

## 多樣訊雜比之訓練語料於降噪自動編碼器其語音強化功能 之初步研究

### A Preliminary Study of Various SNR-level Training Data in the Denoising Auto-encoder (DAE) Technique for Speech Enhancement

李世光<sup>1</sup>、王緒翔<sup>2,3</sup>、曹昱<sup>3</sup>、洪志偉<sup>1</sup>

<sup>1</sup> 國立暨南國際大學電機系

<sup>2</sup> 國立台灣大學通訊工程研究所

<sup>3</sup> 中研院資訊科技創新研究中心

[s105323501@mail1.ncnu.edu.tw](mailto:s105323501@mail1.ncnu.edu.tw), [sypdbhee@gmail.com](mailto:sypdbhee@gmail.com),

[yu.tsao@citi.sinica.edu.tw](mailto:yu.tsao@citi.sinica.edu.tw), [jwhung@ncnu.edu.tw](mailto:jwhung@ncnu.edu.tw)

#### 摘要

在當今普遍的語音應用、諸如語音辨識、語音資訊檢索及聲控機器人等，用以消除雜訊干擾的語音強化技術扮演了相當重要的角色，在眾多語音強化技術中，降噪自動編碼器 (denoising auto-encoder, DAE) 為近年來被廣為探討與使用的方法之一，主因是其使用了當今熱門的深度學習技術、來學習雜訊語音與乾淨語音之間的對應關係，在許多文獻中，DAE 法已被證實可以有效降低雜訊成分、且不至於對原始乾淨語音產生明顯干擾，然而，其效能仍然會隨著訓練語料與模型架構其選擇的不同而有所差異。在本論文中，我們主要是探討不同訊雜比的訓練語料對於 DAE 法其減低雜訊效應的影響。

根據我們初步的評估實驗，主要的發現在於當使用高訊雜比的訓練語料時，所對應的 DAE 法在各種訊雜比的測試語音上，平均而言都能得到顯著的消噪效果，且優於其他種訓練語料所得之 DAE，包含了由多層訊雜比的訓練語料、及近似於訓練語料之測試語音。儘管這似乎與常理不合，我們在論文中提供了可能的解釋，並提及單純使用高訊雜比的訓練語料對於 DAE 訓練效率的優點，其包含了訓練語料數量相對減少、可使用較少隱藏層的簡易 DAE 架構、及調適至其他類型雜訊之可能性等。

## Abstract

Speech enhancement (SE) that reduces the noise effect plays an important role in the current widespread audio applications such as speech recognition, speech-based information retrieval and voice control. Among the various speech enhancement techniques, denoising auto-encoder (DAE) employs the well-known deep learning process to learn the transformation from noisy data to the respective clean noise-free counterpart, and it has been shown to be very effective in reducing the noise component as well as introducing little speech distortion. In this paper, we primarily investigate the influence of the training data with different signal-to-noise ratios (SNRs) for DAE in the corresponding SE capability.

The major finding from our evaluation experiment is that the DAE trained via high-SNR data provides significantly better improvement in speech quality for the noisy testing data over a wide range of noise levels, when compared with the DAE trained via either of multi-SNR data and matched-SNR data. This result somewhat disagrees with the common and instinctive sense that the model created with multi-SNR training data behaves well on average for the testing data at an arbitrary noise level, and the matched-condition model should give the optimal performance. However, we give the possible explanations about the above finding, and explore some advantages of using simply high-SNR training data to prepare the DAE for speech enhancement. These advantages include a smaller amount of training data being required, a simpler DAE structure with fewer hidden layers and higher adaptability to other noisy situations.

關鍵詞：語音強化、時頻圖、降噪自動編碼器。

Keywords: speech enhancement, spectrogram, denoising auto-encoder.

