Improving the performance of hearing aids in noisy environments based on deep learning technology

Ying-Hui Lai, Wei-Zhong Zheng, Shih-Tsang Tang, Shih-Hau Fang, *Member, IEEE*, Wen-Huei Liao, and Yu Tsao, *Member, IEEE*

***Abstract*— The performance of a deep-learning-based speech enhancement (SE) technology for hearing aid users, called a deep denoising autoencoder (DDAE), was investigated. The hearing-aid speech perception index (HASPI) and the hearing-aid sound quality index (HASQI), which are two well-known evaluation metrics for speech intelligibility and quality, were used to evaluate the performance of the DDAE SE approach in two typical high-frequency hearing loss (HFHL) audiograms. Our experimental results show that the DDAE SE approach yields higher intelligibility and quality scores than two classical SE approaches. These results suggest that a deep-learning-based SE method could be used to improve speech intelligibility and quality for hearing aid users in noisy environments.**

I. INTRODUCTION

Hearing aids are common devices used to improve speech audibility for individuals with sensorineural hearing loss (SNHL) [1, 2]. However, previous studies [3, 4] indicated that the performance of current hearing aids still has room for improvement. For example, the poor performance of hearing aids in noisy conditions has deterred people from using them [3]. In order to solve this problem, speech enhancement (SE) aims to improve speech intelligibility and quality for hearing aid users under noisy listening conditions.

The purpose of SE is to enhance the original speech signal and remove background noise, which can be achieved by unsupervised or supervised approaches. A common unsupervised SE approach consists of designing filters that minimize a specific distortion between the original speech signal and its enhanced counterpart. Some well-known filter-based SE approaches are the log-minimum mean squared error (logMMSE) [5] and the Wiener filter [6]. Another unsupervised SE method is the Karhunen-Loève transform (KLT) [7, 8]. This method divides the noisy signal into clean and noisy subspaces and subsequently minimizes the noise components appearing in the clean subspace. Detailed information on the above unsupervised SE method can be found in [5-9]. These well-known unsupervised SE approaches can provide notable benefits under stationary noise conditions. However, there is still room for performance improvement, particularly when the noise signal changes quickly in real-world scenarios [10, 11].

More recently, deep learning methods have demonstrated outstanding performance in a wide variety of regression [10, 12-17] and pattern classification [18] tasks. Following the success of deep learning technology, Chen et al. [19] used 10,000 noises to train deep neural network (DNN) model to estimate the ideal ratio mask [15] to filter out noise from a noisy speech, and the results demonstrated that the proposed method provided better speech intelligibility for normal hearing and hearing-impaired subjects. Lu et al. proposed a deep-denoising-autoencoder-based (DDAE-based) approach for SE based on regression architecture. The DDAE method casts SE as a high dimensional nonlinear encoder-decoder task to remove noise components and achieve clean speech features based on deep neural network techniques. Previous studies indicate that the DDAE SE approach shows better performance than conventional SE approaches [20]. More recently, Lai et al. tested the performance of the DDAE SE approach in cochlear implants based on matched and mismatched DDAE models (i.e., using different noise and speech samples in the training and testing phases) [11, 21]. Objective evaluation and listening tests indicated that, under challenging listening conditions, the DDAE approach yields higher intelligibility scores than the two classical SE techniques (i.e., KLT and logMMSE). This work further investigates the effectiveness of DDAE SE for individuals with high-frequency hearing loss (HFHL) in noisy environments.

The rest of the paper is organized as follows. DDAE techniques are briefly introduced in Section II. Section III presents the methods and results. Finally, Section IV summarizes our findings.

