聞聲制物

2020

目錄

1 作品主題說明

目前 5G 網絡技術的發展與 4G 網絡的全面覆蓋為物聯網產業打下了結實的技術基礎 [1]，但是目前的物聯網設備與使用者的交互方式都還停留在傳統的交互方式上面，通過智能手機輔助交互或物聯網設備强行安插顯示屏都是破壞使用者體驗和增加物聯網設備使用成本的行為 [5]。語音控制雖然可以解決這個問題，但是單純的語音控制無法對使用者的身份進行校驗，導致物聯網設備沒辦法辨識使用者的身份，容易造成物聯網設備管理上的問題。[2] 而目前市面上基於深度學習進行的聲紋辨識系統架設成本高、技術發展不成熟、對物聯網設備硬件要求門檻高和認證延時大也成為限制物聯網設備與語音交互相結合的問題。[3] 而我們此次研究結合了聲音的語音識別與基於分散式資訊系統的說話人確認演算法為物聯網設備量身定制了一套低硬件門檻、低架設成本、低網絡需求、快速響應、可商業化的權限管理系統。這個系統可以應用於各個領域的物聯網設備，使用者通過聲音進行身份認證，使用者說出自己設定的“通關密語”，就可使得物聯網設備認得使用者身份，可以在同一設備下給予不同使用者不同的權限等級，也可限定物聯網設備使用者的身份，真正有效的降低了物聯網設備的管理成本，為物聯網產業的發展提供了强有力的技術支持。

2 作品特色

2.1 功能特色

1. 用聲紋辨識的方式實現對物聯網設備的許可權管控，具有非接觸式遠距離的優點
2. 輸入端設備種類不受限制，只需要符合算力和硬體要求且能安裝軟體即可，不符合硬體要求的物聯網設備則可通過其他符合要求的設備進行管理
3. 對物聯網設備安裝硬體支援分級，自動組網，高階設備對低階設備實現控管
4. 整套系統可分散式可整合，根據使用環境的網路配給和資訊安全要求等因素可實現系統離線化，認證局域網化，滿足不同環境的使用需求
5. 系統無需大樣本訓練模型，擁有根據使用者即時輸入的資訊提高辨識準確率的自我學習演算法，演算法複雜度低，對伺服器要求低，使用成本低
6. 系統介面可開放提供使用者連接有身份確認需求的應用，以更好的提供客制化服務
7. 完全自主研究的聲紋確認演算法，識別率已達到 90% 以上，誤識率降到 10% 以下，具有自我學習功能，根據使用者驗證成功的次數逐漸達到更高的識別率

2.2 作品與市場相關產品差異

2.2.1 與“人臉識別”驗證方式的差異

人臉識別驗證技術需要終端設備帶有攝像頭，且用戶需以特定位置站在攝像頭前才能完成驗證工作，同時人臉識別資料庫需置於雲端伺服器，所有設備均需連入乙太網，對於某些特殊需求不便將資料置於雲端處理的公司來說沒辦法採用。同時人臉識別技術訓練成本高昂，這就體現在幾個大公司的行業壟斷上，不利於小企業實現技術自主化當地語系化和對技術進行進一步的創新應

用。我們的技術只需要部分物聯網設備帶有一定運算能力和麥克風揚聲器即可，不需要高昂的硬體費用，同時資料庫和比對伺服器皆可當地語系化，局域網化，減少公司的頻寬支出，保護資料安全。同時我們的系統根據適配單位的算力分佈情況可分散化也可集中化，各系統模組相對獨立又可自由組裝，在分散的同時又採用了 tls 的加密方式，保證了資料傳輸的安全性。同時我們的系統不需要進行訓練即可使用，成本低可塑性强，非常適合進行二次開發和應用。

2.2.2 與“智慧音箱”系統的差異

智慧音箱系統是指類似“小愛同學”等家用智慧音箱系統，這套系統主要應用場景為家庭娛樂，不需要進行特殊的許可權管理，且帶有對一般家庭物聯網設備的控制功能。但這套系統智慧滿足家庭需要，且輸入端（智慧音箱）往往只有一個，需在特定的範圍內說話才能實現控制功能。我們的系統則是針對物聯網管理而設計的，同時可為企業或公共場所的物聯網設備管理服務。我們通過聲紋辨識以達到物聯網設備控制許可權的管理，使得只有特定的使用者才能控制物聯網設備或獲取物聯網設備的管理許可權。我們的系統輸入端採用分散式的分佈，只需要算力和硬體要求符合的物聯網設備，安裝完軟體之後均可作為高階節點接入我們的系統，不需要額外的定制。我們的系統在確認了接入設備的網路狀況後可實現自動組網，全域控制，保證了輸入端的全面覆蓋，保證了物聯網設備的分級管控和便捷管理。

3 設計理念及架構説明

3.1 硬體架構

我們採用分散式的系統架構，用分散式的運算模式來幫助服務器減少網絡壓力與運算壓力。在我們的系統中終端設備分為兩個等級，第一個等級是有一定運算能力與硬件設備支持的終端設備（比如智能音箱、智能電視、智能手機等），此類型的設備需要擁有一定的運算能力[1]與收發聲音的設備。第二個等級是指可以接入網絡並通過網絡進行操作的一般物聯網設備（不需要其他額外的要求）。我們通過第一個等級的設備與使用者交互，發送信息給伺服器，接收來自伺服器的信息，並通過網絡控制第二個等級的設備，實現自身與第二個等級的物聯網設備權限管控。（圖 1）