### 一、緒論

聲學、語音相關應用如語音通訊、語音辨識等，在現代生活中已無所不在。隨著嵌入式系統不斷演進，以前我們只在移動裝置上傳送或接收聲學訊號，但現在各種物聯網裝置都會是我們說話的對象。從前移動裝置只是傳送聲學訊號至另一端，如今我們有成熟的語音辨識技術可用，其對應的諸多語音應用或功能並已成為不

少人生活中不可或缺的一部分，同時，針對語音處理所面對的各項問題所做的研究與發展至今仍方興未艾，近年研究之相關文獻諸如[1-7]等。

當我們對著錄音裝置說話時，語音經常伴隨著環境雜訊一起傳送。若是傳送此含有雜訊的語音至另一端，顯而易見地環境雜訊會影響語音的品質，若是傳送至伺服器進行語音辨識，環境雜訊則通常會嚴重影響辨識的精確度。

要降低環境雜訊對語音品質或語音辨識結果的影響，其中一個方法就是對參雜環境雜訊的語音（簡稱雜訊語音）進行語音強化（speech enhancement）。語音強化的技術針對不同的應用也可能會有不同的結果。針對提升語音品質而發展的語音強化法，在語音辨識上不一定是正相關的提升效能，反之亦然。

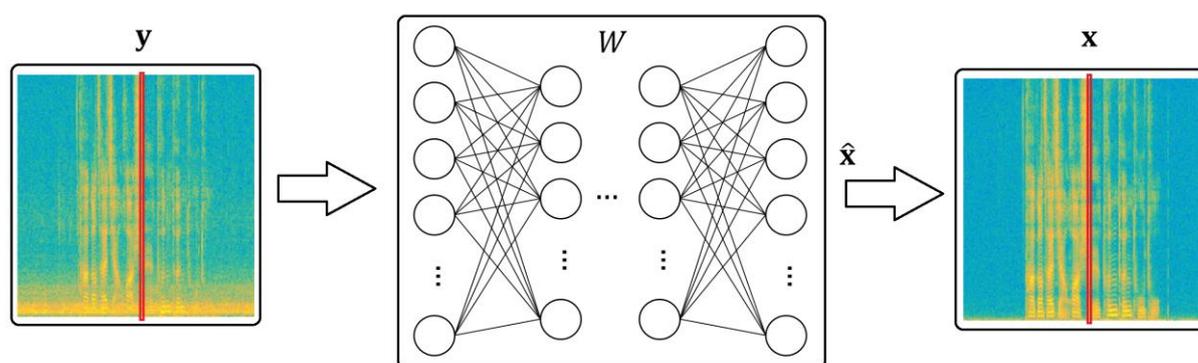
在本篇論文中，我們評估降噪自動編碼技術（denoising auto-encoder, DAE）此一語音強化法的效能。我們把不同訊雜比的訓練語料進行分組，並且針對 DAE 的架構進行調整，期望能完整分析不同訊雜比的訓練語料對訓練 DAE 的影響。

為了評估我們的實驗結果，我們使用語音品質之感知偵測法（Perceptual Estimation of Speech Quality, PESQ）[8]作為評估強化後語音訊號的品質指標。PESQ 使用成對的雜訊語音與乾淨語音，經計算後可以得到給定的雜訊語音接近乾淨語音的程度，換言之即是語音品質。我們把準備的雜訊語音做為測試語料，送入不同經由不同訊雜比訓練語料訓練而得之 DAE 進行強化，即得到語音強化後的測試語料。接著對 DAE 處理前後所得的語音計算其 PESQ 值，藉由二者之差異來評估該 DAE 的效能。

## 二、降噪自動編碼器（denoising auto-encoder, DAE）

自動編碼器（Autoencoder, AE）是神經網路的一種架構，其功能可以看作是非線性的主軸成分分析法（Principal Component Analysis, PCA），但一般而言其具有比 PCA 更好的效能，能有效儲存給定資料集的顯著特性[9]。Vincent 等學者在 AE 的架構上嘗試在訓練資料中引入雜訊[11]，實驗結果指出，在非監督式學習的訓

練上引入雜訊所得之 AE，相較於沒有引入雜訊的 AE 而言，效能會有顯著的提高，並且訓練出來的 AE 將帶有抗雜訊 (denoising) 的能力，因此其也稱為消噪型 AE (denoising AE, 簡稱 DAE)。而 Xugang 等學者在 2013 年引入 DAE 架構在語音強化上[11]，確認了 DAE 的抗雜訊特性能在語音強化任務中實現。圖一為 DAE 架構圖。



