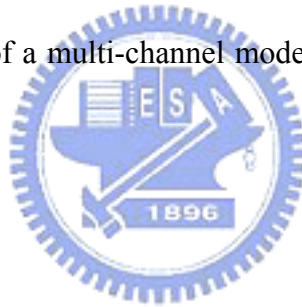


Fig. 1 The block diagram of a multi-channel model-matching problem in the CCS design.



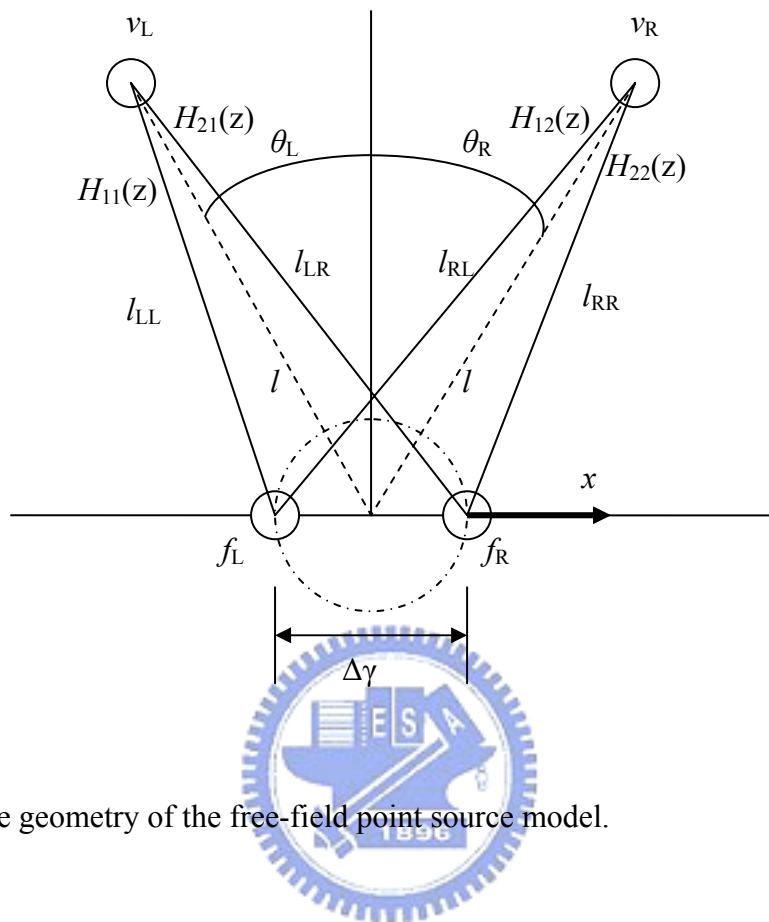
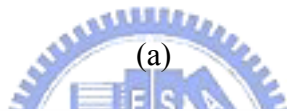
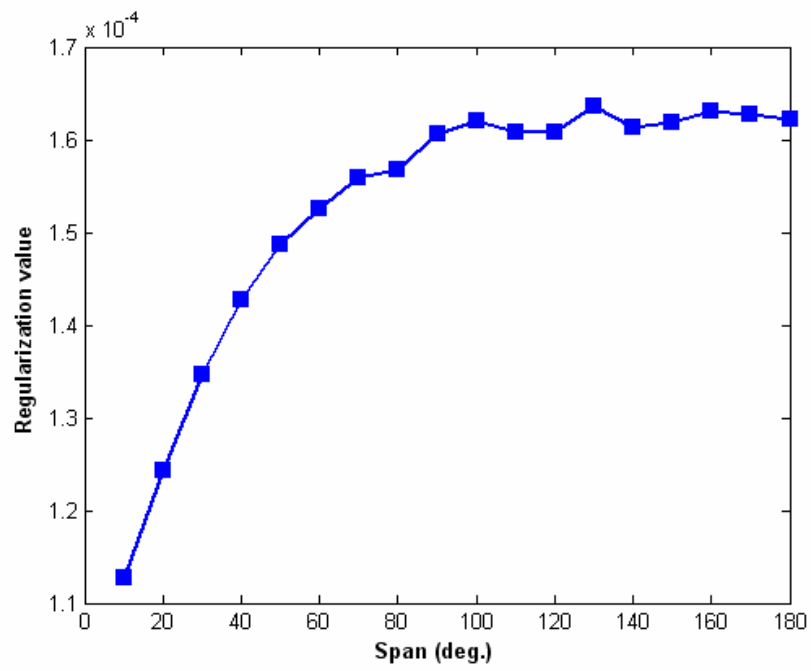
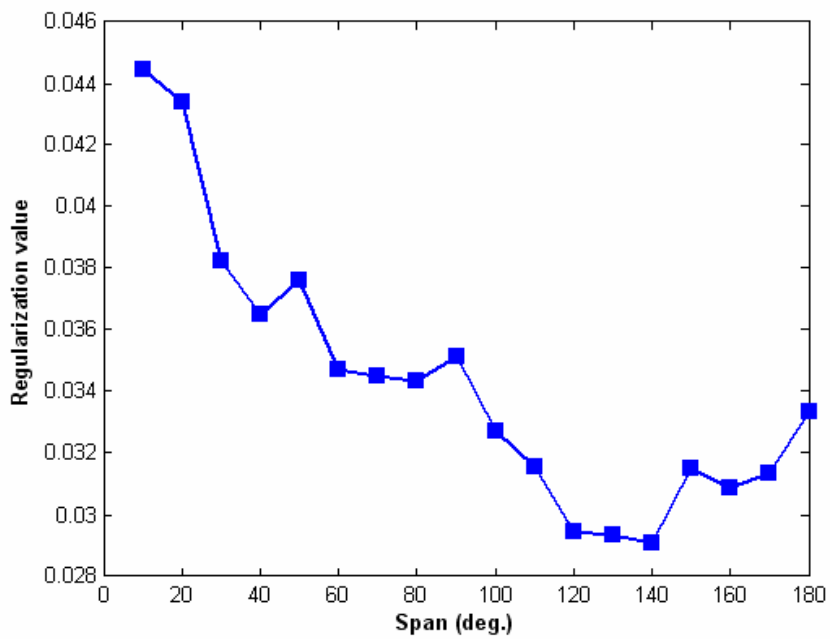


Fig. 2. The geometry of the free-field point source model.



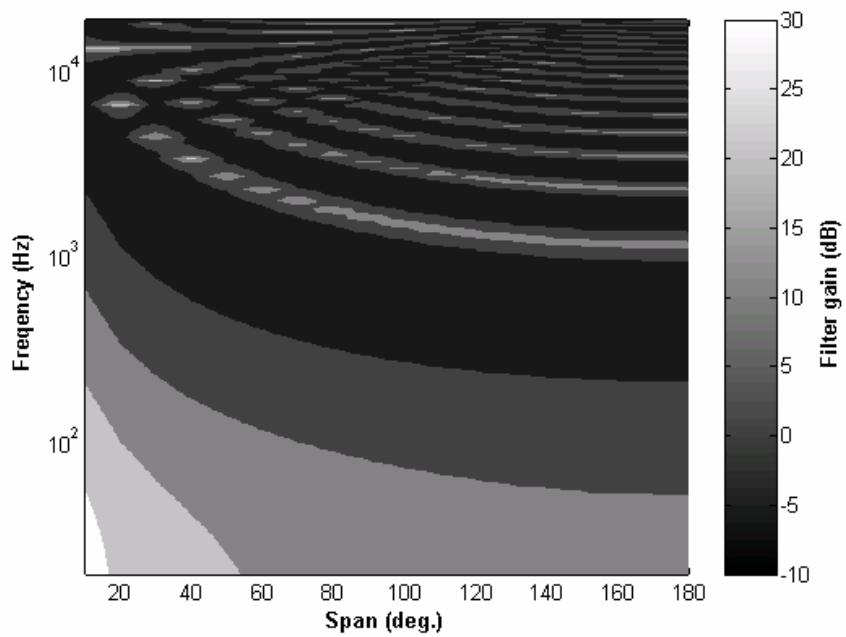
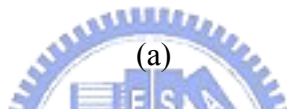
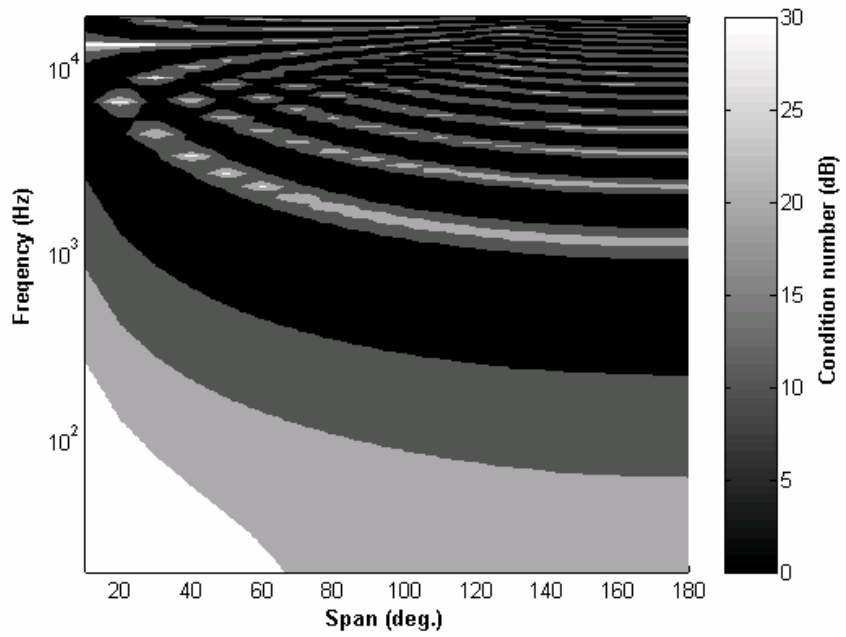
(a)



(b)

Fig. 3. The values of regularization in (a) the free-field point source model and (b) the HRTF model.





(b)

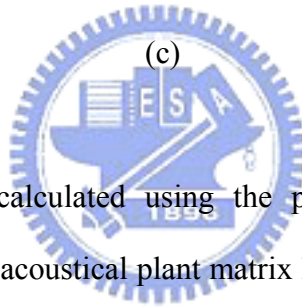
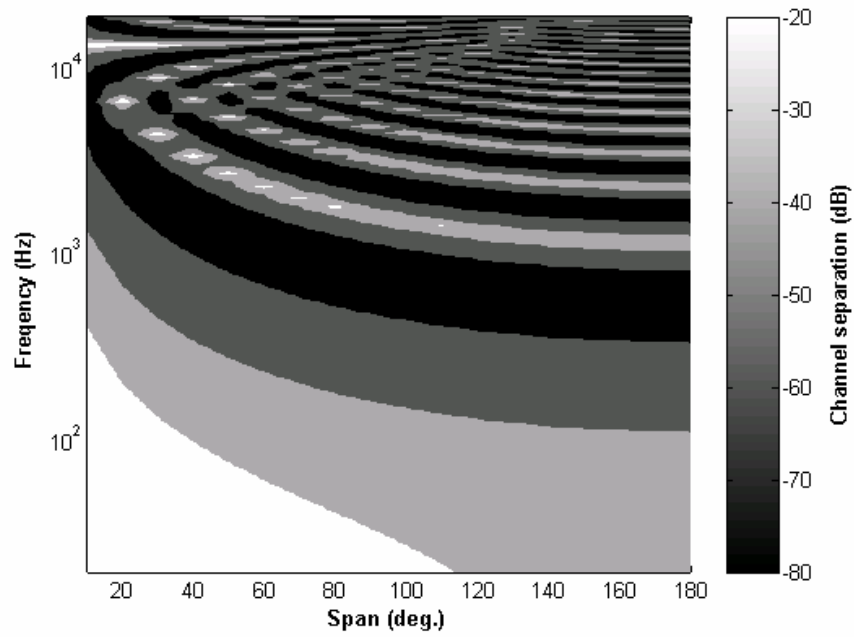
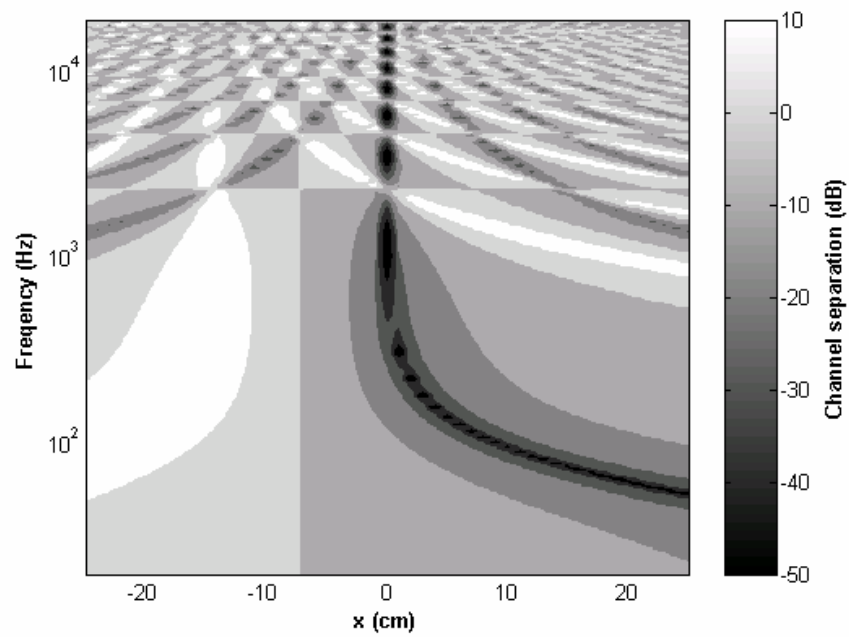
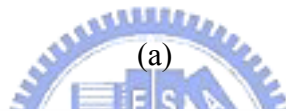
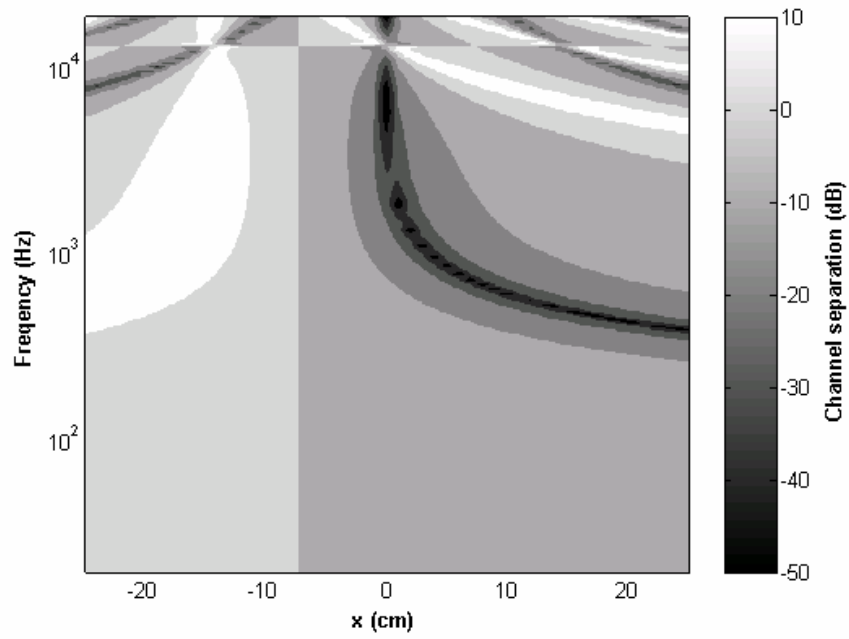


Fig. 4. The contour plots calculated using the point source model of (a) the condition number of acoustical plant matrix  $\mathbf{H}$ , (b) the filter gain, and (c) the channel separation.



(b)

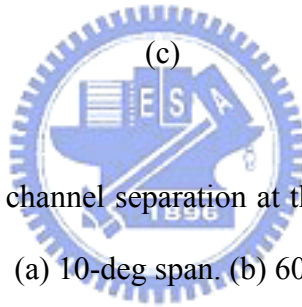
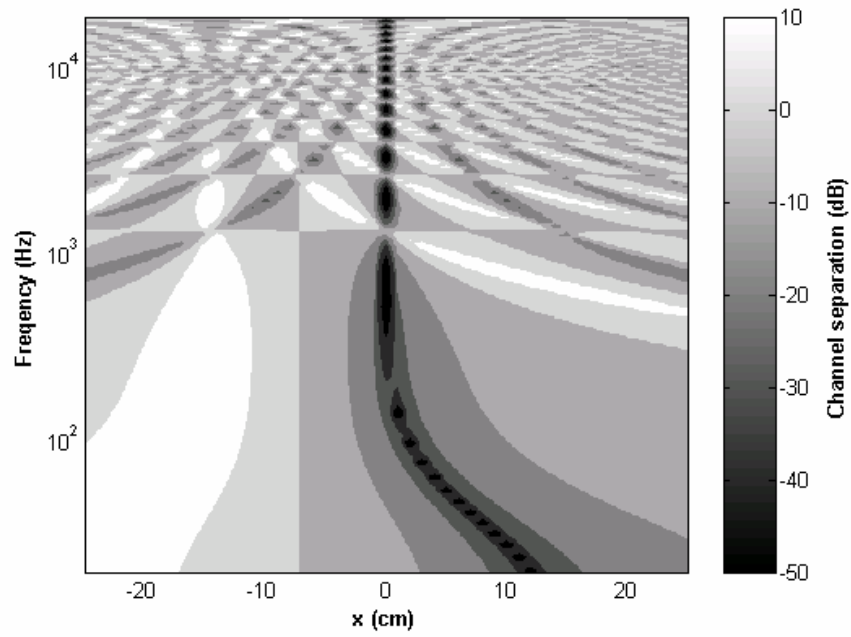
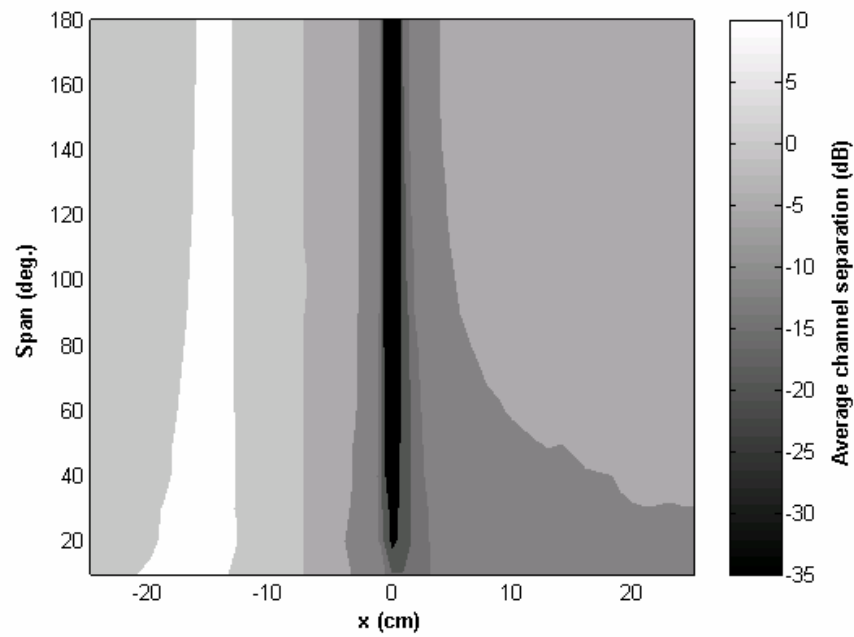
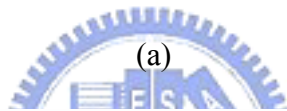
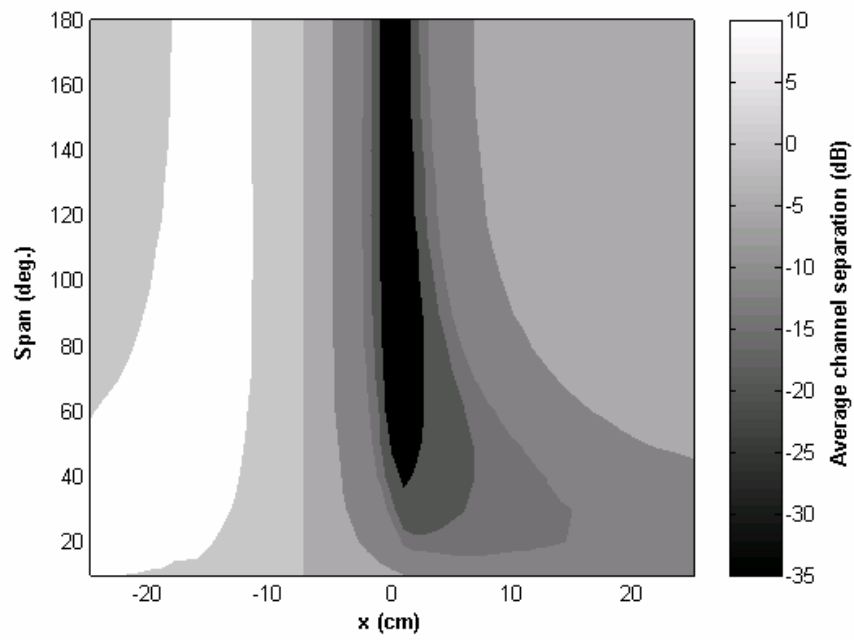


Fig. 5. The contour plots of channel separation at the right ear calculated using the point source model. (a) 10-deg span. (b) 60-deg span. (c) 120-deg span.



(b)

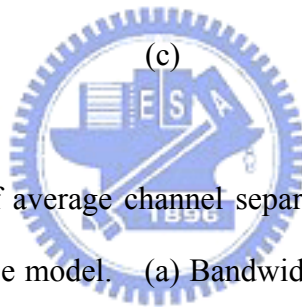
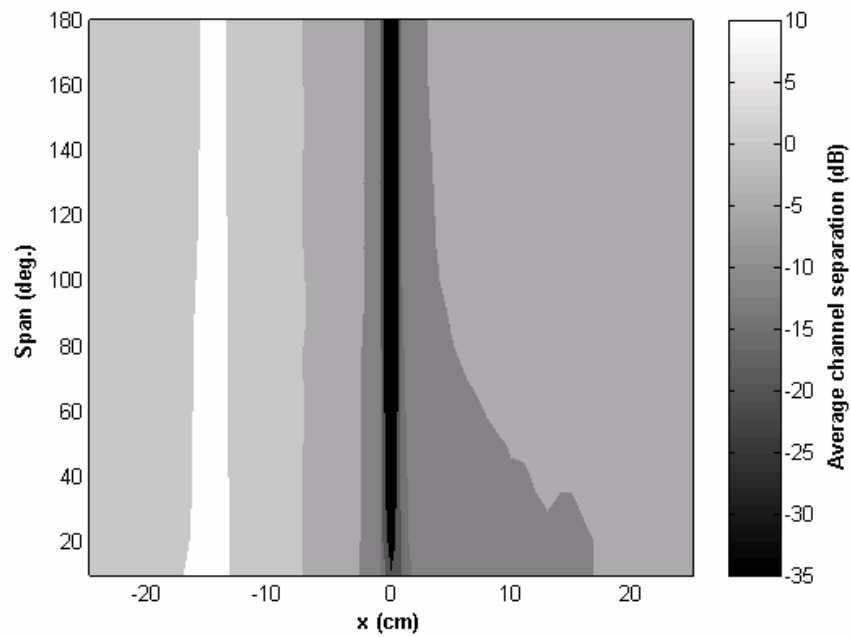
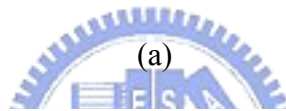
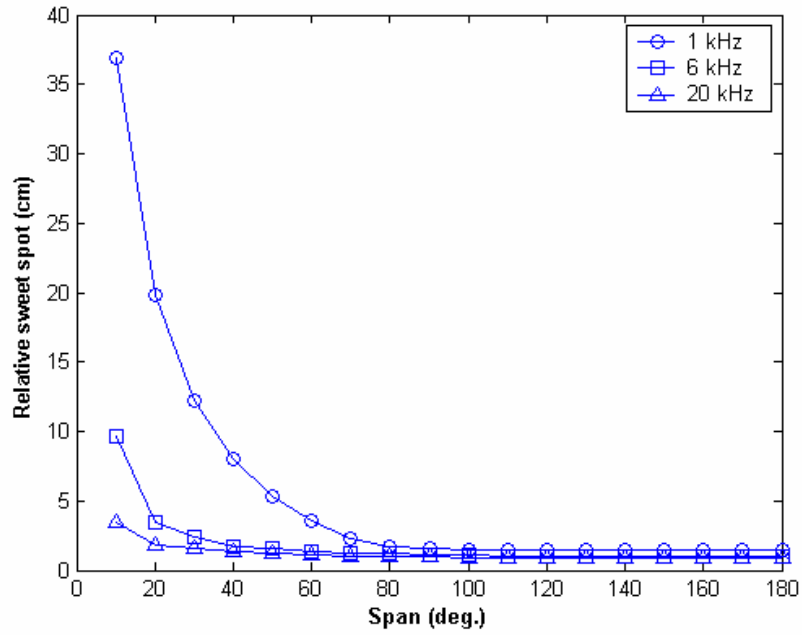
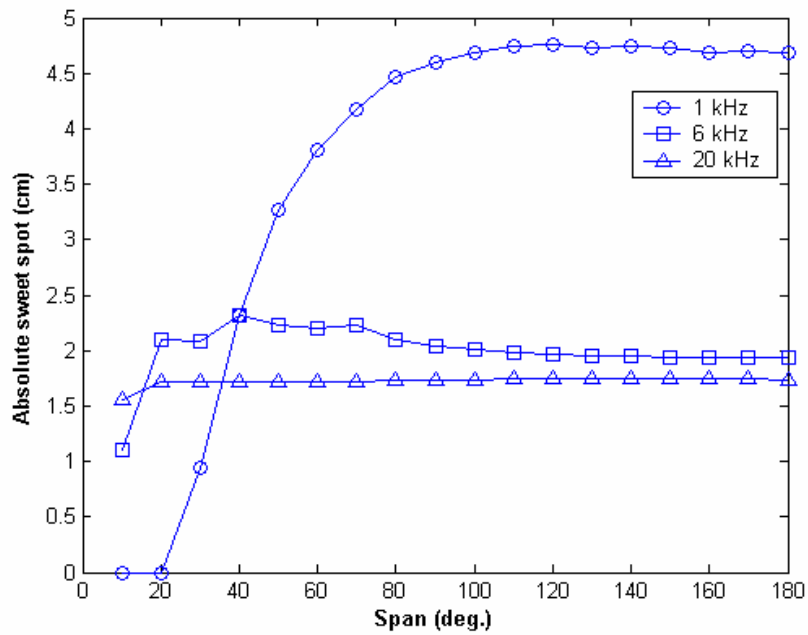


Fig. 6. The contour plots of average channel separation at the right ear calculated using the point source model. (a) Bandwidth to 1 kHz. (b) Bandwidth to 6 kHz. (c) Bandwidth to 20 kHz.

