Interspeech 2024
1-5 September 2024, Kos, Greece

Signal processing algorithm effective for sound quality of hearing loss simulators

Toshio Irino[1], *Shintaro Doan*[1], *Minami Ishikawa*[1]

[1]Faculty / Graduate School of Systems Engineering, Wakayama University, Japan

irino@wakayama-u.ac.jp, doan.shintaro@g.wakayama-u.jp, ishikawa.minami@g.wakayama-u.jp

Abstract

Hearing loss (HL) simulators, which allow normal hearing (NH) listeners to experience HL, have been used in speech intelligibility experiments, but not in sound quality experiments due to perceptible distortion. If they produced less distortion, they might be useful for NH listeners to evaluate the sound quality of, for example, hearing aids. We conducted perceptual sound quality experiments to compare the Cambridge version of HL simulator (CamHLS) and the Wakayama version of the HL simulator (WHIS), which has the two algorithms of filterbank analysis synthesis (FBAS) and direct time-varying filter (DTVF). The experimental results showed that WHIS with DTVF produces less perceptible distortion in speech sounds than CamHLS and WHIS with FBAS, even when the nonlinear process is working. This advantage is mainly due to the use of the DTVF algorithm, which could be applied to various signal synthesis applications with filterbank analysis.

Index Terms: sound quality, hearing loss, filterbank, time-varying filter, nonlinear distortion

1. Introduction

In order to develop the next generation of assistive listening devices that can compensate for the difficulties experienced by elderly people with hearing impairment (HI), it is essential to clarify the dysfunctions. Many psychoacoustic experiments have been conducted by using relatively simple stimulus sounds, such as sinusoids and noise[1]. In addition, many speech sound experiments have been performed, although they have mainly been limited to intelligibility tests, such as the speech-in-noise test. However, it is not easy to determine the degradation factors, whether they are located in the periphery, within the auditory pathway, or due to cognitive factors, because of the considerable variability among HI listeners. To partially resolve this problem, a hearing loss (HL) simulator was developed to specify the effects of peripheral dysfunction, such as absolute threshold (AT) elevation and loudness recruitment [1], on speech intelligibility [2]. Normal-hearing (NH) listeners could evaluate the speech intelligibility of the HL-simulated sounds, which may be similar to what HI listeners perceive.

The development of such HL simulators has a long history. The most popular was developed by Moore's group at Cambridge University [3, 4] (hereafter CamHLS), which has been used in many psychoacoustic studies [5, 6, 7, 8]. There are many other HL simulators, e.g. [9, 10, 11, 12, 13], which are not commonly used in psychoacoustic experiments. With the goal of being used in various psychoacoustic experiments, an HL simulator was also developed by our group at Wakayama University [14, 15, 16] (hereafter WHIS). WHIS has been used for several speech perception experiments [17, 18]. Deutch and Fels [19] evaluated several HL simulators [11, 12, 13] and showed that WHIS simulated the audiogram better.

The psychoacoustic experiments using the HL simulators were mainly limited to speech intelligibility tasks due to the distortion caused by the nonlinear signal processing used to simulate the HL. Although the distortion is unavoidable in nonlinear signal processing, it is desirable to reduce the audible distortion to expand the range of experiments, such as speech and sound quality tasks. For example, the sound quality of hearing aids could be tested by NH listeners using the HL simulator with less distortion. However, it seems that the sound quality of HL simulator has not yet been evaluated, at least not psychoacoustically. Instead, the spectral distortion between the auditory model outputs of the HI condition and of the NH condition with the HL simulator was evaluated, and the latest version of WHIS [16] was shown to outperform CamHLS. However, it is still uncertain whether the spectral distortion is directory corresponding to the sound quality.

In this paper, we evaluate the speech and instrument sound quality of CamHLS and WHIS through subjective listening tests. They employ different signal processing algorithms. Based on the perceptual results, we discuss signal processing algorithm effective for sound quality of HL simulators.

2. Algorithms of HL simulators

This section briefly summarizes the signal processing of the CamHLS and WHIS to clarify the algorithm differences.

