Damascus University Journal For Engineering Sciences
Vol 40 No.2 (2024):225-234

مجلة جامعة دمشق للعلوم الهندسية
المجلد 40– العدد 2 (2024)225–234

Non-reference estimation of intelligibility of noisy speech signals

Dr. Eng. Hassan Ahmad

Assistant professor, Department of communications, Informatics Eng. Faculty, Syrian Private University.

Received: 6/8/2022
Accepted: 20/11/2022

Abstract

A non-reference estimation method of the values of the intelligibility index of noisy speech signals is proposed. It is based on the intelligibility measure SNR (signal-to-noise rate) loss and Scalart speech noise reduction method. The reliability of the obtained non-reference estimates is compared by comparison with the values obtained using the reference method.

The accuracy of the proposed non-reference measure of intelligibility is investigated in the case of the effect of various types of noise on speech signals. The obtained results attest to the relatively high accuracy of the proposed method of non-reference estimation (the average relative error is 1.05–3.55%).

Keywords: speech signal, intelligibility, non-reference estimation, noise reduction, linear regression.

التقييم غير المعياري لوضوح الإشارات الكلامية المضججة

د.م. حسان احمد

أستاذ مساعد، قسم الاتصالات، كلية الهندسة المعلوماتية، الجامعة السورية الخاصة.

الملخص

في هذا العمل تم اقتراح طريقة التقييم غير المعياري (غير مرجعي) لقيم مؤشر وضوح إشارات الكلام المضججة. تعتمد الطريقة على قياس فقدان معدل الإشارة إلى الضجيج (SNR loss) أثناء تقييم وضوح الإشارة الكلامية وعلى طريقة سكالرت (Scalart) لتقليل الضجيج في الكلام. تمت مقارنة موثوقية التقديرات غير المعيارية التي تم الحصول عليها مع القيم التي تم الحصول عليها باستخدام الطريقة المعيارية.

يتم التحقق من دقة مقياس الوضوح غير المعياري المقترح في حالة تأثير أنواع مختلفة من الضجيج على إشارات الكلام. بيّنت النتائج التي تم التوصل إليها دقة عالية نسبياً للطريقة المقترحة للتقدير غير المعياري (متوسط الخطأ النسبي 1.05 – 3.55٪).

تاريخ الإيداع: 2022/**8/6**

تاريخ القبول: 2022/11/20

الكلمات المفتاحية: الإشارة الكلامية، الوضوح، التقدير غير المعياري، تخفيض الضجيج، الارتباط الخطي.

1. Introduction

One of the most important components of the acoustic expertise of speech communication channels is the measurement of speech intelligibility [1, 25]. Intelligibility refers to the degree to which certain speech units (phonemes, syllables, words, phrases) can be correctly perceived by the listener. Various distorting factors (noise, reverberation, non-linear distortion) can lead to a significant reduction in intelligibility and loss of some information. Along with intelligibility, the quality of the speech signal also decreases [2], but these concepts should be distinguished.

The most general approach when classifying methods for estimation of intelligibility is to divide them into subjective and objective. A distinctive feature of subjective methods is the direct participation of speakers and auditors (listeners) in the estimation process. These methods are highly reliable but resource-intensive and unsuitable for the real-time expertise of communication channels. Therefore, approximately since the mid-1950s, objective methods for measuring intelligibility have been actively developed.

Currently, the two largest groups of these methods are distinguished: formant and modulation. The first group includes methods: AI (Articulation Index) and SII (Speech Intelligibility Index), their predecessors and modifications [1, 3-8], as well as Russian methods proposed by Yu .S. Bykov, N. B. Pokrovsky, M. A. Sapozhkov [1, 9-12]. Modulation methods are represented by various modifications of the STI (Speech Transmission Index) method, including the RASTI (Room Acoustic Speech Transmission Index), STIPA (Speech Transmission Index public address), and STITEL (Speech Transmission Index for Telecommunication channels) methods [1, 13-15]. Attempts are also being made to combine the two approaches [16, 17]. A common feature of the above measures of intelligibility is their reference - the need for a "clean" (non-noisy, undistorted) signal, without which the evaluation procedure is impossible. The absence of such signals in real conditions leads to the need to develop non-reference methods [18–21, 23]. This is one of the most complex and actual problems in the considerate area.

