

Original Paper

Improved Environment-Aware–Based Noise Reduction System for Cochlear Implant Users Based on a Knowledge Transfer Approach: Development and Usability Study

Lieber Po-Hung Li^{1,2,3,4}, MD, PhD; Ji-Yan Han⁵, BSc; Wei-Zhong Zheng⁵, MS; Ren-Jie Huang⁵, MS; Ying-Hui Lai⁵, PhD

¹Department of Otolaryngology, Cheng Hsin General Hospital, Taipei, Taiwan

²Faculty of Medicine, Institute of Brain Science, National Yang Ming Chiao Tung University, Taipei, Taiwan

³Department of Medical Research, China Medical University Hospital, China Medical University, Taichung, Taiwan

⁴Department of Speech Language Pathology and Audiology, College of Health Technology, National Taipei University of Nursing and Health Sciences, Taipei, Taiwan

⁵Department of Biomedical Engineering, National Yang Ming Chiao Tung University, Taipei, Taiwan

Corresponding Author:

Ying-Hui Lai, PhD

Department of Biomedical Engineering

National Yang Ming Chiao Tung University

No 155, Sec 2, Linong Street

Taipei, 112

Taiwan

Phone: 886 228267021

Fax: 886 228210847

Email: yh.lai@nycu.edu.tw

Abstract

Background: Cochlear implant technology is a well-known approach to help deaf individuals hear speech again and can improve speech intelligibility in quiet conditions; however, it still has room for improvement in noisy conditions. More recently, it has been proven that deep learning–based noise reduction, such as noise classification and deep denoising autoencoder (NC+DDAE), can benefit the intelligibility performance of patients with cochlear implants compared to classical noise reduction algorithms.

Objective: Following the successful implementation of the NC+DDAE model in our previous study, this study aimed to propose an advanced noise reduction system using knowledge transfer technology, called NC+DDAE_T; examine the proposed NC+DDAE_T noise reduction system using objective evaluations and subjective listening tests; and investigate which layer substitution of the knowledge transfer technology in the NC+DDAE_T noise reduction system provides the best outcome.

Methods: The knowledge transfer technology was adopted to reduce the number of parameters of the NC+DDAE_T compared with the NC+DDAE. We investigated which layer should be substituted using short-time objective intelligibility and perceptual evaluation of speech quality scores as well as *t*-distributed stochastic neighbor embedding to visualize the features in each model layer. Moreover, we enrolled 10 cochlear implant users for listening tests to evaluate the benefits of the newly developed NC+DDAE_T.

Results: The experimental results showed that substituting the middle layer (ie, the second layer in this study) of the noise-independent DDAE (NI-DDAE) model achieved the best performance gain regarding short-time objective intelligibility and perceptual evaluation of speech quality scores. Therefore, the parameters of layer 3 in the NI-DDAE were chosen to be replaced, thereby establishing the NC+DDAE_T. Both objective and listening test results showed that the proposed NC+DDAE_T noise reduction system achieved similar performances compared with the previous NC+DDAE in several noisy test conditions. However, the proposed NC+DDAE_T only required a quarter of the number of parameters compared to the NC+DDAE.

Conclusions: This study demonstrated that knowledge transfer technology can help reduce the number of parameters in an NC+DDAE while keeping similar performance rates. This suggests that the proposed NC+DDAE_T model may reduce the implementation costs of this noise reduction system and provide more benefits for cochlear implant users.

(*J Med Internet Res* 2021;23(10):e25460) doi: [10.2196/25460](https://doi.org/10.2196/25460)

KEYWORDS

cochlear implants; noise reduction; deep learning; noise classification; hearing; deaf; sound; audio; cochlear

Introduction

Cochlear implants (CIs) are implanted electronic medical devices that can enable patients with profound-to-severe hearing loss to obtain a sense of sound. In their study, Gifford et al [1] showed that 28% of individuals equipped with CI achieved 100% speech intelligibility. Sladen et al [2] also reported similar results in their study: after undergoing CI implantation, the word accuracy of CI users was 80% in a quiet environment. Although CI users have few obstacles in a quiet environment, there is still scope for improvement in a noisy environment [2].

Noise reduction (NR) is one of classical methods to alleviate the effect of background noise for CI users. Over the past few decades, many statistical signal processing NR methods have been proposed, such as log minimum mean squared error [3], Karhunen-Loève transform [4], Wiener filter-based on a priori signal-to-noise ratio (SNR) estimation [5], generalized maximum a posteriori spectral amplitude [6], and SNR-based [7] approaches. Loizou et al [8] proposed a single-channel algorithm to conduct NR, and the results showed that the sentence recognition scores in 14 participants with CI improved significantly over their daily performances. Dawson et al [7] evaluated a real-time NR algorithm which used the noise estimation to pick up 1 NR approach out of 2 different levels of NR approaches according to the SNR. The study results showed that the proposed NR algorithm could benefit CI users in speech a reception threshold under 3 kinds of noise. Mauger et al [9] optimized the gain function to achieve a better SNR-based NR, and the results showed that with the optimized gain function, a 27% improvement was achieved for CI users in speech-weighted noise. Although classical NR function can improve speech intelligibility for CI users in stationary noise conditions [7-9], improvements are still needed in nonstationary noise conditions [10].

Deep learning (DL)-based NR methods have recently shown better performance than classical statistical-based NR methods [11-17]. Lai et al [18] used a deep denoising autoencoder (DDAE)-based NR using vocoder simulation to perform NR function for CI users; the listening test showed that the speech intelligibility was better with DDAE-based NR than with conventional single-microphone NR approaches, whether in stationary or nonstationary noise conditions. Goehring et al [19,20] used neural and recurrent neural networks to perform the NR function for CI users, and the results showed that the proposed NR function could significantly improve speech intelligibility in babbling noise conditions. In DL methods, the nonstationary noise can be processed well, but this needs a huge amount of training data in different noise types and SNR levels. However, when a mismatch exists, such as when there is a difference in data between the training and testing phase, the performance of the DL method is usually degraded [10,18].

An environment-aware-based NR system called noise classifier (NC) +DDAE (NC+DDAE) was proposed to alleviate the above issue [21]. The NC+DDAE NR system combines n -specific

noise-dependent (ND)-DDAE NR models and a noise-independent (NI)-DDAE NR model. The NC function (ie, deep neural network model) was used to distinguish n different typical noises and select a suitable DDAE model to perform the NR function for CI users. Hence, the NC function made the NC+DDAE an environment-aware-based NR system. The objective measures and listening test showed that the NC+DDAE model had a much higher performance than did the other NR methods. Although the NC+DDAE model has proven to benefit the CI user and have the flexibility of customization, the NC+DDAE model requires several parameters, which increase the requirements for device implementation. Therefore, the NC+DDAE model needs to be modified to have fewer requirements while maintaining the performance at the same level.

Recently, the knowledge transfer (so called transfer learning) approach [22] has been used in many speech signal processing tasks (eg, speech emotion detection [23], text-to-speech system [24,25], and speech enhancement [26]) and has proven to provide benefits for the DL-based model. Knowledge transfer is a machine learning method developed for a specific task that reuses the initial parameters for a new model for the target task. In other words, the knowledge transfer technology transfers the domain knowledge based on the source domain to the target domain to help the DL-based model achieve better performance; furthermore, it can speed up the time needed to develop and train a model by reusing these pieces or modules that have already been developed [22]. Following the concept of knowledge transfer technology, we proposed an improved NC+DDAE NR model, called NC+DDAE_transfer (NC+DDAE_T). We first analyzed the differences between features in each layer of DDAE to choose the most suitable layer for NR adaptation. Next, we compared the performance between NC+DDAE and NC+DDAE_T with 2 well-known objective metrics: perceptual evaluation of speech quality (PESQ) [27] and short-time objective intelligibility (STOI) [28]. The PESQ shows the result of comparing the clean and processed speech by mean opinion score. In the mean opinion score, 5 is the highest score while 1 is the lowest. According to a previous study [27], a score over 4 is high enough for most people to listen comfortably and a score of 3.6 is an acceptable boundary for those with normal hearing. The STOI represents the speech intelligibility by a correlation coefficient derived from comparing the energy of clean and processed speech in each frame. STOI ranges from 0 to 1, with a higher score representing more clear and understandable speech. Finally, the clinical effectiveness of NC+DDAE_T with the NC+DDAE and DDAE NR systems for patients with CI was evaluated in noisy listening conditions.

