國 立 交 通 大 學

機械工程學系

碩士論文

適應性 3D 音效投影屏幕與藍芽耳機聲學最佳化設計

Adaptive 3D-sound projection screen and optimal acoustical design for Bluetooth earphones

研 究 生：郭育志

指導教授：白明憲

中華民國九十八年八月

適應性 3D 音效投影屏幕與藍芽耳機聲學最佳化設計

Adaptive 3D-sound projection screen and optimal acoustical design for Bluetooth earphones

研 究 生：郭育志　　Student：Yu Chih Kuo

指導教授：白明憲　　Advisor：Mingsian Bai

A thesis
Submitted to Department of Mechanical Engineering
Collage of Engineering
National Chiao Tung University
In Partial Fulfillment of Requirements
for the Degree of Master of Science
in
Mechanical Engineering
August 2009
HsinChu, Taiwan, Republic of China

中華民國九十八年八月

適應性 3D 音效投影屏幕與藍芽耳機聲學最佳化設計

研究生：郭育志　　指導教授：白明憲 教授

國立交通大學 機械工程學系 碩士班

摘　　要

面對耳機設計時各種相互衝突的特性需求，如靈敏度、失真率、頻寬以及微小化等，需要一種有系統且有效率的方法來達成我們的設計並符合設計需求。利用微型揚聲器的參數鑒別法、電機聲模擬電路 (EMA)，建立藍芽耳機的聲學模擬平台。並且以此平台為基礎，利用模擬退火法 (simulated annealing)進行耳機腔體的最佳化設計，並能有效達到期望目標。實驗方面，利用 Type 2 人工耳（IEC 711 ear simulator）來得到實驗結果，並且顯示出經最佳化後的耳機腔體設計，能夠大大的提升其聲學品質，並且能夠符合 3GPP2 的規範。

一般而言，利用串音消除系統（cross talk cancellation system）為基礎的 3D 虛擬環繞音效，皆會產生甜點（sweet spot）的問題，也就是當聽者超出甜點的範圍，就無法聽到良好的音效。本文中提出一個適應性 3D 虛擬聲束操控（beam steering）系統，可使聽者利用雙聲道喇叭，即可在任意聆聽位置感受到絕佳的環繞音效。本系統包含了兩種技術，其中之一是利用攝影機做辨識的頭部追蹤系統，另外是聲束控制演算法，來進行聆聽位置的最佳化設定。利用適應性 3D 虛擬聲束操控系統，一個全新的平面喇叭投影屏幕誕生，並且可以使聆聽者在任何位置享受最佳的 3D 環繞音效。

Adaptive 3D-sound projection screen and optimal acoustical design for Bluetooth earphones

Student: Yu Chih Kuo Advisor: Mingsian R. Bai

Department of Mechanical Engineering

National Chiao-Tung University

ABSTRACT

Faced with various conflicting issues arising from sensitivity, distortion, bandwidth and miniaturization requirements for Bluetooth® earphones, it is desirable to develop a systematic way to attain the design that would meet these requirements. Based on a lumped parameter model, the design of the earphone enclosure is optimized using the simulated annealing (SA) algorithm. Experimental results obtained using a type 2 artificial ear (IEC 711 ear simulator) reveals that the optimized design has resulted in significant enhancement of performance, complying with the frequency response mask dictated in the 3GPP2 standard.

This thesis proposes an adaptive three-dimensional (3D)-sound projection screen aimed at home theater applications. The screen serves both audio and visual rendering system that integrates numerous technologies including panel speakers, cross talk cancellation system (CCS) and video head tracking. In particular, the CCS enhances the localization impression when rendering 3D spatial sound with two loudspeakers. A camera is employed to adaptively track the head movement such that the robustness of CCS is ensured. The algorithms were implemented on a laptop computer in tandem with a digital signal processor, which handled the video and audio signal processing, respectively. Experimental results reveal that the proposed system is capable of rendering immersive spatial sound with robustness against listener's head movement.

致謝

時光飛逝，短短兩年的研究生生涯轉眼就過去了。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。

在論文寫作上，感謝本系李安謙教授和呂宗熙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。

回顧這兩年的日子，承蒙同實驗室的陳榮亮學長、林家鴻、劉青育學長、黃兆民學長、洪志仁學長以及謝秉儒學長在研究與學業上的適時指點，並有幸與王俊仁、何克男、艾學安及劉冠良同學互相切磋討論，每在烏雲蔽空時，得以撥雲見日，獲益甚多。此外學弟廖士涵、廖國志、張濬閣、桂振益、曾智文、陳俊宏及劉嬰婷在生活上的朝夕相處與砥礪磨練，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，獻給我摯愛的雙親郭聰明先生、張麗容女士、姊姊郭育妏。今天我能順利取得碩士學位，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

Table List

Figure List

1. Introduction

In recent years, electro-acoustical transducers have become key components to many 4C (Computer, Communication, Consumer electronics and Car) products. People who use 4C products do not just wish it can see and hear any more. People want to enjoy their lives. Therefore it is a trend for a better 4C products. And sound quality is an important part. This essay introduces the design of TV audio system, earphone and panel speaker projection screen. Measure and simulate electro-acoustical transducer by using EMA analogous circuits. Improve the sound quality by optimization. Then create 3D virtual surround by using CCS with DSP.

The microspeaker discussed in this thesis are primarily dynamic moving-coil type. The electroacoustic model of dynamic microspeaker involves electrical, mechanical, and acoustical domains. At the low-frequency regime, a loudspeaker can be modeled with electro-mechano-acoustical (EMA) analogous circuits and lumped parameters [1]-[4]. For dynamic loudspeakers, vented-box design has traditionally been used as a means for low-frequency enhancement. Thiele [5], [6] and Small [7]-[10] have laid the theoretical foundation for vented-box design in a series of classical papers. Bai and Liao [11] applied the vented-box idea for designing acoustical enclosures of miniature loudspeakers for mobile phones.

Bluetooth® wireless earphones have become an important accessory for reproducing speech in hands-free communication and consumer electronics. Faced with various conflicting issues arising from sensitivity, distortion, bandwidth and miniaturization requirements, it is desirable to develop a systematic and efficient way to attain the design that would meet these requirements.

In this paper, EMA analogous circuit is used to model a Bluetooth earphone. Thiele and Small (T-S) parameters of the microspeaker embedded in the earphone are experimentally identified. Thus, for the microspeaker and the associated enclosure

and casing, a lumped parameter model can be established to predict the frequency response. On the other hand, acoustical impedance of ear canal that is distinct from a free-field environment is also incorporated into the model.

As mentioned previously, it is vital to optimize the enclosure design for a given driver such that the requirements are met. To this end, the Simulated Annealing (SA) [12] is employed to search for the optimal combination of enclosure parameters. SA is a random-search technique which exploits an analogy between the way in which a metal cools and freezes into a minimum energy crystalline structure (the annealing process) and the search for a minimum in a more general system. Using the forgoing lumped parameter model, the sound pressure response at the ear drum position can be simulated for the SA optimization. The cost function alongside the constraints posed by practical application and implementation are formulated in accordance with the frequency response mask dictated in the standard of 3GPP2 C.S0056-0 [13]. A mockup is made based on the optimal design. Experimental results obtained using a type 2 artificial ear [14] are discussed in the conclusions.

In recent years, home theater systems have received a great deal of attention in that people increasingly prefer to experience audiovisual presentation at home rather than at a theater. Generally, a home theater provides three-dimensional (3D) sound by multi-cannel loudspeakers which need a lot of cost and room space. In the reason, the adaptive 3D-sound projection screen for home theater based on the two-channel panel speakers and 3D sound processing is developed. This audiovisual system can offer large image and impressive virtual sound for the viewers.

In three-dimensional (3D) audio reproduction, the crosstalk cancellation system (CCS) is the core technology. The CCS based on inverse filtering are employed to minimize the effects due to crosstalk that can obscure sound image. Deconvolution approach, the frequency-domain method [15] can be utilized to design the required

inverse filters. Conventional inverse filtering leads to reduced crosstalk and equalized ipsilateral response. However, if the CCS is inadequately designed, the latter effect can result in audible high-frequency artifacts. To address the problem, the direct filtering method with the band-limited design is used [16]. The band-limited design limits the crosstalk cancellation to function only within 20-6k Hz. However, the 3D virtual sound system relying on a set of CCS inherently has a sweet spot. The listener can sense the virtual sound effect well only in the area of the sweet spot. Otherwise the listener cannot hear the virtual sound effect well [17]. In order to solve this problem, Kim et al. [18] proposed the smart virtual sound system with a listener position tracking system with infrared and ultrasonic sensors.

In this thesis, the adaptive 3D-sound projection screen consists of the panel speakers, the head tracker with a camera and the CCS using interpolation method is proposed. The head tracking system based on the detection framework [19] that is capable of processing images extremely rapidly while achieving high detection rates. As to the listening positions, most subjects reported that the front back movement of the listener did not cause as much performance degradation as the lateral movement of the listener. In the reason, the inverse filters has been created only considers the listener positions of the lateral movements. For example, in order to cover the listener position in the range from - 2 m to + 2 m in resolutions 20 cm respect to the center of the projection screen, the number of the listening positions is only 20. Furthermore, with symmetry assumption, the CCS filters need to be designed for only half number of the listening positions by a shuffler filter structure [20], [21]. To enhance the performance of the system, the interpolation method is used to create the virtual inverse filters. By using the interpolation method, the adaptive 3D-sound projection screen can renders good 3D sound effect continually for the lateral movement. It only takes low computation and memory for real-time

implementation.

The head tracking system can track the listener’s position rapidly. According to the lateral movement feedback from the video head tracker, the CCS is capable of rendering immersive spatial sound with robustness against listerner’s head movement.

2 Theory and Method

A loudspeaker is an electro acoustic transducer that converts the electrical signal to sound signal. The processes of the transduction are complex. These cover the electrical, mechanical, and acoustical transduction. In order to model the process of the transduction, the EMA analogous circuit can be used to simulate the dynamic behavior of the loudspeaker. The circuit is overall and decomposed to electrical, mechanical, and acoustic part. A loudspeaker is characterized by a mixed of electrical, mechanical, and acoustical parameters.

2.1 Electrical-mechanical-acoustical analogous circuit

The concept of the electric circuit often applied to analyze transducers in the electrical and mechanical system. The technique analysis of the electric circuit can be adopted to analyze the transduction of the mechanical and acoustical system. The simple diagram of EMA analogous circuit is shown in Fig. 1. The subject of EMA analogous circuit is the application of electrical circuit theory to solve the coupling of the electrical, mechanical and acoustical system. The EMA analogous circuit is formulated by the differential equations of the electrical, mechanical, and acoustical system and the differential equations can be modeled by the circuit diagram. The rules of analytic methods are follows. For the electromagnetic loudspeaker, the diaphragm is driven by the voice coil. The voice coil has inductance and resistance which are defined R_E and L_E . The term R_E and L_E are the most common description of a loudspeaker's electrical impedance. In order to model the nonlinearity of inductance, a resistance R'_E can be parallel connected to inductance. Thus, the electrical impedance of loudspeaker is formulated as:

$$Z_E = R_E + (j\omega L_E \,//\, R'_E) \tag{1}$$

When the current (i) is passed through the voice coil, the force (f) is produced and that drives the diaphragm to radiate sound. The voltage (e) induced in the voice coil when it movies with the mechanical velocity (u) . The basic electromechanical

equations that relate the transduction of the electrical and mechanical system are listed.

$$f = Bli \tag{2}$$

$$e = Blu \tag{3}$$

Here, electro-mechanical transduction can be modeled by a gyrator. So, the loudspeaker impedance is formulated as:

$$Z = \frac{e}{i} = Z_E + \frac{Bl^2}{Z_M + Z_{MA}} \tag{4}$$

where Z_M is the mechanical impedance and Z_{MA} is the acoustical impedance reflecting in mechanical system as shown in Fig. 1(b).

A simple driver model is shown in Fig. 2. This simple driver model can be used to describe the mechanical dynamics of the electromagnetic loudspeaker. Force f is produced according to the Eq. (2). Vibration of the diaphragm of the loudspeaker displaces air volume at the interface. The primary parameters of the simple driver are the mass, compliance (compliance is the reciprocal of stiffness) and damping in the mechanical impedance. The acoustical impedance is induced by the radiation impedance, enclosure effect and perforation of the enclosure. f_S is the force that air exerts on the structure. The coupled mechanical and acoustical systems can be simplified as :

$$M_{MD}\ddot{x} = f - \frac{x}{C_{MS}} - R_{MS}\dot{x} - f_S \tag{5}$$

where M_{MD} is the mass of diaphragm and voice coil, f is the force in newtons, f_S is the force that air exert on the structure, C_{MS} is the mechanical compliance, R_{MS} is the mechanical resistance and x is the displacement.

$$M_{MD}(s)(j\omega)^2 x(s) = f(s) - \frac{x(s)}{C_{MS}} - R_{MS} j\omega x(s) - f_S \tag{6}$$

$$M_{MD}(s) j\omega u(s) = f(s) - \frac{u(s)}{j\omega C_{MS}} - R_{MS} u(s) - f_S$$

$$f = (Z_M + Z_A) u(s) \tag{7}$$

where $Z_M = j\omega M_{MD} + R_{MS} + \frac{1}{j\omega C_{MS}}$ is the mechanical impedance and Z_A is the acoustical impedance.

$$f_S = Z_A u \tag{8}$$

The acoustical impedance primarily includes radiation impedance, enclosure impedance, and perforation of the enclosure. The acoustical impedance can be formulated as:

$$Z_A = Z_{AF} + Z_{AB} \tag{9}$$

The general acoustic circuit is shown in Fig. 3(a). The Z_{AF} means the impedance in the front of diaphragm and Z_{AB} means that in the back side. In general, the circuit would turn to Fig. 3(b) the general form in the electronics. The following discussion will use this kind of circuit.

