TELECOMMUNICATIONS AND RADIO ENGINEERING

UDC 621.391.83 (045)
DOI:10.18372/1990-5548.69.16428

Arkadiy M. Prodeus

EQUALIZATION OF THE MEASURING SYSTEM FREQUENCY RESPONSE IN THE OBJECTIVE ASSESSMENT OF SPEECH INTELLIGIBILITY

National Technical University of Ukraine “Ihor Sikorsky Kyiv Polytechnic Institute,” Kyiv, Ukraine
E-mail: aprodeus@gmail.com ORCID 0000-0001-7640-0850

Abstract—*Voice control of an unmanned aerial vehicle has a number of advantages if the operator is indoors. In this case, the distortions of speech commands caused by the influence of noise interference can be significantly reduced. However, the disadvantage of such control is the negative impact of reverberation on speech intelligibility. Therefore, it is advisable to perform a preliminary assessment of speech intelligibility in the room before the session of unmanned aerial vehicle controlling. This assessment can be performed by the modulation method, using the room impulse response estimate. If a non-professional quality loudspeaker and microphone are used to estimate the room impulse response, errors in the room impulse response estimate can affect the results of speech intelligibility assessment. In this paper, two techniques of equalizing of non-professional quality level audio equipment used in assessing the room impulse response have been compared. It is shown that a dividing the frequency response of the “loudspeaker-room-microphone” system into the amplitude frequency response of the “loudspeaker-microphone” subsystem provides almost the same equalization quality as a more complex technique of adaptive filtering. At the same time, studies have shown that such equalization is not necessary, provided that the frequency response unevenness of the “loudspeaker-microphone” subsystem does not exceed 8–10 dB in the frequency range from 100 Hz to 11 kHz.*

Index Terms—Speech intelligibility; room impulse response; frequency response equalization; audio equipment of non-professional quality level.

I. INTRODUCTION

The operator giving voice commands to the unmanned aerial vehicle (UAV) may be indoors [1]. This location of the operator has its advantages over the location in the open space, as it significantly reduces the impact of noise (wind, transport, combat, etc.) on the intelligibility of commands given by the operator [2]. However, the disadvantage of control in the room is the possibility of a significant reduction in speech intelligibility due to the effect of reverberation caused by reflections of sound from reflective surfaces such as walls, ceilings, windows, furniture etc. [3].

Although headsets are commonly used to reduce the effect of reverberation [4], there may be a situation where the microphone must be placed at a certain distance from the speaker. This is an unfavourable situation, from the point of view of UAV control, because the intelligibility of speech distorted by reverberation deteriorates with increasing distance between the speaker and the microphone [5]. This is explained by the fact that the power of the direct signal decreases and becomes close to the power of the reflected signals [6]. Exceptions to this rule are the places in the premises where the microphone receives almost simultaneously a direct signal and strong reflected signals (so-named "early reflections"). Such an exceptional situation, for example, occurs if the microphone is located near the walls [5].

Given the above features of the impact of reverberation on speech intelligibility, it is advisable, before the UAV control session, to perform a preliminary assessment of speech intelligibility in the room for the specified locations of the operator and microphone.

II. PROBLEM STATEMENT

Speech intelligibility predicting is difficult to perform analytically or by computer simulation. It is easier and more reliable to assess speech intelligibility experimentally using the modulation method [7]. With this approach, one can first experimentally evaluate the room impulse response (RIR), using the layout of Fig. 1, and then calculate the speech intelligibility using the Schroeder formula [8].

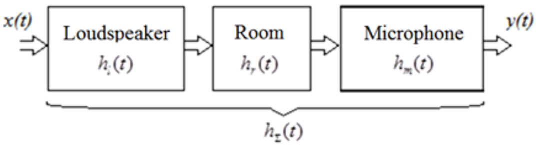


Fig. 1. Layout of RIR evaluating

In the experimental RIR evaluation, the test sound signal $x(t)$ is emitted using a loudspeaker located at the point of the room where the operator is located. The response $y(t)$ of the room to the stimulus $x(t)$ is perceived by a microphone located at another point in the room.

Because the measuring system contains a loudspeaker and a microphone, a convolution

$$h_{\Sigma}(t) = h_l(t) \otimes h_r(t) \otimes h_m(t) = h_r(t) \otimes h_{lm}(t) \qquad (1)$$

will actually be evaluated instead of the RIR $h_r(t)$, where $\otimes$ is convolution symbol, $h_l(t)$ is loudspeaker impulse response, $h_m(t)$ is microphone impulse response, $h_{lm}(t) = h_l(t) \otimes h_m(t)$ is “loudspeaker-microphone” (LM) subsystem impulse response.