2. Research objective

The objective of this work is to study the possibility of obtaining non-reference estimates of the noisy speech signals intelligibility using the reference method and a subsystem that restores the reference (quasi-reference) signal from the distorted signal.

In this article, we consider a technique based on the use of the measure of speech intelligibility SNR loss [7] and the noise reduction method [22, 26]. The chosen approach, which consists of a non-reference estimation of the values of an already existing reference method, can significantly simplify the development process. Since the original reference SNR loss method has been verified (compared in terms of reliability with subjective estimates), the reliability of the proposed non-reference approach can be estimated by the proximity of the obtained estimates to the true SNR loss values.

3. SNR loss original method

The SNR loss method is a development of the formant AI method and, in contrast, takes into account the influence on speech intelligibility of the work of non-linear methods used for noise reduction in speech. Let us briefly consider the SNR loss method described in detail in [7].

Calculations are carried out for individual time intervals (frames). The SNR loss in band j and frame m is defined as follows:

$$\begin{aligned} L(j,m) &= \mathrm{SNR}_X(j,m) - \mathrm{SNR}_{\hat{X}}(j,m) \\ &= 10\log_{10}\frac{X(j,m)^2}{D(j,m)^2} \\ &\quad -10\log_{10}\frac{\hat{X}(j,m)^2}{D(j,m)^2} \\ &= 10\log_{10}\frac{X(j,m)^2}{\hat{X}(j,m)^2} \end{aligned} \tag{1}$$

where $\mathrm{SNR}_X(j,m)$ is the input SNR in band j and interval m, $\mathrm{SNR}_{\hat{X}}(j,m)$ is the effective SNR of the enhanced signal in the jth frequency band, and $\hat{X}(j,m)$ is the excitation spectrum of the processed (enhanced) signal in the jth frequency band at the mth frame, $X(j,m)$ is the excitation spectrum of the clean signal in band *j* at frame *m*, and $D(j,m)$ is the excitation spectrum of the masker (noise). When $\hat{X}(j,m) = X(j,m)$, $\mathrm{SNR}_{\mathrm{LOSS}}$ is zero ($L(j,m) = 0$). In general, the value of $L(j,m)$ can be either positive or negative.

The values of $L(j,m)$ are limited within a certain range $[-\mathrm{SNR}_{\mathrm{lim}}, \mathrm{SNR}_{\mathrm{lim}}]$ dB as follows:

$$\hat{L}(j,m) = \min(\max(L(j,m), -\mathrm{SNR}_{\mathrm{lim}}), \mathrm{SNR}_{\mathrm{lim}}) \qquad (2)$$

The obtained values of $\hat{L}(j,m)$ at the previous stage are subsequently mapped to the range of $[0,1]$ using the following equation:

$$\mathrm{SNR}_{\mathrm{LOSS}}(j,m) = \begin{cases} -\dfrac{C_-}{\mathrm{SNR}_{\mathrm{lim}}}\hat{L}(j,m), & \hat{L}(j,m) < 0 \\ \dfrac{C_+}{\mathrm{SNR}_{\mathrm{lim}}}\hat{L}(j,m), & \hat{L}(j,m) \geq 0 \end{cases} \qquad (3)$$

where C_+ and C_- are parameters (defined in the range of [0, 1]) controlling the slopes of the mapping function.

For each time interval, the $\mathrm{SNR}_{\mathrm{LOSS}}(j,m)$ is averaged over all critical bands as follows:

$$f\mathrm{SNR}_{\mathrm{LOSS}}(m) = \frac{\sum_{j=1}^{K} W(j) \cdot \mathrm{SNR}_{\mathrm{LOSS}}(j,m)}{\sum_{j=1}^{K} W(j)} \qquad (4)$$

where $W(j)$ is a weight function that takes into account the psychoacoustic regularities of speech signals perception.

