**Report**

R04942117 潘郁凱

**Abstract**

聲音的空間定位，是聽覺注意力中很重要的一環。研究主題是以人的生理機制的角度來探討聽覺聲音的空間定位。從外界的聲音訊號，轉換為大腦中神經元的神經訊號，再藉由這些神經訊號中取得一些線索，判斷出聲音的方位。其中，這些線索包含雙耳時間差ITD與雙耳強度差ILD。最終，整合這兩種資訊，來進行聲音來源水平角度的辨識。

1. **Introduction**

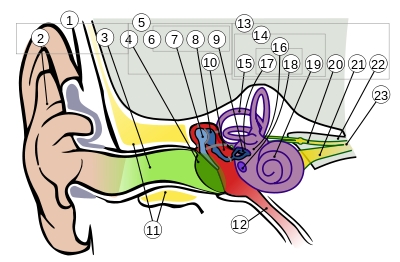
人類的聽覺有著一種聲音空間定位驚人的能力，聽覺系統接收定位聲源所發出的聲波，因為HRTF的效應，從左右兩耳的聽覺訊號中能夠取得各種不同的線索，例如兩耳之間的時間差（Interaural Time Difference，ITD）、兩耳之間的強度差（Interaural Level Difference，ILD），或者是頻譜訊息，從這些線索中，大腦能夠判斷出聲源的位置。

首先，會簡單的介紹人類的聽覺通路，說明人耳的聽覺產生所相關的一系列解剖結構，對於人類的聽覺通路有初步的認識。第二部分，說明HRTF的應用，使得原始單聲道的聲音訊號轉換為左右耳雙聲道的訊號。第三部分，利用Matlab IPEM toolbox中的Auditory Peripheral Module（APM）把左右耳的訊號做分頻處理，依照頻率分為40個頻道，並且再進一步轉換為神經訊號（Neural Rate-Code）的形式。第四部分，將左右兩耳同一頻道的神經訊號同時送到MSO的模型與LSO的模型進行處理，最終再把兩個模型的結果輸入到IC的模型做整合，以判斷出聲音來源水平角度的辨識。

1. **Auditory Pathway**

聽覺通路（Auditory Pathway），是指與聽覺產生相關的一系列解剖結構。聽覺通路在中樞神經系統（腦）之外的部分稱爲週邊聽覺系統，在中樞神經系統内的部分稱爲聽覺中樞或中樞聽覺系統。

耳朵，是動物接收聲波並識別方位，也是維持身體平衡的主要器官，為動物提供聽覺的處理。在大部份的哺乳類動物中，外露在身體外的部分稱為耳殼，作為第一個接收聲音的部分，而人類的耳殼又被稱為耳廓。人類有一雙耳朵，在頭部的兩邊各一隻，通常是左右對稱的，這樣可以幫助判斷聲源的位置。耳朵的生理構造，主要可分為外耳（Outer ear）、中耳（Middle ear）、內耳（Inner ear）三個部份，耳朵整體稱為週邊聽覺系統，如圖（一），最後連接聽神經進入中樞聽覺系統，構成了人類的聽覺系統。



