INTERSPEECH 2021
30 August – 3 September, 2021, Brno, Czechia

Optimising Hearing Aid Fittings for Speech in Noise with a Differentiable Hearing Loss Model

Zehai Tu, Ning Ma, Jon Barker

University of Sheffield, Department of Computer Science, Sheffield, UK
{ztu3, n.ma, j.p.barker}@sheffield.ac.uk

Abstract

Current hearing aids normally provide amplification based on a general prescriptive fitting, and the benefits provided by the hearing aids vary among different listening environments despite the inclusion of noise suppression feature. Motivated by this fact, this paper proposes a data-driven machine learning technique to develop hearing aid fittings that are customised to speech in different noisy environments. A differentiable hearing loss model is proposed and used to optimise fittings with back-propagation. The customisation is reflected on the data of speech in different noise with also the consideration of noise suppression. The objective evaluation shows the advantages of optimised custom fittings over general prescriptive fittings.
Index Terms: Hearing aid speech processing, speech in noise, differentiable hearing loss model

1. Introduction

The most common complaint of hearing aid users is that they struggle to understand speech in noisy situations [1, 2]. This is despite the fact that hearing aids are able to provide sufficient amplification, and despite the fact that modern hearing aids often include noise suppression algorithms. Ultimately better source separation algorithms might solve this problem, but in this paper we investigate whether speech intelligibility in noise can be improved by data-driven approaches to parameter-tuning in current hearing aid designs. In particular we look at the potential for replacing traditional hearing aid fitting formulae with scene-dependent fitting algorithms.

Hearing aids are currently tuned for individuals using one of various prescriptive fitting formulae. These formulae have similar design goals which link the listener's audiogram (frequency dependent hearing thresholds) to a prescribed degree of frequency dependent amplification. The National Acoustic Laboratories' Revised (NAL-R) fitting [3] aims to achieve equal loudness across all frequency bands in linear hearing aids. The more recent NAL-NL1 and NAL-NL2 [4] were designed to work with hearing aids that employ non-linear dynamic range compression algorithms and are designed to theoretically maximise speech intelligibility while not making the signal louder than usual. They achieve this using a loudness model [5] and the speech intelligibility index (SII) [6]. Other fitting formulae include CAMEQ and CAMEQ2-HF [7, 8] and the widely used DSLv5 [9].

The fitting formulae approach to hearing aid gain setting is remarkably successful given that a single formulae is used to cover all listening conditions. However, the question naturally arises as to whether a better result could be achieved by using noise-dependent fittings, and if so, how should these fittings be optimised? This is particularly relevant now that environment classification algorithms are available to automatically detect whether a user is, say, in a domestic living room, in a noisy cafe or standing by a busy road intersection, and hence good be used to switch gain settings. It should also be considered that hearing aids now apply increasingly sophisticated (but imperfect) noise-reduction algorithms (e.g., adaptive filtering [10], spectral subtraction[11, 12], spatial filtering [13]) that can altered the signal in ways that have not been considered in the design of NAL-NL2 etc. Further, recent hearing aids are using environmental classification algorithms [14, 15] to allow the characteristics of the noise suppression algorithms to be tuned separately for different noise types [12]. This further complicates the requirements of a fitting formulae. Logically, hearing aid gains should be optimised in consideration of the perception of the *processed* noise-reduced signal that the hearing aid delivers.

This study explores the possibilities of improving current hearing aid performance by developing fittings specific to different listening environments and noise-reduction processing. This is achieved using a model of hearing loss and a data-driven machine learning method which, inspired by the recent success of deep learning techniques, is optimised with the back-propagation algorithm. Our previous work using a similar data-driven fitting optimisation framework, DHASP [16], has shown that hearing aid fittings optimised by a data-driven approach can achieve high intelligibility for speech in quiet conditions. This work extends the DHASP framework to speech in noise, and proposes a new differentiable hearing loss model for the optimisation. The hearing loss model is approximated to the Cambridge Auditory Group MSBG model [17, 18, 19, 20] from the recent released Clarity Challenge [21]. For evaluation, various objective evaluation metrics were used and the performances of the optimised general fittings and custom fittings along with an open-source well-recognised prescriptive fitting are compared across a variety of noise types and hearing abilities. The results show the advantages of the optimised custom fittings, suggesting the proposed method could bring potential improvements for current hearing aids.