The average $\overline{\mathrm{SNR}}_{\mathrm{LOSS}}$ value is calculated by averaging $f\mathrm{SNR}_{\mathrm{LOSS}}(m)$ over all windows as follows:

$$\overline{\mathrm{SNR}}_{\mathrm{LOSS}} = \frac{1}{M}\sum_{m=0}^{M-1} f\mathrm{SNR}_{\mathrm{LOSS}}(m) \qquad (5)$$

where M is a number of windows into which the signal is divided. The resulting value, denoted as $\mathrm{SNR}_{\mathrm{LOSS}}$ for convenience, varies in the range from 0 to 1. A zero value corresponds to ideal intelligibility, and a single value corresponds to its complete absence. Graphs characterizing the relationship of intelligibility in percent with $\mathrm{SNR}_{\mathrm{LOSS}}$ values are presented in [7].

It should be clarified that in the text of the article, for convenience, two spellings are used: SNR loss refers to the method and $\mathrm{SNR}_{\mathrm{LOSS}}$ - for the values obtained using this method.

4. Non-reference estimation methodology

The idea underlying the proposed non-reference version of the SNR loss criterion is to calculate the $\mathrm{SNR}_{\mathrm{LOSS}}$ value for the estimated (noisy) signal and the signal obtained by noise reduction using the method proposed by Scalart and Filho [22]. In this case, the signal at the output of the noise reduction system is considered as clean (reference). In this case, this is acceptable, since the signal at the output of the noise reduction system is an estimate of a clean (non-noisy) signal.

Let us analyze the relationship between the $\mathrm{SNR}_{\mathrm{LOSS}}$ values calculated in this way (denoted as $\mathrm{SNR}_{\mathrm{LOSS}^1}$) and the true $\mathrm{SNR}_{\mathrm{LOSS}}$ values obtained using a pure signal as a reference. Six non-noisy speech fragments were used for modelling, each of which was noisy with additive white Gaussian noise (AWGN). For each fragment, 230 noisy versions were formed: 5 noise realizations for each SNR from –15 to 30 dB. During the simulation, $\mathrm{SNR}_{\mathrm{LOSS}}$ values were measured for 1380 signals (6 phrases with 230 noise variants for each). The scatter plot between true $\mathrm{SNR}_{\mathrm{LOSS}}$ values and $\mathrm{SNR}_{\mathrm{LOSS}^1}$ values is shown in Figure 1.

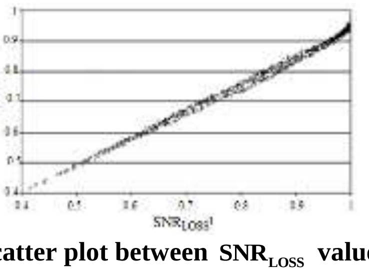

Fig (1) Scatter plot between $\mathrm{SNR}_{\mathrm{LOSS}}$ values and $\mathrm{SNR}_{\mathrm{LOSS}^1}$ for AWGN

From the Figure1, it can be seen that when the SNR changes in the range from –15 to 30 dB, the true (i.e., measured reference - using a pure signal) value of the SNR_{LOSS} criterion changes in the range from approximately 0.4 to 1. The set of points in Figure 1 is distributed in such a way that allows us to assume the possibility of a linear approximation of the dependence of SNR_{LOSS} on SNR_{LOSS^1}. Based on the available data and the least-squares method (LSM), the relationship can be described by the paired linear regression (PLR) equation as follows:

$$SNR_{LOSS} = (b_1 * SNR_{LOSS^1}) + b_0 \qquad (6)$$

It is established that for this type of noise, the coefficients of equation (6) take the following values: $b_1 = 0.8909;\ b_0 = 0.043$. Even when using linear regression, a high value of the determination coefficient (above 0.99) is observed, which characterizes the relationship between the true SNR_{LOSS} values and non-reference SNR_{LOSS} estimates obtained by substituting the SNR_{LOSS^1} values into eq. (6).

The use of power (exponentiation) or polynomial functions to describe the studied dependence leads to an additional slight increase in the accuracy of the approximation. Choosing a linear function was made taking into account the possibility of achieving high accuracy and preserving simplicity.