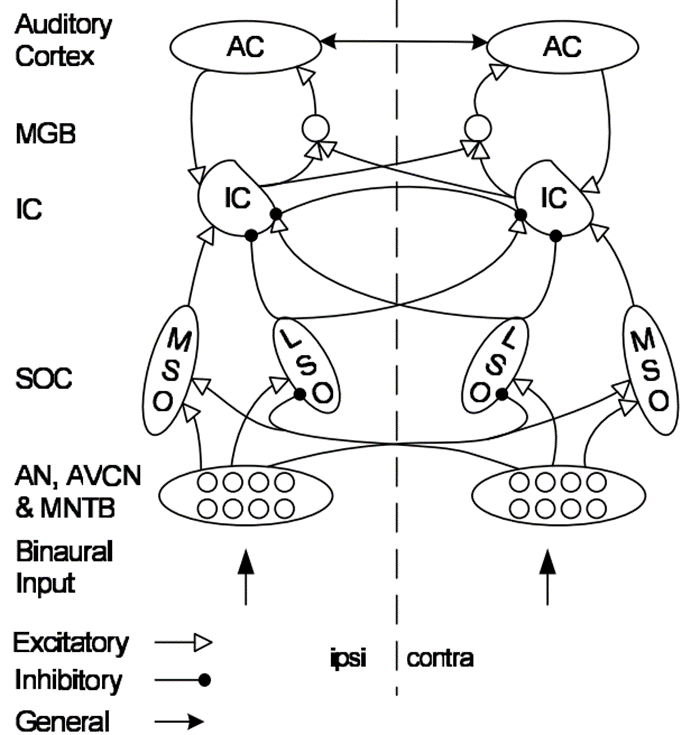
圖（一）: 人耳的解剖構造

1.外耳2.耳廓3.耳道4.鼓膜（耳膜）5.中耳6.聽小骨7.錘骨8.鉆骨9.鐙骨10.鼓室11.顳骨12.耳咽管13.內耳14.半規管15.內耳16.前庭17.卵圓囊18.圓窗19.耳蝸20.前庭神經21.耳蝸神經22.內耳道內部23.前庭耳蝸神經 (https://zh.wikipedia.org/wiki/File:Anatomy\_of\_the\_Human\_Ear\_1\_Intl.svg)

耳朵的生理構造，主要可分為：

* 外耳：耳廓，負責收集聲波傳入耳道；耳道，負責把聲波傳到耳膜，並有擴大效果。
* 中耳：耳膜，作為外耳、中耳之分界線，當聲波撞擊耳膜時，會引起耳膜振動，傳入三小聽骨；聽小骨，包含錘骨、鉆骨、鐙骨，以「槓桿原理」，將聲波的能量轉成「機械能」；卵圓窗，聽小骨的震動導致內部淋巴液產生波動向內傳遞。
* 內耳：耳蝸，與蝸牛殼的形狀相似，淋巴液波動在不同部位形成不同的共振，毛細胞將「液態能」轉換成「電能」連接至聽神經；聽神經，將電能傳送至聽覺中樞，以產生聽覺。

當耳蝸的毛細胞將每個不同頻帶的訊號轉換為神經訊號連接到聽神經後，就進入了到中樞聽覺系統，如圖（二）。聽神經連接至上橄欖體（Superior Olivary Complex），包含了Medial superior olive（MSO）、Lateral superior olive（LSO）兩個部分，可同時接收兩耳的神經訊號，分別取出兩邊訊號的時間資訊，ITD，與強度資訊，ILD。兩種資訊再傳送到下丘（Inferior colliculus, IC），進行聲音的空間定位，最後抵達大腦聽覺皮質負責辨識聲音的意義及種類。



圖（二）: 哺乳動物的聽覺通路的示意圖

AN: Auditory Nerve; AVCN: Anteroventral Cochlear Nucleus; LSO: Lateral Superior Olive; MNTB: Medial Nucleus of the Trapezoid Body; MSO: Medial Superior Olive; IC: Inferior Colliculus; MGB: Medial Geniculate Body; AC: Auditory Cortex. (Liu et al., 2011)

聲波原是以無形的能量存在於空氣中，經耳廓集中收集後，傳遞至外耳、中耳、內耳，在這過程中，能量的形式會不斷轉換，以因應各部份聽覺器官的接收，最後傳遞至大腦，而成為可被理解的有用訊息，如表（一）。

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 聲音 |  | 外耳 |  | 耳膜 | |  | 聽小骨 |  | | 卵圓窗 |  | 耳蝸 |  | | 聽神經 |  | 大腦 |
| ｜----外耳(聲能)---- | | | | | | ｜--中耳(機械能)--｜ | | | | ｜--內耳(液態能)--｜ | | | | | ｜中樞(電能)｜ | | | |