The two basic variables in acoustical analogous circuit are pressure p and volume velocity U . Because of using impedance analogy, the voltage becomes pressure p and current becomes volume velocity U . Therefore, the ground of this circuit showing in Fig. 3 means the pressure of the free air. Thus, it also can employ the concept about the mechanical system and the acoustical system can be coupled by the below two equations.

$$f_S = S_D p \tag{10}$$

$$U = S_D u \tag{11}$$

The equation $f_S = S_D p$ represents the acoustic force on the diaphragm generated by the difference in pressure between its front and back side, where D S is the effective diaphragm area and p is the difference in acoustic pressure across the diaphragm. The volume velocity source $U = S_D u$ represents the volume velocity

emitted by the diaphragm. From the Eq. (10), the pressure difference between the front and rear of the diaphragm is given by

$$p = U(Z_{AF} + Z_{AB}) \tag{12}$$

Using Eqs. (10) and (11), force field can be transformed to pressure field.

2.2 The method of parameter identification

Almost all of the useful loudspeaker parameters had been defined by other researchers before Thiele and Small. However, Thiele and Small made these parameters in a complete design approach and shown how they could be easily determined from impedance data. There are at least four methods for measuring Thiele and Small parameters from driver impedance data. They are:

1. Closed box (Delta compliance method)
2. Added mass (Delta mass method)
3. Open box only
4. Open box/closed box

The first two procedures are the most popular. One is for miniature speakers, the other one is for normal loudspeakers. The closed box method and curve fitting method are adopted to calculate the Thiele and Small parameters. Placing the driver in a closed box will induce the alteration of the resonant frequency. The curve fitting employs the impedance of system to calculate the parameters of Thiele and Small precisely. Both methods are explained in the following section.

Curve fitting method

The curve fitting method is used to calculate Q_{ES} and the result is more accurate. The procedure of the curve fitting method is explained as follows.

(a) Choose the $(\frac{1}{j\omega M + R + \frac{1}{j\omega C}})$ to be become the basic element that it fit a peak of the impedance curve. Because the purpose of the method is to fit the mechanical part, the electrical part can be obtained previously.

(b) Choose the fitting range in the impedance curve. If the range of the impedance curve is chosen broadly, result of the fitting is poor. Therefore, the range that starts and ends both sides of peak enclosures the peak, and it can be chosen. Then, the peak will fit better and it is obtained second order system transfer function.

(c) We compare the coefficient between the second order transfer function and $\frac{1}{s^2 + 2\xi\omega_s + \omega_s^2}$, then the parameters ω_s and Q_{MS} are solved.

$$\omega_s = 2\pi f_S$$

$$Q_{MS} = \frac{1}{2\xi} \tag{13}$$

$$Q_{ES} = Q_{MS}(\frac{R_E}{R_{ES}}) \tag{14}$$

Closed box method

When the impedance of a mechanical system is $Z_M = j\omega M_{MS} + R_{MS} + \frac{1}{j\omega C_{MS}}$ the resonant frequency is $\omega_s = \frac{1}{\sqrt{M_{MD} C_{MS}}}$. When a driver is placed in a closed box, its resonant frequency rises. This is because the inward cone motion is resisted not only by the compliance of its own suspension, but also by the compression of the air in box. The compliance of the driver suspension is reduced by the compliance of the air spring. If the total compliance has decreased, the resonant frequency of the driver will rise. The concept can employed to calculate to the mechanical mass, mechanical compliance and mechanical resistance of the system.

The closed box procedure for determining T-S parameters is given below:

1. Measure f_S and Q_{ES} using the curve fitting method

2. Mount the driver in the test box. Make sure there are no air leaks around the box and speaker. One point must be noticed is that the testing volume for the case of miniature speaker must be less than 0.015L, or you can't measure the realizable T/S parameters.

3. Measure the new in-box resonant frequency and electrical *Q* using the same procedure as that used in step 1. Label these new value f_C and Q_{EC} .

4. Compute the V_{AS} as follows:

$V_{AS} = V_T(\frac{f_C Q_{EC}}{f_S Q_{ES}} - 1)$ Where V_T is the total volume of the tested box

Therefore, the mechanical mass M_{MD} and mechanical compliance C_{MS} can be solved as

$$C_{MS} = \frac{V_{AS}}{\rho_0 c^2 S_D^2} \quad (15)$$

$$M_{MS} = \frac{V_{AS}}{\omega_S^2 C_{MS}} \quad (16)$$

$$M_{MD} = M_{MS} - 2M_1 \quad (17)$$

where M_1 is the air-load impedance at low frequency.

On the other hand, the parameters, and the mechanic resistance (*MS R*) and the motor

constant (*Bl*) can be calculated, using the following formula:

$$R_{MS} = \frac{\omega_S M_{MS}}{Q_{MS}} \quad (18)$$

$$Bl = \sqrt{\frac{\omega_S R_E M_{MS}}{Q_{MS}}} \quad (19)$$

And the lossy voice-coil inductance can be calculated, using the following method:

$Z_E(j\omega) \approx (j\omega)^n L_E$

$$R_E^{'} = \left[\frac{L_e}{\cos(n\pi/2)} \right] \omega^n , L_E = \left[\frac{L_e}{\cos(n\pi/2)} \right] \omega^{n-1} \quad (20)$$

(n=1:inductor;n=0:resistor)

The parameters n and L_e can be determined from one measurement of Z_{VC} at a frequency well above f_s , where the motional impedance can be neglected

$$Z_E = Z_{VC} - R_E$$

$$n = \frac{1}{90} \tan^{-1} \left[\frac{\mathrm{Im}(Z_E)}{\mathrm{Re}(Z_E)} \right] = \frac{\ln|Z_2| - \ln|Z_1|}{\ln \omega_2 - \ln \omega_1} \quad , \quad L_E = \frac{|Z_E|}{\omega^n} \quad (21)$$

The method to calculate lossy voice-coil inductance is described [22].

2.3 Modeling Acoustical Systems

Electroacoustics is using the analogous circuit to model the acoustical behavior including acoustic mass, acoustic resistance and acoustic compliance. The impedance type of analogy is the preferred analogy for acoustical circuits. The sound pressure is analogous to voltage in electrical circuits. The volume velocity is analogous to current.

Acoustic Resistance

Acoustic resistance is associated with dissipative losses that occur when there is a viscous flow of air through a fine mesh screen or through a capillary tube. Fig. 4(a) illustrates a fine mesh screen with a volume velocity U flowing through it. The pressure difference across the screen is given by $p = p_1 - p_2$, where p_1 is the pressure on the side that U enters and p_2 is the pressure on the side that U exits. The pressure difference is related to the volume velocity through the screen by

$$p = p_1 - p_2 = R_A U \quad (22)$$

where R_A is the acoustic resistance of the screen. The circuit is shown in Fig. 4(b). Theoretical formulas for acoustic resistance are generally not available. The values are usually determined by experiments. Table 1 gives the acoustic resistance of

typical screens as a function of the area S of the screen, the number of wires in the screen, and the diameter of the wires.

Acoustic compliance

Acoustic compliance is a parameter that is associated with any volume of air that is compressed by an applied force without an acceleration of its center of gravity. To illustrated an acoustic compliance, consider an enclosed volume of air as illustrated in Fig. 5(a). A piston of area S is shown in one wall of the enclosure. When a force f is applied to the piston, it moves and compresses the air. Denote the piston displacement by x and its velocity by u . When the air is compressed, a restoring force is generated which can be written $f = k_M x$, where k_M is the spring constant. (This assumes that the displacement is not too large or the process cannot be modeled with linear equation.) The mechanical compliance is defined as the reciprocal of the spring constant. Thus we can write

$$f = k_M x = \frac{x}{C_M} = \frac{1}{C_M}\int u dt \tag{23}$$

This equation involves the mechanical variables f and u . We convert it to one that involves acoustic variables p and U by writing $f = pS$ and $u = U/S$ to obtain

$$p = \frac{1}{S^2 C_M}\int U dt = \frac{1}{C_A}\int U dt \tag{24}$$

This equation defines the acoustic compliance C_A of the air in the volume. It is given by

$$C_A = S^2 C_M \tag{25}$$

An integration in the time domain corresponds to a division by $j\omega$ for phasor variable. It follows from Eqs. (24). That the phasor pressure is related to the phasor volume velocity by $p = \frac{U}{j\omega C_A}$. Thus the acoustic impedance of the compliance is

$$Z_A = \frac{p}{U} = \frac{1}{jwC_A} \tag{26}$$

The impedance which varies inversely with $j\omega$ is a capacitor. The analogous

circuit is shown in Fig. 5(b). The figure shows one side of the capacitor connected to ground. This is because the pressure in a volume of air is measured with respect to zero pressure. One node of an acoustic compliance always connects to the ground node. The acoustic compliance of the volume of air is given by the expression derived for the plane wave tube. It is

$$C_A = \frac{V}{\rho c^2} \tag{27}$$

Acoustic mass

Any volume of air that is accelerated without being compressed acts as an acoustic mass. Consider the cylindrical tube of air illustrated in Fig. 6(a) having a length l and cross-section S. The mss of the air in the tube is M 0 $M_M = \rho_0 Sl$. If the air moved with velocity u, the force required is given by $f = M_M \, du/dt$. The volume velocity of the air through the tube is $U = Su$ and the pressure difference between the two ends is $p = p_1 - p_2 = f/S$. It follows from these relations that the pressure difference p can be related to the volume velocity U as follows:

$$p = p_1 - p_2 = \frac{M_M}{S}\frac{du}{dt} = \frac{M_M}{S^2}\frac{dU}{dt} = M_A \frac{dU}{dt} \tag{28}$$

where M_A is the acoustic mass of the air in the volume that is given by

$$M_A = \frac{M_M}{S^2} = \frac{\rho_0 l}{S} \tag{29}$$

A differentiation in the time domain corresponds to a multiplication by $j\omega$ for sinusoidal phasor variable. If follows from Eqs. (28) that the phasor pressure is related to the phasor volume velocity by $p = j\omega M_A U$. Thus the acoustic impedance of the mass is

$$Z_A = \frac{p}{U} = j\omega M_A \tag{30}$$

An electrical impedance which is proportional to $j\omega$ is an inductor. The analogous circuit is shown in Fig. 6(b). For a tube of air to act as a pure acoustic mass,

each particle of air in the tube must move with the same velocity. This is strictly true only if the frequency is low enough. Otherwise, the motion of the air particles must be modeled by a wave equation. An often used criterion that the air in the tube act as a pure acoustic mass is that its length must satisfy $l \le \lambda/8$, where λ is the wavelength.

Radiation impedance of a baffled rigid piston

Radiation impedance can be easily explained by an example of the diaphragm vibration. When the diaphragm is vibrating, the medium reacts against the motion of the diaphragm. The phenomenon of this can be described as there is impedance between the diaphragm and the medium. The impedance is called the radiation impedance.

The detail of the theory of radiation impedance is clearly described by Bernek. The analogous circuit of the radiation impedance for the piston mounted in an infinite baffle is shown in Fig. 7. The acoustical radiation impedance for a piston in an infinite baffle can be approximately over the whole frequency range by the analogous circuit. The parameters of the analogous values are given by

$$M_{A1} = \frac{8\rho_0}{3\pi^2 a} \tag{31}$$

$$R_{A1} = \frac{0.4410\rho_0 c}{\pi a^2} \tag{32}$$

$$R_{A2} = \frac{\rho_0 c}{\pi a^2} \tag{33}$$

$$C_{A1} = \frac{5.94a^3}{\rho_0 c^2} \tag{34}$$

where ρ_0 is the density of air, c is the sound speed in the air, a is the radius of the circuit piston.

Radiation impedance on a piston in a tube

The flat circuit piston in an infinite baffle that is analyzed in the preceding section is commonly used to model the diaphragm of a direct-radiator loudspeaker

when the enclosure is installed in a wall or against a wall. If a loudspeaker is operated away from a wall, the acoustic impedance on its diaphragm changes. It is not possible to exactly model the acoustic radiation impedance of this case. An approximate model that is often used is the flat circuit piston in a tube.

The analogous circuit for the piston in a long tube is the same from as that for the piston in an infinite baffle; only the element values are different. The analogous circuit is given in Fig. 7. The parameters of the analogous values are given by

$$M_{A1} = \frac{0.6133\rho}{\pi a} \tag{35}$$

$$R_{A1} = \frac{0.5045\rho c}{\pi a^2} \tag{36}$$

$$R_{A2} = \frac{\rho c}{\pi a^2} \tag{37}$$

$$C_{A1} = \frac{0.55\pi^2 a^3}{\rho c^2} \tag{38}$$

Transmission line model of a duct

Consider a length of duct, we can write the pressure at any point in the duct as a superposition of two standing waves:

$$p(x) = B_1 \cos kx + B_2 \sin kx \tag{39}$$

(We could also have started with the superposition of two traveling waves, one in each direction.) Using this expression, we will evaluate the constants, *B*1 and *B*2, so that the pressure at $x = 0$ is $p1$ and the volume velocity at $x = 0$ is $U1$. Find $U(x)$ from Newton's second law in a fluid,

$$-\frac{dp}{dx} = j\omega\rho v$$

$$U(x) = -v \cdot A = \frac{A}{j\omega\rho}\frac{dp}{dx} \tag{40}$$

$$U(x) = \frac{Ak}{j\omega\rho}[-B_1 \sin kx + B_2 \cos kx] \tag{42}$$

Replace B_1 and B_2 by using the pressure and volume velocity at $x = 0$ and evaluate

the pressure and volume velocity at $x = L$.

$$p_2 = p_1 \cos kL + U_1 (\frac{j\omega\rho}{Ak} \sin Kl) \tag{43}$$

$$U_2 = p_1 (-\frac{Ak}{j\omega\rho} \sin kL) + U_1 \cos kL \tag{44}$$

$$\begin{bmatrix} p_2 \\ U_2 \end{bmatrix} = \begin{bmatrix} \cos kL & -\frac{Ak}{j\omega\rho} \sin kL \\ \frac{j\omega\rho}{Ak} \sin kL & \cos kL \end{bmatrix} \begin{bmatrix} p_1 \\ U_1 \end{bmatrix} \tag{45}$$

These equations are in transfer matrix form.

