Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2016) 24: 4193 – 4207

doi:10.3906/elk-1411-135

Evaluation of hearing loss simulation using a speech intelligibility index

Umut ARIÖZ[1,*], **Banu GÜNEL**[2]
[1]Department of Computer Engineering, Faculty of Engineering, Gazi University, Ankara, Turkey
[2]Informatics Institute, Middle East Technical University, Ankara, Turkey

Received: 19.11.2014 • **Accepted/Published Online:** 28.07.2015 • **Final Version:** 20.06.2016

Abstract: Hearing loss simulation (HLS) is a useful tool for hearing studies, since finding large numbers of hearing-impaired volunteers with various levels of hearing loss is usually a problem. Correct simulation of the hearing impaired ear should take into account different suprathreshold effects, such as reduced frequency selectivity, reduced audibility, and loudness recruitment. Although these effects can be implemented using various algorithms, so far they have not been evaluated in detail taking into account subject dependency, effect of noisy environment, and phonetic characteristics of test words. In this study, for the purpose of evaluating the HLS, a speech intelligibility index (SII) was used as the objective measure and a modified rhyme test (MRT) was used as the subjective measure. Hearing loss of 12 subjects was simulated using various stimuli and 36 subjects with normal hearing were used as a control group. The factors that affect the performance of HLS algorithms were initially determined using the SII metric. Three other factors (sex of speaker, background noise level, and characteristics of the words) were tested both by the SII metric and the MRT. The lists contained word groups designed according to Turkish phonetic characteristics. In addition, the results of the MRT were compared with the results of the SII for both original and simulated sounds. According to the mean values and correlation analysis, both measures gave similar and reliable results for the HLS (58.40 and 57.37 for MRT mean values of unprocessed and simulated sounds, respectively; 0.23 for MRT mean values of unprocessed and processed sounds; 58% and 74% correlation coefficients for MRT and SII, respectively). When statistical significances of the measures were taken into account, the MRT gave more reliable results than the SII. While sounds were simulated in a satisfactory manner in both the noise-free environment and noisy environment for the MRT, similar results for the SII were obtained only in the noise-free environment. Because of high sensitivity of the SII to noise, the results were not satisfactory for the noisy cases. After the simulation, sex of the speaker and test list factors were found to be significant with the noise factor for the MRT.

Key words: Modified rhyme test, speech intelligibility index, hearing loss simulation, suprathreshold effects

1. Introduction

In order to understand the nature of hearing impairments and recommend hearing aid solutions, simulators can be used as important and useful research tools. Subjects with normal hearing listen to simulated sounds and respond to tests like hearing impaired subjects, so that the auditory changes can be observed. Successful hearing loss simulation (HLS) provides an opportunity for carrying out large-scale subjective testing for various levels of hearing loss as an alternative to recruiting hearing impaired subjects. As an added benefit, individuals with normal hearing can hear sounds as if they have a hearing loss, which might raise awareness of this disability.

---

[*] Correspondence: umut.arioz@gazi.edu.tr

However, the primary goal of simulation studies is to achieve similar intelligibility scores between hearing impaired subjects and subjects with normal hearing using simulated sounds.

The first studies on HLS date back to 1974 [1], when the first simulation was done for loudness recruitment by reducing the span between hearing thresholds and equal loudness contours. Nowadays, the research is still ongoing. In the work of Desloge et al. [2], a general overview of the HLS studies was presented. According to that study, there are two main approaches in the literature. In the first approach, the additive masking noise method [3–5] and multiband amplitude expansion methods [1,6–8] are used for simulating the effect of hearing loss. In the second approach, suprathreshold effects are reproduced using specific methods alone [9], or in combination with various methods [10,11]. In our study, the second approach was adopted and a combination of the effects has been used for implementing HLS.

Many researchers have attempted to specify the rationale behind hearing loss and to simulate it. The main reasons for difficulty in understanding speech and similar problems in the case of hearing loss are reduced audibility, reduced frequency selectivity, loudness recruitment, and dead regions. Therefore, by simulating these phenomena, an estimated response of real hearing can be achieved.

There are different methods for simulating different suprathreshold effects. Reduced audibility can be simulated by threshold elevation according to the subject's audiogram; reduced frequency selectivity can be simulated by spectral smearing, and loudness recruitment can be simulated by the loudness recruitment algorithm [12].

Reduced audibility is the fundamental effect of hearing loss that occurs at different levels for each frequency, which can be measured by audiometry and recorded in an audiogram. This effect can be simulated by decreasing the dB amount of the input signal according to the appropriate hearing threshold level.

Frequency selectivity is referred to as an ability to resolve the spectral components in the auditory system, which can be characterized as a bandpass filter with its structure. In subjects with normal hearing, auditory filters have constant bandwidths. However, in hearing impaired subjects, those filters are broader than normal [13]. Thus, selectivity of specific frequencies is reduced. To get an effect similar to that of a hearing impaired ear, smearing or broadening the spectral contents of the input signal can be used in both analog and digital signal processing methods [9,14–20].

Another common effect is loudness recruitment. If sound level is increased, the rate of growth of the loudness level occurs more rapidly in hearing impaired subjects than in subjects with normal hearing after a specific level. When the sound level reaches a sufficiently high level, such as 90–100 dB, loudness reaches its normal value as in subjects with normal hearing. For simulating loudness recruitment, there are some approaches that use applying dynamic range expansion to each of the frequency bands [1,4,7,8,10,14,21–24].

1.1. Evaluation of hearing loss simulation

There are both subjective and objective speech intelligibility measures for evaluating the effectiveness of HLS. The most widely used subjective speech intelligibility tests are the diagnostic rhyme test (DRT), modified rhyme test (MRT), and mean opinion score (MOS). On the other hand, generally, the articulation index (AI), speech transmission index (STI), and speech intelligibility index (SII) are used for objective speech intelligibility tests.

