語音資料增量技術應用於構音障礙輔具之效益

Data Augmentation Technology for Dysarthria Assistive Systems

朱唯中 Wei-Chung Chu
國立陽明交通大學生物醫學工程學系
Department of Biomedical Engineering
National Yang Ming Chiao Tung University
abc55169@gmail.com

洪瑛秀 Ying-Hsiu Hung
國立陽明交通大學生物醫學工程學系
Department of Biomedical Engineering
National Yang Ming Chiao Tung University
wer850718906@gmail.com

鄭惟中 Wei-Zhong Zheng
國立陽明交通大學生物醫學工程學系
Department of Biomedical Engineering
National Yang Ming Chiao Tung University
s1010654@gm.ym.edu.tw

賴穎暉 Ying-Hui Lai
國立陽明交通大學生物醫學工程學系
Department of Biomedical Engineering
National Yang Ming Chiao Tung University
yh.lai@nycu.edu.tw

摘要

以語音驅動之溝通輔具是構音障礙患者常用的方法之一。然而這類輔具需要患者大量的錄製語音來提升系統效益，而常造成使用上的困難。有鑑於此，本研究提出語音增量技術來試圖減少患者錄音負擔並提升溝通輔具辨識效益。於研究結果證明，所提出之構音障礙語音生成系統能產生類患者語音並提升溝通輔具於重複語句之辨識率。此外，於患者 Free-talk 情況下的字詞錯誤率可由 64.42%降至 4.39%。而這些成果也證實本論文提出之方法將對溝通輔具發展有所幫助。

Abstract

Voice-driven communication aids are one of the methods commonly used by patients with dysarthria. However, this type of assistive devices demands a large amount of voice data from patients to increase the effectiveness. In the meantime, this will sink patients into an overwhelming recording burden. Due to those difficulties, this research proposes a voice augmentation system to conquer the aforementioned concern. Furthermore, the system can improve the recognition efficiency. The results of this research reveal that the proposed speech generator system for dysarthria can launch corpus to be more similarities to the patient's speech. Moreover, the recognition rate, in duplicate sentences, has been improved and promoted to the higher level. The word error rate can be reduced from 64.42% to 4.39% in the case of patients with Free-talk. According to these results, our proposed system can provide more reliable and helpful technique for the development of communication aids.

關鍵字：構音障礙、溝通障礙系統、資料增量、深度學習
Keywords: dysarthria, communication assistance system, data augmentation, deep learning

一、 緒論

構音障礙為腦部(或神經)受損而導致患者無法良好的控制發聲肌肉而影響語音清晰度。根據美國語言學學會(2020)的資料顯示，全美大約有 4000 萬位溝通障礙患者，且每年還約有 100 萬人因罹患疾病而可能導致構音障礙問題。有鑑於此，我們需要投入更多研究資源來幫助患者有更好的溝通效益，進而提升他們的生活品質。

對於構音障礙患者來說，溝通輔具系統(augmentative and alternative communication, AAC)是提升溝通效率的常見方法。目前常見的 AAC 包括：溝通字板(Calculator et al., 1983)、眼動追蹤(Lin et al., 2006)等。但上述 AAC 溝通速率(約每分鐘 2~5 個字)仍遠不及透過言語驅動的方法 (約每分鐘 127±46 個字)(Murdoch & Theodoros, 2001)。換言之，以語音驅動的溝通輔具應是更有效率的方式。基於此概念並伴隨著語音訊號處理技術發展下，目前已有許多方法被提出，例如：語音轉換(voice conversion, VC)及自動語音辨識(automatic speech recognition, ASR)，來試圖改善構音障礙患者的溝通效益。

以 VC 技術用於構音障礙語音轉換來說，Yang et al.(2020)使用生成對抗網路 cycle-consistent generative adversarial network (cycle-GAN)，將構音障礙語音轉換成正常語者語音。於實驗結果證明，此方法可降低患者語音達 33.4%的字詞錯誤率(word error rate, WER)。Wang et al.(2020)提出結合文字轉語音(text-to-speech, TTS)與 knowledge distillation (KD)技術的 end-to-end VC 方法來試圖改善患者語音清析度。於實驗結果發現，此方法能讓較嚴重之構音異常患者分別降低 35.4%和 48.7%的 WER。