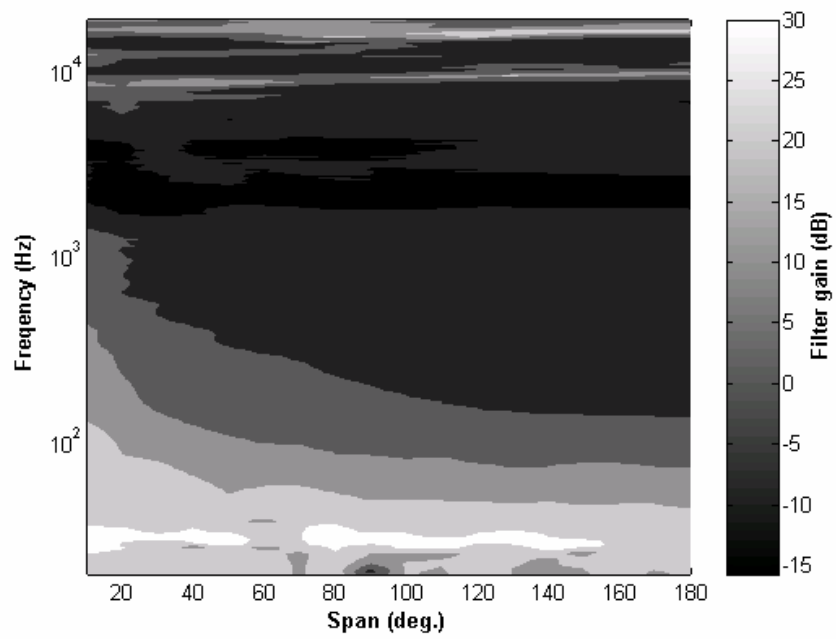
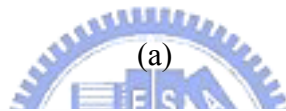
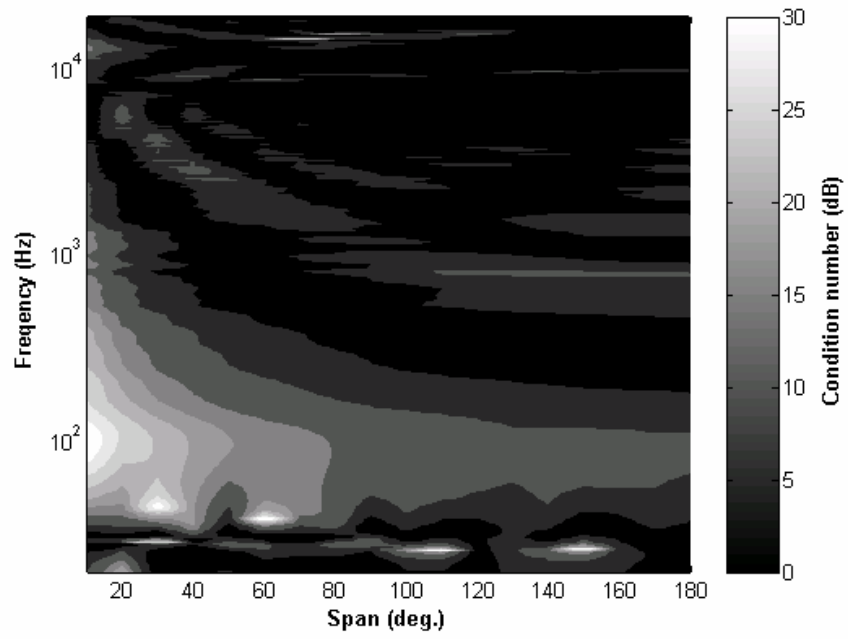


(a)



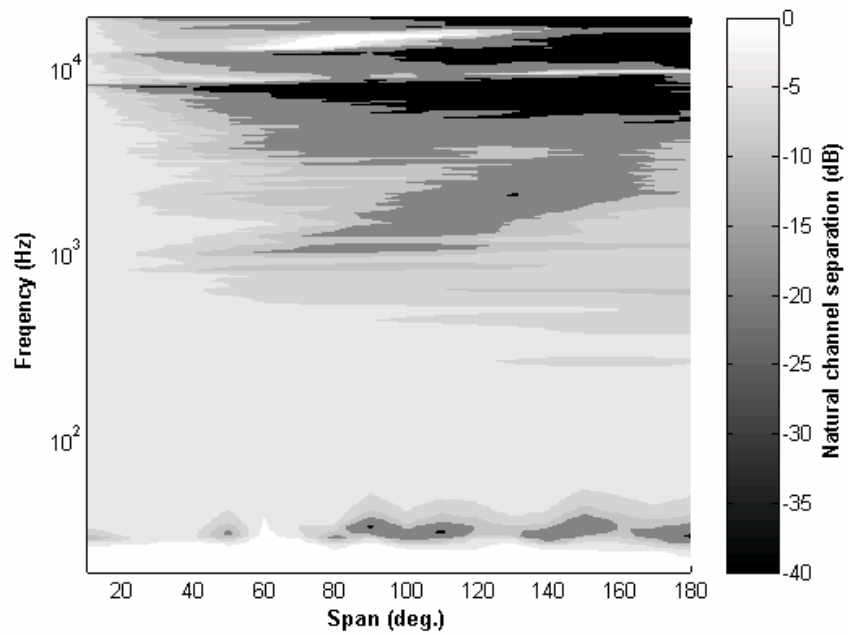
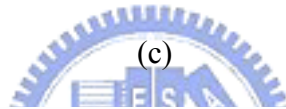
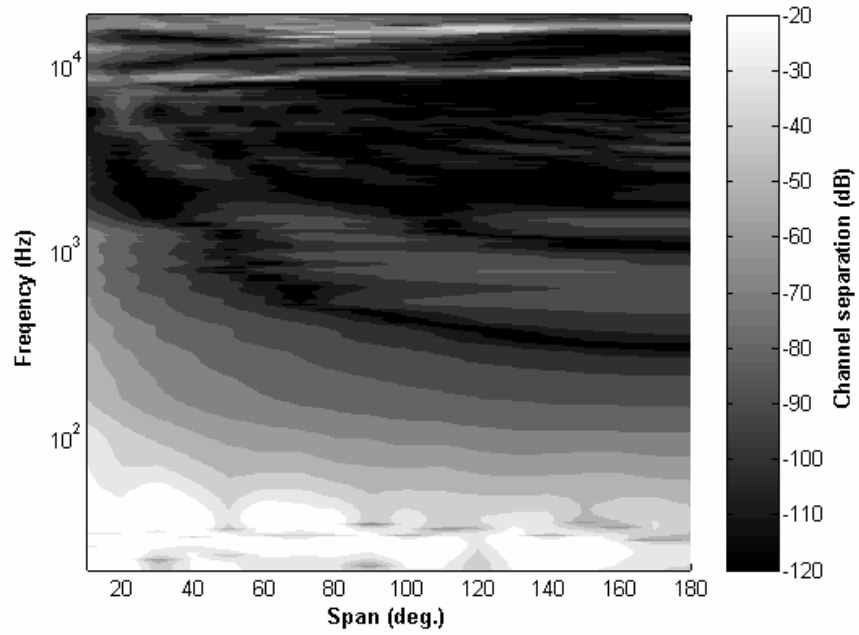
(b)

Fig. 7. Two sweet spot definitions calculated using the point source model for 1, 6, and 10 kHz bandwidths. (a) Relative sweet spot. (b) Absolute sweet spot.



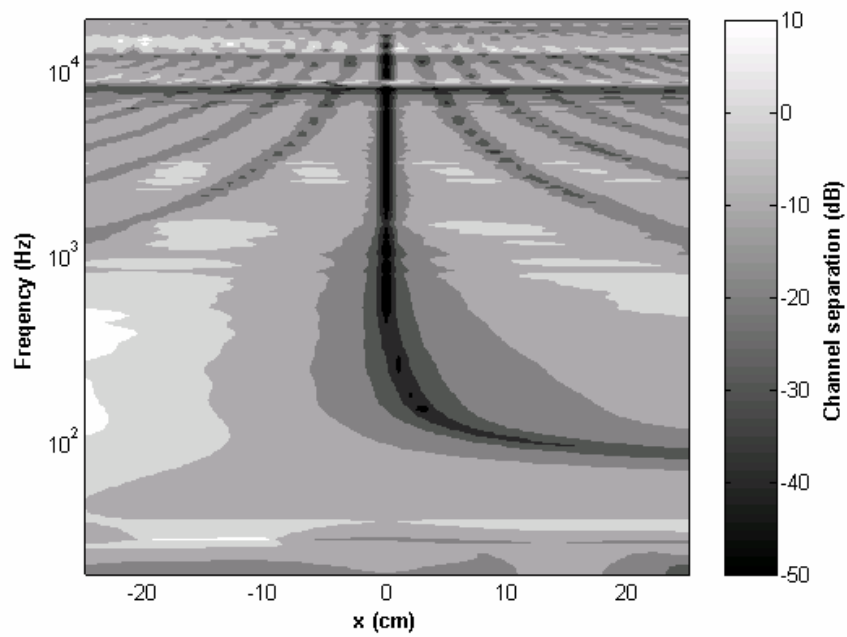
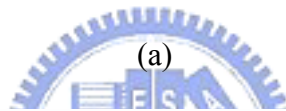
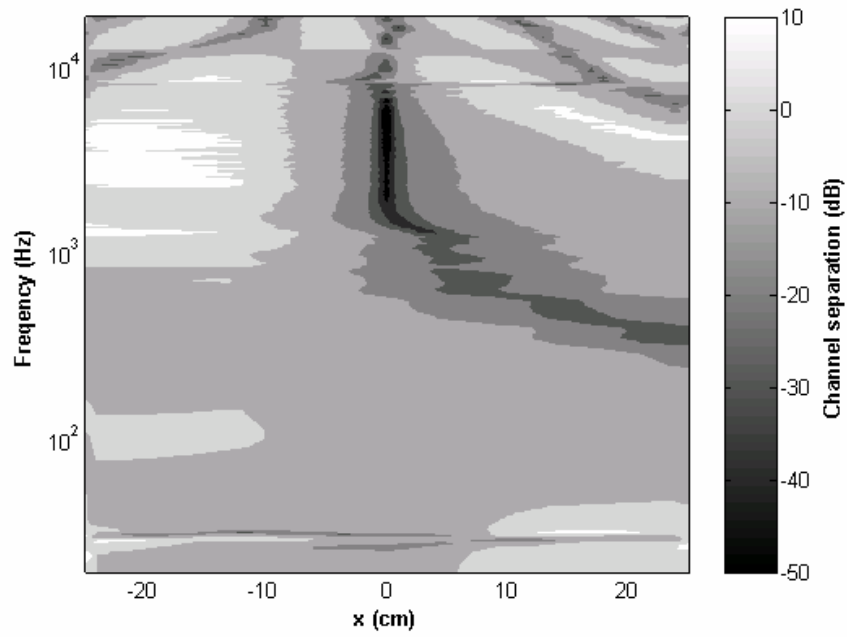
(b)





(d)

Fig. 8. The contour plots calculated using the HRTF model of (a) the condition number of acoustical plant matrix  $\mathbf{H}$ , (b) the filter gain, (c) the channel separation, and (d) the uncompensated natural channel separation.



(b)

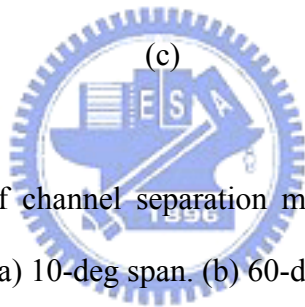
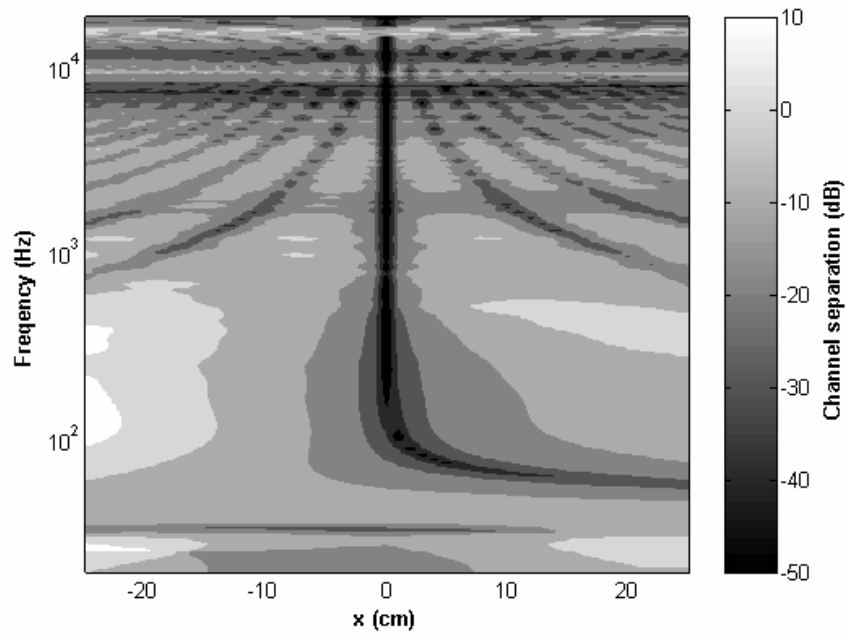
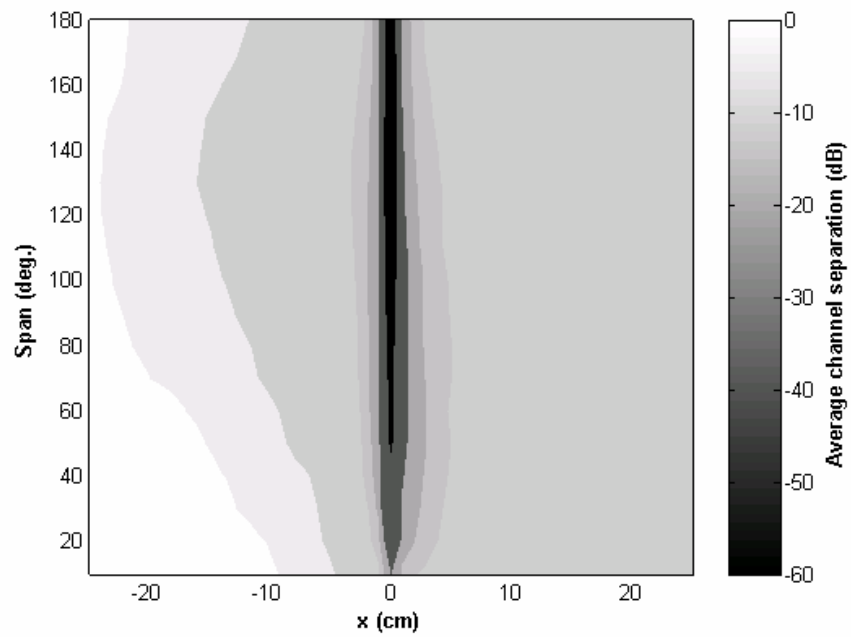
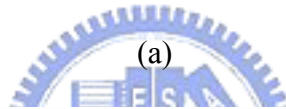
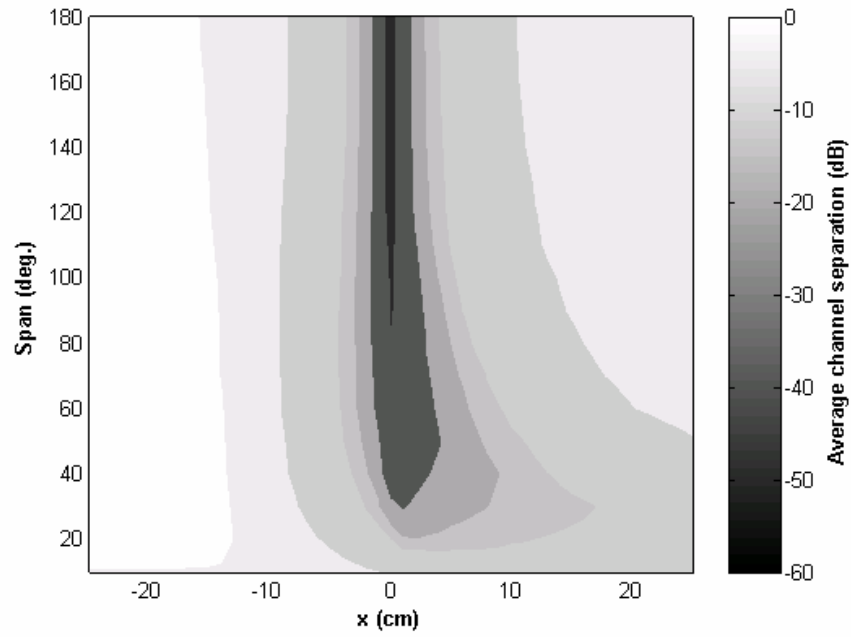


Fig. 9. The contour plots of channel separation measured at the right ear of the acoustic manikin. (a) 10-deg span. (b) 60-deg span. (c) 120-deg span.



(b)

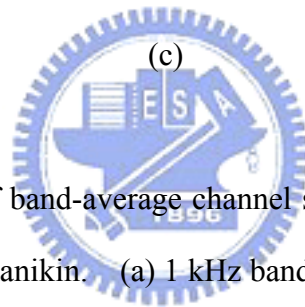
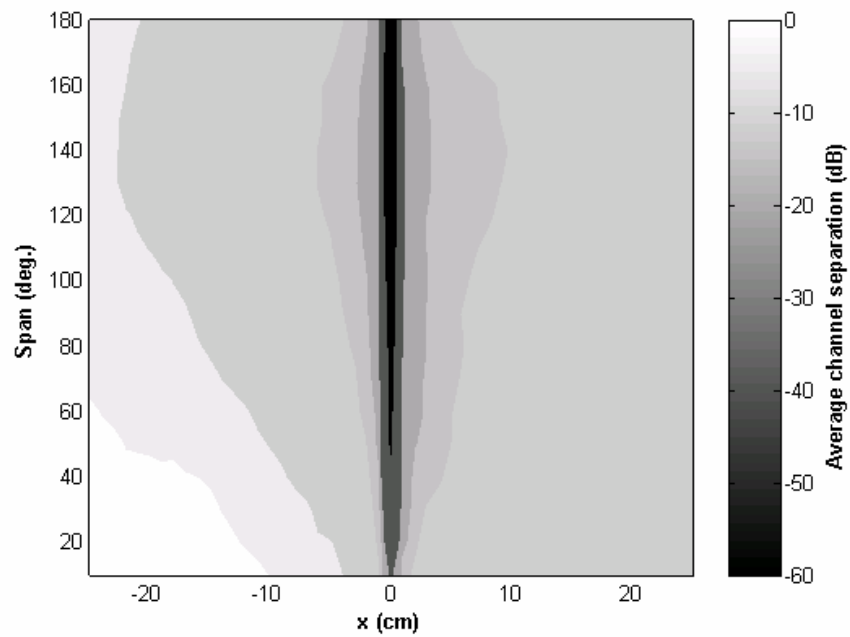
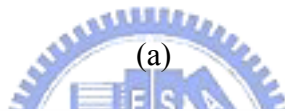
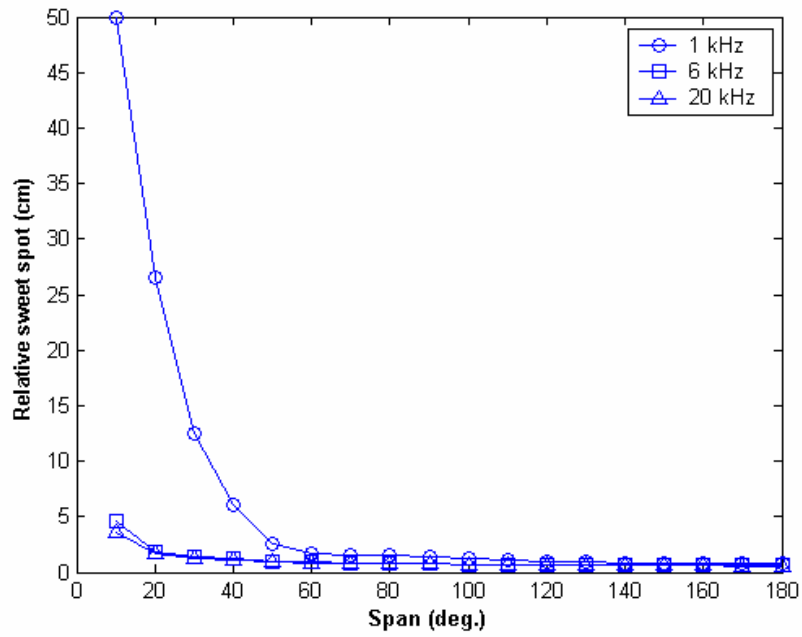
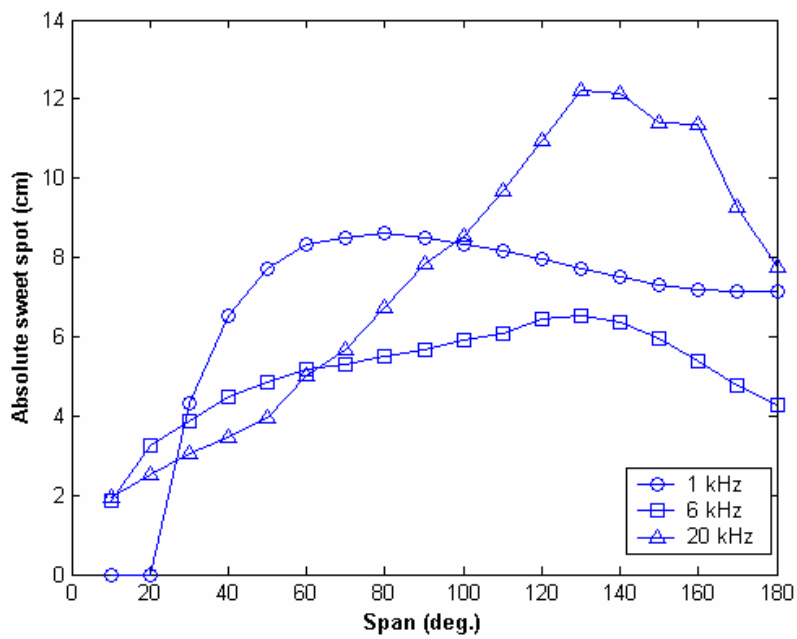


Fig. 10. The contour plots of band-average channel separation measured at the right ear of the acoustic manikin. (a) 1 kHz bandwidth. (b) 6 kHz bandwidth. (c) 20 kHz bandwidth.



(a)



(b)

Fig. 11. Two sweet spot definitions calculated using the HRTF model for 1, 6, and 10 kHz bandwidths. (a) Relative sweet spot. (b) Absolute sweet spot.