The first study about the rhyme test was done by Fairbanks [25], and with the inspiration of this study the MRT was designed by House et al. [26]. In this test a spoken word is played for the subject and the subject is asked to choose the spoken word among a written list of six words, one of which is the spoken word and the other five rhyme with it. Rhyming can be achieved either by selecting words so that they start with the same consonant or end with the same consonant of the spoken word. In a session, usually a total of 50 spoken words are presented, –25 word groups having the same first consonant and 25 word groups having the same last consonant. The MRT score of a subject is then reported as a percentage. There are MRTs in different languages constructed by considering the phonetic characteristics of that language (House et al. [26] for English, Li et al. [27] for Chinese, Tihelka et al. [28] for Czech).

The DRT, which was developed in 1977 [29], has a closed response method and is similar to the MRT, but it is based on word pairs only. The DRT has a simple training session for the listeners, while the MRT requires slightly more training.

The MOS is commonly used for both evaluating the effectiveness of speech coding algorithms and assessing the quality of synthesized speech [30,31]. MOS scores are generally given in three categories: intelligibility, naturalness, and pleasantness.

Because of the challenges of subjective speech intelligibility tests, such as the testing taking too long, high costs, difficulty of finding subjects, etc., alternative objective measurement methods have also been developed for evaluating speech intelligibility.

Studies on measuring objective speech intelligibility started in Bell Laboratories in 1940. In 1969, the AI was developed by the American National Standards Institute (ANSI) [32]. Then the STI was proposed in 1980 [33]. The SII was defined in the ANSI standards in 1997 [34]. Since then, it has become the most widely used objective method. The SII gives a metric between 0 and 1, which mean completely unintelligible and perfectly intelligible, respectively.

1.2. Properties of HLS

In our study, hearing loss was simulated by taking into account a combination of suprathreshold effects. Reduced frequency selectivity was simulated by spectral smearing, and both the reduced audibility and loudness recruitment were simulated by the loudness recruitment algorithm.

For subjective evaluation, Turkish was the chosen language. Therefore, Turkish phoneme sets and Turkish word pairs or word groups [35] were used. In the literature, there are some studies for developing Turkish intelligibility tests [35,36]. However, those studies were developed only for the DRT [37]. Our study is the first attempt for creating an MRT list for Turkish [38].

For objective evaluation, the SII was used. The SII was calculated for both original (i.e. unprocessed) sounds before HLS and processed sounds after HLS.

Differences from the previous studies and questions addressed by our study are:

- What is the accuracy of hearing loss simulation when comparing the results of subjects with normal hearing with those of hearing impaired subjects? In related studies, tests were done on subjects with normal hearing only, and they listened to the processed sounds. In our study, the results of one hearing impaired subject were compared with the results of three subjects with normal hearing with HLS.

- What is the reliability of HLS-based intelligibility measurement with combined effects measured by both MRT and SII? In our study, the evaluation of hearing loss simulation was carried out not only with a subjective measure, but also with an objective measure.

- Is there any relationship between the MRT and SII for different noise contents for both unprocessed sounds and simulated sounds?

- Which of the factors (sex of the speaker, noise, MRT list) are significant for the MRT and SII for both unprocessed and simulated sounds? In our study, sex (female, male), three different noise levels (no noise, 0 dB SNR, and –3 dB SNR), and three different MRT lists (list with the same first consonant, list with the same last consonant, and list grouped according to the Turkish phonetic characteristics) were used. Their effects on the MRT and SII were investigated for both unprocessed and simulated sounds.
- Which factors (sex of the speaker, noise, MRT list) are affecting the HLS performance according to the differences between unprocessed and simulated sounds? In our study, the difference occurred after the simulation was analyzed for all factors and subjects.

2. Methods

2.1. Subjects

Twelve (ten male and two female) hearing impaired subjects and 36 (25 male and 11 female) subjects with normal hearing participated in the study. Three subjects with normal hearing were used as a control group for each hearing impaired subject in the study. The average value of three subjects with normal hearing was compared with one hearing impaired subject.

Hearing impaired subjects were selected to provide different demographics and audiological properties. Audiogram values and properties of the hearing impaired subjects can be seen in Tables 1 and 2. Subjects were asked to remove their hearing aids during the tests. Generally, all subjects listened to words from 8 to 12 lists in the tests and it took 45–60 min per subject. For the study, approvals were obtained from the Ethics Committee of Middle East Technical University. Each participant signed an informed consent form before participating in the study.

2.2. Experimental design

In designing the MRT, only well-known and commonly used words were used. The MRT consisted of a number of subtests where each subtest measured the subject's ability to use acoustic information mainly along one signal dimension like nasality, sustention, sibilation, compactness, and graveness effects for better modeling of Turkish according to the phonetic characteristics [36].

For the MRT, two 25-word grouped lists and one 50-word grouped list were prepared. While the 50-word grouped list was designed according to the Turkish phonetic characteristics, two extra subtests with 25 words were also prepared that started with the same consonant (for example: yat-yaz-yay-yan-yar-yas) and ended with the same consonant (example: yay-vay-fay-hay-tay-çay). To prevent memorization of words, the order of words both in the list and within the word groups was randomly mixed. As a result, there were six versions of each 25-word grouped list and four versions for 50-word grouped list.

2.3. Stimuli

Words in the MRT were spoken by a male and a female native Turkish speaker and recorded in an acoustically treated studio at the Department of Electrical and Electronics Engineering of Middle East Technical University. Since a carrier sentence is required by the MRT, for each word group, the speakers repeated the sentence "Aşağıdakilerden <Word> kelimesini seçer misiniz?", which can be translated into English as "Could you choose the word <Word>?"