另一部份以 ASR 為基礎之溝通輔具也持續發展中。例如 Shor et al. (2019)提出在有限的構音障礙語音資料條件仍有準確的構音障礙語音辨識能力。於此研究中，他們先用巨量資料(1000 小時正常語者資料)來預訓練 RNN-T 模型。接著，再用 36.7 小時構音障礙語者資料進行微調(finetune)。於實驗結果顯示，在輕度和重度患者上分別降低 22.3%和 38%的 WER。Takashima et al. (2019) 透過遷移學習(transfer learning)技術來善用不同語言資料對於語音辨識模型系統之效益提升。從實驗結果也證明能使 ASR 的音素錯誤率降低(從 38.49%降至 25.69%)。

上述的多項研究顯示現今的語音訊號處理模型可以有效提升構音障礙語音的理解度，但是模型的強健度往往會受到訓練資料量和品質影響。然而對於構音障礙患者來說，要大量錄得訓練語料將十分困難且花費成本。有鑑於此，資料增量技術將顯得十分重要。目前已有許多研究嘗試使用資料增量方法來克服資料不足的問題。舉例來說， Vachhani et al. (2018)藉由調整語音速度和節奏，將正常人語音的速度調整至類患者語速作為增量資料。隨後，再將其丟入 ASR 進行訓練。由結果證明，透過提出方法能提升 ASR 系統辨識率約 3%。但在此研究中仍呈現出當患者資料產生不穩定音素情況，其生成之資料的效果仍有待加強。而 Jiao et al. (2018)使用 deep convolutional generative adversarial network 技術來設計語音轉換模型，進而將正常人語音轉換成患者語音作為增量資料。此外，他們再使用二分類網路模型來比較傳統混噪增量和上述語音轉換增量技術間的差異。於結果顯示，使用語音轉換系統作為增量手法明顯優於混噪增量方法來幫助模型學習。然而上述這些典型的資料增量技術仍無法產生高維之訓練資料分布特性，進而難以更廣泛的模擬出患者日常說出語音可能的變異性，使得以語音驅動為基礎之溝通輔具效益受到限制。

基於上述的研究結果和討論，我們假設一個具備多語者轉換能力的模型在資料增量任務中，將更能模擬出更多元之患者語音變異性，進而提升溝通輔具採用模型效益。有鑑於此，本論文將提出以多語者轉換模型為基礎之構音障礙語音生成系統，並探討生成類患者語料的品質與相似度。隨後，我們將更進一步的探討提出方法所增量之資料對於溝通輔具系統模型訓練之效益。

二、 構音障礙語音生成系統

鑒於上述討論，我們提出一個以多對多為基礎之語音轉換技術來設計構音障礙語音生成系統，稱構音障礙語音生成器(dysarthric

speech generator, DSG)。於本論文提出之DSG(如圖 1)主要是使用 StarGAN-VC (Kameoka et al., 2018)作為核心模型架構，透過此模型特性來學習多位語者下轉到患者語音之模型參數，進而生成出更廣泛之患者語音特性(例如：不同語氣和語速)。此外，此提出之 DSG 將僅需患者錄製少量訓練語料(約 288 句話)來學習如何將正常語者轉換成構音障礙患者之語音，接著透過此模型來將大量正常語者語音轉成類患者語音。換言之，我們透過訓練出來之模型來將大量正常語者語音轉成大量類患者語音訓練資料，進而減少患者錄音的負擔。

StarGAN-VC 是由二類別轉換 CycleGAN-VC (Kaneko & Kameoka, 2018)所延伸出來的多類別轉換架構，在二類別轉換(CycleGAN)中若要生成k個語者類別就需要k個模型，而使用多類別轉換(StarGAN)只需要一個模型，便能大幅減少模型的訓練數目。此外，多語者特徵所訓練出的模型更具多變性，對於因為缺乏患者資料而缺少變異性的問題來說非常有幫助。StarGAN-VC 主要由三個模型組成，生成器(Generator, G)、鑑別器(Discriminator, D)和輔助分類器(Classifier, C)，其架構如圖 1 所示。x為輸入頻譜、c為語者標記類別、y為目標頻譜，生成器(G)會根據輸入的x和c輸出預測頻譜$\hat{y}$，鑑別器(D)則根據輸入的$\hat{y}$和y輸出兩者之間的相似性；輔助分類器(C)根據輸入的$\hat{y}$輸出預測類別c'。三個模型所對應的損失函數(loss function)分別為$\mathcal{L}_G(G)$、$\mathcal{L}_D(D)$和$\mathcal{L}_C(C)$，公式如下：