圖一、DAE 架構， $y$  與  $x$  分別代表雜訊語音與乾淨語音時頻圖(spectrogram)中之對應行向量。 $W$  代表 DAE 的模型參數， $\hat{x}$  為行向量  $y$  經正向傳遞算法而得之行向量。 $\hat{x}$  的數值會隨著參數  $W$  的更新變得更接近乾淨語音時頻圖行向量  $x$ 。

在本論文中使用的 DAE 架構組合可分為輸入層、隱藏層及輸出層。假設隱藏層層數為一，即第零層為輸入層、第二層為為隱藏層、第三層為輸出層。雜訊語音特徵  $y$  經隱藏層及輸出層正向傳遞計算後得到一向量  $\hat{x}$ ，如下式(1)與(2)所示，其中 $\{W_1, \mathbf{b}_1\}$ 與 $\{W_2, \mathbf{b}_2\}$ 分別為第一層與第二層的權重變數， $\sigma(\cdot)$ 為啟動函數(activation function)，通常採用 S 型函數(sigmoid function)，在訓練時，我們把輸出之  $\hat{x}$  與理想值  $x$  的差值用倒傳遞算法(backward propagation) 計算得到梯度(gradient)即可更新 DAE 之權重變數 $\{W_1, \mathbf{b}_1\}$ 與 $\{W_2, \mathbf{b}_2\}$ 。

$$\text{第一層輸出： } h(\mathbf{y}) = \sigma(W_1 \mathbf{y} + \mathbf{b}_1). \quad (1)$$

$$\text{第二層輸出： } \hat{x} = W_2 h(\mathbf{y}) + \mathbf{b}_2. \quad (2)$$

為避免訓練 DAE 時有過度擬合 (overfitting) 的狀況發生，在定義所需的損失函

數(loss function)時，通常會引入正則項 (regulation)，如式(3)中的 $g(\Theta)$ 項，倒傳遞運算即是使用到此損失函數相對於各變數項的微分，亦即其梯度。一般而言，在引入正則項後，對應之 DAE 的效能會隨著訓練次數增加而漸趨穩定，我們也能有效地比較不同 DAE 架構配置對效能帶來的影響。

$$\text{損失函數：} \quad J(\Theta) = \sum_{\mathbf{x}} \|\mathbf{x} - \hat{\mathbf{x}}\|^2 + g(\Theta). \quad (3)$$

通過引入正則項於損失函數，所訓練得出的 DAE 模型會有穩定的效能。進而使不同架構配置的 DAE 間會有明顯的效能區別，使我們能更能有效比較不同 DAE 實驗設置所得到的結果。

### 三、訓練降噪自動編碼器之配置及評估方式

為了評估 DAE 在語音強化上的效能，我們在訓練過程中設置了各種變因，包含不同訊雜比之訓練語料、DAE 中隱藏層包含的神經元個數、以及隱藏層之數量。

#### (一)、不同訊雜比之訓練語料

我們在上一章提到，根據 Vincent 等學者的實驗結果[10]，在訓練 AE 時引入雜訊，所得出的神經網路將擁有抗雜訊的效能，並且對於原乾淨語料能有更佳代表性。因此，這裡我們使用不同訊雜比(signal-to-noise ratio, SNR)之語料來訓練 DAE，其訓練語料分類如下：

- 1、完整之全訊雜比(all-SNR)之訓練語料：SNR 範圍為-12 dB 至 12 dB，間隔 3 dB，即 -12 dB, -9 dB, -6 dB, -3 dB, 0 dB, 3 dB, 6 dB, 9 dB, 12 dB，共 9 種訊雜比。
- 2、高訊雜比(high-SNR)之訓練語料：SNR 範圍為 6 dB 至 12 dB，間隔 3 dB，即 6 dB, 9 dB, 12 dB，共 3 種訊雜比。
- 3、中訊雜比(median-SNR)之訓練語料：SNR 範圍為 -3 dB 至 3 dB，間隔 3 dB，即 -3 dB, 0 dB, 3 dB，共 3 種訊雜比。
- 4、低訊雜比(low-SNR)之訓練語料：SNR 範圍為 -12 dB 至 -6 dB，間隔 3 dB，即 -12 dB, -9 dB, -6 dB，共 3 種訊雜比。