2.1. CamHLS

Figure 1(a) shows the block diagram of CamHLS. There are two stages: the frequency smearing stage using the Short-Time Fourier Transform (STFT) and the stage for simulating loudness recruitment and absolute threshold elevation. These are basically an analysis/synthesis method based on frequency component decomposition and summation with overlap-and-add. The frequency smearing process in the first stage is introduced to simulate bandwidth widening in HI listeners [5, 6]. This process produces perceptual distortion (i.e.,'grittiness' sound as described in the CamHLS software). This makes difficult to use CamHLS in sound quality experiments. The algorithm in the second stage is based on a loudness model [21] proposed to reflect the HL caused by the dysfunction of outer hair cell (OHC, HL_{OHC}) and inner hair cell (IHC, HL_{IHC}). They defined that the total HL, HL_{total}, that appears on the audiogram is assumed to be their sum on a dB scale:

$$HL_{total} = HL_{OHC} + HL_{IHC}. \quad (1)$$

Since the HL may not be caused by OHC and IHC alone, we have preferred to use active HL ($HL_{act} \geq 0$) and passive HL ($HL_{pas} \geq 0$) as

$$HL_{total} = HL_{act} + HL_{pas}. \quad (2)$$

10.21437/Interspeech.2024-111

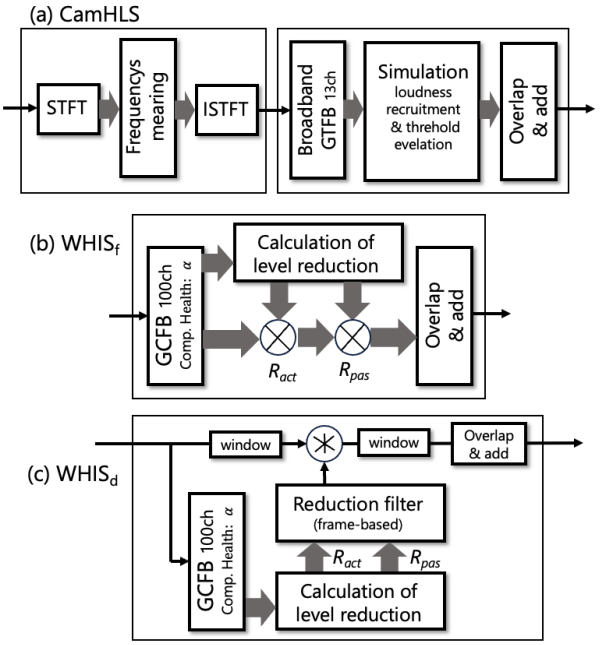


Figure 1: *Block diagrams of hearing loss simulators. See Section 2 for detail.*

2.2. WHIS

Figure 1(b) shows WHIS using the FilterBank Analysis/Synthesis (FBAS) method (hereafter, WHIS$_{\mathrm{f}}$) as in CamHLS. Figure 1(c) shows WHIS using the Direct Time-Varying Filter (DTVF) method (hereafter, WHIS$_{\mathrm{d}}$) [18, 16].

2.2.1. Analysis algorithm

The analysis part is the same for both types. The gammachirp filterbank (GCFB) is used to reflect the dysfunction of the active cochlear process via a parameter called compression health, α. The HL simulation is performed by calculating the level reduction of the active process, R_{act}, and the passive process, R_{pas}, in the cochlear input-output function. R_{act} depends on the input sound level and can be written as $R_{act}(f_c, L(\tau))$, where $L(\tau)$ is the estimated level at frame time τ in a filter frequency, f_c. $R_pas(f_c)$ is a constant. The total level reduction R_{total} is represented as the sum of these values on a dB scale:

$$R_{total}(f_c, L(\tau)) = R_{act}(f_c, L(\tau)) + R_{pas}(f_c). \quad (3)$$

It is similar to Eq. 2. In fact, at a filter frequency f_c, $R_{pas} = HL_{pas}$ because it is a constant and $R_{act}(L_{AT}) = HL_{act}$ where L_{AT} is the absolute threshold (AT) level.