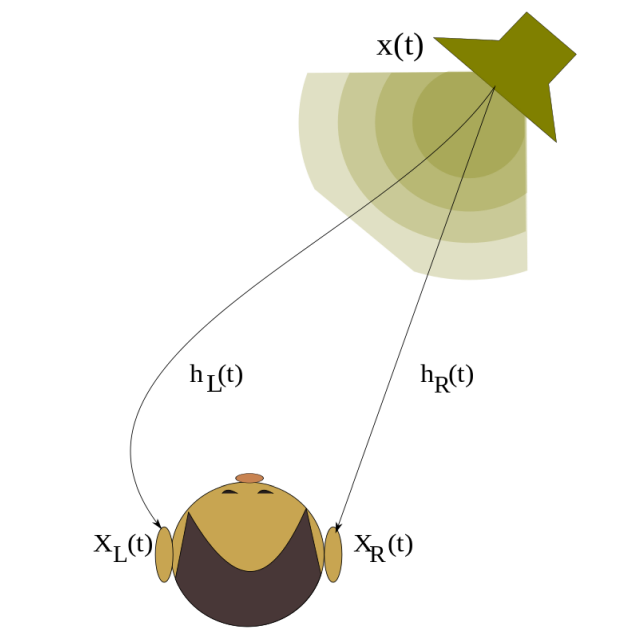
表（一）: 聽覺的傳導途徑

(http://www.ear.com.tw/CGMH-WEB/earinfo.htm)

1. **Head-Related Transfer Function**

Head-related transfer functions（HRTFs）能夠描述從聲源發出的聲波，經由空氣的傳遞產生衰減後，再抵達人類的外耳所得到的訊號，是兩者訊號之間的轉換關係。這種轉換也包括了對我們身體的各個部分所產生的繞射與反射的影響，比如我們的頭、耳朵、肩膀和身體。因此，上面這兩種功能，能夠為我們人類對聲音創造出有距離感、方向性、空間上的錯覺。

HRTFs是頭部相關脈衝響應（Head-Related Impulse Responses，HRIRs）的傅利葉轉換。這一個複數函數的定義強調了對於每一耳（左耳或右耳）都同時擁有振幅大小的資訊與相位位移的資訊。HRTF根據相對於我們的耳朵的聲源方向、角度及位置，而產生不同的響應。所以，透過不同方向的HRTF，我們就能夠模擬來自於不同方向的聲音，圖（三）為HRTF的示意圖。



圖（三）: HRTFs（表達為 HRIRs）的示意圖

(https://en.wikipedia.org/wiki/Head-related\_transfer\_function)

首先從時域上定義，左右兩耳的Impulse Responses分別可以標記為 和 。而從頻域上定義，左右兩耳的Transfer Functions可以用大寫來表示，分別為 和 。函數 用來描述因聲源物體的振動所產生的原始聲音訊號，而函數 與 則分別代表由聲源所發出的聲音，傳遞到左右耳之後所接收到的聲音訊號。

從時域來分析，人耳所接收到的聲音訊號，可以寫成原始的聲音訊號與HRIR的Convolution結果：

(3.-a)

(3.-b)

從頻域來分析，Convolution可以從積分簡化為乘法運算：

(3.2-a)

(3.2-b)

1. **Experiments & Results of HRTFs**

在這個實驗中所採用的資料庫為IRCAM Listen Database。此Database中左右兩耳的資料各別儲存，其中包含了幾個欄位，取樣頻率、仰角、水平角以及相對應的FIR（有限脈衝響應），代表在不同位置的HRIR。

在應用上，取樣頻率設定在44.1 kHz，仰角與水平角都是每15度做一次量測，仰角的範圍由下至上為-45度到90度，而水平角的範圍則是由正前方逆時針方向0度到345度，聲源的距離是固定在1.95公尺，由仰角和水平角共同決定左右耳的HRIR。HRIR的長度皆為8192個點數，輸入的聲音訊號分別與兩耳的HRIR進行運算，計算結果就會產生左右不同的聲音，透過耳機播放，使得原來只有單聲道聲音訊號能夠產生出聲音從不同方向來的錯覺。