3.1.1 物聯網終端設備

在第一個等級的物聯網設備上，我們將搭載兩個模塊的程式。第一個模塊是具有生成聲音密碼、語音識別功能及發送驗證請求與信息給伺服器的程式，實現物聯網終端採集使用者音頻信息與聲音特征，並發送聲紋辨識請求給伺服器。其發送的内容有作為“賬號信息”的驗證文字與提取聲音特征信息生成的聲音密碼文字。第二個模塊則是與使用者的語音互動與控制模塊，此模塊將運用語音合成技術實現與使用者的互動，並接受來自雲計算服務器的比對結果，根據比對結果執行相關的操作，發送指令給自身或由其控制的物聯網硬件，並在此過程中語音回復使用者需要重新講一次“操作秘語”或進行其他的提示。

[1]目前是用樹莓派為推薦算力配置

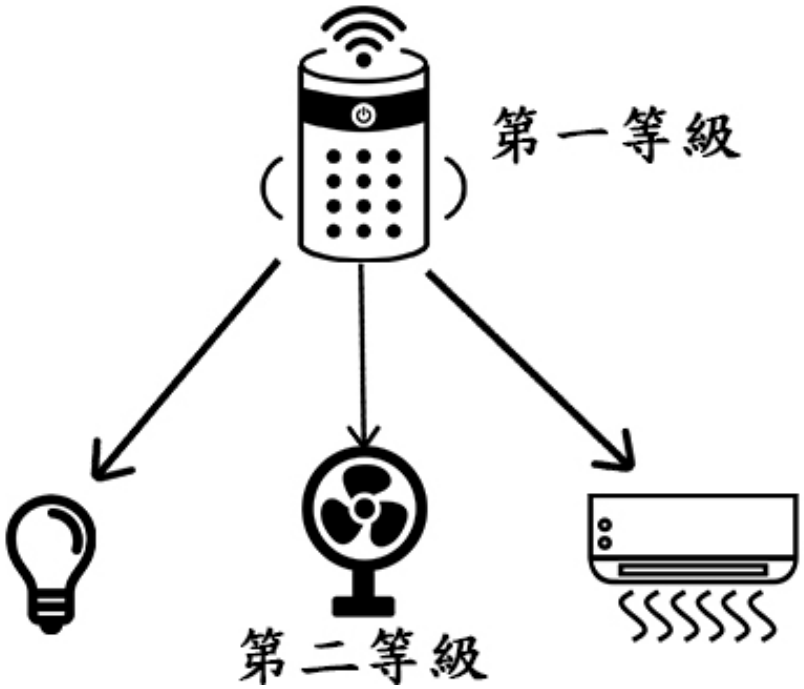

图 1: 終端分級

3.1.2 雲端計算服務器

我們位於雲端[2]的計算服務器則搭載了聲音密碼比對模塊，它會使用終端設備發送過來的驗證文字來向雲端數據庫發送查詢請求，查詢對應的聲音密碼。再將數據庫的聲音密碼與使用者說話生成的聲音密碼進行比對，若相似度較高[3]，則回傳驗證成功的信息和使用者的姓名或驗證文字，若相似度較低則回傳驗證失敗的信息。

3.1.3 雲端數據庫

我們的數據庫模塊也將架設在雲端服務上面，具有可拓展性和可轉移性，並視使用情況進行熱備份服務的架設。雲端服務器存取使用者的姓名、驗證文字、對應的聲音密碼等使用者信息，主要職責是提供對應的聲音密碼給雲端計算服務器進行比對的工作。雲端數據庫將可與使用者信息數據庫分開架設，保證第三方需要使用我們服務的時候擁有獨立自主的數據庫，對於數據庫的訪問與存儲可通過雲計算服務器或終端設備來完成。

3.2 軟體架構

我們的聲紋辨識技術主要為三大模塊，第一個是聲音密碼生成模塊，第二個是聲音密碼比對模塊，第三是語音識別模塊，其他子模塊則是進行訊息的傳遞和保存。其中我們的研究集中在聲音密碼生成模塊與聲音密碼比對模塊，而語音識別模塊我們運用較為成熟的第三方技術支持[4]。我們使用語音識別模塊生成的文字來確認使用者將要登錄的賬號信息，再通過說話人確認技術確認即將登錄的使用者為本人。(圖 3)