Compare to transfer matrix for the T-circuit shown in Fig. 8. Its transfer matrix is shown below:

$$\text{T} = \begin{bmatrix} 1 + \frac{Z_a}{Z_b} & Z_a \left(2 + \frac{Z_a}{Z_b} \right) \\ \frac{1}{Z_b} & 1 + \frac{Z_a}{Z_b} \end{bmatrix} \tag{46}$$

Then Z_a and Z_b can be got as:

$$Z_a = jZ_0 \tan(\frac{kL}{2})$$

$$Z_b = \frac{Z_0}{j \sin kL} \tag{47}$$

$$Z_0 \equiv \frac{\rho c}{A}$$

Other acoustic elements

Perforated sheets are often used as an acoustic resistance in application where an acoustic mass in series with the resistance is acceptable. Fig. 9(a) illustrates the geometry. If the holes in the sheet have centers tat are spaced more than on diameter apart and the radius a of the holes satisfies the inequality $0.01/\sqrt{f} < a < 10/f$, where f is the frequency and a is in m, the acoustic impedance of the sheet is given by

$$Z_A = \frac{\rho_0}{N\pi a^2} \left\{ \sqrt{2\omega u} \left[\frac{t}{a} + 2\left(1 - \frac{\pi a^2}{b^2} \right) \right] + j\omega \left[t + 1.7\left(1 - \frac{a}{b} \right) \right] \right\} \tag{48}$$

where N is the number of holes. The parameters μ is the kinematic coefficient of viscosity. For air at $20^{\circ}C$ and 0.76 mHg, $\mu \approx 1.56\times10^{-5}\ m^2/s$. This parameter value approximately as $T^{1.7}/P_0$, where T is the Kelvin temperature and P_0 is the atmospheric pressure.

A tube having a very small diameter is another example of an acoustic element which exhibits both a resistance and a mass. If the tube radius a in meters satisfies the inequality $a < 0.002/\sqrt{f}$, the acoustic impedance is given by

$$Z_A = \frac{8\eta l}{\pi a^4} + j\omega\frac{4\rho_0 l'}{3\pi a^2} \tag{49}$$

where l is the actual length of the tube and l' is the length including end corrections. The parameter η is the viscosity coefficient. For air, $\eta = 1.86\times10^{-5}\ N\cdot S/m^2$ at $20^{\circ}C$ and 0.76 mHg. This parameter varies with temperature as $T^{0.7}$, where T is the Kelvin temperature. If the radius of the tube satisfies the inequality $0.01/\sqrt{f} < a < 10/f$, the acoustic impedance is given by

$$Z_A = \frac{\rho_0}{\pi a^2}\sqrt{2\omega u}\left(\frac{l}{s}+2\right) + j\omega\frac{\rho_0 l'}{\pi a^2} \tag{50}$$

For a tube with a radius such that $0.002/\sqrt{f} < a < 0.01/\sqrt{f}$, interpolation must be used between the two equations.

A narrow slit also exhibits both acoustic resistance and mass. Fig. 9(b) shows the geometry of such a slit. If the height t of the slit in meters satisfies the inequality $t < 0.003/\sqrt{f}$, the acoustic impedance of the slit, neglecting end corrections for the mass term, is given by

$$Z_A = \frac{12\eta l}{t^3\omega} + j\omega\frac{\rho_0 l}{5\omega t} \tag{51}$$

2.4 Crosstalk Cancellation Systems

Fig. 10 shows a two-channel loudspeaker reproduction scenario, where H_{11} and

H_{22} are ipsi-lateral transfer functions, and H_{12} and H_{21} are contra-lateral transfer functions from the loudspeakers to the listener's ears. The contra-lateral transfer functions, also known as the crosstalks, interfere with human's localization of sound sources when the binaural signals are reproduced by loudspeakers. In order to mitigate the effects of crosstalk, the crosstalk canceller is chosen to be the inverse of the acoustic plants such that the overall response becomes a diagonalized and distortion less response.

$$\begin{bmatrix} \delta(n-m) & 0 \\ 0 & \delta(n-m) \end{bmatrix} = \begin{bmatrix} h_{11}(n) & h_{12}(n) \\ h_{21}(n) & h_{22}(n) \end{bmatrix} \otimes \begin{bmatrix} c_{11}(n) & c_{12}(n) \\ c_{21}(n) & c_{22}(n) \end{bmatrix}, \qquad (52)$$

where $\otimes$ denotes convolution operation and $h_{ij}(n)$, $c_{ij}(n)$ and $\delta(n-m)$ represent the impulse responses of the respective acoustic paths, the inverse filters, and the discrete delta function delayed by m samples delay to ensure a causal inverse filter. The architecture of the crosstalk canceller is shown in Fig. 11. In the rationale of the inversion problem, it can be regarded as a model-matching system as shown in Fig.12. The recorded input signals are denoted by $\mathbf{u}(z)$, the speaker input signals by $\mathbf{v}(z)$,the reproduced signals by $\mathbf{w}(z)$, the desired signals by $\mathbf{d}(z)$, and the performance error signals by $\mathbf{e}(z)$. The matrices $\mathbf{M}(z)$, $\mathbf{H}(z)$, and $\mathbf{C}(z)$ represent multi-channel filters. $\mathbf{M}(z)$ is the target matrix, $\mathbf{H}(z)$ is the plant transfer matrix, and $\mathbf{C}(z)$ is the matrix of CCS filters. The term z^{-m} is referred to as a modeling delay to ensure that it is possible to achieve a good performance from the CCS filter under the constraint that they are causal.

For the system, it is straightforward to establish the following relationships:

$$\mathbf{v}(z) = \mathbf{C}(z)\mathbf{u}(z) \qquad (53)$$

$$\mathbf{w}(z) = \mathbf{H}(z)\mathbf{v}(z) \qquad (54)$$

$$\mathbf{d}(z) = z^{-m}\mathbf{M}(z)\mathbf{u}(z) \qquad (55)$$

$$\mathbf{e}(z) = \mathbf{d}(z) - \mathbf{w}(z). \tag{56}$$

Ideal model matching requires that $\mathbf{H}(z)\mathbf{C}(z) = \mathbf{M}(z)$. $\mathbf{H}(z)$ is generally non-invertible because it is usually ill-conditioned and even non-square. To overcome this difficulty, we employ the Tikhonov regularization [23] in the matrix inversion process. In the method, a frequency-domain cost function J is defined as the sum of the "performance error" $\mathbf{e}^{\mathrm{H}}\mathbf{e}$ and the "input power" $\mathbf{v}^{\mathrm{H}}\mathbf{v}$:

$$J(e^{j\omega}) = \mathbf{e}^{\mathrm{H}}(e^{j\omega})\mathbf{e}(e^{j\omega}) + \beta^2(\omega)\mathbf{v}^{\mathrm{H}}(e^{j\omega})\mathbf{v}(e^{j\omega}). \tag{57}$$

A regularization parameter $\beta(\omega)$ weighs the input power against the performance error. If β is too small, there will be sharp peaks in the frequency responses of the CCS filters, whereas if β is too large, the cancellation performance will be rather poor. The optimal input $\mathbf{v}_{\mathrm{opt}}(e^{j\omega})$ can be obtained by minimizing J

$$\mathbf{v}_{\mathrm{opt}}(e^{j\omega}) = \left[\mathbf{H}^H(e^{j\omega})\mathbf{H}(e^{j\omega}) + \beta^2(\omega)\mathbf{I}\right]^{-1}\mathbf{H}^H(e^{j\omega})\mathbf{M}(e^{j\omega})\mathbf{u}(e^{j\omega}). \tag{58}$$

This solution always exists for $\beta \neq 0$ irrespective of the dimensions and rank of $\mathbf{H}(e^{j\omega})$. Consequently, the CCS matrix can be readily identified as

$$\mathbf{C}(e^{j\omega}) = \left[\mathbf{H}^H(e^{j\omega})\mathbf{H}(e^{j\omega}) + \beta^2(\omega)\mathbf{I}\right]^{-1}\mathbf{H}^H(e^{j\omega})\mathbf{M}(e^{j\omega}). \tag{59}$$

In the case when the desired signals $\mathbf{d}(z)$ are identical to the binaural signals $\mathbf{u}(z)$, the matrix $\mathbf{M}(z)$ is an identity matrix of order $R = B$ and the corresponding optimal filters are given by

$$\mathbf{C}(e^{j\omega}) = \left[\mathbf{H}^{\mathrm{H}}(e^{j\omega})\mathbf{H}(e^{j\omega}) + \beta^2(\omega)\mathbf{I}\right]^{-1}\mathbf{H}^H(e^{j\omega}). \tag{60}$$

While the expression in Eq. (60) may look similar to that in Ref. [24], there is a distinction in the choice of $\beta(\omega)$. In our approach, the parameter $\beta(\omega)$ is frequency dependent. This is in contrast to the approach in Ref. [24], where a constant $\beta(\omega)$ applied to all frequencies.

Next, the frequency response matrix $\mathbf{C}(e^{j\omega})$ is sampled at N_c equally spaced frequencies

$$\mathbf{C}(k)=\left[\mathbf{H}^{H}(k)\mathbf{H}(k)+\beta^{2}(k)\mathbf{I}\right]^{-1}\mathbf{H}^{H}(k),\quad k=1,2,\ldots N_c \quad (61)$$

The impulse responses of the inverse filters can be obtained using inverse FFT of the frequency samples of Eq. (61) in conjunction with appropriate windowing. In order to guarantee the causality of the CCS filters, cyclic shift of the impulse response matrix is generally needed, hence the modeling delay z^{-m} in Fig. 12.

In designing the inverse filter, it is crucial to apply suitable regularization with varying degree according to frequency bands. In this paper, two strategies are employed to choosing the frequency-dependent parameter $\beta(k)$. The first strategy is based on to the condition number $c_d(k)$ of the plant matrix $\mathbf{H}(k)$, in which case $\beta(k)$ is set to be a fraction r of the condition number, i.e., $\beta(k)=r\cdot c_d(k)$. Next, the resulting impulse response is examined if the premasking limit (-15 ms) in psychoacoustics [25] is violated. If violated, go back and adjust the parameter r until the limit is satisfied. The second strategy of choosing a frequency-dependent $\beta(k)$ is based on a gain threshold on the maximum of the absolute values of all entries in $\mathbf{C}(k)$. In the paper, the threshold is selected to be 12 dB. If the threshold is exceeded, a larger β should be chosen. The binary search method can be used to accelerate the search.

2.5 Rectangular Haar-like features

A simple rectangular Haar-like feature can be defined as the difference of the sum of pixels of areas inside the rectangle, which can be at any position and scale within the original image. This modified feature set is called 2 rectangle feature. *Viola and Jones* also defined 3 rectangle features and 4 rectangle features. The values indicate certain characteristics of a particular area of the image. Each feature type can indicate the existence (or not) of certain characteristics in the image, such as edges or changes in texture.

2.6 AdaBoost

AdaBoost, short for Adaptive Boosting, is a machine learning algorithm, formulated by Yoav Freund and Robert Schapire. It is a meta-algorithm, and can be used in conjunction with many other learning algorithms to improve their performance.

AdaBoost is adaptive in the sense that subsequent classifiers built are tweaked in favor of those instances misclassified by previous classifiers. AdaBoost is sensitive to noisy data and outliers. Otherwise, it is less susceptible to the overfitting problem than most learning algorithms.

3. Acoustical design of a Bluetooth® Earphone

3.1 Electroacoustic modeling of earphone

3.1.1 EMA analogous circuit of earphone

In this section, a lumped parameter model based on EMA analogy is established for the earphone. The cross-section of a Bluetooth earphone connected to a type 2 artificial ear is shown in Fig. 13. In Fig. 3 (a), the coupling of the electrical and the mechanical domains is modeled by a gyrator, whereas the coupling of the mechanical and the acoustical domains is modeled by a transformer. The T-S parameters of the microspeaker identified via an electrical impedance measurement are summarized in Table 1. A distinct feature of the electroacoustic modeling of the earphone as compared with the other free-field loudspeaker systems is the ear canal impedance. To gain an appreciation of the main difference of these two environment, the ear canal is approximated as a cavity with acoustic compliance C_A. Simplifying the circuit in the acoustical domain leads to the pressure response at the ear canal

$$p_c = \frac{C'Bl}{C_A S_D R_E} \frac{1}{\left(\frac{s}{\omega_s}\right)^2 + \left(\frac{1}{Q_{AT}}\right)\left(\frac{s}{\omega_s}\right) + 1} e_g, \tag{62}$$

where $C' = \frac{C_A C_{AS}}{C_A + C_{AS}}$, $\omega_s = \frac{1}{\sqrt{M_{AD} C'}}$, $Q_{AT} = \frac{1}{R_{AT}} \sqrt{\frac{M_{AD}}{C'}}$, C_{AS} is the acoustic compliance of the suspension, M_{AD} is the acoustic mass of the diaphragm, $R_{AT} = \frac{(Bl)^2}{S_D^2 R_E} + \frac{R_{MS}}{S_D^2}$, Bl is the force factor of the voice coil, R_E is the voice-coil resistance, R_{MS} is the mechanical resistance of the suspension, and S_D is the effective area of the diaphragm. It follows that the pressure response has a second-order lowpass characteristic, which is quite different from a direct radiator typically having a second-order highpass characteristic. Note that, since the

combined acoustic compliance C' is effectively decreased due to the ear canal impedance, the system quality factor Q_{AT} will be increased. Thus, it is crucial in designing earphones to "shape" the resonant peak at ω_s for an acceptable Q_{AT} value.

Apart from the approximation above, a more detailed circuit model of the acoustical system is illustrated in Fig. 14 (a). The acoustical system primarily consists of tree parts: a cavity (C_{AF}) and a duct in front of the speaker, the artificial ear, and a cavity (C_{AB}) with a leakage hole behind the speaker.