The sounds were recorded at 48,000 Hz sampling frequency with 16-bit resolution using a Sennheiser M64 prepolarized condenser microphone and EDIROL UA-1000 Audio Capture device. Speakers were told to

Table 1. Audiometric measurements of the hearing impaired subjects.

	250	500	1000	2000	4000	6000	8000	Ear tested
Subject 1 L	55	55	55	50	55	70	85	
Subject 1 R	45	50	55	60	65	85	85	■
Subject 2 L	10	10	10	35	60	70	80	■
Subject 2 R	10	15	15	30	55	65	80	
Subject 3 L	30	30	25	55	80	87	95	■
Subject 3 R	40	40	35	55	70	75	80	
Subject 4 L	15	10	10	10	30	45	55	■
Subject 4 R	15	10	10	10	30	45	55	
Subject 5 L	45	70	75	85	90	95	100	■
Subject 5 R	25	25	20	40	70	75	85	
Subject 6 L	25	20	20	30	55	57	60	
Subject 6 R	70	80	90	100	100	100	100	■
Subject 7 L	40	50	70	75	70	65	70	■
Subject 7 R	45	55	70	75	65	75	80	
Subject 8 L	25	40	50	50	80	85	100	■
Subject 8 R	20	30	35	50	80	85	100	
Subject 9 L	30	45	55	50	50	40	50	
Subject 9 R	30	45	60	45	45	35	55	■
Subject 10 L	20	30	30	45	55	50	65	
Subject 10 R	55	55	50	55	50	70	60	■
Subject 11 L	40	40	40	45	45	50	50	■
Subject 11 R	40	40	45	45	50	50	50	■
Subject 12 L	20	20	30	85	110	110	110	■
Subject 12 R	20	15	20	70	110	110	110	

maintain a constant level of speech throughout the recording. Noisy sounds were constructed using a restaurant noise for all sounds.

All subjects listened to different subsets of the testing material to reduce testing time. Selected tests were arranged to get the same number of listened cases. All cases were listened to seven times by different subjects, except for one case (6 times in that case), in order to obtain meaningful comparisons and statistics. There were three different factors: sex of the speaker (male/female), SNR (no noise, 0 dB, –3 dB), and word lists (same first consonant, same last consonant, and list according to Turkish phonetics). According to these combinations, 18 different test cases were constructed for the MRT.

Hearing impaired subjects listened to the stimuli using a headphone in a noise-free environment. The stimuli was only provided to the selected ear of each hearing impaired subject. The stimuli were only provided to the right ear of each subject with normal hearing. The speech level was adjusted to give a maximum of 80 dB at the ear location for all tests.

All subjects were trained until they got used to the MRT procedures. The MRT was started with a no-noise case for all subjects. At the beginning of the experiment, each participant received written instructions, and their questions were answered by the experimenter if they had any.

Table 2. Significance and correlation analysis for each subject (subjects' mean values show the mean of all tests; correlation is significant at the 0.05 level for Pearson coefficients and 0.01 level for Spearman's rho; bold values are statistically significant).

	MRT results				SII			
	Unprocessed case (mean value)	Simulated case (mean value)	Paired samples test (P-value)	Correlation (Pearson coefficients)	Unprocessed case (mean value)	Simulated case (mean value)	Wilcoxon Signed ranks test (P-value)	Correlation (Spearman's rho)
Subject 1	73.50	48.92	**0**	0.63	0.24	0.31	**0.02**	0.77
Subject 2	59.50	55.89	0.23	0.85	0.29	0.29	0.5	0.74
Subject 3	45.40	53.97	0.1	0.49	0.16	0.29	**0.005**	0.52
Subject 4	64.00	67.33	0.47	0.55	0.25	0.14	**0.008**	0.83
Subject 5	36.17	51.22	**0**	0.73	0.2	0.23	0.29	0.7
Subject 6	49.17	56.61	0.08	0.71	0.1	0.24	0.15	0.58
Subject 7	59.80	54.27	0.22	0.5	0.24	0.31	**0.02**	0.73
Subject 8	66.40	62.47	0.23	0.67	0.32	0.2	**0.04**	0.78
Subject 9	75.00	58.07	**0.001**	0.39	0.19	0.12	**0.01**	0.69
Subject 10	54.80	61.40	0.15	0.55	0.22	0.2	0.12	0.79
Subject 11	69.33	62.22	0.12	0.5	0.3	0.19	**0.01**	0.94
Subject 12	50.50	60.75	0.17	0.42	0.22	0.16	**0.01**	0.86
Grand mean	**58.40**	**57.37**		**0.58**	**0.23**	**0.23**		**0.74**

2.4. Hearing loss simulation implementation

For determining the exact values for real testing, an offline simulation study was done with sample sound files. For the offline simulation study, sex (male, female), standard audiograms (normal, mild, moderate, moderate to severe, severe, profound), factors for loudness recruitment, smearing factor (3, 4, 5, 6), hamming window size of spectral smearing (128, 256, 512), and different combinations of suprathreshold effects (threshold elevation, loudness recruitment, spectral smearing) were considered as different parameters. These parameters were investigated by the SII. ANOVA testing was applied to the results to determine significant factors that affect intelligibility. As a result, sex, audiogram type, and smearing factor were found significant.

2.4.1. Spectral smearing

A schematic diagram of the steps of the spectral smearing process, which were identical to our algorithm's steps, can be found in the study done by Baer et al. [9]. The smearing function is based on human auditory filters and requires as input only the smearing factor that determines the width of these filters. Auditory filters are characterized by an intensity weighting function that determines the filter shape [39]. That function includes the sharpness of the filter and the deviation amount from the filter's center frequency. By changing the sharpness of the filter from both sides, the effect of smearing can be changed. For subjects with normal hearing, sharpness values on both sides are approximately equal. For hearing impaired subjects, the only parameter that affects the smearing is the smearing factor.