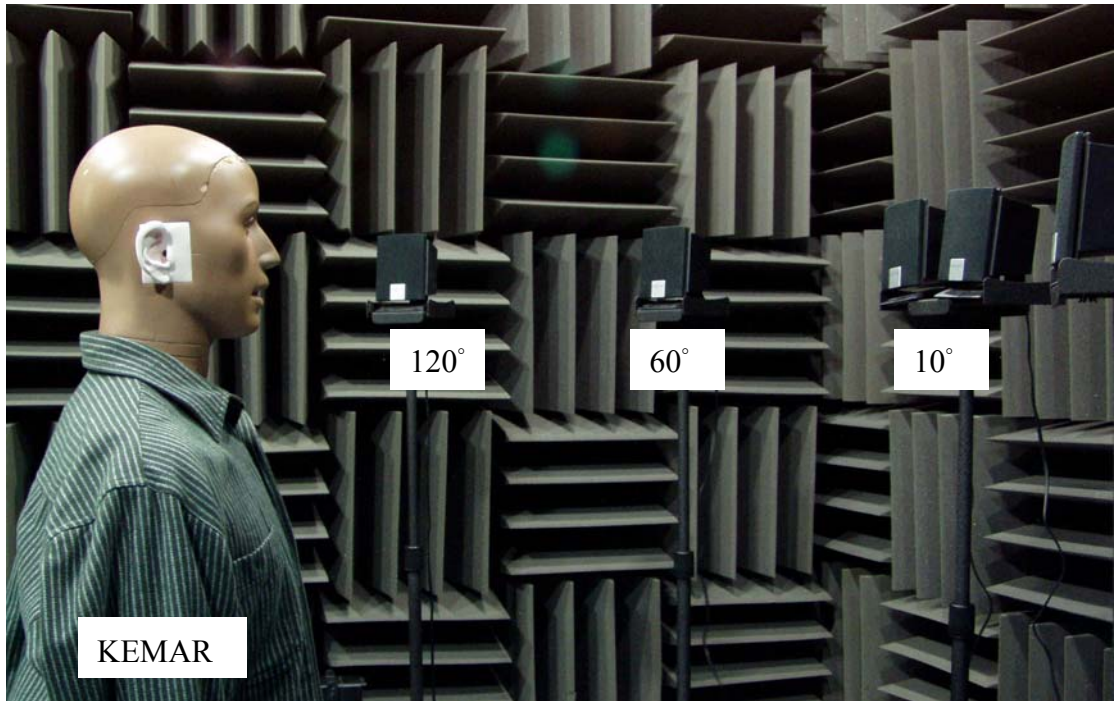
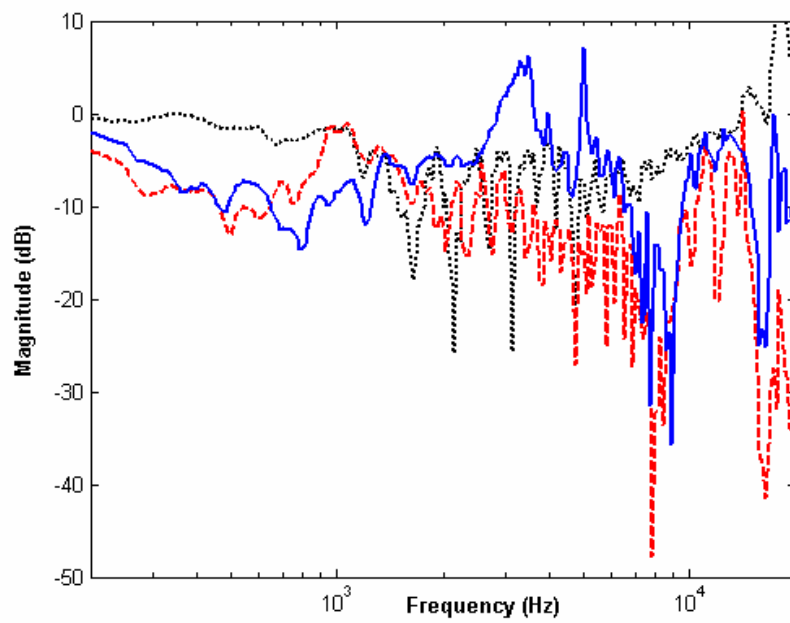
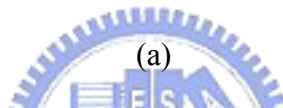
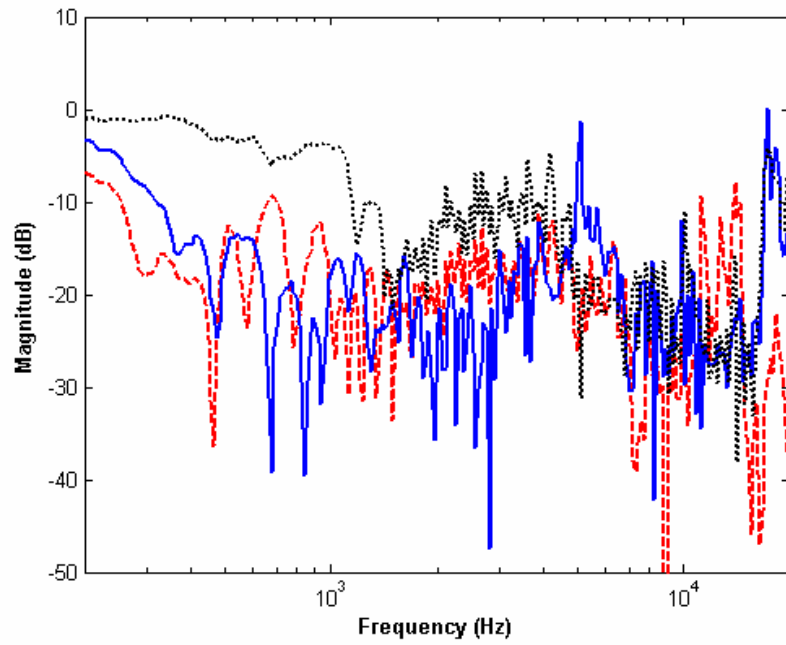


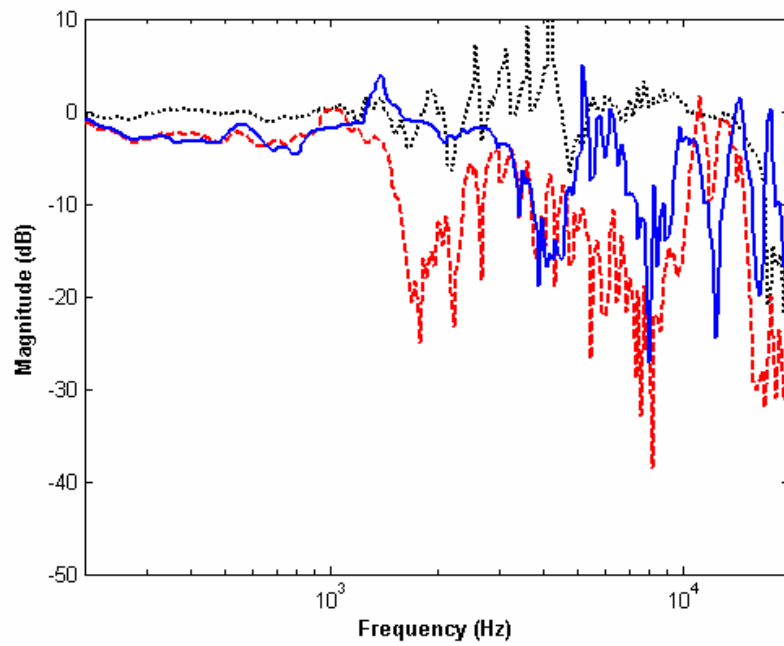
Fig. 12. Photo of the experimental arrangement.





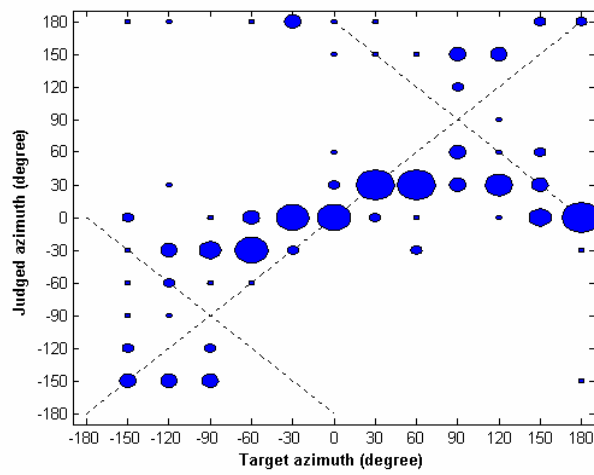
(b)



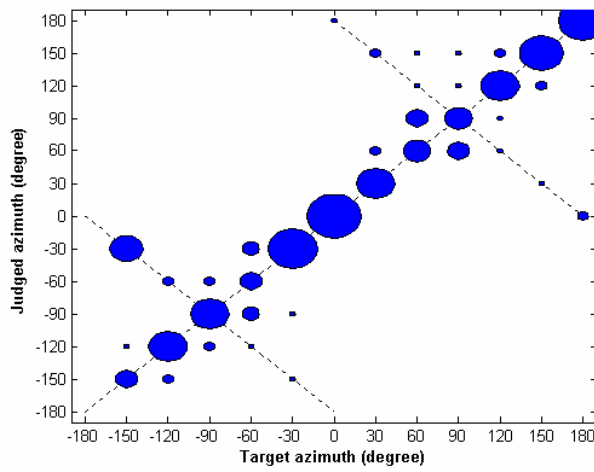


(c)

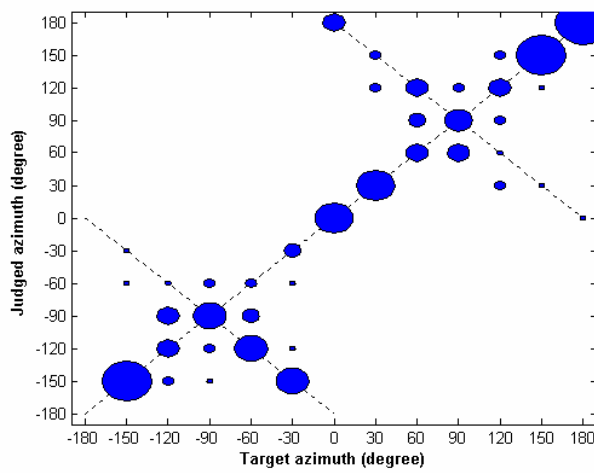
Fig. 13. Channel separations measured at the right ear of the acoustic manikin. The dotted lines, solid lines, and dashed lines represent 10-deg, 60-deg, and 120-deg spans, respectively. (a) In the nominal position ( $x = 0\text{cm}$ ). (b) Rightward 5cm displacement. (c) Rightward 10cm displacement.



(a)

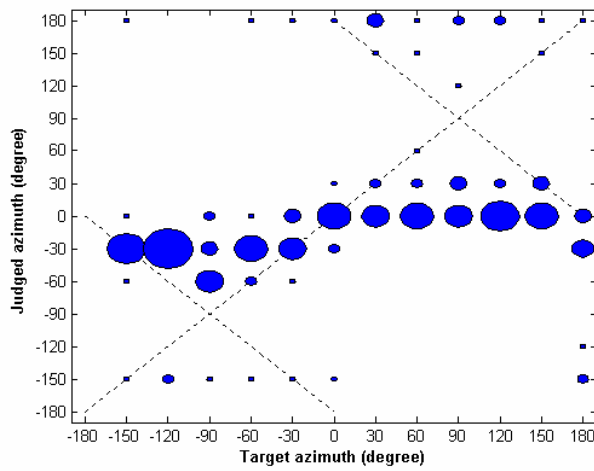


(b)

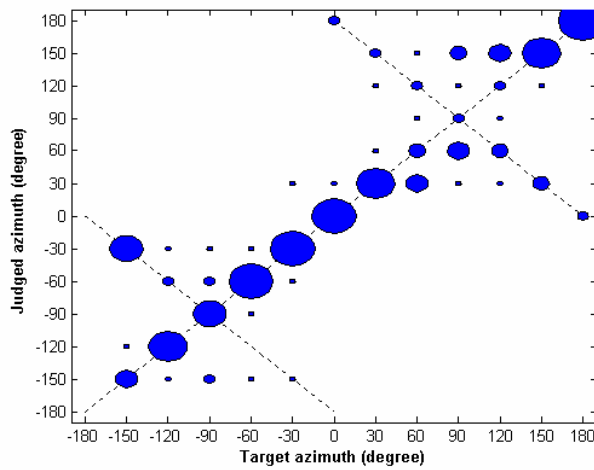


(c)

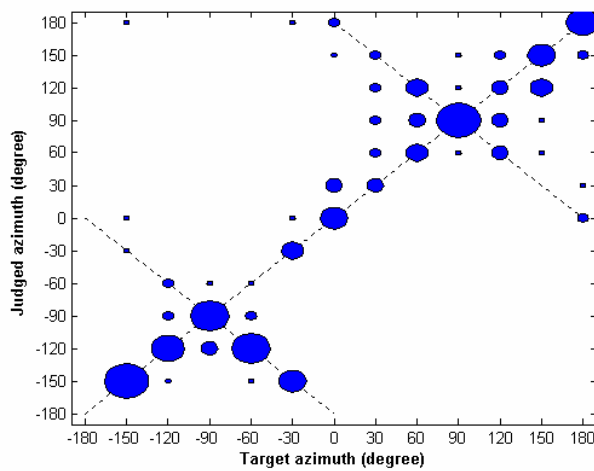
Fig. 14. Results of the subjective localization test of azimuth angles with no head displacement. (a) 10-deg span. (b) 60-deg span. (c) 120-deg span.



(a)

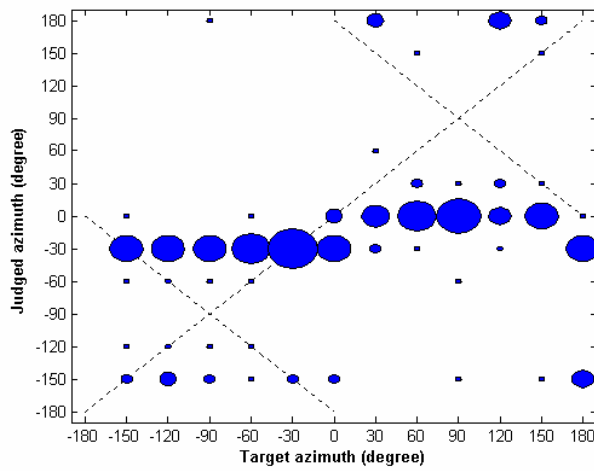


(b)

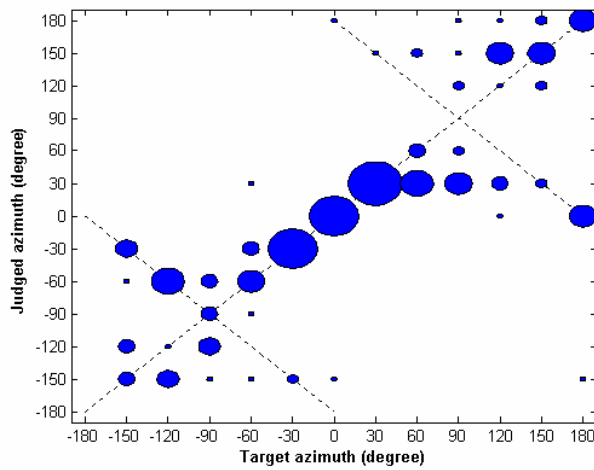


(c)

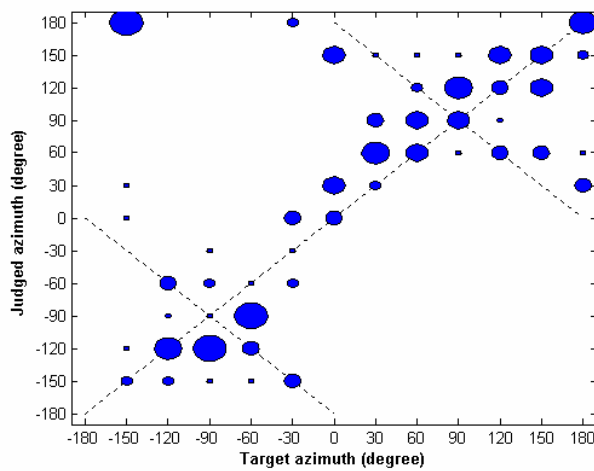
Fig. 15. Results of the subjective localization test of azimuth angles with 5-cm head displacement to the right. (a) 10-deg. (b) 60-deg. (c) 120-deg. span.



(a)

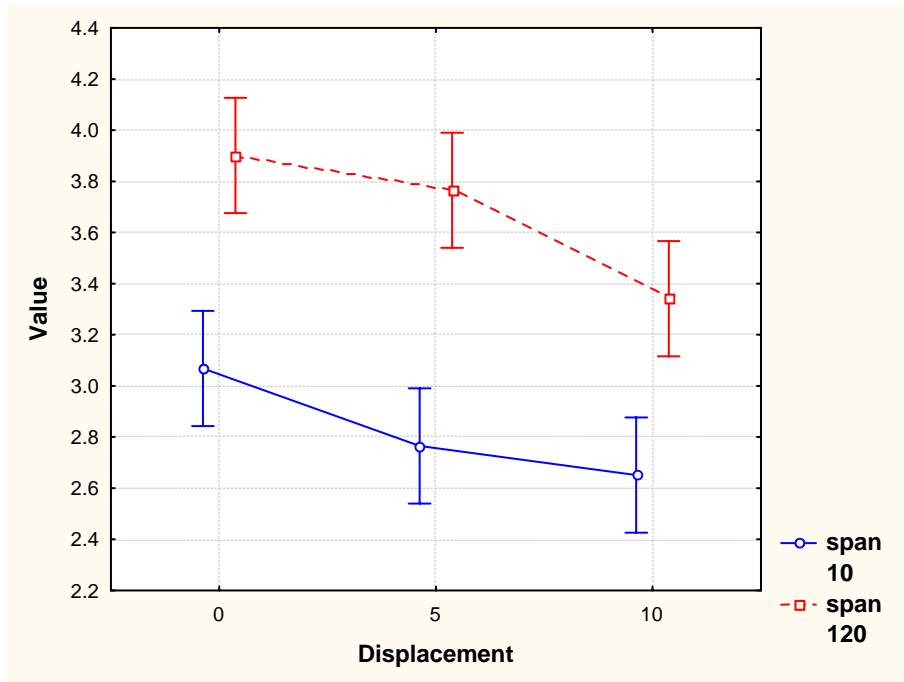


(b)

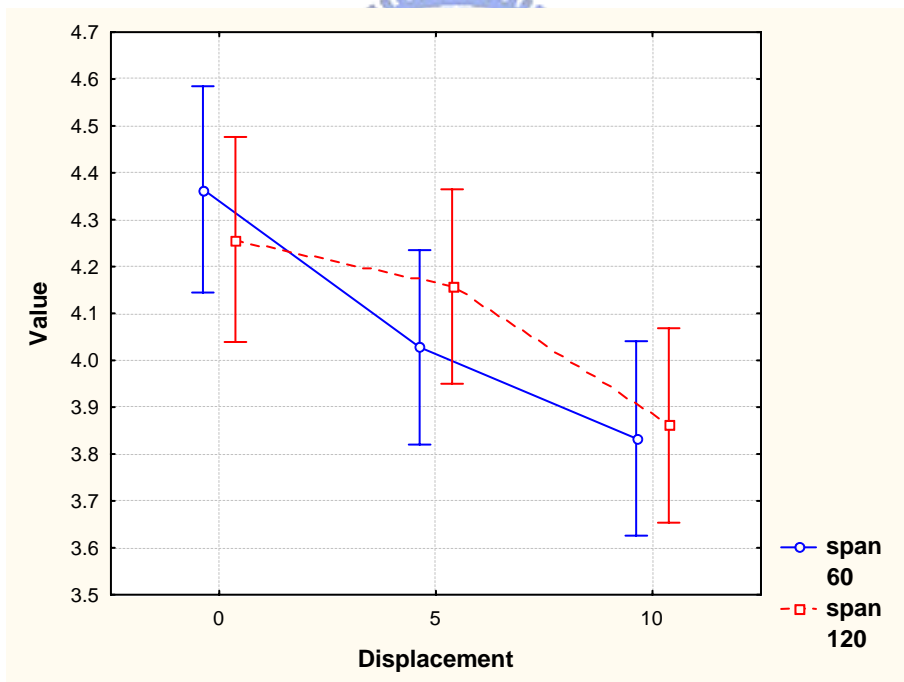


(c)

Fig. 16. Results of the subjective localization test of azimuth angles with 10-cm head displacement to the right. (a) 10-deg (b) 60-deg (c) 120-deg span.



(a)



(b)

Fig. 17. Means and spreads (with 95% confidence intervals) of the grades for three kinds of head displacements. (a) Grades of the 10-deg arrangement (solid line) and the 120-deg arrangement (dotted line). (b) Grades of the 60-deg arrangement (solid line) and the 120-deg arrangement (dotted line).

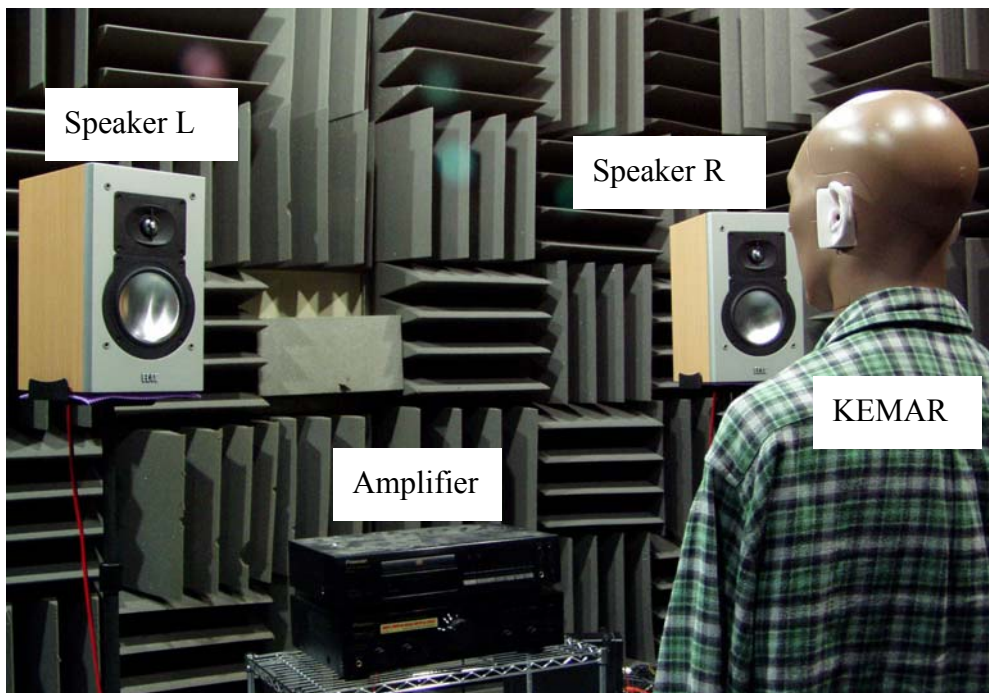


Fig. 18 The experimental configuration.

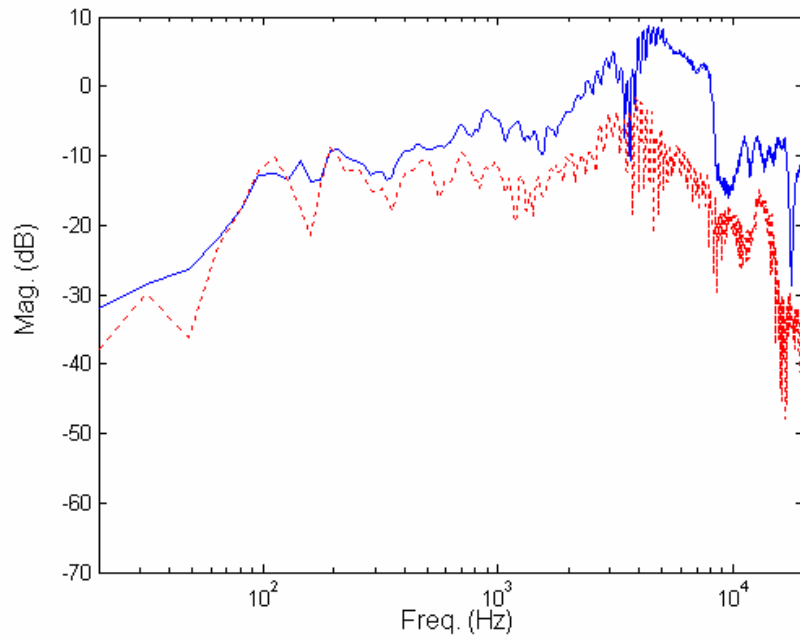
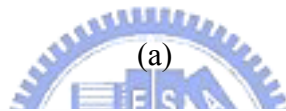
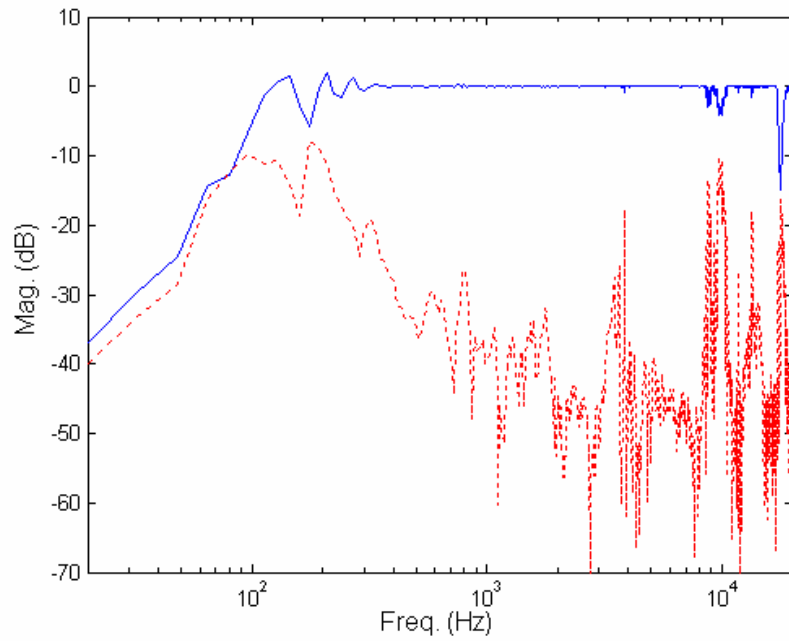
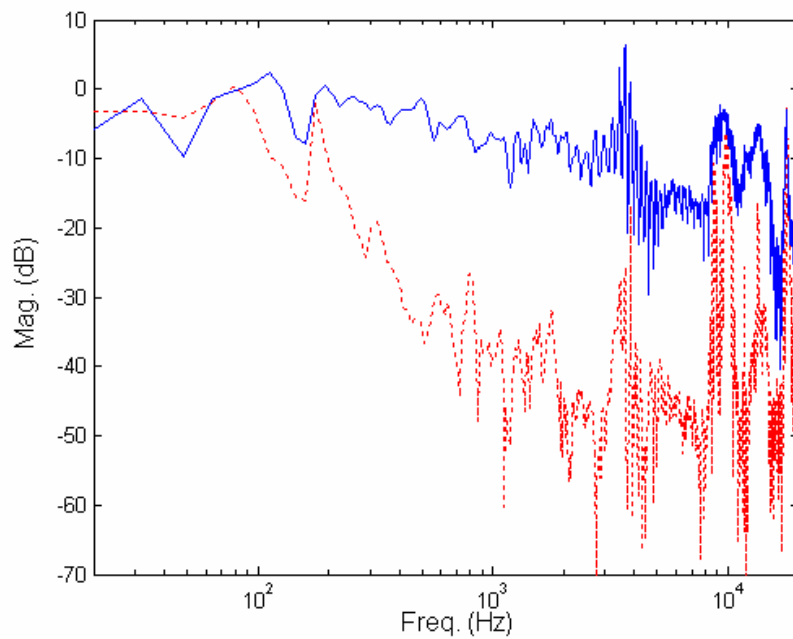


Fig. 19 The frequency responses of the plants including ipsilateral (solid line) and contralateral paths (dotted line).