這裡值得注意的是，由於個別訊雜比的語料量相同，所以全訊雜比使用的訓練語料量是其他三類（高、中、低訊雜比）訓練語料量的 3 倍。

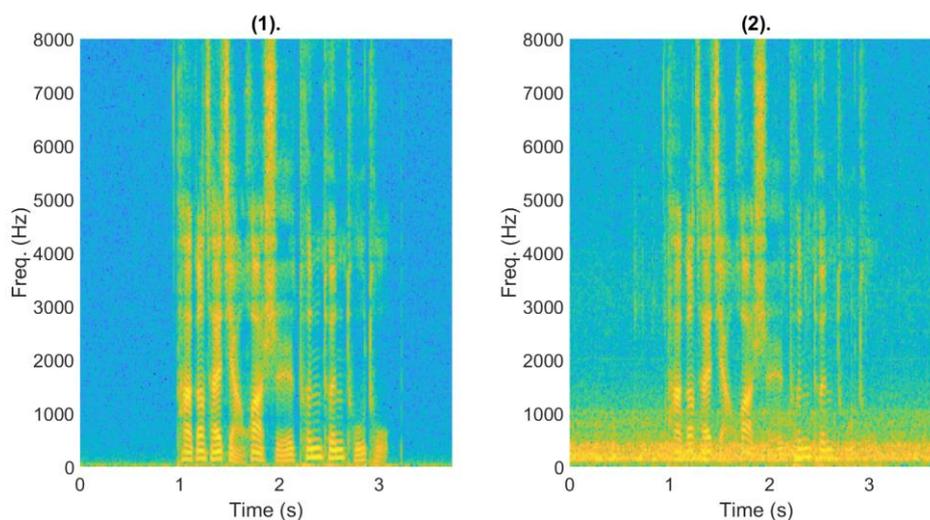
### (二)、降噪自動編碼器神經元數量及隱藏層深度

DAE 模型的大小可經由隱藏層之神經元數量及隱藏層深度做調整。我們參考了 Xugang 等學者的實驗配置[11]，將評估實驗所用的 DAE 架構設置變化如下：

- 1、各隱藏層包含的神經元數量：100、300、500、1000、3000
- 2、隱藏層深度：1 至 6 層

### (三)、特徵 (Features)

如同 Vincent 等學者的實驗[10]，我們對 DAE 神經網路的訓練資料是使用成對的雜訊語音及乾淨語音(noisy-clean pair)，而實際運用的特徵單位則是每個音框(frame)的對數頻譜(logarithmic spectrum)，其具體求法為語音之短時間音框訊號作傅立葉轉換後，將得到的頻譜之強度(magnitude)取對數，一段語音之時頻圖(spectrogram)如圖二所示，其中每一行向量即為各音框之對數頻譜，我們把雜訊語音的對數頻譜當成輸入特徵，其對應之乾淨語音的對數頻譜作為監督式學習中的理想輸出(desired output)特徵。



圖二、語句經短時距傅立葉變換並取對數後所得之時頻圖(spectrogram)。(1) 代表乾淨語音、(2) 代表雜訊語音，其中橫軸為時間、縱軸為聲學頻率(acoustic frequency)。

#### (四)、Perceptual Estimation of Speech Quality (PESQ)

我們把前述不同類別訊雜比訓練語料訓練所得之 DAE 模型，對測試語料進行測試。測試語料經過 DAE 運算後所得之強化語音，理想上會趨近其對應之乾淨語音。為了評估 DAE 強化語音的效能，我們使用 PESQ 做為評估效能的指標。欲計算一語音之 PESQ，與 DAE 一樣需要其對應的乾淨語音。PESQ 值介於-0.5 至 4.5 之間，越高的值代表測試語音越接近其包含的乾淨語音、亦即語音品質越佳。