Methods

In this section, we describe first the NC+DDAE approach. We then introduce the NC+DDAE_T method, the transfer learning-based NC+DDAE NR modified in this study. Finally, we describe the experimental setting and material to prove the

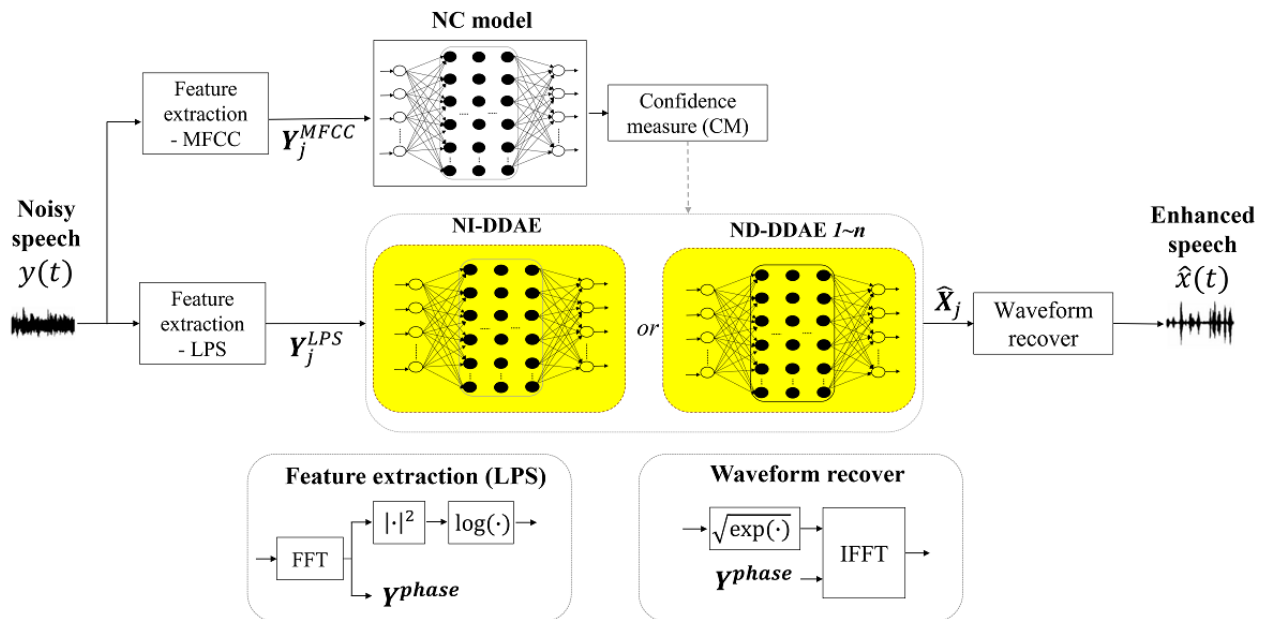
benefits of the proposed NC+DDAE_T compared to 2 well-known DL-based NR systems (ie, DDAE and NC+DDAE).

NR Based on the NC+DDAE Approach

Figure 1 shows the proposed NC+DDAE model in our previous study [21], where 2 critical units, NC and DDAE, were included. In this approach, first, the noisy speech signals $y(t)$ are processed by feature extraction units to obtain Y_j^{MFCC} and Y_j^{LPS} , which denote log power spectra (LPS) [29] and Mel-frequency cepstral coefficients [30], respectively, with j denoting the frame in the short-time Fourier transform. Y_j^{MFCC} is the input of the NC

model to determine the current type of background noise and to select a suitable DDAE model for NR, which includes multiple ND-DDAE models each trained by a model-specific noise type and a single NI-DDAE model trained by 120 noise types [15]. When the noisy input signal is similar to one of the specific noise types, the specific ND-DDAE model is chosen for NR; otherwise, the NI-DDAE is used. Afterward, the selected DDAE model processes Y_j^{LPS} to obtain the enhanced features. \hat{x}_j is combined with the noisy phase Y^{phase} to finally reconstruct the enhanced speech $\hat{x}(t)$. The NC+DDAE NR system has been defined in detail previously [21].

Figure 1. Structure of the noise classifier with a deep denoising autoencoder (NC+DDAE) system. DDAE: deep denoising autoencoder; FFT: fast Fourier transform; IFFT: inverse fast Fourier transform; LPS: log power spectra; NC: noise classifier; ND: noise-dependent; NI: noise-independent; MFCC: Mel-frequency cepstral coefficient.

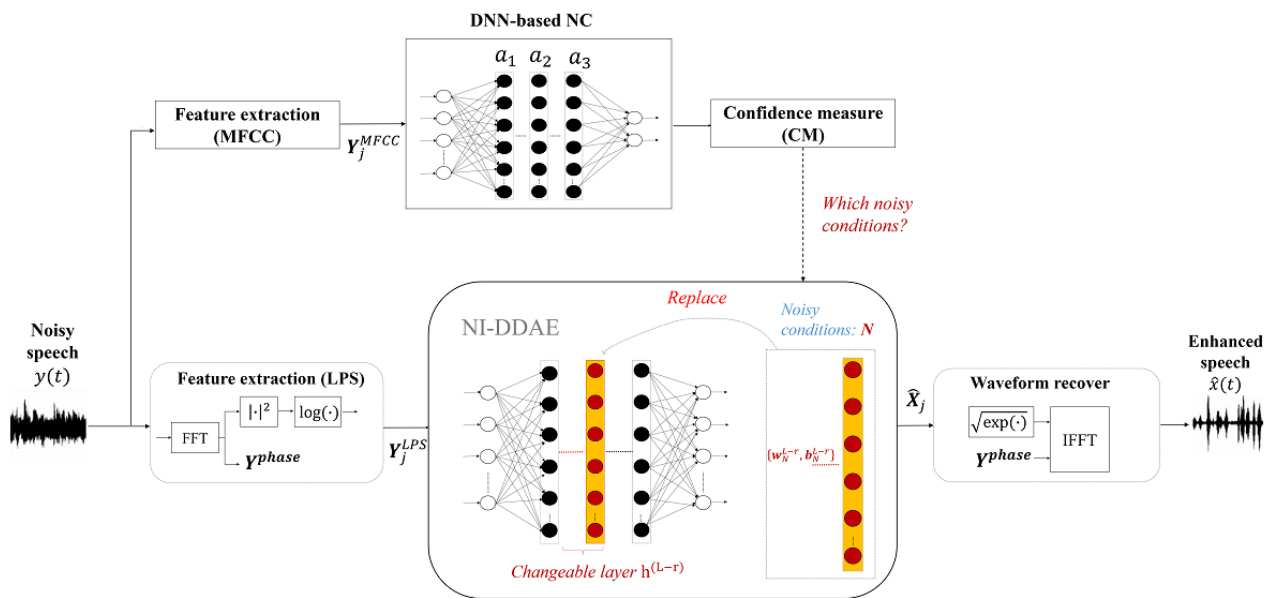


NR With the Proposed NC+DDAE_T Approach

Figure 2 shows the pipeline of the NC+DDAE_T NR approach proposed in this study. The signal processing procedure of the

NC+DDAE_T is similar to that of the above-mentioned NC+DDAE. The major difference lies in the NR model as described in the following sections.

Figure 2. Structure of the proposed noise classifier system with DDAE and knowledge transfer. DDAE: deep denoising autoencoder; DNN: deep neural network; FFT: fast Fourier transform; IFFT: inverse fast Fourier transform; LPS: log power spectra; NC: noise classifier; NI: noise-independent; MFCC: Mel-frequency cepstral coefficient.



NC Model

The NC model of the proposed NC+DDAE_T is the same as that in our previously described system. Initially, the system receives a noisy speech $y(t)$ and computes the Y_j^{MFCC} and Y_j^{LPS} features separately. Y_j^{MFCC} is then sent to the NC model. The NC model is a deep neural network (DNN) composed of 3 hidden layers. Each layer consists of 100 neurons and an output layer adapting the softmax function [30]. The output at the j -th node of the l -th layer in a DNN $h_j^{(l)}$ is produced according to equation 1:

$$h_j^{(l)} = \sigma \left(\sum_i h_i^{(l-1)} W_{ij}^{(l-1)} + b_j^{(l-1)} \right)$$

(1)

where the term $h_j^{(l-1)}$ denotes the output from the i -th node in the $(l-1)$ -th layer, $b_j^{(l)}$ is the bias of index j , and W_{ij}^l is the weight between hidden unit j and i . $\sigma(\cdot)$ is the activation function [30], which is the logistic function described in equation 2:

$$\sigma(t) = \frac{1}{1 + e^{-t}}$$

(2)

Next, the trained DNN model is used in the NC function. The output of the last layer is converted into the probability by the softmax function [31] to obtain the normalized probability-based output. The back propagation algorithm [32,33] is then applied to parameter set θ in equation 3, where $L(\cdot)$ is the loss function, N_i denotes the correct noise class, and \hat{N}_i is the output class of the DNN-based NC.

$$\theta = \arg \min_{\theta} \{L(N_i, \hat{N}_i; \theta)\}$$

(3)

To avoid substantial variance in the DNN output, we use the confidence measurement [34] to analyze the output of the DNN-based NC. Based on the confidence measurement score, a threshold is used to determine the classification results. In other words, when the confidence measurement score is higher than the threshold, the result predicted by the NC model is considered trustworthy. Nevertheless, if the confidence measurement score is not concrete to one noise type, then the NI-DDAE is chosen for NR; on the other hand, if the confidence measurement is solid, the ND-DDAE is selected.