(a)



(b)

Fig. 20 (a) The frequency responses of  $Q_{11f}$  (solid line) and  $Q_{12f}$  (dotted line). (b) Natural channel separation (solid line) and compensated channel separation (dotted line).



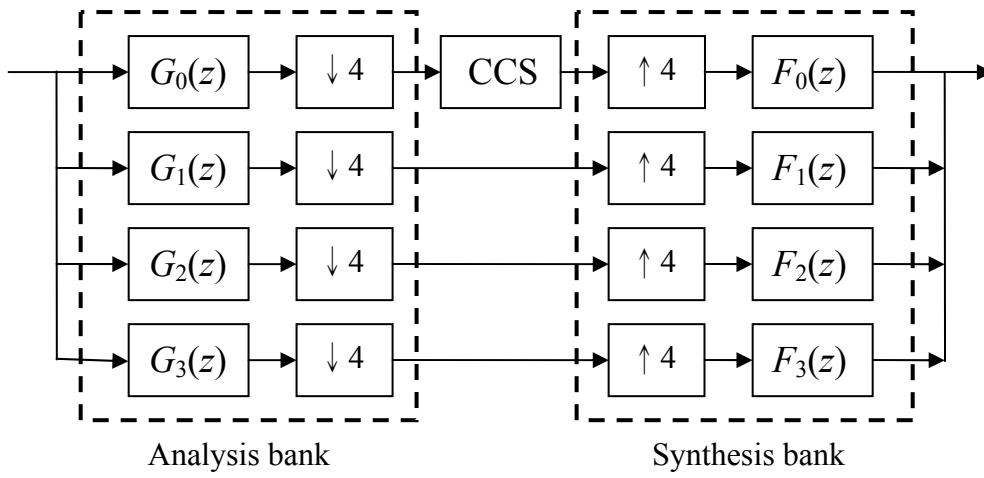
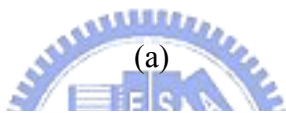
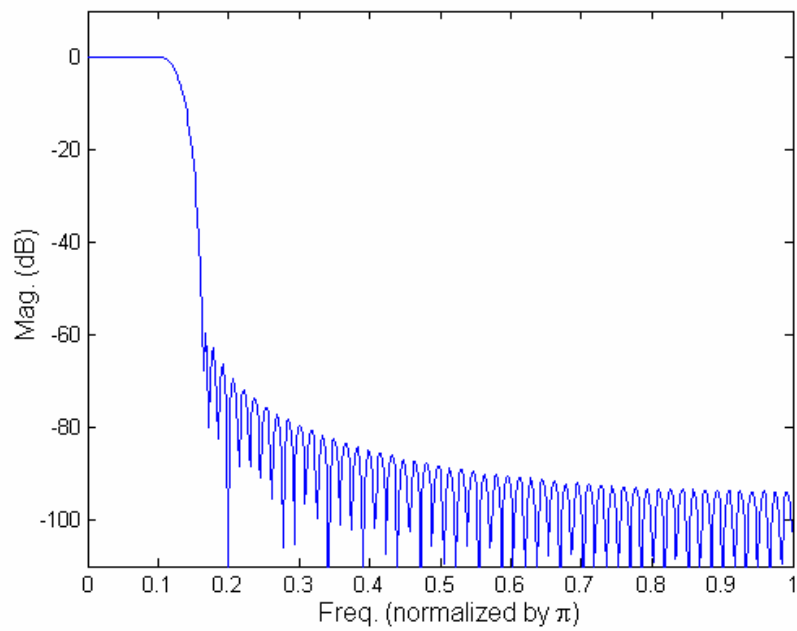
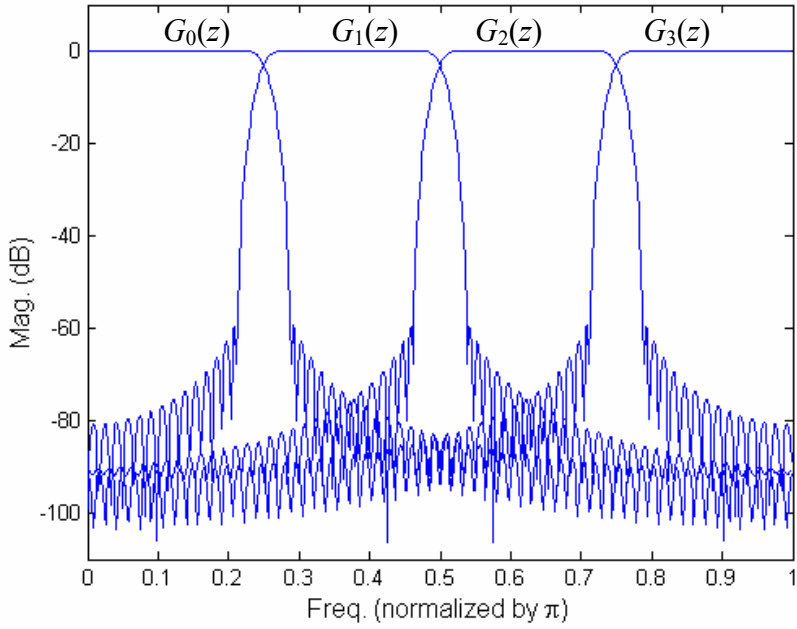


Fig. 21 The block diagram of the bandlimited CCS.



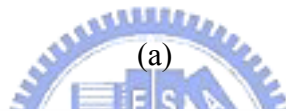
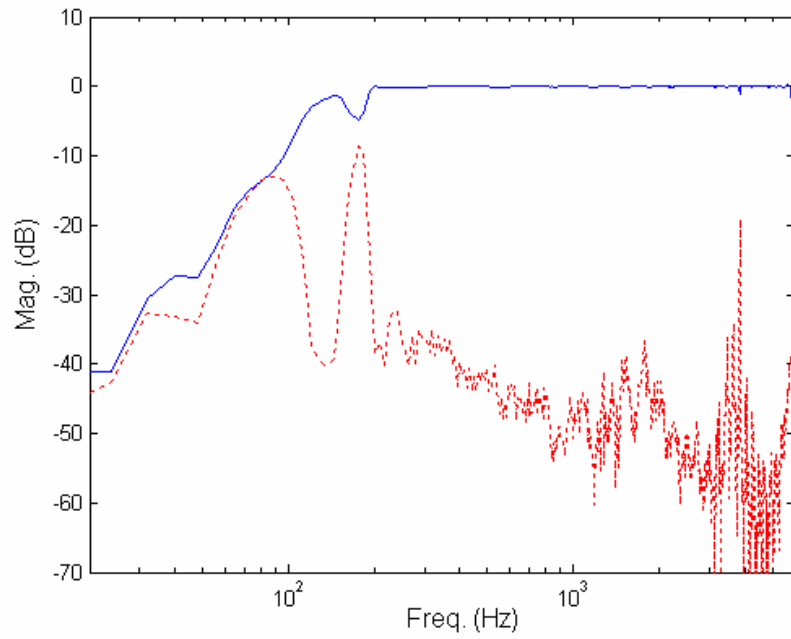


(a)

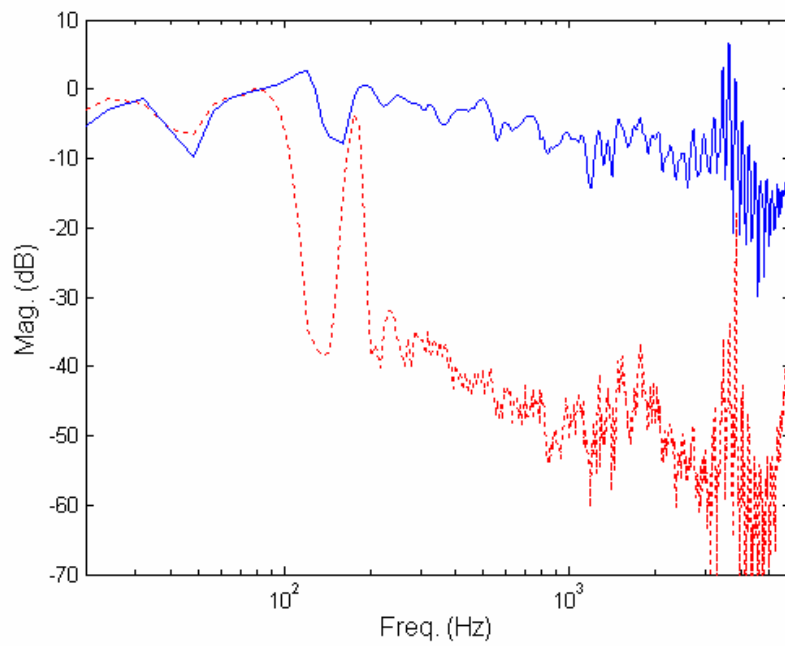


(b)

Fig. 22 The magnitude responses of (a) Prototype FIR filter and (b) Analysis bank.

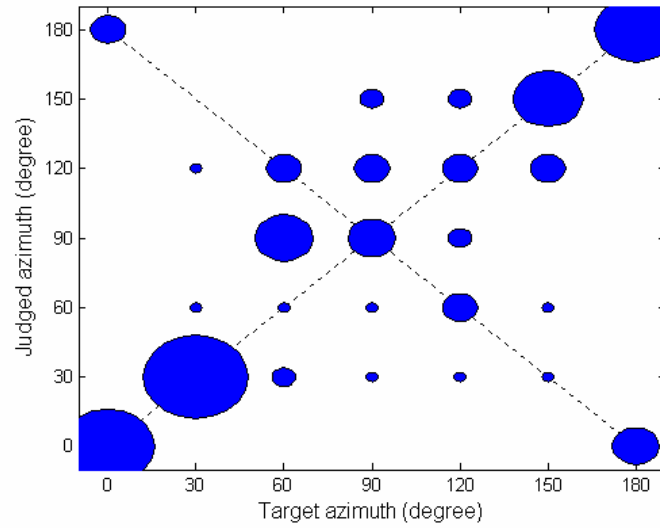


(a)

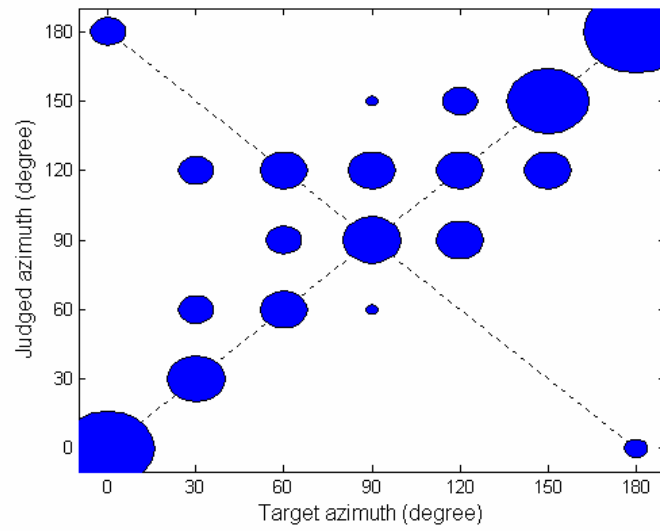


(b)

Fig. 23 (a) The frequency responses of  $Q_{11b}$  (solid line) and  $Q_{12b}$  (dotted line). (b) Natural channel separation (solid line) and compensated channel separation (dotted line).

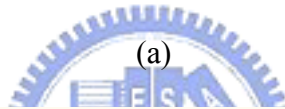
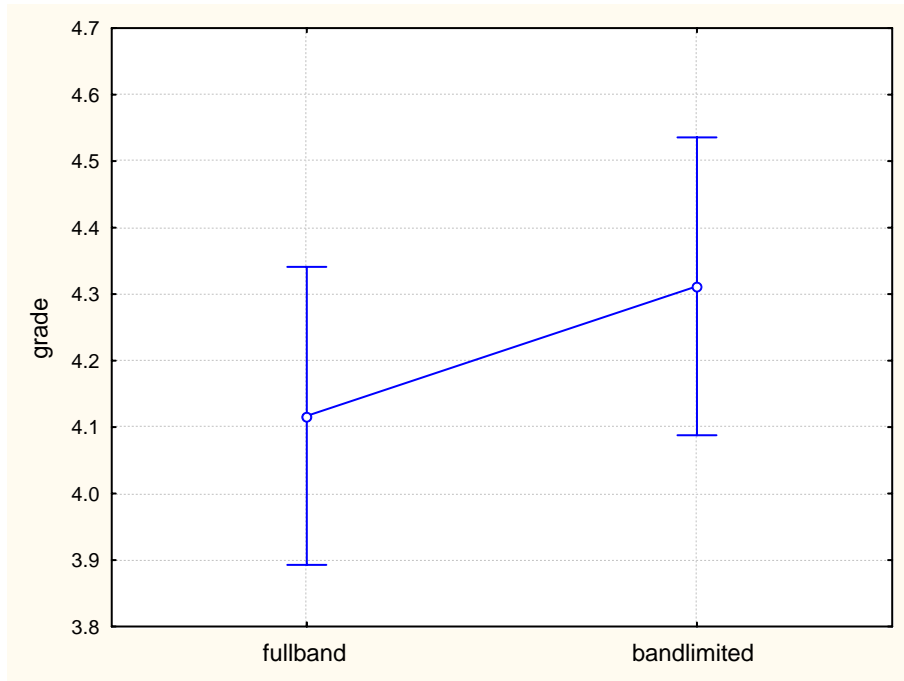


(a)

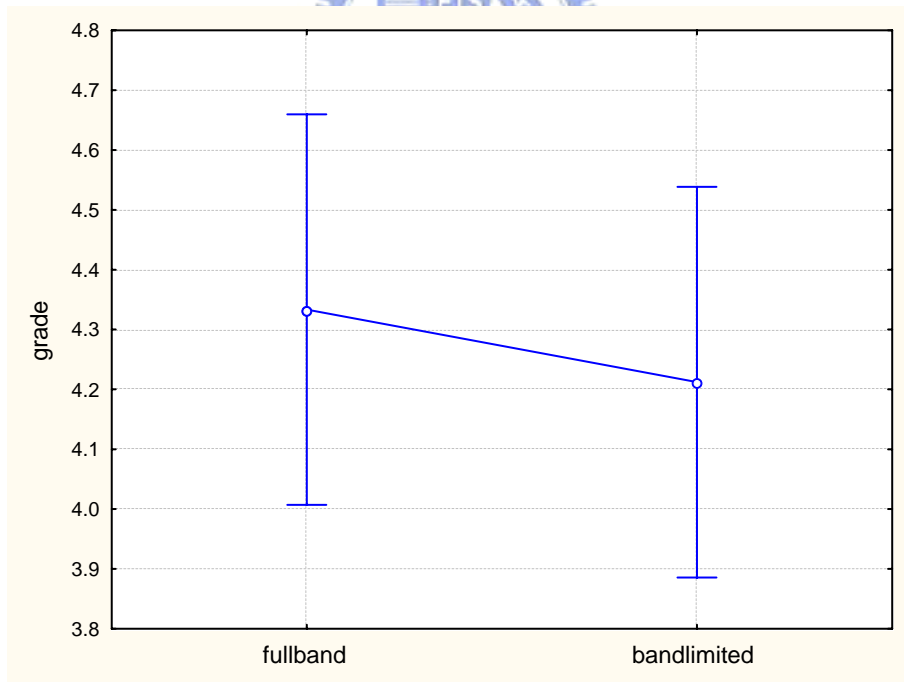


(b)

Fig. 24 Results of the subjective localization test of azimuth. (a) Fullband CCS. (b) Bandlimited CCS.



(a)



(b)

Fig. 25 Means and spreads (with 95% confidence intervals) of the grades for two kinds of CCS approaches. (a) Grades of the source localization experiment. (b) Grades of the sound quality tests.

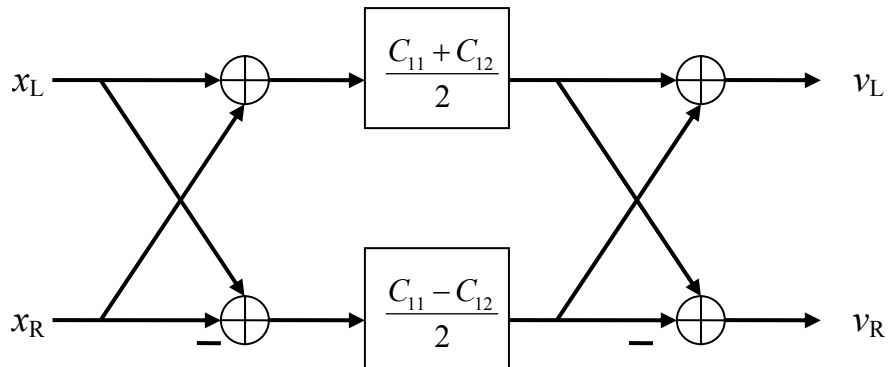


Fig. 26 Shuffler filter structure for 2x2 CCS.



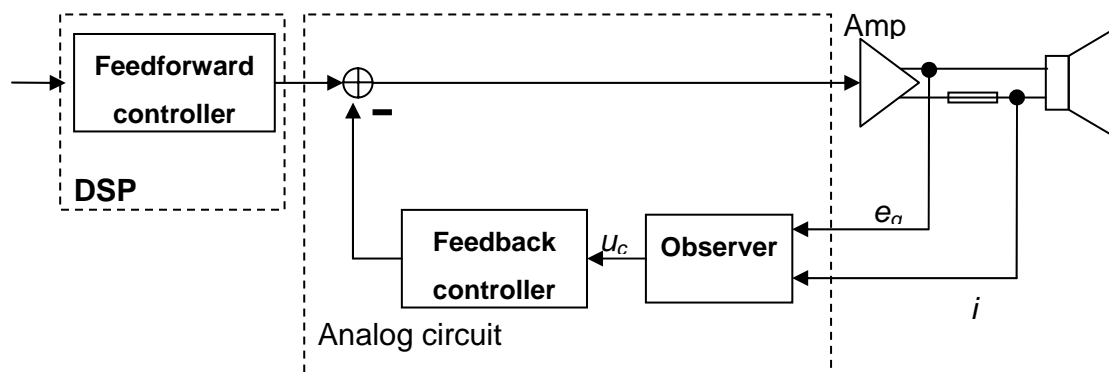


Fig. 27 Hybrid control architecture



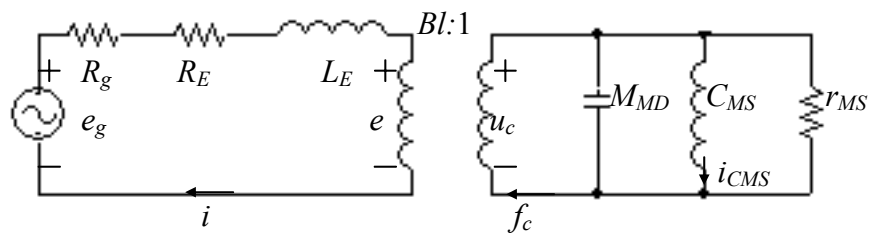


Fig. 28 Electromechanical analogous circuit.





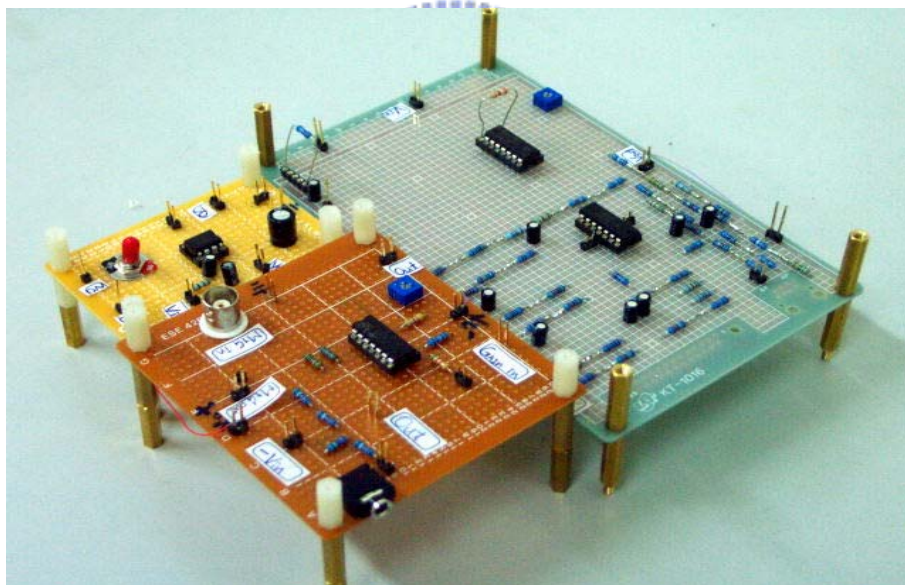
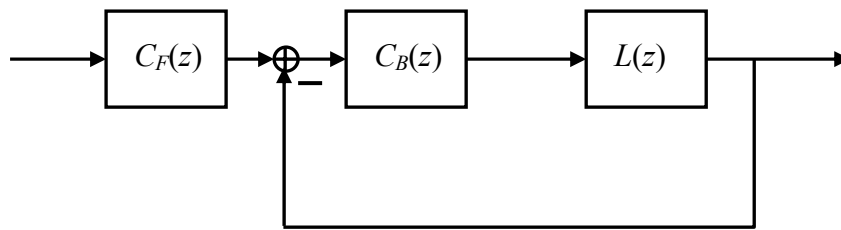
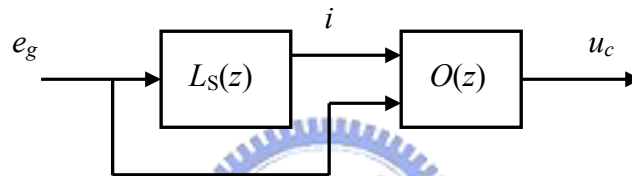


Fig. 29 The experimental arrangement of the cone velocity observer.



(a)



(b)

Fig. 30 (a) The hybrid structure composed of a feedforward  $C_F(z)$  controller and a feedback controller  $C_B(z)$ . (b) The detail structure of the plant.

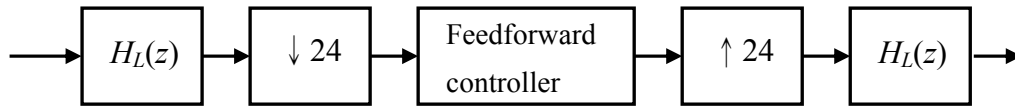


Fig. 31 Block diagram of the down-sampling procedure.