This paper is organised as follows. Section 2 presents the proposed fitting optimisation method. Section 3 describes the evaluation metrics, the databases and the experimental setup. The results comparing fittings in three contrasting noise environments, with and without noise reduction are presented and discussed in Section 4. Section 5 concludes the paper and presents future directions.

2. Method

The overall workflow of the proposed method is shown in Fig. 1. The degraded signal represents a noisy speech signal processed with or without a noise suppression algorithm. To simulate the typical signal pathway of a hearing aid (HA) user, the degraded signal is enhanced by a HA processor before being processed by a hearing impaired (HI) model. Its difference to a reference signal processed by a normal hearing (NH) model is

Copyright © 2021 ISCA

http://dx.doi.org/10.21437/Interspeech.2021-1613

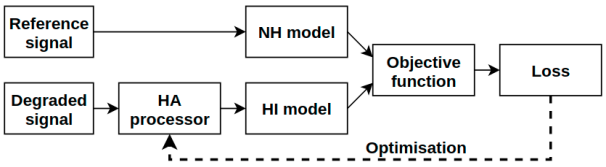

Figure 1: *Overall workflow*

measured and used as loss to optimise the HA processor with back-propagation. The NH and HI models are represented by a hearing loss model, whose characteristics are based on a listener's audiogram. Both the HA processor and the hearing loss models operate at a 44.1 kHz sampling rate, and all signals are presented at 65 dB as the sound pressure level (SPL) of normal conversation. A high-performance deep learning library PyTorch [22] is used for the implementation to retrieve the gradients for the optimisation.

2.1. Hearing-aid processor

As the SPL of input speech is constant at 65 dB, a finite impulse response (FIR) filter is used as the HA processor providing constant amplification excluding dynamic-range compression features. The amplification provided depends on six trainable parameters, which represent the insertion gains at [250, 500, 1000, 2000, 4000, 6000] Hz to be consistent with the audiograms. The frequency response is then obtained with linear interpolation, and iFFT is applied to retrieve the impulse response. A Hann window is subsequently multiplied with the impulse response.

2.2. Hearing loss model

The MSBG model can be considered as a hearing impairment simulator, which consists of the simulations of acoustic transformation between sound source and cochlea, spectral smearing, and loudness recruitment. The structure of the model is shown in Fig. 2. The hearing loss model proposed in this work is a differentiable approximation to the MSBG model, and the differences are in filter implementation and envelope retrieval. All infinite impulse response (IIR) filters in the MSBG model are approximated with FIR filters, and Hilbert transformation is used to extract the envelopes, so that the computation can be performed in parallel for fast optimisation using GPUs.

2.2.1. Source to cochlea transformation

The transformation of the sound pressure level from a sound source to the cochlea is derived from the combination of a free field and a middle ear transfer function. The free field transfer function [23] approximates the acoustic changes during sound propagation from the free field to the eardrum. The middle ear transfer function [24] simulates the acoustic alterations of sound in the middle- and inner-ear before arriving at the cochlea. The overall transformation is implemented using an FIR filter, whose frequency response is the combination of the free field and the middle ear frequency-gain tables.

2.2.2. Spectral smearing

Spectral smearing [17] is used in the hearing loss model to simulate reduced frequency selectivity, which is one of the major deficits in the sound analysis ability of cochlear hearing loss. Experimental results showed that this technique leads to little effect on the intelligibility of speech in quite, but a large effect on speech in noise [17] or interfering speech [18]. This is generally consistent with the phenomenon that impaired frequency selectivity contributes largely to the difficulty of understanding speech in noise for listeners with cochlea hearing loss.

