聲控音樂盒

蔡文宗
朝陽科技大學資訊與通訊系
助理教授
azongtsai@cyut.edu.tw

施又郡
朝陽科技大學資訊與通訊系
學生
asd8730873@yahoo.com.tw

摘要

科技的日新月異，人們已不再滿足於傳統的操作方式，進而追求更方便的方法，所以聲音辨識這項技術也越來越受到人們的重視。因此本論文以 FPGA 實作聲音辨識系統，並結合音樂播放器，使用者可以直接拍手控制音樂盒，無須透過其他任何的遙控器模組，此一做法減少了遙控器的實現成本，並同時增加了操作的便利性。

關鍵詞：現場可編輯邏輯閘陣列；聲音辨識; 特徵值; 梅爾倒頻譜係數

Abstract

Along with technology advances, people are no longer being satisfied with traditional operation methods for consumer electronics, and then pursue a more convenient way of control. Recently, voice recognition technology are capturing more and more people's attentions. As such, this paper implemented a voice recognition system for the control of a music player based on a FPGA embedded system. Users can just clap to control music box, without using any device of remote control. Our proposed approach reduces the cost of remoted-control device and increases the operation convenience of the implemented music player.

Keywords: Field-Programmable Gate Array; Signal Processing; Eigenvalue; Mel-Scale Frequency Cepstral Coefficients

1. 前言

當看報告看的很煩悶時，如果有音樂的陪伴總能舒解心煩意亂，而現今的音響大多搭配遙控器來使用，所以為了聽音樂，你就必須先找到遙控器才能聽的到音樂，但遙控器總是會因為各種原因而找不到，甚至有時找到了遙控器，卻發現電池沒電了。想像如果音響能支援聲音控制是不是就能除去這些麻煩，想聽音樂時只要拍個手音響自動撥放音樂，讓使用者不會因為繁雜的手續而降低聽音樂的樂趣。

近幾年嵌入式系統以非常快的速度取代了傳統單一功能的電路，嵌入式系統在生活中幾乎隨處可見。其架構則主要圍繞 32 位元 RISC 處理器，多核心架構、整合週邊以及可配置處理等方面的發展，以快速適應不斷變化的應用要求。對嵌入式 CPU 核心市場來說，近年來興起的可重規劃 (Re-Configurable)現場可編輯邏輯閘陣列(Field-Programmable Gate Array；FPGA)[1]軟核心技術，如 ALTERA NIOS II [2]，已逐漸受到市場的關注。相對於傳統單一功能的電路來說，可配置與規畫之軟式核心能彈性選擇所需的軟硬體功能，以達成系統成本與效能間的最佳化，且其所需時間和成本也降低相當多。另外 32 位元的嵌入式處理器則可利用 C/C++來進行應用程式開發，例如 NIOS II Software Build Tools for Eclipse [2]。

隨著時間的推進，聲音辨識已成為現今發展的趨勢。因此本論文以「FPGA 開發板」發展「聲音辨識系統」結合音樂播放器完成「聲控音樂盒」。且該系統未來也能移植成為冷氣機、電冰箱、洗衣機、電風扇、與抽風機控制系統的關鍵技術。

2. 背景

「沒有音樂，生活將是一種錯誤」- 尼采」

不論在讀書、做報告、玩遊戲、或是清晨剛睡醒時，如果有音樂的陪伴，總是會讓人心情愉快許多。此時想要聽到音樂，就必須先找到遙控器，或者起身將音響打開。假使音響有支援聲音控制的話，就能省下這些麻煩。所以我們結合了聲音辨識與音樂播放，運用科技使生活更便利。

現今市面上之聲控作品，大多都是控制電燈或是 LED 燈，例如聲控檯燈或是 LED 流水燈，在音響操作方面還是使用傳統控制方法(操作按鍵或使用遙控器)，但這些操作方式皆需移動身體，因此產生了使用上的衝突，減損聽音樂時的愉悅感受。

2.1 相關產品與技術

傳統的音響(如圖 1)在操作方面，還是使用傳統控制方法(操作按鍵或使用遙控器)，但這些操作方式皆需移動身體，因此產生了使用上的衝突。而為了收到遙控器之訊號(紅外線或是藍芽)，還必須在音響裡加裝接收訊號的模組，很明顯的此一音響需要花費額外的遙控器模組故額外增加整個系統之成本。