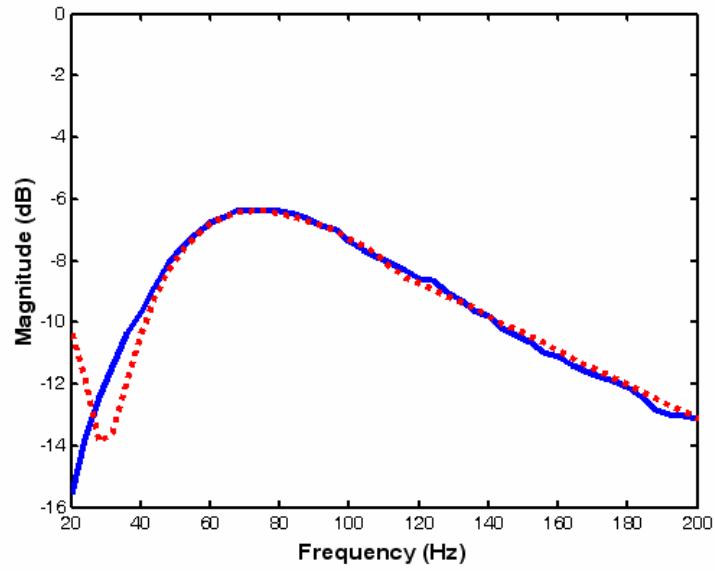
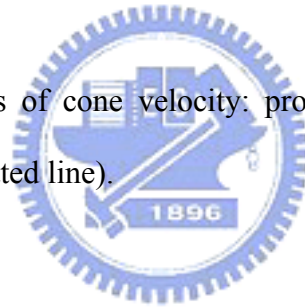


Fig. 32 Frequency responses of cone velocity: proposed observer (solid line) vs. laser vibrometer (dotted line).



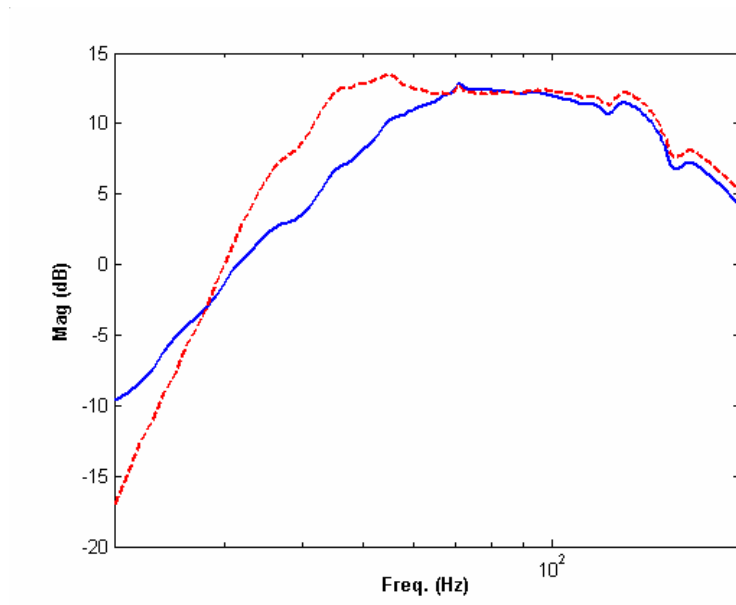


Fig. 33 Frequency responses of sound pressure: uncompensated (solid line) vs. compensated (dotted line).



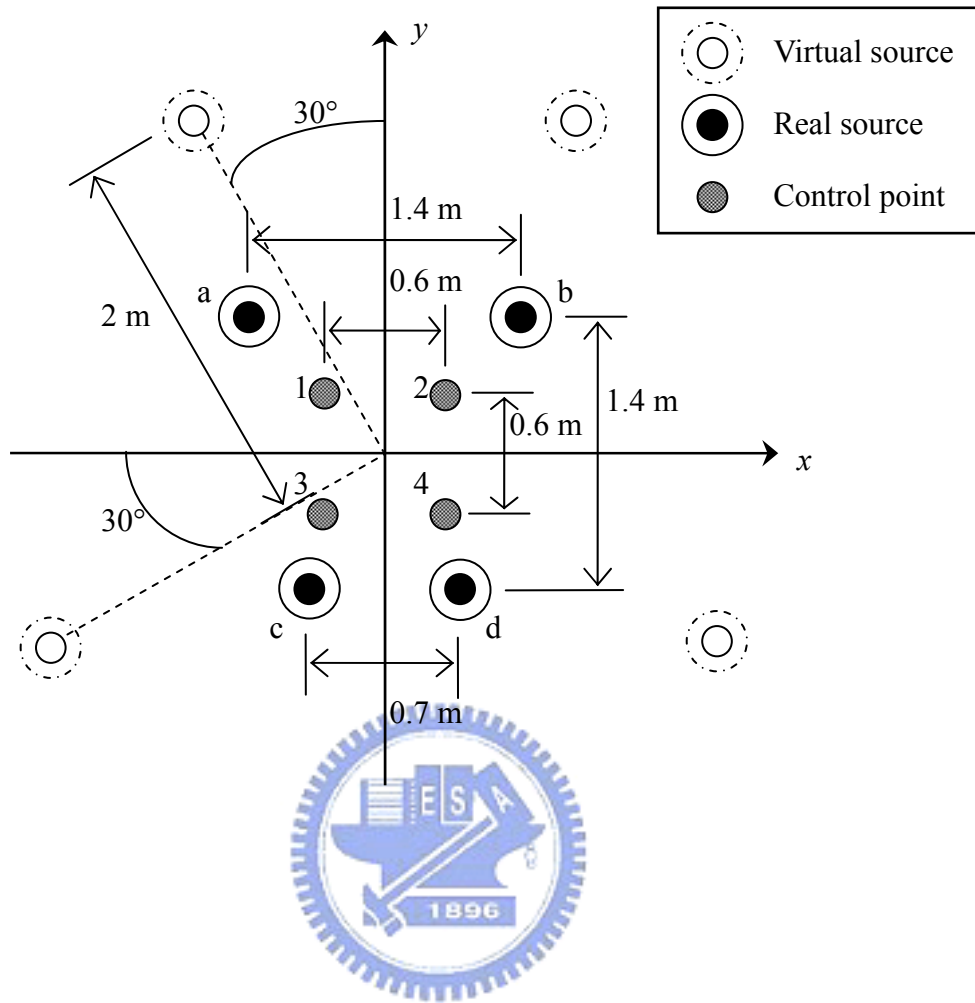
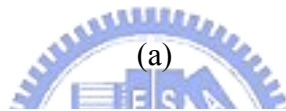
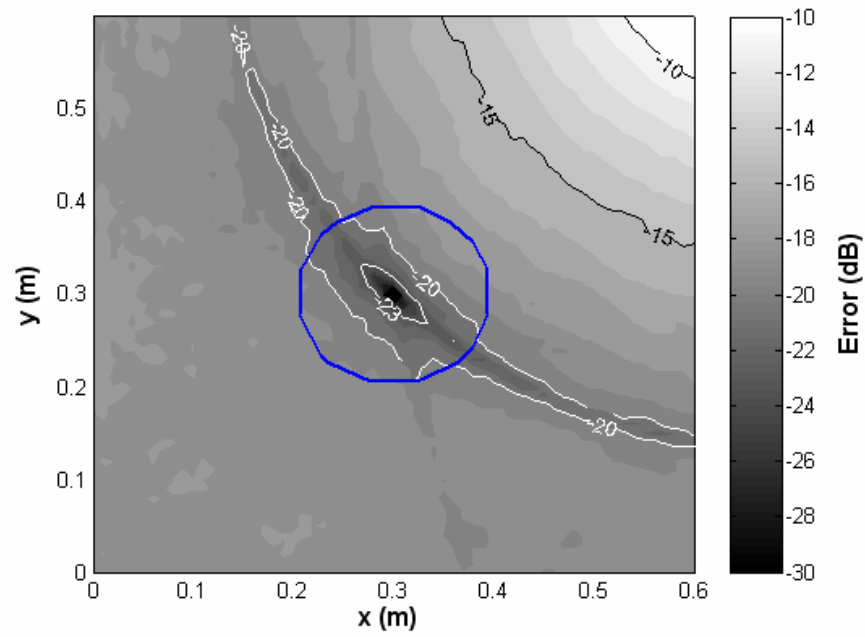
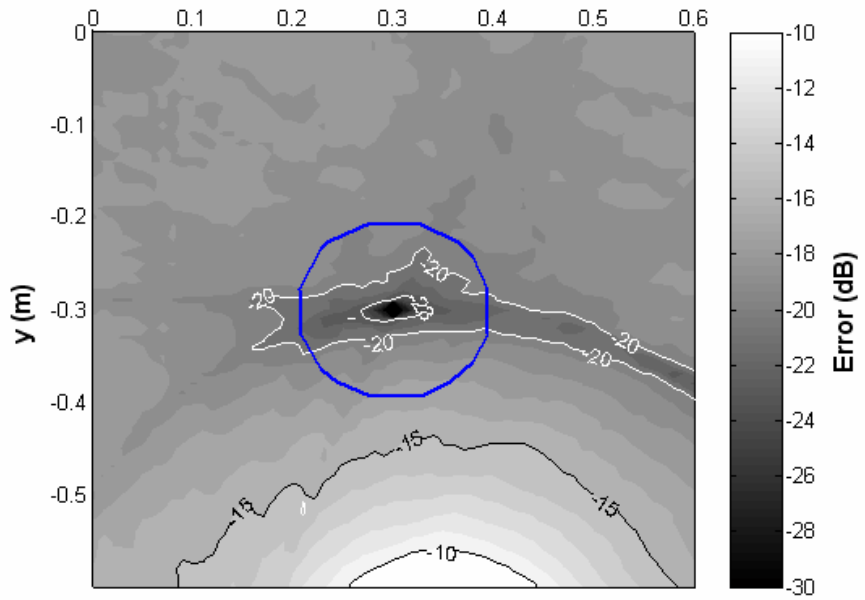


Fig. 34 The geometry of the free-field point source model.

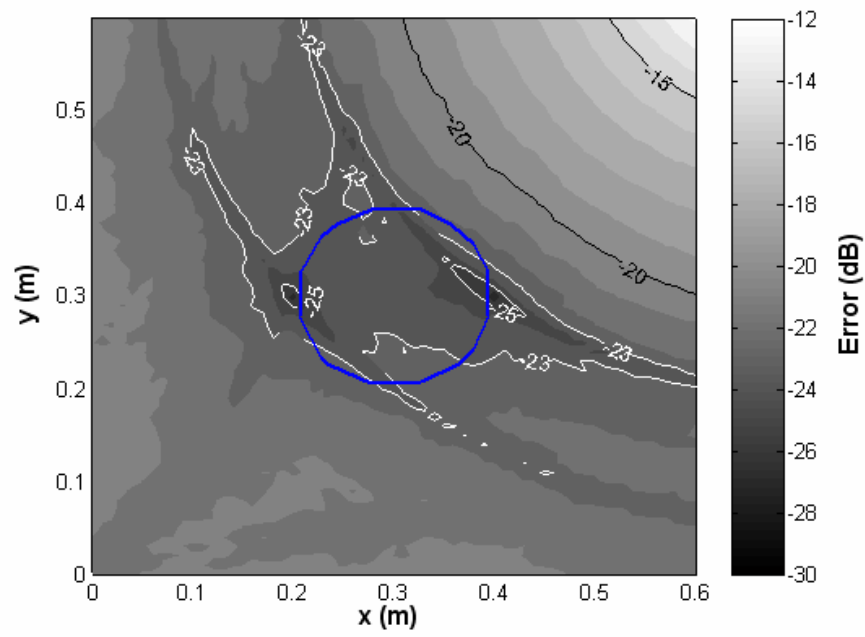


(a)

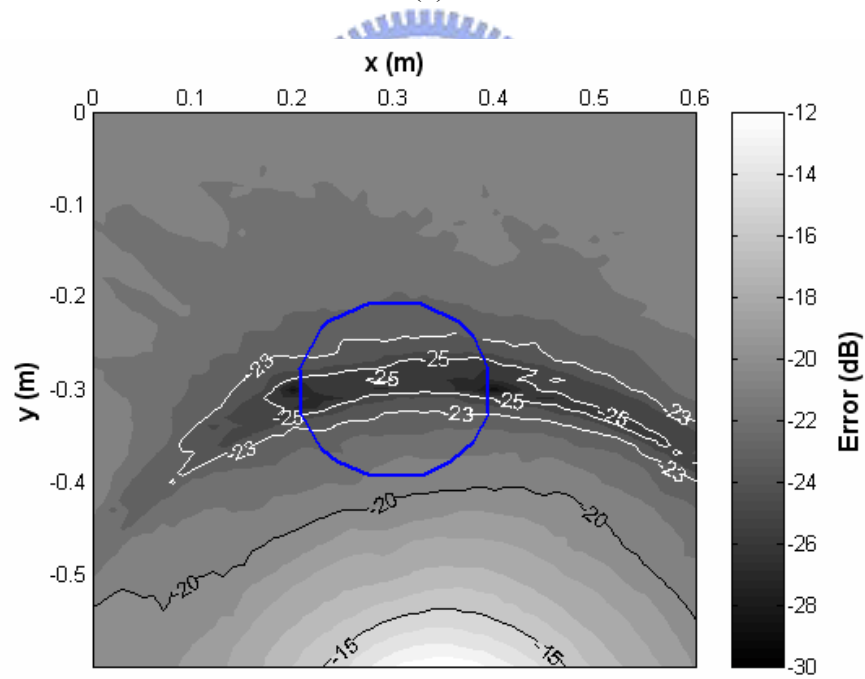


(b)

Fig. 35 The contour plots of the performance of Case I. (a) Front-right seat (b) Rear-right seat.



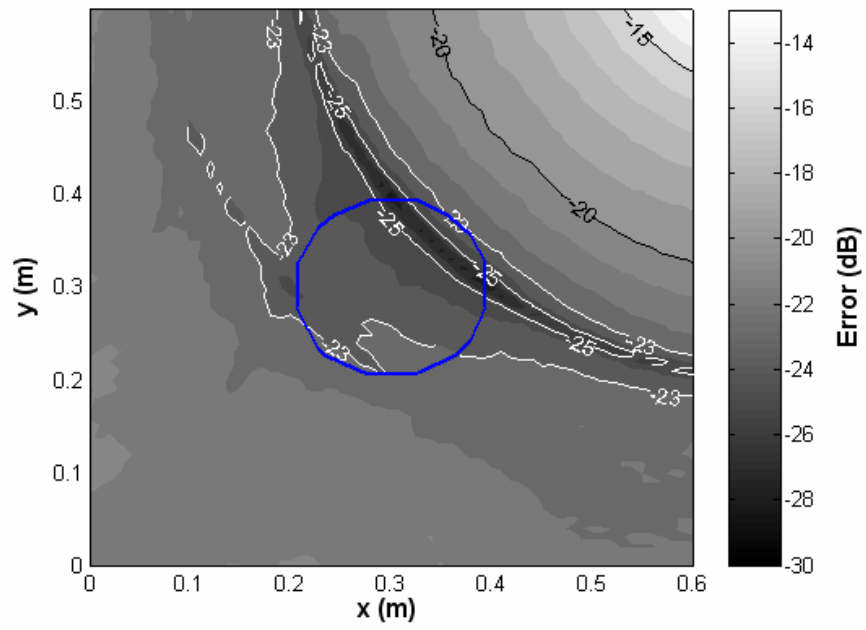
(a)



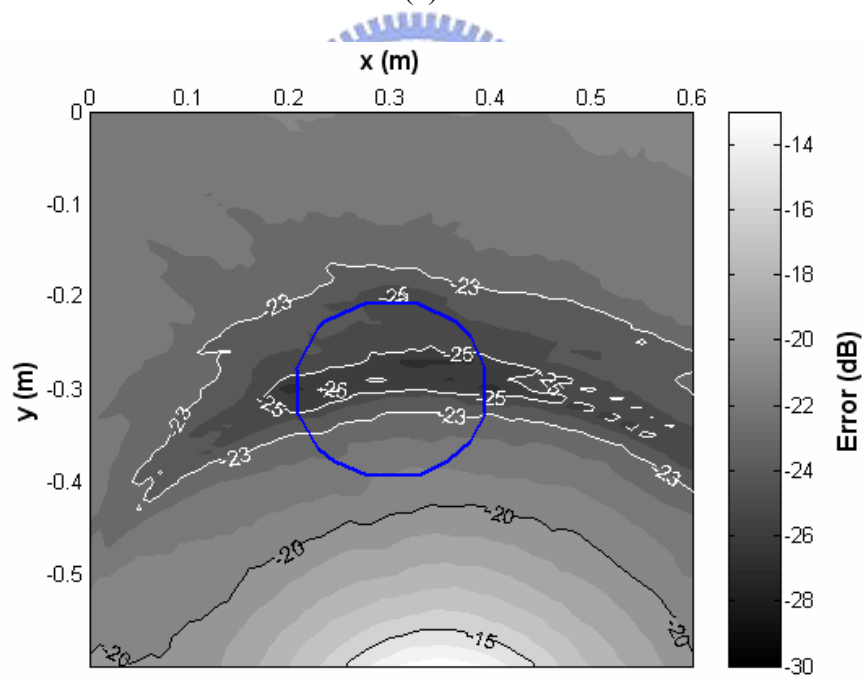
(b)

Fig. 36 The contour plots of the performance of Case II. (a) Front-right seat (b) Rear-right seat.





(a)



(b)

Fig. 37 The contour plots of the performance of Case III. (a) Front-right seat (b) Rear-right seat.

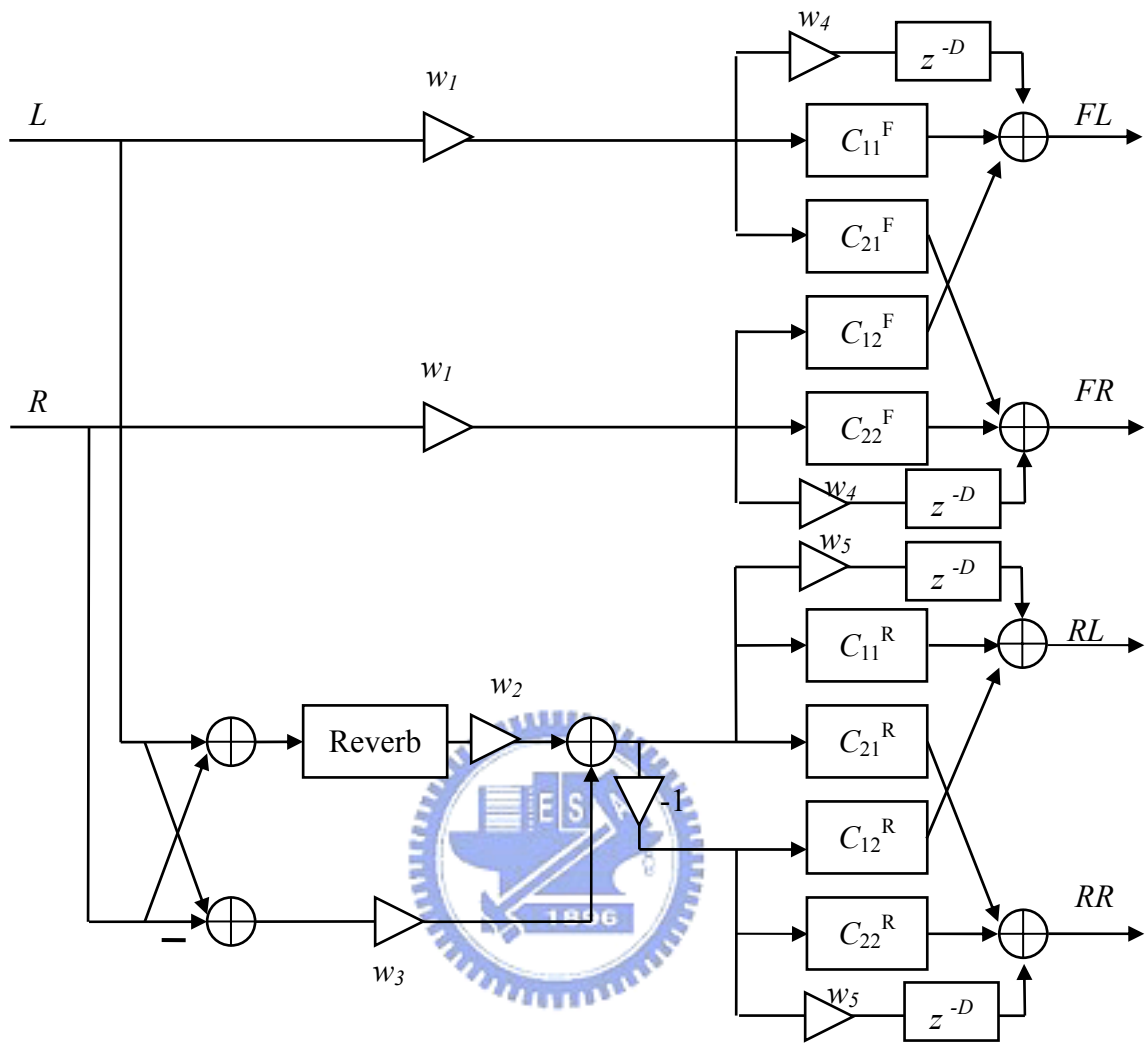
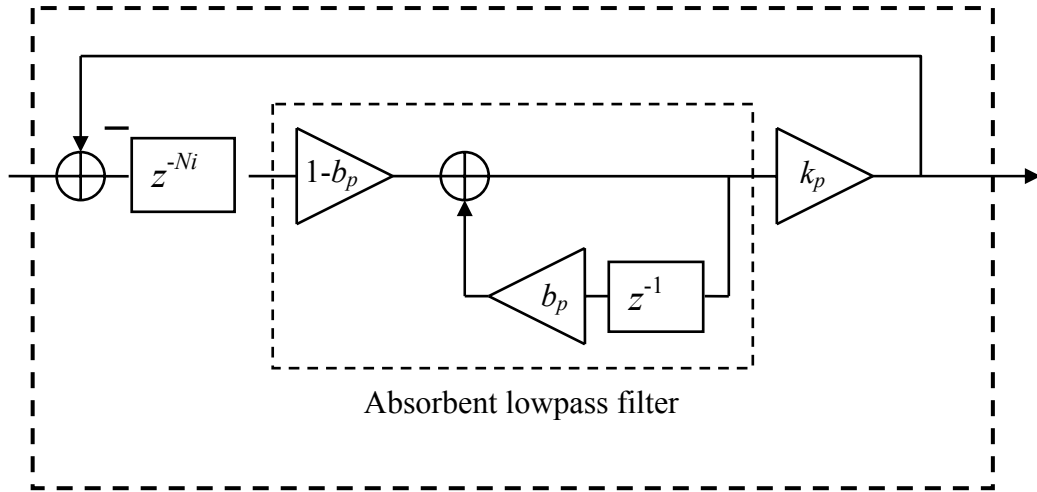
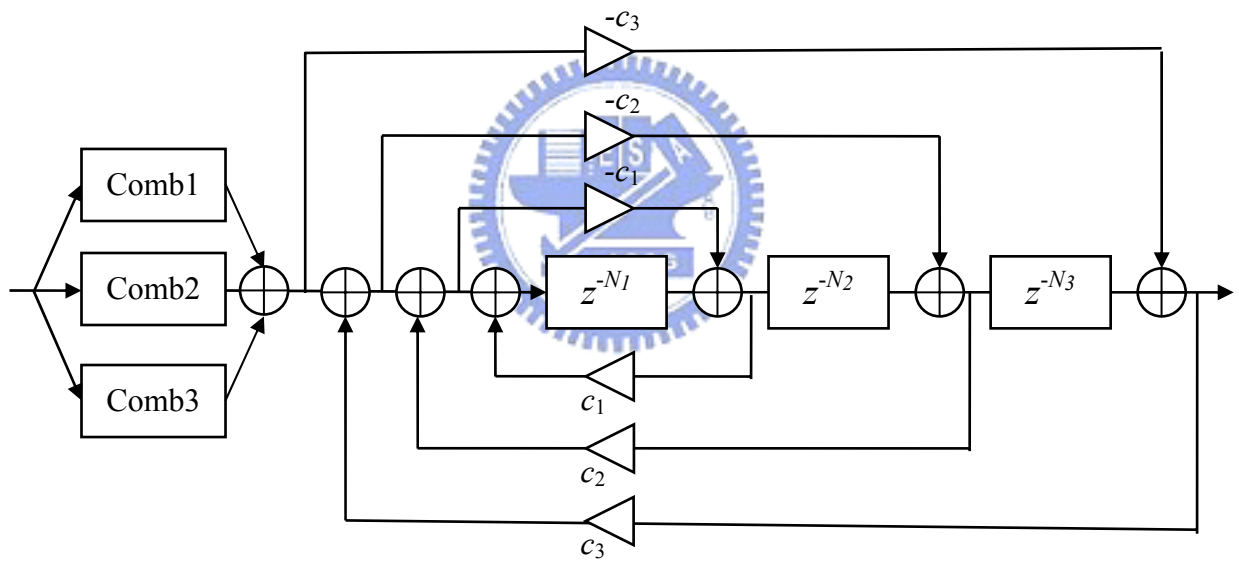


Fig. 38 The block diagram of Method I.



(a)



(b)

Fig. 39 The block diagrams of (a) comb filter and (b) overall reverberator.