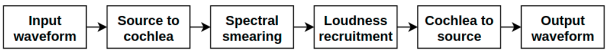

Figure 2: *Structure of the hearing loss model*

Input waveform signals are first processed by STFT with Hamming windows. Smearing is then performed to the power spectrogram, and the phase remains unchanged for iSTFT after smearing. Given the input spectrogram X and the output spectrogram Y, the spectrum smearing function is expressed as:

$$Y = A_N^{-1} A_W X, \tag{1}$$

where A_N and A_W are the matrices representing the normal and the widened auditory filterbanks, respectively. For each auditory filter within the filterbank, the form is given by:

$$W(g) = (1 + pg)\exp(-pg), \tag{2}$$

where $W(g)$ is the intensity weighting function describing the filter shape in the frequency domain, g is the frequency difference from the center frequency f_c divided by f_c, and p is the parameter determining the sharpness of the auditory filter [25]. The value of p is computed as:

$$p = \frac{4 f_c}{r \times \mathrm{ERB}}, \tag{3}$$

where ERB is the equivalent rectangular bandwidth [26] calculated as $24.7 \times (0.00437 f_c + 1)$, and the widening factor r differs for the lower and upper sides of the filter, denoted as r_l and r_u, respectively. The values of r_l and r_u are dependent on the degree of hearing loss. Each auditory filter is at last calibrated by dividing $24.7 \times \mathrm{ERB}(r_l + r_u)/2$ to remove an upward tilt in the excitation pattern, which is caused by the increasing of the bandwidth as the f_c grows.

2.2.3. Loudness recruitment

It is observed that the response of a damaged cochlea to low-level sounds is much smaller than a normal one, while the response to high-level sounds is roughly the same as normal [19]. This is simulated by a dynamic range compression mechanism.

A group of Gammatone filters are firstly used to extract the fine structures $x(n)$ of the smeared waveform signal. The i^{th} filter $h^{(i)}$ of filterbank is expressed as:

$$h^{(i)}(t) = A^{(i)} t^{\left(N^{(i)}-1\right)} e^{-2\pi b^{(i)} t} \cos\left(2\pi f^{(i)} t\right), \tag{4}$$

where $f^{(i)}$ is the centre frequency, $b^{(i)}$ is the bandwidth computed as $1.019 \times \mathrm{ERB}$, N is the order of the Gammatone filter (4 in this study), and $A^{(i)}$ is the amplitude to normalise the filters. The number of filters I differs according to the hearing abilities. The bandwidths of the filters are broadened two or three times for moderate or moderate to severe hearing loss, respectively. The outputs of the filters are aligned in the time domain to ensure the peaks for all channels are coincident for a pulse input. This alignment will make the mixture of the outputs of the filters generally sound almost identical to the input signal. The envelope $E(n)$ of each channel is retrieved with Hilbert transformation followed by a group of low pass filters for smoothing. The waveform output signal $y(n)$ is then recruited as:

$$y(n) = \sum_{i=1}^{I} \left(\frac{E^{(i)}(t)}{E_\theta} \right)^{\left(\frac{\theta}{\theta - \mathrm{HL}^{(i)}} - 1\right)} x^{(i)}(n), \tag{5}$$

where HL is the audiometric hearing loss in dB, θ is the maximal loudness threshold which is set 105 dB, and E_θ is the corresponding envelope magnitude.

2.2.4. *Cochlea-to-source transformation*

The recruited signal is lastly processed by a cochlea-to-source transfer FIR filter, whose frequency response is the additive inverse in dB of the frequency response of the source-to-cochlea transformation filter.