Figure 2 illustrates the relationship between HL_{total}, HL_{act}, and HL_{pas} on the audiogram. In this case, HL_{total} is the average hearing level of 70-year-old male (hereafter, 70-yr) listeners [20]. HL_{act} ($\alpha = 0.5$) increases the hearing level up to the HL_{total} line. HL_{pas} fills the gap between HL_{total} and HL_{act}. If $\alpha = 1$, the cochlear active process is healthy and $HL_{act} = 0$. Therefore, $HL_{total} = HL_{pas}$ and it becomes a constant level reduction that can be simulated by a linear filter. If $\alpha < 1$, an nonlinear level-dependent filtering is required. When $\alpha = 0$, the active process is completely damaged, HL_{act} becomes maximum and HL_{pas} becomes relatively smaller.

Table1 shows how the 70-yr hearing level is divided into the sum of HL_{act} and HL_{pas} depending on the value of α. Note that the upper limit of HL_{act} is HL_{total} even if α is set to a low value in the software. For example, HL_{act}+HL_{pas} = 8+0

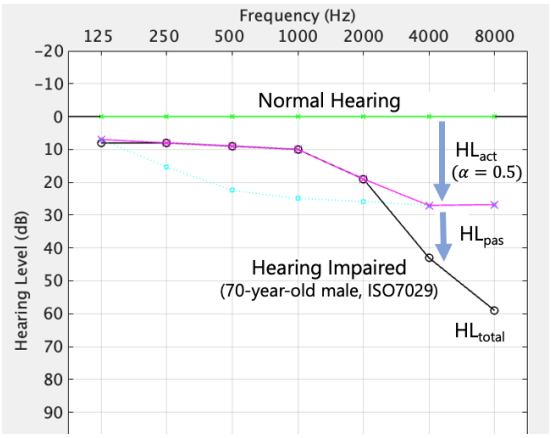


Figure 2: *Relationship between the hearing level, the active HL, and the passive HL in the audiogram (adapted from the GUI version of WHIS [16]). The green line shows the hearing level of NH; the black line shows the average hearing level of the 70-year-old male (70-yr) listeners [20]; the magenta line represents the active HL, HL_{act}, that appears on the audiogram. The difference between the magenta and black lines is the passive HL, HL_{pas}.*

Table 1: *Average hearing level (dB) of 70-yr [20] (i.e., HL_{total}). The lower three rows show the sum of the active and passive level reductions by WHIS on the audiogram: HL_{act} + HL_{pas}, which are dependent on the α values.*

Freq.	125	250	500	1000	2000	4000	8000
70-yr	8	8	9	10	19	43	59
$\alpha = 1$	0+8	0+8	0+9	0+10	0+19	0+43	0+59
$\alpha = 0.5$	8+0	8+0	9+0	10+0	19+0	27+16	27+32
$\alpha = 0$	8+0	8+0	9+0	10+0	19+0	43+0	44+15

at 125 Hz with $\alpha = 0.5$ and $\alpha = 0$. The difference between the α values of 0.5 and 0 appears in this case at high frequencies (4000 and 8000 /Hz).

2.2.2. Synthesis algorithm

The synthesis procedures of WHIS$_{\mathrm{f}}$ (Fig. 1(b)) and WHIS$_{\mathrm{d}}$ (Fig. 1(c)) are completely different. WHIS$_{\mathrm{f}}$ uses a common filterbank analysis/synthesis method. It is simple and widely used. However, it is important to align the phase lag, which depends on filter characteristics such as frequency and bandwidth, before performing the overlap and add (e.g. [22]). Incomplete alignment can cause distortion. WHIS$_{\mathrm{d}}$ was proposed to avoid this problem. A minimum phase filter which corresponds the frequency distribution of $R_{total}(f_c)$ is calculated for each time frame τ. The filter is convolved with the windowed input signal and then overlapped and added frame by frame. This method seems to be advantageous for sound quality because there is a single filter applied for each frame and it is not necessary to compensate the phase lag of multiple filters. However, it has not been confirmed in perceptual sound quality experiments.

