Article

A Machine Learning Approach to Fitting Prescription for Hearing Aids

S I M M Raton Mondol and Sangmin Lee *

Department of Electronic Engineering, Inha University, Incheon 22212, Korea

* Correspondence: sanglee@inha.ac.kr

Received: 8 May 2019; Accepted: 24 June 2019; Published: 28 June 2019

Abstract: A successful Hearing-Aid Fitting (HAF) is more than just selecting an appropriate Hearing Aid (HA) device for a patient with Hearing Loss (HL). The initial fitting is given by the prescription based on user's hearing loss; however, it is often necessary for the audiologist to readjust some parameters to satisfy the user demands. Therefore, in this paper, we concentrated on a new application of Neural Network (NN) combined with a Transfer Learning (TL) strategy to develop a fitting algorithm with the prescription database for hearing loss and readjusted gain to minimize the gap between fitting satisfaction. As prior information, we generated the data set from two popular hearing-aid fitting software, then fed the training data to our proposed model, and verified the performance of the architecture. Pondering real life circumstances, where numerous fitting records may not always be accessible, we first investigated the number of minimum fitting records required for possible sufficient training. After that, we evaluated the performance of the proposed algorithm in two phases: (a) NN with refined hyper parameter showed enhanced performance in compare to state-of-the-art DNN approach, and (b) the TL approach boosted the performance of the NN algorithm in a broad way. Altogether, our model provides a pragmatic and promising tool for HAF.

Keywords: hearing-aid fitting; neural network; transfer learning; NAL-NL1; NAL-NL2

1. Introduction

Hearing-Aid Fitting (HAF) formulas estimate the required electro-acoustic amplification for individuals with hearing loss [1]. The National Acoustic Laboratories (NAL) procedures (NAL-NL1 [2,3]; NAL-NL2 [4,5]) and the Desired Sensation Level (DSL) [6] procedures (DSL v4.1 [7] and DSL m[i/o] [8]) have been used extensively for prescribing hearing-aid gain in recent time.

NAL-NL1 [2,3], the first prescription fitting procedure for prescribing nonlinear gain was developed by National Acoustic Laboratories (NAL) in 1999. The purely theoretical fitting formula aims at maximizing speech intelligibility while ensuring the overall loudness does not exceed the loudness that perceived by a normal-hearing person. Like its progenitor NAL-NL1, the second generation prescription procedure of NAL-NL2 [4,5] also aims to make speech intelligible and overall loudness comfortable. However, the theoretical derivation of NAL-NL2 differs from that of NAL-NL1 in two major points: (a) Intelligibility model and (b) gain constraint. The result of this is that NAL-NL2 prescribes relatively more gain in high and low frequencies than at the mid frequencies. Moreover, NAL-NL2 takes the profile of the hearing-aid user's age, gender, experience, language, and compressor speed into consideration [4]. On the other hand, the DSL fitting [6] aims at comfortable listening levels to maximize the speech recognition performance in each frequency region. The procedure uses the desired sensation levels to calculate its empirical real-ear aided gain.

Standard machine learning approaches to predict hearing aid fitting could be a potential solution for audiologists in clinical applications. Recently, deep learning is becoming a mainstream technology in solving major issues that have been stacked for ages in the artificial intelligence (AI) community.

It has turned out to be very promising at discovering complicated high-dimensional data structures and is, thus, applicable to the domains of business, administration, and science at the industrial scale [9]. In addition to beating records in image recognition [10,11] and speech recognition [12–14], it has beaten other machine-learning-based approaches at predicting the activity of living organism, DNA sequencing to bioimaging, and neuroimaging and in different biological applications [15]. Researchers believe that, in the near future, deep machine learning will have many more successes because it requires very light engineering by hand, so it can easily take advantage of the increasing amount of available data and computing. New learning algorithms that are currently being developed for deep neural networks will only accelerate this progress to a large extent. However, in many real world applications, like the availability of hearing-aid fitting data is limited due to privacy concerns. In addition, the data can be in a different feature space or follow a different data distribution. In such cases, the knowledge transfer would significantly improve the learning performance by avoiding very costly data labeling efforts [16].

