**三、研究計畫內容（以中文或英文撰寫）：**

1. 研究計畫所要探索的課題。
2. 研究方法及進行步驟。
3. 預期達成目標及可行性評估指標。
4. 研究計畫所要探索的課題
5. 說明機器人辩識聲音來源方位的重要性
6. 說明機器人辨識聲音來源方位方法上的限制
7. 指出探索課題

何謂複雜環境?

何謂仿生計算?

聽覺週邊圖

腦幹聽覺訊號路徑圖

系統方塊圖

Nengo概念圖

1. 既有的方位辨識方法

4.1 麥克風陣列方法

4.2 雙耳辨識方法

4.3 Simple ASA

Multiple sources

Binaural cues

Onset detection

Harmonic structure

Streaming

4.4 仿生方法

聽覺訊號傳播處理路徑

HRTF

外耳

中耳

內耳

MSO

LSO

IC

ASA

* 1. 神經系統模擬軟體Nengo, SPA, 及 Spaun

Neural Engineering Framework

Nengo

SPA

1. 研究方法及進行步驟
2. HRTF from Brian
3. 聽覺周邊 : IPEM 模型
4. Jeffrey model in Nengo
5. LSD model in Nengo
6. IC model in Nengo
7. **預期達成目標及可行性評估指標**

預期目標

1. 本研究預期產生一篇投稿國際會議論文。
2. 所完成的軟體系統，預備以下列類型測試場景，在單一聲源的假設下，進行數值實驗。
3. 同時將完成的Python程式，移植到Robot Operating System (ROS)(官網http://www.ros.org/)，以所附機器人模擬系統Gazebo(官網http://gazebosim.org/)呈現3D場景，顯示機器人在無窮空間/長方形空房間內，靜止或運動時，轉頭面向聲音來源的行為。

**測試場景類型1**

探討無反射(anechoic)環境下的靜止機器人聲源定位以及語

意指標建立準確率，與文獻已知數值實驗結果或神經生理特性比較，並討

論對應物理生理機制。

聲源：為簡化問題，設定俯仰角(elevation) 0o; 水平角(azimuth) -90o至

90o， 間隔30o，正前方為0o。

與機器人頭部中心距離1.28 m

雜訊: 不同訊雜比(Signal to Noise Ratio，SNR)的White Gaussian、冷氣

聲、與交通背景雜訊。

機器人頭部:

應用Brian Hear Python程式庫(官網http://www.briansimulator.org/docs/hears.html#head-related-transfer-functions)，取用其中的擬真人頭HRTF。

數值實驗預定結果:

場景1.1~1.4: 求得單聲源方位水平角，畫出類似

(Liu, Erwin, &Wermter, 2008)之圖9(d)的圖表(複製於本計畫書圖 1)，

並探討其準確性及相關物理生理機制。



圖 1.聲音方位辨識結果呈現，圖中方塊大小代表Estimated Angle附近的神經脈衝數與整體脈衝速率的比例大小。本圖取自(Liu, Erwin, &Wermter, 2008)文中的Fig. 9(d)。

場景1.4: 對於國語數字0到10的語音，以及特殊聲音，產生各聲音訊號的語意指標，估算判定語意指標的正確率並對較重要的參數如聲源方向角度作圖。

場景1.1:

聲音訊號：低頻弦波(100 Hz、100 x 151/15 Hz、100 x 152/15 Hz、…、1500 Hz)，預期ITD影響較ILD為大。

場景1.2:

聲音訊號：高頻弦波(1200 Hz、1200 x 151/15 Hz, 1200 x 152/15 Hz、…、

18,000 Hz)，預期ILD影響較ITD為大。

場景1.3:

聲音訊號：自然語音(五個英語字：”hello,” “look,” “fish,”

“coffee,” “tea”)，(Liu et al., 2008) 指出這五個單字包含相當大

範圍的日常語音型態)。雖然我們人形機器人的頭部形狀與

(Liu et al., 2008)的機器人頭部不同，語音特徵也有差異，但是仍可與

其結果參考比較。

場景1.4:

聲音訊號：國語數字0到10的發音、老人低沉呼救聲、重物落地聲(模仿老

人摔倒發出的聲音)、門窗開啟聲、玻璃破裂聲。

**測試場景類型2**

探討回音環境(reverberant environment)下，靜止機器人聲源定位以及語

意指標建立的正確率，與文獻結果及測試場景類型1的各對應場景所得比較，討論相關物理與神經生理機制。

多重反射回音環境:

長方形空房間(shoebox 模型)，使用與

(Youssef, Argentieri, &Zarader, 2012)相同的設定。主要參數如下：

5 x 4 x 2.7 m，灰泥(acoustic plaster)牆、木頭地板、混凝土天花板；

濕度50%；溫度20oC，同時考慮空氣吸收率及聲音強度隨距離產生的衰減。

使用roomsim (Campbell, Palomaki, &Brown, 2005) (MATLAB程式原始碼

位於https://sourceforge.net/projects/roomsim/)產生雙耳室內脈衝響

應函數BRIR(Binaural Room Impulse Responses)，與聲源發出之聲音訊號

摺積(convolution)計算後，即得接收位置收到，包含多重反射回音效果的

聲音。

此時回音時間。

牆壁的吸收係數可以修改，得到RT60等於0.45s與0.7s的回音，一併加入數值實驗，比較不同RT60的結果。

機器人頭部:

中心位置: (2, 2, 1.5)m。

應用CIPIC資料庫(Algazi, Duda, Thompson, &Avendano, 2001)(包含

MATLAB軟體的完整資料庫檔案位於

http://interface.cipic.ucdavis.edu/sound/hrtf.html)，取用其中的擬

真人頭HRTF。

聲源: 使用與(Youssef et al., 2012)相同的設定。為簡化問

題，設定俯仰角0o; 水平角 -45o至45o，間隔5o，正前方為0o。與頭部中心的距離由1m 到2.8m，間隔0.45m。

雜訊：與測試場景類型1相同。

數值實驗預定結果:

場景2.1~2.2: 以不同RT60之值，畫出聲源方位水平角誤差與RT60的相關

曲線，並探討其準確性及相關物理生理機制。

場景2.2: 對於國語數字0到10的語音，以及特殊聲音，產生各聲音訊號

的語意指標，估算判定語意指標的正確率並對RT60統計作圖。

場景2.1: 聲音訊號為自然語音(五個英語字：”hello,” “look,”

“fish,” “coffee,” “tea”)，結果可以參考(Youssef et al., 2012)(Youssef, Argentieri, &Zarader, 2013)(Mroueh, Marcheret, Goel, &Ibm, 2015)。

場景2.2: 國語數字1到10的發音、老人低沉呼救聲、重物落地聲(模仿

老人摔倒發出的聲音)、門窗開啟聲、玻璃破裂聲。

**測試場景類型3**

探討無反射環境下，機器人在不同速率下直線或繞圈運動，求

出聲源定位以及語意指標建立的正確率，相對於時間的變化，與文獻結果及

測試場景類型1的各對應場景所得比較，討論相關物理與神經生理機制。

聲源：為簡化問題，設定俯仰角(elevation) 0。

測試開始時，位於機器人正前方，與機器人頭部中心距離1.28 m。

雜訊: 不同訊雜比(Signal to Noise Ratio，SNR)的White Gaussian、冷氣

聲、與交通背景雜訊。

機器人頭部:

應用Brian Hear Python程式庫中的擬真人頭HRTF。

數值實驗預定結果:

場景3.1~3.3: 求得聲源方位水平角，畫出各弦波頻率下，誤差與時間的

關係圖，並探討其準確性及相關物理生理機制。

場景3.4: 求得聲源方位水平角，畫出聲源定位以及語意指標建立的正

確率，相對於時間的變化，並探討其準確性及相關物理生理機制。

場景3.1: 聲音訊號：低頻弦波，同場景1.1。

場景3.2: 聲音訊號：高頻弦波，同場景1.2。

場景3.3: 聲音訊號：自然語音，同場景1.3。

場景3.4: 聲音訊號：同場景1.4。

**測試場景類型4**

探討回音環境下，直線或繞圈運動機器人聲源定位以及語意指標建立的正確

率，與文獻結果及測試場景類型2的各對應場景所得比較，討論相關物理與神經生理機制。

多重反射回音環境: 同測試場景類型2。

機器人頭部: 起始中心位置: (2, 2, 1.5)m。應用CIPIC的擬真人頭HRTF。

聲源: 起始時位於機器人正前方，與機器人頭部中心距離1.28 m。

雜訊: 不同訊雜比(Signal to Noise Ratio，SNR)的White Gaussian、冷氣

聲、與交通背景雜訊。

數值實驗預定結果:

場景4.1~4.2: 以不同RT60之值，畫出聲源方位水平角誤差與RT60的相關

曲線，並探討其隨時間變化的準確性及相關物理生理機制。

場景4.2: 對於國語數字0到10的語音，以及特殊聲音，產生各聲音訊號

的語意指標，估算判定語意指標的正確率隨時間變化的關物理生理機制，並

對RT60統計作圖。

場景4.1: 聲音訊號：自然語音，同場景3.1。

場景4.2: 聲音訊號：同場景3.2。

**可行性**

**評估指標**

請參見以上各測試場景之說明。

參考文獻

Algazi, V. R., Duda, R. O., Thompson, D. M., &Avendano, C. (2001). THE CIPIC HRTF DATABASE Creative Advanced Technology Center 1500 Green Hills Road Scotts Valley , CA 95066. *Signal Processing*, (October), 99–102.

Campbell, D. R., Palomaki, K. J., &Brown, G. (2005). A MATLAB simulation of“ shoebox” room acoustics for use in research and teaching. *Computing and Information Systems J*, *9*, 48–51.

Liu, J., Erwin, H., &Wermter, S. (2008). Mobile robot broadband sound localisation using a biologically inspired spiking neural network. *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, 2191–2196. http://doi.org/10.1109/IROS.2008.4650760

Mroueh, Y., Marcheret, E., Goel, V., &Ibm, M. I. T. (2015). Deep Multimodal Learning for Audio-Visual Speech Recognition, 2130–2134. http://doi.org/10.1109/ICASSP.2015.7178347

Youssef, K., Argentieri, S., &Zarader, J. (2012). Towards a Systematic Study of Binaural Cues, 1004–1009.

Youssef, K., Argentieri, S., &Zarader, J. L. (2013). A learning-based approach to robust binaural sound localization. *IEEE International Conference on Intelligent Robots and Systems*, 2927–2932. http://doi.org/10.1109/IROS.2013.6696771