圖 1 傳統音響

2.2 運用科技帶來更多生活的便利性

相較於傳統遙控或按鍵式開關，而是使用聲音感測(如圖 2)來取代實體接觸，讓使用者不必再為了尋找遙控器或是不願移動身體去切換開關而感到困擾。在想聽音樂時，只要拍個手，音響立即撥放，既簡單又方便。

我們希望透過利用現有元件之功能實現聲音控制之功能達到降低成本之目標，經過我們多次的實測發現由於音響為了能達到高音質，在音響內部會有一顆音訊晶片，因此將聲控功能與此音訊晶片結合，就能達到降低成本之目的。

圖 2 聲控示意圖

3. 方法與架構

本章節說明我們所使用之聲音辨識系統與相關之 FPGA 系統晶片整合工作。

3.1 實現方法

本論文主要使用 Altera DE2-115 可程式化嵌入式系統開發板進行聲音辨識系統程式的撰寫，且搭配開發板本身的麥克風接收器、A/D 轉換模組等相關硬體設備，來構成聲音辨識系統的基本架構。辨識方面先做聲音特徵參數的擷取，使用「梅爾倒頻譜係數」。「梅爾倒頻譜係數」是一種利用聲音在頻率上的變化，以求取聲音在頻域上的特徵值，以作為語音辨識之用。

3.2 系統晶片功能方塊圖

如圖 3 所示，本論文所實現之聲音辨識系統經由音頻解碼晶片整合至 FPGA 系統晶片之中。其他之功能模組尚包含嵌入式微處理器，記憶體控制器，時脈產生器等等。

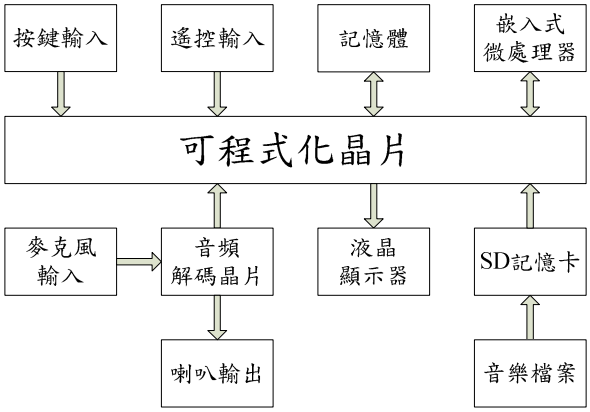

圖 3 系統晶片功能方塊圖

3.3 嵌入式系統開發板

如圖 4 所示，「聲控音樂盒」所使用的嵌入式系統開發板為 Altera DE2-115。為一個整合式的 FPGA 架構，包含軟式核心 32 位元的嵌入式處理器(NIOS II)以及其他硬式核心的硬體加速 IP，可提高設計的重複使用性。預估此類型元件在今後 10 年中將會得到廣泛應用，為系統設計人員提供更多的選擇。

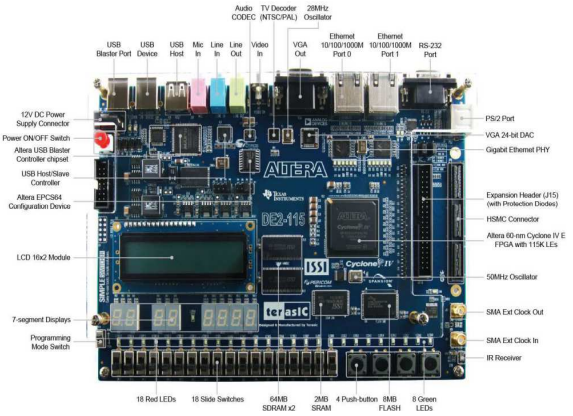

圖 4 Altera DE2-115 嵌入式系統開發板

3.4 聲音辨識系統

本論文的參數擷取方法使用「梅爾倒頻譜係數」(Mel-Frequency Cepstral Coefficients，簡稱 MFCC）[3][4]，MFCC 常被廣泛的使用在特徵參數擷取[5]。每一個音框基本的聲音特徵就有 13 維，包含了 1 個對數能量和 12 個倒頻譜參數。13 維的梅爾倒頻譜特徵是由 20 個梅爾頻譜上濾波器組的輸出經餘弦轉換求得，以下為梅爾倒頻譜參數求取過程：