DDAE-based NR Model

In the training phase, the noisy LPS feature Y_j^{LPS} and clean LPS feature X_j^{LPS} are the input and output, respectively, of the DDAE-based NR model. The details for training the DDAE NR model with L hidden layers mapping Y_j^{LPS} to X_j^{LPS} are available elsewhere [21]. The difference between NC+DDAE and NC+DDAE_T is that only the parameters of a specific layer (ie, w^{L-r} and b^{L-r}) are trainable as shown in equation 4, whereas the other parameters remain untrainable in the fine-tuning process. The constant L denotes the number of layers, and we used 5 layers (ie, $L=5$) in this study.

$$h^{(L-r)}(Y_j^{LPS}) = Relu(w^{L-r} h^{L-(r+1)}(Y_j^{LPS}) + b^{L-r})$$

(4)

where $\{W^1 \dots W^{L-r} \dots W^L\}$ and $\{b^1 \dots b^{L-r} \dots b^L\}$ are the matrices of weights and bias vectors of the DDAE NR model, respectively, whereas $Relu$ represents the activation function rectified linear unit [35]. The constant r is the index to identified the specific trainable layer. In this study, the second layer (ie, $r=3$) was chosen because, on average, substituting the second

layer achieved the best performance in our pilot study. The detailed experimental results are shown in [Multimedia Appendix 1](#).

Based on the above idea, the original NI-DDAE, trained with a huge database of noise samples, can be transformed into many ND-DDAE models according to the type of background noise. In this study, 12 common types of background noise were used; hence, 12 ND-DDAE models were derived from the NI-DDAE model. More specifically, each ND-DDAE model was determined by optimizing the following objective function:

$$\varphi = \arg \min_{\varphi} (F(\varphi) + \eta^1 \|W^1\|_2^2 + \dots + \eta^L \|W^L\|_2^2) \tag{5}$$

$$F(\varphi) = \frac{1}{M} \sum_{m=1}^M \|\hat{X}_j - X_j^{LPS}\|_2^2 \tag{6}$$

where M is the total number of training samples and $F(\varphi)$ is the loss function derived from \hat{x}_j and X_j^{LPS} . \hat{x}_j is the vector that contains the logarithmic amplitudes of the enhanced speech corresponding to the paired noisy LPS feature Y_j^{LPS} . Subsequently, the trained NI-DDAE provides the initial parameters for the ND-DDAE model, and the noise data of the specific environment are used to fine-tune this ND-DDAE model. Finally, the transformed LPS feature \hat{x}_j is sent to the waveform recovery unit to reconstruct the waveform. More specifically, \hat{x}_j is first processed using square root and exponential operations. The waveform recovery function then reconstructs the enhanced speech \hat{x}_j with the noisy phase Y^{phase} .

Training and Evaluation Procedure

In this section, we show how the NC, DDAE, and NC+DDAE_T models were trained. First, we trained a new NC model according to the 12 common background noises, 2talker_unseen1, 2talker_unseen2, Construction Jackhammer (CJ), 2 Talker, Cafeteria, MRT (Mass Rapid Transit), cafeteria, Toy-Squeeze-Several, speech shape noise from the Institute of Electrical and Electronics Engineers (SSN_IEEE), Siren, Multiple type noise 1, and Multiple type noise 2, which are shown in [Figure 3](#). Note that the training approach is described in the previous section “NC Model”. After the training, the prediction accuracy of the 12 noises was 100%. The detailed results of the confusion matrix are shown in [Multimedia Appendix 2](#).

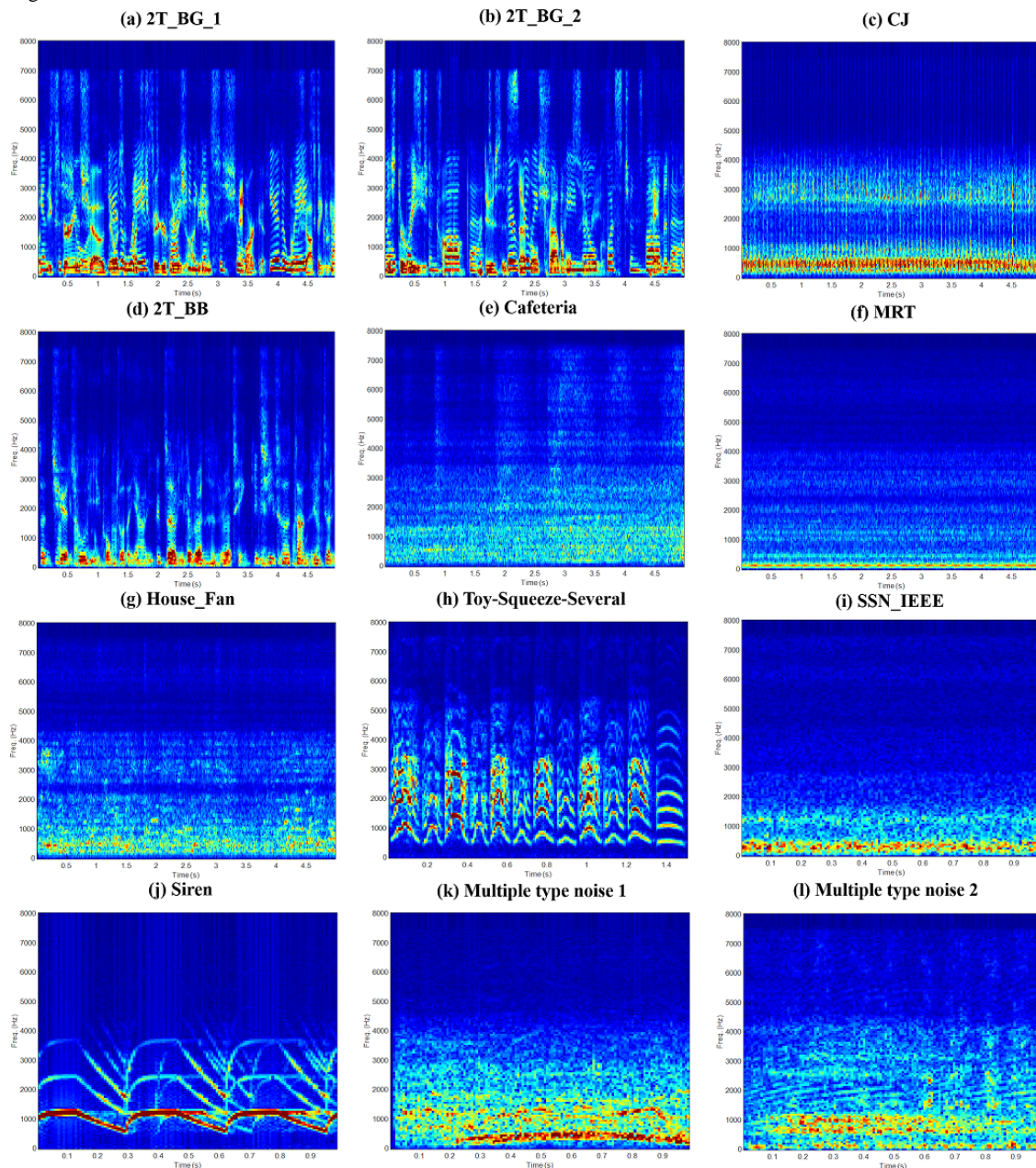
To train the DDAE NR model, the Taiwan Mandarin version of the hearing in noise test (TMHINT) corpus [36] was selected to conduct all experiments, including the training and evaluation parts. All 320 sentences, each consisting of 10 characters, were recorded at a 16 kHz sampling rate, after which 120 utterances among the TMHINT corpus were selected and corrupted by 120 noise types [15] at 7 SNR levels (−10, −7, −4, −1, 1, 4, 7, and 10 dB) as the training set for the DDAE model. The other

200 utterances were also corrupted with the 12 common background noises—as mentioned in the description of NC training—at 6 SNR levels (−6, −3, 0, 3, and 6 dB) as the outside testing set. In our previous study, this trained model was defined as the NI-DDAE.

Next, we combined the NC with NI-DDAE and fine-tuned the model with each noise type in the NC, and the NI-DDAE was transformed into NC+DDAE_T. In the fine-tuning step, we could freeze or adopt each layer in the NI-DDAE. Previously, we had studied which layer of the NI-DDAE model had to be replaced to achieve the best performance. We substituted each layer by modifying r in the range from 1 to 5; meanwhile, we conducted 2 well-known objective speech evaluations, PESQ [27] and STOI [28], to identify the most appropriate layer. On average, replacing the middle layer of the NI-DDAE model (ie, the second layer this study) achieved a better performance than did substituting other layers. The detailed results can be found in [Multimedia Appendix 1](#). Hence, we uniformly replaced the parameters of the third layer in all subsequent tests. As the 2 DL-based NR systems, DDAE and NC+DDAE, achieved better performances in our previous studies [18,21] than did the well-known unsupervised NR algorithms, the log minimum mean squared error [3] and Karhunen-Loève transform [37], we used the DDAE and NC+DDAE algorithms for comparisons to evaluate the NC+DDAE_T in this study.

Subsequently, we enrolled 10 CI users to conduct speech intelligibility tests, and details of these subjects are shown in the [Multimedia Appendix 3](#). This study protocol was approved by the Research Ethics Review Committee of Cheng Hsin Hospital under the following approval number: CHGH-IRB (645) 107A-17-2. The first author, LPHL, explained the study to the patients and collected the signed institutional review board informed consent before the experiment. All participants used their own clinical speech processors and temporarily disabled the built-in NR functions during the test. The test signals of noisy and enhanced speech were played at 65 dB sound pressure level by a speaker and were then processed through a CI processor to simulate the performance of each NR approach for CI users. To ensure that fatigue did not affect the study participants, each individual only heard a total of 16 test conditions (2 background noise [2 talker and CJ] × 2 SNR levels [0 and 3 dB] × 4 signal processing systems [noisy, DDAE, NC+DDAE, and NC+DDAE_T]) with 10 sentences of 10 words in each test condition. The participants were instructed to repeat verbally what they had heard. We evaluated the speech intelligibility under each test condition using the word correct rate (WCR) [38-42] calculated as the ratio between the number of correctly identified words and the total number of words. To further prevent participant fatigue, tests were paused for 5 minutes every 30 minutes. Moreover, we calculated the statistical power to see whether the sample size (10 patients in this study) was large enough to obtain a significant difference in the result. The statistical power of this study is 1. According to Cohen et al [43] a statistical power over 0.8 is sufficiently high to conclude that there is a significant difference in the hypothesis.