3. Perceptual experiments

Subjective listening tests were conducted using a paired comparison test that is well established in psychoacoustics because our goal was to evaluate HL simulators that can be used in formal experiments. MUSHRA [23] is not suitable for this purpose because it was designed to quickly assess the quality of sound encoding, but sacrifices accuracy and reliability, as stated in the recommendation itself. In addition, MUSHRA requires

a hidden reference of the original or "high quality" sample and a hidden anchor filtered by a 3.5kHz low-pass filter. These requirements are unreasonable for the current evaluation because the sound pressure levels of the original samples are different and the HL simulation already has a low-pass characteristic.

3.1. Sound material

3.1.1. Speech sounds

We used the male voice from the FW07 database for Japanese four-mora (roughly four-syllable) words with word familiarity control [24, 25]. The words with the highest familiarity were used for this experiment, as no lexical judgement was made. We randomly selected 32 words: 10 for the main experiment, 10 for the practice session, and 12 for the training session. After preliminary listening, we decided to apply a room impulse response to the word sounds to make it easier to detect the distortion. Moreover, people with HI usually hear sounds in everyday environments, not just dry sources. The impulse response was derived from the database of the University of Aachen [26]. The conditions were an office room, speaker to microphone distance of 3m and a reverberation time of 0.48s. The duration of the word sounds was about 0.8 s.

3.1.2. Instrument sounds

The original music consisted of simple three consecutive triads (three-note chords) with the same instrument to allow the listener to focus on detecting the distortion. The duration of each chord was 1 s and the total stimulus was 3 s. Four instrument sounds were created in GarageBand (Apple): Tuba, Grand Piano, Cheap Organ, and French Horn. The three triads in Roman numeral analysis are $I^6 - V^{+6} - I^6$ chords in the major key [27]. The highest note of these chords was F3 for Tuba and C4 for the other instruments. These four stimuli were used for the main session. In addition, tones with one or more lower semitones were created for the training and practice sessions. Note that the total duration of the triads was 3 s, which was longer than that of the speech sounds, so the number of sound pairs was reduced to keep the duration of the experiment within a reasonable range.

3.1.3. HL simulation

The speech and instrument sounds were processed using CamHLS, $\mathrm{WHIS_d}$, and $\mathrm{WHIS_f}$ with the 70-yr condition shown in Table 1. The reason for choosing this condition was to ensure effective experiments by balancing the increase in nonlinear distortions and the decrease in the audibility of small distortions with simulated age. Although the condition is limited due to the time-consuming paired comparison experiments, the results would be valid for other HL profiles since the theoretical study has already shown that the spectral distortion measures for a wide range of HL profiles are smaller for WHIS than for CamHLS [16]. We have used the CamHLS default settings as supplied with the software. For $\mathrm{WHIS_d}$ and $\mathrm{WHIS_f}$, it is possible to control the degree of dysfunction of the nonlinear active process by the α value, as described in Section 2.2. After some preliminary listening, we set $\mathrm{WHIS_d}$ to α values of 1, 0.5, and 0 (hereafter, $\mathrm{WHIS_d^{(1)}}$, $\mathrm{WHIS_d^{(0.5)}}$, $\mathrm{WHIS_d^{(0)}}$) and $\mathrm{WHIS_f}$ to the α value of 0.5 ($\mathrm{WHIS_f^{(0.5)}}$). Note that $\mathrm{WHIS_d^{(1)}}$ is virtually equivalent to a linear low-pass filter with the 70-yr hearing level frequency response as described in Section 2.2. Therefore, there is no nonlinear distortion due to the HL simulation. As the α value decreases, the distortion increases due to the effect of the nonlinear process. Our concern was how much of the nonlinear distortion would be perceived by the listener.

The source sound level was set to L_{eq} of 70 dB to produce sufficient level for subjective judgments of distortion after the HL simulation. The output levels were dependent on the simulation conditions and could affect the judgments. Therefore, the sound levels for each word and instrument were normalized to the level processed by $\mathrm{WHIS_d^{(1)}}$. The final levels were approximately the same: L_{eq} was 60.8 (SD ± 0.1) dB for speech and 61.1 (SD ± 0.2) dB for instrument.