2.3. Objective function

Given the simulated reference signal $y_r(n)$, i.e. the clean normal hearing signal, and the processed hearing impaired signal $y_p(n)$, STFT is firstly applied to retrieve the corresponding spectrograms $Y_r(m,k)$ and $Y_p(m,k)$. The objective function consists of a spectrogram reconstruction loss L_{spec} and a sound pressure level loss L_{spl}. L_{spec} is expressed as:

$$L_{spec} = 20\log_{10}\left(\frac{1}{mk}\sum_{m,k}|Y_p(m,k) - Y_r(m,k)|\right), \quad (6)$$

and L_{spl} is computed as:

$$L_{spl} = 20\log_{10}\left(\sqrt{\frac{1}{n}\sum_n y_p(n)^2} - \sqrt{\frac{1}{n}\sum_n y_r(n)^2}\right). \quad (7)$$

The overall objective function is expressed as:

$$L = \begin{cases} L_{spec} + \alpha L_{spl}, & \text{if } L_{spl} \geq 0 \\ L_{spec}, & \text{otherwise} \end{cases}, \quad (8)$$

where α is a weighting coefficient. L_{spl} is used to prevent over-amplification, which could lead to listening discomfort for the listeners, and thus it is not applied if negative.

3. Experiments

3.1. Evaluation

The hearing-aid speech perception index (HASPI) [27] and hearing-aid speech quality index (HASQI) [28], which are developed based on a physiological auditory model, are used to evaluate the speech intelligibility and quality, respectively. Segmental frequency weighted signal-to-noise (FWSNR) ratio is also used for evaluation, and the implementation is from [29]. The FWSNRs are measured after the amplified signals processed by the MSBG model.

The open-source NAL-R prescription is used as the baseline prescription, which is widely recognised and more importantly tested in subjective experiments. To be consistent with the HA processor, the derivation of the NAL-R frequency response is based on the hearing losses at [250, 500, 1000, 2000, 4000, 6000] Hz. The Wiener filtering algorithm, popular for noise suppression in hearing aids, is used in this work as the noise suppression front end. Fittings optimised on clean data are regarded the general fittings (G). Meanwhile, the custom fittings include the fittings optimised on noisy data (Cn) and the fittings optimised on the Wiener filtering enhanced noisy data (Cw).

3.2. Databases

Three standard audiograms N1, N2, and N4 from [30] are used in this study, which represent three hearing loss categories: mild (N1), moderate (N2), and moderate to severe (N4). Their hearing losses at different frequencies are shown in Table 1.

Table 1: *Hearing losses of the audiograms used in this study.*

	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	6 kHz
N1	10 dB	10 dB	10 dB	15 dB	30 dB	40 dB
N2	20 dB	20 dB	25 dB	35 dB	45 dB	50 dB
N4	55 dB	55 dB	55 dB	65 dB	75 dB	80 dB

The noisy speech corpus from [31] is used, which was mixed using speech utterances from the the Voice Bank Corpus [32], and noises from the recordings of the first channel of the Demand database [33]. A set of 56 speakers is used for training, and another set of 28 speakers is divided into the validation and test sets. There are around 400 utterances from each speaker. 10 types of noises are mixed with the utterances, at the SNRs of 0, 5, 10, and 15 dB according to ITU-T P.56 standard [34]. All signals are filtered with a high pass filter whose cut-off frequency is 80 Hz to eliminate non-speech interference. Three categories of noises are selected in the experiments: traffic, kitchen, and babble. The traffic noise and kitchen noise are comparatively more stationary, and mainly distributed in low and high frequencies, respectively. The babble noise is less stationary, and its spectrum overlaps clean speech.

3.3. Experimental setup

The Wiener filter implementation is based on the method proposed in [35]. The parameter setting in the differentiable hearing loss model is consistent with the MSBG model. In the spectral smearing, $[r_l, r_u]$ are set as [4.0, 2.0], [2.4, 1.6], and [1.6, 1.1] for the mild, moderate, and moderate to severe hearing loss, respectively. In the loudness recruitment, the Gammatone filterbank consists of 36, 28, or 19 filters respectively for the three types of hearing loss. The HA processors are trained with a batch size of 128 for 500 epochs using the Adam optimiser with a learning rate of $1e$-2. The parameters are initialised as the NAL-R fitting. The weighting coefficient in Eq. 8 is set to 5.