$$\mathcal{L}_G(G) = \mathcal{L}_{adv}^G(G) + \lambda_{cls}\mathcal{L}_{cls}^G(G) + \lambda_{cyc}\mathcal{L}_{cyc}(G) + \lambda_{id}\mathcal{L}_{id}(G) \quad (1)$$

$$\mathcal{L}_D(D) = \mathcal{L}_{adv}^D(D) \quad (2)$$

$$\mathcal{L}_c(C) = \mathcal{L}_{cls}^C(C) \quad (3)$$

$\mathcal{L}_{adv}^G(G)$和$\mathcal{L}_{adv}^D(D)$為對抗損失(adversarial loss)，藉由對抗式訓練讓模型學習x和y之間的特徵差異，使得$G(x)$的特徵趨近於y；$\mathcal{L}_{cls}^G(G)$和$\mathcal{L}_{cls}^C(C)$為類別分類器損失(domain classification loss)，$\mathcal{L}_{cls}^G(G)$越小表示分類器C對於生成器所輸出語者c的頻譜$G(x,c)$的準確度越高，$\mathcal{L}_{cls}^C(C)$越小表示分類器C對於目標語者的頻譜y的準確度越高；$\mathcal{L}_{cyc}(G)$為循環一致損失(cycle consistency loss)，確保生成器建構(construct)之後能重建(reconstruct)回語音頻譜，

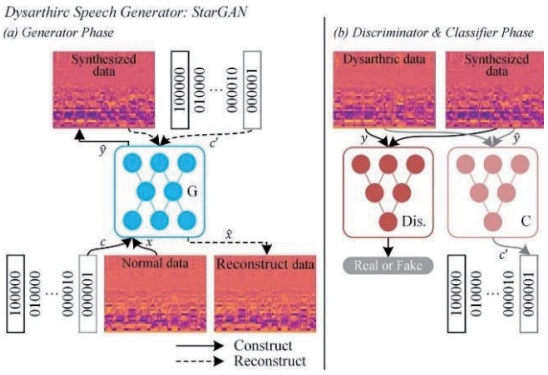


圖 1. StarGAN-VC 架構圖

強化不同語者生成器之間的特徵轉換同時也保留語音訊息。$\mathcal{L}_{id}(G)$為身分映射損失(identity loss)，當生成器(G)輸入頻譜x和其目標類別c為同一類別時，使其輸出保持不變，強化生成器(G)的語者特性。λ_{cls}、λ_{cyc}、λ_{id}為三個損失函數的權重值，本實驗的參數設定為 3、10、5。三個模型皆使用二維 gated CNN (Dauphin et al., 2017)結構，此結構中利用堆疊 sigmoid 門控單元(gated linear units, GLU)而有序列記憶的架構，並且在語言轉換任務上比起 long short-term memory (LSTM)架構有更精簡、好訓練、推算速度快等優點。

當上述轉換模型訓練完成後，我們還需要 vocoder 把聲學特徵轉換為聲音訊號。本實驗中，使用 WORLD (Morise et al., 2016)作為聲學特徵，其中包含三種特徵：基頻$(F0)$、頻譜包絡線(spectral envelope, SP)、非週期參數(aperiodic, AP)。在訓練階段時，使用頻譜包絡線(SP)作為 StarGAN-VC 模型的訓練資料。在轉換階段時除了用 StarGAN-VC 轉換SP，我們也將基頻$(F0)$和非週期參數(AP)用正規化線性映射轉換成目標語者的資料分布，使音調更接近目標語者。

三、 研究方法

3.1 實驗材料

我們採用 TMHINT (Huang et al., 2005)文本(共 320 句，每句 10 字)來對患者進行錄音，錄音的設定為：取樣率 16k、位元率 16bit、單聲道、wav 檔案格式。本研究共有三位構音障礙語者$(D_n, n = 3)$和三位正常語者$(N_m, m = 3)$，來錄製至少一套 TMHINT 語料(詳細資料如表 1)。實驗中我們也再使用 Microsoft Speech API