If it is possible to calculate the impulse response $h_{lm}^{-1}(t)$ of the system inverse to LM subsystem, then the LM subsystem can be equalized:

$$h_r(t) = h_{\Sigma}(t) \otimes h_{lm}^{-1}(t) = h_r(t) \otimes h_{lm}(t) \otimes h_{lm}^{-1}(t) \qquad (2)$$

because $h_{lm}(t) \otimes h_{lm}^{-1}(t) = \delta(t)$, where $\delta(t)$ is Dirac delta function.

The calculation of the $h_{lm}^{-1}(t)$ can be performed in different ways. In this paper, two methods are considered. The first way is to use an expression

$$h_{lm}^{-1}(t) = \mathbb{F}^{-1}\left\{ \frac{1}{|H_{lm}(f)|} \cdot M_R(f) \right\} \qquad (3)$$

where $\mathbb{F}^{-1}$ is the symbol of the inverse Fourier transform, $H_{lm}(f)$ is frequency response of the LM subsystem, $|\cdot|$ is module symbol, $M_R(f)$ is regularization factor [9].

The problem with the expression (3) is the division operation, because the amplitude-frequency response $H_{lm}(f)$ of the LM subsystem may contain small numerical values, which will lead to emergency shutdown of the computer application. The presence of a multiplier $M_R(f)$ neutralizes this effect, although the main task of this multiplier is to reduce the variance of the $h_{lm}^{-1}(t)$ estimate [9]. The disadvantage of this method is that it does not take into account the properties of the phase-frequency response $\theta_{lm}(f)$ of the LM subsystem.

The second method of calculating the inverse filter $h_{lm}^{-1}(t)$ is deprived of this disadvantage and consists in the use of an adaptive filter [10].

Thus, the purpose of this work is to compare two methods of correction of the frequency response of the LM subsystem. The first method is reduced to the calculation of the amplitude-frequency response $|H_{lm}(f)|$, and the second method is reduced to the calculation, by the method of adaptive filtering, the coefficients of the inverse filter $h_{lm}^{-1}(t)$.

In addition, it is interesting to evaluate the difference between the intelligibility scores for cases where the frequency response is equalized and in the absence of such equalization.

III. Organization of Studies

When comparing the two methods of correction of the frequency response of the LM subsystem it is advisable to use speech intelligibility as a measure of the correction quality.

After calculating the RIR, the assessment of speech intelligibility can be performed by modulation method [7]. The first step in this evaluation is to calculate, according to Schroeder's formula [8], the modulation transfer coefficients:

$$m_{ki} = \left| \int_0^{\infty} h_k^2(t) \exp(-j2\pi F_i t) dt \right| \Big/ \int_0^{\infty} h_{ki}^2(t) dt, \qquad (4)$$

where $h_k(t)$ is the result of filtering the function $h_r(t)$ by a k th bandpass filter (seven octave filters with central frequencies from 125 Hz to 8 kHz are used); F_i is modulation frequency (fourteen F_i values are used, in the range from 0.63 Hz to 12.5 Hz).

The next step is calculation of the Speech Transmission Index (STI) as speech intelligibility measure [11].

The LM subsystem used in experimental studies contained electroacoustic equipment of different quality. These are a household active loudspeaker Genius SP-HF 2.0 500 (14 W, 65–20000 Hz, SNR 73 dB) and omnidirectional measuring condenser microphones Superlux ECM-999 (20-20000 Hz, dynamic range 106 dB, SNR 70 dB).

Recording signals from the microphone output was performed with a sampling frequency of 44.1 kHz and a quantization depth of 24 bits. The test signal was based on a maximum length sequence (MLS) contained 216 samples, which corresponds to a signal length of 1.49 s at a sampling rate of 44.1 kHz. This MLS was repeated 17 times during radiation, which allowed to average the last 16 bursts of the RIR estimate to increase the signal-to-noise ratio by 12 dB.

Auditorium №209 of the Department of Acoustic and Multimedia Electronic Systems of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute" was selected for measurements. This is a medium-sized auditorium with the following characteristics: dimensions 10x15x3.1 m, four windows, door, two bookcases, wardrobe, three rows of 9 desks in each row, a teacher's desk and 2 additional free tables. Nine students and one teacher were in the room during the recording of signals. The distances from the loudspeaker to the points 1, 2, 3, 4, 5 and 6, where the microphone was placed, were 3 m, 6 m, 9 m, 14 m, 9 m and 9 m, respectively. The first four points were located on a straight line between the front and back walls of the room. The fifth and sixth points were located at a distance of 1.5 m from the left and right side walls [13].