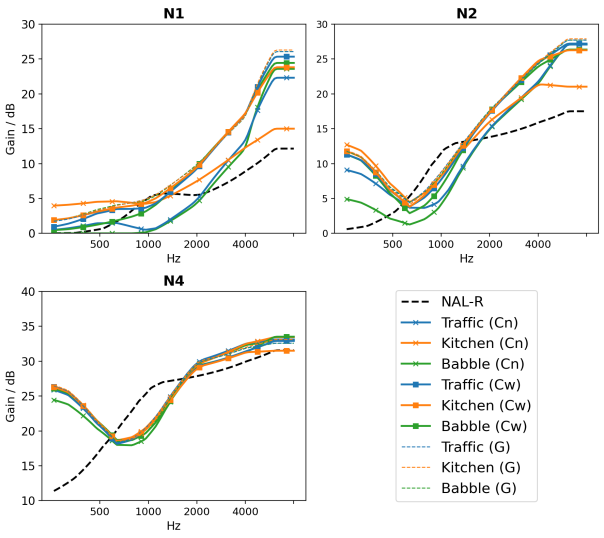

Figure 3: *Frequency responses of NAL-R fitting, custom fittings optimised with noisy data (Cn), custom fittings optimised with Wiener filtering enhanced noisy data (Cw), and the general fittings (G) for different hearing losses.*

Table 2: *Evaluation scores of various fittings applied to noisy speech before and after the enhancement of Wiener filtering. N: NAL-R prescriptive fittings, G: optimised general fittings, +W: using enhanced noisy speech by Wiener filtering, Cn: custom fittings optimised on noisy data, Cw+W: custom fittings optimised on Wiener filtering enhanced noisy data with Wiener filtering. The single best score in each group is indicated in bold.*

		Traffic						**Kitchen**						**Babble**					
		N	N+W	G	G+W	Cn	Cw+W	N	N+W	G	G+W	Cn	Cw+W	N	N+W	G	G+W	Cn	Cw+W
N1	HASPI	0.92	0.93	0.92	0.95	0.93	0.95	0.99	0.98	0.99	0.99	0.99	0.99	0.68	0.65	0.66	0.67	0.68	0.67
	HASQI	0.25	0.23	0.26	0.27	**0.28**	0.26	0.42	0.30	0.42	0.35	**0.44**	0.35	0.15	0.13	0.15	0.14	**0.16**	0.14
	FWSNR	5.40	6.08	5.82	6.40	**7.59**	6.22	9.54	8.77	7.42	8.65	**10.71**	8.76	4.26	4.21	4.41	4.36	**6.00**	4.23
N2	HASPI	0.87	0.81	0.88	0.91	0.89	0.91	0.97	0.91	0.98	0.98	0.98	0.98	0.57	0.50	0.59	0.59	**0.60**	0.59
	HASQI	0.18	0.13	0.20	0.19	**0.21**	0.19	0.23	0.17	0.30	0.24	**0.30**	0.24	0.11	0.08	0.12	0.10	0.12	0.10
	FWSNR	5.63	6.00	7.08	6.31	**7.61**	6.30	**9.64**	8.54	7.84	7.77	8.39	7.65	4.35	4.07	5.38	4.35	**5.48**	4.24
N4	HASPI	0.21	0.16	0.52	0.40	0.52	0.39	0.23	0.20	0.61	0.53	0.61	0.51	0.13	0.09	**0.31**	0.24	0.30	0.23
	HASQI	0.03	0.02	0.08	0.05	0.08	0.05	0.03	0.03	0.08	0.07	0.08	0.07	0.02	0.02	**0.06**	0.04	0.05	0.04
	FWSNR	0.91	2.06	5.12	4.44	**5.24**	4.40	1.86	2.21	6.55	5.71	**6.63**	5.65	0.37	1.01	3.52	2.81	**3.66**	2.80