II. DEEP-LEARNING-BASED SE APPROACH

Fig. 1 shows the structure of the DDAE SE model. The DDAE model consists of an offline phase and an online phase. In the offline phase, a set of noisy-clean speech pairs are prepared. The fast Fourier transform (FFT) is used to convert the signal from the time domain to the frequency domain, and the log power spectrum (LPS) features are obtained. The noisy ($\boldsymbol{Y}_n^{\mathrm{LPS}}$) and clean ($\boldsymbol{X}_n^{\mathrm{LPS}}$) LPS features are subsequently placed at the input and output sides of the DDAE model, and *n*

Ying-Hui Lai is with the Department of biomedical engineering, National Yang-Ming University, Taipei, Taiwan. (e-mail: yh.lai@ym.edu.tw).

Shih-Hau Fang and Wei-Zhong Zheng is with the Department of Electrical Engineering, Yuan Ze University (YZU), and MOST Joint Research Center for AI Technology and All Vista Healthcare, Taiwan.

Shih-Tsang Tang is with the Department of Biomedical Engineering, Ming Chuan University, Taoyuan, Taiwan.

Wen-Huei Liao is with the Department of Otolaryngology-Head and Neck Surgery, Taipei Veterans General Hospital, Taipei, Taiwan and the Department of Otolaryngology-Head and Neck Surgery, School of Medicine, National Yang-Ming University, Taipei, Taiwan.

Yu Tsao is with the Research Center for Information Technology Innovation at Academia Sinica, Taipei, Taiwan.

978-1-5386-3646-6/18/$31.00

denotes the frame index for the short-time Fourier transform [14]. For a DDAE model with k hidden layers, we have:

$$\begin{gathered} h^1(\boldsymbol{Y}_n^{\text{LPS}}) = \sigma(\boldsymbol{W}^0\boldsymbol{Y}_n^{\text{LPS}} + \boldsymbol{b}^0), \\ \vdots \\ h^k(\boldsymbol{Y}_n^{\text{LPS}}) = \sigma(\boldsymbol{W}^{k-1}h^{k-1}(\boldsymbol{Y}_n^{\text{LPS}}) + \boldsymbol{b}^{k-1}), \\ \widehat{\boldsymbol{X}}_n^{\text{LPS}} = \boldsymbol{W}^k h^k(\boldsymbol{Y}_n^{\text{LPS}}) + \boldsymbol{b}^k, \end{gathered} \tag{1}$$

where $\{\boldsymbol{W}^0 \ldots \boldsymbol{W}^k\}$ are the matrices of the connection weights, $\{\boldsymbol{b}^0 \ldots \boldsymbol{b}^k\}$ are the bias vectors for the k hidden layers, and $\widehat{\boldsymbol{X}}_n^{\text{LPS}}$ is the vector containing the logarithmic enhanced speech amplitudes corresponding to their noisy counterpart $\boldsymbol{Y}_n^{\text{LPS}}$. In addition, the sigmoid activation function [22] was used in this study:

$$\sigma(z) = \frac{1}{1+e^{-z}}, \tag{2}$$

where z is the output of each unit in this neural network. If z is a very large negative number, then the output is approximately 0, and if z is a very large positive number, then the output is approximately 1.

Finally, the DDAE parameters are determined by optimizing the following objective functions:

$$\begin{gathered} \theta^* = \arg\min_{\theta}(F(\theta) + \eta^0\|\boldsymbol{W}^0\|_2^2 + \cdots + \eta^k\|\boldsymbol{W}^k\|_2^2), \\ F(\theta) = \tfrac{1}{N}\sum_{n=1}^{N}\left\|\boldsymbol{X}_n^{\text{LPS}} - \widehat{\boldsymbol{X}}_n^{\text{LPS}}\right\|_2^2, \end{gathered} \tag{3}$$

where N is the total number of training samples (noisy-clean pairs). In the online phase, the DDAE parameters (obtained from the training phase) are used to transfer noisy speech ($\tilde{y}$) to the enhanced speech signal ($\tilde{x}$). Details of DDAE SE can be found in [20].