Many applications can be found where transfer learning can be truly beneficial. One example is cancer drug sensitivity prediction [17], where a subset of transferred data from the auxiliary data of the related task and the target training data is used to build a model. The resulting transfer learning model is then applied on the test set of the target task to make predictions. In Reference [18], transfer learning with deep network is used to predict salient area in natural video. The knowledge learned with the deep network on a large data set is transferred to train the network on a small data set to predict salient areas. In Reference [19], a Deep Neural Network based fitting prescription for hearing aids has been offered, but the information about detailed methodology, techniques, and how this is accomplished are proprietary and unpublished. In this paper, we propose a transfer learning approach for the application of hearing-aid fitting, where our proposed approach combined several techniques. Our contribution and the structure of the paper are as follows:

- First, we extracted the fitting data set from two popular nonlinear hearing-aid fitting software (NAL-NL1 and NAL-NL2) and divided the randomly shuffled data set in two groups for training two different models.
- Second, we investigated the number of minimum fitting records required for sufficient training and observed that more than 1500 records are essential.
- Third, we trained the NN model with improved hyper parameters where we used random weight initialization and exponential weight decay concept and the approach shows enhanced performance.
- Fourth, we applied the inductive parameter transfer learning approach in our second model in a smaller data set condition with the same hyper parameter. We transferred the final weights from the NN model as the initial weights for the TL model, and it performed surprisingly well over the traditional NN model.
- Finally, we compared and analyzed the output of the two models for verification.

The model acquired in this research was selected by reviewing the top performing machine learning algorithm as applied to hearing-aid fitting prescription. As no other papers were found on the topic of transfer learning based fitting algorithm, all of our research on engineering features to estimate the expected gain for hearing loss patient is fancy.

2. Background

In this section, we introduced basic concepts of neural network and transfer learning, as the background of our proposed NN-TL based algorithm.

2.1. Fully Connected Neural Networks

Deep Neural Networks (DNN) are distinguishable from other types of Neural Networks by their depth, computational complexity, and performance. In a fully connected feed-forward DNN

model [20], nodes are fully connected with all the nodes in the previous layer. The computational procedure of a fully connected DNN involved with matrix–vector arithmetics and transformation by the activation function can be expressed as follows:

$$Y = \sigma(W^T \cdot X + b) \quad (1)$$

where Y and X are the outputs of the current layer and previous layer, respectively; W is the weight matrix; b is the bias vector; and σ is the activation function.

Each node in a neural network does a weighted sum of all of its inputs, adds a constant named bias, and then feeds the result through some nonlinear activation function (e.g., sigmoid, softmax, or relu). Nodes from the final layer produces the prediction, and a loss function measures the distance between the actual output and predicted output. The connections in a neural network can be shown in Figure 1.

Figure 1. Neural network connections.

2.2. Transfer Learning

In most applications in the real world, it is expensive or impossible to collect the necessary training data and to rebuild different models. Here comes the knowledge transfer or transfer learning approach that could reduce the effort of data collection. The main aim is to extract the knowledge from one domain and to apply the knowledge in another domain.

In transfer learning *what to transfer*, *when to transfer*, and *how to transfer* are three main research issues. Inductive transfer learning, transductive transfer learning, and unsupervised transfer learning are three popular sub-settings in this approach. Based on “what to transfer”, the approaches to transfer learning in the above sub-settings can be further classified into four classes [16]. The specific class used in this paper can be referred to as the parameter-transfer approach. The parameters (weights) sharing concepts have been widely used in different machine learning models. In general, very common hard weight sharing uses previously trained weights as the initial weights of a neural network model. A general parameter-based transfer learning method can be shown as in Figure 2.

Figure 2. Parameter-based transfer learning method (weight transfer).

3. Proposed Architecture

A simple block diagram of our proposed hearing-aid fitting model is shown in Figure 3.

Figure 3. Block diagram of the proposed hearing-aid fitting.

The data set contains hearing loss information for 6 different frequencies for each of the individual subjects. Other than hearing loss and insertion gain information for the corresponding loss, the data set does not contain any other features. Therefore, we did a binary conversion of the hearing loss data as pre-processing and then positioned them based on frequency hierarchy to make more features (6 frequencies * 7 bit binary conversion * 6 positions) as shown in Figure 4.

Figure 4. Binary conversions of input features.