Figure 3. Spectrograms of the 12 noise signals: (a) 2T_BG_1, (b) 2T_BG_2, (c) CJ, (d) 2T_BB, (e) Cafeteria, (f) MRT, (g) House Fan, (h) Toy-Squeeze-Several, (i) SSN_IEEE, (j) Siren, (k) Multiple type noise 1, and (l) Multiple type noise 2. 2T_BG_1 is a noise that mixes the speech of a girl and a boy both speaking repeatedly in English. 2T_BG_2 are the same as those in 2T_BG_1 but with different sentences. 2T_BB is a noise that overlays 2 sentences in Chinese spoken by the same male speaker. Multiple type noise 1 is a mix of the sound of sirens and cheering crowd, whereas Multiple type noise 2 is a sound combining scratching and boing. The other samples are common background noises from daily life. 2T_BB: 2 Talker; 2T_BG_1: 2 talker_unseen1; 2T_BG_2: 2 talker_unseen2; CJ: Construction Jackhammer; MRT: Mass Rapid Transit; SSN_IEEE: speech shape noise from the Institute of Electrical and Electronics Engineers.



Results

Objective Evaluation Using PESQ and STOI Scores

We compared the newly proposed NC+DDAE_T with the previously established NR systems, DDAE and NC+DDAE. The PESQ and STOI scores of these tests are shown in Figures 4 and 5, respectively. As demonstrated in Figure 4, the PESQ scores of the proposed NC+DDAE_T are generally similar to those of the NC+DDAE. The details regarding the average

scores of each approach (ie, noisy, DDAE, NC+DDAE, and NC+DDAE_T) for the 12 background noises at 6 different SNR levels can be found in Table A1 of Multimedia Appendix 4. In the STOI scores, the NC+DDAE_T model also achieved the same level as did the NC+DDAE (Figure 5). The detailed STOI scores are listed in Table A2 of Multimedia Appendix 4. These objective evaluation results proved that the NC+DDAE_T could provide almost the same speech intelligibility performance as the NC+DDAE.

Figure 4. Mean perceptual evaluation of speech quality (PESQ) scores of the 4 noise reduction approaches. 2T_BB: 2 Talker; 2T_BG_1: 2 talker_unseen1; 2T_BG_2: 2 talker_unseen2; CJ: Construction Jackhammer; dB: decibel; DDAE: deep denoising autoencoder; NC: noise classifier; NC+DDAE_T: noise classifier + deep denoising autoencoder with knowledge transfer; MRT: Mass Rapid Transit; PESQ: perceptual evaluation of speech quality; SNR: signal-to-noise ratio; SSN_IEEE: speech shape noise from the Institute of Electrical and Electronics Engineers.

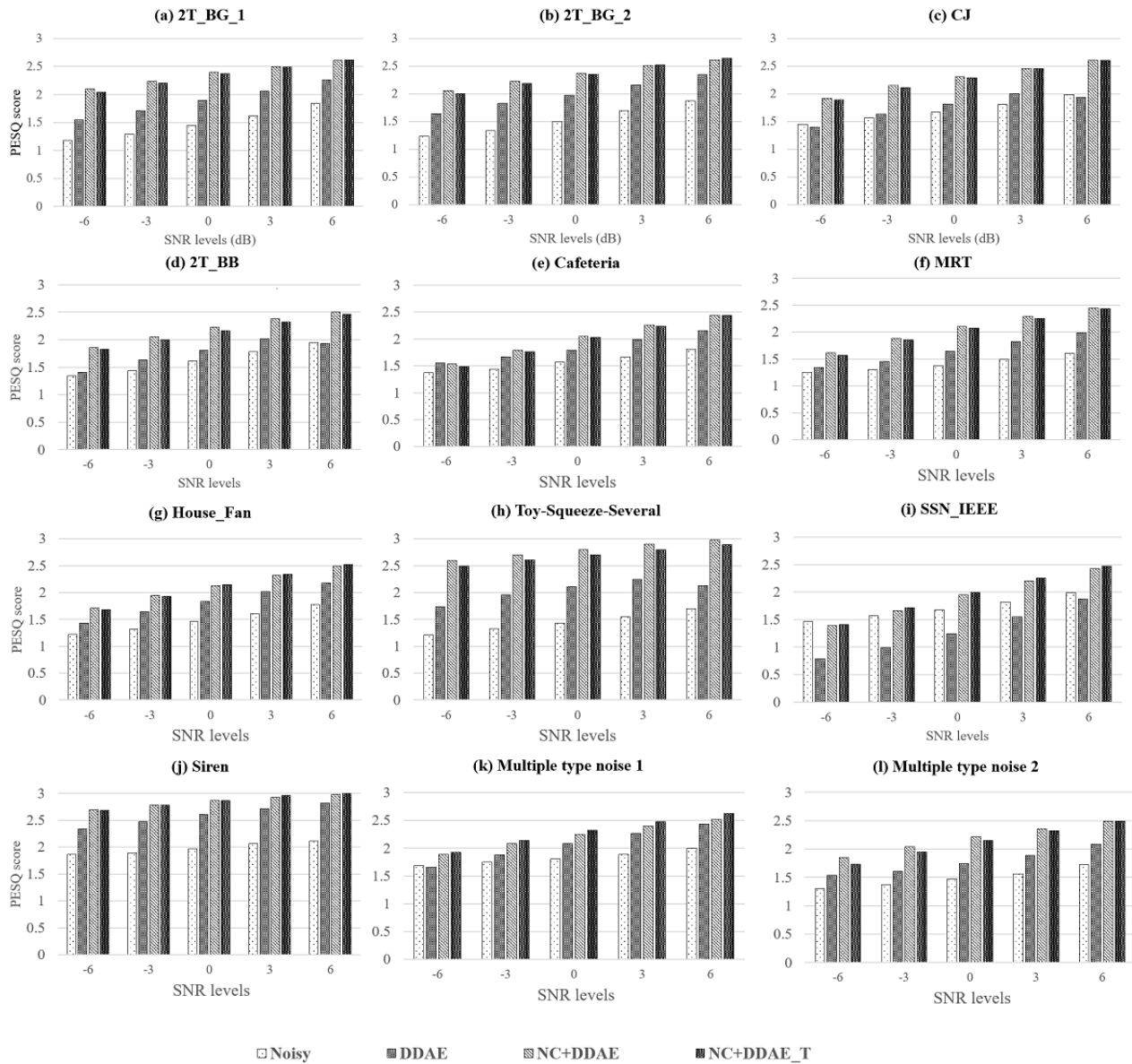
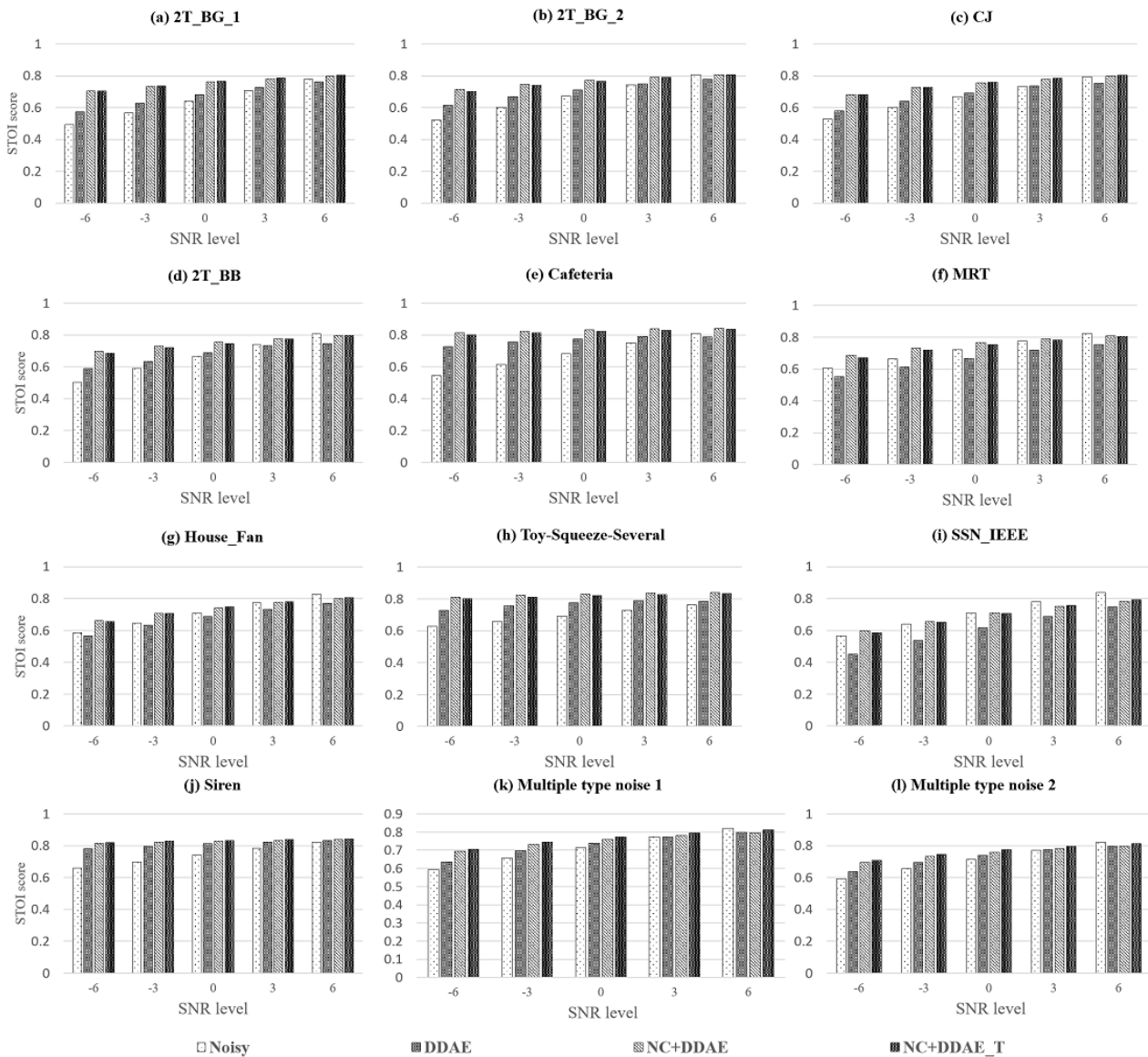


