

中文語音清晰度與訊號自函數之關聯性研究以台灣地區華語為主

A study of the Chinese speech intelligibility in relation to the autocorrelation function model: the standard phonics of Taiwanese

陳炯堯¹、詹旻憲²

1 朝陽科技大學 建築系 專任助理教授

2 朝陽科技大學 建築及都市計劃研究所 研究生

412 台中縣霧峰鄉吉峰東路 168 號

04-3323000-4349

E-mail : chychen@mail.cyut.edu.tw s9112617@mail.cyut.edu.tw

摘要

本研究之目的在建立一個適用在中文（台灣地區）使用的室內語音清晰度（speech intelligibility）測定法。進而對語音之本身特性或文化表現以計量方式來預測其影響程度。在我們先期之研究中，證實了中文語音與語音傳輸指標（speech transmission index, STI）在整體問卷實驗中有良好的相容性。

在室內語音清晰度的研究中考慮的因素應包括（1）發聲（單音節字音及語者）（2）擴聲系統（擴聲器及喇叭等）（3）建築空間特性（形狀及面材等）（4）空氣條件（溫濕度及氣流等）（5）聽者（聽力、注意力及理解力等）（6）噪音等因素。在上述因子中可由物理測量 STI 得到預測的有 2, 3, 4 及 6，但對於 1 與 5 兩者是無法預測的。本研究假設單音節之原聲樣本（source template）與空間實際錄製之空間樣本其訊號之自函數（autocorrelation function）之間之比較中，以幫助瞭解因子 1 與 5 以及它和 2, 3, 4 及 6 間的關聯性。因自函數對於訊號之時間特性有著良好的解析能力。

本研究所錄製聲場為空席狀態，背景噪音均為在 NC-40 以下，並依據先期研究結果【1】選取中文音韻錯聽實驗（Articulation Test）中較有代表性之單音節（如空韻），作為訊號之自函數演算，並依各聲場所量測之 STI 平均值與自函數各物理因子比較，來檢測上述之對應關係，並發現各聲場測點間之 τ_c 與 τ_d 值，有明顯時間點之差異性。

關鍵詞：語音清晰度、中文語音、語音傳輸指標、自函數

壹、前言

語音聽取的清晰度，不論是教室、會議室、通道以及廣場等所有的建築空間中，是有考量的必要性，特別是在災害時的避難誘導設施，對於聽力較弱和高齡者的考量最為重要。

語音清晰度是聲場品質評價標準之一。在建築空間語音清晰度預測法的研究中，人耳對於聲音的清晰度多以主觀的判斷及空間之物理評價指標間之相關性為重心。影響語音清晰度之物理因子，在長期的發展過程中已獲得不少的知識。

目前國際上製定完成的語音清晰度客觀物理量測（Indirect Method），主要分為三種類型

【2】:第一種是利用訊號及噪音的關係來評價，如構音指標(Articulation Index)。第二種是利用直接音與殘響時間的能量比來計算，如子音損失百分比(%ALcons)。第三種則是利用特定的聲壓級的噪音，內含一組經過調變後的訊號，於聽者位置量測調變後的訊號，位於聽者位置量測調變訊號的程度而導出的評估值，如利用調變轉移函數(MTF)導出語音傳輸指標(STI)及快速語言傳輸指標(RASTI)。

快速語言傳輸指標為目前評價聲場語音清晰度最常使用的客觀工具。此指標除荷蘭語以外，歐洲各國的語音清晰度之評價也都對應得相當好。但以東方語言為主之一的日文卻無法完全對應。此現象在我們的先期研究中【3】已經證實台灣地區的中文語法與 STI 之對應就整體之結果而論可說明以下幾個事實。