Then we applied the NN and TL concept as two separate models. Finally, we compare and verify the results of these two models.

3.1. Neural Network with Refined Hyper-Parameter

In the first part of our proposed design, we considered a simple 4-layer (252 units of input layer, 85 units of first and 30 units of second hidden layers, and 6 units of output layer) Multilayer Perceptron (MLP) Neural Network as shown in Figure 5.

Figure 5. Multilayer Perceptron Neural Network (MLP-NN) Model.

The architecture adopted here is similar to the DNN regression approach [21] that is based on a feed-forward neural network [20,22], having many levels of nonlinearities and allowing them to densely illustrate an immensely varying and distinctly nonlinear function. In Reference [21], the training involves an unsupervised pre-training and supervised fine-tuning. However, in our approach, we only considered the simple supervised training with a refined hyper-parameter.

We choose random weight initialization, ensuring the weights varied in the range of -0.5 to 0.5 (± 0.1) from input layer to first hidden layer, -1.0 to 1.0 (± 0.1) from first hidden layer to second hidden layer, and -1.5 to 1.5 (± 0.1) from second hidden layer to output layer. We also considered the exponential weight decay concept in our approach. After initializing all the parameters, in the training session, the feed forward propagation is calculated first, the back propagation is calculated second, then the parameter is updated, and loss function is calculated at the end.

3.1.1. Feed Forward Propagation

Feed forward neural network learned to map a fixed size input to a fixed size output. The results of weighted sums from hidden layers to the output layer passes through a nonlinear activation function. We use sigmoid ($\sigma = 1/(1 + e^{-x})$), the most classical activation function in all the layers. In the feed forward propagation, the output of two hidden layers and output layer can be expressed as follows:

First hidden layer:

$$H_j^1 = \sigma\left(\sum_{i=1}^{252} w_{ij} \cdot I_i\right) \quad (2)$$

Second hidden layer:

$$H_k^2 = \sigma\left(\sum_{j=1}^{85} w_{jk} \cdot H_j^1\right) \quad (3)$$

Output layer:

$$O_l = \sigma\left(\sum_{k=1}^{30} w_{kl} \cdot H_k^2\right) \quad (4)$$

where I_i is the input; w_{ij} , w_{jk} , and w_{kl} are the corresponding weight matrices; H_j^1 , H_k^2 , and O_l are the outputs of the first hidden layer to the output layer respectively; and σ is the activation function.

3.1.2. Error Calculation

The error is calculated with the simple subtraction as follows:

$$e_l = y_l - O_l \quad (5)$$

where y_l is the real sample output.

3.1.3. Back Propagation and Weight Update

To properly adjust the weight vector, we considered the mini-batch Stochastic Gradient Descent (SGD) [23] optimizer. To lower the computation cost and to get lower variance, we used 100 subset of our data set each time we trained the network. First, we get the gradients of each layer of the networks for the current mini-batch, and then, we use that gradients to update the weights of each layers of the networks. To make the update operation very easy, we just add the gradient of particular weight matrix to our existing weight matrix. In addition, to make the learning better, we scale the gradients with a suitable learning rate. This less-complicated procedure usually finds a good set of weights surprisingly well in comparison to far more elaborate optimization techniques [9].

3.2. Transfer Learning Approach

In the second part of our proposed design, we use the concept of simple inductive parameter transfer learning approach [16]. We use the same NN model that includes the same hyper-parameter here in the TL approach. However, instead of random weight initialization, we retained the final weights from NN model and employed them here as the initial weights of the Transfer Learning model as shown in Figure 6.

Figure 6. Parameter transfer learning model.

4. Experimental Results and Evaluation

The performance of our proposed fitting algorithm have been evaluated in different phases. In the first subsection, the complete information about the data set has provided. The the next section, we tried to find the minimum number of fitting records requires for sufficient training. After that, performance of the NN algorithm with a refined hyper-parameter and the performance of the TL approach have been evaluated. Evaluation with a Mean Squared Error (MSE) analysis and some statistical parameter has also been considered.