Different combinations of smearing factors (0, 3, 6) for both sides were tested to get the effect of the smearing on speech intelligibility. In our study, for evaluating the effect of smearing, the smearing factor and the Hamming window size were selected as 3 and 512, respectively, based on the results of an offline simulation study. Normal hearing and hearing impaired (widened) auditory filters were calculated with intensity weighting function and equivalent rectangular bandwidth (ERB). In the widened auditory filter, the sharpness parameter was determined by the ratio of normal hearing sharpness parameter to the smearing factor. Then the smearing function matrix was obtained by dividing the normal auditory filters by the widened auditory filters.

To get smeared components of the power spectrum, fast Fourier transform (FFT) with Hamming window (size: 512) was used. For the whole process, an overlap and add method was implemented.

2.4.2. Loudness recruitment and threshold elevation

Suprathreshold effects of reduced audibility and loudness recruitment were simulated together with a single processing block [7]. While simulating the loudness recruitment, the constant values for each of the spectrum bands were calculated according to the threshold elevations of each subject. Thus, there was no need for an extra step to simulate the reduced audibility.

There are six main steps for simulating the loudness recruitment [7]. In the first step, the input signal is filtered according to 13 center frequencies (100, 190, 306, 452, 640, 879, 1184, 1579, 2067, 2698, 3503, 4529, and 5837 Hz). An auditory filter, which is composed of four first-order gammatone filters, is applied to achieve similarity with moderate to severe cochlear hearing loss. In the second step, time alignment is applied to all outputs of the filter to get compatible peaks. Then Hilbert transform of the input signal is extracted from the input signal to obtain the analytical signal. In the third step, the input signal is decomposed into an envelope and a fine structure using the input signal and the Hilbert transform of the input signal. In the main processing step, the fourth step, the outputs of the filter are normalized so that the unity peak corresponds to 100 dB SPL (complete loudness recruitment level). To achieve the final output for each channel, the processed envelope is multiplied by its fine structure, which is the fifth step. The final sixth step is combining all the channels to obtain the output sound.

3. Results

3.1. Analysis of measure reliability

3.1.1. Comparison of measures' distributions

In this analysis, a 95% confidence interval of the difference was applied to all subjects for MRT and SII results for unprocessed (presimulation) and simulated (postsimulation) sounds. While the results of the MRT showed normal distribution, the SII did not show such a distribution. Thus, for the purpose of comparing the significance of the mean distributions of the MRT, the paired samples test and the Pearson coefficients for correlation were used. The paired samples test gives the differences between the values of the two variables and tests whether the average is different from zero. To calculate the Pearson coefficients, variables should be normally distributed. Because SII results did not have a normal distribution, the Wilcoxon signed rank test and Spearman's rho were calculated for the SII. The Wilcoxon signed rank test can give information about the differences between the pairs, and Spearman's rho is used for abnormally distributed variables in the correlation analysis. The results are shown in Table 2.

3.1.2. Interaction analysis between MRT and SII according to noise

Four different analyses were carried out to investigate the relationships between MRT and SII results for unprocessed and simulated sounds (pre_MRT/pre_SII, post_MRT/post_SII, pre_MRT/post_MRT, pre_SII/post_SII). Analyses were done according to noise amounts, because noise was found to be a significant factor for both measures. Figures 1–4 show the scatter plots of MRT and SII results.

3.2. Analysis of HLS reliability

Univariate analysis of HLS was carried out. Mean values for each factor and significant factors of the tests are shown in Table 3.

4. Discussion

4.1. Distribution measures

For MRT results, according to the P-values of the paired samples test, there is no statistical significance between unprocessed and simulated results for the subjects, except for three subjects (subject 1, subject 5, and subject 9). Subject 1 performed much better in the MRT test than was expected considering his high level of hearing loss. This unexpected result can possibly be explained by a coping mechanism developed by this subject as a result of the long duration of hearing loss, taking into account the age of this subject (30 years of age with 20 years of hearing loss history). For subject 5, education level could have resulted in lower MRT scores than expected. Subject 9 is a trained musician, which could have contributed to higher than expected MRT scores.

According to the Pearson coefficients, on average, there is a 58% correlation between unprocessed and simulated values. The smallest Pearson coefficient is obtained for subject 9. These results show the reliability of the MRT for our simulation program.

For the SII, an opposite situation to that of the MRT was observed. Excluding three subjects (subject 2, subject 5, subject 10), the P-values of the paired samples test gave statistically significant values, which shows that there is a difference between unprocessed and simulated for the SII. However, in correlation results, excluding subject 3 and subject 6, all subjects had correlation values greater than 69% and the mean value for all subjects was 74%. This situation shows that the SII is not a reliable measure.

Table 3. Mean values and significant factors for the MRT and SII for both unprocessed and simulated (asterisk specifies statistically significant factors).

	Pre_MRT	Pre_SII	Post_MRT	Post_SII
General values				
Grand mean	58.40	0.23	57.37	0.23
Std. error	1.50	0.01	0.86	0.01
Significant factors				
Sex			*	*
Test			*	
Noise	*	*	*	*
Noise × test	*			
Parameters' mean values				
Sex - female	59.02	0.21	61.89	0.25
Sex - male	57.78	0.24	52.86	0.21
No noise	66.38	0.46	64.91	0.48
0 dB SNR	57.95	0.15	57.31	0.13
–3 dB SNR	50.86	0.07	49.90	0.08
Test list - FC	57.52	0.20	60.00	0.24
Test list - LC	54.76	0.25	50.52	0.22
Test list - 50	62.91	0.23	61.60	0.23

According to the paired samples test applied to the MRT results, subjects generally showed the same performance. On the other hand, the SII gave statistically significant results, except for four subjects. For distribution comparisons, mean values of all no-noise and noisy cases were used. However, SII results were very low for the noisy cases since the SII is very sensitive to noise and this might have affected the Wilcoxon signed rank test. For these analyses, MRT results gave a more reliable evaluation according to each subject than the SII in our study.