[2]可租用安全性高的雲服務提供商的服務器快速架設

[3]根據演算法設定閾值

[4]目前是採用百度雲計算提供的語音識別服務

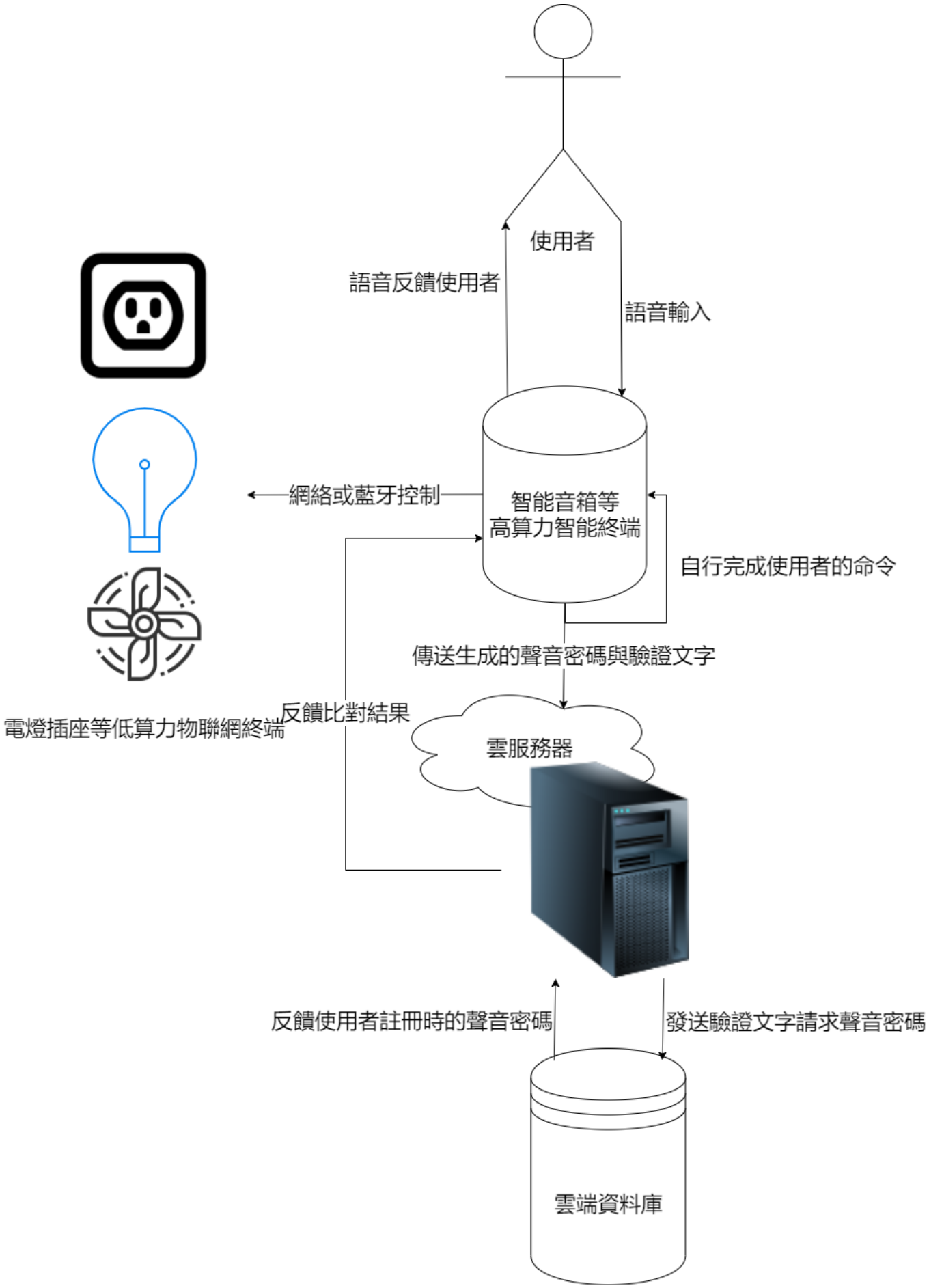
使用者
語音反饋使用者
語音輸入
網絡或藍牙控制
智能音箱等
高算力智能終端
自行完成使用者的命令
傳送生成的聲音密碼與驗證文字
反饋比對結果
電燈插座等低算力物聯網終端
雲服務器
反饋使用者註冊時的聲音密碼
發送驗證文字請求聲音密碼
雲端資料庫