下圖（四）為原始單聲道音檔的波形：



圖（四）: 原始單聲道音檔的波形

利用 Head Related Transfer Function（HRTF）濾波器，產生來自不同方向的聲音訊號在左右耳得到的結果，如圖（五）。

|  |  |
| --- | --- |
| （a） | |
|  |  |
| （b） | |
|  |  |
| （c） | |
|  |  |

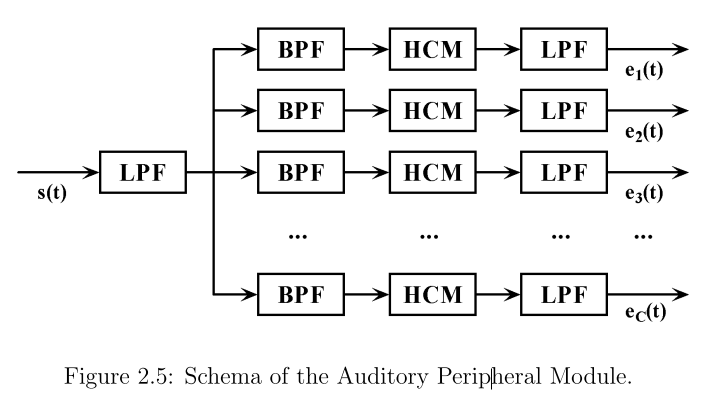
圖（五）: 經過HRTF之後，所產生的左右耳輸出結果

（a）水平角30度，仰角0度，距離1.95公尺；（b）水平角90度，仰角0度，距離1.95公尺；（c）水平角-90度，仰角0度，距離1.95公尺。

1. **Auditory Peripheral Module（APM）**

MATLAB程式庫IPEM Toolbox中的Auditory Peripheral Module模組能夠模擬週邊聽覺系統，將聲音訊號轉換為神經訊號。由低通濾波器LPF模擬外耳、以帶通濾波器BPF陣列模擬耳蝸、由非線性半波整流放大器 HCM（Hair Cell Model）模擬纖毛、以及另一個擷取訊號Envelope的低通濾波器LPF，構成一個週邊聽覺系統的架構，如圖（六）。

此模組將聲音訊號經過BPF陣列分為40個頻率通道，再進一步轉換為神經訊號（Neural Rate-Code）的形式。



圖（六）: IPEM Toolbox中的Auditory Peripheral Module模組

1. **Experiments & Results of APM**

這個模擬是延續上一個主題HRTF的結果，聲音經過HRTF後，分離為左右兩耳的訊號，再分別進入Auditory Peripheral Module（APM），來模擬週邊聽覺系統轉換到神經訊號的過程。

|  |  |
| --- | --- |
| （a） | |
|  | |
|  | |
| （b） | |
|  |  |

圖（七）: 左右耳訊號經過APM之後，所產生40個頻道的神經訊號

（a）分別為左右耳訊號；（b）經過APM之後，所產生的神經訊號。

1. **Auditory Spatial Localization**

聽覺空間定位，人類的聽覺空間感，是聽覺感知中很重要的一環，可以使我們定義出我們周圍環境中物體的方向、位置及任何的移動。聽覺器官對聲源空間位置判斷，必須依賴於雙耳的聽覺，由左右兩耳接收到的聲音訊號，比較兩耳的差異來進行判斷。有時候某些情況下，也可以藉助頭部的轉動，來得到第二種不同的HRTF，這樣就能夠精確的判斷出聲源的位置。