4. Results and Discussions

Fig. 3 shows the frequency responses of the optimised general (G) and custom (Cn, Cw) fittings along with the baseline NAL-R prescription for the three hearing loss categories in various noise conditions. First, it can be observed that in general, the frequency responses of optimised fittings have higher gains in low and high frequencies and lower gains around 1 kHz than the NAR-R prescription. This is broadly consistent with the results of subjective hearing experiments reported in [36] and our previous findings using DHASP [16]. The subjective experiments for hearing-aid self-fitting showed that hearing impaired listeners prefer higher gain in high frequency, and lower gain around 1 kHz for speech in noise compared to the NAL-NL2 fitting, which provides even more gain in low and high frequencies and less gain in middle frequency than NAL-R.

Second, compared to the frequency responses of the processors optimised using noisy data (Cn), those optimised using Wiener filtering denoised data (Cw) are more similar to those optimised using clean data (G). This is expected as Wiener filtering is able to suppress the noise to some extent. Among the three noise types, the kitchen noise energy is mainly distributed in high frequencies, and thus the processors optimised in kitchen noise provide more gain in low frequency and less gain in high frequency. On the contrary, the processors optimised in traffic and babble noises, whose energy is mostly in low frequencies, show more gain in high frequency and less in low frequency. It can also be observed that as the hearing loss severity increases from N1 to N4, the differences among processors optimised with different settings are smaller.

Table 2 presents the HASPI, HASQI, and FWSNR scores of the various fittings evaluated in this study. The optimised fittings specific to different listening environments and noise-reduction processing outperform NAL-R in most of the evaluation scores. The improvement over NAL-R is in general larger for more severe hearing loss than that for mild hearing loss, but the benefit can be seen for all three listeners' audiograms and across different noise conditions. This shows there are potential advantages in listening-condition specific fittings that can be learned using the proposed data-driven approach.

Comparing the scores between the optimised custom fittings and the general prescriptive fittings, it is clear that the custom fittings produced higher FWSNRs than the general fittings in all cases The custom fittings also achieved overall higher or approximately equal HASPI and HASQI scores than the general fittings for mild and moderate hearing losses. For moderate to severe hearing loss in the babble noise condition, the general fitting scores are marginally better than the custom fitting scores. This could be due to the fact that the custom fittings provided overall higher insertion gain, and HASPI and HASQI are sensitive to signal presentation level when the hearing threshold is high, i.e., when the hearing loss is more severe.

It can also be seen that Wiener filtering does not lead to better performance for hearing loss based on the objective evaluation. Our direct measurement without the hearing loss model suggests that Wiener filtering can improve the SNR of noisy speech on average by 6.18, 8.62, and 5.06 dB for traffic, kitchen and babble noise, respectively. However, the HASPI scores suggest that only listeners with mild and moderate hearing losses can gain intelligibility benefit from Wiener filtering in the environment with traffic noise, where the noise is relatively stationary and distributed in low frequencies. For more severe hearing loss, or in the kitchen and babble noise, Wiener filtering did not improve the HASPI score. The SNR improvement by Wiener filtering also did not translate into improvement to the FWSNR in all the tested conditions. While it improves the FWSNR for NAL-R fittings in the traffic noise, there is no clear performance pattern for other noise environments. Overall, the objective experiments suggest the potential advantage of the custom hearing aid fitting optimisation, while challenging the intelligibility benefit of Wiener filtering for environmental noise suppression.