图 2: 完整系統運作流程

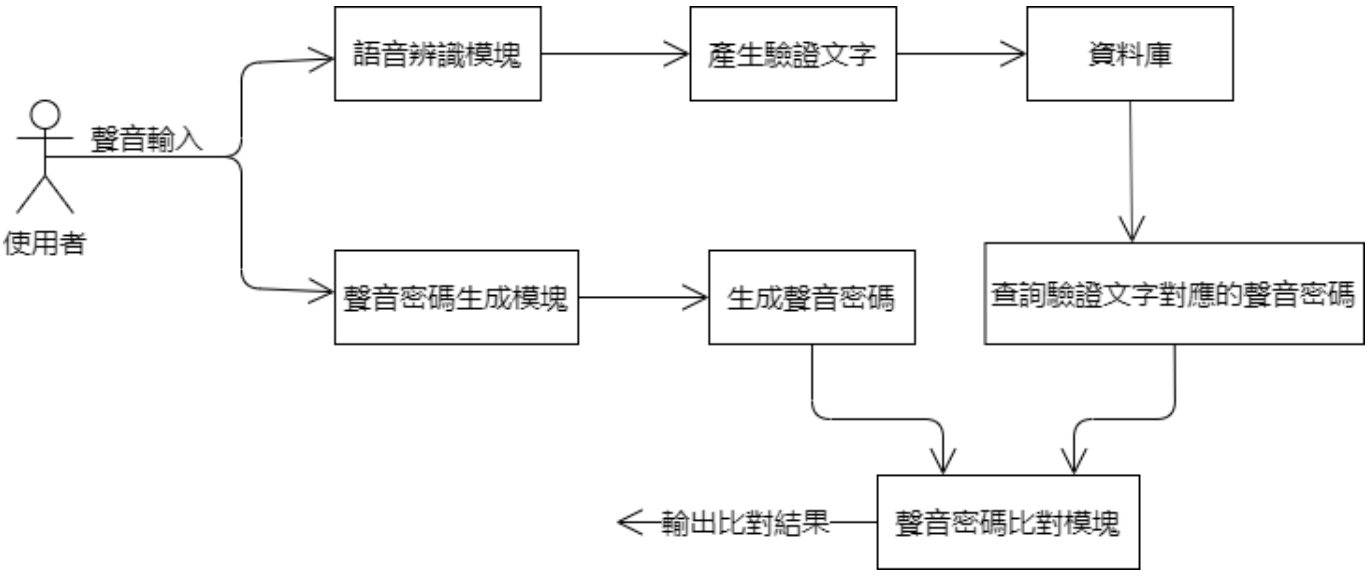

图 3: 聲紋辨識三大模塊運作關係

3.2.1 聲音密碼生成模塊

此程序模塊獲取由系統聲音輸入裝置輸入的聲音，通過傅裡葉轉換將聲音量化到 1024 個不同的頻率段上，並用輸入聲音與背景聲音[5]强度的比值來定義輸入聲音的强度，我們每秒鐘約取樣 110 次，即可得到一個不定長度的三維聲音數據模型。(圖 4) 根據我們的實驗結果[6]，我們採用前 256 頻率段上的聲音强度作為聲紋辨識的特征值取樣範圍，因為其幾乎包含了所有人類的聲音特征。我們通過判斷輸入聲音的强度來確定使用者輸入聲音的起始時間與終止時間，截取出聲音信息有效的部分。(起始點如圖 4中第 80 個聲音信息左右的位置) 然後我們對有效聲音的所有時刻中的聲音信息進行處理。處理方法為提取出此時刻聲音信息分佈在 256 個頻率段上的聲音强度極大值（圖 5)，並取前六大極大值的頻率記錄下來，作為此時刻的聲音密碼。有效聲音信息從起始時刻到終止時刻所有時刻的聲音密碼則為此段聲音信息的密碼。(圖 6) 聲音密碼進行處理後每時刻將有 18 個十進制數字，每秒鐘有約 110 個採樣時刻。

3.2.2 聲音密碼比對模塊

此程序模塊接收來自聲音密碼生成模塊生成的聲音密碼以及訪問並讀取數據庫中的密碼，將密碼信息進行切割後時間、頻率、音高三個維度上進行比較。在兩個時刻的聲音信息的相似性比較中，使用需要比較的兩個時刻各自的頻率與音高的維度上使用音高最高的三個頻率位置信息與對面所有的六個頻率位置信息進行比較，每個頻率位置信息將會尋找與自己最為接近的頻率位置，並計算與其的差值，並加總進差值和[7]。若差值和小於設定的閾值[8]，則說明這兩個時刻的聲音信息具有相似性，在後文中我們使用“頻率位置分析演算法”為名稱取代之。(公式 1)（圖 7）