過去有許多心理聲學的研究，根據Duplex Theory實驗證明了主要有兩種資訊提供大腦來進行聽覺定位：

* The Interaural Time Diﬀerence（ITD）：兩耳時間差

由於聲源與兩耳間的距離大多情況下不相同，導致聲波達到兩耳時會有時間上的差異，訊號會有相位上的差別。較先聽到聲音的一耳，代表離聲源較近。若是兩耳幾乎同時聽到，則代表聲源落在兩耳的中介面上。對於低頻的聲音，大腦主要依賴於ITD的資訊進行判斷（水平角）。

* The Interaural Level Diﬀerence（ILD）：兩耳強度差

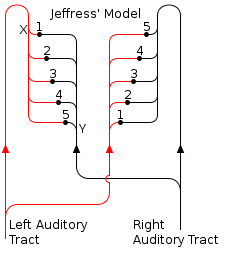
假如聲音在我們的左方，因為比較靠近左耳，左耳聽到的強度大於右耳，所以我們會判斷出聲音在偏左的位置。對於較高頻的聲音，大腦主要依賴於ILD的資訊進行判斷（水平角）。

人類在進行聽覺空間定位時，大腦會利用ITD與ILD來進行判斷。Medial superior olive（MSO）主要是利用兩耳的神經訊號取得雙耳時間差ITD，而Lateral superior olive（LSO）則是利用兩耳的神經訊號取得雙耳強度差ILD。最終，兩邊的訊號會送至聽覺定位信息整合的主要部位，下丘（Inferior colliculus, IC）。在低頻帶IC主要根據ITD進行聲源水平角度的辨識，在高頻帶IC主要則根據ILD進行聲源水平角度的辨識，而中頻帶IC會同時利用ITD與ILD來進行聲源水平角度的辨識。

1. **MSO Neuron Model**

目前MSO的模型使用最多的是Jeffress Model，如圖（八）、圖（九）。在這模型中，來自左右兩耳的神經訊號是延著相反方向輸入。其中有三個主要的組成成分，分別為時間延遲線（tapped delay lines）、重合神經元（coincidence neurons）、重合計數器（coincidence counters）。

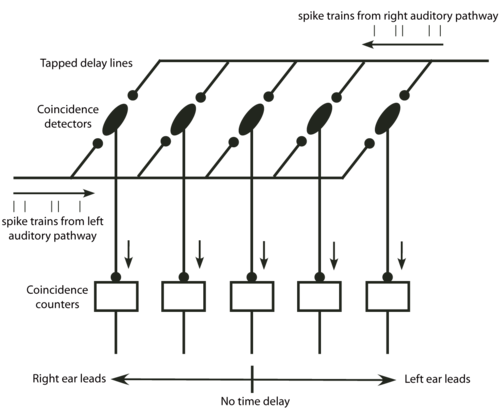
時間延遲線（Tapped delay lines），左右兩邊的輸入訊號以相反方向的路徑排列，並且依據神經纖維的路徑長短不同，造成傳遞神經訊號時產生時間延遲，得以抵銷原來輸入訊號的時間差，因此左右兩邊的神經訊號會同時抵達某一個位置。重合神經元（Coincidence neurons），接受左右兩邊經過不同時間延遲的神經訊號，並且以並聯方式排列。在不同的時間延遲下，各別連接到不同的Coincidence neurons上，只有在左右兩個輸入同時刺激的時候才會被激發。因為有Tapped delay lines的機制，使得在適當的時間延遲下，兩邊的輸入能夠同時抵達某個Coincidence neuron產生反應。重合計數器（coincidence counters），類似加法器或積分器的作用，記錄在一個短時間之內，不同Coincidence neurons產生反應的次數或Firing rate做為兩邊訊號重合的程度。