4.2. Interaction analysis

Distinct regions for each noise case can be seen for the pre_MRT vs. pre_SII scatter plot in Figure 1. For the no-noise case, SII and MRT values increase and decrease together. However, for noisy cases, denser groups of SII values occur and no relation can be observed between these two metrics. According to these results, MRT and SII metrics are more consistent for the no-noise case. Because of the high sensitivity of the SII to noise, the SII and MRT give different results for noisy cases. However, in general, both the SII and the MRT scores decrease with noise, as expected.

Distinct regions between noisy cases and no-noise cases can be seen for the post_MRT vs. post_SII scatter plot in Figure 2. For no-noise cases, SII and MRT values increase and decrease together. However, for noisy cases, denser groups of SII values occur and no relation can be observed between these two metrics. However, for noisy cases, more dense groups occur. Also, the size of the groups for the MRT is larger than for the SII for the noisy cases. After the simulation, the compactness of the SII decreases. This shows the effect of the simulation on the noise sound according to the SII calculations. After the simulation, SII values become similar and the difference between 0 dB SNR and –3 dB SNR cases diminishes. Also, the SII for 0 dB SNR and the

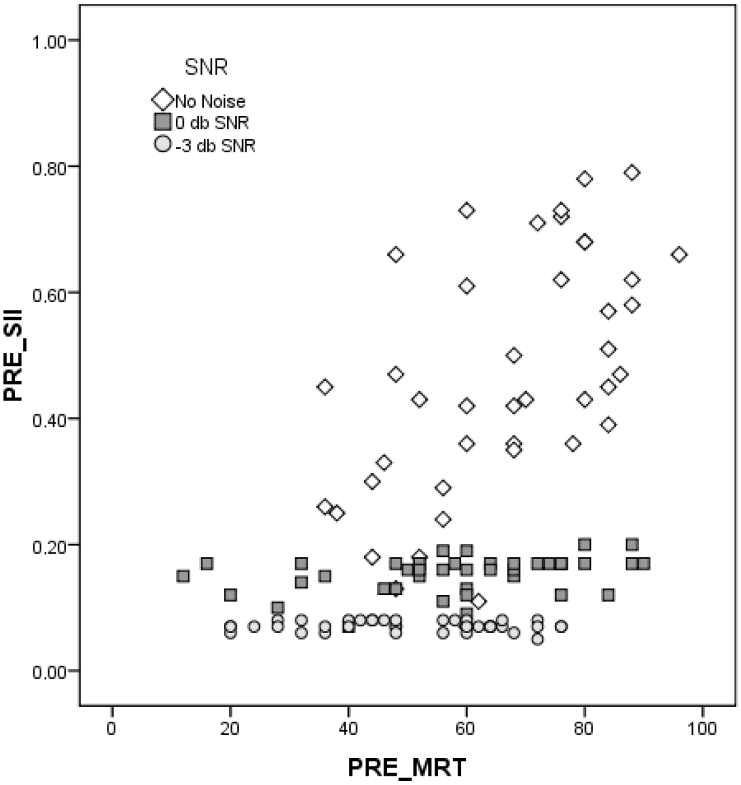

Figure 1. Scatter plot of pre_MRT and pre_SII.

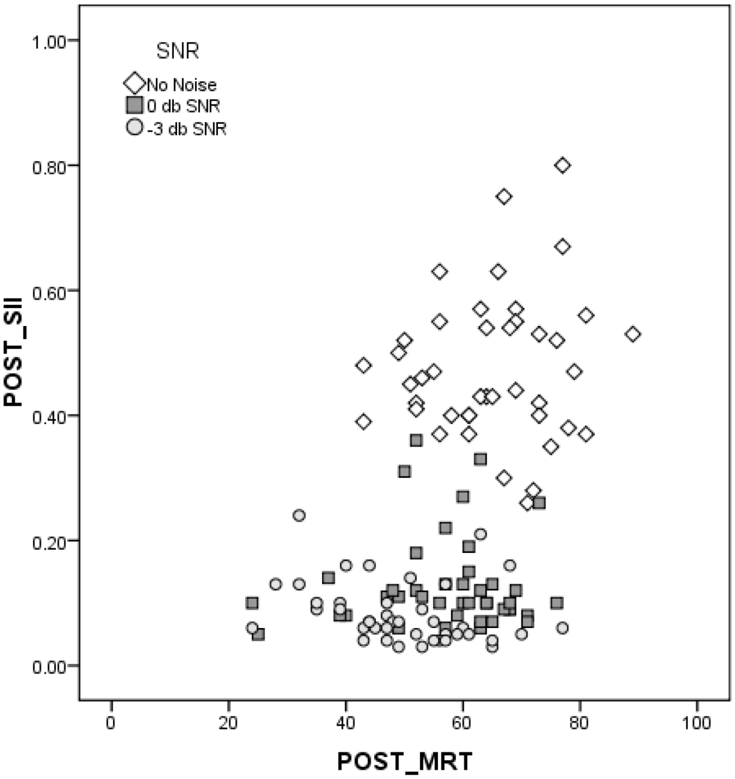

Figure 2. Scatter plot of post_MRT and post_SII.