3.2. Experimental procedure

The sound quality evaluation was performed using a paired comparison method. The speech experiment was conducted first, followed by the instrument experiment for the same participants. The presentation of the sound pairs and the response collection were performed using a set of web pages [28]. Participants had to respond to the interval in which the sound contained more distortion. For each of the speech and instrument sounds, all combinations and their reverse order of pair sounds for five HL simulation conditions (${}_5C_2 \times 2 = 20$) were prepared. Thus, in the main experiment, the participants judged 200 pairs of 10 word sounds and 80 pairs of 4 instruments in random order.

3.2.1. Training and practice sessions

It was necessary to familiarize the participants with the sound quality judgment because they were naive to such experiments. In the training session for the speech experiment, we used a total of 12 pairs: 3 pairs of $\mathrm{WHIS_d^{(1)}}$ and $\mathrm{WHIS_f}$, 3 pairs of $\mathrm{WHIS_d^{(1)}}$ and CamHLS, and their reverse order. This is because we clearly perceived nonlinear distortions in $\mathrm{WHIS_f}$ and CamHLS, while $\mathrm{WHIS_d^{(1)}}$ contains no nonlinear distortions. In the training session for the instrument experiment, we used a total of 8 pairs due to the relatively long sounds: 4 pairs of $\mathrm{WHIS_d^{(1)}}$ and $\mathrm{WHIS_f}$, 4 pairs of $\mathrm{WHIS_d^{(1)}}$ and CamHLS. Participants judged which sound had more distortion and received feedback from the collect response. After listening to the all the sound pairs, the test score was calculated and the participants who scored 83% (=10/12) for speech and 88% (=7/8) for instrument passed this session and proceeded to the practice session. If they did not pass, the training session was repeated. In the practice session, they listened to sound pairs that were the same as in the main session, except that different words or different sounds of the same instruments were used. There were 20 speech sound pairs and 10 instrument sound pairs.

3.2.2. Participants and audio settings

The experiment was conducted with fifteen young NH listeners (aged 21–23 years) who read information about the experiments and gave informed consent. The experiments were approved by the Ethics Committee of Wakayama University (No. 2015-3, Rei01-01-4J, and Rei02-02-1J). Participants were seated in a sound-attenuated room with a background noise level of approximately 26dB in L_{Aeq}. They had a hearing level of less than 20 dB between 125 Hz and 8,000 Hz. The sounds were presented diotically through a DA-converter (SONY, NW-A55) via headphones (SONY, MDR-1AM2). The sound was presented as 48 kHz and 16-bit wav files through the web page using Google Chrome. Level calibration was performed using an artificial ear (Brüel & Kjær, Type 4153), a microphone (Brüel & Kjær, Type 4192), and a sound level meter (Brüel & Kjær, Type 2250-L).