圖（八）: Jeffress 所提出的MSO模型

在人體腦幹的解剖結構中，左右各有一個MSO的腦區。此圖表示左MSO與右MSO。

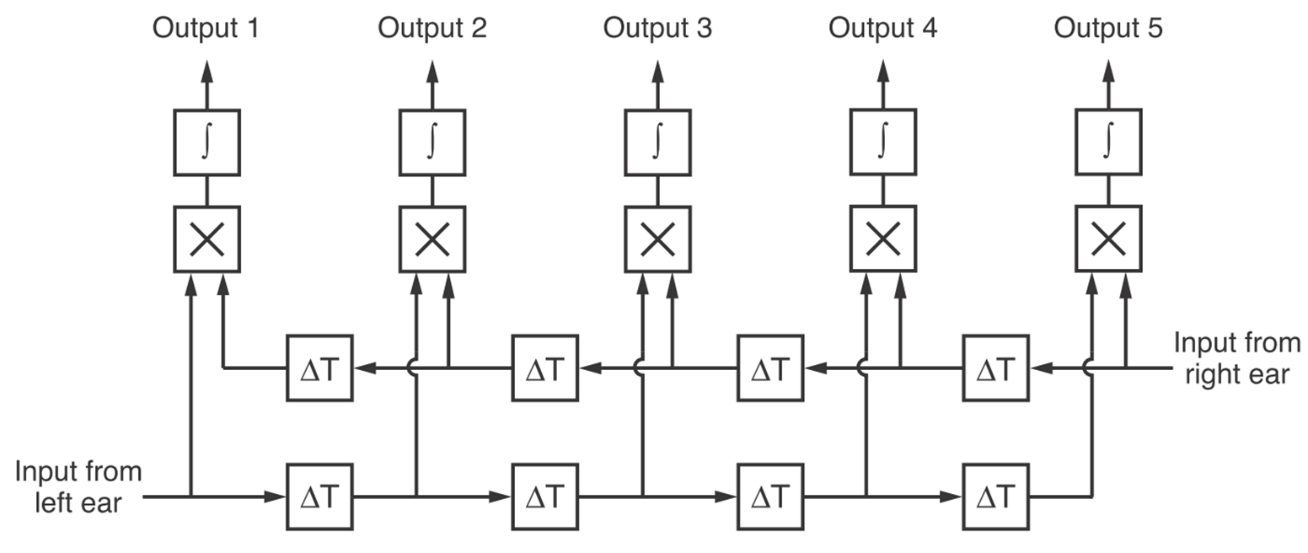
(https://en.wikipedia.org/wiki/Coincidence\_detection\_in\_neurobiology)



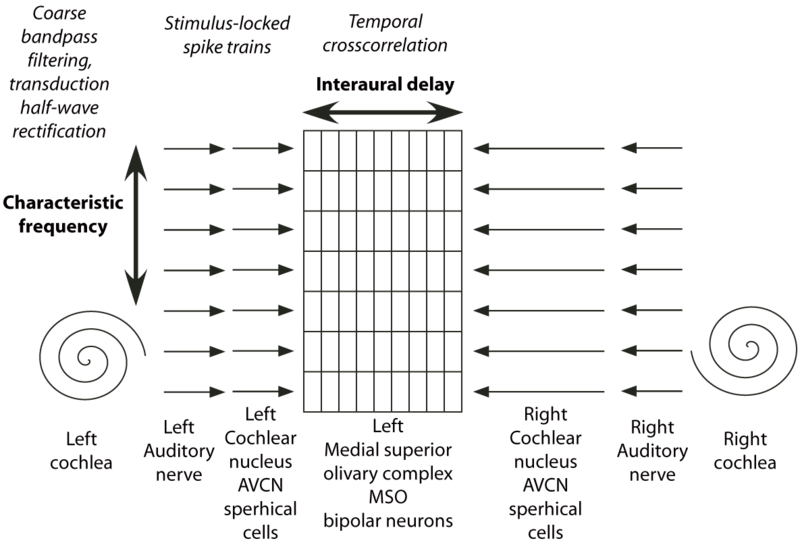
圖（九）: Jeffress Model的時間延遲神經架構。主要有三個組成成分，tapped delay lines、coincidence neurons/detectors、coincidence counters。

(http://www.scholarpedia.org/article/Jeffress\_model)