Figure 5. Mean short-time objective intelligibility (STOI) scores of the different noise reduction approaches. 2T_BB: 2 Talker; 2T_BG_1: 2 talker_unseen1; 2T_BG_2: 2 talker_unseen2; CJ: Construction Jackhammer; DDAE: deep denoising autoencoder; NC: noise classifier; NC+DDAE_T: noise classifier + deep denoising autoencoder with knowledge transfer; MRT: Mass Rapid Transit; SNR: signal-to-noise ratio; SSN_IIEE: speech shape noise from the Institute of Electrical and Electronics Engineers; STOI: short-time objective intelligibility.



Recognition in Listening Tests

Figure 6 shows the average WCR scores of 10 individuals with CI in the 2 Talker and CJ noise conditions each at 0- and 3-dB SNR levels. The detailed results are as follows: The respective average WCR scores and standard error of the mean (SEM) for noisy, DDAE, NC+DDAE, and NC+DDAE_T with 2 Talker background noise were 4.1 (SEM 1.87), 27.8 (SEM 5.42), 38.9 (SEM 8.83), and 43.2 (SEM 9.33) at the 0-dB SNR level; and 10.3 (SEM 3.84), 27.7 (SEM 5.24), 48.2 (SEM 9.69), and 50.3 (SEM 8.98) at the 3-dB SNR level. In the CJ background noise, the respective average scores and SEMs were 19.3 (SEM 5.76), 27.7 (SEM 5.24), 42.2 (SEM 9.64), and 50.6 (SEM 10.0) at the 0-dB SNR level; and 37.1 (SEM 9.84), 38.8 (SEM 8.41), 49.3 (SEM 9.31), and 50.9 (SEM 10.13) at the 3-dB SNR level. These results demonstrated that the NC+DDAE_T provided better

speech intelligibility scores than did noisy speech. Moreover, the newly developed NC+DDAE_T model achieved slightly higher intelligibility performances than did the NC+DDAE approach under most test conditions. The 1-way analysis of variance (ANOVA) [44] with least significant difference post hoc comparison [45] was used to analyze the results of the 4 NR systems (noisy, DDAE, NC+DDAE, and NC+DDAE_T) in the 4 test conditions. The 1-way ANOVA result confirmed that the WCR scores differed significantly among the 4 systems ($F=13.256$; $P<.001$). The least significant difference post hoc comparisons (Table 1) further revealed that the noisy condition was significantly different from the other 3 systems (DDAE: $P=.16$; NC+DDAE: $P<.001$; NC+DDAE_T: $P<.001$). Meanwhile, the differences between the NC+DDAE and NC+DDAE_T models were not significant ($P=.50$).

Figure 6. Mean intelligibility scores of 10 participants with cochlear implants in 4 types of simulated test conditions. 2T_BB: 2 Talker; CJ: Construction Jackhammer; dB: decibel; DDAE: deep denoising autoencoder; NC: noise classifier; NC+DDAE_T: noise classifier + deep denoising autoencoder with knowledge transfer.

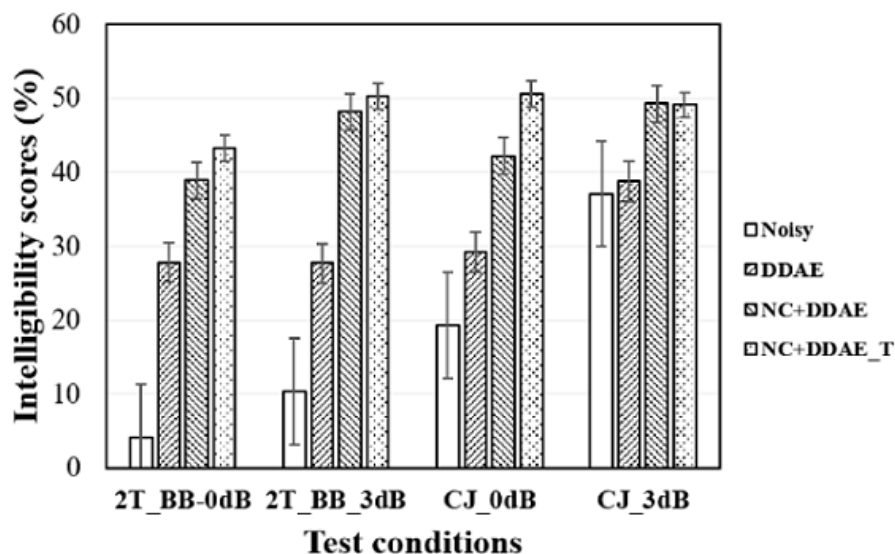


Table 1. The mean difference, standard error, and significance of the listening test in each noise reduction system.

| Method (I) by test (J) | Mean difference (I-J) (standard error) | P value ^a |
|----------------------------|--|----------------------|
| Noisy (I) | | |
| DDAE ^b (J) | -13.18 (5.428) | .016 ^c |
| NC ^d +DDAE (J) | -26.95 (5.428) | < .001 |
| NC+DDAE_T ^e (J) | -30.60 (5.428) | < .001 |
| DDAE (I) | | |
| Noisy (J) | 13.18 (5.428) | .02 |
| NC+DDAE (J) | -13.78 (5.428) | .01 |
| NC+DDAE_T (J) | -17.43 (5.428) | .002 |
| NC+DDAE (I) | | |
| Noisy (J) | 26.95 (5.428) | < .001 |
| DDAE (J) | 13.78 (5.428) | .01 |
| NC+DDAE_T (J) | -3.65 (5.428) | .50 |
| NC+DDAE_T (I) | | |
| Noisy (J) | 30.60 (5.428) | < .001 |
| DDAE (J) | 17.43 (5.428) | .002 |
| NC+DDAE (J) | 3.65 (5.428) | .50 |

^aP values are significant at $\alpha = .05$. Least significant difference was selected to conduct post hoc testing.

^bDDAE: deep denoising autoencoder.

^cValues in italics represent significant values.

^dNC: noise classifier.

^eNC+DDAE_T: noise classifier + deep denoising autoencoder with knowledge transfer.

Comparison of the Numbers of Parameters

The original structure of the NC+DDAE system used 12 ND+DDAEs and 1 NI+DDAE for the NR. In this study, the

newly developed NC+DDAE_T system only needed 1 NI+DDAE and 12 different layer parameters to achieve the same performance as the previous NC+DDAE system. We further compared the numbers of parameters between the

NC+DDAE and NC+DDAE_T approaches. The NC+DDAE_T approach required only 0.1 million parameters while the previous NC+DDAE system needed 4.4 million parameters. The number of parameters was thus reduced by 76.5% compared to the previous approach.