The duct in front of the speaker can be modeled as an acoustic resistance [9] R_{ST} cascaded with a transmission line which can be simulated with a T-circuit with parameters Z_{STA} and Z_{STB}. The T- circuit with parameters is given by

$$Z_A = jZ_0 \tan(\frac{kL}{2}) \tag{63}$$

$$Z_B = \frac{Z_0}{j \sin kL} \tag{64}$$

$$Z_0 = \frac{\rho_0 c}{\pi a^2}, \tag{65}$$

where $Z_A = Z_{STA}$, $Z_B = Z_{STB}$, $a = a_{ST}$ is the radius of the cross-section of the duct, $L = L_{ST}$ is the length of the duct, and $k = \frac{\omega}{c}$ is the wave number.

The analogous circuit of the type-2 artificial ear using IEC 711 ear simulator [26] is shown in Fig. 14 (a). Instead of an acoustic mass, the duct embedded in the artificial ear is modeled with a transmission line T-circuit. The artificial ear can be modeled as a transmission line T-circuit in series with the circuit of an IEC 711 simulator [26], as shown in Fig. 14(b). The parameters of the transmission line T-circuit have been defined in Eqs. (63) - (65), where $Z_A = Z_{AEA}$, $Z_B = Z_{AEB}$, $a = a_{AE}$ and $L = L_{AE}$.

The leakage hole impedance is given in the acoustical domain as [9]

$$Z_{LK} = R_{LK} + j\omega M_{LK} \tag{66}$$

The leakage hole has radiation of air loading. The acoustic element parameters calculated according to the formulas given in Ref. 9 are summarized in Table 2. This analogous circuit serves as the simulation platform for the earphone.

3.1.2 Verification of the lumped parameter model

The volume velocity U_{AE} flowing through C_{A8} in the artificial ear can be obtained from solving the analogous circuit in Fig.2. From U_{AE}, the pressure p_{CA8} at the element C_{A8} can be calculated by

$$p_{CA8} = U_{AE} \frac{1}{j\omega C_{A8}} \tag{67}$$

The ear canal and the eardrum are simulated by the transmission line T-circuit shown in Fig. 14 (b). Assume the eardrum here is rigid. Hence, the eardrum impedance $Z_{ED} \to \infty$, which renders the Z_{ED} in the ear canal circuit in Fig. 14 (b) an open circuit. Therefore, the sound pressure at the tympani position p_{ED} is given by

$$p_{ED} = p_{CA8} \frac{Z_{ECA}}{Z_{ECA} + Z_{ECB}}, \tag{68}$$

where Z_{ECA} and Z_{ECB} denote the impedance elements of the duct represented by the transmission line T-circuit, as defined in Eqs. (2)-(4), and $Z_{ECA} = Z_A$, $Z_{ECB} = Z_B$. $L_{EC} = L$ the length of the duct and $a_{EC} = a$ is the radius of the duct.

Experiments were undertaken to validate the aforementioned earphone simulation model. The dimensions of the earphone enclosure are shown in Table 2. It can be observed from Fig. 15 that SPL response predicted by lumped parameter model (denoted as original simulation) is in good agreement with the measurement (denoted as original experiment) up to approximately 16 kHz.

3.2 Optimization of the enclosure design of earphone

The SPL response of the earphone shown in Fig. 15 apparently did not meet the requirement in 3GPP2 C.S0056-0 [13]. The peak at 1 kHz exceeded the frequency

response mask. This calls for optimization of the enclosure design, where Simulated annealing (SA) is exploited in this study.

SA is a generic probabilistic meta-algorithm for the global optimization problem, namely, locating a good approximation to the global optimum of a given function in a large search space. SA's major advantage over other methods is the ability to avoid becoming trapped at local minima. The algorithm employs a random search which, in the initial stage, not only accepts changes that decrease cost function Q but also changes (bad solutions) that increase it. The acceptance of the bad solutions is determined by the probability

$$P = \exp\left(-\frac{\Delta Q}{T}\right) > \gamma(0,1), \tag{69}$$

where ΔQ is the increase in Q and T is a control parameter, which by analogy is known as the system "temperature" irrespective of the cost function involved. $\gamma(0,1)$ is a random number generated uniformly in the interval (0,1).

In the earphone optimization, we choose the initial temperature $T_i = 200$ and the final temperature $T_f = 10^{-3}$. The temperature decrement rule is given by

$$T_{k+1} = \alpha T_k, \tag{70}$$

where the annealing coefficient $\alpha = 0.95$ when $T > 20$, $\alpha = 0.99$ when $T < 20$. The solution will be rejected if it fails to comply with the the frequency response mask. The cost function for the SA optimization is chosen to be

$$Q = \sum_{n=1}^{95} \left[SPL_{new}(n) - L_{ref}(n) \right]^2, \tag{71}$$

where SPL_{new} is the SPL of the current design, while L_{ref} is a smooth reference SPL curve passing the central region of the mask. The parameter n is the frequency index within the band 20–4500 Hz. The design variables and the associated constraints are given in the following inequalities:

$$\begin{cases} 2\times10^{-4} \le a_{SP} \le 3\times10^{-3} \\ \quad 10^{-3} \le L_{SP} \le 10^{-2} \\ 2\times10^{-9} \le V_{AF} \le 9\times10^{-8} \\ 2\times10^{-9} \le V_{AB} \le 9\times10^{-8} \end{cases}, \tag{72}$$

where a_{SP} is the radius of the duct, L_{SP} is the length of the duct, V_{AF} is the volume of the front cavity, and V_{AB} is the volume of the back cavity.

With the SA procedure, the optimized enclosure parameters are compared with the original non-optimized ones in Table 3. Note that the front port radius has the largest design change to make. The port radius has to be reduced to 13.33% of its original size. The SPL responses of the optimal and original designs are compared in simulation and experiment, as shown in Fig. 15. As opposed to the original non-optimized design, the optimized design effectively lowers the resonance peak to be within the frequency response mask of 3GPP2. This improvement also comes at the price of steeper roll-off above 3 kHz. Nevertheless, this should not be a problem since the present study is aiming at only the speech application in which the upper cutoff frequency at 3 kHz is deemed sufficient.

4. Adaptive CCS Panel Screen Design

4.1 Band Limited CCS

Use direct filtering methods, crosstalk cancellation is carried out by direct filtering using inverse filters. However, crosstalk cancellation can be demanded either for a full-band (20~24k Hz) performance or just a band-limited performance (20-6k Hz) in the design stage of inverse filters. The reason for the latter design is twofold. First, the sweet spot in which CCS is effective becomes impractically small at high frequencies. Second, a listener's head provides natural shadowing at high frequencies so that the need for cancellation becomes less important. The match equation appropriate for the band-limited design is written as

$$\begin{bmatrix} \delta(n-m) & 0 \\ 0 & \delta(n-m) \end{bmatrix} = \begin{bmatrix} h_{11}(n) & h_{12}(n)\otimes f_{LP}(n) \\ h_{21}(n)\otimes f_{LP}(n) & h_{22}(n) \end{bmatrix} \otimes \begin{bmatrix} c_{11}(n) & c_{12}(n) \\ c_{21}(n) & c_{22}(n) \end{bmatrix}, \tag{73}$$

where $f_{LP}(n)$ denotes the impulse response function of a low pass filter. Thus, the inverse filters should in principle give rise to a flat response within the intended band after compensation.

4.2 Adaptive Head tracking system

There are three contributions of the adaptive head tracking system. The first contribution of this paper is a new image representation called an *integral image* that allows for very fast feature evaluation. The second contribution of this paper is a simple and efficient classifier that is built by selecting a small number of important features from a huge library of potential features using AdaBoost [28]. The third major contribution of this paper is a method for combining successively more complex classifiers in a cascade structure which dramatically increases the speed of the detector by focusing attention on promising regions of the image.

4.2.1 Features

The head tracking procedure classifies images based on the value of Haar features. There are two reasons for using Haar features. The first reason is that features can act to encode ad-hoc domain knowledge that is difficult to learn using a finite quantity of training data. The second reason is that the feature-based system operates much faster than pixel-based system. There are three kinds of features: two-rectangular features, three-rectangular features and four-rectangular features, which are shown in Fig. 16. The values indicate certain characteristics of a particular area of the image. Each feature type can indicate the existence (or not) of certain characteristics in the image, such as edges or changes in texture. For example, a 2 rectangle feature can indicate where the border between a dark region and a light region lies.

Rectangle features can be computed very rapidly using an integral image [19]. In order to achieve true scale invariance, almost all face detection systems must operate on multiple image scales. The integral image, by eliminating the need to compute a multi-scale image pyramid, reduces the initial image processing required for face detection significantly. Using the integral image, face detection is completed in almost the same time as it takes for an image pyramid to be computed.

4.2.2 Learning Classification Function

For the original face detector, the exhaustive set of rectangle feature is quite large. Even though each feature can be computed very efficiently, computing the complete set is prohibitively expensive. There is a solution that a very small number of these features can be combined to form an effective classifier. The main challenge is to find these features.

In head tracking system, AdaBoost is used both to select the features and to train the classifier. In its original form, the AdaBoost learning algorithm is used to boost the classification performance of a simple learning algorithm. It does this by

combining a collection of weak classification functions to form a stronger classifier. So, for example the perceptron learning algorithm searches over the set of possible perceptrons and returns the perceptron with the lowest classification error. In order for the weak learner (boosting the simple learning algorithm) to be boosted, it is called upon to solve a sequence of learning problems. After the first round of learning, the examples are re-weighted in order to emphasize those which were incorrectly classified by the previous weak classifier. The final strong classifier takes the form of a perceptron, a weighted combination of weak classifiers followed by a threshold.

The conventional AdaBoost procedure can be easily interpreted as a greedy feature selection process. Consider the general problem of boosting, in which a large set of classification functions are combined using a weighted majority vote. The challenge is to associate a large weight with each good classification function and a smaller weight with poor functions. AdaBoost is an aggressive mechanism for selecting a small set of good classification functions which nevertheless have significant variety. Drawing an analogy between weak classifiers and features, AdaBoost is an effective procedure for searching out a small number of good "features" which nevertheless have significant variety.

4.2.3 Constructing a Cascade of classifiers

This section describes an algorithm for constructing a cascade of classifiers which achieves increased detection performance while radically reducing computation time. The key insight is that smaller, and therefore more efficient, boosted classifiers can be constructed which reject many of the negative sub-windows while detecting almost all positive instances. Simpler classifiers are used to reject the majority of sub-windows before more complex classifiers are called upon to achieve low false positive rates.

The overall form of the detection process is that of a degenerate decision tree, what we call a “cascade” [29] shown in Fig. 17. A positive result from the first classifier triggers the evaluation of a second classifier which has also been adjusted to achieve very high detection rates. A positive result from the second classifier triggers a third classifier, and so on. A negative outcome at any point leads to the immediate rejection of the sub-window.

The structure of the cascade reflects the fact that within any single image an overwhelming majority of sub-windows are negative. As such, the cascade attempts to reject as many negatives as possible at the earliest stage possible. While a positive instance will trigger the evaluation of every classifier in the cascade, this is an exceedingly rare event

4.3 The Adaptive 3D-sound projection screen

A fixed CCS suffers from the sweet spot problem that severely limits its practicality. To combat the problem, we propose an adaptive interpolated CCS alongside the video head tracker as follows.

4.3.1 CCS with Interpolation Method

The inverse filter array is designed for 11 listening positions equally spaced by 20 cm. At each listening position, four inverse filters are needed. By assuming two symmetric listening positions in Fig. 19, the ipsi-lateral transfer functions of the left position H_{L11} and H_{L22} are identical to the ipsi-lateral transfer functions of the right position H_{R22} and H_{R11}, respectively. Likewise, the contra-lateral transfer functions of the left position H_{L12} and H_{L21} are identical to the contra-lateral transfer functions of the right position H_{R21} and H_{R12}, i.e.,

$$\begin{cases} H_{L11} = H_{R22} \\ H_{L22} = H_{R11} \\ H_{L12} = H_{R21} \\ H_{L21} = H_{R12} \end{cases} \tag{74}$$

It turns out that only four CCS filters need to be designed for the two symmetric listening positions. In total, 6 fixed CCS filters are needed for 11 listening positions.

The fixed CCS filters are designed for 11 discrete listening positions. In the presence of lateral movement of listener's head, a mixed filter is generated by using linear interpolation to ensure the continuity of reproduction for the intermediate position between two nearest nominal positions.

$$c_{\text{int}}(n) = \frac{(L - L_1)c_1(n) + L_1 c_2(n)}{L}, \tag{75}$$

where $c_{\text{int}}(n)$ is the impulse response of the interpolated filter of the intended interval and n is the time index. $c_1(n)$ and $c_2(n)$ are the impulse responses of the fixed CCS filters corresponding to the left and the right end positions of the intended interval, and L is the interval length. L_1 is the distance from the listener's head center to the left end position associated with $c_1(n)$.

4.3.2 System Implementation

The adaptive 3D-sound projection screen consists of the panel speakers and the head tracking system with a camera. Fig. 20 shows the arrangement of the adaptive 3D-sound projection screen where the camera is in the middle of the projection screen. The two-channel panel speakers are placed on the both two sides of the screen symmetrically. Fig. 21 shows the GUI of the head tracking system. According to the imported distance from the screen and the angle of the camera lens, the CCS can get the coordinates of the listener's lateral movements by the feedback of the head tracker. Then the appropriate CCS filters will be obtain to render the immersive spatial sound.

A panel loudspeaker consists primarily of a light panel that replaces the conical diaphragm. The size of loudspeaker is smaller in applying such design. Resonance of flexural motion is encouraged such that the panel vibrates as randomly as possible.

The exciter is a moving coil actuator that can obtain the better sound quality. The other advantages of the panel loudspeakers which are compared with conventional loudspeaker are less beaming problem, compactness, linear attenuation, insensitivity to room conditions, and bi-polar radiation.