#### 四、實驗結果與分析

在評估實驗中，我們使用了 Mandarin hearing in noise test (MHINT) 語料庫的子集(subset) [13]，分別取其 360 句及 120 句作為原始訓練與測試語料、且二者並不重複，同時，我們對二者加入汽車引擎雜訊，建構成各種訊雜比(SNR)的雜訊語音。訓練語料的訊雜比介於 -12 至 12 dB 之間、間隔 3 dB、共 9 種，而測試語料的訊雜比介於 -10 dB 至 15 dB 之間、間隔 5 dB、共 6 種，訓練與測試語料在訊雜比上互相不匹配。另外，我們所使用的語料，其取樣頻率為 16 kHz。使用短時距傅立葉變換時的取樣長度為 512 維，因此特徵長度為 257 維。

在使用對數頻譜特徵進行 DAE 的訓練時，我們把訓練的特徵總數量限制為 8 萬個。因此，雖然所配置的不同訓練資料大小不一，但我們可以用同樣的神經網路架構去比較不同訊雜比的配置所帶來的影響。

接下來我們將從多樣訊雜比之訓練語料、神經元數量及隱藏層深度對應之 DAE 來呈現並探討我們的實驗結果：

##### (一)、多樣訊雜比之訓練語料比較

表一列出了測試語料經過以不同訊雜比組合訓練語料訓練出的 DAE 處理前後的 PESQ 分數。如第三節所述，這裡共有四種不同的訊雜比組合，分成「全訊雜比」、「低訊雜比」、「中等訊雜比」以及「高訊雜比」。這裡我們使用的 DAE，包含有一個隱藏層，且此隱藏層有 300 個神經元，基準值(baseline)表示測試語料未經 DAE 處理前的 PESQ 分數。

從表一的結果，我們可以有以下幾點觀察：

- 1、對比基準值，經過 DAE 處理的測試語料，其 PESQ 分數不一定會上升。特別在由低訊雜比訓練語料所得的 DAE。此結果大致符合一般對於映射(mapping)函數的看法，亦即要從嚴重失真的資料中找出對應到無失真資料的映射關係是相對困難的。這也可以部份解釋為什麼低訊雜比（包含了-12 dB, -9 dB 與-6 dB 訊雜比的訓練語料）之 DAE 對於訊雜比 -10 dB 之測試語料的語音強化結果相當有限，即使其相對於其他 DAE 而言，與訊雜比-10 dB 之測試語料是處在相對匹配(match)的情況。
- 2、對於訊雜比 0 dB 與 5 dB 的測試語料，全訊雜比的 DAE 優於中訊雜比的 DAE，對於訊雜比 -10 dB 與 -5 dB 的測試語料，全訊雜比的 DAE 也優於低訊雜比的 DAE。然而，對於訊雜比 10 dB 與 15 dB 的測試語料，全訊雜比的 DAE 卻不能提升其語音品質、得到的 PESQ 比基準值(baseline)還低。可見全訊雜比之訓練語料並不能保證其 DAE 一定能提升任何訊雜比之語音的品質。
- 3、中訊雜比訓練所得的 DAE 在低訊雜比(-10 dB, -5 dB)的測試語料上有最佳的品質提升結果。同時，高訊雜比的 DAE 幾乎在所有訊雜比的測試語料上都能提升 PESQ 分數（除了訊雜比 15 dB 的語料外）。這種由特定訊雜比訓練出的 DAE、可更有效強化相對較低訊雜比之測試語料的現象，與學者 Minje Kim 其文獻[12]所呈現的實驗十分相似。對這一實驗結果，我們有以下解釋：
  - (1) 相較於其他訊雜比資料而言，高訊雜比語料與乾淨語料兩者之間的距離較小，因此，高訊雜比的 DAE 比其他訊雜比的 DAE 在訓練上較為容易且精確。
  - (2) 語音強化法效能反映在雜訊抑制的程度與對於乾淨語音的干擾(speech distortion)程度，高訊雜比的 DAE 由於是學習如何將高訊雜比語音轉換至乾淨語音，即使無法有效降低測試語音中的雜訊、其對於乾淨語音的干擾程度應比其他 DAE 來的低。
  - (3) 從機器學習的角度來看，監督式學習的整體誤差有兩種：偏差值(bias)與變