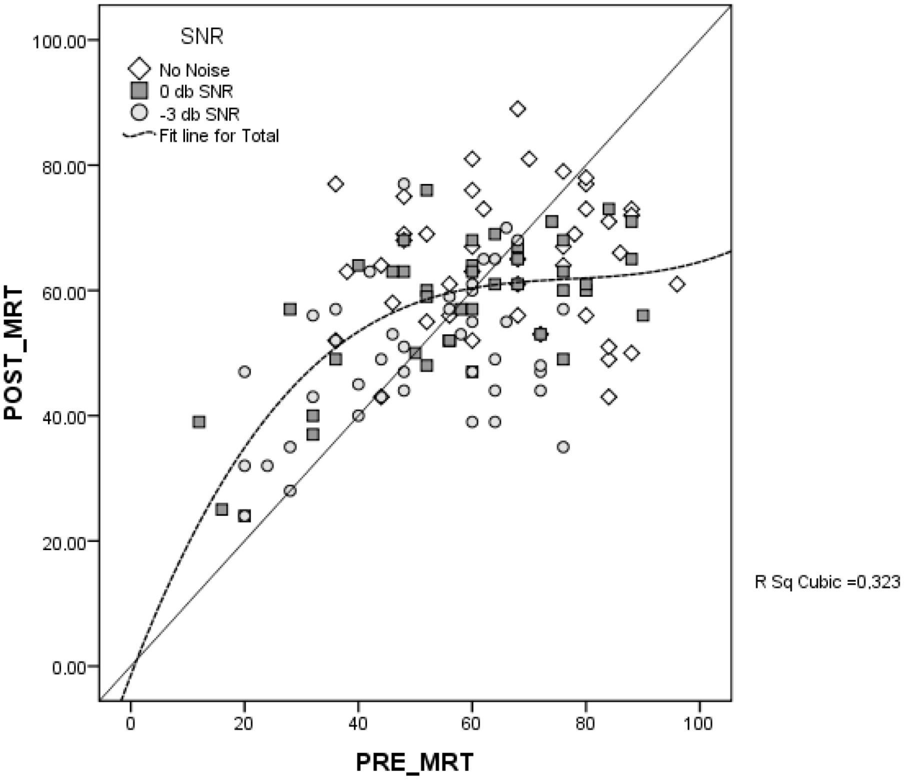

Figure 3. Scatter plot and curve fitting plot (R2: 0.323) of pre_MRT and post_MRT.

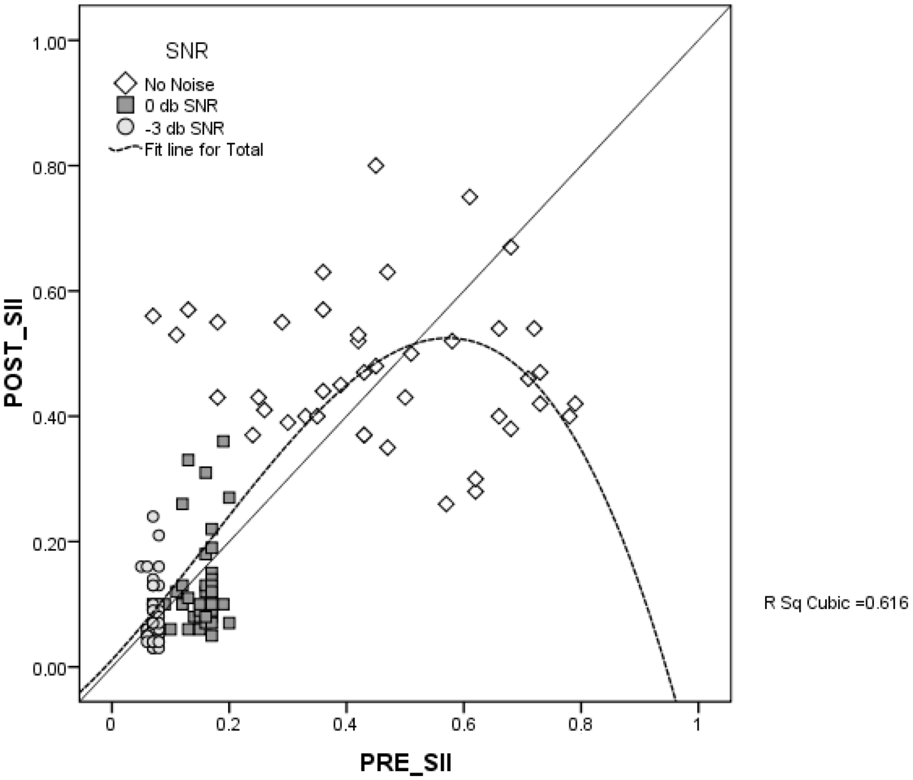

Figure 4. Scatter plot and curve fitting plot (R2: 0. 616) of SII for pre_SII and post_SII.

SII for –3 dB SNR cases are mixed with each other. However, the expected decreasing behavior with noise is also present here.

According to these findings, it can be said that the simulation gives reliable and consistent results between the MRT and the SII values, especially for the no-noise cases.

A pre_MRT vs. post_MRT values scatter plot can be seen in Figure 3. As expected, when the noise increases, the correct percentages decrease for both pre_MRT and post_MRT results. The difference between no-noise and 0 dB SNR cases is lower than the difference between 0 dB SNR and –3 dB SNR cases. Generally, the results are on a line and display adjacent groups. This effect can be seen easily from the line drawn by curve fitting.

Distinct regions for each noise case can be seen for the pre_SII vs. post_SII values scatter plot in Figure 4. Generally, the results are on a line and display adjacent groups. This effect can be seen easily from the line drawn by curve fitting. For no-noise cases, there is a similar distribution between pre_SII and post_SII. However, for noisy cases, dense groups occur. Also, the size of the groups for post_SII is larger than for pre_SII for noisy cases. These dense groups confirm that the SII is very sensitive to noise and presents nonlinearity for noisy cases. While the difference of the results of no-noise cases and noisy cases is clear and easily seen, noisy cases are concentrated in a small range. As expected, when noise increases, values for both pre_SII and post_SII decrease.

The effect of the noise, especially on the SII, for both unprocessed and simulated results can be seen in Figures 1 and 2. The consistency of the results of no-noise cases of HLS can also be seen in these figures. For noisy cases, although the SII gives results in a narrow range, the MRT gives results in a wider range. However, especially for the simulated cases, the difference and spread of the SII range diminishes. This shows the averaging effect of HLS on the noisy sounds. Thus, from Figures 1 and 2, one can say that speech intelligibility is not very reliable, especially in noisy environments.