All calculations were performed in Matlab R2015b, using the RLS method for adaptive filtering.

IV. Results of Studies

The results of spectral analysis of the signal $y(t)$ recorded in a muffled room are shown in Fig. 2. The spectrum estimate Type 0 graph is the result of averaging 16 raw periodograms of the signal $y(t)$. Calculation algorithms of the Types 1–3 estimates can be found in [12], they are variations of smoothing the Type 0 estimate with a triangular window with a width of 50 Hz.

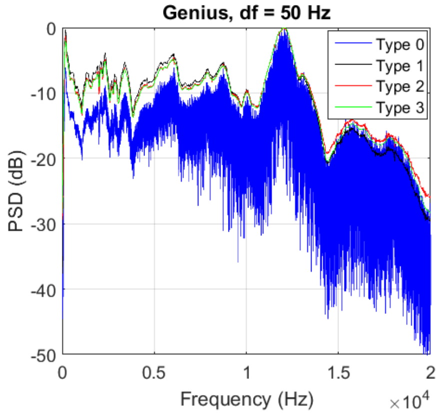


Fig. 2. Estimates $|H_{lm}(f)|$ of the LM subsystem

The estimates $|H_{lm}(f)|$ of the LM subsystem show that the non-uniformity of the frequency response in the frequency range 100–15000 Hz is significant and reaches 20 dB. However, in the frequency range 100–5000 Hz, where the majority of the energy of the speech signal is concentrated, the non-uniformity of the frequency response does not exceed 8–10 dB. Moreover, it can be seen (Fig. 2) that for the used equipment such unevenness remains in a wider frequency range from 100 Hz to 11 kHz.

The results of $h_{lm}(t)$ estimation, for studies in muffled room, before and after frequency equalization are shown in Fig. 3. Compensation of the non-uniformity of the frequency response of the LM subsystem was performed here in the first way, using (2) – (3).

As can be seen, the equalization significantly reduced the level of lateral petals of the $h_{lm}(t)$ estimate by 4 dB (Fig. 3a), although there are residual periodic bursts in the range of 0–0.01 s, the period of which is close to 0.003 s (Fig. 3b).

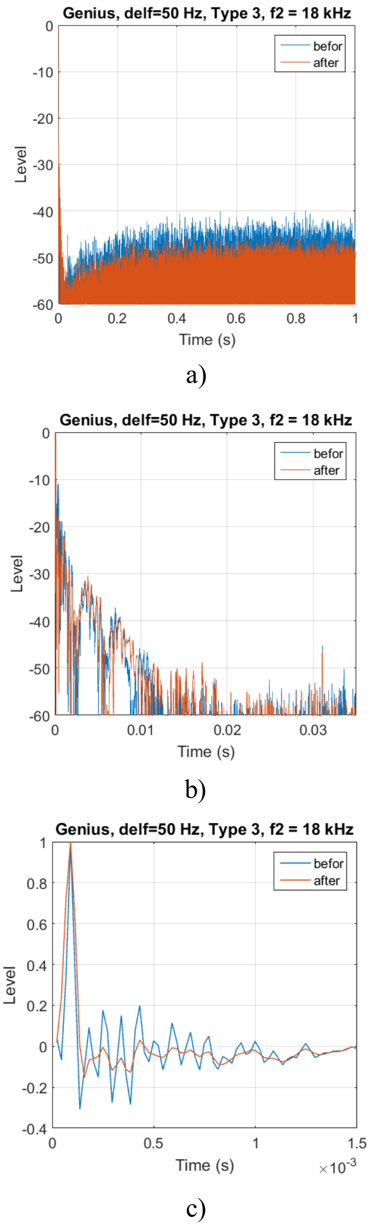
Increasing the resolution over time allows one to notice the presence of another residual periodicity with a period of 8×10^{-5} s (Fig. 3c). Thus, the above results indicate insufficient compensation of $|H_{lm}(f)|$ spectral bursts at frequencies of 300 Hz and 12 kHz. It is obvious that insufficient compensation of the $|H_{lm}(f)|$ spectral burst at a frequency of 300 Hz is highly undesirable.

The results of the $h_{lm}(t)$ evaluation, where the equalization of the LM subsystem was performed in the second way, are shown in Fig. 4 (for the adaptive filter order 150). Comparison of Figs 3 and 4 graphs indicates a certain advantage of the second method.