異數(variance)。對於各種訊雜比的測試語料，這裡所使用的各種 DAE 其對應的變異數誤差可能大致相近，但高訊雜比 DAE 對應的偏差值誤差則應該比其他 DAE 來的小，這是因為高訊雜比 DAE 學習的是高訊雜比語料與乾淨語料的差距關係，此差距量明顯較低。

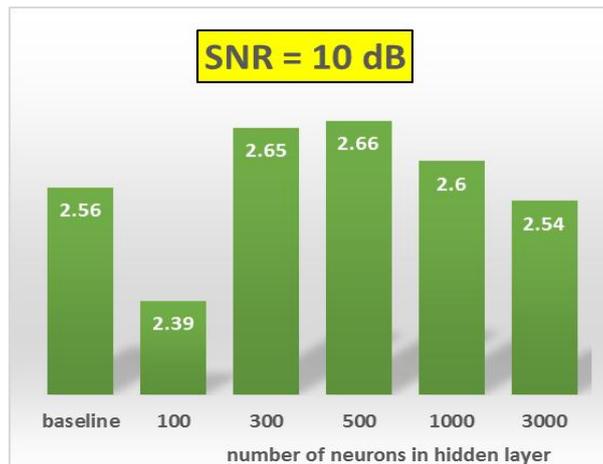
表一、多樣訊雜比訓練語料之 DAE 在各種訊雜比之測試語料上的強化結果  
(PESQ 分數)

		基準值	全訊雜比 DAE (-12 to 12 dB)	低訊雜比 DAE (-12 to -6 dB)	中訊雜比 DAE (-3 to 3 dB)	高訊雜比 DAE (6 to 12 dB)
測 試 語 音	-10dB	1.27	1.36	1.28	<b>1.40</b>	1.30
	-5dB	1.54	1.69	1.48*	<b>1.74</b>	1.73
	0dB	1.86	1.98	1.55*	1.97	<b>2.31</b>
	5dB	2.20	2.22	1.57*	2.11*	<b>2.59</b>
	10dB	2.56	2.33*	1.57*	2.14*	<b>2.76</b>
	15dB	2.90	2.38*	1.55*	2.16*	<b>2.82*</b>

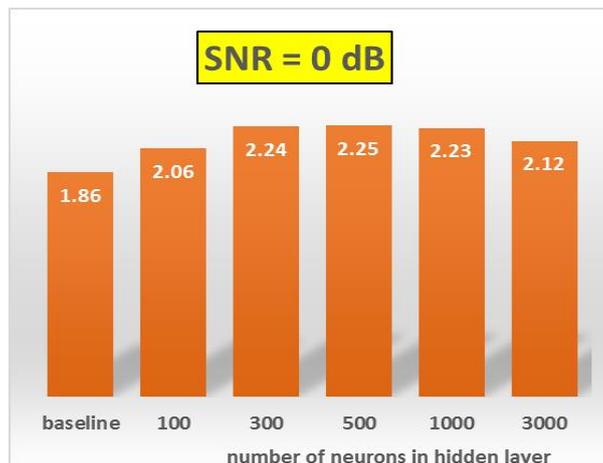
## (二)、隱藏層的神經元數量

在 DAE 架構配置上，我們為隱藏層配置的神經元數量有 100, 300, 500, 1000, 3000 五種，而如前所述，高訊雜比訓練語料的 DAE 初步呈現有最佳效能，因此這裡我們單純選擇高訊雜比語料對應之 DAE，而隱藏層的數目固定為一。