To test the proposed non-reference methodology for estimating the SNR_{LOSS} values, additional modelling was carried out. For ensuring reliability, a set of speech fragments was chosen that is different from that used in the first part of the research. For 24 speech fragments, a total of 5520 pairs of SNR_{LOSS} on SNR_{LOSS^1} values were measured (for different implementations of noise and SNR). The obtained SNR_{LOSS^1} values were substituted into expression (6). Thus, 5520 non-reference estimates of SNR_{LOSS} values were found. Comparison of non-reference estimates with the true values of the criterion obtained by the reference method makes it possible to foresee the sufficiently high accuracy of the proposed method of non-reference evaluation, as shown in Figure 2.

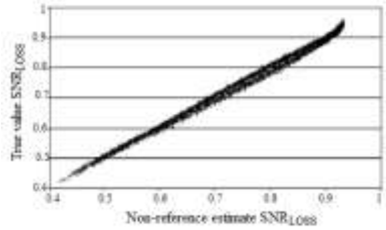

Fig (2) Scatter plot of true SNR_{LOSS} values and its non-reference estimates for AWGN

When using a linear function, the average value of the absolute error is approximately 0.008, and its maximum value is 0.036. The average value of the relative error is approximately 1.05%, and its maximum value is 4.72%. we can note that the error rates can be further reduced by using polynomial or piecewise linear approximations. The obtained results indicate the perspective of applying the proposed approach to the non-reference estimation of the SNR loss values.

The next step of the research is to test the applicability of this approach to other types of noise that are most often encountered in solving practical problems. For this purpose, a specialized Signal Processing Information Base (SPIB) was used [24], containing different noise data measured in the field by Speech Research Unit (SRU). Since the sampling frequency of the test speech signals was 8 kHz, the sampling frequency of the noise recordings was also reduced to 8 kHz. The original titles of the noise recordings are given in the paper.

To perform a non-reference estimation of SNR_{LOSS} values in the case when an arbitrary type of noise influences the signal, three approaches are possible:

Approach 1: use of regression expression (6) with parameters obtained for AWGN;

Approach 2: use the regression expression (6) with the parameters obtained for a mixed sample consisting of speech signals noisy with different types of noise;

Approach 3: finding new parameters of dependence (6), which individual for each specific type of noise.

The first two approaches are more universal and convenient in practice, but the third approach has the potential to provide greater accuracy.

To implement the second and third approaches, it is necessary to estimate the new coefficients of equation (6). To find the coefficients for the implementation of the second approach, a mixed sample was used, consisting of speech signals distorted by noises of various types. It is established that in this case, the coefficients of equation (6) take the following values: $b_1 = 0.8879;\ \ b_0 = 0.0479$. At the same time, a high value of the coefficient of determination is observed.

To implement the third approach, it is necessary to find the coefficients of the regression equation, which individual for each type of noise. As with AWGN, linear regression is well suited to describe the relationship between $\mathrm{SNR}_{\mathrm{LOSS}}$ and $\mathrm{SNR}_{\mathrm{LOSS}^1}$ values. The simulation was carried out to find the coefficients. For each type of noise, 1380 signals were formed (6 phrases with 230 variants of the noise of each phrase). Based on these samples, for each type of noise, we found the coefficients of linear regression that relates the values of $\mathrm{SNR}_{\mathrm{LOSS}}$ and $\mathrm{SNR}_{\mathrm{LOSS}^1}$, and also the root-mean-square deviation (RMSD) and the determination coefficient ($R^2 = 1 - \frac{\mathrm{SNR}_{\mathrm{LOSS}}}{\mathrm{SNR}_{\mathrm{LOSS}^1}}$) were estimated and given in Table 1. The values of the determination coefficient close to unity indicate the applicability of the linear model and the close relationship between the values of $\mathrm{SNR}_{\mathrm{LOSS}}$ and $\mathrm{SNR}_{\mathrm{LOSS}^1}$. It has been established that the highest RMSD and the lowest value of the determination coefficient are observed in the case of finding the linear regression parameters for the "Speech babble" noise.