Discussion

Layers for Substitution

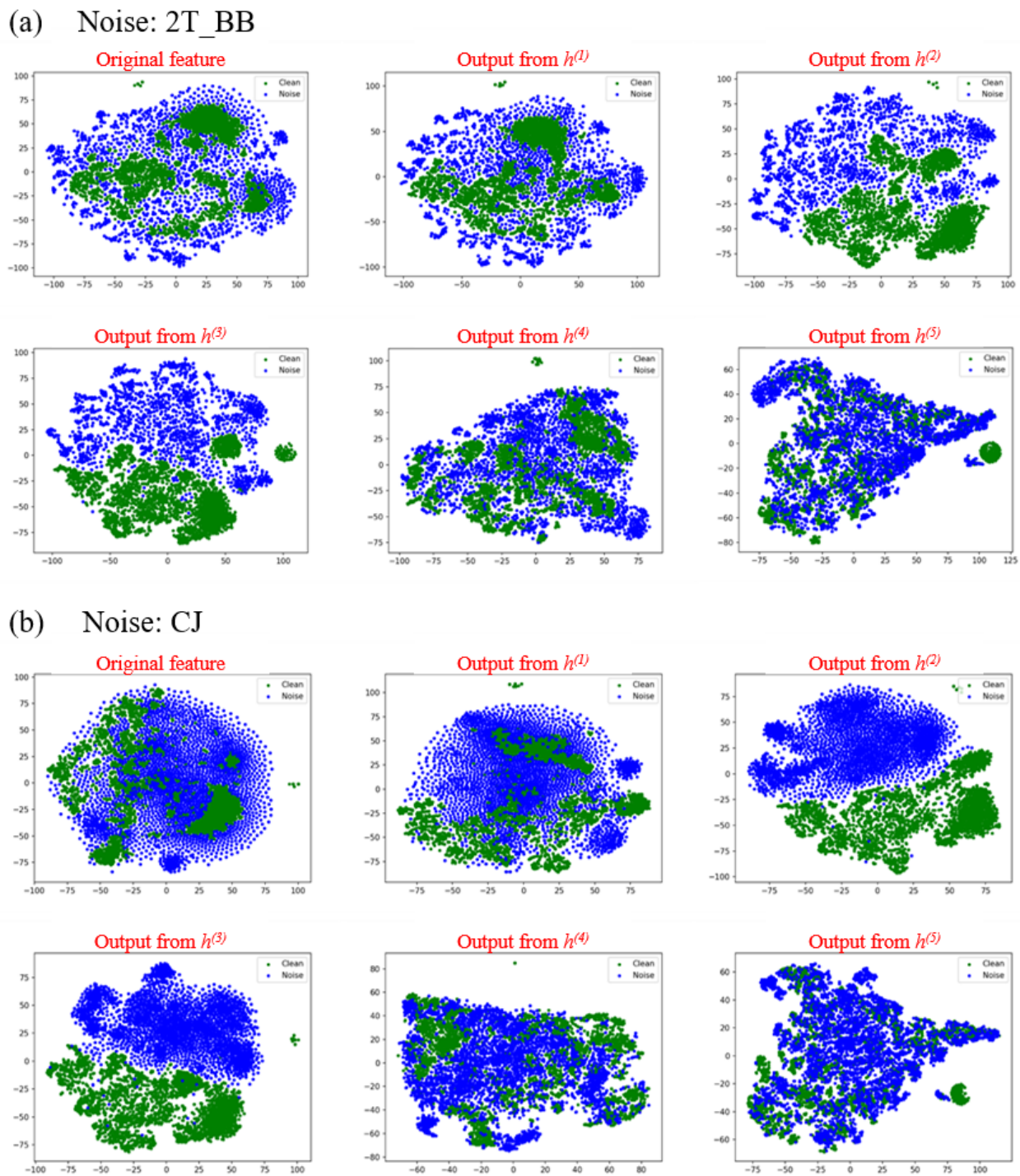
This study proposed a new NC+DDAE_T NR model that helps CI users to improve speech intelligibility in noisy listening conditions. Knowledge transfer technology was used to reduce the parameter requirements in comparison to the previous NC+DDAE approach. The experimental results of the objective evaluation and the subjective listening tests demonstrated that the NC+DDAE_T achieved performances comparable to those of the NC+DDAE approach, while the number of parameters used by the NC+DDAE_T was reduced by 76.5% compared to the NC+DDAE. Therefore, knowledge transfer technology could be a useful approach to further improve the benefits of NC+DDAE in reducing the cost of implementation in the future.

The architecture of the NC+DDAE_T, (ie, which layer is substituted) is the basis for achieving higher performance with this novel system compared to the NC-DDAE. According to the objective evaluation by PESQ and STOI scores ([Multimedia Appendix 1](#)), the substitution of the middle layer can achieve better performances. To further analyze why the middle layer was so important, *t*-distributed stochastic neighbor embedding (*t*-SNE) [46] was used to visualize the features that output by each layer. The acoustic features of noisy and clean speech (ie, LPS) were the inputs for the trained NI-DDAE NR model. The

output features of each NI-DDAE layer were analyzed using *t*-SNE, which can project the distribution of each layer onto a 2D plane. [Figure 7](#) shows the results of this feature visualization. Green dots represent the output features of clean speech, whereas blue dots indicate features of noisy speech. The less overlap is apparent between the green and blue areas, the better the layer can separate the features. These results indicate that clean and noisy data were primarily separated in the output from $h^{(2)}$ and $h^{(3)}$, implying that the front layers help to distinguish noisy speech from clean features and thus could be the most important layers. This interpretation is also consistent with the objective evaluation results in [Multimedia Appendix 1](#).

To explain the phenomenon illustrated in [Figure 7](#), we suggest that the NC+DDAE_T model may work similarly to the human brain. The first layers of the model may try to separate the noise from the speech features. Therefore, these features would diverge completely in the middle layers of this NR model. The model would then try to reconstruct the enhanced speech and lower the volume of the noise in the final layers of the model; hence, the features would converge again in the *t*-SNE analysis. Based on these hypotheses, the second layer may be the key to feature separation because the features are well separated after the second layer. Therefore, to adapt the NR model to a specific type of noise, substituting the second layer would be the best choice, which corresponds to the results of the objective evaluation. The other parts of the NC+DDAE_T model may work as preprocessing and vocoder units. These parts are common units of all NR models; thus, different ND-DDAEs can share the same weight and bias values. Therefore, the concept of knowledge transfer can be used in this part to decrease the size of each model.

Figure 7. *t*-distributed stochastic neighbor embedding (*t*-SNE) feature analysis of each layer in the noise-independent deep denoising autoencoder (NI-DDAE) model with noisy and clean speech data. The green dots represent the output features of clean speech and the blue dots indicate features of noisy speech. 2T_BB: 2 Talker; CJ: Construction Jackhammer.

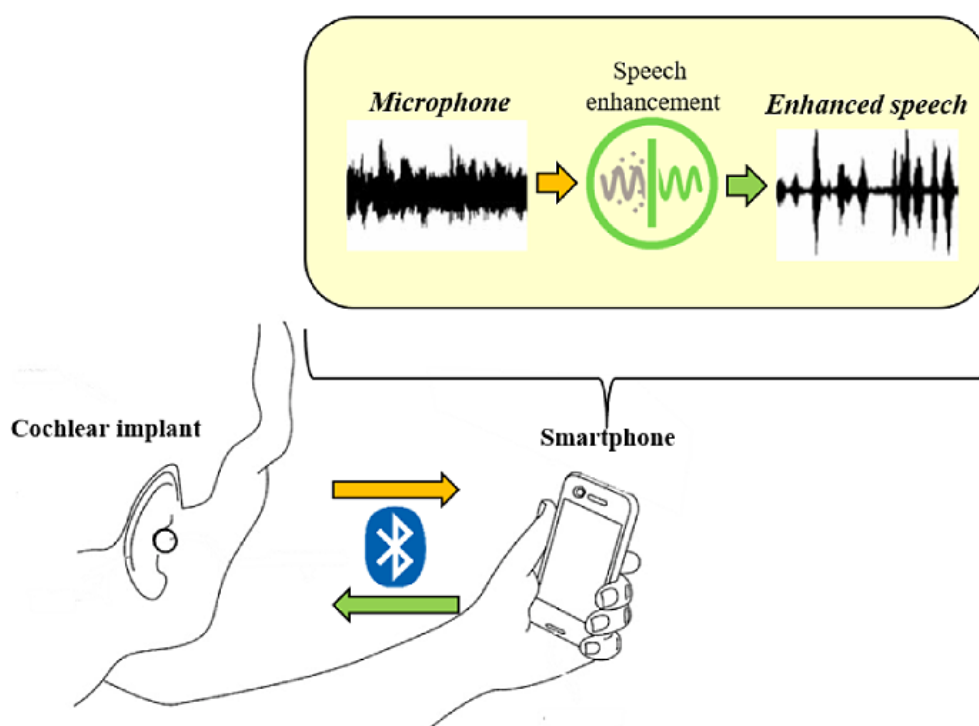


Future Perspectives

Based on previous and current results of objective evaluation and listening tests, we can conclude that the proposed NC+DDAE_T performs comparably to the NC+DDAE. In addition, the NC+DDAE_T needs only a quarter of the number of parameters compared to the 12 ND-DDAE models. These characteristics suggest a great potential for future implementation of the NC+DDAE_T model. With the decreased

number of parameters, an implemented device would require less memory. To prove this concept, we have implemented the NC+DDAE_T architecture in an app on an iPhone XR mobile phone (Apple Inc) as shown in Figure 8. The processing time could satisfy the maximum group delay requirement of assistive listening devices. With this advantage of edge computing, the proposed NC+DDAE_T may become a new kind of hearing assistive technology in the near future.

Figure 8. Schematic of the noise classifier deep denoising autoencoder with knowledge transfer (NC+DDAE_T) implementation.