III. Methods and Results

A. Materials

Two hundred and fifty IEEE sentences were used in this study and were obtained from [9]. From these 250 sentences, 200 sentences were used for training, and the other 50 sentences were used for testing. In addition, 100 noise types [10, 23] were used in the training set to train DDAE by corrupting the 200 training sentences with these 100 noise types at three different signal-to-noise ratio (SNR) levels (i.e., −4, 1, and 6 dB). In this study, fast-changing non-stationary noise (N1), shown in Fig. 2 (B), was used to corrupt the test sentences at five SNR levels (−3, 0, 3, 6, and 9 dB).

B. Experimental Setup and Procedure

Two typical HFHL audiograms [24, 25] were used in this study to test the performance of the DDAE SE method: Audiogram 1 = {0, 0, 0, 60, 80, 90} and Audiogram 2 = {0, 15, 30, 60, 80, 85} at {0.25 kHz, 0.5 kHz, 1 kHz, 2 kHz, 4 kHz, 6 kHz}. It should be noted that these two audiograms correspond to common HFHL types in individuals with SNHL [26]. The DDAE SE results were compared with two other conventional SE approaches (i.e., KLT and logMMSE). The first step in this study was to train the DDAE SE model. The DDAE SE model used in this study had five hidden layers with 1024 neurons in each layer. Note that a speech framing strategy with a 16 ms window and an 8 ms frame shift were applied to each speech sample. Each windowed speech segment was processed with a 256-point FFT and then converted to an LPS feature vector with 129 dimensions.

We compared the performance of the DDAE and the conventional SE methods on speech intelligibility and quality based on two well-known evaluation metrics: the hearing-aid speech perception index (HASPI) [27] and the hearing-aid sound quality index (HASQI) [28]. Previous studies have shown that these two evaluation metrics can capture speech

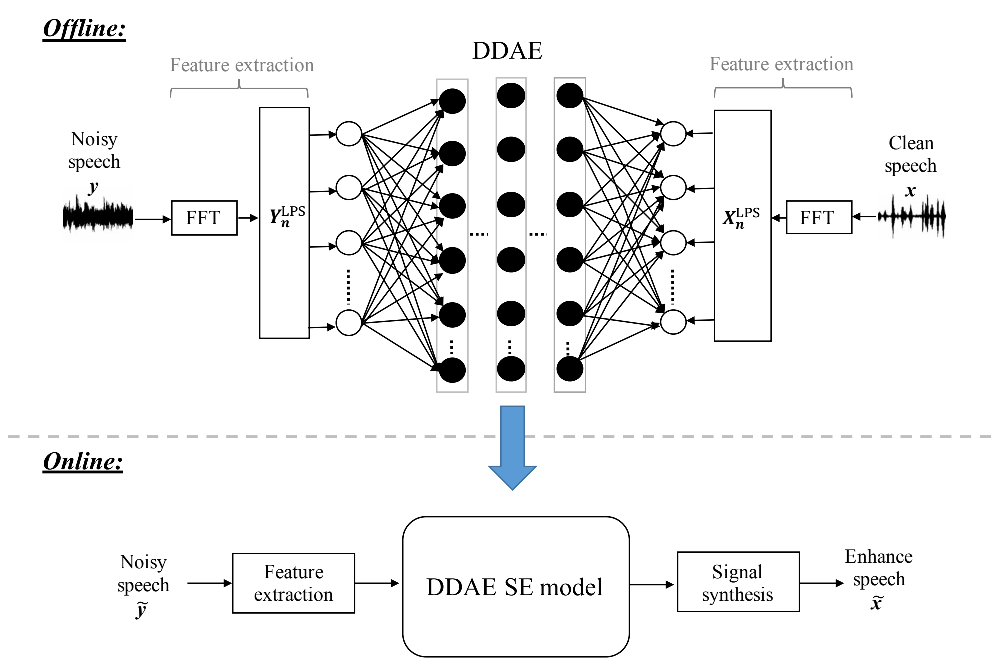

Figure 1. *DDAE-based speech enhancement system.*

intelligibility and quality when speech is subjected to a variety of distortions commonly found in hearing aids [27, 28].