(Wikipedia contributors)中的 TTS 來產生輸入文本[1]之語料，進而節省患者大量錄音時間。

表 1. 資料使用表

	訓練資料數量與合成資料數量	
原始語料總數	D_1 (男，中風)=320 句×2 套 D_2 (女，腦性麻痺)=320 句×2 套 D_3 (女，聽力損失)=320 句×1 套 N_1 (男)=320 句×2 套 N_2 (女)=320 句×2 套 N_3 (女)=320 句×1 套 TTS (女)=320 句×1 套	
訓練語料	D_{1_2}=288 句 D_{2_2}=288 句 D_3=288 句	N_{1_2}=288 句 N_{2_2}=288 句 N_3=288 句 TTS=288 句
合成語料	N_{1_2} ➔ D_n=320 句 N_{2_1}, N_{2_2} ➔ D_n=640 句 N_3 ➔ D_n=320 句 TTS ➔ D_n=News:2880 句	
測試語料	重複語句測試 (Duplicate test) D_{n_1} =288 句	
	外部測試 (Outside test) D_n =32 句	

註：D_{n_b}、N_{m_a}之n、m為不同語者，a、b為各語者的第幾套語料。

3.2 實驗設計

本研究主要目的為設計一個構音障礙語音生成系統並透過以 ASR 為基礎之溝通輔具進行效益驗證。有鑑於此目標，我們透過以下實驗來證明提出系統的效益。於實驗一中，我們比較「雙語者轉換模型」和「多語者轉換模型」間所合成出之語音與構音障礙患者語音的相似度情況。我們使用 Mel-cepstral distortion (MCD)[2] (Kominek et al., 2008)來評估 CycleGAN-VC 和 StarGAN-VC 二個方法之效益。隨後，我們透過實驗二來對表現較佳之 DSG 來進行系統的實作，其完整實驗流程如圖 2 所示。

圖 2 左側為 DSG 之流程，使用的訓練語料為每位語者中同一套 TMHINT 288 句(如表 1)。在完成訓練後，我們用其生成了不同數量的患者語料。並在後續效益驗證中，使用不同倍數之增量資料各別訓練多個 ASR，觀

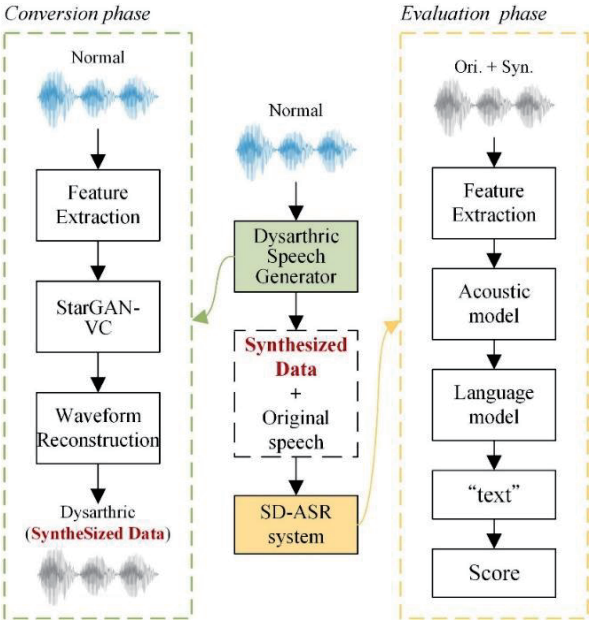


圖 2.實驗二流程圖

察增量資料對其辨識率之影響。我們使用 Kaldi ASR (Povey et al., 2011)作為驗證工具並建立每位構音障礙患者專用的 ASR，稱為 speaker dependent-ASR (SD-ASR)。而此 SD-ASR 系統是由聲學模型和語言模型組成(如圖 2 右側)，聲學模型主要學習將聲學特徵轉換為語言特徵，我們使用 time delay neural network (TDNN) (Peddinti et al., 2015)作為模型架構，聲學特徵為 MFCC，語言特徵為音素後驗機率 (phonetic posteriorgrams, PPGs) (Hazen et al., 2009)。而語言模型使用 N-gram 語言概率模型，它是一種基於馬爾可夫鏈的機率統計模型，藉由統計方法推算可以在音素序列中排列出最合適的文字。在本研究使用的 N-gram 為三元模型(N=3)。