$$sum_{distance} = \sum_{k=1}^{3} |x_k - y_{find}| + |y_k - x_{find}| \tag{1}$$

[5]使用者不說話時系統會自動計算背景聲音

[6]聲音信息 1.0

[7]每次兩時刻進行比較時都會有此次比較的差值和

[8]實驗時是以 20 為閾值

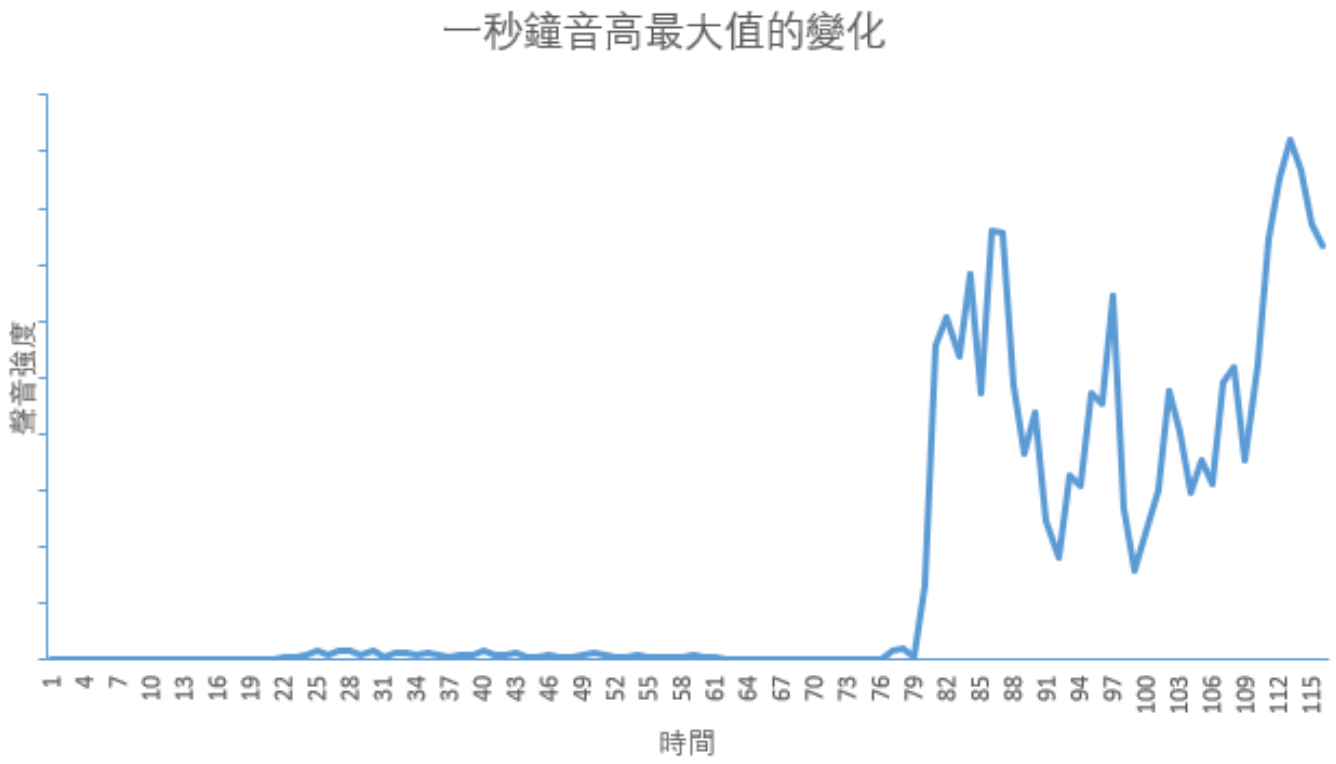

图 4: 一秒鐘最高聲音强度變化

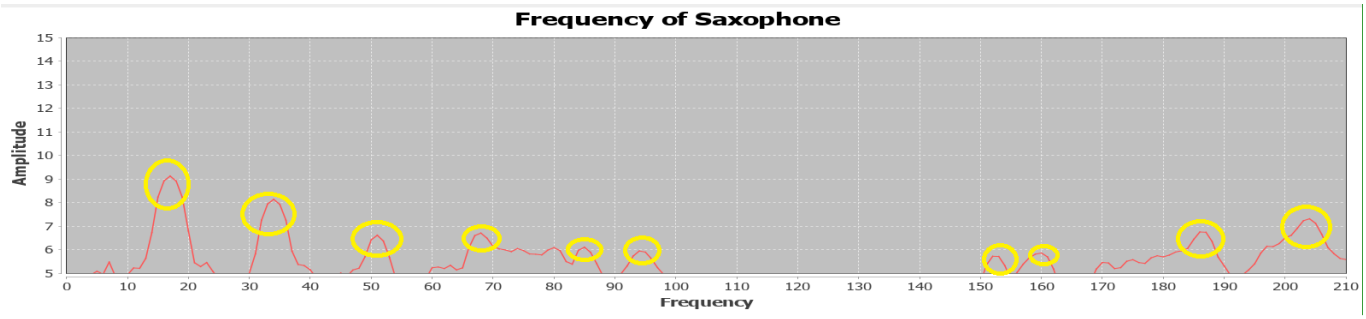

图 5: 聲音的峰值

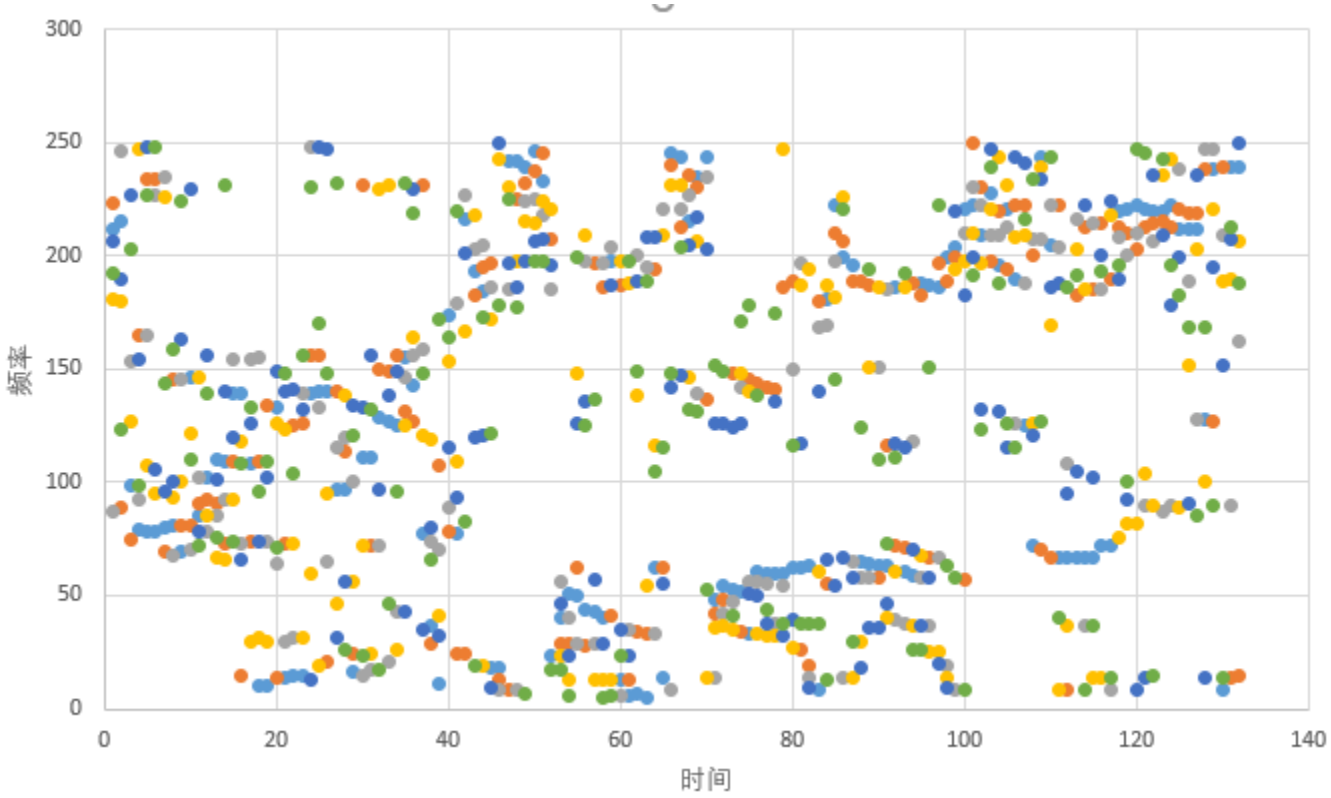

图 6: 某段聲音密碼對應的點陣圖

在兩段聲音信息的相似性比較中，我們使用最長子序列演算法 [4] 進行比較。判定需要對比的兩個位置的相似性時，我們使用前文所述的頻率位置分析演算法進行比較，此演算法會回傳相似/不相似給最長子序列演算法進行判定。最長子序列演算法運行完畢後則會回傳最長子序列的長度，我們再使用最長子序列的長度除以比較短的聲音信息長度最後得出兩段聲音的相似度。我們再根據實驗結果對聲音相似度設定閾值，即可回傳兩段聲音信息是否相似。此演算法的實際使用效果將在後文的實驗記錄中呈現出來。