表 1 華語單音節錯聽評估音表

(一)台灣地區的中文語音清晰度問卷(語表，見表 1) 調查結果與預測法 STI 測量

結果之對應就空間效應而言是一致的。

(二)證實台灣地區的中文語音清晰度問卷(語表)對於聲場語音清晰度之測定是有效的。

(三)從中文語音清晰度問卷(語表)調查結果顯示對於不同空間之錯聽率就不同單音節而言有個別差異(individual difference)存在。表示影響錯聽率之因子除空間特性(STI 測量結果)外，還有多元以上之因子存在。

因此，本研究針對上述第三項內容進行後續研究計畫，並假設在傳統聲場語音清晰度之測定時所使用之物理量測 STI 可以被單音節之原聲樣本 (source template) 或空間實際錄製之空間樣本之自函數 (autocorrelation function, ACF)所替代。

由於 Ando【4】已證實可將音樂訊號之自函數 (autocorrelation function, ACF)用來預測音樂廳各席次上聽眾之心理主觀喜好產生的過程。我們也證實【5】音樂廳之聆聽的明瞭性也可由音樂訊號之自函數來預測。並且對於航空器噪音所造成的生活交談干擾也由於不同航空器噪音之自函數演算證實自函數在聲音訊號上的解析能力。因此本研究也將以中文語音單音節之原聲樣本(source template)與空間樣本(sound-field template) 作自函數演算，找出在正規自函數初期衰減的四個指標與錯聽率是否有特殊的關聯性作為此研究之目的。

第一組		第二組	
ㄉㄛˇ	poo	ㄇㄚ	Ma
ㄌㄣˋ	len	ㄉㄛˊ	Por
ㄉㄟˊ	pie	ㄉㄛˊ	'de
ㄟ	sy	ㄉㄟˊ	Bie
(一)ㄟ	ier	ㄟ	Jr
ㄅㄣˋ	bin	ㄟ	Ell
ㄑㄟˊ	quei	ㄉㄟˊ	Di
ㄉㄞˊ	tai	ㄑㄟˊ	goei
ㄉㄞˊ	dau	ㄉㄞˊ	bai
ㄌㄧㄡˊ	liu	ㄉㄞˊ	puh
ㄉㄞˊ	doou	ㄌㄧㄡˊ	neu
ㄇㄟㄠ	myau	ㄉㄞˊ	moou
ㄏㄢ	han	ㄉㄟㄠ	biau
ㄇㄣˊ	'men	ㄉㄞ	pan
ㄉㄤ	dang	ㄉㄞ	pen
ㄉㄟㄠ	ping	ㄉㄞ	bang
ㄌㄨㄥˊ	loong	ㄉㄟㄠ	bing
ㄊㄞˊ	taa	ㄉㄟㄠ	tong

貳、研究方法

一、本計畫採用之研究方法與原因

本計畫將採以計量方式觀察時時刻刻變動中的聲訊特性，以瞭解中文語音之物理特性。在錯聽實驗中得到之結果與之比較後，可預測語音之自函數(autocorrelation function, ACF)可以把錯聽率之預測方法求得。由於自函數是把時間點上前後聲能量變化情形的相似性導出，因此對於時刻變動之事件特別有效。自函數定義為：

$$\Phi(\tau) = \frac{1}{2T} \int_{-T}^{+T} f(t) f(t + \tau) dt \dots \dots \dots (1)$$

$2T$ 為自函數之積分長度

而正規化自函數為：

$$\phi(\tau) = 10 \log \frac{\Phi(\tau)}{\Phi(0)} [dB] \dots \dots \dots (2)$$

有關自函數的應用已有很多報導，如 Brown and Zhang 【6】利用樂器之樂音來計算其自函數，可將樂音之基礎頻率(fundamental frequency)肆出。Brown 【7】利用樂音的自函數來分辨音樂作曲的時期。Srodecki 【8】提出利用自函數來評估殘響衰減斜率之可能性。原理是利用反射音之先後秩序，稱時間音響擴散(temporal diffuse)來瞭解空間之幾何環境。自函數最佳之運用便是可將事件在時軸上之出現頻率、相似程度清楚地表達出來。