圖三與圖四分別為不同神經元數量之 DAE 在訊雜比 10 dB 與 0 dB 測試語料的語音強化結果。我們藉此配置觀察神經元數量對 DAE 效能的影響。從兩張圖可看出在神經元數量為 500 時，對應之 PESQ 值為最高、亦即有最佳的語音強化效能。若再增加神經元數量則未明顯提升 PESQ 值，當神經元數量高至 3000 時，DAE 效能甚至會降低。其中原因可能是訓練語料的不足，導致過度擬合 (overfitting) 的狀況發生。



圖三、高訊雜比訓練語料與一層隱藏層及 500 個神經元配置之 DAE 在不同神經元數量設定下對訊雜比 10 dB 測試語料的語音強化結果，數值為 PESQ 分數。



圖四、高訊雜比訓練語料與一層隱藏層及 500 個神經元配置之 DAE 在不同神經元數量設定下對訊雜比 0 dB 測試語料的語音強化結果，數值為 PESQ 分數。

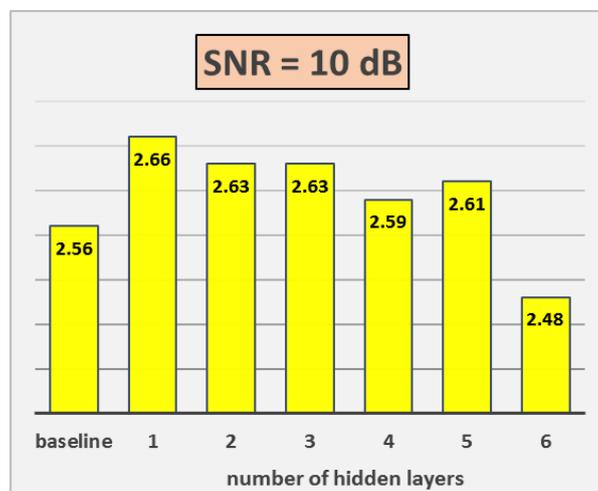
### (三)、隱藏層深度

我們在這裡探討隱藏層深度對 DAE 效能的影響，隱藏層深度設置為 1 至 6 層。如同上一節探討神經元數量的配置，我們同樣使用高訊雜比訓練語料做為不同隱藏層深度 DAE 的訓練語料，同時神經元的數量固定為 500，其在訊雜比 10 dB 與 5 dB 的測試語料之實驗結果如圖五與圖六所示。

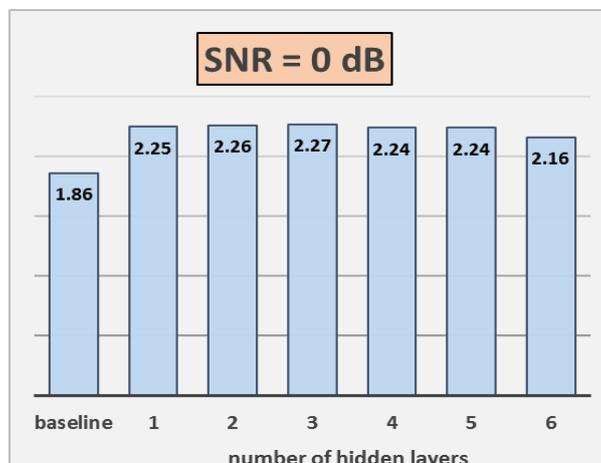
從圖五中可以觀察到，對於訊雜比 10 dB 的測試語料，最佳的結果是出現於隱藏層層數設定為 1 時，同時 DAE 的效能會隨著隱藏層的增加而降低。相對而言，

在圖六中，對於訊雜比 5 dB 的測試語料，隱藏層層數 1 至 3 的 DAE，其效能都十分接近，在隱藏層層數增加到 4 至 6 時才明顯降低。這樣的結果可能來自於我們在 DAE 中使用的啟動函數 (activation function) 為 sigmoid、其輸出為一平滑曲線。由於其平滑的輸出值，使多層隱藏層的重疊無法反映相對高波動的輸入值、且誤差隨著隱藏層的增加而不斷累積。

簡單來說，我們的實驗結果指出，一個相對淺層 (shallow) 的 DAE 擁有比深層 DAE 更好的效能。此一結果與其他研究者的實驗結果[12]有著相同的趨勢，這表示我們可以使用較低計算複雜度之 DAE 達到更好的語音強化效能。