我們又引入了自我學習模塊配合原算法以加强比對識別率，降低比對誤識率，自我學習模塊將會在後文單獨講述，並在實驗記錄中體現出來。

3.2.3 註冊模塊

此程式模塊多次接收來自聲音密碼生成模塊生成的聲音密碼，達到指定次數[9]後生成對應的密碼母本存入資料庫中。在生成密碼母本時，我們使用自己研發的加權演算法來降低註冊時單次輸入時因環境聲音或者輸入者帶來的誤差。

加權演算法的比對基礎是之前由頻率位置分享演算法與最長公共子序列演算法組合成的聲音密碼比對算法。輸入的聲音為母本聲音密碼與子本聲音密碼，輸出為新的母本聲音密碼。其中母本聲音密碼都含加權值，子本聲音密碼都不含加權值。我們將子本聲音密碼與母本聲密碼進行聲音密碼的比對，並用指針將比對成功的密碼時刻使用指針進行標記，最後將所有母本比對成功的時刻權值 $+1$，最後得到新的母本。若輸入的母本與子本均為同一人説的同一句話對應的不同聲音密碼，新的母本有更大的概率與之後真實使用者輸入的聲音密碼相似。(圖 8)

在註冊模式中，我們採用多次調用加權演算法的方式生成註冊時的聲音密碼母本。我們先通

[9]目前實驗時是五次

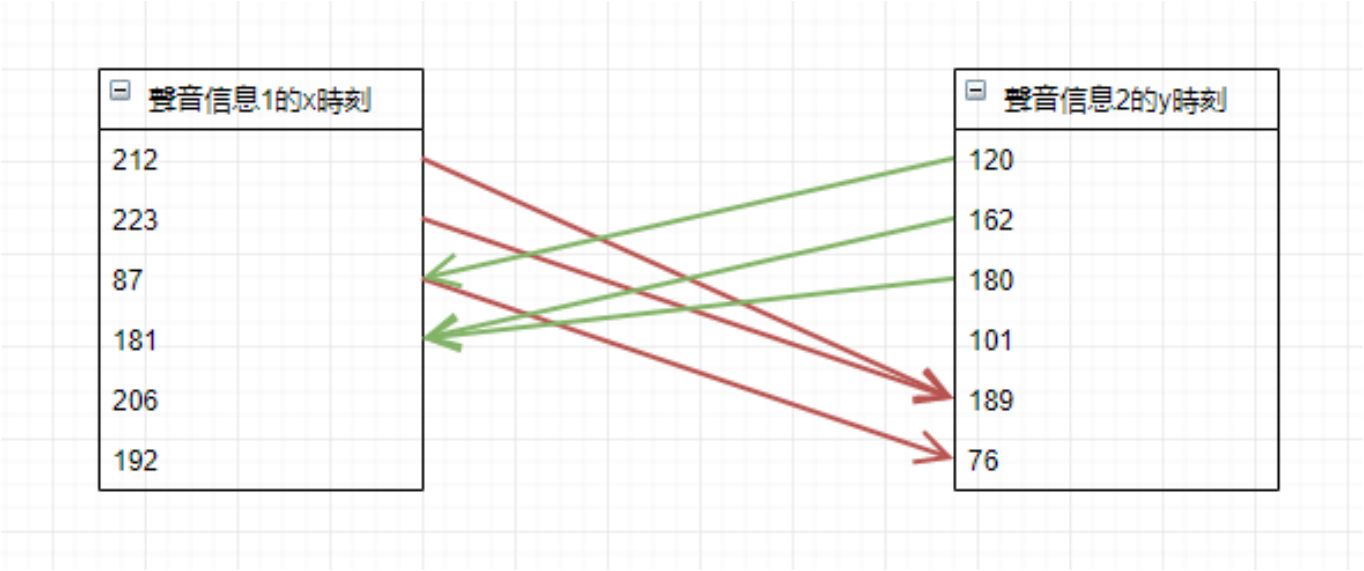