從數學的角度上，Jeffress認為這個架構可以用交相關函數（Cross-Correlation）的電路來實現，如圖（十）。從耳蝸開始的分頻結果，會一路延續到高級的聽覺處理，稱作Tonotopy。因此MSO模型也會依照頻率通道的不同而單獨做相關的處理過程，如圖（十一）。



圖（十）: Jeffress Model的Cross-Correlation Functions (DL Wang et al, 2006)



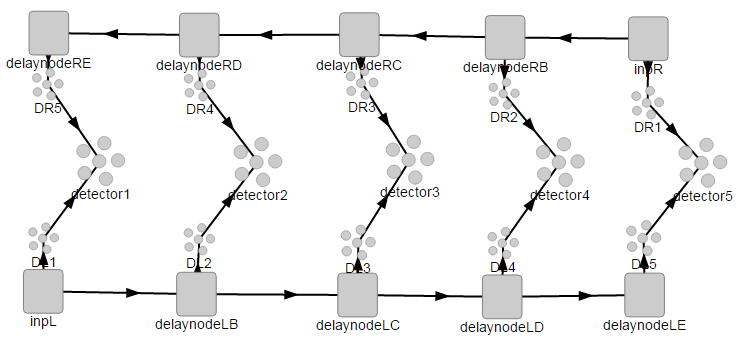
圖（十一）: Jeffress Model的分頻處理架構

(http://www.scholarpedia.org/article/Jeffress\_model)

1. **Experiments of MSO Model in Nengo**

從上一個模擬中，我們得到APM輸出的40個頻道的神經訊號，以此作為MSO Model中左右兩耳各個頻率通道的輸入，我們將利用神經科學模擬軟體Nengo來模擬神經網路的架構。

如圖（十一）所示，依照Jeffress Model的架構，以Nengo中的相關物件所形成的MSO Model。每一個除了Input Node作為訊號輸入端之外，其他的Node都是一個Delay Node，可以將輸入訊號作時間上的延遲，達到時間延遲線（Tapped delay lines）的效果。每一個延遲結果都會送到不同的Ensemble，作為Jeffress Model的重合神經元（Coincidence neurons）。最終比較這些Ensemble的Output大小來決定兩邊訊號重合的程度。每個頻率通道都經過這樣的MSO Model處理過程，如此來實現Jeffress Model的分頻處理架構。



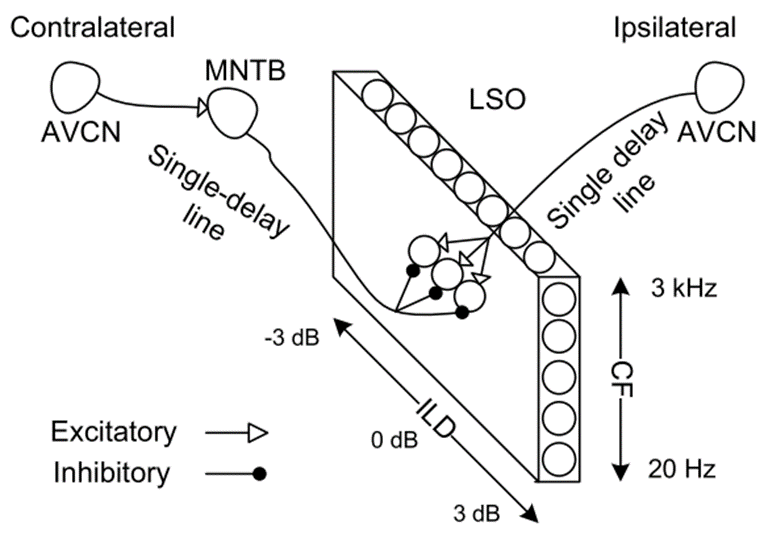
圖（十一）: Nengo的MSO Model

此為每個頻率通道的MSO model。其中inpL與inpR作為左右兩耳的輸入，其他Node為Delaynode產生不同的時間延遲，detector為重合神經元。