隨後我們更進一步比較，在 SD-ASR 中加入透過 DSG 轉換的合成語料與直接加入原始的正常人語音(意即未經過 DSG 轉換的正常語者語料)之差異。因為在構音障礙 ASR 系統中加入多位正常語者資料是經典的增量手法，因此本研究希望從實驗中證實合成語料比其它資料增量方法更能幫助 ASR 系統的學習。在實驗二使用之增量資料以$n\hat{Y}$、$n\widehat{W}$表示其增量資料及類別，n表示增量套數(以 TMHINT

[1] TTS 系統採用之文本為網路新聞(共 2880 句)做為 TTS 增量文本。

[2] MCD 值越小表示合成出的語音與患者越相似。

288 句為一套)，$\hat{Y}$表示真實錄音相關語料、$\widehat{W}$表示 TTS 相關語料。

四、結果與討論

實驗一結果如表 2 所示，在三位構音障礙患者語料分別使用 CycleGAN 和 StarGAN 架構建立語音轉換模型，再使用模型各別生成D_1、D_2、D_3構音障礙語音。可以觀察到在三位患者中，有兩位患者的 StarGAN 合成語料 MCD 數值較 CycleGAN 低，意即整體平均表現來看，StarGAN 的架構所轉換出之語音與患者語音較為相似。且在此實驗設計下要生成相同數量的語料，使用 CycleGAN 要比 StarGAN 多上 6 倍的模型參數和訓練時間。有鑑於上述觀察，不論是在訓練時間以及合成資料品質上，採用 StarGAN 架構做為 DSG 模型生成合成語料將是較具潛力的方法。

表 2. CycleGAN 與 StarGAN 合成語料之 MCD

MCD results of D_1 synthesized data				
	$N_1 \rightarrow D_1$(320)	$N_2 \rightarrow D_1$(640)	$N_3 \rightarrow D_1$(320)	Average
StarGAN	0.814±0.06	0.891±0.07	0.913±0.07	**0.872**
CycleGAN	0.779±0.05	1.064±0.16	1.03±0.13	0.958
MCD results of D_2 synthesized data				
	$N_1 \rightarrow D_2$ (320)	$N_2 \rightarrow D_2$ (640)	$N_3 \rightarrow D_2$(320)	Average
StarGAN	1.311±0.16	1.324±0.15	1.240±0.14	**1.292**
CycleGAN	1.516±0.50	1.348±0.19	1.168±0.17	1.344
MCD results of D_3 synthesized data				
	$N_1 \rightarrow D_3$ (320)	$N_2 \rightarrow D_3$ (640)	$N_3 \rightarrow D_3$(320)	Average
StarGAN	1.064±0.21	1.008±0.24	1.178±0.23	1.083
CycleGAN	1.039±0.19	0.974±0.18	1.069±0.15	**1.027**

接著在實驗二中的實作結果如圖 3、圖 4 所示，我們使用兩位構音障礙語者D_1、D_2訓練個人語音辨識器 SD-ASR 來驗證所提出 DSG 是否能提升溝通輔助系統的語音辨識度，並且使用字詞錯誤率(character error rate, CER)來做為評估指標。在測試資料中分為重複語句測試(duplicate test)和半外部測試(half outside test)。Duplicate test 為患者重複語句(TMHINT 語料 288 句)，資料未參與 ASR 模型訓練。Half outside test 測試語句為患者的 TMHINT 語料 32 句，資料未參與 StarGAN 訓練，也沒有放入 ASR 模型訓練，但和語料N有關。在 half outside test 結果中，當訓練資料從原始(original)資料量到增加額外 2 倍合成資料時，可將 CER 從 60%左右降至 13%左右。且後續的 4 倍、9 倍增量也都能持續降低錯誤率，最終增量至 14 倍合成資料時，分別在患者D_1、D_2的 Half outside test 中得到 3.76%與 5.02% CER。