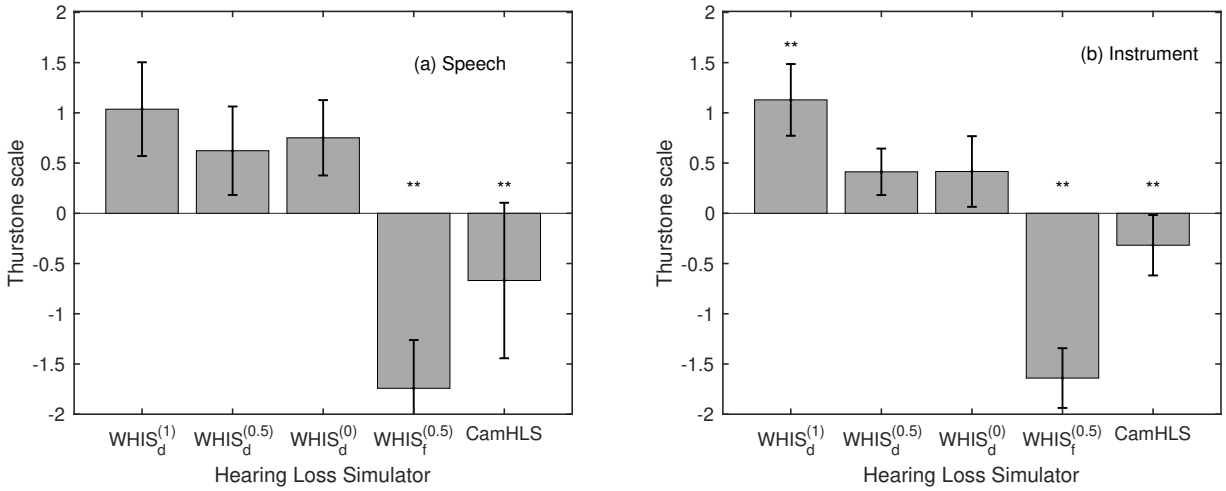


Figure 3: *Mean (bar) and 99% confidence interval (error bar). Double asterisks (**) indicate the HL conditions whose means were significantly different from* $\mathrm{WHIS}_\mathrm{d}^{(0.5)}$ *using Tukey's HSD tests at the 0.01 significance level.*

3.3. Result

For each listener and for each HL simulator condition, responses were aggregated for 10 words and for 4 instruments. Scores on the Thurstone scales were calculated from these. Figure 3 shows the mean and 99% confidence interval across listeners for the speech (a) and instrument (b) experiments. The perceived distortion is smaller when the score is higher. The trend across the HL simulator conditions was similar in both experiments. Mean scores were positive in the three $\mathrm{WHIS_d}$ conditions and negative in $\mathrm{WHIS_f}$ and CamHLS. The score for $\mathrm{WHIS}_\mathrm{f}^{(0.5)}$ is the most negative, meaning that the distortion was perceived the most.

Tukey's HSD tests were performed at a significance level of 0.01. Double asterisks (**) in Fig. 3(a)(b) indicate the HL conditions whose means were significantly different from $\mathrm{WHIS}_\mathrm{d}^{(0.5)}$. There were significant differences between $\mathrm{WHIS}_\mathrm{d}^{(0.5)}$ and $\mathrm{WHIS}_\mathrm{f}^{(0.5)}$ and between $\mathrm{WHIS}_\mathrm{d}^{(0.5)}$ and CamHLS in the both experiments. This was also the case for $\mathrm{WHIS}_\mathrm{d}^{(1)}$ and $\mathrm{WHIS}_\mathrm{d}^{(0)}$. In the speech experiment (Fig. 3(a)), there were no significant difference between $\mathrm{WHIS}_\mathrm{d}^{(1)}$, $\mathrm{WHIS}_\mathrm{d}^{(0.5)}$, and $\mathrm{WHIS}_\mathrm{d}^{(0)}$. This means that the distortion was not very noticeable even as the α value decreased and the nonlinear process was in operation. In contrast, in the instrument experiment (Fig. 3(b)), there were significant differences between $\mathrm{WHIS}_\mathrm{d}^{(1)}$ and $\mathrm{WHIS}_\mathrm{d}^{(0.5)}$, as indicated by **, and also between $\mathrm{WHIS}_\mathrm{d}^{(1)}$ and $\mathrm{WHIS}_\mathrm{d}^{(0)}$. The nonlinear distortion became noticeable when the value of α was less than 1, i.e. with the nonlinear process. However, these distortions were significantly less pronounced than those of $\mathrm{WHIS}_\mathrm{d}^{(0.5)}$ and CamHLS. In summary, WHISd was found to be generally superior to WHISf and CamHLS.

4. Effective algorithm for sound quality

Here we discuss the relationship between the algorithms of the HL simulators and the results of the sound quality assessment. Figure 1 in Section 2 show the algorithms of CamHLS (a), $\mathrm{WHIS_d}$ (b), and $\mathrm{WHIS_f}$ (c).

As described in Section 2.1, the degradation of sound quality in CamHLS is mainly caused by the frequency smearing process to simulate the band bandwidth widening in HI listeners. By removing this process, the quality can be improved but the effect of bandwidth widening cannot be simulated anymore.