因此，利用自函數來觀察本計畫之先期研究中制定出之中文單音節表(表1)各中文音韻(Chinese Consonant and Vowel)將是本計畫之基本方法。首先應制定觀察計劃，在聲場中，由於麥克風收錄之語音樣本包含了背景噪音與其它因素，我們將無法把此因子排除，所以在比較中的空間樣本(sound-field template)，事實上是包括已被包含空間因素與背景噪音因素等之自函數，我們將之記號為 $\Phi(\tau)_{S+F}^x$ ，而無響室中的錄製音韻(dry source)其自函數將是 $\Phi(\tau)_s^x$ ，而 x 將是中文音韻單音節之編號，因此假設錯聽率將反應在：

$$d(x) = \frac{1}{n} E_x^n (\phi(\tau)_s^x - \phi(\tau)_{S+F}^x) \dots \dots \dots (3)$$

n 表示自函數中4個指標的編號，錯聽率將集中此四指標方可得到對單音節音韻的特性回應。

在此自函數的四大指標定義如下(圖1)：

- (一) 規一化自函數(normalized ACF)之有效遲延時間(effective duration, e)
在遲延時間軸上圖形輪廓降至初期($\tau = 0$)的1/10時點。

- (二) 規一化自函數之等效聲能，即式(2)中 $\phi(\tau) = 0$ 時的自函數計算， τ_e ，回歸是根據 $\phi(\tau) < -10$ dB 範圍內輪過廓線頂點。
- (三) 規一化自函數除 $\tau = 0$ 之外第一個圖形上的極大值， τ_1 。
- (四) 規一化自函數於 τ_1 處之時間遲延， τ_1 。

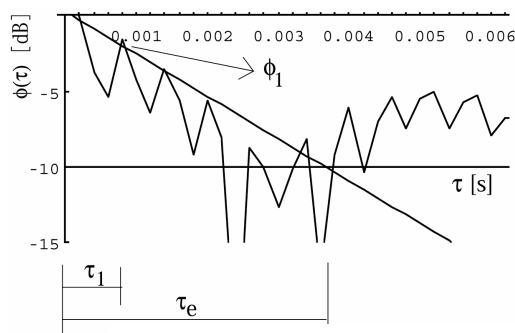


圖 1 自函數之特徵指標之定義

此處由於要得到時刻之聲訊變動情報，在樣本被 A/D 轉換後將進行連續自函數(running ACF)演算，其積分長度 $2T$ 將設定為 30 毫秒(ms)，而連續間隔(running step)設定為 5 毫秒。所以式(3)可變成為下式：

$$c(x) = \frac{1}{N} \prod_{n=1}^4 (\log \phi(\tau)_s^x - \log \phi(\tau)_{s+F}^x) \dots \dots (4)$$

N 表示每個單音節可能連續演算之次數

二、聲場語音清晰度之試驗規劃

(一) 原聲樣本量測方法

於無響室中錄製量測所需男、女生發音之語音樣本(dry-source)，作為進行語音清晰度檢測實驗之語音訊號來源。錄音前告知發音者(男一員為教師；女一員曾擔任電台播音員)以標準發音的方式穩定唸出評估音表之內容，錄製完成後，進行語音之自函數運算。

(二) 空間樣本量測方法

本研究使用之物理量測系統如(圖 2)所示，用以紀錄空間樣本。藉由先前研究之檢測音表【2】，作為進行發音之語音樣本 (dry-source)，它是由數位錄音機(TEAC R-9)透過揚聲器(Martin audio EM-76)於聲場發音位置，固定面向於座席中軸線(指向角= $150^\circ H \times 80^\circ V$)發送語音訊號，而在測點上有接收麥克風(B & K, Type 4007)其高度為模擬聽眾