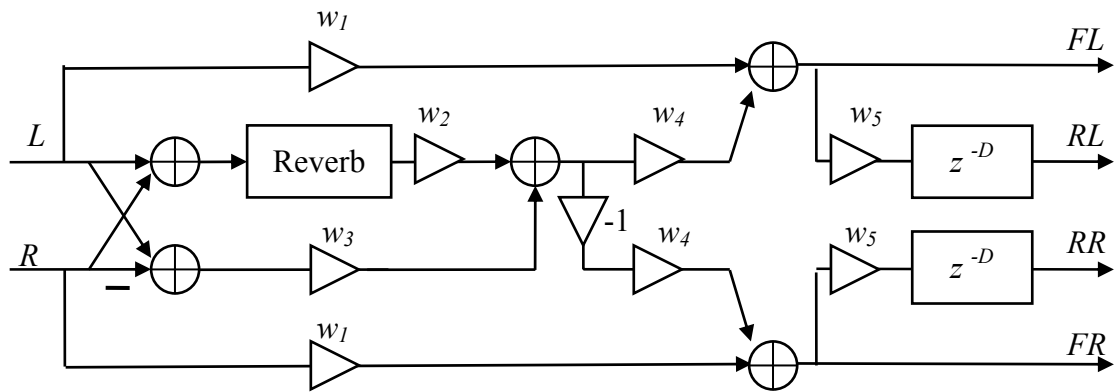


Fig. 40 The block diagram of Method II.



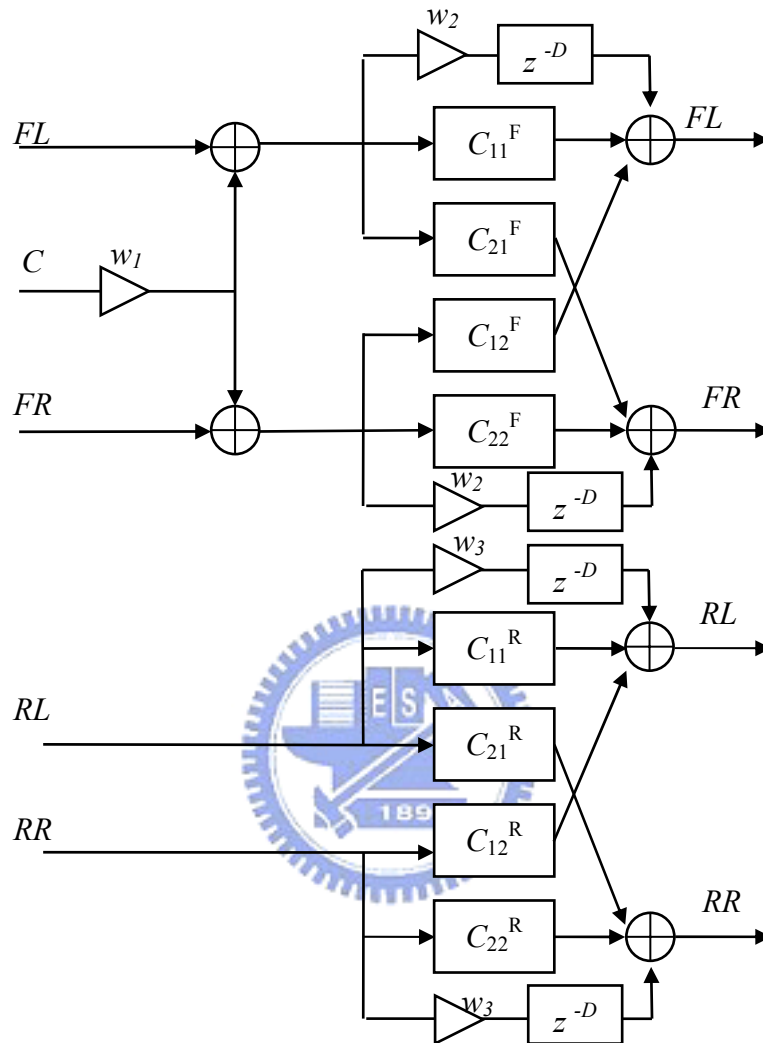


Fig. 41 The block diagram of Method III.

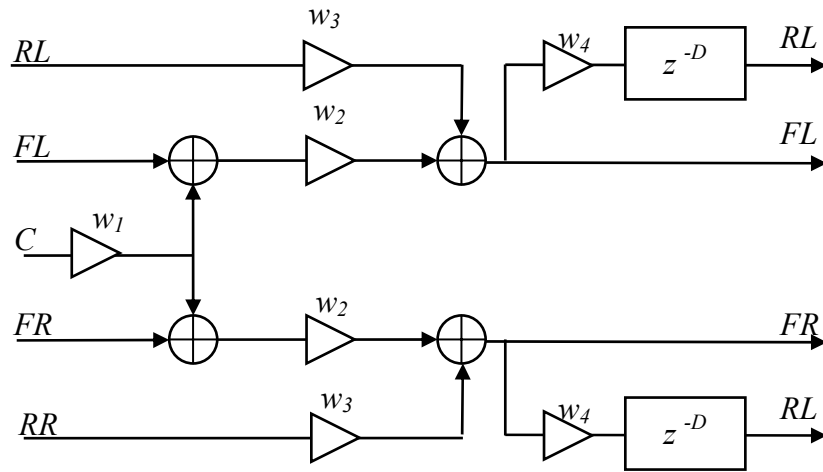
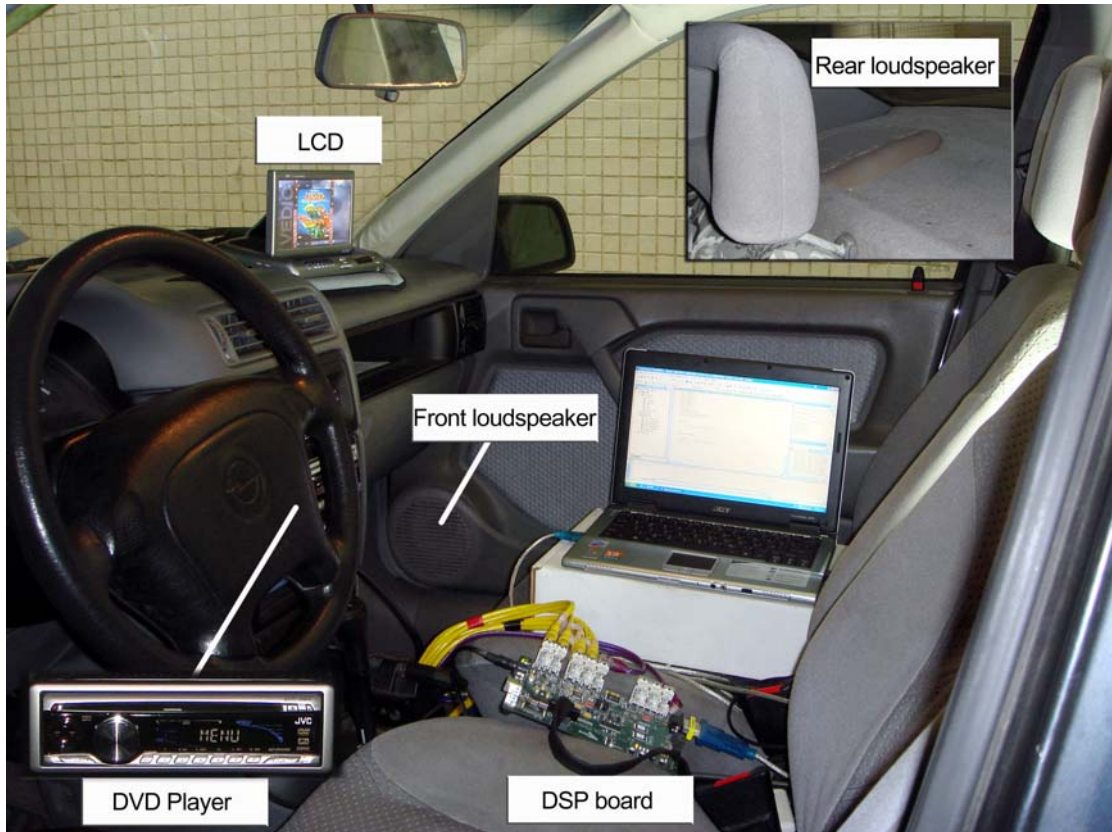


Fig. 42 The block diagram of Method IV.



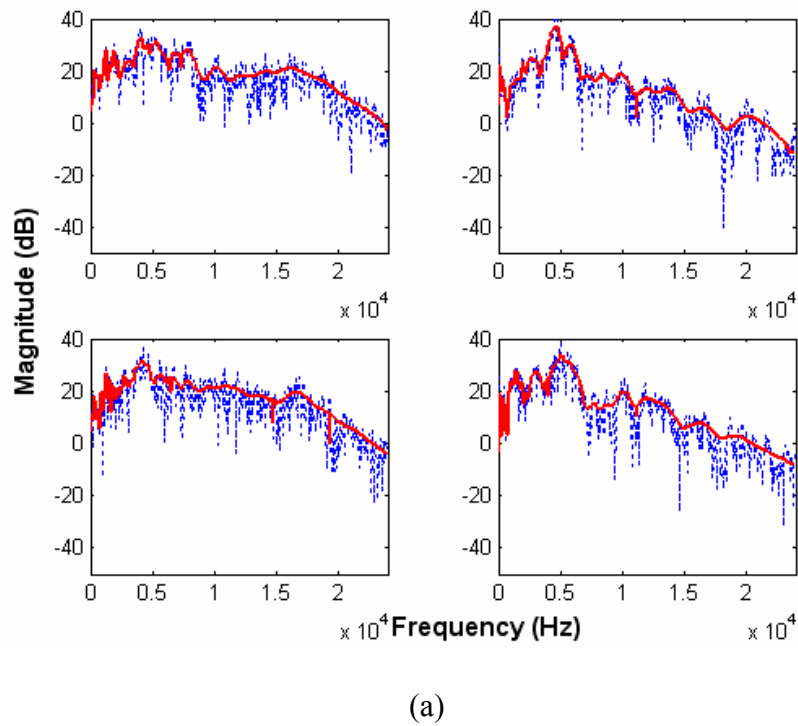


(a)

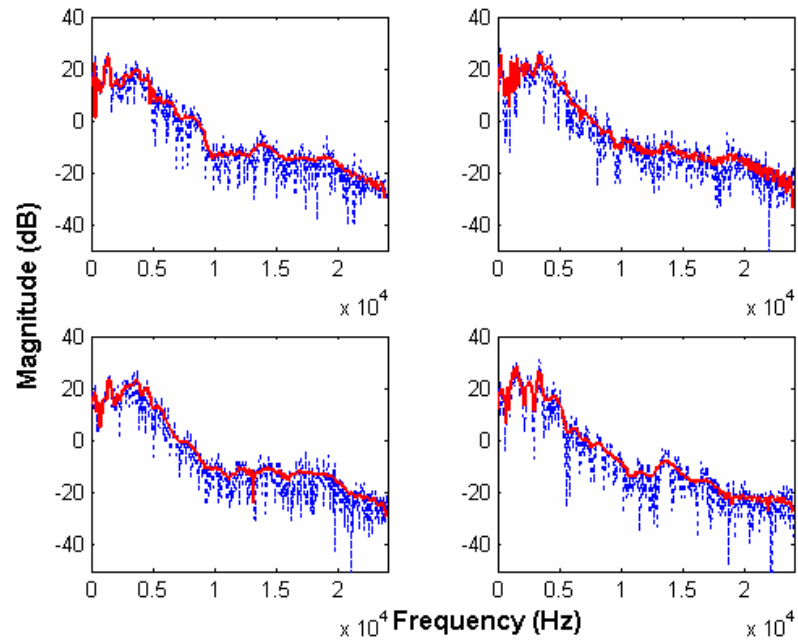


(b)

Fig. 43 (a) A 2-liter sedan. (b) The photo of the experimental arrangement.



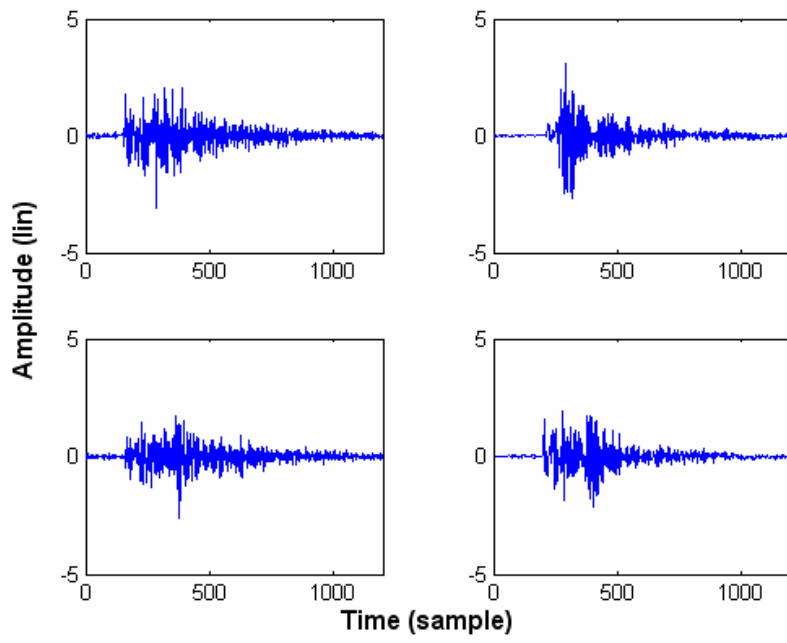
(a)



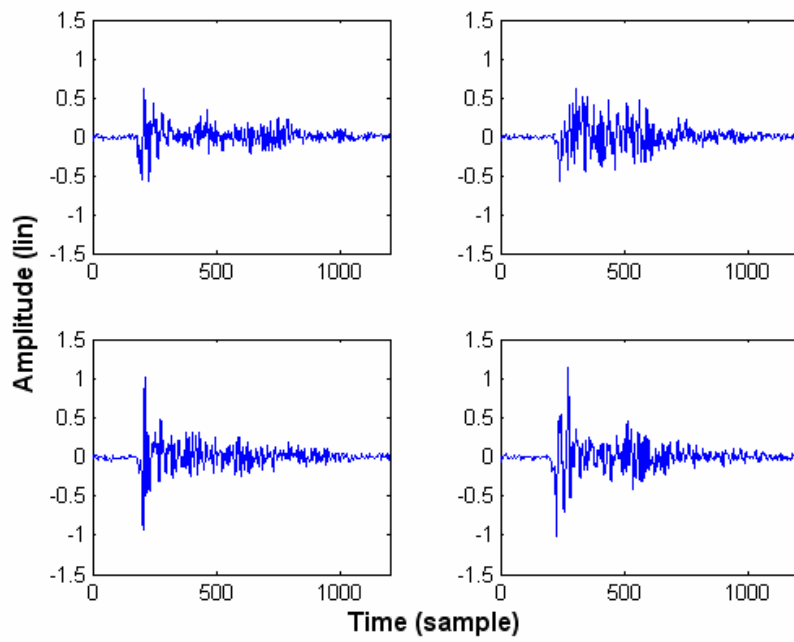
(b)

Fig. 44 The frequency responses of the plants at the front-left seat with (a) Front loudspeaker and (b) Rear loudspeaker. The dotted lines represent original measurement and solid lines represent smoothed measurement.

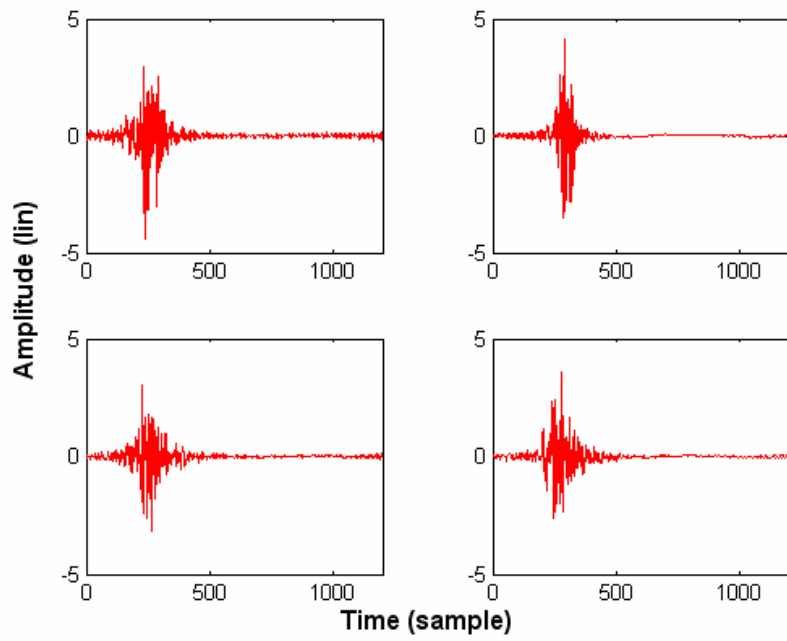




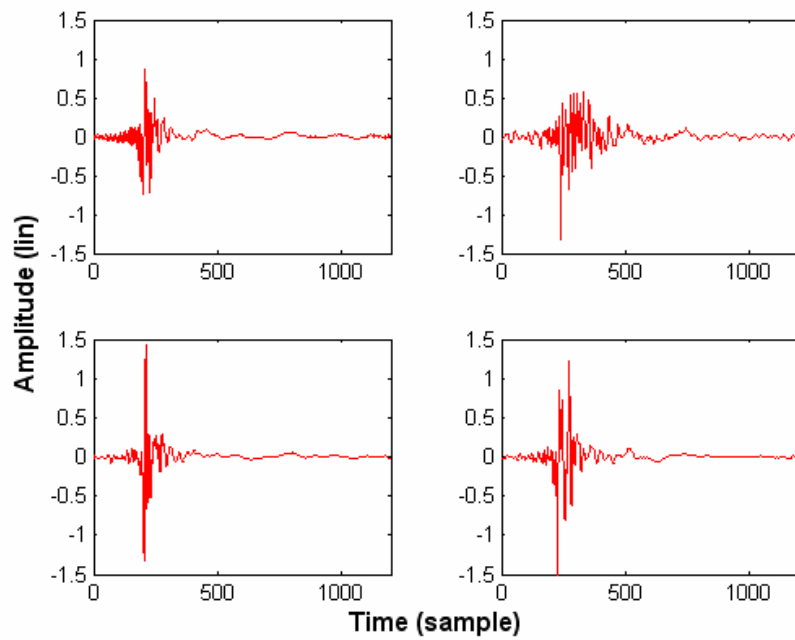
(a)



(b)

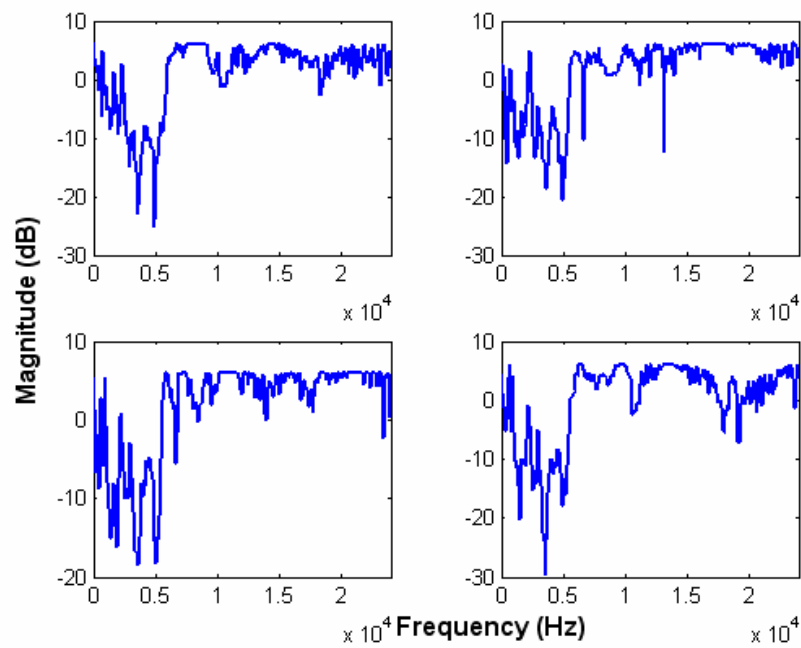
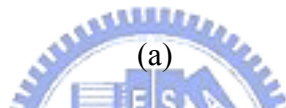
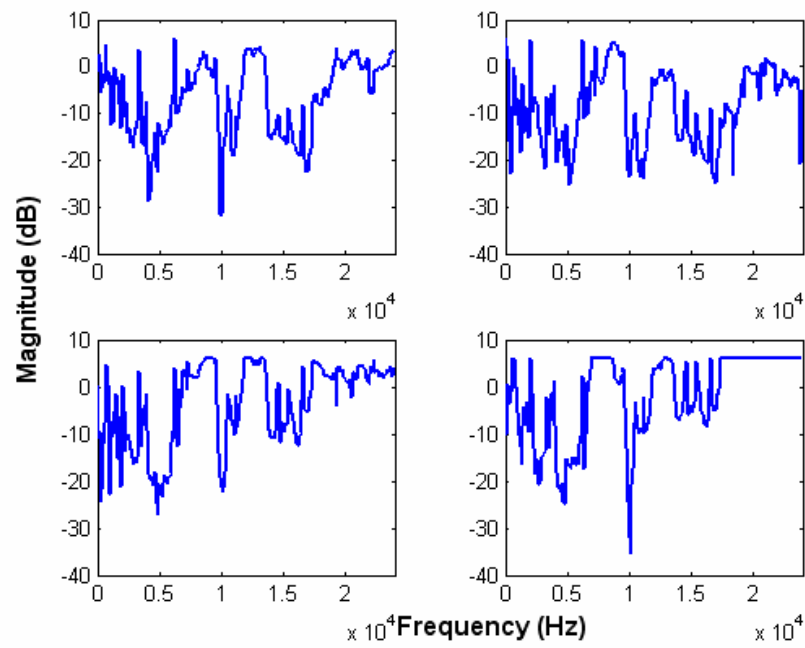


(c)



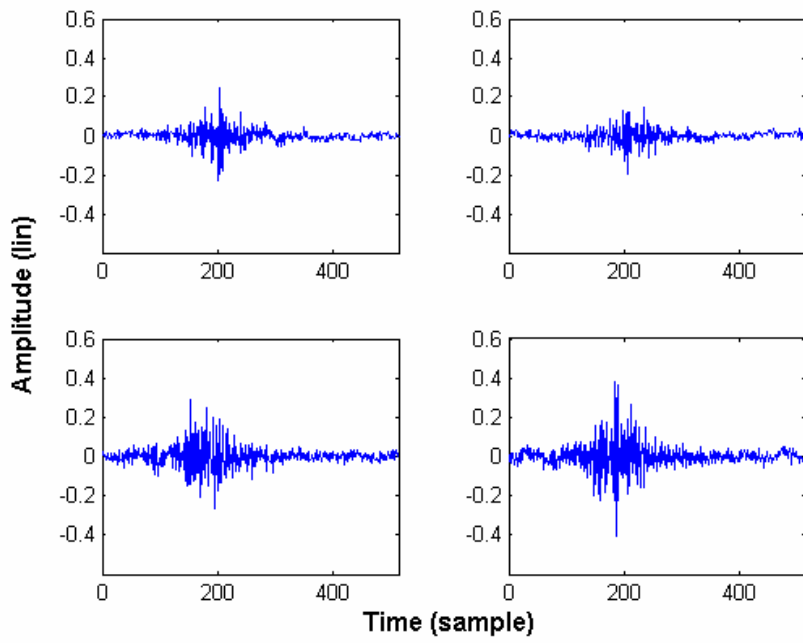
(d)

Fig. 45 The impulse responses of the plants at front-left seat. (a) Original measurement from front loudspeakers. (b) Original measurement from rear loudspeakers. (c) Smoothed measurement from front loudspeakers. (d) Smoothed measurement from rear loudspeakers.

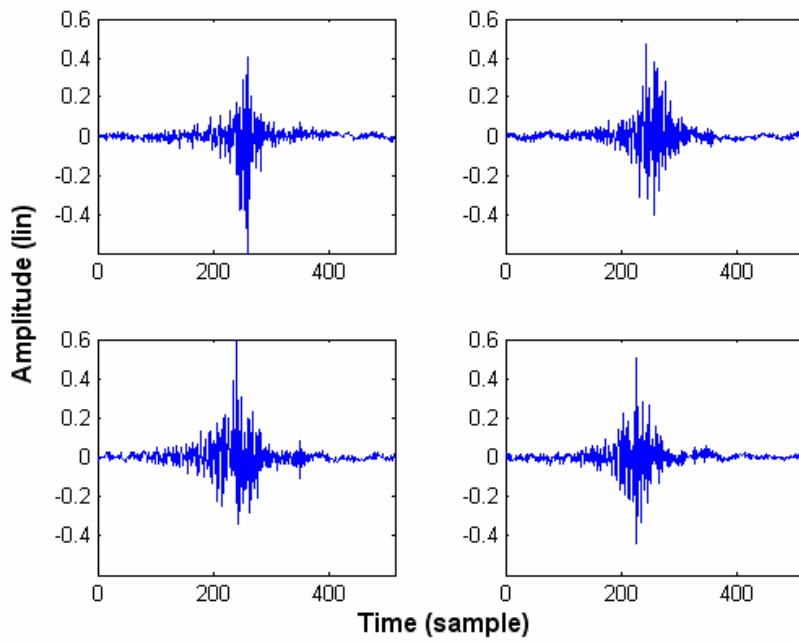


(b)

Fig. 46 Frequency responses of the inverse filters at front-left seat. (a) For front sound image. (b) For rear sound image.



(a)



(b)

Fig. 47 Impulse responses of the inverse filters at front-left seat. (a) For front sound image. (b) For rear sound image.