HASPI uses an auditory model to predict intelligibility that incorporates the aspects of normal and impaired peripheral auditory functions [29]. The first step requires calculating the correlation values c between the enhanced spectral shape of the signal over time and the clean signal. Auditory coherence was used to measure cross-correlation in the high-level portions (denoted a_{high}) of the enhanced and the clean signals in each frequency band. The envelope is sensitive to the dynamic signal behavior associated with consonants, while the cross-correlation tends to preserve the harmonics in steady-state vowels. Finally, the HASPI score was calculated based on c and a_{high}. More detailed information on HASPI can be found in [27]. HASQI is the product of two independent components. The first component, called Q_{nonlin}, captures nonlinear distortion and noise. The second component, called Q_{lin}, captures linear filtering and spectral changes. These two components quantify specific differences in cochlear model representations of the clean reference signal and the enhanced signal. In order words, HASQI can predict the device's performance on sound quality based on the hearing threshold of individuals with hearing loss. More detailed information on HASQI can be found in [28, 30].

In this study, conventional SE results from the KLT and logMMSE methods were used to compare the performance of DDAE SE approaches in hearing aid users. For each approach, 250 test sentences (50 utterances at five different SNR levels) were used to compare speech intelligibility and quality based on the HASPI and HASQI metrics.

C. Comparison of Spectrograms

Spectrogram is a common tool used to analyze the spectral-temporal representations of a time-varying signal [31]. Fig. 2 contains six sub-figures showing the spectrograms of (A) clean speech, (B) noise, and (C) noisy speech at an SNR of 0 dB, and the enhanced speech processed by the (D) KLT, (E) logMMSE, and (F) DDAE methods. The clean speech used was extracted from a male voice saying "*The juice of lemons makes fine punch.*" in English. Fig. 2 shows that the DDAE SE method can effectively remove residual noise (present between 1.2 and 2.9 seconds).

It is possible to obtain a worse speech intelligibility performance using conventional SE methods due to inaccurate noise estimation (so-called: noise tracking), which is used to compute noise statistics [32, 33]. Problematic residual noise or speech distortions may occur if the noise statistics are not accurately estimated, thereby resulting in poor speech intelligibility and quality for hearing aid users. Numerous well-known noise estimation algorithms, such as voice activity detection [34], the minima-controlled recursive algorithm (MCRA) [35], Martin's algorithm [36], and the improved MCRA (IMCRA) [37] have been proposed over the past two decades [32]. These noise estimation algorithms provide satisfactory SE performance against stationary noise. However, most of these algorithms do not perform well in challenging conditions (e.g., nonstationary and fast-changing noise) [32]. The above-mentioned reasons may explain the unsatisfactory performance achieved by conventional NR methods under nonstationary conditions, especially for fast-changing background noise, as shown in Fig. 2 (B).

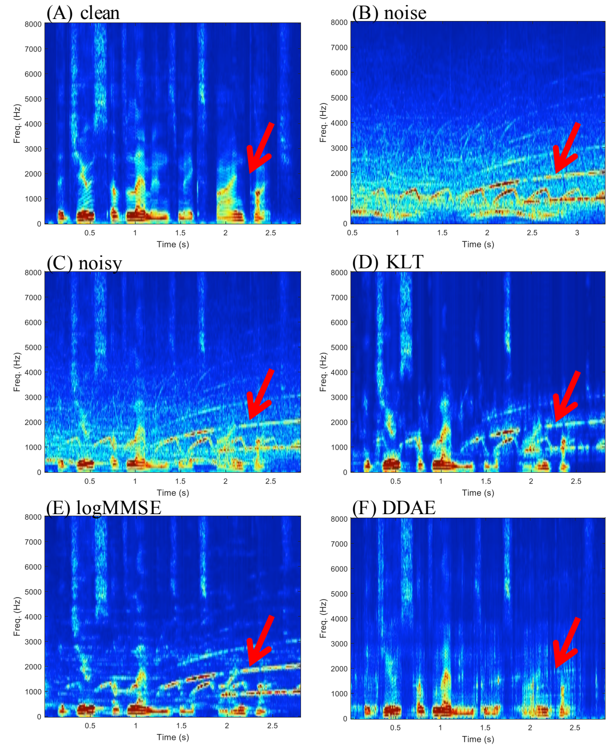

Figure 2. Spectrograms (A) to (C) show the clean speech signal, N1 background noise, and a noisy sentence (0 dB SNR), respectively. (D) to (F) show the sentence enhanced by KLT, logMMSE, and DDAE, respectively.