1. **LSO Neuron Model**

相對於MSO提取雙耳訊號中的時間資訊，LSO是用來提取雙耳訊號中的強度資訊，MSO與LSO一同用來幫助聽覺的空間定位。LSO神經元接收左右兩耳的輸入訊號，可以分為興奮性輸入（Excitatory Input）與抑制性輸入（Inhibitory Input），在數學的意義上，等效於兩個輸入作減法的運算，以此來得到雙耳強度差ILD。在人體腦幹的解剖結構中，左右各有一個LSO的腦區（MSO亦同），同側的輸入信號作為興奮性輸入（Excitatory Input），對側的信號則作為抑制性輸入（Inhibitory Input）。例如，以左側LSO來說，左耳的神經信號為Excitatory Input，而右耳的神經信號為Inhibitory Input。這就是LSO最基本的神經生理機制。

在Tonotopy的分頻處理架構上，Liu等人提出了ILD Map的模型，如圖（十二）。在此模型中，一軸為頻率分佈由20Hz到3kHz，一軸為ILD range由-3dB到+3dB，兩軸構成LSO神經元的分佈情形。每一個LSO神經元都有不同的特徵雙耳強度差（Characteristic ILD），即為此LSO神經元對某一特定的ILD刺激最為敏感，只會在Characteristic ILD的刺激下產生反應。依照此神經機制，在不同頻率通道下的ILD都能被偵測出來。



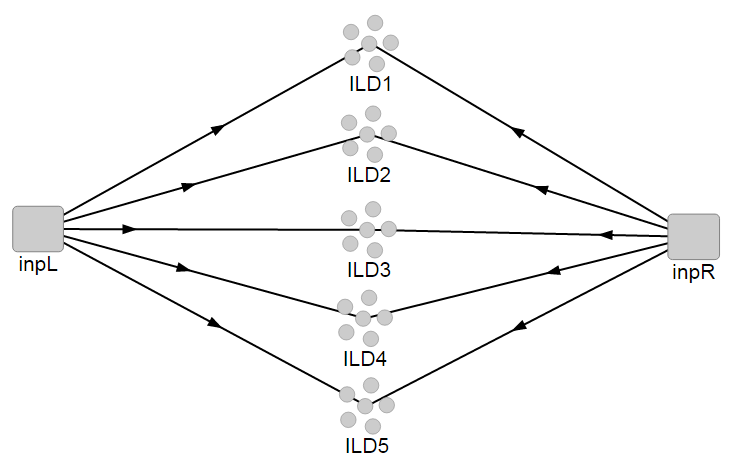
圖（十二）:人類LSO處理ILD的示意圖

Note:人類ILD實際範圍為30dB，此為3dB代表ILD range可以伸縮調整。(Liu et al, 2008)

1. **Experiments of LSO Model in Nengo**

與MSO Model的模擬相同，由之前的模擬中APM輸出40個頻道的神經訊號，以此作為LSO Model中左右兩耳各個頻率通道的輸入，並且同樣利用神經科學模擬軟體Nengo來模擬其神經網路的架構。

如圖（十一）所示，依照Liu等人所提出的ILD Map模型架構，以Nengo中的相關物件形成的LSO Model。以單一頻率通道來看，Input Node作為左右兩耳訊號輸入端，兩邊的輸入直接送到作為LSO神經元的Ensemble中，依照不同的Characteristic ILD排列，一邊作為興奮性輸入（Excitatory Input），另一邊作為抑制性輸入（Inhibitory Input），並且設定這些Ensemble只會在特定ILD的刺激下才會產生反應，如此偵測出ILD的資訊。最後延伸到多個頻率通道，所有頻道都經過這樣的LSO Model處理過程，形成Liu等人所提出的模型架構。



圖（十三）: Nengo的LSO Model

此為每個頻率通道的LSO model。其中inpL與inpR作為左右耳的輸入，中間的Ensemble作為LSO神經元，分別對應不同的ILD。

1. **IC Model**
2. **Experiments of IC Model in Nengo**