The stereo panel speaker array of the projection screen is constructed using PU foam panels. The size of the screen frame is $1.8\,\mathrm{m} \times 2\,\mathrm{m}$. The size of each panel is $33.75\,\mathrm{cm} \times 25\,\mathrm{cm}$. Each panel is driven by two electromagnetic exciters and fixed on a wood frame and coved by a plastic skin.

4.3.3 Experimental Investigations

To verify the efficacy of the adaptive 3D virtual sound beam steering system, the head-related transfer functions (HRTF) are measured by using the dummy head. An objective performance index, channel separation, is employed to assess the cancellation performance.

$$S_{ep}(k) = H_c(k) / H_i(k), \tag{76}$$

where $H_c(k)$ and $H_i(k)$ represent the contralateral and ipsilateral frequency responses, respectively. According to the definition, smaller value of channel separation implies better cancellation performance. Processed and unprocessed channel separation results are shown in Fig. 60 Figure 60(a) shows the channel separations when a dummy sits at the nominal center position. Figures 60(b) and (c) show the channel separations when the dummy is displaced to the right by 7 cm and 30 cm, respectively, from the centerline. In is clear that the channel separation with CCS activated is significantly lower than the natural separation from 100 Hz to 8k Hz. The maximum difference of the channel separations reaches approximately 30 dB. These experimental results reveal that the interpolated CCS is effective in rendering spatial sound with robustness against the lateral movements of listener's head.

5. Conclusion

Earphones are faced with ear canal impedance, which is fundamentally different from direct radiator loudspeakers exposed in a free-field environment. EMA analogous circuits have been developed to model the Bluetooth earphone. On the basis of this simulation model, the enclosure design of the earphone has been optimized using the SA technique. The SPL response resulting from the optimized design has been significantly enhanced and the 3GPP2 standard has been met.

In this thesis, the adaptive 3D-sound projection screen composed of panel speakers, a video head tracker and an adaptive interpolated CCS is proposed. The system is capable of rendering immersive spatial sound with robustness against listener's movement. In order to cope with the lateral movement of the listener's head, we introduce the adaptive CCS based on video head tracking technique. The head tracking system based on the detection framework is capable of processing images extremely rapidly with high detection rates. The system is successfully implemented on the platform of a laptop computer and a digital signal processor. Experimental results reveal that the proposed system is capable of rendering immersive spatial sound with robustness against listener's movement.

REFERENCES

[1] I. Chun, P. A. Nelson and J. T. Kim, "Numerical models of miniature loudspeakers," in *The 32nd International Congress and Exposition on Noise Control Engineering*, Jeju Island, Korea, Aug. 2003.

[2] S. J. Oh, H. R. Lee, S. W. Yoon and J. S. Park, "Study of the Acoustical Properties as a Function of Back Cavity for Loudspeaker," in *The 32nd International Congress and Exposition on Noise Control Engineering*, Jeju Island, Korea, Aug. 2003.

[3] C. H. Choi, H. S. Yoon, "Acoustic and Vibration Characteristics of a Micro Speaker through the Electro-Magnetic field Analysis," in *The 32nd International Congress and Exposition on Noise Control Engineering*, Jeju Island, Korea, Aug. 2003.

[4] S. H. Lee, J. H. Kim, J. T. Kim, O. S. Kwon and C. H. Choi, "Development of the simulation program to analyze acoustic characteristics of a miniature type loudspeaker," in *The 32nd International Congress and Exposition on Noise Control Engineering*, Jeju Island, Korea, Aug. 2003.

[5] A. Bright, "Simplified Loudspeaker Distortion Compensation by DSP," in *The 23rd AES International Conference*, Copenhagen, Denmark, May 2003.

[6] M. R. Bai and R. L. Chen, "Optimal Design of Loudspeaker Systems Based on Sequential Quadratic Programming (SQP)," *J. Audio Eng. Soc.*, Vol. 55, No. 1/2, pp. 44-54, 2007.

[7] H. Olson, *Acoustical Engineering*, Van Nostrand, New York, 1957. Reprinted by Professional Audio Journals, Philadelphia, PA, 1991.

[8] L. L. Beranek, *Acoustics*, Acoustical Society of America, Woodbury, NY. 1996.

[9] W. M. Leach, Jr., *Introduction to Electroacoustics and Audio Amplifier Design*, Kendall-Hunt, Dubuque, IA, 2003.

[10] N. Thiele and R. Small, in *AES Loudspeaker Anthologies*, Vols. 1–3, Audio Engineering Society, New York, 1978, 1984, 1996.

[11] M. R. Bai and J., Liao, "Acoustic Analysis and Design of Miniature Loudspeakers for Mobile Phones," *Audio Engineering Society*, Vol. 53, No. 11, pp. 1061-1076, 2005.

[12] L. Ingber, "Simulated annealing: practice versus theory," *Mathematical and Computer Modelling* **18**, 11, 29-57 (1993).

[13] 3GPP2 C.S0056-0, "Electro-Acoustic Recommended Minimum Performance Specification for cdma2000 Mobile Stations" (3rd Generation Partnership Project 2, 2005).

[14] ITU-T Recommendation P.57, "Artificial ear" (International Telecommunication Union, 2006).

[15] Kirkeby, P.A. Nelson, and H. Hamada, "Fast deconvolution of multichannel systems using regularization," *IEEE Trans. Speech and Audio Processing*, vol. 6, pp. 189-195, 1998.

[16] W. G. Gardner, "3-D Audio Using Loudspeakers," Kluwer Academic Publishers, 1998.

[17] J. Rose, P. A. Nelson, B. Rafaely, and T. Takeuchi, "Sweet Spot Size of Virtual Acoustic Imaging Systems at Asymmetric Listener Locations," *J. Acoust. Soc. Am.*, vol.112, pp. 1992–2002 (2002).

[18] S. Kim, D. Kong, and S. Jang," Adaptive Virtual Surround Sound Rendering System for an Arbitrary Listening Position," *J. Audio Eng. Soc.*, Vol. 56, No. 4, 2008 April.

[19] P. Viola, M. Jones, "Rapid object detection using boosted cascade of simple features," Computer Vision and Pattern Recognition, 2001.

[20] D. H. Cooper and J. L. Bauck, "Prospects for Transaural Recording," *J. Audio*

Eng. Soc., vol. 37, pp. 3–19, 1989 Jan./Feb.

[21] J. M. Jot, “Etude et re´alization d’un spatialisateur de sons par mode`les physiques et perceptifs,” Ph.D. dissertation, Telecom, Paris, France, 1992.

[22] A. N. Thiele, “Loudspeakers in Vented-boxes: Part II,” *Audio Engineering Society*,Vol. 19, No. 6, pp. 471-483 (1971).

[23] A. Schuhmacher, and J. Hald, Sound Source Reconstruction Using Inverse Boundary Element Calculations, J. Acoust. Soc. Am. 113 (2003) 114-127.

[24] O. Kirkeby, P. A. Nelson, and H. Hamada, Fast Deconvolution of Multichannel Systems Using Regularization, IEEE Trans. Speech Audio Processing 6 (1998) 189-194.

[25] L. D. Fielder, Analysis of Traditional and Reverberation-Reducing Methods of Room Equalization, J. Audio Eng. Soc. 51 (2003) 3-26.

[26] IEC 711, “Occluded-ear simulators for the measurement of earphones coupled to the ear by ear inserts”(International Electrotechnical Commission, 1981).

[27] C. A. Poldy, “Headphones,” in *Loudspeaker and Headphone Handbook*, edited by J. Borwick (Focal Press, London, 1994).

[28] Y. Freund and, R.E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” *Computational Learning Theory: Eurocolt 95*, Springer-Verlag, pp. 23–37, 1995.

[29] J. Quinlan, “Induction of decision trees,” *Machine Learning*, 1:81–106, 1986.

Table 1 Experimentally identified lumped-parameters of the microspeaker.

$R_E\,(\text{ohm})$	31.75	Q_{TS}	5.96
$R_E^{'}\,(\text{m}^2)$	70.033	Q_{ES}	15.11
$L_E\,(\text{mH})$	6.49×10^{-2}	Q_{MS}	9.85
$S_D\,(\text{m}^2)$	7.07×10^{-6}	$Bl\,(\text{T}\cdot\text{m})$	0.34
$F_S\,(\text{Hz})$	1100.42	$V_{AS}\,(\text{L})$	1.85×10^{-5}

Table 2 The dimensions of the earphone and the parameters of acoustic analogous circuit.

Parameter	Value	Parameter	Value	Parameter	Value	Parameter	Value
R_{ST}	1.67×10^{5}	$a_{ST}\,(\mathrm{m})$	1.5×10^{-3}	M_{A4}	78.8	C_{A4}	9×10^{-13}
R_{LK}	3.85×10^{5}	$L_{ST}\,(\mathrm{m})$	3×10^{-3}	M_{A5}	9.4×10^{3}	C_{A5}	1.9×10^{-12}
M_{LK}	766.5	$a_{EC}\,(\mathrm{m})$	1.3×10^{-2}	M_{A6}	132.3	C_{A6}	1.5×10^{-12}
R_{A}	2×10^{8}	$L_{EC}\,(\mathrm{m})$	1.45×10^{-2}	M_{A7}	983.8	C_{A7}	2.1×10^{-12}
M_{A}	232.69	$a_{AE}\,(\mathrm{m})$	5×10^{-3}	M_{A8}	153.5	C_{A8}	2.1×10^{-12}
C_{AF}	2×10^{-13}	$L_{AE}\,(\mathrm{m})$	1×10^{-3}	R_{A5}	5.06×10^{7}	R_{A7}	3.11×10^{7}
C_{AB}	3.5×10^{-13}						

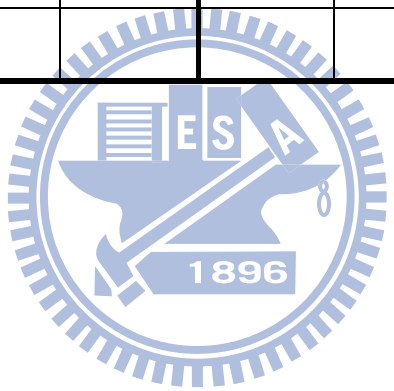
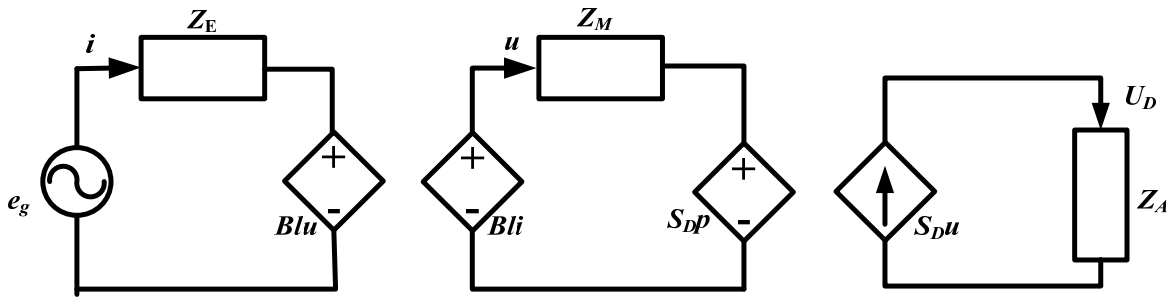
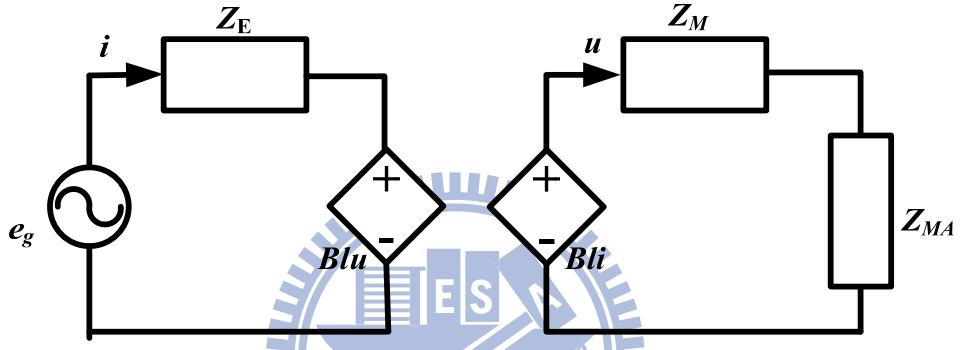


Table 3 Parameters of the optimized design versus the original non-optimized design.

	Original design(1)	Optimal design(2)	(2)/(1) %
$a_{ST}\,(\mathrm{m})$	1.5×10^{-3}	2×10^{-4}	13.33%
$V_{AF}\,(\mathrm{m}^3)$	2.83×10^{-8}	2.8×10^{-8}	98.94%
$L_{ST}\,(\mathrm{m})$	3×10^{-3}	2.9×10^{-3}	96.67%
$V_{AB}\,(\mathrm{m}^3)$	4.85×10^{-8}	6×10^{-8}	123.71%



(a)



(b)

Figure 1. (a) Electro-mechano-acoustical analogous circuit of loudspeaker. (b) Same circuit with acoustical impedance reflecting to mechanical system.