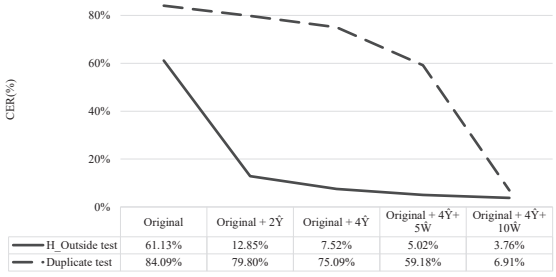


圖 3.實驗二中患者D_1之增量測試結果

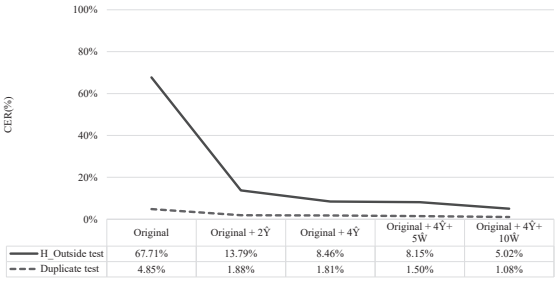
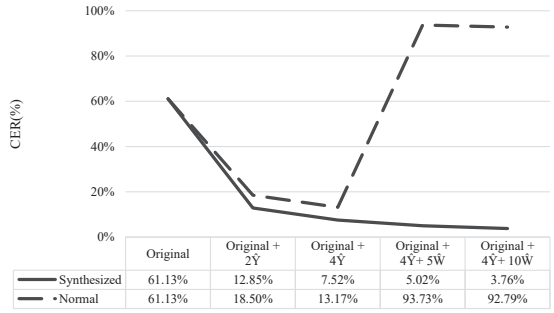


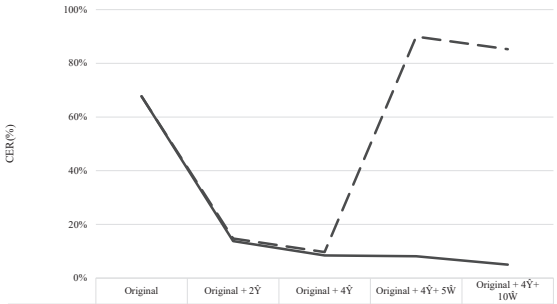
圖 4. 實驗二中患者D_2之增量測試結果

隨後，實驗二更進一步探討所提出 DSG 系統是否優於常見之增量手法(使用正常人語料作為增量資料)，而在 half outside test 中的結果如圖 5、圖 6 所示。其中可以觀察到，前 4 倍增量($4\hat{Y}$)在兩種增量系統中皆能幫助 ASR 降低錯誤率，但使用 DSG 方法仍比直接加入正常語料的 CER 更低一些。而再加入 5~10 倍的 TTS 語料($5\,\widehat{W}$、$10\,\widehat{W}$)，則可以從兩位語者的結果中發現，未轉換成合成語料的 TTS 會讓測試語料的 CER 有大幅度的上升現象。這是因為大量的 TTS 語料使 ASR 系統的辨識整體的偏向辨識 TTS。這也表示當加入的增量語料若未轉換成合成語料，一旦資料量增加到超越患者的原始語料量許多，會使系統辨識產生偏移。而正常人的發音與患者有著巨大的差異，若未將轉換之合成語料用於 ASR 系統，將導致系統辨識時把患者的發音誤判成正常人的發音，進而無法分類正確的音素，所以錯誤率也隨之大幅提升。而上述的實驗也再進一步的證明本論文提出之方法的效益。



	Original	Original + 2Ŷ	Original + 4Ŷ	Original + 4Ŷ+ 5Ŵ	Original + 4Ŷ+ 10Ŵ
Synthesized	61.13%	12.85%	7.52%	5.02%	3.76%
•Normal	61.13%	18.50%	13.17%	93.73%	92.79%

圖 5. 實驗二中，比較患者D_1使用合成語料與正常語料增量差異



	Original	Original + 2Ŷ	Original + 4Ŷ	Original + 4Ŷ+ 5Ŵ	Original + 4Ŷ+ 10Ŵ
Synthesized	67.71%	13.79%	8.46%	8.15%	5.02%
•Normal	67.71%	14.73%	9.72%	89.97%	85.27%

圖 6. 實驗二中，比較患者D_2使用合成語料與正常語料增量差異

五、 結論

本論文主要目的為探討資料增量技術對於以語音驅動為基礎之構音障礙溝通輔具的效益。於實驗結果發現，我們所提出之 DSG 系統所生成之大量患者語料能有效的提升構音障礙溝通輔具的效益。由實驗中我們也證明多語者轉換系統在生成語音質量上優於雙語者轉換技術，並且在訓練成本上有巨大的優勢。而在實驗中也證明了增量系統 DSG 應用在 SD-ASR 上可以明顯的降低 CER，且隨著增量句數的提升可以使錯誤率持續下降。此外，比起僅加入多位正常語者來訓練模型之增量方法，本論文所提出之增量系統更能使 SD-ASR 更專注辨識單一語者。有鑑於本論文之成果，未來我們將基於 DSG 資料增量方式來生成更大量語料，進而期望幫助患者在 Free-talk 情況下有更佳之辨識效益。