图 7: 頻率位置分析演算法

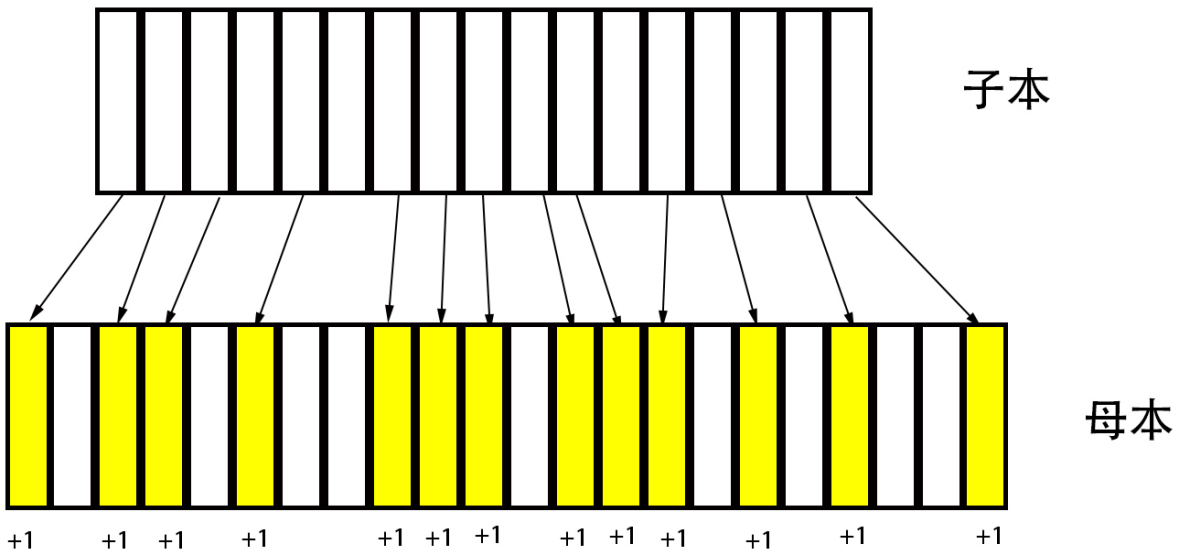

图 8: 加權演算法

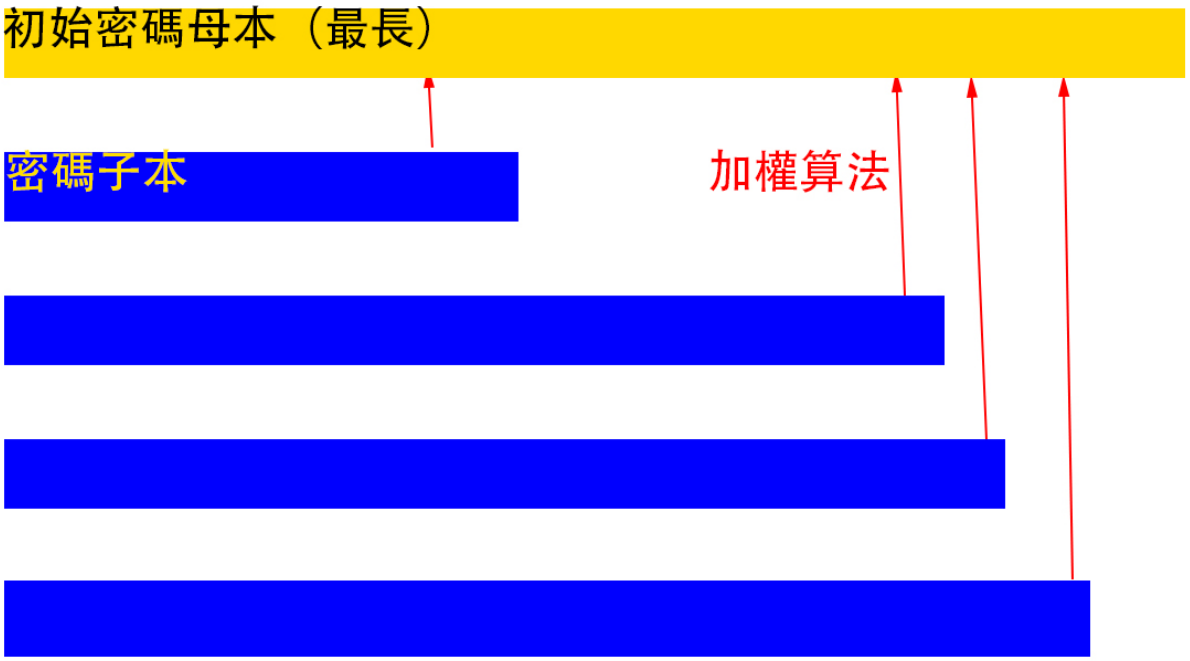

图 9: 註冊算法

過聲音密碼生成模塊獲取到了五個註冊模式輸入的聲音密碼，然後選擇長度最長的那段聲音密碼作為待加權的聲音密碼母本，再依次使用其他的聲音密碼與之進行加權演算，最後可以輸出一個最大全權值為 4 的聲音密碼母本。

使用這種方法生成的聲音密碼母本可最大程度的降低註冊時使用者的輸入誤差，減少當時環境聲音等因素的影響，把使用者具有其聲音特點的密碼模塊突出出來，為提高辨識的識別率提供了可靠的基礎。（圖 9）

3.2.4 自我學習模塊

這個系統的自我學習模塊的目的是為了讓使用者資料庫中的密碼與使用者實際聲音生成的密碼更加相似，提高識別率，降低誤識率。這個程序模塊由聲音密碼比對模塊調用，在每次比對成功后對聲音密碼母本與輸入的聲音密碼子本調用加權演算法，使得每次母本被比對到的時刻權值都會上升，讓會被頻繁比對到的聲音信息影響最後計算出的相似度的佔比更多，減少了無關信息的干擾與環境信息的干擾。（圖 10）