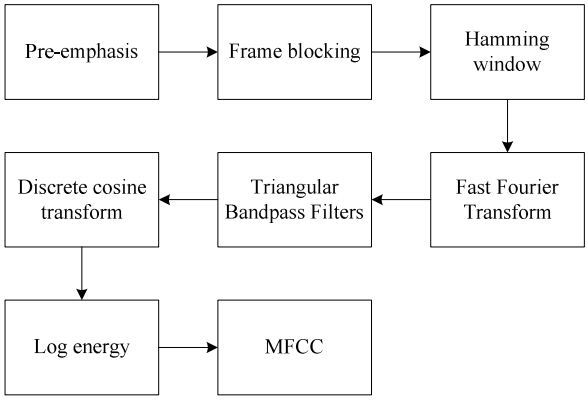

圖 5 MFCC 流程圖

3.5 梅爾倒頻譜係數(MFCC)

聲音訊號在提取特徵參數之前，為了使訊號能夠正確得到聲音訊號的特性，因此會先將訊號做前處理(Preprocess)，其流程包含預強調(Pre-Emphasis)、音框化(Frame Blocking)、加窗(Hamming Window)、接著是取特徵參數(Feature Extraction)以下將針對各項做詳細說明。

Pre-Emphasis：將聲音訊號通過一個高通濾波器，目的是要突顯在高頻的共振峰。運算式如下:

$$s_2(n) = s(n) - a * s(n-1) \quad (1)$$

其中 a 為一介於 0.9~1.0 之間的值，s(n)為原始時域訊號。

Frame Blocking：由於聲音信號是快速變化的，因此通常是將聲音訊號以 N 個取樣點為單位切割成許多小塊的連續訊號集合而這些小塊稱為音框(Frame) ，如圖 6，利用短時距處理(Short-Time Processing)的概念，使得聲音訊號更易於處理。

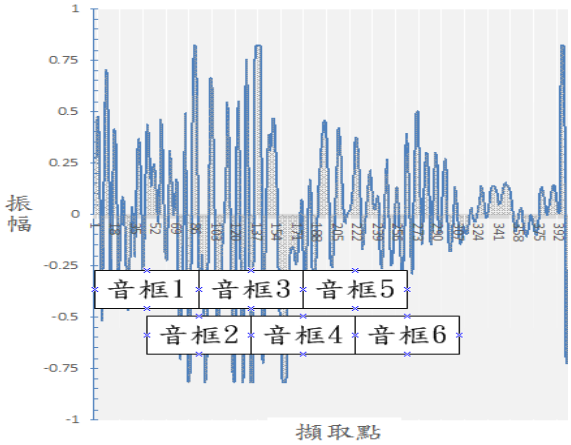

圖 6 音框取樣示意圖

為了避免相鄰兩音框的變化過大，因此會讓相鄰音框之間有一段重疊區域，通常是 N 的 1/2 或 1/3。以此方式重覆直到訊號結束，便能得到一序列的音框。

Hamming Window：每個聲音訊號通常要與一個平滑的窗函数相乘，讓音框兩端平滑地衰減到零，這樣可以降低傅利葉轉換後兩邊的訊號，取得更高質量的頻譜。運算式如下[3]：

$$W(n,\alpha) = (1-\alpha) - \alpha\cos\left(\frac{2\pi n}{(N-1)}\right),\ 0 \le n \le N-1,\ \alpha = 0.46 \quad (2)$$

Fast Fourier Transform：由於訊號在時域（Time Domain）上的變化通常很難看出訊號的特性，所以通常將它轉換成頻域（Frequency Domain）上的能量分佈來觀察。