5. Conclusions

This paper has presented a data-driven method to include the effects of different noises and noise suppression techniques into the development of hearing aid fittings. It was argued that a data-driven hearing aid fitting algorithm can be more flexible than current prescribed fitting formulae. A differentiable hearing loss model was proposed for the data-driven optimisation and evaluated in different noise conditions both with and without noise reduction processing enabled. The objective evaluation metrics suggest that the optimisation based approach has potential to outperform prescribed fittings, and that noise-dependent optimisation is particularly promising, with greatest benefits for mild and moderate hearing losses.

So far the results have been evaluated using objective measurements alone. Based on the promise of these early-stage results, subjective evaluations with hearing impaired listeners are planned for the next phase of the work, including subjective comparisons between the optimisation objectives of the hearing loss model proposed in this work and HASPI based model in the DHASP [16].

6. References

[1] N. A. Lesica, "Why do hearing aids fail to restore normal auditory perception?" *Trends in Neurosciences*, vol. 41, no. 4, pp. 174–185, 2018.

[2] I. Brons, R. Houben, and W. A. Dreschler, "Acoustical and perceptual comparison of noise reduction and compression in hearing aids," *Journal of Speech, Language, and Hearing Research*, vol. 58, no. 4, pp. 1363–1376, 2015.

[3] D. Byrne and H. Dillon, "The national acoustic laboratories'(NAL) new procedure for selecting the gain and frequency response of a hearing aid," *Ear and Hearing*, vol. 7, no. 4, pp. 257–265, 1986.

[4] G. Keidser, H. Dillon, M. Flax, T. Ching, and S. Brewer, "The nal-nl2 prescription procedure," *Audiology Research*, vol. 1, no. 1, pp. 88–90, 2011.

[5] B. Moore and B. Glasberg, "A model of loudness perception applied to cochlear hearing loss," *Auditory Neuroscience*, vol. 3, no. 3, pp. 289–311, 1997.

[6] A. ANSI, "S3. 5-1997, methods for the calculation of the speech intelligibility index," *New York: American National Standards Institute*, vol. 19, pp. 90–119, 1997.

[7] B. Moore, B. Glasberg, and M. Stone, "Use of a loudness model for hearing aid fitting: III. a general method for deriving initial fittings for hearing aids with multi-channel compression," *British Journal of Audiology*, vol. 33, no. 4, pp. 241–258, 1999.

[8] B. C. Moore, B. R. Glasberg, and M. A. Stone, "Development of a new method for deriving initial fittings for hearing aids with multi-channel compression: CAMEQ2-HF," *International Journal of Audiology*, vol. 49, no. 3, pp. 216–227, 2010.

[9] S. Scollie, R. Seewald, L. Cornelisse, S. Moodie, M. Bagatto, D. Laurnagaray, S. Beaulac, and J. Pumford, "The desired sensation level multistage input/output algorithm," *Trends in Amplification*, vol. 9, no. 4, pp. 159–197, 2005.

[10] P. Vary and R. Martin, *Digital speech transmission: Enhancement, coding and error concealment*. John Wiley & Sons, 2006.

[11] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 27, no. 2, pp. 113–120, 1979.

[12] R. Bentler and L.-K. Chiou, "Digital noise reduction: An overview," *Trends in Amplification*, vol. 10, no. 2, pp. 67–82, 2006.

[13] H. Levitt, "Noise reduction in hearing aids: A review," *Journal of Rehabilitation Research and Development*, vol. 38, no. 1, pp. 111–122, 2001.

[14] L. Lamarche, C. Giguère, W. Gueaieb, T. Aboulnasr, and H. Othman, "Adaptive environment classification system for hearing aids," *The Journal of the Acoustical Society of America*, vol. 127, no. 5, pp. 3124–3135, 2010.

[15] P. Nordqvist and A. Leijon, "An efficient robust sound classification algorithm for hearing aids," *The Journal of the Acoustical Society of America*, vol. 115, no. 6, pp. 3033–3041, 2004.

[16] Z. Tu, N. Ma, and J. Barker, "DHASP: Differentiable hearing aid speech processing," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 296–300.