The results of the MRT are generally positioned on or adjacent to the 45° line. This is especially observed for the noisy cases, which is important for showing the effect of HLS in a noisy environment. Also, for the no-noise cases, pre_MRT values are higher than post_MRT values, which means that hearing impaired subjects perform much better than subjects with normal hearing with simulated hearing loss. This is expected, because hearing impaired people develop coping mechanisms over time to compensate for their hearing loss. However, the simulated sounds were new and strange for subjects with normal hearing. These results show that according to the MRT scores, HLS is reliable.

The inconsistency and unreliability of the SII with HLS can be seen from Figure 4. The expected effect of the noise is only present for unprocessed and simulated tests and the effect of the noise diminishes, yet without the averaging effect as with the MRT. Results positioned on or adjacent to the 45° line do not have any meaningful states because of the high sensitivity of the SII to noise.

4.3. Univariate analysis

The grand mean values for pre_MRT and post_MRT are very close to each other (58.4 and 57.37, respectively). While effective factors for pre_MRT are noise and the combined effect of the noise and test list, all factors are effective for post_MRT.

The mean values for the noise cases decreased in parallel with the noise amount and there was an approximately equal decrement between the noise cases for both measures. Thus, noise is statistically significant for all measures for both unprocessed and simulated cases.

Grand mean values for pre_SII and post_SII are the same (0.23). Noise is the only effective factor for pre_SII. After the simulation, the sex factor became effective besides the noise.

The performance of HLS was also shown by both MRT and SII via univariate analysis. HLS gave consistent grand means for the MRT and SII. In the literature, noise is identified as the main effective factor. In our study, HLS reflected reliable results according to the noise content. There are similar decrements among the results of noisy cases for the MRT, showing that the response reliability is preserved under noise. Also, HLS gave more distinguishable results for sex and test list according to the MRT and gave more distinguishable results for sex according to the SII when the significant factors were taken into account.

General decrease according to noise amount was not observed for the 25-word grouped list with the same first consonant for both sexes for the unprocessed case. This may show that noise is not effective for discriminating Turkish words starting with the same consonant.

While results of the MRT of the unprocessed case are higher for the female voice than the male voice, the SII values of unprocessed sounds are higher for the male voice than the female voice.

In general, results obtained for female speech were higher than for male speech for all processed sounds. This shows that the female voice used in this study was more intelligible than the male voice.

5. Conclusion

In our study, hearing loss with combined suprathreshold effects was simulated. In order to implement the suprathreshold effects, spectral smearing was used for simulating the reduced frequency selectivity, and a loudness recruitment simulation algorithm was used for simulating the reduced audibility and loudness recruitment together. Twelve hearing impaired subjects for unprocessed sounds and 36 subjects with normal hearing for simulated sounds were involved in the testing. For the evaluation of the HLS, both subjective and objective speech intelligibility tests, namely MRT and SII, respectively, were used. An MRT word list was developed in Turkish for the first time in literature. For the MRT, three different word lists were used for getting detailed information. These are a six-word list with the same first consonant, a six-word list with the same last consonant, and a six-word list redesigned according to the phonetics of the Turkish language.

In our study, MRT results showed more reliable behavior than the SII. Although both measures gave correlated results, only the MRT gave similar results for presimulation (i.e. for testing with hearing impaired subjects) and postsimulation (i.e. for testing with subjects with normal hearing with HLS).

Subject selection is an important factor when evaluating HLS studies. In our study, we tried to find subjects with different hearing loss levels. Also, the demographic and educational properties of the subjects may affect the results, like in our study. That effect can be easily seen from the difference of the presimulation and postsimulation results.

The main factor for the SII is the noise content. With decreasing SNR, the SII declines rapidly. Thus, the SII may be reliable and can be compared with the results of the MRT only for no-noise cases.

While in the presimulation cases only noise was a significant factor, after the simulation, all factors (sex, test list, and noise) became significant. In the grand means of both the MRT and SII, similar results were obtained. This shows the reliability and usability of HLS.

For further investigations, comparisons can be made with more subjects and other intelligibility measures can be used for the validation of the HLS. Also, in the future, the MRT for the Turkish language can be studied together with HLS to determine the effect of Turkish phonetics on speech intelligibility for hearing impaired people.

Acknowledgments

We thank Dr Özlem Konukseven and Aynur Akdoğan for providing hearing impaired subjects, supplying special test rooms, and making useful comments. Brian CJ Moore provided helpful comments on the algorithms. This study was a part of Umut Arıöz's doctorate thesis, which was completed at the Institute of Informatics, Middle East Technical University, Ankara.

References

[1] Villchur E. Simulation of the effect of recruitment on loudness relationships in speech. J Acoust Soc Am 1974; 56: 1601-1611.

[2] Desloge JG, Reed CM, Braida LD, Perez ZD, Delhorne LA. Temporal modulation transfer functions for listeners with real and simulated hearing loss. J Acoust Soc Am 2011; 129: 3884-3996.

[3] Milner P, Braida LD, Durlach NI, Levitt H. Perception of filtered speech by hearing-impaired listeners. In: Elkins E, editor. Speech Recognition by the Hearing Impaired. Rockville, MD, USA: American Speech Language Hearing Association, 1984. pp. 30-41.

[4] Zurek PM, Delhorne LA. Consonant reception in noise by listeners with mild and moderate sensorineural hearing impairment. J Acoust Soc Am 1987; 82: 1548-1559.

[5] Florentine M, Reed CM, Rabinowitz WM, Braida LD, Durlach NI. Intensity perception. XIV. Intensity discrimination in listeners with sensorineural hearing loss. J Acoust Soc Am 1993; 94: 2575-2586.