Triangular Bandpass Filters：將能量頻譜能量乘以一組 20 個三角帶通濾波器，求得每一個濾波器輸出的對數能量（Log Energy）。必須注意的是：這 20 個三角帶通濾波器在「梅爾頻率」（Mel Frequency）上是平均分佈的，而梅爾頻率和一般頻率 f 的關係式如下：

$$mel(f) = 2595 * \log_{10}(1 + f/700) \quad (3)$$

Discrete Cosine Transform：將求出的對數能量代入離散餘弦轉換(Discrete Cosine Transformation; DCT)成為梅爾倒頻譜係數。

4. 成果展示

本章節說明我們所使用之聲控音樂盒系統流程，其中包含聲音辨識系統執行數據，系統流程分述於下列章節之中。

4.1 系統運行數據

由於音框的擷取點會影響系統最終的結果輸出，當擷取點較大時，所需的計算量也會相對減少，但對於訊號特性改變的情形也將較難以精確呈現，使得較不易觀測到聲音訊號變化的特性。而擷取點較小時，在分析時會因為使用的點數變少，使得結果易受到訊號突然變化的影響，較不具代表性，計算量也會變大。故我們在分析聲音時，通常以「短時距處理」（Short-Time Processing）為主，因為音訊在短時間內是相對穩定的，所以將音框取 64、128、256 與 512 點。

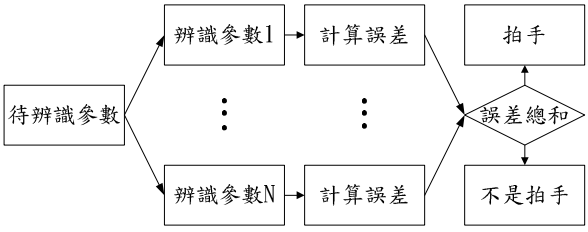

圖 7 辨識流程圖

如圖 7 所示，當系統分析出特徵值時，會與多個辨識參數作誤差值運算，以其結果分析是否為拍手，辨識參數之個數的由最少到辨識效能最佳為止，實驗結果又分為拍手與彈手，如表 1、2 所示。

表 1 彈手判斷錯誤率

彈手(錯誤率%)		音框擷取點			
		64點	128點	256點	512點
辨識個數	1	49	46	42	33
	2	39	22	20	16
	3	24	20	15	10
	4	20	19	15	11
	5	21	17	13	9
	6	12	9	5	2
	7	11	9	6	3

表 2 拍手判斷錯誤率

拍手(錯誤率%)		音框擷取點			
		64點	128點	256點	512點
辨識個數	1	73	60	58	53
	2	65	56	55	48
	3	66	63	40	26
	4	56	55	35	21
	5	33	24	22	11
	6	11	9	7	4
	7	13	12	7	4

在相同系統環境下，不同的辨識個數實驗結果如圖 8 所示，在辨識個數為 1 時，錯誤率為最高，而在辨識個數為 6 時，錯誤率為最低。

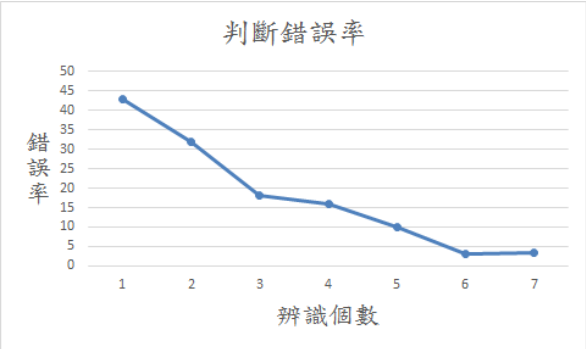

圖 8 彈手判斷錯誤率(辨識個數)

在相同系統環境下，不同的音框擷取點實驗結果如圖 9 所示，在音框擷取點為 64 時，錯誤率為最高，而在辨識個數為 512 時，錯誤率為最低。

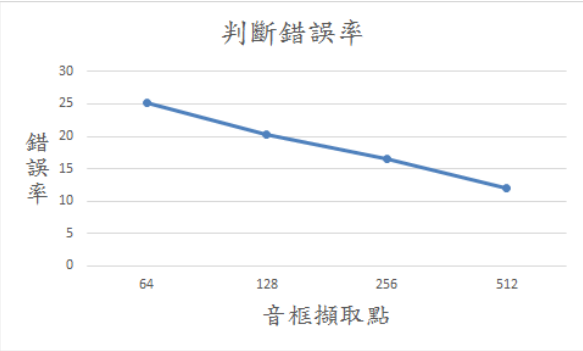

圖 9 彈手判斷錯誤率(音框擷取點)