$\mathrm{WHIS_d}$ and $\mathrm{WHIS_f}$ use the same analysis algorithm with GCFB and differ only in the synthesis algorithm, as described in Section 2.2. GCFB can also reflect the effect of bandwidth widening. Although the degree is generally smaller than in CamHLS, the difference is not expected to be as large for the 70-yr simulation. The sound quality of $\mathrm{WHIS_d}$ using DTVF was significantly better than that of $\mathrm{WHIS_f}$ using FBAS. The degradation factor in FBAS seems to be the misalignment of the phase lag before the overlap-and-add. The current implementation compensates for a fixed amount of lag for each filter, although the filter response depends on the α value. Introducing the α dependency can solve the problem, but makes the algorithm more complex. In contrast, the DTVF algorithm simply applies a single filter per frame, which is calculated from the amplitude frequency response of R_{total}, in Eq. 3, which does not require precise phase alignment. As a result, the DTVF algorithm is advantageous for the sound quality of the HL simulator.

5. Summary

We conducted perceptual experiments to evaluate the sound quality of the HL simulators and argued the relationship to the signal processing algorithms. We compared the widely used CamHLS and the recently proposed WHIS, which has two algorithms using DTVF ($\mathrm{WHIS_d}$) and FBAS ($\mathrm{WHIS_f}$). The perceived sound quality of $\mathrm{WHIS_d}$ was better than that of $\mathrm{WHIS_f}$ and CamHLS. $\mathrm{WHIS_d}$ seems to produce little perceptible distortion in speech sounds, even when the nonlinear process is working. This could be useful for evaluating the speech quality of hearing aids that have more nonlinear distortion. This advantage is mainly due to the use of the DTVF algorithm, which can be applied to various signal synthesis applications with filterbank analysis, not only to HL simulators.

6. Acknowledgements

This research was supported by JSPS KAKENHI: Grant Numbers JP21H03468, JP21K19794, and JP24K02961. The authors would like to thank Prof. Toshie Matsui for her help in describing the instrument sounds.

7. References

[1] B. C. J. Moore, *An introduction to the psychology of hearing*, 6th ed. Leiden, The Netherlands: Brill, 2013. [Online]. Available: https://brill.com/view/title/24210

[2] E. Villchur, "Simulation of the effect of recruitment on loudness relationships in speech," *J. Acoust. Soc. Am.*, vol. 56, no. 5, pp. 1601–1611, 1974.

[3] B. Moore and B. Glasberg, "Simulation of the effects of loudness recruitment and threshold elevation on the intelligibility of speech in quiet and in a background of speech," *J. Acoust. Soc. Am.*, vol. 94, no. 4, pp. 2050–2062, 1993.

[4] Y. Nejime and B. Moore, "Simulation of the effect of threshold elevation and loudness recruitment combined with reduced frequency selectivity on the intelligibility of speech in noise," *J. Acoust. Soc. Am.*, vol. 102, no. 1, pp. 603–615, 1997.

[5] T. Baer and B. Moore, "Effects of spectral smearing on the intelligibility of sentences in noise," *J. Acoust. Soc. Am.*, vol. 94, no. 3, pp. 1229–1241, 1993.

[6] ——, "Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech," *J. Acoust. Soc. Am.*, vol. 95, no. 4, pp. 2277–2280, 1994.

[7] B. Glasberg and B. Moore, "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," *J. Acoust. Soc. Am.*, vol. 79, no. 4, pp. 1020–1033, 1986.

[8] M. Stone and B. 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.

[9] P. Zurek and J. Desloge, "Hearing loss and prosthesis simulation in audiology," *The Hearing Journal*, vol. 60, no. 7, pp. 32–38, 2007.

[10] H. Hu, J. Sang, M. Lutman, and S. Bleeck, "Simulation of hearing loss using compressive gammachirp auditory filters," in *ICASSP 2011*. IEEE, 2011, pp. 5428–5431.

[11] N. Grimault, T. Irino, S. Dimachki, A. Corneyllie, R. Patterson, and S. Garcia, "A real time hearing loss simulator," *Acta Acustica united with Acustica*, vol. 104, no. 5, pp. 904–908, 2018.

[12] M. Cuevas-Rodríguez, D. González-Toledo, E. De La Rubia-Cuestas, C. Garre, L. Molina-Tanco, A. Reyes-Lecuona, D. Poirier-Quinot, and L. Picinali, "The 3d tune-in toolkit–3d audio spatialiser, hearing loss and hearing aid simulations," in *2018 IEEE 4th VR workshop on sonic interactions for virtual environments (SIVE)*. IEEE, 2018, pp. 1–3.