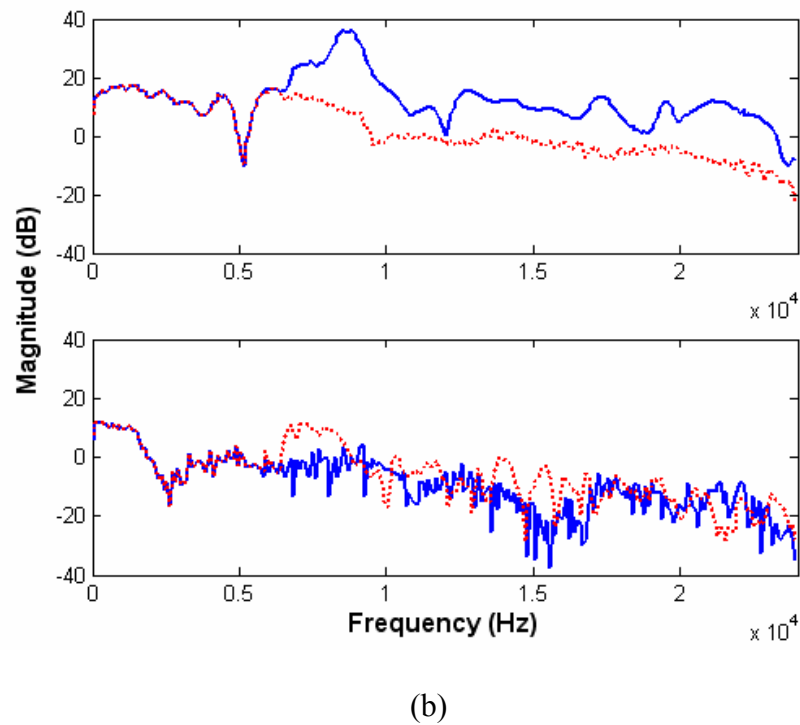
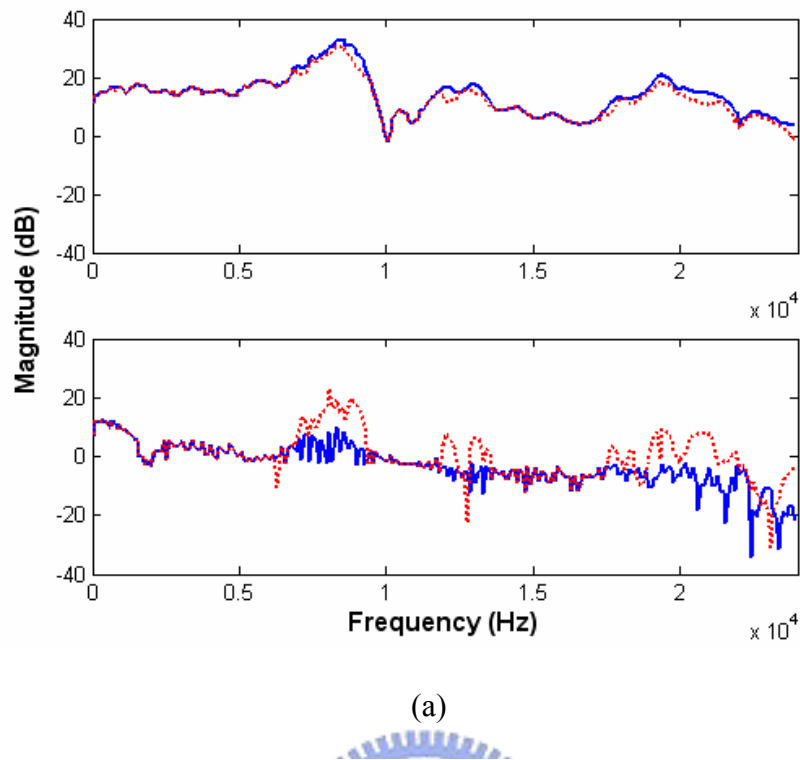
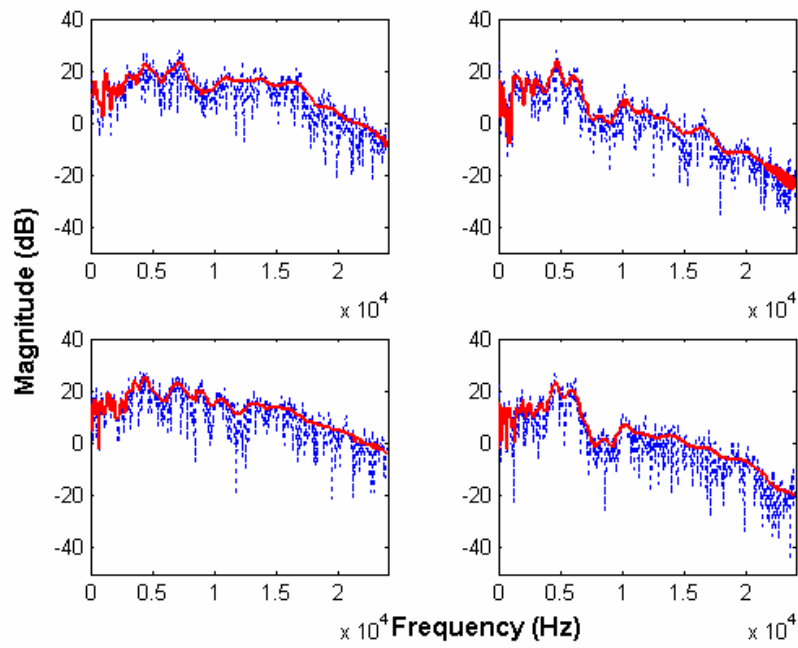
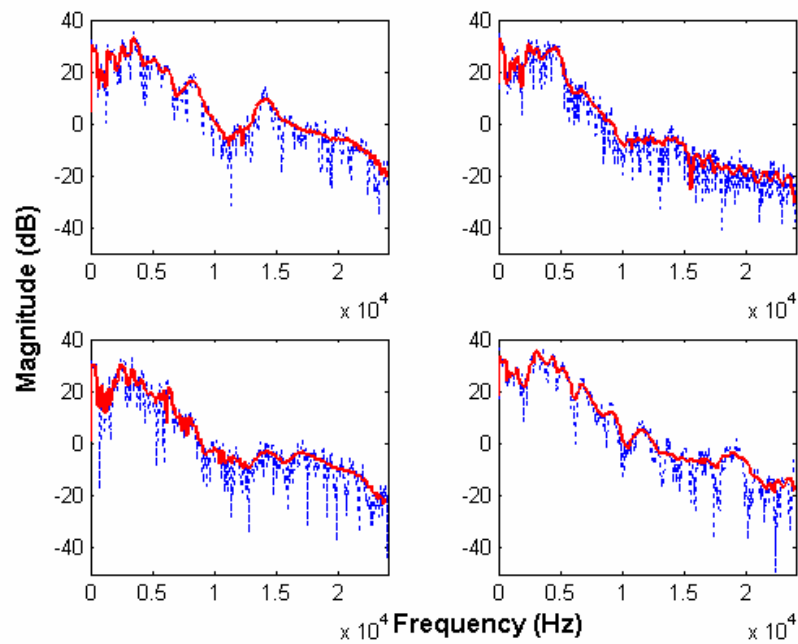


Fig. 48 The frequency responses of the virtual sound images. The solid lines represent desired responses  $\mathbf{M}$  and the dotted lines represent the multi-channel filter-plant product  $\mathbf{HC}$ . (a)  $\pm 30^\circ$  HRTF (b)  $\pm 110^\circ$  HRTF.

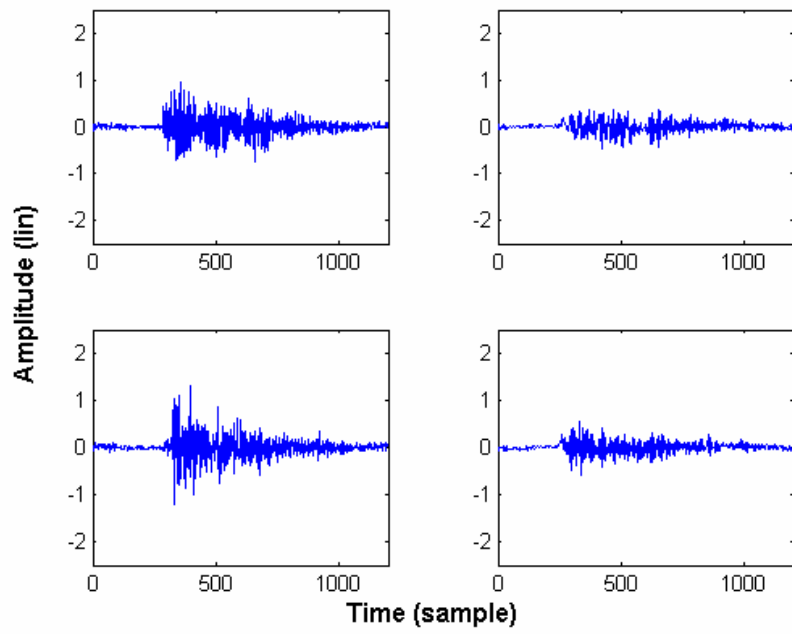


(a)

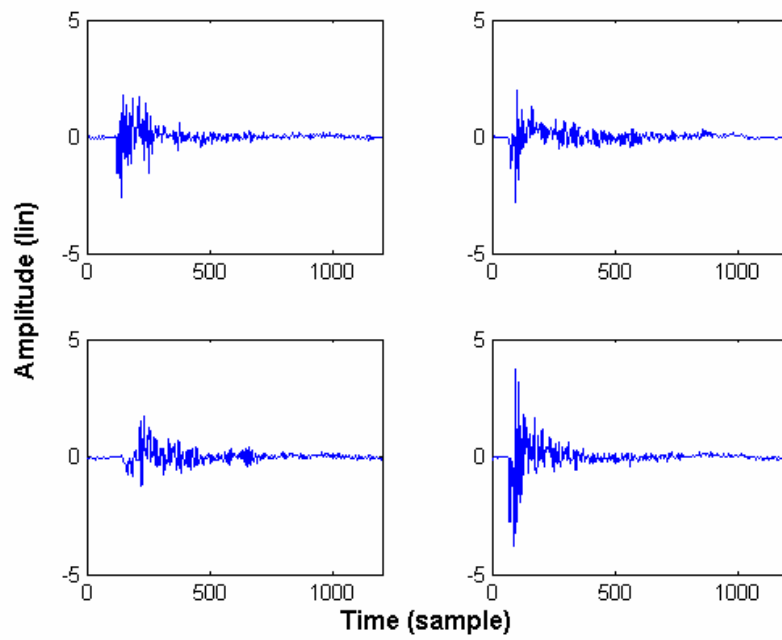


(b)

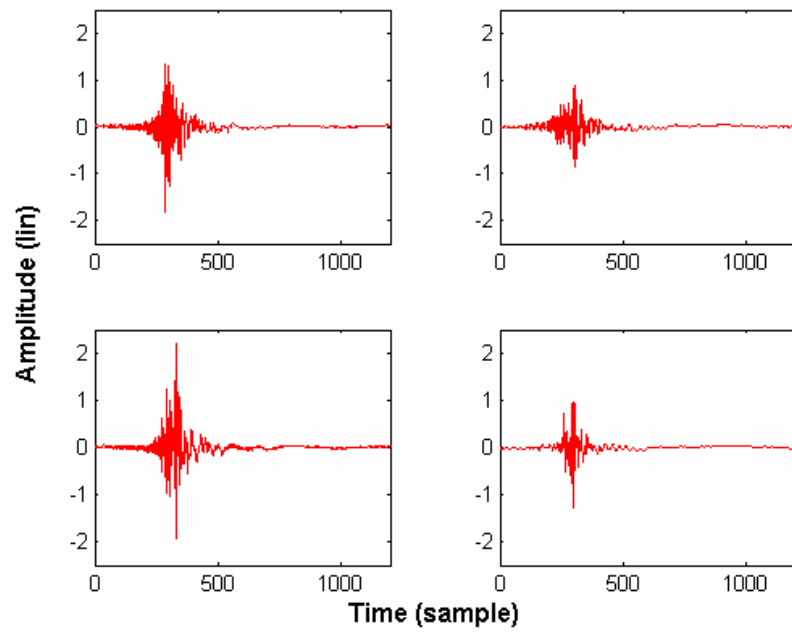
Fig. 49 The frequency responses of the plants at rear-right seat with (a) Front loudspeaker and (b) Rear loudspeaker. The dotted lines represent original measurement and solid lines represent smoothed measurement.



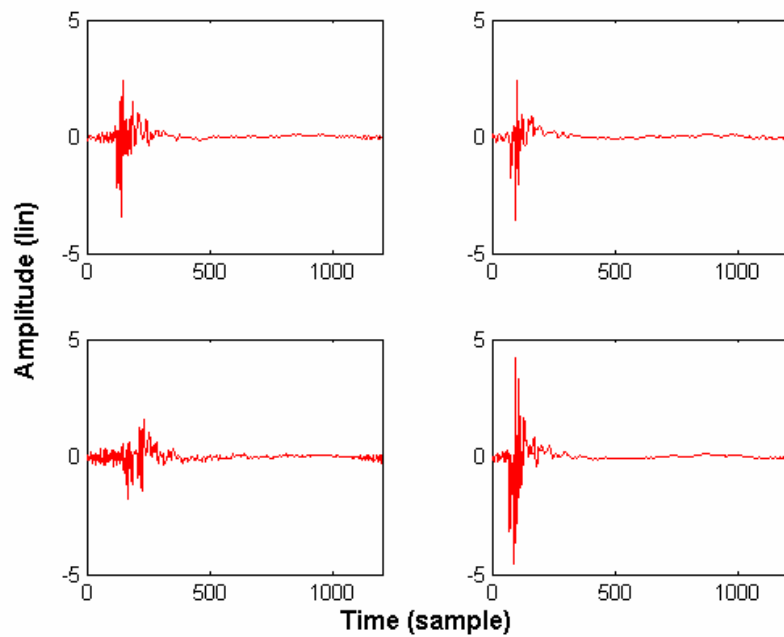
(a)



(b)



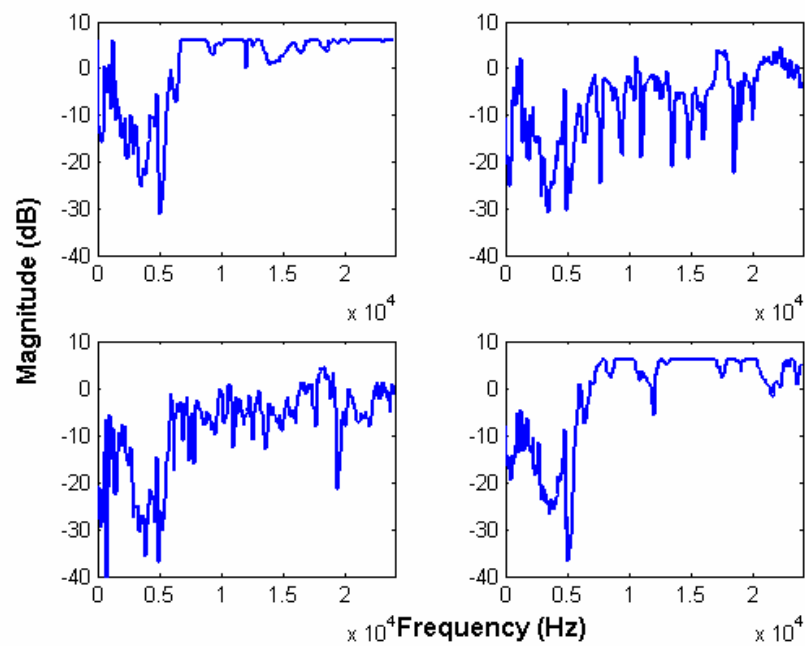
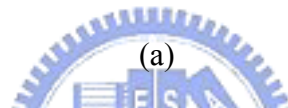
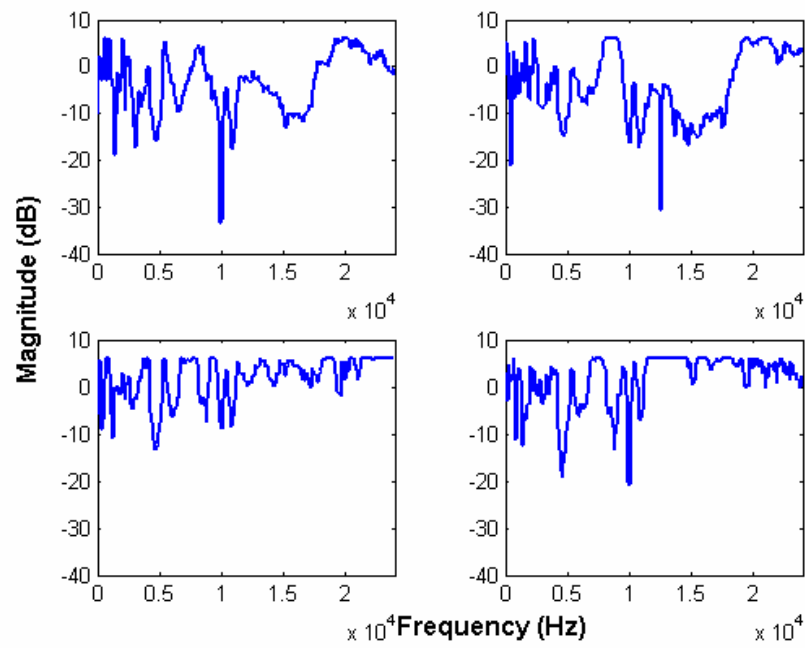
(c)



(d)

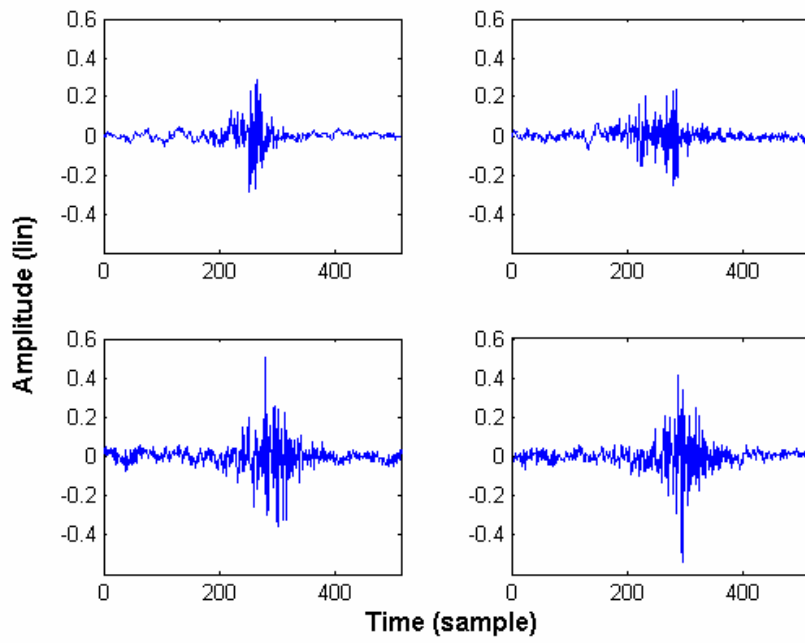
Fig. 50 The impulse responses of the plants at rear-right seat. (a) Original measurement from front loudspeakers. (b) Original measurement from rear loudspeakers. (c) Smoothed measurement from front loudspeakers. (d) Smoothed measurement from rear loudspeakers.



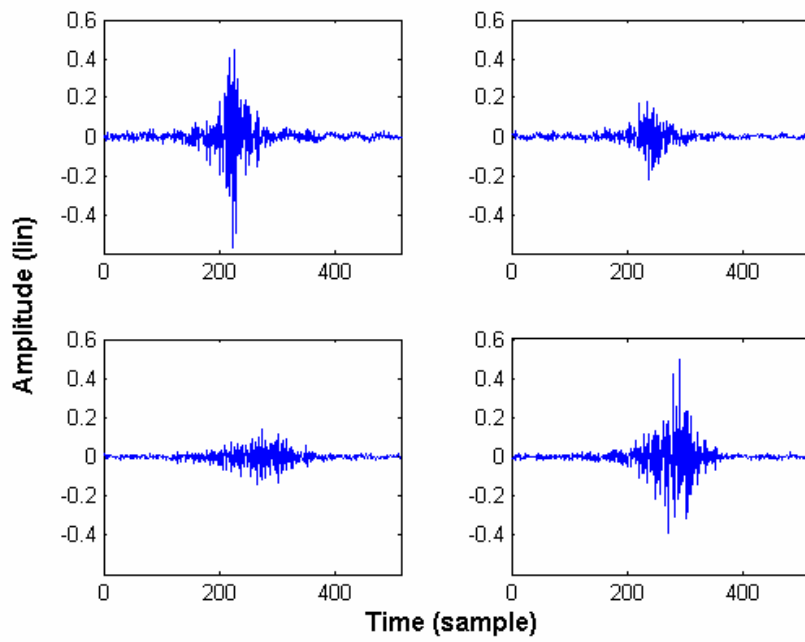


(b)

Fig. 51 Frequency responses of the inverse filters at rear-right seat. (a) For front sound image. (b) For rear sound image.



(a)



(b)

Fig. 52 Impulse responses of the inverse filters at rear-right seat. (a) For front sound image. (b) For rear sound image.