4.2 整體系統流程與功能

如圖 10 所示，本系統包含一聲音辨識系統。聲音經由麥克風接取，將聲音訊號傳遞至 FPGA 嵌入式系統板，該系統板將接收到之聲音訊息經過取樣與 MFCC 運算處理後，再將結果做特徵值對比，系統依據對比結果做出相對應之操作。

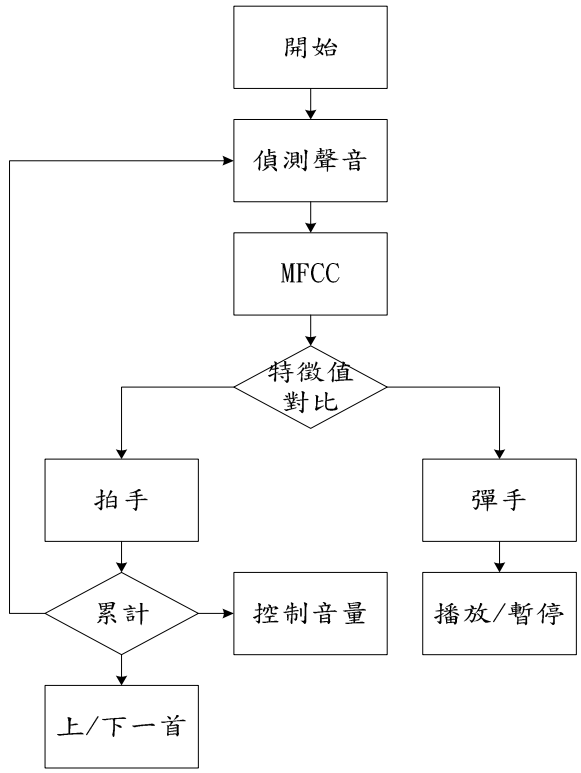

圖 10 系統整體流程圖

已完成之功能整理如下：

- ◈ 彈一次手：開始/結束 播放
- ◈ 拍一次手：下一首歌
- ◈ 拍二次手：上一首歌
- ◈ 連續快拍：音量往上調
- ◈ 連續慢拍：音量往下調

4.3 系統開發環境

軟體：Quartus II Web Edition 13.0.1.232、Nios II EDS 13.0.1.232

硬體：Altera DE2-115 FPGA 開發板、麥克風，擴音器

5. 結論

相較於傳統遙控或按鍵式開關，我們使用聲音感測來取代實體接觸，讓使用者不必再為了尋找遙控器或是不願移動身體去切換開關而感到困擾。且利用現有元件所開發之 FPGA 嵌入式系統，其功能所需的 FPGA 成本對於整體系統生產所佔的成本比例不大。使用 FPGA 所開發的產品具有另一絕對的優勢，即是當產品需要大量商品化時，同一設計能改以 ASIC 實現，可進一步有效降低其製造成本與產品功耗。

6. 致謝

感謝朝陽科技大學 104 年度校級重點特色計畫「智慧物聯網核心技術之建立-子計畫三-智慧家庭物聯網」之支持。感謝科技部計畫編號：MOST 104-2221-E-324 -001-MY2 與 MOST 105-2623-E-324-001-D 之支持。

參考文獻

[1] ALTERA FPGAs, http://www.altera.com/products/fpga.html

[2] NIOS II Processor: The World's Most Versatile Embedded Processor, http://www.altera.com/devices/processor/nios2/ni2-index.html

[3] Mel-Frequency Cepstral Coefficients, http://mirlab.org/jang/books/audiosignalprocessing/speechFeatureMfcc_chinese.asp?title=12-2%20MFCC

[4] Mel-Frequency Cepstral Coefficients, https://en.wikipedia.org/wiki/Mel-frequency_cepstrum

[5] 李健平，”語音辨認應用於 PDA 之作業控制研究”中原大學資訊工程學系碩士學位論文，2001。