4 使用情境

因為我們系統可拆分可整合的特點，所以適用於多種使用情境，在此選取幾種有代表性的進行說明

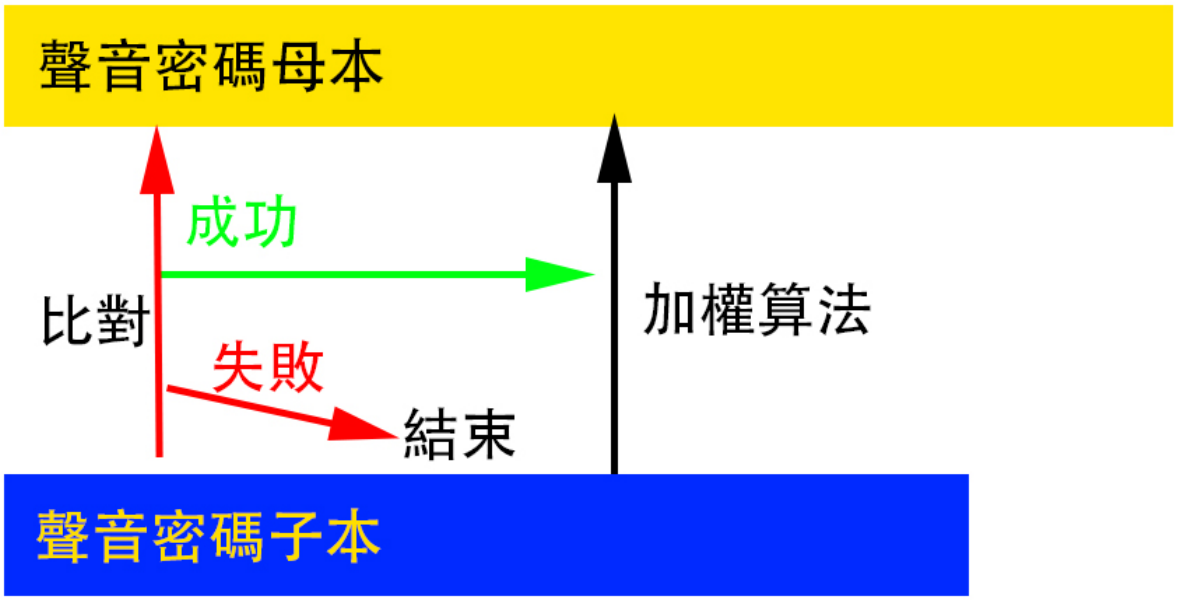

图 10: 自我學習算法

4.1 直連以太網的公共網絡物聯網設備：以餐廳中的物聯網設備管理為例

在具備無線網絡或是高强度的 4G 網絡環境中，系統採用雲端驗證的方式。在智能化餐廳中，有很多物聯網設備需要進行管理，服務員通過手機或是電子熒幕進行操作又顯得繁瑣。搭載了我們聞聲制物系統的物聯網設備只需要管理人員説一句口令，即可對其進行控制。管理人員説“把空調打開”餐廳内的空調即可開始運行，而顧客講相同的話卻不能開啓空調，聞聲制物系統實現了權限控管的功能。

4.2 位於局域網内的設備：以互聯網公司管理系統為例

在公司局域網中，系統不具備以太網，我們則將資料庫與比對服務器架設與公司的局域網中，物聯網設備通過公司無線網絡接入局域網，保障資料安全只需要保護資料庫與比對服務器。公司員工到達單位後説了一聲“我要簽到”，打卡系統便知悉了員工的身份信息為他簽到。公司領導説了一聲“今天十點安排工作會議”，系統即刻識别出這個領導具有安排會議的權限，並創建了提醒，到時間便提醒員工參加會議。

4.3 位於山區、自然保護區等網絡條件差的特殊物聯網監測設備：以移動式的氣象監測設備為例

在網絡條件差的情況下，我們使用 NB-IoT 模式連入物聯網。在只有微弱的 4G 信號覆蓋的山區中，放置有昂貴的移動式氣象監測設備，科研人員需要投放大量設備對自然保護區中的氣候條件進行研究。實驗設備價格昂貴，為了防止他人隨意碰觸，科研人員使用了聲紋鎖。設備採用鋰電池供電，常年在低功耗環境下運行，科研人員先按了系統觸發按鈕，並説出“芝麻開門”，系統成功認證後便為他打開了聲紋鎖。而有愛玩的小朋友去對著系統説“芝麻開門”系統卻毫無反應。

5 商業模式

我們系統的通用性使得商業模式也可呈現多樣化。

5.1 B2C: 普通消費者購買的智能家庭管理中心

消費者可購買我們的聞聲制物系統搭建智能家庭，通過語音即可控制家庭內的物聯網設備運作，同時可只賦予家長使用權限，防止兒童進行誤操作。還可配合網絡訂餐 APP 實現聲紋支付快速訂餐等需要進行使用者身份認證的功能。在系統發展成熟後搭配物聯網路由器實現 mesh 組網即可完美兼容家庭環境的使用，因家庭設備的管控對安全性要求低，比對模塊可置入本地以提高系統的運作速度。

5.2 B2B: 公司管理、公共場所設備管理、戶外物聯網設備管理

在使用情境中我提到了這三種使用情境，我們的系統可針對商業客戶不同的使用需求為其定制適合的系統架構方案，將系統授權給相關單位使用。在這種商業模式中，我們系統的高兼容性得到了完美體現，搭配中華電信提供的 LTE-M 與 NB-IoT 連接模式，為聲紋認證提供了多樣化的解決方案。

6 開發工具及其他相關說明

6.1 作業系統環境

Linux 與 Windows 均可

6.2 主要開發程式語言

JAVA

6.3 專案支援語言

中文

6.4 開發環境及工具

1. Apache NetBeans IDE
2. Arduino IDE
3. IntelliJ IDEA Community Edition
4. Raspberry Pi
5. Mariadb
6. baidu api

参考文献

[1] 曾剑秋. 5g 移动通信技术发展与应用趋势. 电信工程技术与标准化, 30(2):1–4, 2017.

[2] 肖爱民. 基于语音识别技术的智能家居控制系统的设计. Master's thesis, 南昌大学, 2018.

[3] 莫于攀. 社保声纹认证的研究与实现. Master's thesis, 电子科技大学, 2017.

[4] 郑翠玲. 最长公共子序列算法的分析与实现. 武夷学院学报, 29(2):44–48, 2010.

[5] 鹿曼. 基于 *Android* 的智能家居控制系统的设计与实现. PhD thesis, 济南: 山东建筑大学信息与电气工程学院, 2013.