4.1. Data Sets

Hearing loss vs. insertion gain data have been extracted manually for 1100 subjects from the two renowned nonlinear hearing-aid fitting softwares; National Acoustic Laboratories' nonlinear fitting procedure, version 1 (NAL-NL1) and version 2 (NAL-NL2). We considered flat hearing loss type for each subject, and the hearing loss varied maximally by 3 dB within 6 frequency bands (250 Hz, 500 Hz, 1 KHz, 2 KHz, 4 KHz, and 8 KHz). The maximum and minimum hearing loss we considered were in between 10 dB and 70 dB respectively. We manually put the hearing loss information in the software and noted the corresponding insertion gain information for the 50 dB, 65 dB, and 80 dB input levels. The 1000 training and 100 test sets were selected randomly in such a way that no data appears more than once in the combined training and test sets. There were no correlation between the two data sets; hence, we considered them separately. The full data sets can be found in the Supplementary Materials.

4.2. Number of Fitting Records for Sufficient Training

Number of training data always plays an important role in the performance of the machine learning based algorithm. In general, the more the data set, the better the training [9]. However, pondering real life situations, a big number of fitting records from audiologist may not always be practical. Therefore, we did an experiment to find the minimum number of fitting data set required for sufficient training by varying the amount of fitting data from 2800 to 1000. It was noticed that, the more training data used, the smaller the gain difference became. Furthermore, a monotonic difference in the gain with the increase in hearing loss was examined. Then, we were convinced that about 1500 data are necessary for sufficient training.

4.3. Efficiency of Predicting Fitting Records

The algorithm performance for 3 different input levels (50 dB, 65 dB, and 80 dB) for NAL-NL1 are shown in Figures 7–9 and for NAL-NL2 are shown in Figures 10–12 correspondingly. The solid green line illustrates the insertion gain extracted from the fitting software, and the solid red and blue line illustrate the gain predicted by the NN and TL algorithms. On the other hand, the dashed line indicates the corresponding gain differences. Many of the acronyms used in the performance comparison are defined in the legends of the figures.

Figure 7. NAL-NL1: Performance comparison for NN and TL vs. fitting formula for the 50 dB input level. Comparisons are made for 250 Hz, 500 Hz, 1 KHz, 2 KHz, 4 KHz, and 8 KHz frequencies from top left to right to bottom left to right.

Even though the proposed NN approach have followed the trend of the fitting formula pretty well compared to the state-of-the-art approach [19], it is clearly visible that the TL approach has outperformed it and that the predictions were very close to the original fitting records extracted by the formula. In most of the cases, the differences were in between -2 dB and $+2$ dB. However, some unexpected gain fluctuation can be noticed in some cases but can be neglected.

Figure 8. NAL-NL1: Performance comparison for NN and TL vs. fitting formula for the 65 dB input level. Comparisons are made for 250 Hz, 500 Hz, 1 KHz, 2 KHz, 4 KHz, and 8 KHz frequencies from top left to right to bottom left to right.

Figure 9. NAL-NL1: Performance comparison for NN and TL vs. fitting formula for the 80 dB input level. Comparisons are made for 250 Hz, 500 Hz, 1 KHz, 2 KHz, 4 KHz, and 8 KHz frequencies from top left to right to bottom left to right.

Figure 10. NAL-NL2: Performance comparison for NN and TL vs. fitting formula for the 50 dB input level. Comparisons are made for 250 Hz, 500 Hz, 1 KHz, 2 KHz, 4 KHz, and 8 KHz frequencies from top left to right to bottom left to right.

Figure 11. NAL-NL2: Performance comparison for NN and TL vs. fitting formula for the 65 dB of nput level. Comparisons are made for 250 Hz, 500 Hz, 1 KHz, 2 KHz, 4 KHz, and 8 KHz frequencies from top left to right to bottom left to right.

Figure 12. NAL-NL2: Performance comparison for NN and TL vs. fitting formula for the 80 dB input level. Comparisons are made for 250 Hz, 500 Hz, 1 KHz, 2 KHz, 4 KHz, and 8 KHz frequencies from top left to right to bottom left to right.

4.4. Mean Square Errors Comparison

Predictive mean square error results are shown in Table 1. In the table, we described the prediction error rate of our proposed TL method compared with NN method for three different input levels, and the results were taken after the 3rd epoch for each cases. From the table, it is clear that the TL algorithm for each of the input level significantly outperforms the results of the NN algorithm. In addition, our proposed transfer-learning-based fitting algorithm obtains better predictions and thus had smaller MSE for all the cases. The average MSE for TL and NN were 0.8662% and 1.1792% respectively.