Acknowledgments

本研究由科技部射月計畫(MOST110-2218-E-A49A-501)及宇康生科股份有限公司計畫(YM109J052) 支持，特此致謝。

References

American Speech-Language-Hearing Association. (2020, November 02). *Quick Facts About ASHA*. https://www.asha.org/about/press-room/quick-facts/

Calculator, S., Luchko, C. D. A. J. J. o. s., & Disorders, H. (1983). Evaluating the effectiveness of a communication board training program. *48*(2), 185-191.

Dauphin, Y. N., Fan, A., Auli, M., & Grangier, D. (2017). Language modeling with gated convolutional networks. International conference on machine learning,

Hazen, T. J., Shen, W., & White, C. (2009). Query-by-example spoken term detection using phonetic posteriorgram templates. 2009 IEEE Workshop on Automatic Speech Recognition & Understanding,

Huang, M. J. D. o. s. l. p., audiology, N. T. U. o. N., & science, H. (2005). Development of taiwan mandarin hearing in noise test.

Jiao, Y., Tu, M., Berisha, V., & Liss, J. (2018). Simulating dysarthric speech for training data augmentation in clinical speech applications. 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP),

Kameoka, H., Kaneko, T., Tanaka, K., & Hojo, N. (2018). Stargan-vc: Non-parallel many-to-many voice conversion using star generative adversarial networks. 2018 IEEE Spoken Language Technology Workshop (SLT),

Kaneko, T., & Kameoka, H. (2018). Cyclegan-vc: Non-parallel voice conversion using cycle-consistent adversarial networks. 2018 26th European Signal Processing Conference (EUSIPCO),

Kominek, J., Schultz, T., & Black, A. W. (2008). Synthesizer voice quality of new languages calibrated with mean mel cepstral distortion. Spoken Languages Technologies for Under-Resourced Languages,

Lin, C.-S., Ho, C.-W., Chen, W.-C., Chiu, C.-C., & Yeh, M.-S. J. O. A. (2006). Powered wheelchair controlled by eye-tracking system. *36*.

Morise, M., Yokomori, F., Ozawa, K. J. I. T. o. I., & Systems. (2016). WORLD: a vocoder-based high-quality speech synthesis system for real-time applications. *99*(7), 1877-1884.

Murdoch, B. E., & Theodoros, D. G. (2001). Traumatic brain injury: Associated speech, language, and swallowing disorders.

Peddinti, V., Povey, D., & Khudanpur, S. (2015). A time delay neural network architecture for efficient modeling of long temporal contexts. Sixteenth annual conference of the international speech communication association,

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., . . . Schwarz, P. (2011). The Kaldi speech recognition toolkit. IEEE 2011 workshop on automatic speech recognition and understanding,

Shor, J., Emanuel, D., Lang, O., Tuval, O., Brenner, M., Cattiau, J., . . . Nollstadt, M. (2019). Personalizing ASR for Dysarthric and Accented Speech with Limited Data. *arXiv preprint arXiv:1907.13511*.

Takashima, Y., Takashima, R., Takiguchi, T., & Ariki, Y. (2019). Knowledge transferability between the speech data of persons with dysarthria speaking different languages for dysarthric speech recognition. *IEEE Access*, *7*, 164320-164326.

Vachhani, B., Bhat, C., & Kopparapu, S. K. (2018). Data Augmentation Using Healthy Speech for Dysarthric Speech Recognition. Interspeech,

Wang, D., Yu, J., Wu, X., Liu, S., Sun, L., Liu, X., & Meng, H. (2020). End-to-end voice conversion via cross-modal knowledge distillation for dysarthric speech reconstruction. ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),

Wikipedia contributors. (27 July 2021 19:52 UTC). *Microsoft Speech API*. https://en.wikipedia.org/w/index.php?title=Microsoft_Speech_API&oldid=1035806404

Yang, S. H., & Chung, M. (2020). Improving dysarthric speech intelligibility using cycle-consistent adversarial training. *arXiv preprint arXiv:2001.04260*.