Table (1) Linear regression parameters for relationship between $\mathrm{SNR}_{\mathrm{LOSS}}$ and $\mathrm{SNR}_{\mathrm{LOSS}^1}$ values.

Noise type	b_1	b_0	RMSD	R^2
AWGN	0.891	0.043	0.010	0.995
Car interior noise	0.894	0.039	0.010	0.997
Engine Room Noise (Destroyer)	0.892	0.42	0.015	0.991
Operations Room Background Noise (Destroyer)	0.879	0.56	0.011	0.995
Cockpit Noise 3 (F-16)	0.891	0.040	0.011	0.995
Factory Floor Noise 1 (plate-cutting and electrical welding equipment)	0.877	0.068	0.015	0.990
Factory Floor Noise 2 (car production hall)	0.872	0.057	0.012	0.995
Cockpit Noise 1 (Buccaneer Jet Traveling at 190 knots)	0.891	0.042	0.010	0.995
Cockpit Noise 2 (Buccaneer Jet Traveling at 450 knots)	0.889	0.046	0.010	0.995
Military Vehicle Noise (Leopard)	0.870	0.057	0.011	0.996
Military Vehicle Noise (M109 tank)	0.869	0.055	0.010	0.997
Speech babble	0.889	0.070	0.022	0.982

Next, the above approaches were compared in terms of accuracy. For this purpose, for each type of noise, as in the case of AWGN, 5520 pairs of $\mathrm{SNR}_{\mathrm{LOSS}}$ and $\mathrm{SNR}_{\mathrm{LOSS}^1}$ values were measured for 24 phrases (for different implementations of noise and SNR). The obtained values $\mathrm{SNR}_{\mathrm{LOSS}^1}$ were substituted into expression (6) with the corresponding coefficients, depending on the applied approach. After that, the obtained estimates of the $\mathrm{SNR}_{\mathrm{LOSS}}$ values were compared with the true values obtained using the reference method. The values of the mean absolute error (MAE), the root of the mean square error (RMSE), and the mean relative percentage error (MRPE) were calculated as follows:

$$\left.\begin{aligned}
\mathrm{MAE} &= \frac{1}{N_T}\sum_{l=1}^{N_T}\left|\mathrm{SNR}_{\mathrm{LOSS}_l}^{\ \mathrm{est}} - \mathrm{SNR}_{\mathrm{LOSS}_l}^{\ \mathrm{true}}\right| \\
\mathrm{RMSE} &= \sqrt{\frac{1}{N_T}\sum_{l=1}^{N_T}(\mathrm{SNR}_{\mathrm{LOSS}_l}^{\ \mathrm{est}} - \mathrm{SNR}_{\mathrm{LOSS}_l}^{\ \mathrm{true}})^2} \\
\mathrm{MRPE} &= \frac{100\%}{N_T}\sum_{l=1}^{N_T}\frac{\left|\mathrm{SNR}_{\mathrm{LOSS}_l}^{\ \mathrm{est}} - \mathrm{SNR}_{\mathrm{LOSS}_l}^{\ \mathrm{true}}\right|}{\mathrm{SNR}_{\mathrm{LOSS}_l}^{\ \mathrm{true}}}
\end{aligned}\right\} \quad (7)$$

where $\mathrm{SNR}_{\mathrm{LOSS}_l}^{\ \mathrm{est}}$ is a non-reference estimate of the $\mathrm{SNR}_{\mathrm{LOSS}}$ values for the l-th signal, $\mathrm{SNR}_{\mathrm{LOSS}_l}^{\ \mathrm{true}}$ is the true value of $\mathrm{SNR}_{\mathrm{LOSS}}$ values for the l-th signal, N_T is the total number of signals for which $\mathrm{SNR}_{\mathrm{LOSS}}$ values are measured.

The results obtained using the simulation allow us to compare the accuracy of the proposed approaches, given in Table 2. As expected, the third approach provides the best accuracy. The first and second approaches have comparable accuracy. A version of the method using individual coefficients of the regression equation should be used in cases where it is possible to estimate the coefficients of equation (6) corresponding to specific noise conditions. In cases where the noise conditions are not known in advance or the type of noise may change significantly over time, the third approach is not possible and the first or second approach is recommended.