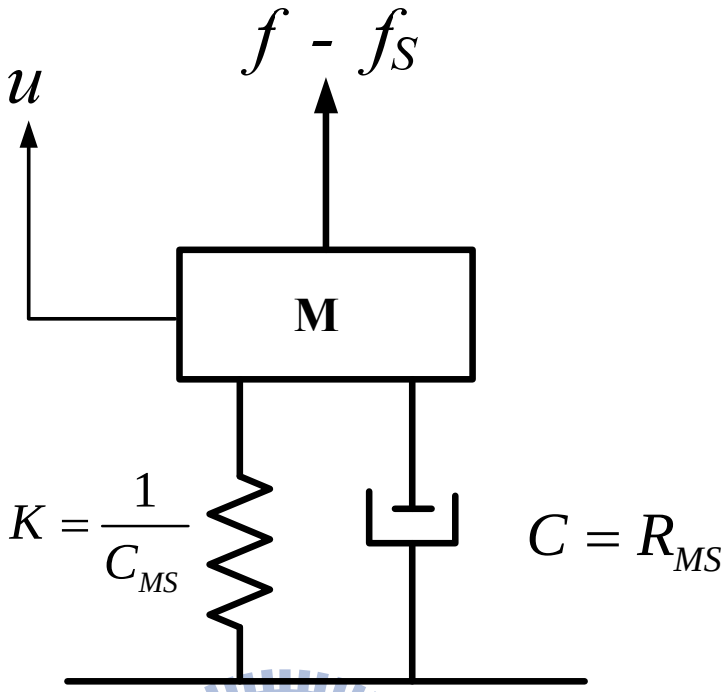
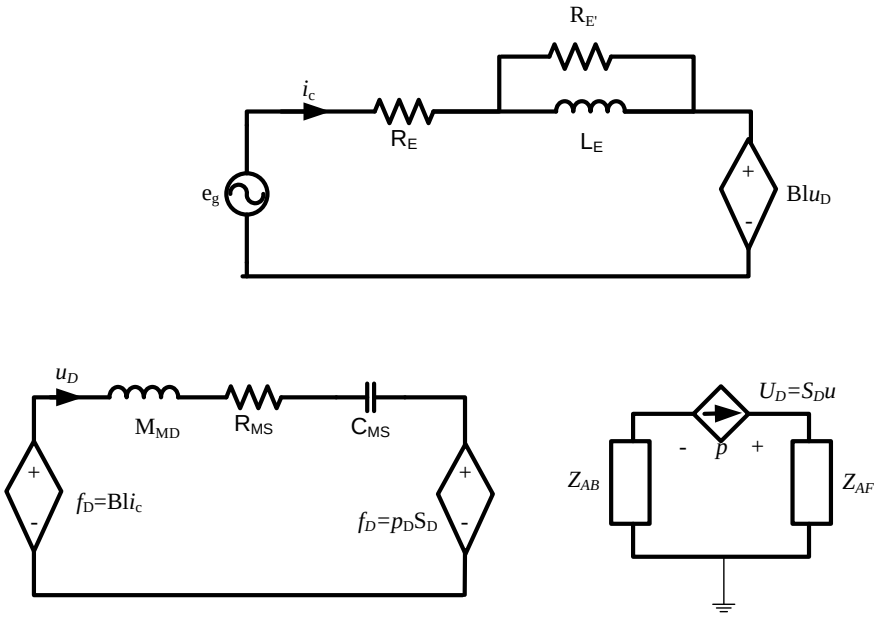
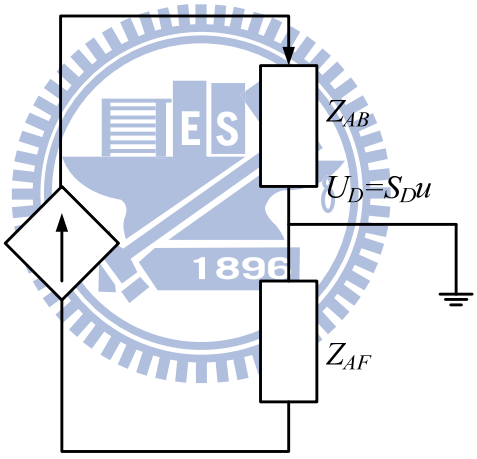


Figure 2. The mechanical system of loudspeaker. (M is diaphragm and voice coil mass, k is stiffness of suspension, C is damping factor)

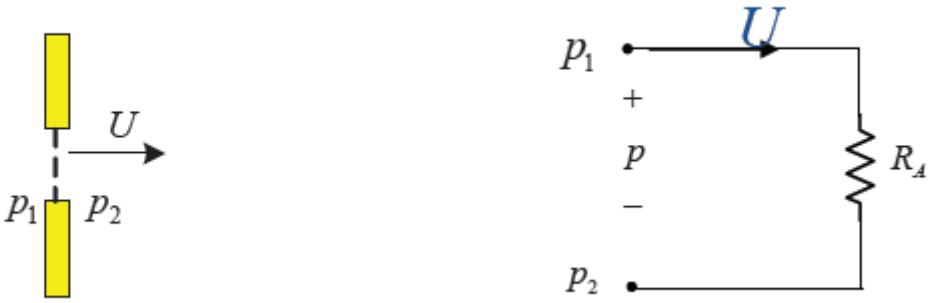


(a)



(b)

Figure 3. (a) Detailed Electro-mechanical-acoustical analogous circuit of loudspeaker. (b) Another form of acoustic system.



(a) (b)

Figure 4. (a) An acoustic resistance consisting of a fine mesh screen.
(b) Analogous circuit.

(a) (b)

Figure 5. (a) Closed volume of air that acts as acoustic compliance.
(b) Analogous circuit.

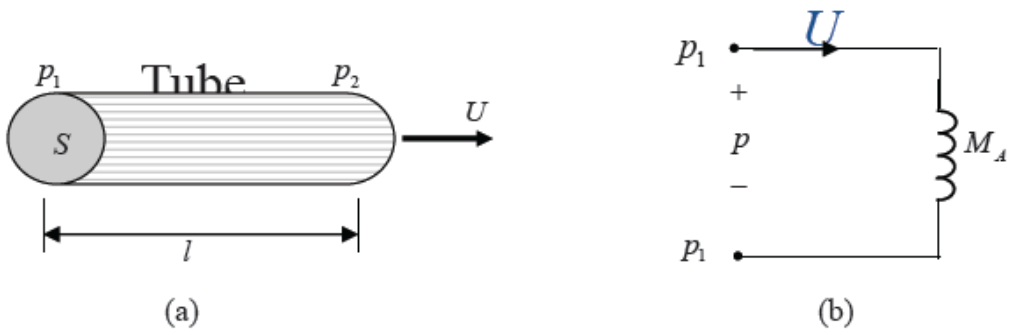


Figure 6. (a) Cylindrical tube of air which behaves as acoustic mass.
(b) Analogous circuit.

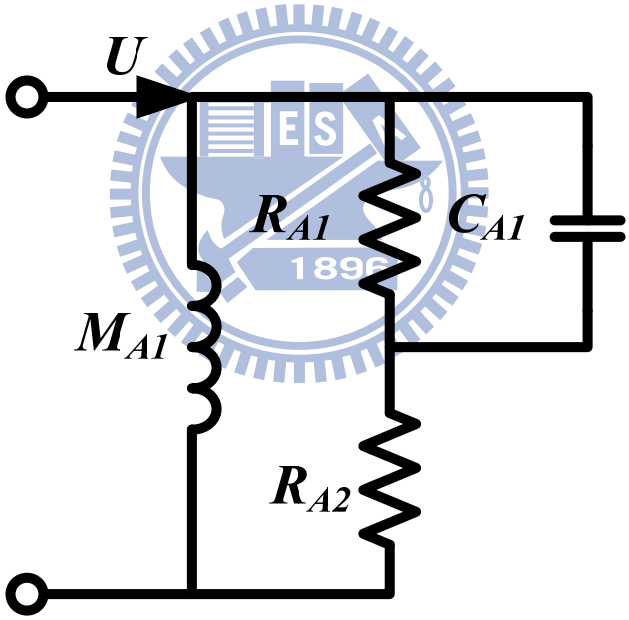


Figure 7. Analogous circuit for radiation impedance on a piston in a infinite baffle.
Analogous circuit for radiation impedance on a piston in a tube.

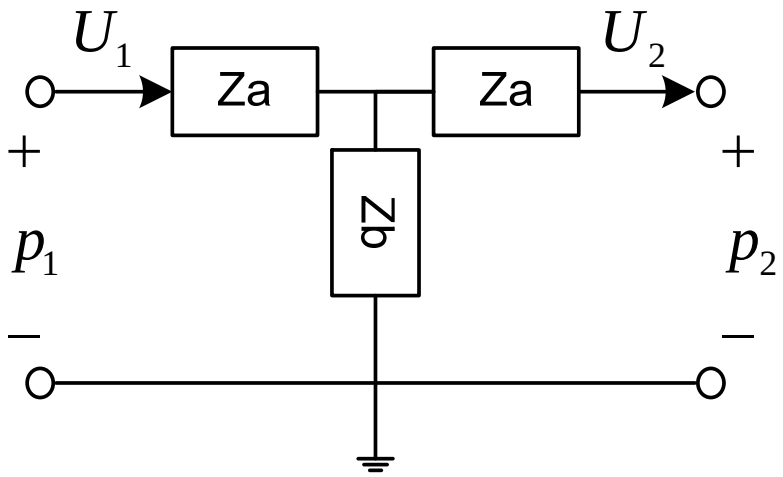


Fig. 8. T-circuit of transmission line

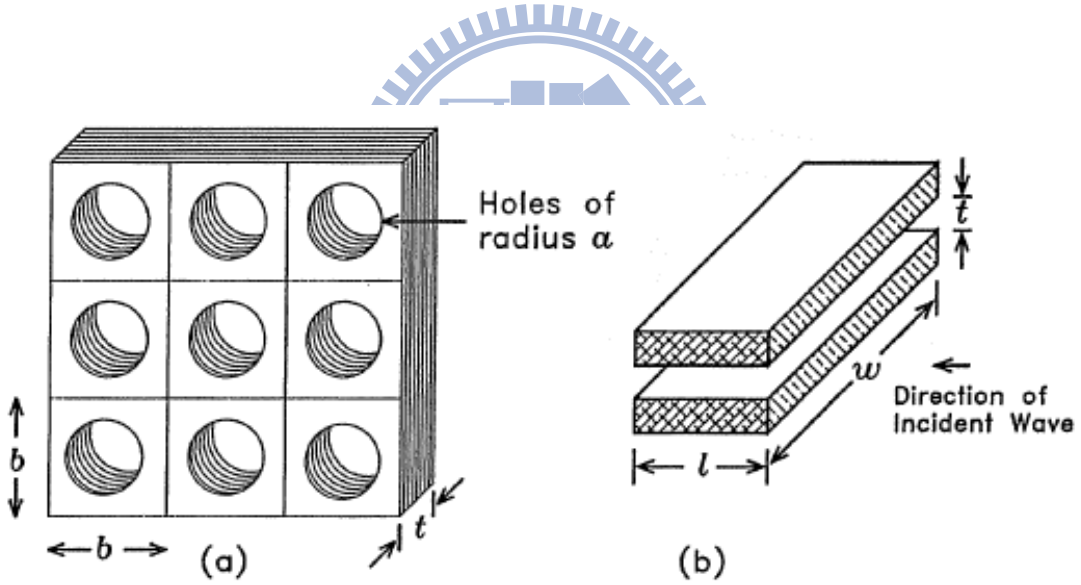


Fig. 9. (a) Perforated sheet of thickness t having holes of radius a spaced a distance (b) Geometry of the narrow slit.

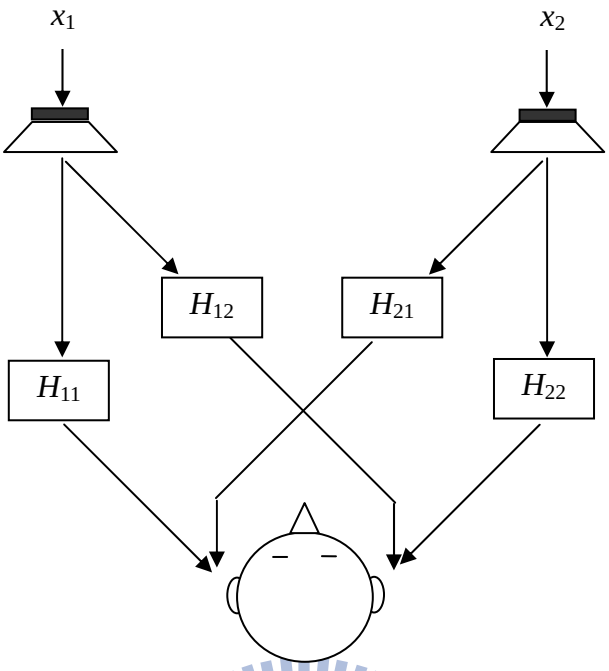


Fig. 10 Schematic diagram showing an audio reproduction system using two-channel stereo loudspeakers. Acoustic transfer functions between the loudspeakers and the listener's ears are indicated in the figure.

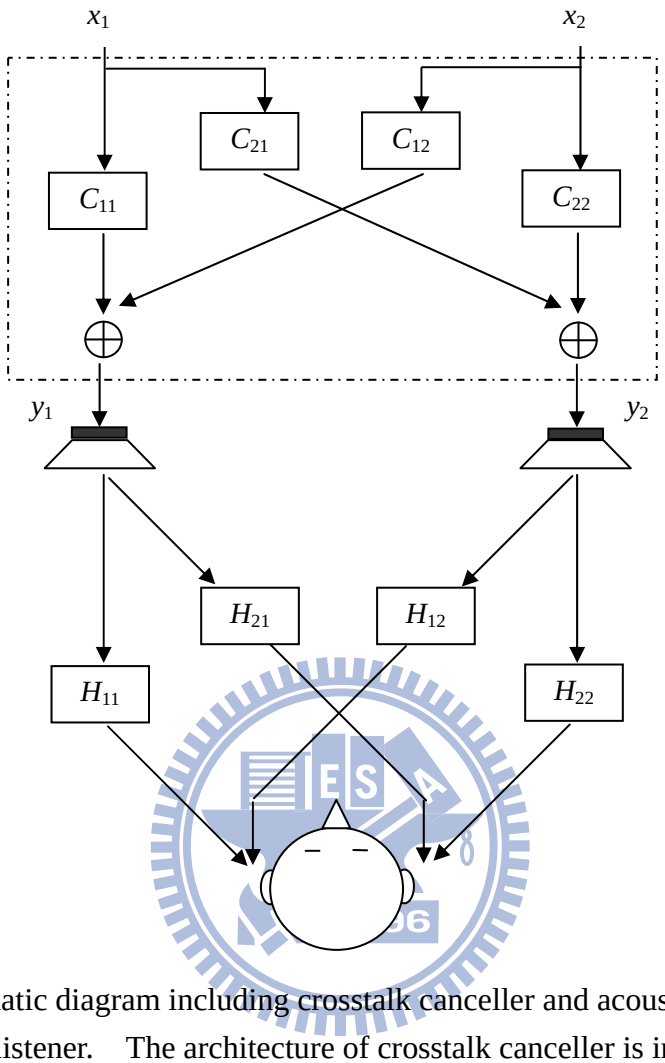


Fig. 11 Schematic diagram including crosstalk canceller and acoustic transfer functions to the listener. The architecture of crosstalk canceller is indicated in the dotted line.