D. Analysis of t-SNE

We used the *t*-distributed stochastic neighbor embedding *(t*-SNE) [38] method to compare the performance of each SE approach (i.e., noisy, KLT, logMMSE, and DDAE). The *t*-SNE method is a machine learning algorithm for dimensionality reduction and is based on a nonlinear dimensionality reduction technique. It is particularly well-suited for embedding high-dimensional data into a space of two (or three) dimensions, which can then be visualized in a scatter plot. Details on this method can be found in [38]. We examined the speech processed by different SE methods using *t*-SNE on the 50 test sentences at an SNR level of 0 dB, as shown in Fig. 3. Each red and yellow point represents clean and processed speech information during a given timeframe (i.e., 16 ms), respectively. The method projected the 129-dimension LPS features onto two dimensions. In Fig. 3 (A), one can observe that the LPS features shown in red (clean speech) and the yellow (noisy speech) point clouds are split, and the processed speech (i.e., processed by KLT and logMMSE) is similar to clean speech. This can be seen as an overlap between the red and yellow points. However, many red and yellow points are still separated. In other words, the conventional SE methods could not enhance the noisy speech to the point where it is similar to clean speech. On the other hand, the DDAE SE method provided an LPS feature distribution of the processed speech closer to that of clean speech when compared with the conventional SE methods shown in Fig. 3. This implies that the DDAE SE method yielded results that are closer to the clean speech signals than

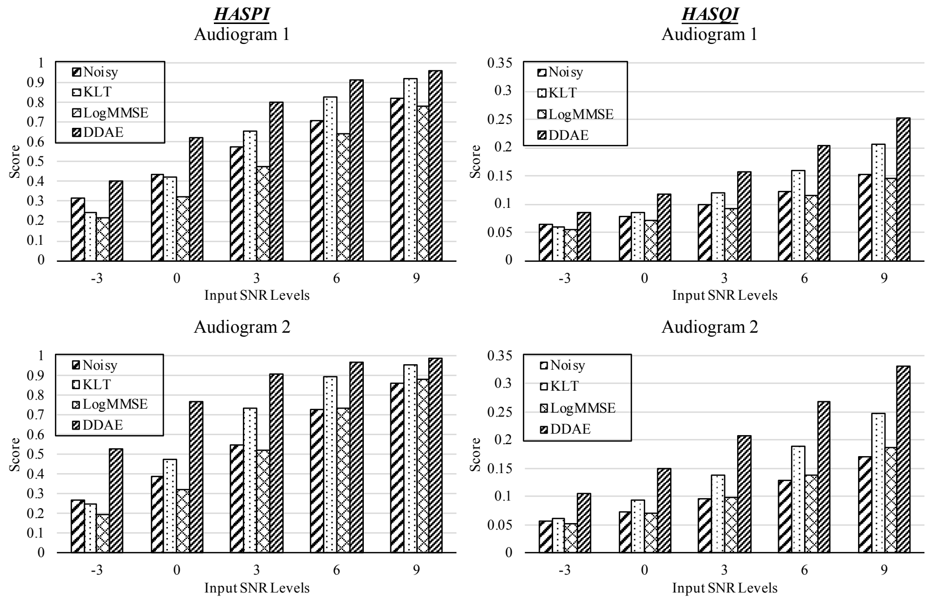

Figure 4. Subfigures (A) and (C) show the results of HASPI evaluation, and subfigures (B) and (D) show the results of HASQI evaluation, for Audiograms 1 and 2.

the conventional SE approaches. Therefore, this method could provide better speech intelligibility and quality for hearing aid users in noisy environments. Note that these four subfigures in Fig. 3 were obtained using the *t*-SNE method, trained four times individually. The objective function of *t*-SNE was minimized using a gradient descent optimization method that was initiated randomly. Therefore, the clean speech distributions in these four subfigures are different, despite the fact that they used the same clean data. Further details can be found in [38-40].