[17] T. Baer and B. C. Moore, "Effects of spectral smearing on the intelligibility of sentences in noise," *The Journal of the Acoustical Society of America*, vol. 94, no. 3, pp. 1229–1241, 1993.

[18] ——, "Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech," *The Journal of the Acoustical Society of America*, vol. 95, no. 4, pp. 2277–2280, 1994.

[19] B. C. Moore and B. R. Glasberg, "Simulation of the effects of loudness recruitment and threshold elevation on the intelligibility of speech in quiet and in a background of speech," *The Journal of the Acoustical Society of America*, vol. 94, no. 4, pp. 2050–2062, 1993.

[20] M. A. Stone and B. C. Moore, "Tolerable hearing aid delays. i. estimation of limits imposed by the auditory path alone using simulated hearing losses," *Ear and Hearing*, vol. 20, no. 3, pp. 182–192, 1999.

[21] S. Graetzer, J. Barker, T. J. Cox, M. Akeroyd, J. F. Culling, G. Naylor, E. Porter, and R. Viveros Muñoz, "Clarity-2021 challenges: Machine learning challenges for advancing hearing aid processing," *Interspeech*, 2021.

[22] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in Neural Information Processing Systems*, vol. 32, pp. 8026–8037, 2019.

[23] E. A. Shaw, "Transformation of sound pressure level from the free field to the eardrum in the horizontal plane," *The Journal of the Acoustical Society of America*, vol. 56, no. 6, pp. 1848–1861, 1974.

[24] M. C. Killion, "Revised estimate of minimum audible pressure: Where is the"missing 6 db"?" *The Journal of the Acoustical Society of America*, vol. 63, no. 5, pp. 1501–1508, 1978.

[25] B. C. Moore and B. R. Glasberg, "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," *The Journal of the Acoustical Society of America*, vol. 74, no. 3, pp. 750–753, 1983.

[26] B. R. Glasberg and B. C. Moore, "Derivation of auditory filter shapes from notched-noise data," *Hearing Research*, vol. 47, no. 1-2, pp. 103–138, 1990.

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

[28] ——, "The hearing-aid speech quality index (hasqi) version 2," *Journal of the Audio Engineering Society*, vol. 62, no. 3, pp. 99–117, 2014.

[29] P. C. Loizou, *Speech Enhancement: Theory and Practice*. CRC press, 2013.

[30] N. Bisgaard, M. S. Vlaming, and M. Dahlquist, "Standard audiograms for the iec 60118-15 measurement procedure," *Trends in Amplification*, vol. 14, no. 2, pp. 113–120, 2010.

[31] C. Valentini-Botinhao, X. Wang, S. Takaki, and J. Yamagishi, "Investigating rnn-based speech enhancement methods for noise-robust text-to-speech." in *SSW*, 2016, pp. 146–152.

[32] C. Veaux, J. Yamagishi, and S. King, "The voice bank corpus: Design, collection and data analysis of a large regional accent speech database," in *2013 international conference oriental COCOSDA held jointly with 2013 conference on Asian spoken language research and evaluation (O-COCOSDA/CASLRE)*. IEEE, 2013, pp. 1–4.

[33] J. Thiemann, N. Ito, and E. Vincent, "Demand: a collection of multi-channel recordings of acoustic noise in diverse environments," in *Proc. Meetings Acoust*, 2013.

[34] P. ITU-T, "Objective measurement of active speech level," *ITU-T Recommendation*, 1993.

[35] C. Plapous, C. Marro, and P. Scalart, "Improved signal-to-noise ratio estimation for speech enhancement," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 6, pp. 2098–2108, 2006.

[36] C. L. Mackersie, A. Boothroyd, and H. Garudadri, "Hearing aid self-adjustment: Effects of formal speech-perception test and noise," *Trends in Hearing*, vol. 24, 2020.