圖五、高訊雜比訓練語料與 500 個神經元配置之 DAE 在不同隱藏層深度設定下對訊雜比 10 dB 測試語料的語音強化結果，數值為 PESQ 分數。



圖六、高訊雜比訓練語料與 500 個神經元配置之 DAE 在不同隱藏層深度設定下

對訊雜比 0 dB 測試語料的語音強化結果，數值為 PESQ 分數。

## 五、結論

在這篇論文中，我們探討了降噪自動編碼器(DAE)在不同配置的模型下，其語音強化的效能。我們發現高訊雜比訓練語料與其他訊雜比訓練語料相比，所得到的 DAE 在整體上有著最佳的語音強化表現。此外，用高訊雜比訓練語料學習之 DAE，只需具有少量的隱藏層與中量的神經元個數即可，代表其計算複雜度相對較低。在未來的工作上，我們將在雜訊語音的準備上增加雜訊的種類，觀察在雜訊種類上不匹配時，高訊雜比的 DAE 是否仍能有好的表現，同時在探討其他 DAE 架構配置對語音強化效能的影響，例如不同的啟動函數、不同的語音特徵等。

## 參考文獻 [References]

- [1] T.-S. Chi, T.-H. Lin, and C.-C. Hsu, "Spectro-temporal modulation energy based mask for robust speaker identification," *Journal of the Acoustical Society of America*, vol. 131, no. 5, pp. EL368-EL374, 2012.
- [2] T.-C. Wu, T.-S. Chi, and C.-F. Lee, "Simulations of high-frequency vocoder on Mandarin speech recognition for acoustic hearing preserved cochlear implant," in *Proceedings of Annual Conference of International Speech Communication Association (Interspeech) 2017*.
- [3] Jen-Tzung Chien and Yuan-Chu Ku, "Bayesian recurrent neural network for language modeling", *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 2, pp. 361-374, February 2016.
- [4] Jieh-weih Hung, Hsin-Ju Hsieh, Berlin Chen, "Robust speech recognition via enhancing the complex-valued acoustic spectrum in modulation domain," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 2, pp. 236-251, February 2016.

- [5] Kuan-Yu Chen, Shih-Hung Liu, Berlin Chen, Hsin-Min Wang, Hsin-Hsi Chen, "Exploring the use of unsupervised query modeling techniques for speech recognition and summarization," *Speech Communication*, Vol. 80, pp. 49-59, June 2016.
- [6] Chien-Yao Wang, Jia-Ching Wang, Andri Santoso, Chin-Chin Chiang, Chung-Hsien Wu, "Sound Event Recognition Using Auditory-Receptive- Field Binary Pattern and Hierarchical-Diving Deep Belief Network," Accepted by *IEEE/ACM Trans. Audio, Speech, and Language Processing*, 2017
- [7] Chia-Ping Chen, Yi-Chin Huang, Chung-Hsien Wu, and Kuan-De Lee, Polyglot Speech Synthesis Based on Cross-lingual Frame Selection Using Auditory and Articulatory Features, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22(10), pp. 1558-1570, 2014
- [8] W. Rix et al, "Perceptual evaluation of speech quality (PESQ) – a new method for speech quality assessment of telephone networks and codecs," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 749-752, 2001
- [9] G. Hinton and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, July 2006.
- [10] P. Vincent, H. Larochelle, Y. Bengio, and P. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th International Conference on Machine Learning (ICML-08)* 2008
- [11] X. Lu, Y. Tsao, S. Matsuda and C. Hori, "Speech enhancement based on deep denoising autoencoder," in *Proceedings of Interspeech*, pp. 436-440, 2013
- [12] M. Kim, "Collaborative deep learning for speech enhancement: A run-time model selection method using autoencoders", in *Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*
- [13] L. L. Wong, S. D. Soli, S. Liu, N. Han, and M. -W. Huang, "Development of the Mandarin hearing in noise test (MHINT)," *Ear and hearing*, 28 (2), pp. 70-74, 2007