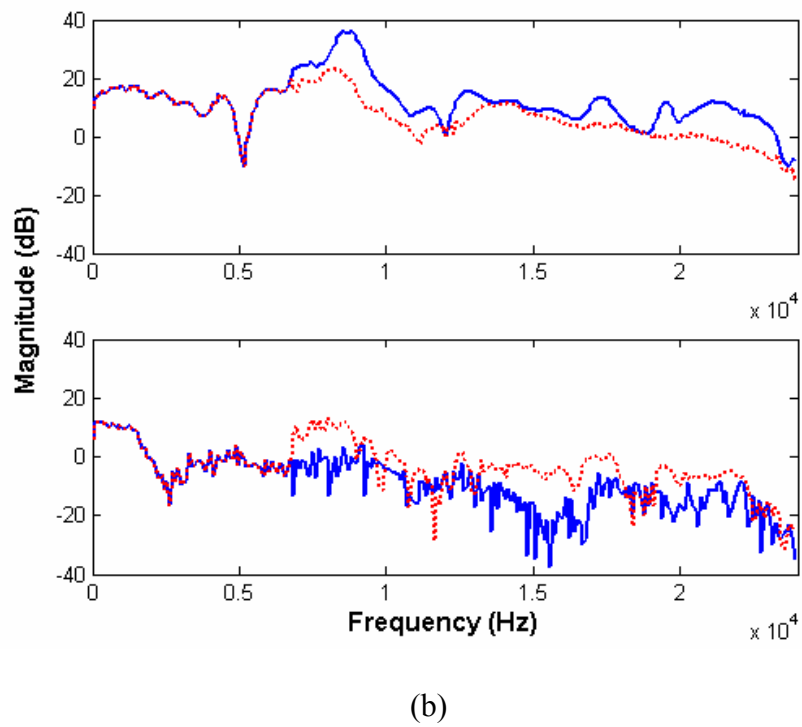
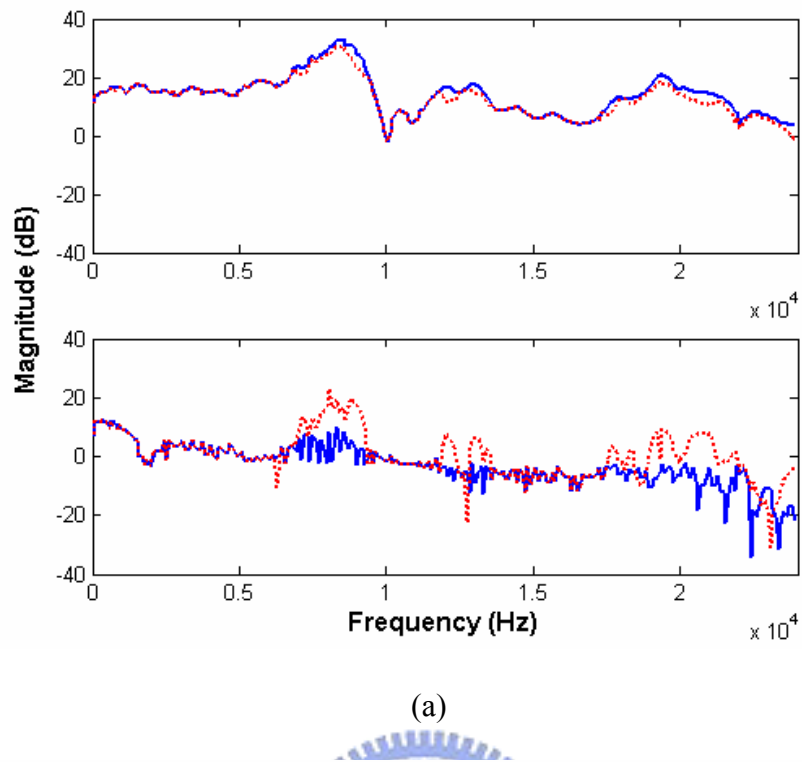
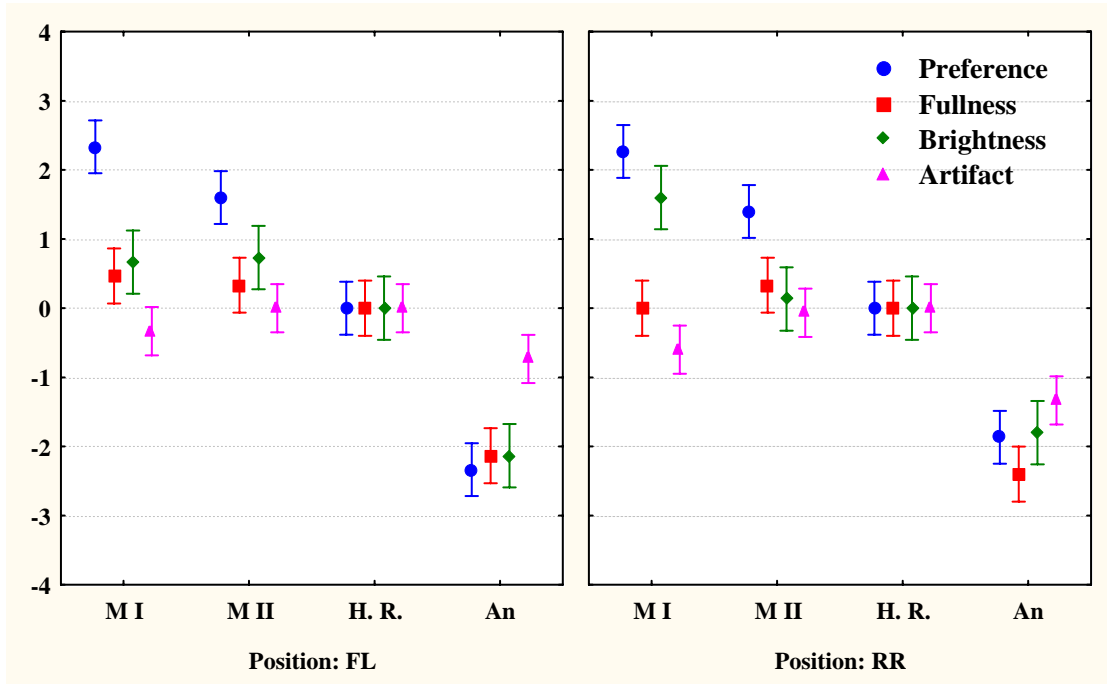
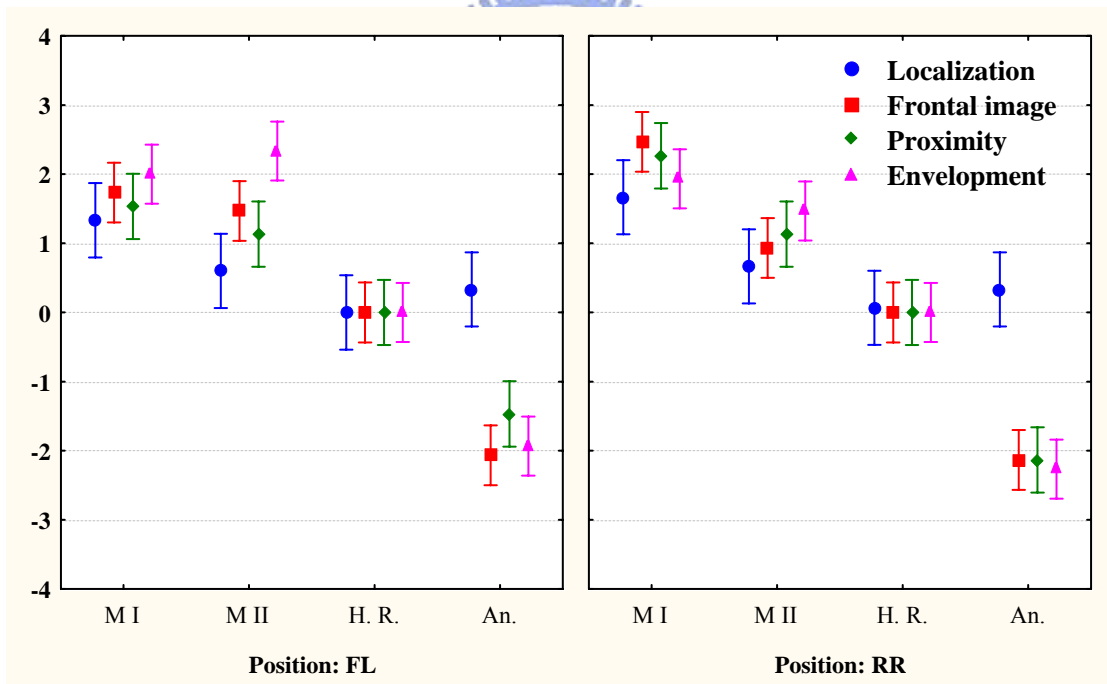


Fig. 53 The frequency responses of the virtual sound images. The solid lines represent desired responses  $\mathbf{M}$  and the dotted lines represent multi-channel filter-plant product  $\mathbf{HC}$ . (a)  $\pm 30^\circ$  HRTF (b)  $\pm 110^\circ$  HRTF.



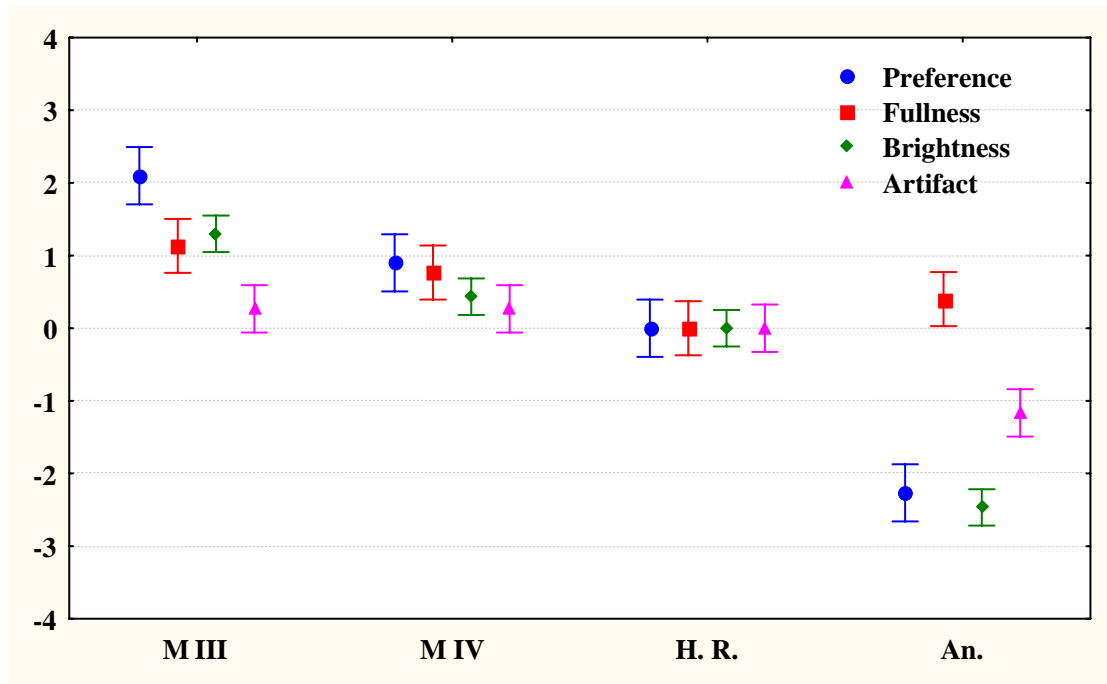
(a)



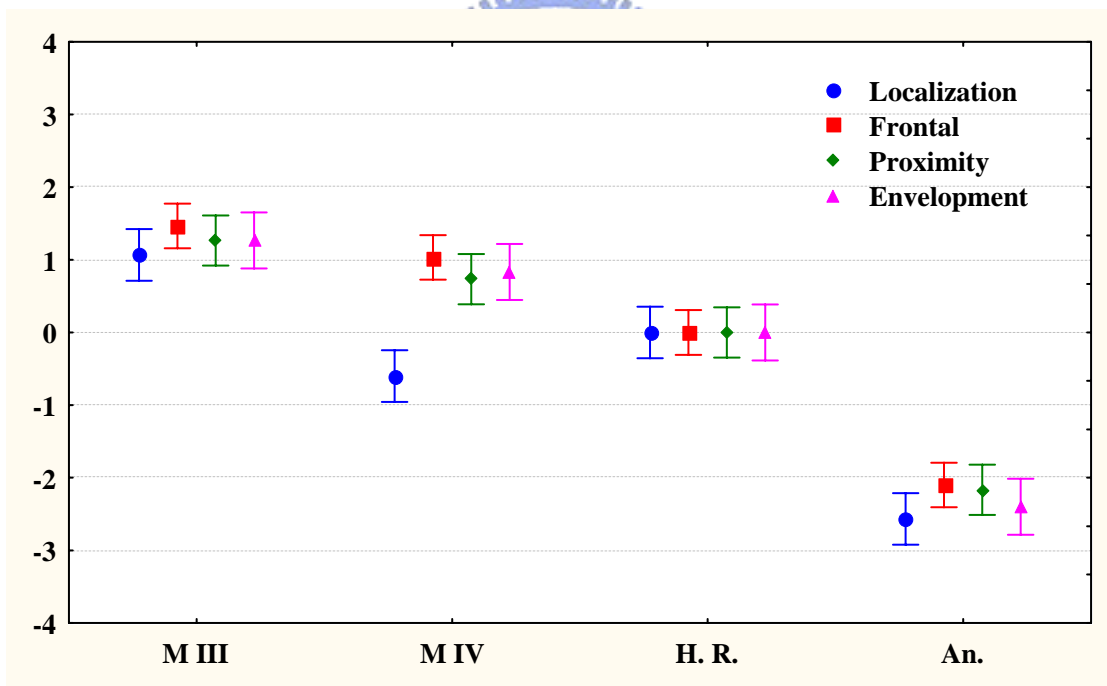
(b)

Fig. 54 The means and spreads (with 95% confidence intervals) of the grades in Exp.

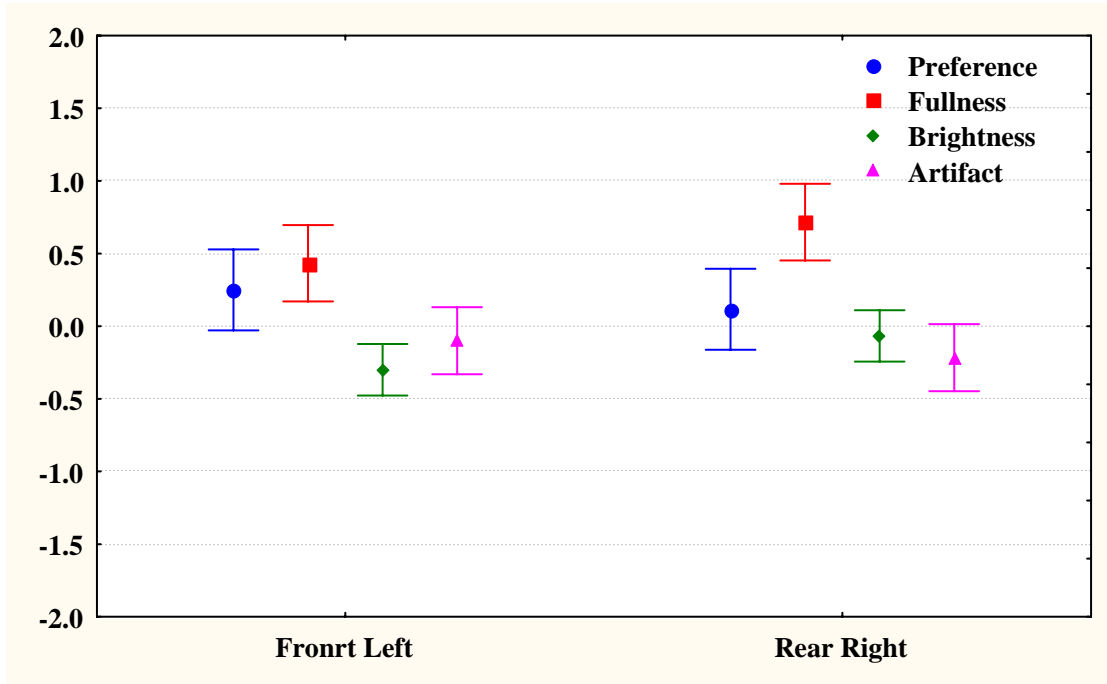
I. (a) The first four attributes (b) The last four attributes.



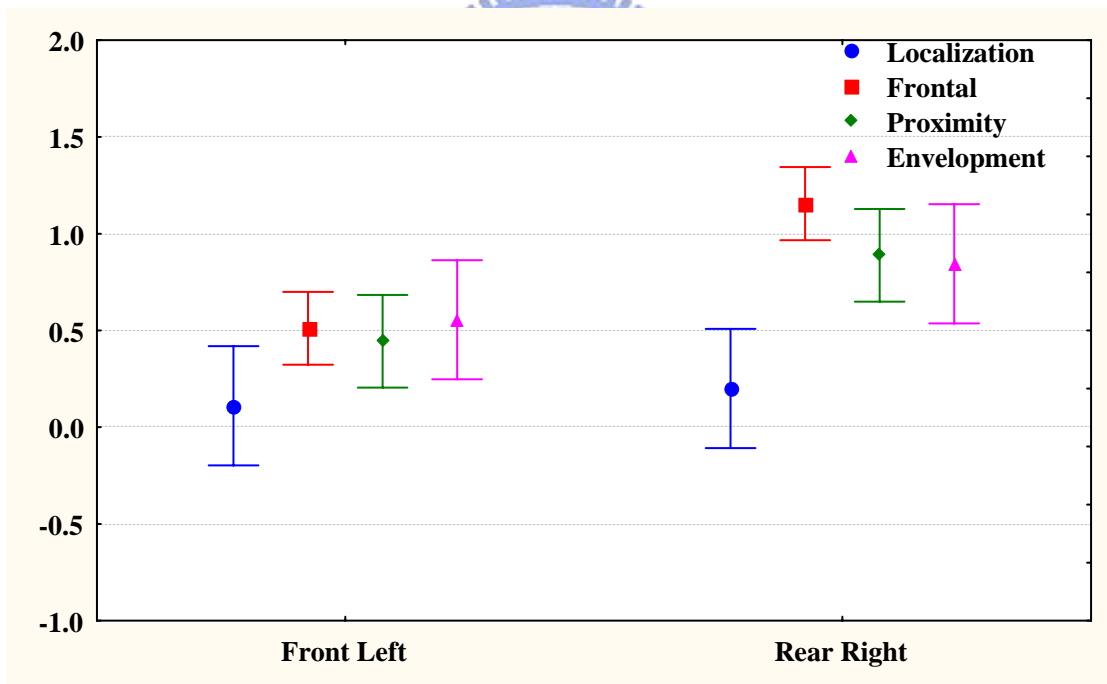
(a)



(b)

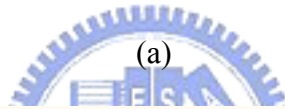
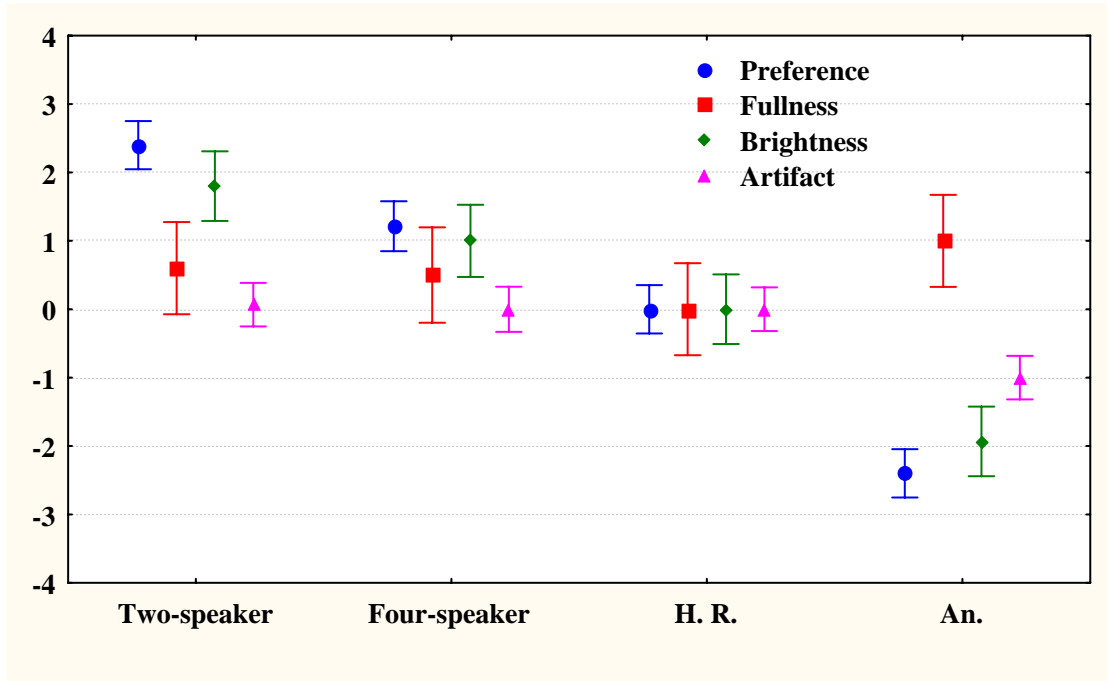


(c)

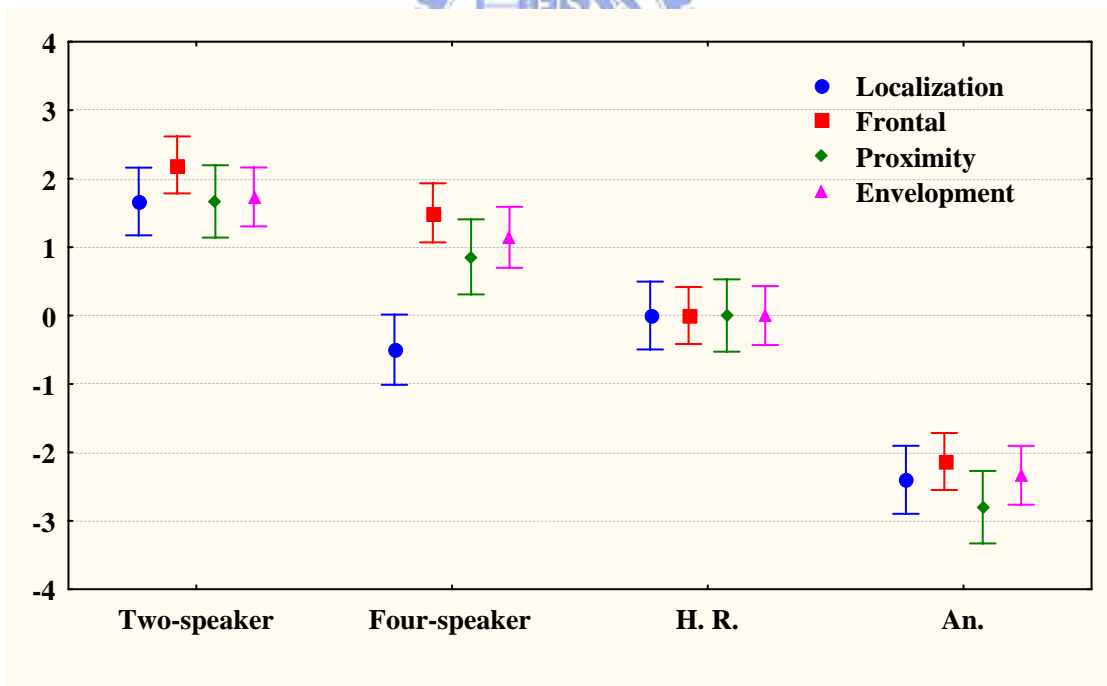


(d)

Fig. 55 The means and spreads (with 95% confidence intervals) of the grades in Exp. II. (a) The first four attributes for the methods (b) The last four attributes for the methods. (c) The first four attributes for the positions (d) The last four attributes for the positions.

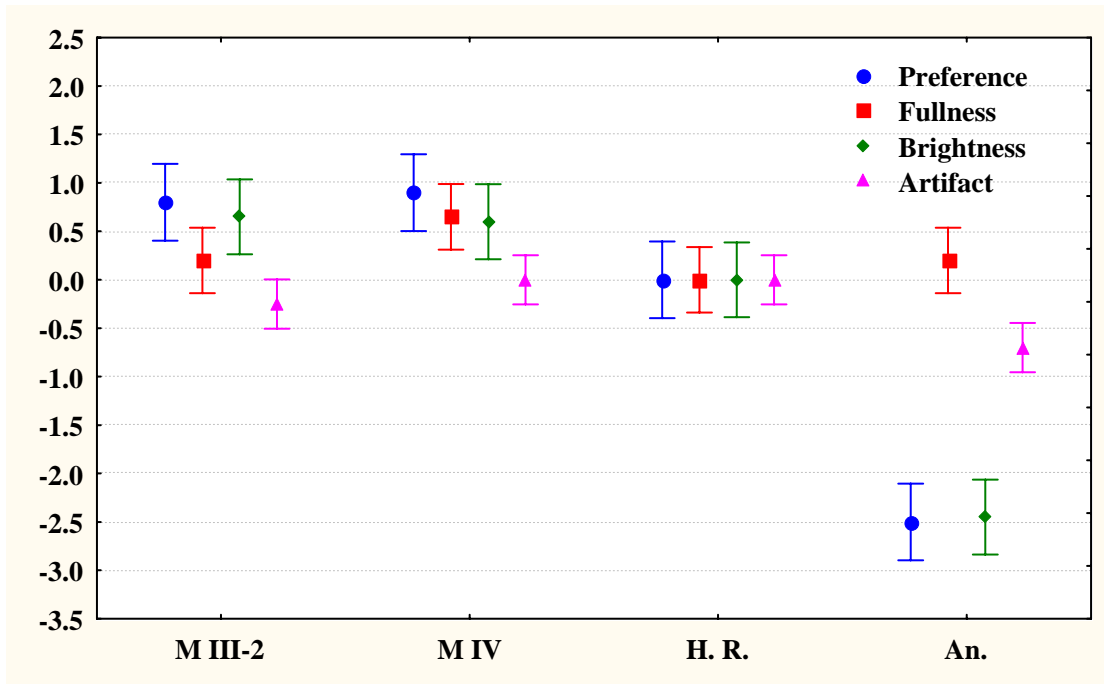


(a)

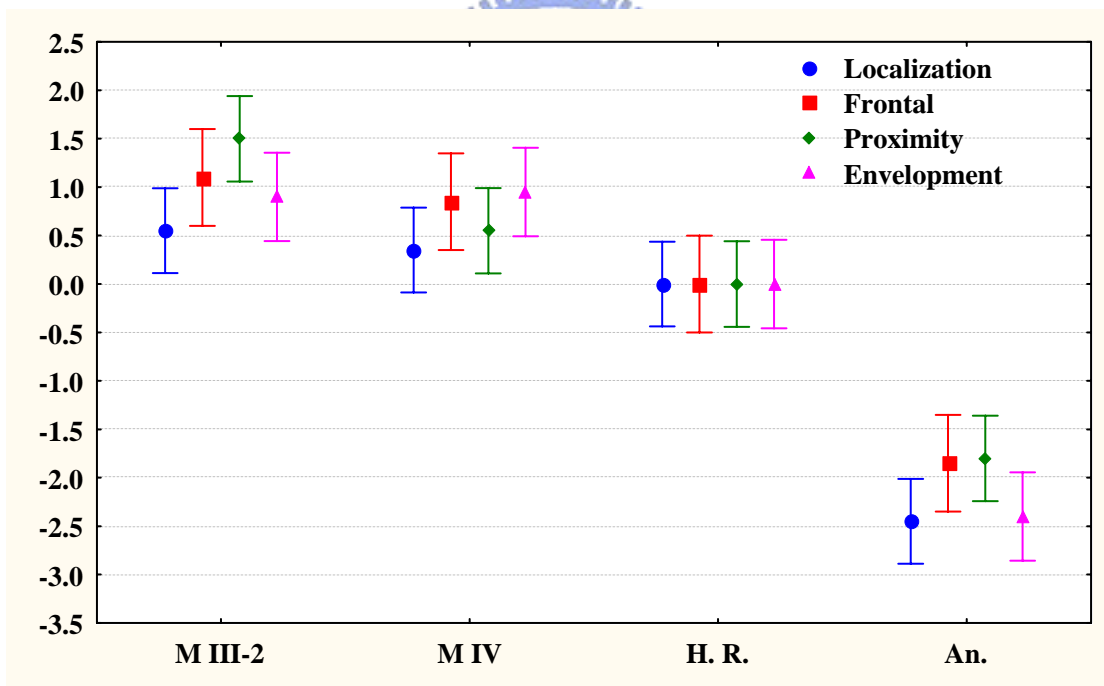


(b)

Fig. 56 The means and spreads (with 95% confidence intervals) of the grades in Exp. III. (a) The first four attributes (b) The last four attributes.



(a)



(b)

Fig. 57 The means and spreads (with 95% confidence intervals) of the grades in Exp.

IV. (a) The first four attributes (b) The last four attributes.