Table (2) Accuracy of non-reference estimation of $\mathrm{SNR}_{\mathrm{LOSS}}$ values using different approaches

Noise type	Mean relative percentage error (MRPE), %		
	Approach 1	Approach 2	Approach 3
AWGN	1.05	1.07	1.05
Car interior noise	1.87	2.43	1.60
Engine Room Noise (Destroyer)	1.58	1.65	1.57
Operations Room Background Noise (Destroyer)	1.27	1.17	1.14
Cockpit Noise 3 (F-16)	1.18	1.34	1.13
Factory Floor Noise 1 (plate-cutting and electrical welding equipment)	2.11	1.87	1.56
Factory Floor Noise 2 (car production hall)	1.42	1.40	1.28
Cockpit Noise 1 (Buccaneer Jet Traveling at 190 knots)	1.06	1.12	1.06
Cockpit Noise 2 (Buccaneer Jet Traveling at 450 knots)	1.09	1.09	1.08
Military Vehicle Noise (Leopard)	1.35	1.36	1.17
Military Vehicle Noise (M109 tank)	1.43	1.48	1.26
Speech babble	3.55	3.22	2.20
Average	1.58	1.60	1.34

5. Improvement of the proposed methodology

After modelling with various types of noise, the viability of the proposed non-reference technique is confirmed, but the question of its further improvement is actual. An analysis of the number of addition and multiplication operations required for a non-reference estimation of the intelligibility of a speech signal showed that a significant part of them are based on the fast Fourier transform (FFT) and inverse FFT (IFFT). However, the used calculation block diagram, shown in Figure 3 [27], contains operations that can be eliminated. In this case, it is extremely important to add the parameters of the spectral transformation in the noise reduction block to those used in the SNR loss method, since they affect the estimated reliability of the intelligibility. The block diagram of the proposed modification, based on the elimination of repeated and mutually exclusive operations, has a simpler structure and contains one FFT calculation block as shown in Figure 4.

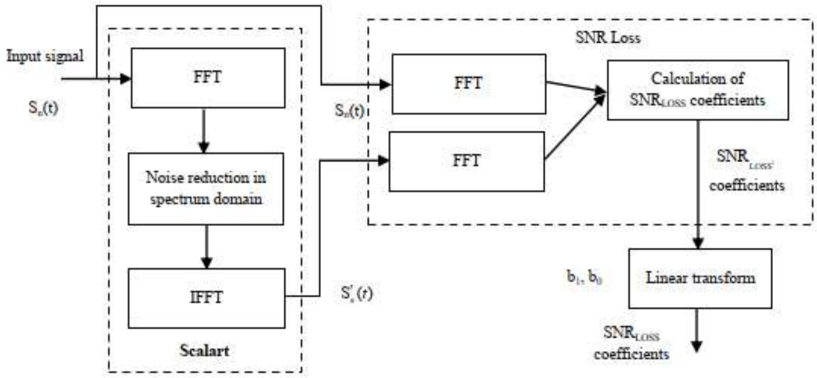

Fig (3) Block diagram of the original non-reference methodology for estimating intelligibility

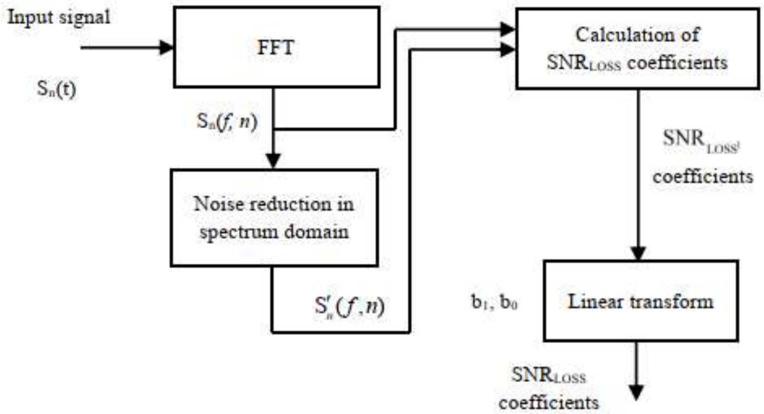

Fig (4) Block diagram of the modified methodology

The proposed modification is analyzed and compared with the original method in terms of the number of required computational operations. It was found that the proposed modification allows reducing the number of operations by about 30.5% compared to the original method.