在座席上耳朵的位置 (距樓板面 120cm)，語音訊號再由數位錄音機 (SONY PC-2) 錄製完成後，進行語音之自函數運算。

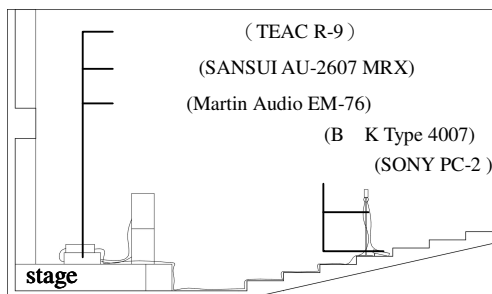
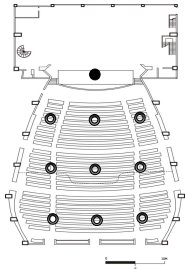
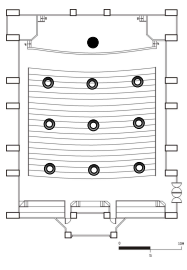
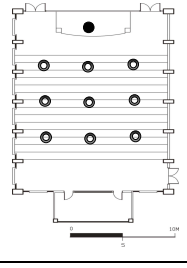
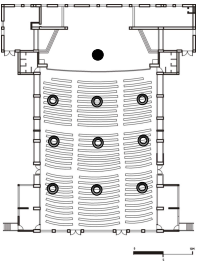
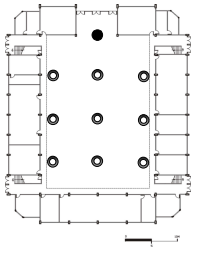
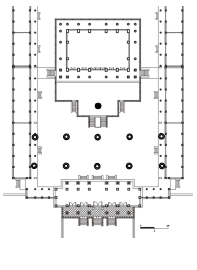


圖 客觀物理量測系統示意圖

(三) 各聲場之建築基本 與測點分
平均值

場所	測試基本		測點分 圖
台中中	室容積(m^3)	14613.1	
	S/N ratio (dB)	5.33	
	RASTI	0.68	
	測點分 劃 說明	此聲場為大型多能之演廳，在座席面積中，平均每 $30 m^2$ 分一測點。	
朝陽科大	室容積(m^3)	5194.8	
	S/N ratio (dB)	5.12	
	RASTI	0.67	
	測點分 劃 說明	在此聲場中，主要以演、表演活動為主，在座席中，平均每 $25 m^2$ 設置一測點。	
朝陽科大演廳	室容積(m^3)	2034.7	
	S/N ratio (dB)	2.39	
	RASTI	0.86	
	測點分 劃 說明	提學演活動之場所，在座席區中，每 $12 m^2$ 平均分一點。	
台中子場所	室容積(m^3)	13061.9	
	S/N ratio (dB)	6.32	
	RASTI	0.71	
	測點分 劃 說明	以表演為主之多能集會，在座席區中，每 $25 m^2$ 平均分一點。	
台中子場所	室容積(m^3)	19237.2	
	S/N ratio (dB)	-1.06	
	RASTI	0.46	
	測點分 劃 說明	提類運動之多能體，每 $30 m^2$ 平均分一點於場內。	
台中子場所	廣場面積(m^2)	863.7	
	S/N ratio (dB)	-3.51	
	RASTI	0.39	
	測點分 劃 說明	此外聲場將聲源置於台上，測點則平均每 $8 m^2$ 設置一點，分於廣場與台上，如圖所示。	