Limitations

The proposed NC+DDAE_T is an adaptable NR system, which means that the system benefits may be affected by the training data (eg, background noise types, speakers). Therefore, if the proposed system faces noisy conditions that are very different from the training data (ie, mismatch conditions), the proposed system would require major improvements, and new recordings of noise data may be needed. Overcoming this issue requires future study. Additionally, although the proposed system was implemented in an app, the full implementation of the proposed system in the hardware of currently used CI devices is still a way off. However, as studies increasingly focus on the acceleration of DL-based models in microprocessors [47,48],

there is a greater chance that DL technologies may be implemented into CI devices in the near future.

Conclusions

This study proposed a novel NC+DDAE_T system for NR in CI devices. The knowledge transfer approach was used to lower the number of parameters of the DDAE model. The experimental results of the objective evaluations, along with the listening tests, showed that the proposed NC+DDAE_T model provided comparable performance to the previously established NC+DDAE NR model. These results suggest that the proposed NC+DDAE_T model may be a new NR system that can enable CI users to hear well in noisy conditions.

Acknowledgments

This study was supported by the Ministry of Science and Technology of Taiwan (project #110-2218-E-A49A-501, #110-2314-B-350-003, #109-2218-E-010-004, and #108-2314-B-350-002-MY2) and Cheng Hsin General Hospital (#CY10933).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Results Following Replacement of Each Layer of weight and bias of the Deep Denoising Autoencoder Model.
[\[DOCX File, 126 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Confusion matrix of the 12 noise classifications.
[\[DOCX File, 46 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Individual biographical data of the attended cochlear implant subjects.

[\[DOCX File , 17 KB-Multimedia Appendix 3\]](#)

Multimedia Appendix 4

Perceptual evaluation of speech quality and short-time objective intelligibility scores of different noise reduction systems.

[\[DOCX File , 29 KB-Multimedia Appendix 4\]](#)

References

1. Gifford RH, Shallop JK, Peterson AM. Speech recognition materials and ceiling effects: considerations for cochlear implant programs. *Audiol Neurootol* 2008;13(3):193-205. [doi: [10.1159/000113510](https://doi.org/10.1159/000113510)] [Medline: [18212519](https://pubmed.ncbi.nlm.nih.gov/18212519/)]
2. Sladen DP, Ricketts TA. Frequency importance functions in quiet and noise for adults with cochlear implants. *Am J Audiol* 2015 Dec;24(4):477-486. [doi: [10.1044/2015_AJA-15-0023](https://doi.org/10.1044/2015_AJA-15-0023)] [Medline: [26650018](https://pubmed.ncbi.nlm.nih.gov/26650018/)]
3. Ephraim Y, Malah D. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. *IEEE Trans. Acoust., Speech, Signal Process* 1985 Apr;33(2):443-445. [doi: [10.1109/tassp.1985.1164550](https://doi.org/10.1109/tassp.1985.1164550)]
4. Rezaeey A, Gazor S. An adaptive KLT approach for speech enhancement. *IEEE Trans. Speech Audio Process* 2001;9(2):87-95. [doi: [10.1109/89.902276](https://doi.org/10.1109/89.902276)]
5. Scalart P. Speech enhancement based on a priori signal to noise estimation. 1996 Presented at: Speech enhancement based on a priori signal to noise estimation. In *IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*(Vol. 2, pp.). IEEE; May 5, 1996; Atlanta, GA, USA p. 629-632. [doi: [10.1109/icassp.1996.543199](https://doi.org/10.1109/icassp.1996.543199)]
6. Lai YH, Su YC, Tsao Y, Young ST. Evaluation of generalized maximum a posteriori spectral amplitude (GMAPA) speech enhancement algorithm in hearing aids. 2013 Presented at: 2013 IEEE International Symposium on Consumer Electronics (ISCE); June 2013; Hsinchu, Taiwan p. 245-246. [doi: [10.1109/isce.2013.6570208](https://doi.org/10.1109/isce.2013.6570208)]
7. Dawson PW, Mauger SJ, Hersbach AA. Clinical evaluation of signal-to-noise ratio-based noise reduction in Nucleus cochlear implant recipients. *Ear and hearing* 2011;32(3):382. [doi: [10.1097/aud.0b013e318201c200](https://doi.org/10.1097/aud.0b013e318201c200)]
8. Loizou PC, Lobo A, Hu Y. Subspace algorithms for noise reduction in cochlear implants. *The Journal of the Acoustical Society of America* 2005 Nov;118(5):2791-2793. [doi: [10.1121/1.2065847](https://doi.org/10.1121/1.2065847)] [Medline: [16334894](https://pubmed.ncbi.nlm.nih.gov/16334894/)]
9. Mauger SJ, Dawson PW, Hersbach AA. Perceptually optimized gain function for cochlear implant signal-to-noise ratio based noise reduction. *The Journal of the Acoustical Society of America* 2012 Jan;131(1):327-336. [doi: [10.1121/1.3665990](https://doi.org/10.1121/1.3665990)] [Medline: [22280595](https://pubmed.ncbi.nlm.nih.gov/22280595/)]
10. Tu Y, Du J, Lee C. Speech enhancement based on teacher? Student deep learning using improved speech presence probability for noise-robust speech recognition. *IEEE/ACM Trans. Audio Speech Lang. Process* 2019 Dec;27(12):2080-2091. [doi: [10.1109/taslp.2019.2940662](https://doi.org/10.1109/taslp.2019.2940662)]
11. Wang D, Chen J. Supervised speech separation based on deep learning: an overview. *IEEE/ACM Trans Audio Speech Lang Process* 2018 Oct;26(10):1702-1726 [FREE Full text] [doi: [10.1109/TASLP.2018.2842159](https://doi.org/10.1109/TASLP.2018.2842159)] [Medline: [31223631](https://pubmed.ncbi.nlm.nih.gov/31223631/)]
12. Wu M, Wang D. A one-microphone algorithm for reverberant speech enhancement. 2003 Presented at: IEEE International Conference on Acoustics, Speech, and Signal Processing; May 2003; Hong Kong, China. [doi: [10.1109/icassp.2003.1198925](https://doi.org/10.1109/icassp.2003.1198925)]
13. Healy EW, Delfarah M, Vasko JL, Carter BL, Wang D. An algorithm to increase intelligibility for hearing-impaired listeners in the presence of a competing talker. *The Journal of the Acoustical Society of America* 2017 Jun;141(6):4230-4239. [doi: [10.1121/1.4984271](https://doi.org/10.1121/1.4984271)] [Medline: [28618817](https://pubmed.ncbi.nlm.nih.gov/28618817/)]
14. Kumar A, Florencio D. Speech enhancement in multiple-noise conditions using deep neural networks. arXiv:1605.02427. 2016. URL: <https://arxiv.org/abs/1605.02427> [accessed 2016-05-09]
15. Xu Y, Du J, Dai L, Lee C. A regression approach to speech enhancement based on deep neural networks. *IEEE/ACM Trans. Audio Speech Lang. Process* 2015 Jan;23(1):7-19. [doi: [10.1109/taslp.2014.2364452](https://doi.org/10.1109/taslp.2014.2364452)]
16. Lu X, Tsao Y, Matsuda S, Hori C. Speech enhancement based on deep denoising autoencoder. In: *Interspeech*. 2013 Aug Presented at: Interspeech 2013; 25-29 August 2013; Lyon, France.
17. Xu Y, Du J, Dai LR, Lee CH. Dynamic noise aware training for speech enhancement based on deep neural networks. 2014 Presented at: Fifteenth Annual Conference of the International Speech Communication Association; Sept 14, 2014; Singapore.
18. Lai Y, Chen F, Wang S, Lu X, Tsao Y, Lee C. A deep denoising autoencoder approach to improving the intelligibility of vocoded speech in cochlear implant simulation. *IEEE Trans. Biomed. Eng* 2017 Jul;64(7):1568-1578. [doi: [10.1109/tbme.2016.2613960](https://doi.org/10.1109/tbme.2016.2613960)]
19. Goehring T, Bolner F, Monaghan JJ, van Dijk B, Zarowski A, Bleeck S. Speech enhancement based on neural networks improves speech intelligibility in noise for cochlear implant users. *Hearing Research* 2017 Feb;344:183-194. [doi: [10.1016/j.heares.2016.11.012](https://doi.org/10.1016/j.heares.2016.11.012)] [Medline: [27913315](https://pubmed.ncbi.nlm.nih.gov/27913315/)]
20. Goehring T, Keshavarzi M, Carlyon RP, Moore BCJ. Using recurrent neural networks to improve the perception of speech in non-stationary noise by people with cochlear implants. *The Journal of the Acoustical Society of America* 2019 Jul;146(1):705-718. [doi: [10.1121/1.5119226](https://doi.org/10.1121/1.5119226)] [Medline: [31370586](https://pubmed.ncbi.nlm.nih.gov/31370586/)]