[6] Villchur E. Signal processing to improve speech intelligibility in perceptive deafness. J Acoust Soc Am 1973; 53: 1646-1657.

[7] Moore BCJ, Glasberg BR. 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 1993; 94: 2050-2062.

[8] Duchnowski P, Zurek PM. Villchur revisited: another look at AGC simulation of recruiting hearing loss. J Acoust Soc Am 1995; 98: 3170-3181.

[9] Baer T, Moore BCJ. Effects of spectral smearing on the intelligibility of sentences in the presence of noise. J Acoust Soc Am 1993; 94: 1229-1241.

[10] Moore BCJ, Glasberg BR, Vickers DA. Simulation of the effects of loudness recruitment on the intelligibility of speech in noise. Brit J Audiol 1995; 29: 131-143.

[11] Nejime Y, Moore BCJ. 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 1997; 102: 603-615.

[12] Moore BCJ. Speech processing for the hearing-impaired: successes, failures, and implications for speech mechanisms. Speech Commun 2003; 41: 81-91.

[13] Glasberg BR, Moore BCJ. Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments. J Acoust Soc Am 1986; 79: 1020-1033.

[14] Villchur E. Electronic models to simulate the effect of sensory distortions on speech perception by the deaf. J Acoust Soc Am 1977; 62: 665-674.

[15] Summers IR, Al-Dabbagh AD. Simulated loss of frequency selectivity and its effects on speech perception. Acoust Lett 1982; 5: 129-132.

[16] Summers IR. Electronically simulated hearing loss and the perception of degraded speech. In: Wise DL, editor. Bioinstrumentation and Biosensors. New York, NY, USA: CRC Press, 1991. pp. 589-610.

[17] Celmer RD, Bienvenue GR. Critical bands in the perception of speech signals by normal and sensorineural hearing loss listeners. In: Schouten MEH, editor. The Psychophysics Speech Perception. Dordrecht, the Netherlands; Nijhoff, 1987. pp. 473-480.

[18] Howard-Jones PA, Summers IR. Temporal features in spectrally degraded speech. Acoust Lett 1992; 15: 159-163.

[19] ter Keurs M, Festen JM, Plomp R. Effect of spectral envelope smearing on speech reception. J Acoust Soc Am 1992; 91: 2872-2880.

[20] Moore BCJ, Glasberg BR, Simpson A. Evaluation of a method of simulating reduced frequency selectivity. J Acoust Soc Am 1992; 91: 3402-3423.

[21] Dubno JR, Schaefer AB. Comparison of frequency selectivity and consonant recognition among hearing-impaired and masked normal-hearing listeners. J Acoust Soc Am 1992; 91: 2110-2121.

[22] Phillips DP. Stimulus intensity and loudness recruitment: neural correlates. J Acoust Soc Am 1987; 82: 1-12.

[23] Moore BCJ, Glasberg BR. A model of loudness perception applied to cochlear hearing loss. Audit Neurosci 1997; 3: 289-311.

[24] Fairbanks G. Test of phonemic difference: the rhyme test. J Acoust Soc Am 1958; 30: 596-600.

[25] House AS, Williams CE, Hecker MHL, Kryter KD. Articulation testing methods: consonantal differentiation with a closed response set. J Acoust Soc Am 1965; 37: 158-166.

[26] Li Z, Tan EC, McLoughlin I, Teo TT. Proposal of standards for intelligibility tests of Chinese speech. IEE P-Vis Image Sign 2000; 147: 254-260.

[27] Tihelka D, Matoušek J. The design of Czech language formal listening tests for the evaluation of TTS systems. In: Fourth International Conference on Language Resources and Evaluation Conference, 26–28 May 2004; Lisbon, Portugal. pp. 1099-2002.

[28] Voiers WD. Diagnostic evaluation of speech intelligibility. In: Hawley ME, editor. Speech Intelligibility and Speaker Recognition. Stroudsburg, PA, USA: Hutchinson and Ross Inc., 1977.

[29] ITU. ITU-T Recommendation P.85. Telephone Transmission Quality Subjective Opinion Tests. A Method for Subjective Performance Assessment of the Quality of Speech Voice Output Devices. Geneva, Switzerland: ITU, 1994.

[30] Salza PL, Foti E, Nebbia L, Oreglia M. MOS and pair comparison combined methods for quality evaluation of text to speech systems. Acta Acust 1996; 82: 650-656.

[31] American National Standards Institute. ANSI S3.5. American National Standard Methods for the Calculation of the Articulation Index. New York, NY, USA: American National Standards Institute, 1969.

[32] Steeneken H, Houtgast T. A physical method for measuring speech transmission quality. J Acoust Soc Am 1980; 67: 318-326.

[33] American National Standards Institute. ANSI S3.5. American National Standard Methods for the Calculation of the Speech Intelligibility Index. New York, NY, USA: American National Standards Institute, 1997.

[34] Sak H, Gungor T, Safkan Y. A corpus-based concatenative speech synthesis system for Turkish. Turk J Electr Eng Co 2006; 14: 209-223.

[35] Palaz H, Bicil Y, Kanak A, Dogan MU. New Turkish intelligibility test for assessing speech communication systems. Speech Commun 2005; 47: 411-423.

[36] American National Standards Institute. ANSI S3.2. Method for Measuring the Intelligibility of Speech over Communication Systems. New York, NY, USA, American National Standards Institute, 1989.

[37] Arıöz U, Günel B. Modified rhyme test for evaluating Turkish speech intelligibility. Gazi Medical Journal 2015; 26: 75-81.

[38] Patterson RD, Nimmo-Smith I, Weber DL, Milroy R. The deterioration of hearing with age: frequency selectivity, the critical ratio, the audiogram, and speech threshold. J Acoust Soc Am 1982; 72: 1788-1803.