E. Speech Intelligibility and Quality Analysis Based on HASPI and HASQI

Higher HASPI and HASQI scores (the maximum is 1 and the minimum is 0) represent better speech intelligibility and quality for hearing aid users, respectively. Fig. 4 shows the average HASPI and HASQI scores of the original noisy signal and the KLT, logMMSE, and DDAE signals at −3, 0, 3, 6, and 9 dB. As shown in Fig. 4 (A) and (C), the proposed DDAE SE approach achieved higher HASPI scores than the conventional SE approaches (KLT and logMMSE) for HFHL at all five SNR levels. Regarding the sound quality results, Fig. 4 (B) and (D) show that the DDAE SE approach provides better speech quality performance than the conventional SE approaches. The HASPI and HASQI evaluation metrics further confirm that the DDAE SE approach provides better speech intelligibility and quality for individuals with HFHL in non-stationary noise environments, such as N1 noise shown in Figure 2 (B).

IV. Conclusions

In this study, the speech intelligibility and quality produced by the DDAE SE method were analyzed based on two well-known evaluation metrics applied to two typical HFHL audiograms. The results were compared with those of the classic KLT and logMMSE approaches for SE. The results show that the DDAE method yields higher HASPI and HASQI scores than those of the KLT and logMMSE methods under all test conditions. Thus, the DDAE SE method could provide better speech intelligibility and quality for individuals with hearing loss. These findings show that deep-learning-based SE (e.g., DDAE) can be used to improve speech intelligibility and quality for hearing aid users.

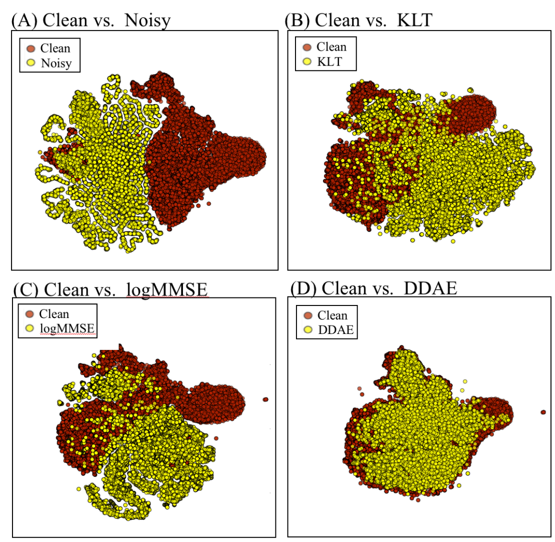

Figure 3. *t*-SNE embedding using 129-dimension LPS features to analyze (A) clean-noisy pairs, (B) clean-KLT pairs, (C) clean-logMMSE pairs, and (D) clean-DDAE pairs. Note that each of the red and yellow points represents the clean and the processed speech information during a given timeframe (i.e., 16 ms), respectively.

Acknowledgment

This work was supported by the Ministry of Science and Technology, Taiwan, under Grant MOST 105-2218-E-010-005-MY2, 106-2221-E-010-021 and MOST 107-2634-F-155-001. This work was also supported by VTA107-V1-8-1 and VTA107-V1-8-2.

REFERENCES

[1] H. Dillon, *Hearing aids (2nd), Chapter 1: introductory concepts*, pp. 1-19: Thieme, 2012.

[2] S. Gatehouse, G. Naylor, and C. Elberling, "Benefits from hearing aids in relation to the interaction between the user and the environment," *International Journal of Audiology,* vol. 42, no. sup1, pp. 77-85, 2003.