Table 1. MSE Comparison for NN and TL.

Input Level	NN (in %)	TL (in %)
50 dB	1.0914	0.7101
65 dB	1.0871	0.8330
80 dB	1.3593	1.0555
Average	1.1792	0.8662

4.5. Statistical Evaluation

Tables 2 and 3 use three statistical indicator (average, minimum, and maximum) to analyze the performance of TL algorithm over the NAL-NL1 and NAL-NL2 data set, respectively. In both of the tables, it is clearly noticeable that the differences were quite negligible except some cases that were highlighted in bold.

Table 2. Statistical analysis of NAL-NAL1.

Group		Input Level								
		50 dB			65 dB			80 dB		
		TL	NL1	Diff	TL	NL1	Diff	TL	NL1	Diff
250 Hz	Average	4.85	4.76	−0.09	4.10	4.02	−0.08	2.84	2.74	−0.10
	Min	0.49	0.50	0.01	0.63	1.00	0.37	0.93	0.50	−0.43
	Max	15.18	18.50	3.32	11.72	15.00	3.28	7.50	10.00	2.50
500 Hz	Average	13.04	13.04	0.00	9.14	9.19	0.05	4.03	4.15	0.12
	Min	3.39	2.00	−1.39	2.58	2.00	−0.58	0.84	1.00	0.16
	Max	23.32	26.00	2.68	16.93	19.50	2.57	10.96	13.50	2.54
1 KHz	Average	23.46	23.68	0.21	19.60	19.76	0.16	10.75	11.01	0.25
	Min	10.72	10.00	−0.72	8.38	7.00	−1.38	3.10	2.50	−0.60
	Max	34.66	37.50	2.84	29.73	32.50	2.77	20.15	23.00	2.85
2 KHz	Average	23.48	24.03	0.55	18.37	18.49	0.12	10.58	10.69	0.11
	Min	12.29	12.50	0.21	7.54	7.00	−0.54	2.48	2.00	−0.48
	Max	33.56	37.00	3.44	27.97	30.50	2.53	21.04	23.50	2.46
4 KHz	Average	19.87	20.16	0.29	15.22	15.05	−0.17	9.83	9.89	0.06
	Min	10.88	10.00	−0.88	6.05	4.00	−2.05	2.02	1.50	−0.52
	Max	27.80	30.00	2.20	22.78	24.50	1.72	18.74	20.00	1.26
8 KHz	Average	5.52	5.69	0.17	7.13	7.26	0.13	6.94	7.06	0.12
	Min	3.61	3.00	−0.61	4.47	4.50	0.03	4.53	4.00	−0.53
	Max	8.98	14.00	5.02	10.80	15.00	4.20	10.66	15.00	4.34

Table 3. Statistical analysis of NAL-NAL2.

Group		Input Level								
		50 dB			65 dB			80 dB		
		TL	NL2	Diff	TL	NL2	Diff	TL	NL2	Diff
250 Hz	Average	10.77	10.20	−0.56	5.40	4.96	−0.44	0.72	0.50	−0.22
	Min	1.78	0.00	−1.78	0.46	0.00	−0.46	0.05	0.00	−0.05
	Max	21.17	22.47	1.30	15.47	16.74	1.27	6.14	7.15	1.01
500 Hz	Average	12.03	11.44	−0.59	5.44	4.98	−0.46	0.72	0.48	−0.24
	Min	1.77	0.00	−1.77	0.43	0.00	−0.43	0.06	0.00	−0.06
	Max	24.48	26.36	1.88	16.46	17.94	1.48	6.06	7.16	1.10
1 KHz	Average	16.34	15.66	−0.68	10.58	10.55	−0.04	4.12	4.17	0.05
	Min	3.83	0.00	−3.83	2.21	0.00	−2.21	0.55	0.00	−0.55
	Max	28.97	32.06	3.09	20.84	23.38	2.54	11.72	13.15	1.43
2 KHz	Average	18.48	17.73	−0.74	12.09	11.94	−0.15	4.81	4.80	−0.02
	Min	6.05	1.72	−4.33	3.07	0.00	−3.07	0.69	0.00	−0.69
	Max	30.88	33.31	2.43	23.28	25.39	2.11	14.24	15.50	1.26
4 KHz	Average	18.91	18.46	−0.45	13.22	13.30	0.07	6.08	6.20	0.12
	Min	6.49	2.77	−3.72	3.40	0.00	−3.40	0.76	0.00	−0.76
	Max	30.54	33.63	3.09	24.03	27.10	3.07	15.74	17.75	2.01
8 KHz	Average	19.06	18.54	−0.51	14.97	14.91	−0.06	8.63	8.89	0.26
	Min	5.11	1.48	−3.63	3.39	0.00	−3.39	1.44	0.00	−1.44
	Max	32.51	35.78	3.27	27.27	30.31	3.04	19.67	22.00	2.33