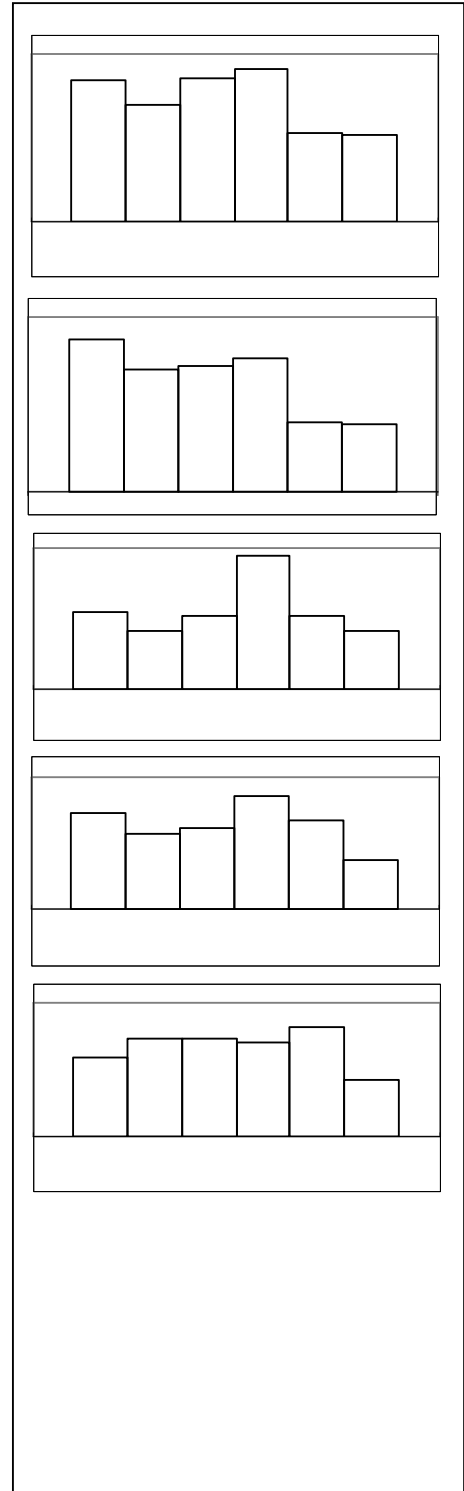
、分析結果

由上述之研究方法，我們將錄製完成之空間樣本（朝陽科技大學設計大樓、朝陽科技大學教學大樓演廳、台中中、中中會、中少年活動中心、台中）74個測點之連續自函數的平均值計算完成，並將自函數的四大指標之結果一與語音清晰度討論其相關之對應程度。其結果如（圖3）顯示，六個聲場之空間樣本於自函數的四大指標中之 τ_1 、 τ_e 及 τ_2 值分均與 intelligibility Score（）有良好的相關性，以中中會有相當特殊之一性，其語音清晰度值最高，其中以均能聲量為代表的因子指標 τ_2 值為最佳（r 值為 0.93； R^2 值為 0.86），其結果與語音清晰度是相反的向，且也獲得良好的相關，是以初期能量為代表之中，其意義指示比語音在時間軸上似程度的高，即同頻率者連續不斷出現造成的結果。 τ_1 則較不理（r 值為 0.22； R^2 值為 0.05）。

而各聲場之自函數的四大指標，為 τ_1 、 τ_e 及 τ_2 值分與 RASTI 值有良好的相關之相關係數（r 值）於 0.97~0.78，由此可知，可判定三個自函數指標與 RASTI 值之變數之間向正相關線性分情形，且在單判定係數（ R^2 ）值分數於 0.61~0.93。由此可知變數間之關係可度高。

上述所分析之結果顯示，已有反應出有長自函數初期分之有效性演無。

一方面依字音空韻及空韻與聲場之自函數的四大指標相對應而言（圖5），有 τ_1 （ $T=5.3$ ， $p < 0.05$ ）及 τ_2 值（ $T=9.61$ ， $p < 0.05$ ）有明顯之差異。顯示各聲場語音之單音節自函數的運算結果，可由時時刻刻變動中的聲訊特性，發現其差異性，且實語音清晰度的評價根據，除空間物理因子影響之外，還有其固定之字音發音特性因子存在。



ERROR: rangecheck
OFFENDING COMMAND: .buildcmap

STACK:

-dictionary-
/WinCharSetFFFF-VTT491A9C96t
/CMap
-dictionary-
/WinCharSetFFFF-VTT491A9C96t