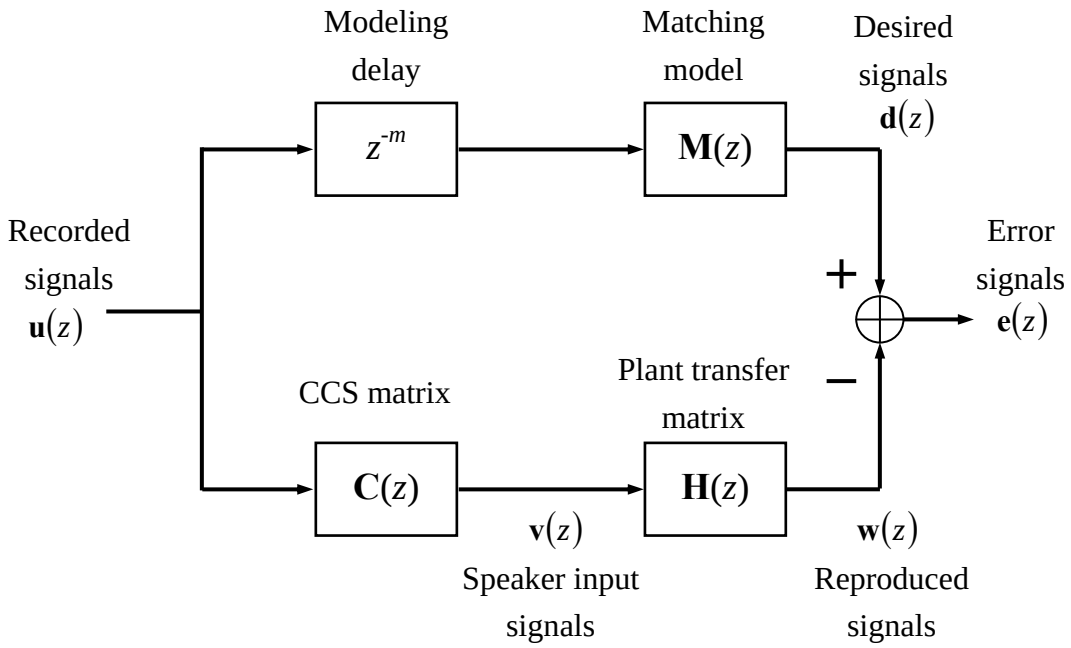


Fig. 12 The discrete time inverse filtering problem in block diagram form.

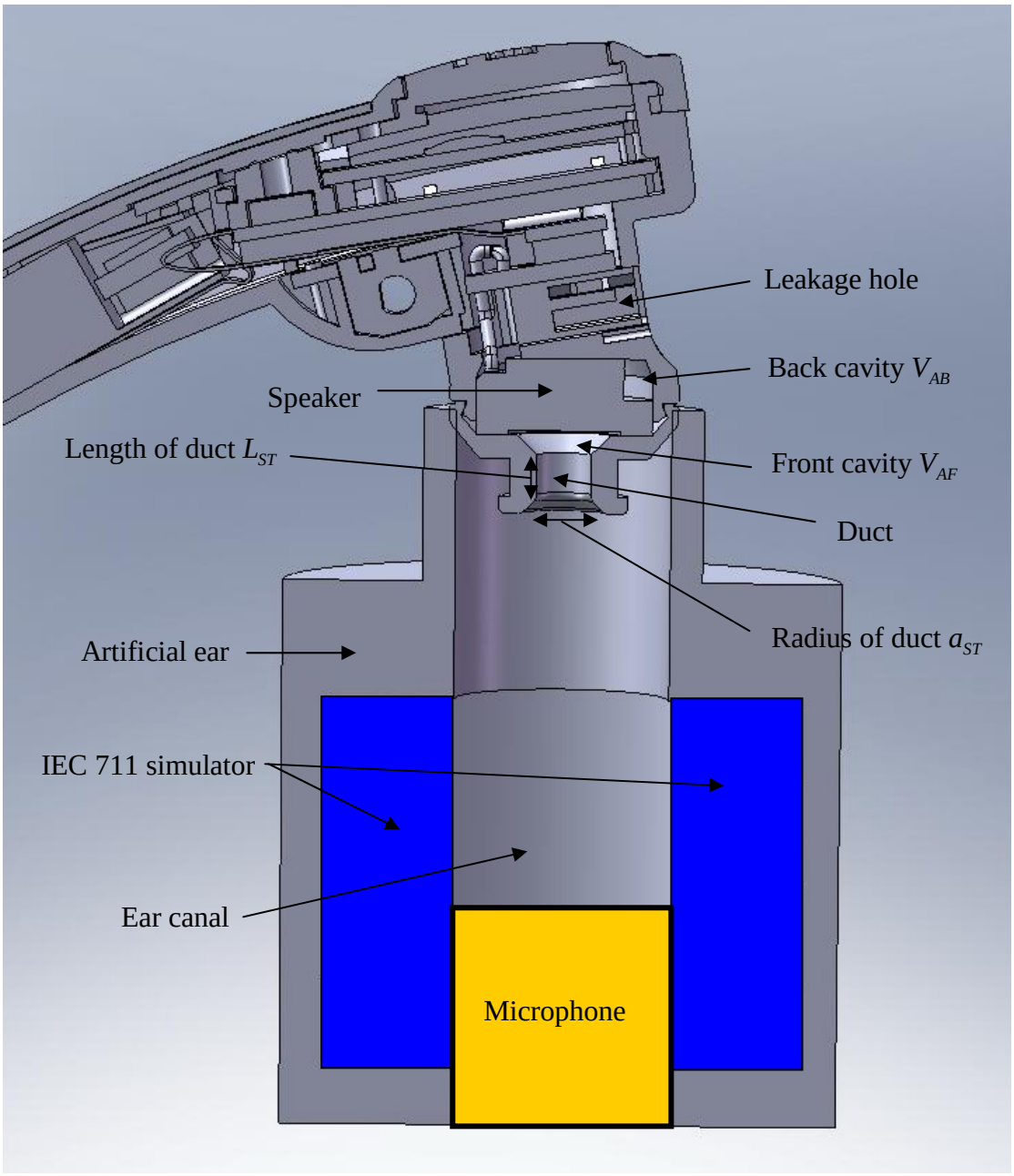


Fig. 13 The sectional drawing of earphone connecting with artificial ear.

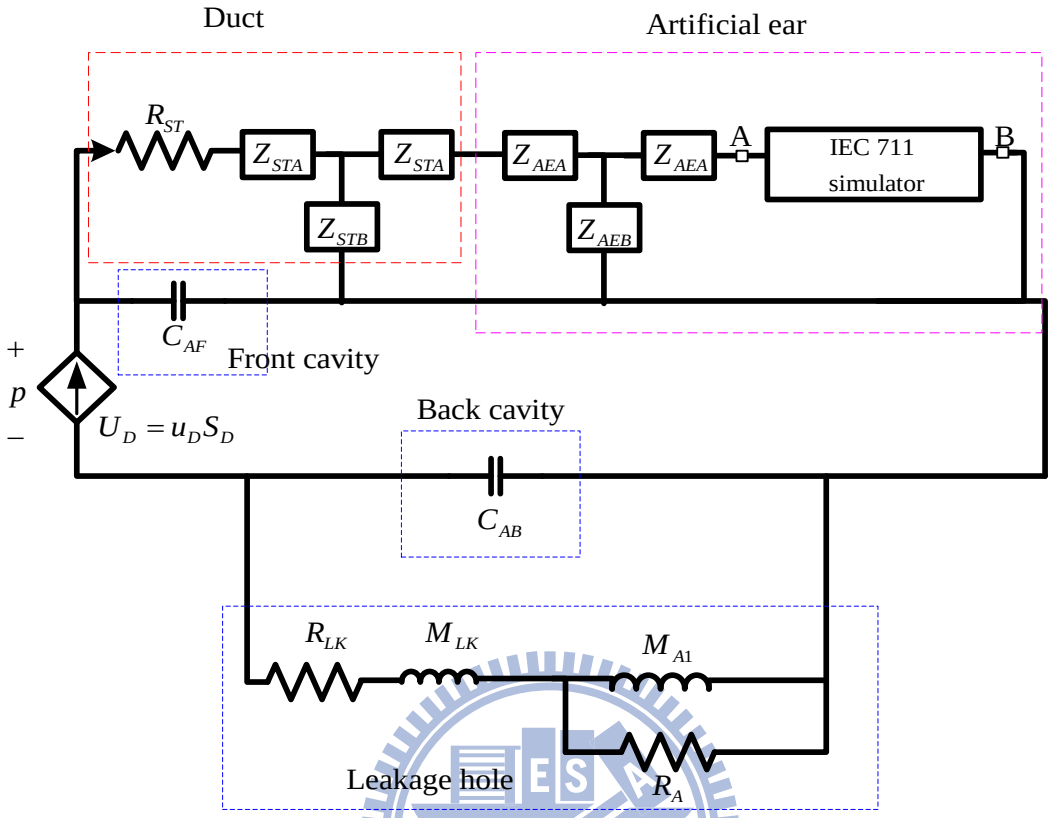

Duct
Artificial ear
R_{ST}
Z_{STA}
Z_{STA}
Z_{AEA}
Z_{AEA}
A
IEC 711
simulator
B
Z_{STB}
Z_{AEB}
C_{AF}
Front cavity
+
p
–
$U_D = u_D S_D$
Back cavity
C_{AB}
R_{LK}
M_{LK}
M_{A1}
Leakage hole
R_A

Fig. 14 (a)

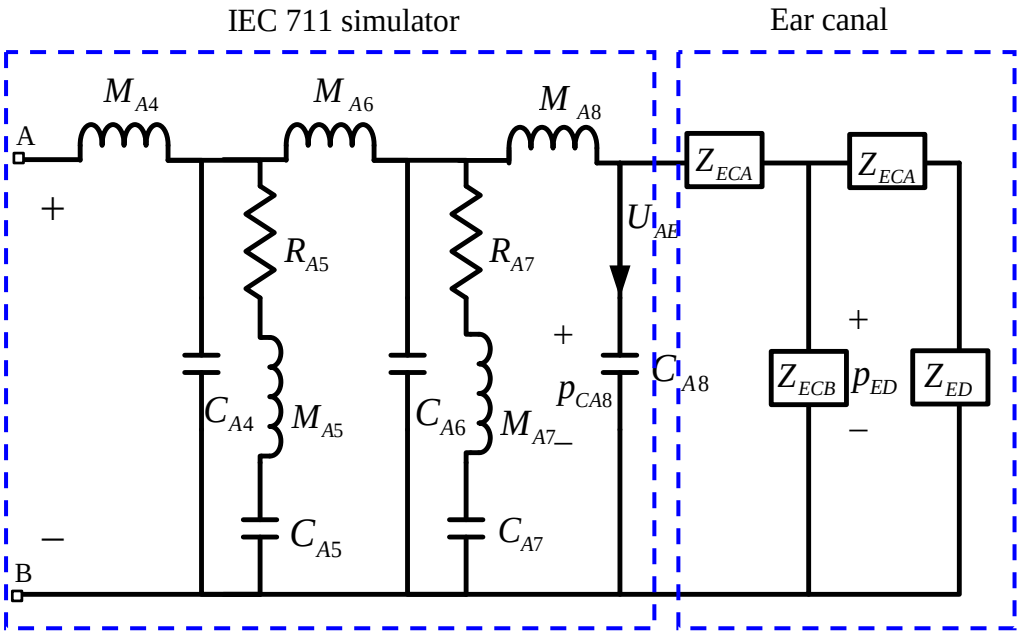


Fig. 14 (b)

Fig. 14. The Bluetooth earphone. (a) The analogous circuit of the acoustical system. (b) The analogous circuit of IEC 711 simulator connecting with the transmission line.