5. Discussion

All the advanced and modern fitting methodologies are fortunate to provide at least some corrective gain to the hearing-impaired ear. Although not yet developed, the optimum fitting method would seek to restore all the dynamic acoustic properties lost through cochlear and conductive causes.

However, each patient functions as a unique experiment with variable and unpredictable outcomes. Therefore, the insertion gain calculated by fitting formulas might not be the preferred gain for a hearing loss patient, and there is no potential benchmark for an acceptable prediction error. The preferred gain can vary largely from subject to subject, male to female, young to old, or one geographical region to another. In addition, we also do not know how each fitting software calculates the insertion gain for a corresponding hearing loss information, thus it remains a black box.

Therefore, in this paper, we extracted the patient hearing loss and insertion gain data from fitting software and applied the machine learning strategies to see if they can follow the trend of fitting formulas. By applying this approach with the clinical data from audiologist, we can mitigate the gap between fitting satisfaction of a hearing loss patient. The collection of large hearing aid fitting data from audiologist for further investigation will be a immense challenge due to the regulations and privacy concerns. That is why we considered parameter transfer learning approach, so that we can deal with the smaller data set problem. The TL prediction error for insertion gain is basically the calculated difference between predictive gain and the original gain from fitting formula. Therefore, very small or no difference will be expected in ideal case.

6. Conclusions and Future Work

MLP-NN model with refined hyper-parameters empowers computational prototypes consisting of multiple processing layers to learn data representation in multiple levels of abstraction and TL approach speeds up the training and enhances the performance of the model even in a smaller data set condition. For future work, we will be using the hearing-aid fitting data that would be collected from the health care professionals. We are expecting more features in the data set that includes the patients' personal preference to the hearing-aid type. We will be considering more challenging and smaller data set conditions to get a more realistic impression about the performance of the proposed algorithm. More consideration will have to be provided to optimize and evaluate the proposed approach.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2079-9292/8/7/736/s1>. A "Readme.txt" file has also been included mentioning all the details about the included files.

Author Contributions: Conceptualization, S.I.M.M.R.M. and S.L.; data curation, S.I.M.M.R.M. and S.L.; formal analysis, S.I.M.M.R.M. and S.L.; funding acquisition, S.L.; investigation, S.I.M.M.R.M. and S.L.; methodology, S.I.M.M.R.M.; project administration, S.L.; resources, S.L.; software, S.I.M.M.R.M.; supervision, S.L.; validation, S.I.M.M.R.M.; visualization, S.I.M.M.R.M.; writing—original draft, S.I.M.M.R.M.; writing—review and editing, S.I.M.M.R.M. and S.L.