[13] A. Mourgela, J. Reiss, and T. Agus, "Investigation of a real-time hearing loss simulation for use in audio production," in *Proc. Audio Eng. Soc. Conv. 149*. Audio Engineering Society, 2020, pp. 50–57.

[14] T. Irino, T. Fukawatase, M. Sakaguchi, R. Nisimura, H. Kawahara, and R. Patterson, "Accurate estimation of compression in simultaneous masking enables the simulation of hearing impairment for normal-hearing listeners," in *Basic Aspects of Hearing*. Springer, 2013, pp. 73–80.

[15] M. Nagae, T. Irino, R. Nisimura, H. Kawahara, and R. Patterson, "Hearing impairment simulator based on compressive gammachirp filter," in *Proc. APSIPA ASC 2014*, 2014, pp. 1–4.

[16] T. Irino, "Hearing impairment simulator based on auditory excitation pattern playback: Whis," *IEEE access*, vol. 11, pp. 78 419–78 430, 2023.

[17] T. Matsui, T. Irino, M. Nagae, H. Kawahara, and R. Patterson, "The effect of peripheral compression on syllable perception measured with a hearing impairment simulator," in *Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing*. Springer, Cham, 2016, pp. 307–314.

[18] T. Irino, H. Tamaru, and A. Yamamoto, "Speech intelligibility of simulated hearing loss sounds and its prediction using the Gammachirp Envelope Similarity Index (GESI)," in *Proc. Interspeech 2022*, 2022, pp. 3929–3933.

[19] T. Deutsch and J. Fels, "Evaluating models for the simulation of cochlear hearing loss," in *DAGA 2023 Hamburg*, 2023, pp. 200–203.

[20] ISO 7029:2017, "Acoustics — statistical distribution of hearing thresholds related to age and gender," *ISO*, 2017. [Online]. Available: https://www.iso.org/standard/42916.html

[21] B. Moore, B. Glasberg, and T. Baer, "A model for the prediction of thresholds, loudness, and partial loudness," *J. Audio Eng. Soc.*, vol. 45, no. 4, pp. 224–240, 1997. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=10272

[22] T. Irino and M. Unoki, "An analysis/synthesis auditory filterbank based on an IIR implementation of the gammachirp," *Journal of the Acoustical Society of Japan (E)*, vol. 20, no. 6, pp. 397–406, 1999.

[23] ITU-R BS.1534-3, "Method for the subjective assessment of intermediate quality level of audio systems," *ITU*, pp. 1–34, 2015. [Online]. Available: https://www.itu.int/rec/R-REC-BS.1534-3-201510-I

[24] S. Sakamoto, T. Yoshikawa, S. Amano, Y. Suzuki, and T. Kondo, "New 20-word lists for word intelligibility test in japanese," in *Proc. 9th ICSLP*, 2006, pp. 2158–2161. [Online]. Available: https://www.isca-speech.org/archive_v0/archive_papers/interspeech_2006/i06_1517.pdf

[25] T. Kondo, S. Sakamoto, S. Amano, and Y. Suzuki, "Spoken word intelligibility tests under noisy conditions," in *Proc. 40th INTER-NOISE 2011*, 2011, pp. 1640–1646.

[26] M. Jeub, M. Schafer, and P. Vary, "A binaural room impulse response database for the evaluation of dereverberation algorithms," in *2009 16th International Conference on Digital Signal Processing*. IEEE, 2009, pp. 1–5.

[27] S. Kostka and B. Almn, *Tonal Harmony: With an Introduction to Post-tonal Music*, 9th ed. New York, USA: McGraw-Hill College, 2023.

[28] A. Yamamoto, T. Irino, K. Arai, S. Araki, A. Ogawa, K. Kinoshita, and T. Nakatani, "Comparison of remote experiments using crowdsourcing and laboratory experiments on speech intelligibility," in *Proc. Interspeech 2021*, 2021, pp. 181–185. [Online]. Available: https://www.doi.org/10.21437/Interspeech.2021-174