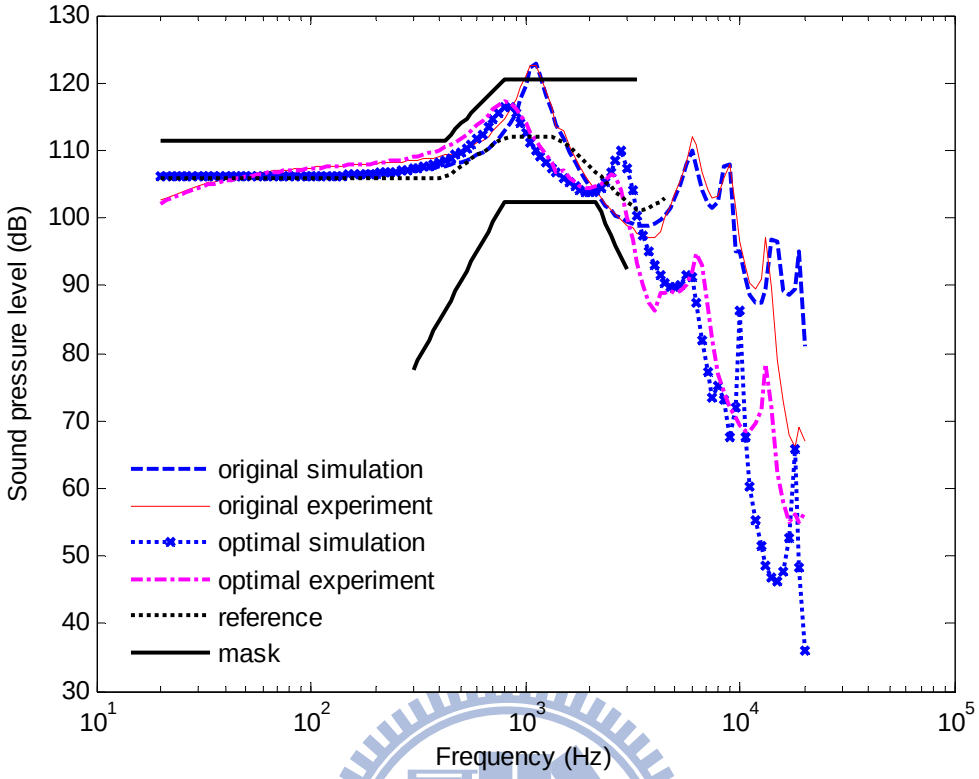


Fig. 15 The measured and simulated SPL responses for the optimal design and the original non-optimal design. The frequency response mask and a central reference curve are also sown in the figure.

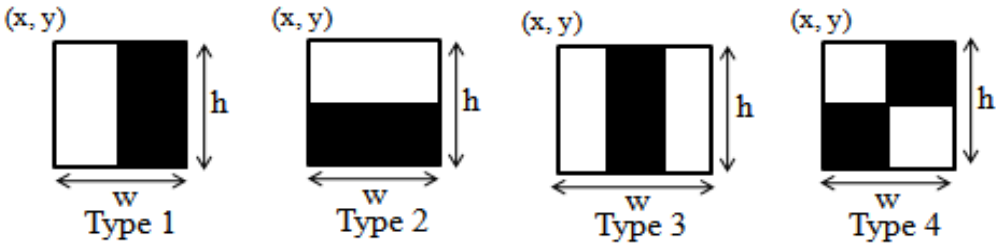


Fig. 16. Four types of Rectangular Haar-like features.

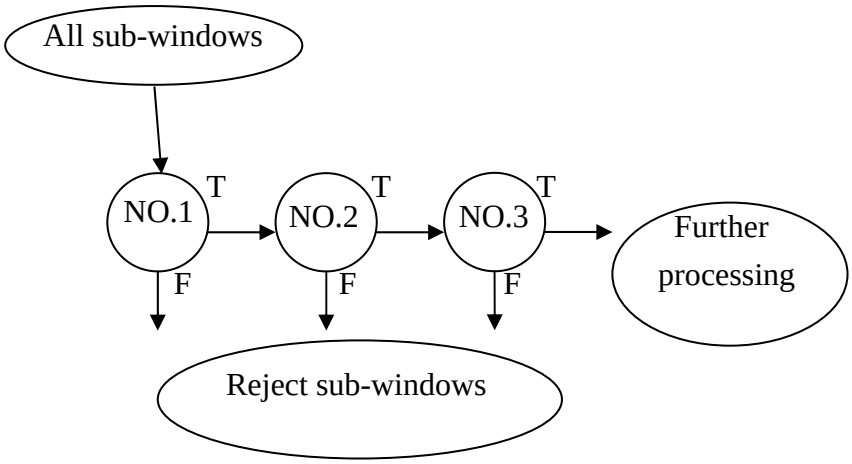


Fig. 17. Schematic depiction of a the detection cascade. A series of classifiers are applied to every sub-window. The initial classifier eliminates a large number of negative examples with very little processing. Subsequent layers eliminate additional negatives but require additional computation. After several stages of processing the number of sub-windows have been reduced radically. Further processing can take any form such as additional stages of the cascade (as in our detection system) or an alternative detection system.

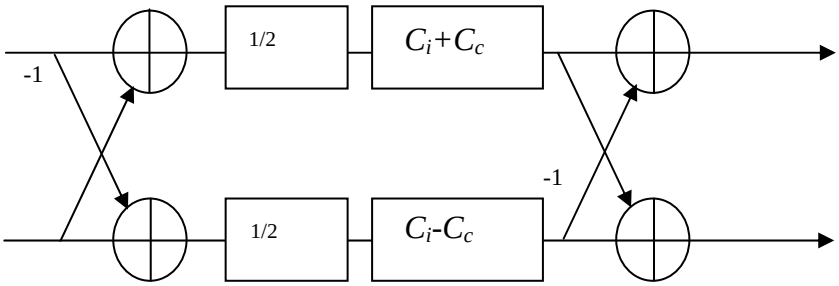


Fig. 18. The shuffler filter structure.

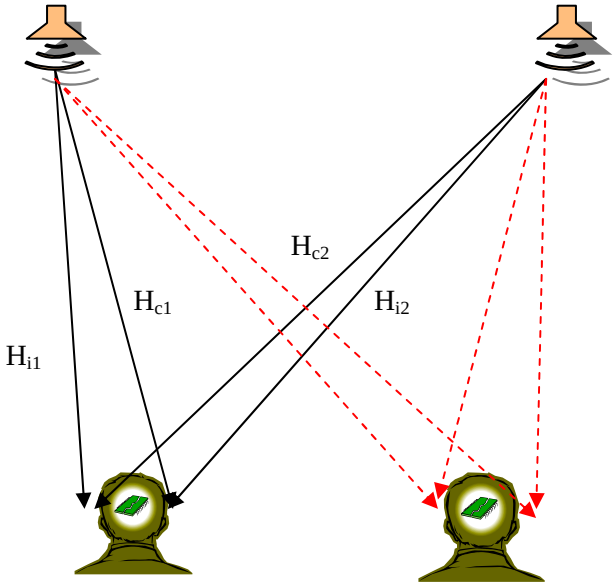


Fig. 19. The two listener positions are symmetric with each other, where H_{i1}, H_{c1}, H_{i2}, and H_{c2} are the acoustic transfer functions between the loudspeakers and the listener's ears.

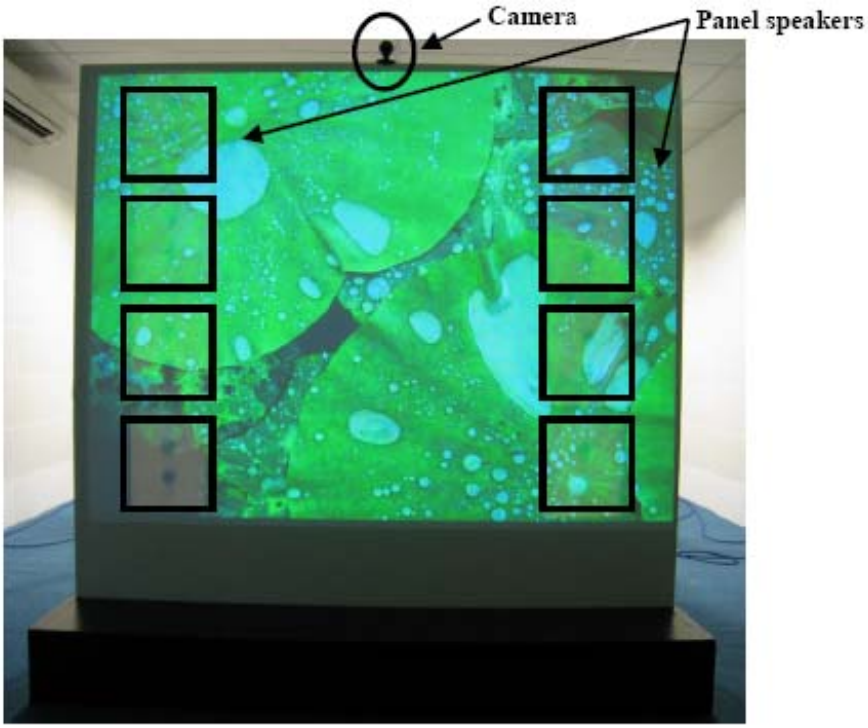
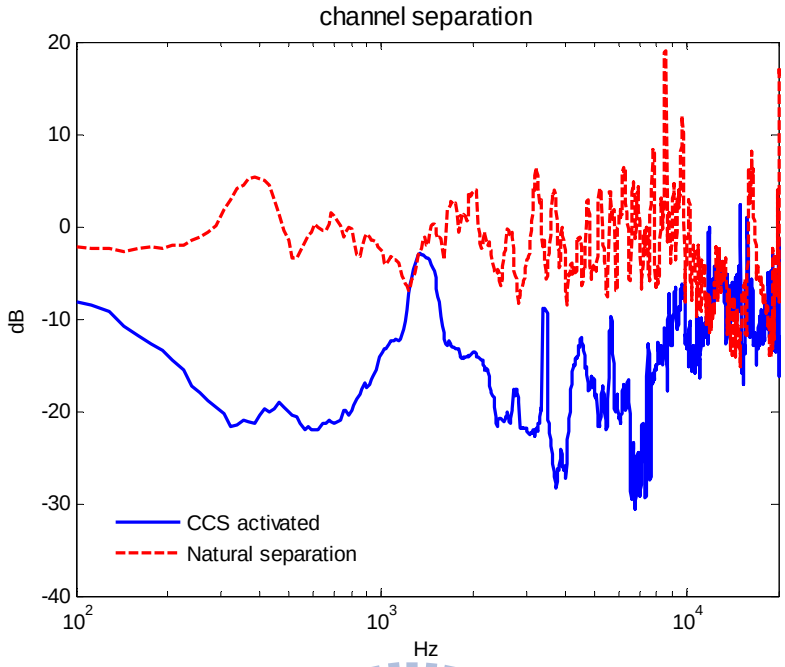
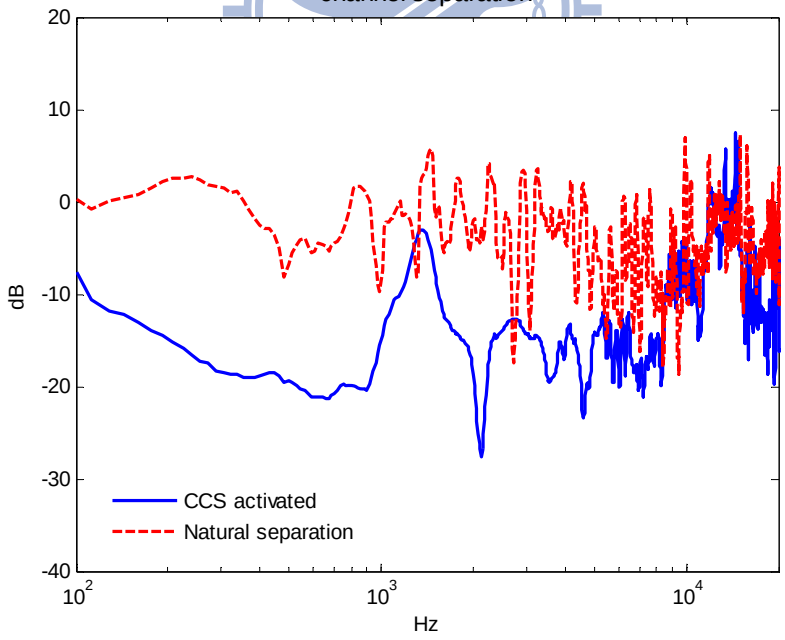


Fig. 20. The arrangement of the adaptive 3D-sound projection screen where the camera is in the middle of the projection screen. The stereo panel speaker array of the projection screen is constructed using PU foam panels. The size of each panel is $33.75\ \text{cm} \times 25\ \text{cm}$**.**

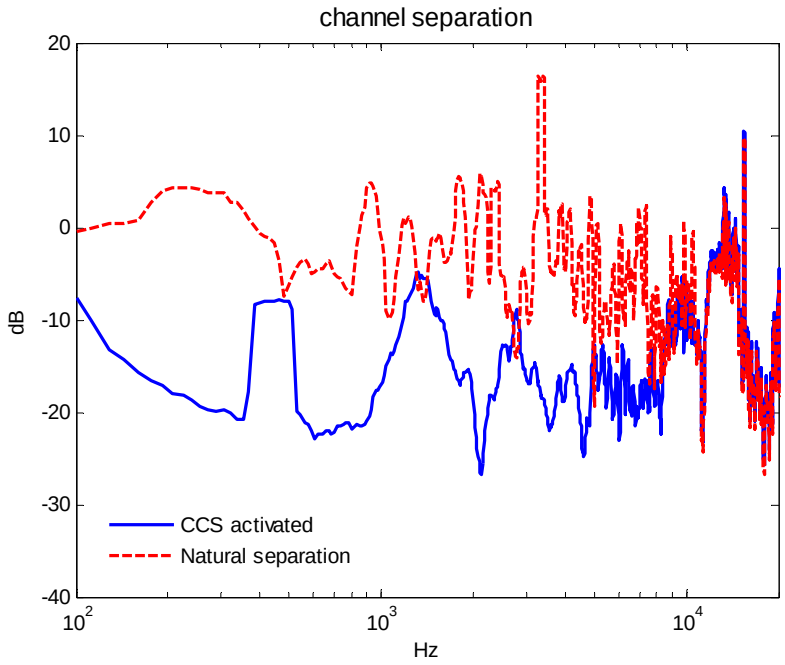
Fig. 21. The GUI of the head tracking system. The system can detect the face and send the coordinate to the beam steering system.


channel separation
20
10
0
-10
-20
-30
-40
dB
10^2
10^3
10^4
Hz
CCS activated
Natural separation

(a)


channel separation
20
10
0
-10
-20
-30
-40
dB
10^2
10^3
10^4
Hz
CCS activated
Natural separation

(b)



(c)

Fig. 22. The channel separations. (a) The channel separation when the dummy is at the centerline. (b) The channel separation when the dummy moves rightward 7 cm. (c) The channel separation when the dummy moves rightward 30 cm.