We also compared the speed of work of two versions of methods based on modelling in the Matlab platform, and the personal computer with the following configuration was used: AMD Ryzen 5 3500U with Radeon Vega Mobile Gfx 2.10 GHz, RAM 8.00GB, 64-bit operating system, x64-based processor, Windows 10 Pro. As a result, the time gain of the proposed modification was 29.9%.

As the simulation showed, the proposed modification has slightly lower accuracy. The comparison results for MAE, RMSE, and MRPE are shown in Figure 5. The graphs show averaged characteristics over all types of noise.

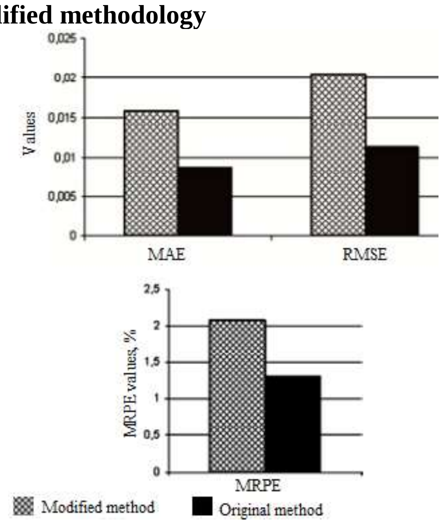

Fig (5) Comparison of methods accuracy based on MAE, RMSE, and MRPE

The presented diagrams demonstrate that the accuracy of estimates of the proposed modification is somewhat inferior to the accuracy of the original method. The ratio of accuracy for the original method and its modification is not as informative as the absolute change in these indices. For example, the mean percentage error (MAPE) increased from about 1.3% to 2.1%. The relative change of 1.6 times seems significant.

6. Conclusion

In this article, a non-reference version of the measure of intelligibility of noisy speech signals is proposed – SNR loss. It is based on the use of the original (reference) version of the SNR loss method, the noise reduction method, and pairwise regression.

The accuracy of the proposed non-reference measure of intelligibility is studied in the case when various types of noise are affected by speech signals. The obtained results indicate the relatively high accuracy of the proposed method of non-reference estimation (the average relative error is 1.05–3.55%).

A faster version of the non-reference methodology is also presented. It follows from the research results that the modification has a fairly high accuracy, which is slightly inferior to the original, while the speed of work has increased due to a significant reduction in the number of computational operations performed. The proposed modified method can be used to automatically control noise reduction systems, as well as to select a transmission mode that provides an acceptable level of intelligibility in communication systems.

Further work consists in conducting experiments using video recordings with speech phrases.

Funding information: this research is funded by Damascus university – funder No. (501100020595).

References

1. Didkovskiy V.S., Didkovskaya M.V., Prodeus A.N. Acoustic assessment of speech communication channels. Kiev. Imex-LDT, 2008. 420 p.
2. K. Mohanaprasad, Anjali Singh, Karishma Sinha & Tejal Ketkar. Noise reduction in speech signals using adaptive independent component analysis (ICA) for hands free communication devices. International Journal of Speech Technology volume 22, pages 169–177 (2019)
3. Collard J. A theoretical study of the articulation and intelligibility of a telephone circuit // Electrical Communication (1929). V. 7. p. 168.
4. French N.R., Steinberg J.C. Factors governing the intelligibility of speech sounds // The journal of the Acoustical Society of America (1947) V. 19. Is. 1. pp. 90-119.
5. Kryter K.D. Methods for the calculation and use of the articulation index // The Journal of the Acoustical Society of America (1962) V. 34. Is. 11. pp. 1689-1697.
6. Kryter K.D. Validation of the articulation index // The Journal of the Acoustical Society of America (1962) V. 34. Is. 11. pp. 1698-1702.
7. Ma J., Loizou P. SNR loss: a new objective measure for predicting the intelligibility of noise-suppressed speech // Speech Communication (2011) V. 53. Is. 3. pp. 340-354.
8. Loizou P., Ma J. Extending the articulation index to account for non-linear distortions introduced by noise-suppression algorithms // Journal of the Acoustical Society of America (2011) V. 130. Is. 2. pp. 986–995.
9. Bykov Y.S. The theory of speech intelligibility and efficiency of radio communication – Moscow-Leningrad: Gosenergoizdat, 1959. 350 p.