21. Lai YH, Tsao Y, Lu X, Chen F, Su YT, Chen KC, et al. Deep learning-based noise reduction approach to improve speech intelligibility for cochlear implant recipients. *Ear and hearing* 2018;39(4):795-809. [doi: [10.1097/aud.0000000000000537](https://doi.org/10.1097/aud.0000000000000537)]
22. Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng* 2010 Oct;22(10):1345-1359. [doi: [10.1109/tkde.2009.191](https://doi.org/10.1109/tkde.2009.191)]
23. Latif S, Rana R, Younis S, Qadir J, Epps J. Transfer learning for improving speech emotion classification accuracy. arXiv. 2018. URL: <https://arxiv.org/abs/1801.06353> [accessed 2018-01-19]
24. Fan Y, Qian Y, Soong FK, He L. Multi-speaker modeling and speaker adaptation for DNN-based TTS synthesis. 2015 Presented at: International Conference on Acoustics, Speech and Signal Processing; April 2015; Toronto, Ontario, Canada p. 4475-4479. [doi: [10.1109/icassp.2015.7178817](https://doi.org/10.1109/icassp.2015.7178817)]
25. Jia Y, Zhang Y, Weiss R, Wang Q, Shen J, Ren F, et al. A transfer learning and progressive stacking approach to reducing deep model sizes with an application to speech enhancement. *Advances in neural information processing systems* 2018:4480-4490 [FREE Full text] \
26. Wang S, Li K, Huang Z, Siniscalchi SM, Lee CH. Transfer learning and progressive stacking approach to reducing deep model sizes with an application to speech enhancement. 2017 Presented at: International Conference on Acoustics, Speech and Signal Processing; March 2017; New Orleans, LA, USA p. 5575-5579. [doi: [10.1109/icassp.2017.7953223](https://doi.org/10.1109/icassp.2017.7953223)]
27. Rix AW, Beerends JG, Hollier MP, Hekstra AP. Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs. 2001 May Presented at: International Conference on Acoustics, Speech, and Signal Processing; May 7-11, 2001; Salt Lake City, UT, USA p. 749-752. [doi: [10.1109/icassp.2001.941023](https://doi.org/10.1109/icassp.2001.941023)]
28. Taal CH, Hendriks RC, Heusdens R, Jensen J. A short-time objective intelligibility measure for time-frequency weighted noisy speech. 2010 Presented at: International Conference on Acoustics, Speech and Signal Processing; April 2010; Dallas, TX, USA p. 4214-4217. [doi: [10.1109/icassp.2010.5495701](https://doi.org/10.1109/icassp.2010.5495701)]
29. Du J, Huo Q. A speech enhancement approach using piecewise linear approximation of an explicit model of environmental distortions. 2008 Presented at: Ninth Annual Conference of The International Speech Communication Association; Sept 2008; Brisbane, Australia.
30. Davis S, Mermelstein P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Trans. Acoust., Speech, Signal Process* 1980 Aug;28(4):357-366. [doi: [10.1109/tassp.1980.1163420](https://doi.org/10.1109/tassp.1980.1163420)]
31. Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network. arXiv. 2015. URL: <https://arxiv.org/abs/1503.02531> [accessed 2015-03-09]
32. Bengio Y. Learning deep architectures for AI. In: *FNT in Machine Learning*. Norwell, MA: Now Publishers Inc; 2009:1-127.
33. Mohamed A, Dahl GE, Hinton G. Acoustic modeling using deep belief networks. *IEEE Trans. Audio Speech Lang. Process* 2012 Jan;20(1):14-22. [doi: [10.1109/tasl.2011.2109382](https://doi.org/10.1109/tasl.2011.2109382)]
34. Mengusoglu E, Ris C. Use of acoustic prior information for confidence measure in ASR applications. 2001 Presented at: Seventh European Conference on Speech Communication and Technology; Sept 13, 2021; Virtual.
35. Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines. *ICML*. 2010 Jan. URL: <https://openreview.net/forum?id=rkb15iZdZB> [accessed 2019-07-17]
36. Wong LLN, Soli SD, Liu S, Han N, Huang MW. Development of the Mandarin Hearing in Noise Test (MHINT). *Ear Hear* 2007 Apr;28(2 Suppl):70S-74S. [doi: [10.1097/AUD.0b013e31803154d0](https://doi.org/10.1097/AUD.0b013e31803154d0)] [Medline: [17496652](https://pubmed.ncbi.nlm.nih.gov/17496652/)]
37. Mittal U, Phamdo N. Signal/noise KLT based approach for enhancing speech degraded by colored noise. *IEEE Trans. Speech Audio Process* 2000;8(2):159-167. [doi: [10.1109/89.824700](https://doi.org/10.1109/89.824700)]
38. Chen F, Loizou PC. Predicting the intelligibility of vocoded and wideband Mandarin Chinese. *The Journal of the Acoustical Society of America* 2011 May;129(5):3281-3290. [doi: [10.1121/1.3570957](https://doi.org/10.1121/1.3570957)] [Medline: [21568429](https://pubmed.ncbi.nlm.nih.gov/21568429/)]
39. Chen F, Wong LLN, Qiu J, Liu Y, Azimi B, Hu Y. The contribution of matched envelope dynamic range to the binaural benefits in simulated bilateral electric hearing. *J Speech Lang Hear Res* 2013 Aug;56(4):1166-1174. [doi: [10.1044/1092-4388\(2012/12-0255\)](https://doi.org/10.1044/1092-4388(2012/12-0255))] [Medline: [23926330](https://pubmed.ncbi.nlm.nih.gov/23926330/)]
40. Chen F, Hu Y, Yuan M. Evaluation of noise reduction methods for sentence recognition by Mandarin-speaking cochlear implant listeners. *Ear Hear* 2015 Jan;36(1):61-71. [doi: [10.1097/AUD.0000000000000074](https://doi.org/10.1097/AUD.0000000000000074)] [Medline: [25127321](https://pubmed.ncbi.nlm.nih.gov/25127321/)]
41. Lai Y, Tsao Y, Chen F. Effects of adaptation rate and noise suppression on the intelligibility of compressed-envelope based speech. *PLoS ONE* 2015 Jul 21;10(7):e0133519. [doi: [10.1371/journal.pone.0133519](https://doi.org/10.1371/journal.pone.0133519)]
42. Wang SS, Tsao Y, Wang HLS, Lai YH, Li PH. A deep learning based noise reduction approach to improve speech intelligibility for cochlear implant recipients in the presence of competing speech noise. 2017 Presented at: Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC); Dec 2017; Kuala Lumpur, Malaysia p. 808-812. [doi: [10.1109/APSIPA.2017.8282144](https://doi.org/10.1109/APSIPA.2017.8282144)]
43. Cohen J. Statistical power analysis for the behavioural sciences. In: Hillsdale, NJ: Laurence Erlbaum Associates. Cambridge, Massachusetts: Academic press; 1988:273-406.
44. Dien J. Issues in the application of the average reference: Review, critiques, and recommendations. *Behavior Research Methods, Instruments, & Computers* 1998 Mar;30(1):34-43. [doi: [10.3758/bf03209414](https://doi.org/10.3758/bf03209414)]
45. Williams LJ, Abdi H. Fisher's least significant difference (LSD) test. *Encyclopedia of research design* 2010:840-853. [doi: [10.1007/978-1-4020-6754-9_9279](https://doi.org/10.1007/978-1-4020-6754-9_9279)]

46. Williams LVD, Hinton G. Visualizing data using t-SNE. *Journal of machine learning research*. *Journal of Machine Learning Research* 2008;2579-2605.
47. Georgiev P, Lane ND, Mascolo C, Chu D. Accelerating mobile audio sensing algorithms through on-chip gpu offloading. 2017 Presented at: Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services; June 2017; New York, NY, USA p. 306-318. [doi: [10.1145/3081333.3081358](https://doi.org/10.1145/3081333.3081358)]
48. Chen J, Ran X. Deep learning with edge computing: a review. *Proc. IEEE* 2019 Aug;107(8):1655-1674. [doi: [10.1109/jproc.2019.2921977](https://doi.org/10.1109/jproc.2019.2921977)]

Abbreviations

ANOVA: analysis of variance
CI: cochlear implant
CJ: Construction Jackhammer
DDAE: deep denoising autoencoder
DL: deep learning
DNN: deep neural network
LPS: log power spectra
MRT: Mass Rapid Transit
NC: noise classifier
NC+DDAE_T: noise classifier + deep denoising autoencoder with knowledge transfer
ND: noise-dependent
NI: noise-independent
NR: noise reduction
PESQ: perceptual evaluation of speech quality
SEM: standard error of the mean
SNR: signal-to-noise ratio
SSN_IIEEE: speech shape noise from the Institute of Electrical and Electronics Engineers
STOI: short-time objective intelligibility
TMHINT: Taiwan Mandarin version of the hearing in noise test
t-SNE: *t*-distributed stochastic neighbor embedding
WCR: word correct rate

Edited by R Kukafka; submitted 09.11.20; peer-reviewed by YC Chu, ST Tang; comments to author 30.11.20; revised version received 11.02.21; accepted 27.04.21; published 28.10.21

Please cite as:

Li LPH, Han JY, Zheng WZ, Huang RJ, Lai YH

Improved Environment-Aware-Based Noise Reduction System for Cochlear Implant Users Based on a Knowledge Transfer Approach: Development and Usability Study

J Med Internet Res 2021;23(10):e25460

URL: <https://www.jmir.org/2021/10/e25460>

doi: [10.2196/25460](https://doi.org/10.2196/25460)

PMID:

©Lieber Po-Hung Li, Ji-Yan Han, Wei-Zhong Zheng, Ren-Jie Huang, Ying-Hui Lai. Originally published in the *Journal of Medical Internet Research* (<https://www.jmir.org>), 28.10.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the *Journal of Medical Internet Research*, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.