Funding: This work was supported by INHA UNIVERSITY Research Grant and by Basic Science Research Program (NRF-2016R1A2B4015370) through the National Research Foundation of Korea (NRF).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sandlin, R.E. *Textbook of Hearing Aid Amplification*; Singular Publishing Group, Thomson Learning: San Diego, CA, USA, 2000.
2. Dillon, H. NAL-NL1: A new procedure for fitting non-linear hearing aids. *Hear. J.* **1999**, *52*, 10–16. [[CrossRef](#)]
3. Byrne, D.; Dillon, H.; Ching, T.; Katsch, R.; Keidser, G. NAL-NL1 procedure for fitting nonlinear hearing aids: Characteristics and comparisons with other procedures. *J. Am. Acad. Audiol.* **2001**, *12*, 37–51. [[PubMed](#)]
4. Keidser, G.; Dillon, H.; Flax, M.; Ching, T.; Brewer, S. The NAL-NL2 prescription procedure. *Audiol. Res.* **2011**, *1*, 88–90. [[CrossRef](#)] [[PubMed](#)]
5. Keidser, G.; Dillon, H.; Carter, L.; O'Brien, A. NAL-NL2 Empirical Adjustments. *Trends Amplif.* **2012**, *16*, 211–223. [[CrossRef](#)] [[PubMed](#)]
6. Dillon, H. *Hearing Aids*; Boomerang Press: Turramurra, Australia, 2012; Volume 30, p. 169. [[CrossRef](#)]
7. Seewald, R.; Cornelisse, L.; Ramji, K.; Sinclair, S.; Moodie, K.; Jamieson, D. *DSL 4.1 for Windows*; Hearing Health Care Research Unit, University of Western Ontario: London, ON, Canada, 1997.
8. Seewald, R.; Moodie, S.; Scollie, S.; Bagatto, M. The DSL method for pediatric hearing instrument fitting: Historical perspective and current issues. *Trends Amplif.* **2005**, *9*, 145–157. [[CrossRef](#)] [[PubMed](#)]

9. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
10. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]
11. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 1–9. [[CrossRef](#)]
12. Hinton, G.; Deng, L.; Yu, D.; Dahl, G.; Mohamed, A.R.; Jaitly, N.; Senior, A.; Vanhoucke, V.; Nguyen, P.; Sainath, T.; et al. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Process. Mag.* **2012**, *29*, 82–97. [[CrossRef](#)]
13. Graves, A.; Mohamed, A.R.; Hinton, G. Speech Recognition with Deep Recurrent Neural Networks. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013. [[CrossRef](#)]
14. Zhang, Y.; Pezeshki, M.; Brakel, P.; Zhang, S.; Bengio, C.L.Y.; Courville, A. Towards End-to-End Speech Recognition with Deep Convolutional Neural Networks. *arXiv* **2017**, arXiv:1701.02720.
15. Mahmud, M.; Kaiser, M.S.; Hussain, A.; Vassanelli, S. Applications of Deep Learning and Reinforcement Learning to Biological Data. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 2063–2079. [[CrossRef](#)] [[PubMed](#)]
16. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [[CrossRef](#)]
17. Turki, T.; Wei, Z.; Wang, J.T.L. A transfer learning approach via procrustes analysis and mean shift for cancer drug sensitivity prediction. *J. Bioinform. Comput. Biol.* **2018**, *16*, 1840014. [[CrossRef](#)] [[PubMed](#)]
18. Chaabouni, S.; Benois-Pineau, J.; Ben Amar, C. Transfer learning with deep networks for saliency prediction in natural video. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1604–1608. [[CrossRef](#)]
19. Haruta, C.; Kamimura, Y.; Sunohara, M.; Tatenno, M.; Co, R. Deep Neural Network-based fitting prescription for hearing aids: The efficacy of the proposed system. In Proceedings of the 2016 International Hearing Aid Research Conference (IHCON2016), Tahoe City, CA, USA, 10–14 August 2016; p. 1.
20. Khadivi, P.; Tandon, R.; Ramakrishnan, N. Flow of Information in Feed-Forward Deep Neural Networks *arXiv* **2016**, arXiv:1603.06220.
21. Xu, Y.; Du, J.; Dai, L.R.; Lee, C.H. A Regression Approach to Speech Enhancement Based on Deep Neural Networks. *IEEE Trans. Audio Speech Lang. Process.* **2015**, *23*, 7–19. [[CrossRef](#)]
22. Zemouri, R. An evolutionary building algorithm for Deep Neural Networks. In Proceedings of the 2017 12th International Workshop on Self-Organizing Maps and Learning Vector Quantization, Clustering and Data Visualization (WSOM), Nancy, France, 28–30 June 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–7. [[CrossRef](#)]
23. Li, M.; Zhang, T.; Chen, Y.; Smola, A.J. Efficient mini-batch training for stochastic optimization. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining—KDD’14, New York, NY, USA, 24–27 August 2014; ACM Press: New York, NY, USA, 2014; pp. 661–670. [[CrossRef](#)]