10. Pokrovskiy N.B. Calculation and measurement of speech intelligibility. – Moscow: Svyaz'izdat, 1962. 390 p.
11. Sapozhkov M.A. Speech signal in cybernetics and communication. – Moscow: Svyaz'izdat, 1963. 452 p.
12. Hiroto Takahashi1, Kazuhiro Kondo. On non-reference speech intelligibility estimation using DNN noise reduction. 23rd International Congress on Acoustics, Aachen, Germany, (2019).
13. Houtgast T., Steeneken H.J.M. Evaluation of speech transmission channels by using artificial signals // Acta Acustica united with Acustica (1971) V. 25. Is. 6. pp. 355-367.
14. Steeneken H.J.M., Houtgast T. A physical method for measuring speech transmission quality // The Journal of the Acoustical Society of America (1980) V. 67. Is. 1. pp. 318-326.
15. Steeneken H.J.M., Houtgast T. Validation of the revised STIr method // Speech Communication (2002) V. 38. Is. 3. pp. 413-425.
16. Prodeus A. On possibility of advantages join of formant and modulation methods of speech intelligibility evaluation // Proceedings of VII International Conference on Perspective Technologies and Methods in MEMS Design (MEMSTECH) (2010) pp. 254-259.
17. Prodeus A. Assessment of speech Intelligibility by formant-modulation method // Journal of Basic and Applied Physics (2013) V. 2 Is. 5. pp. 10-18.
18. Falk T.H., Zheng C., Chan W.Y. A non-intrusive quality and intelligibility measure of reverberant and dereverberated speech // IEEE Transactions on Audio, Speech, and Language Processing (2010) V. 18. Is. 7. pp. 1766-1774.
19. Santos J.F., Senoussaoui M., Falk T.H. An improved non-intrusive intelligibility metric for noisy and reverberant speech // 14th International Workshop on Acoustic Signal Enhancement (IWAENC) (2014), pp. 55-59.
20. Li F.F. Speech intelligibility of VoIP to PSTN interworking – a key index for the QoS. // Proceedings of International Conference "Telecommunications Quality of Service: The Business of Success" (2004), pp. 104-108.
21. Chen F., Hazrati O., Loizou P. C. Predicting the intelligibility of reverberant speech for cochlear implant listeners with a non-intrusive intelligibility measure // Biomedical signal processing and control (2013), V. 8. Is. 3. pp. 311-314.
22. Scalart P., Filho J. Speech enhancement based on a priori signal to noise estimation // IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-96). 1996. V. 2. pp. 629-632.
23. Wissam A. Jassima and Naomi Harte. Estimation of a priori signal-to-noise ratio using neurograms for speech enhancement. The Journal of the Acoustical Society of America 147, 3830 (2020).
24. Don H. Johnson. Signal Processing Information Base (SPIB) (1990). http://spib.linse.ufsc.br/noise.html
25. Tassadaq Hussain1, Muhammad Diyan et al. A Novel Speech Intelligibility Enhancement Model based on Canonical Correlation and Deep Learning. Cornell university, (2022)
26. Mbachu C. B1., Akaneme S. A. Noise reduction in speech signals using recursive least adaptive algorithm. International Journal of Engineering and Advanced Technology Studies, Vol.8, No.2, pp.1-17 (2020).
27. Nesterov M.S., Topnikov A.I. Modifikatsiya metodiki neetalonnoy otsenki razborchivosti rechevykh signalov // 16-ya mezhdunarodnaya konferentsiya «Tsifrovaya obrabotka signalov i ee primenenie – DSPA-2014». M., 2014. T. 1. – S. 208-212. [Rus]