[3] S. Kochkin, "MarkeTrak VIII: Consumer satisfaction with hearing aids is slowly increasing," *The Hearing Journal,* vol. 63, no. 1, pp. 19-20, 2010.

[4] D. Wang, "Deep learning reinvents the hearing aid," *IEEE Spectrum,* vol. 54, no. 3, pp. 32-37, 2017.

[5] Y. Ephraim, and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech and Signal Processing,* vol. 33, no. 2, pp. 443-445, 1985.

[6] P. Scalart, "Speech enhancement based on a priori signal to noise estimation." pp. 629-632, 1996.

[7] A. Rezayee, and S. Gazor, "An adaptive KLT approach for speech enhancement," *IEEE Transactions on Speech and Audio Processing,* vol. 9, no. 2, pp. 87-95, 2001.

[8] Y. Hu, and P. C. Loizou, "A generalized subspace approach for enhancing speech corrupted by colored noise," *IEEE Transactions on Speech and Audio Processing,* vol. 11, no. 4, pp. 334-341, 2003.

[9] P. C. Loizou, *Speech enhancement: theory and practice*: CRC press, 2013.

[10] Y. Xu, J. Du, L. R. Dai, and C. H. Lee, "A Regression Approach to Speech Enhancement Based on Deep Neural Networks," *IEEE Transactions on Audio, Speech, and Language Processing,* vol. 23, no. 1, pp. 7-19, 2015.

[11] Y. H. Lai, F. Chen, S. S. Wang, X. Lu, Y. Tsao, and C. H. Lee, "A Deep Denoising Autoencoder Approach to Improving the Intelligibility of Vocoded Speech in Cochlear Implant Simulation," *IEEE Transactions on Biomedical Engineering*, vol. 64, pp. 1568-1587, 2017.

[12] X. Lu, Y. Tsao, S. Matsuda, and C. Hori, "Ensemble modeling of denoising autoencoder for speech spectrum restoration," *in Proc. Interspeech*, pp. 885-889, 2014.

[13] X. Lu, M. Unoki, S. Matsuda, C. Hori, and H. Kashioka, "Controlling tradeoff between approximation accuracy and complexity of a smooth function in a reproducing kernel Hilbert space for noise reduction," *Transactions on Signal Processing,* vol. 61, no. 3, pp. 601-610, 2013.

[14] Y. Wang, A. Narayanan, and D. Wang, "On training targets for supervised speech separation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing,* vol. 22, no. 12, pp. 1849-1858, 2014.

[15] A. Narayanan, and D. Wang, "Ideal ratio mask estimation using deep neural networks for robust speech recognition," *in Proc. ICASSP*, pp. 7092-7096, 2013.

[16] S. W. Fu, Y. Tsao, and X. Lu, "SNR-Aware Convolutional Neural Network Modeling for Speech Enhancement," *in Proc. Interspeech*, pp. 3768-3772, 2016.

[17] T. Goehring, F. Bolner, J. J. Monaghan, B. van Dijk, A. Zarowski, and S. Bleeck, "Speech enhancement based on neural networks improves speech intelligibility in noise for cochlear implant users," *Hearing Research*, 2016.

[18] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, and T. N. Sainath, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Processing Magazine,* vol. 29, no. 6, pp. 82-97, 2012.

[19] J. Chen, Y. Wang, S. E. Yoho, D. Wang, and E. W. Healy, "Large-scale training to increase speech intelligibility for hearing-impaired listeners in novel noises," *The Journal of the Acoustical Society of America,* vol. 139, no. 5, pp. 2604-2612, 2016.

[20] X. Lu, Y. Tsao, S. Matsuda, and C. Hori, "Speech enhancement based on deep denoising autoencoder," *in Proc. Interspeech*, pp. 436-440, 2013.

[21] Y. H. Lai, Y. Tsao, X. Lu, F. Chen, Y. T. Su, K. C. Chen, Y. H. Chen, L. C. Chen, Lieber P. H. Li, and C. H. Lee, "Deep Learning based Noise Reduction Approach to Improve Speech Intelligibility for Cochlear Implant Recipients," *Ear and Hearing, 2018.*

[22] X. Glorot, A. Bordes, and Y. Bengio, "Deep Sparse Rectifier Neural Networks." *In Proc. 14th International conference on artificial intelligence and statistics*, pp. 315-323, 2011.

[23] G. Hu. "100 nonspeech environmental sounds.", 2004. [http://web.cse.ohio-state.edu/pnl/corpus/HuNonspeech/HuCorpus.html]

[24] Y. H. Lai, P. C. Li, K. S. Tsai, W. C. Chu, and S. T. Young, "Measuring the Long-Term SNRs of Static and Adaptive Compression Amplification Techniques for Speech in Noise," *Journal of the American Academy of Audiology,* vol. 24, no. 8, pp. 671-683, 2013.

[25] D. Byrne, H. Dillon, T. Ching, R. Katsch, and G. Keidser, "NAL-NL1 procedure for fitting nonlinear hearing aids: characteristics and comparisons with other procedures," *Journal of the American Academy of Audiology,* vol. 12, no. 1, 2001.

[26] L. Ciletti, and G. A. Flamme, "Prevalence of Hearing Impairment by Gender and Audiometric Configuration: Results From The National Health and Nutrition Examination Survey (1999-2004) and The Keokuk County Rural Health Study (1994-1998)," *Journal of the American Academy of Audiology,* vol. 19, no. 9, pp. 672-685, 2008.

[27] J. M. Kates, and K. H. Arehart, "The hearing-aid speech perception index (HASPI)," *Speech Communication,* vol. 65, pp. 75-93, 2014.

[28] J. M. Kates, and K. H. Arehart, "The hearing-aid speech quality index (HASQI)," *Journal of the Audio Engineering Society,* vol. 58, no. 5, pp. 363-381, 2010.

[29] J. Kates, "An auditory model for intelligibility and quality predictions." in Proc. Mtgs. Acoust., p. 050184, 2013.

[30] A. Kressner, D. Anderson, and C. Rozell, "Evaluating the Generalization of the Hearing Aid Speech Quality Index (HASQI)," *IEEE Transactions on Audio, Speech, and Language Processing,* vol. 21, no. 2, pp. 407-415, 2013.

[31] S. Haykin, *Advances in spectrum analysis and array processing (vol. III)*: Prentice-Hall, Inc., 1995.

[32] S. Rangachari, and P. C. Loizou, "A noise-estimation algorithm for highly non-stationary environments," *Speech communication,* vol. 48, no. 2, pp. 220-231, 2006.

[33] Y. Tsao, and Y. H. Lai, "Generalized maximum a posteriori spectral amplitude estimation for speech enhancement," *Speech Communication,* vol. 76, pp. 112-126, 2016.

[34] J. Sohn, N. S. Kim, and W. Sung, "A statistical model-based voice activity detection," *IEEE Signal Processing Letters,* vol. 6, no. 1, pp. 1-3, 1999.

[35] I. Cohen, and B. Berdugo, "Noise estimation by minima controlled recursive averaging for robust speech enhancement," *Signal Processing Letters, IEEE,* vol. 9, no. 1, pp. 12-15, 2002.

[36] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *Speech and Audio Processing, IEEE Transactions on,* vol. 9, no. 5, pp. 504-512, 2001.

[37] I. Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," *Speech and Audio Processing, IEEE Transactions on,* vol. 11, no. 5, pp. 466-475, 2003.

[38] L. v. d. Maaten, and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research,* vol. 9, no. Nov, pp. 2579-2605, 2008.

[39] L. v. d. Maaten. "*t*-SNE," https://lvdmaaten.github.io/tsne/.

[40] L. Van Der Maaten, "Accelerating t-SNE using tree-based algorithms," *Journal of machine learning research,* vol. 15, no. 1, pp. 3221-3245, 2014.