Indeed, the level of the side petals decreased by 6 dB, approaching the theoretical level of -48 dB (Fig. 4a). In Figures 4b and 4c, one can see that the periodicity of the $h_{lm}(t)$ structure is practically absent which indicates good compensation of $|H_{lm}(f)|$ bursts.

The results of the LM subsystem equalization for two different microphones of the same type are shown in Fig. 5. As can be seen, the non-uniformity of the adjusted frequency response, for both microphones, does not exceed 8 dB in the range of 100-5000 Hz and 3 dB in the band 5–10 kHz.

A comparison of speech intelligibility estimates for the first (Fig. 6a) [13] and second (Fig. 6b) equalization techniques shows that the maximum discrepancy of the results does not exceed 2.25%, and the average discrepancy is 0.54%. Thus, a simple, from a computational point of view, the first way of equalization the measuring system is almost not inferior to the second way, based on the use of the adaptive filtering technique.

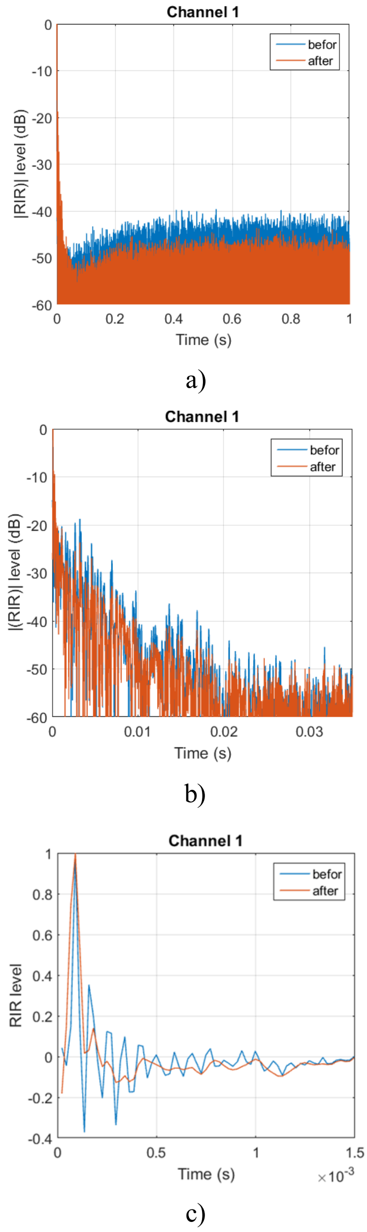


a)

b)

c)

Fig. 3. Estimate of $h_{lm}(t)$ for the first way of equalization at time intervals: 0–1 s (a); 0–0.033 s (b); 0–1.5 ms (c) [12]

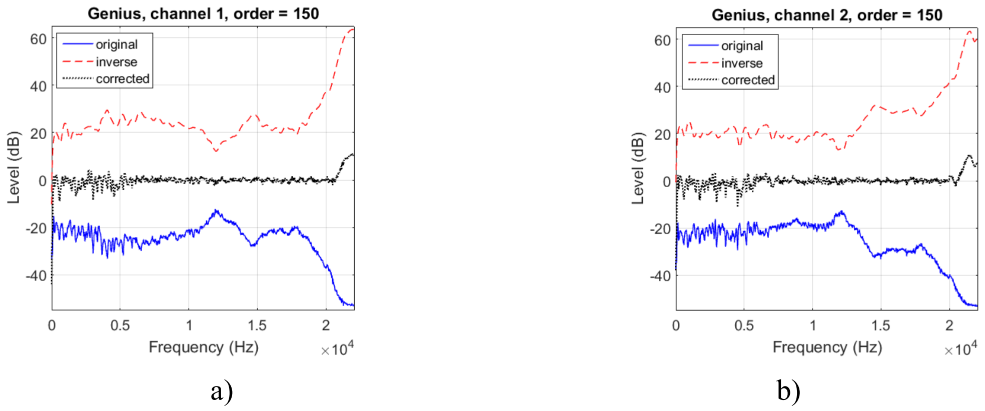


a)

b)

c)

Fig. 4. Estimate of $h_{lm}(t)$ for the second way of equalization at time intervals: 0–1 s (a); 0–0.033 s (b); 0–1.5 ms (c)



a)

b)

Fig. 5. Results of $|H_{lm}(f)|$ non-uniformity compensation in the first (a) and second (b) microphones

The results obtained can be explained as follows. As mentioned above, the non-uniformity of the frequency response of the studied LM subsystem did not exceed 8–10 dB in the frequency band 100–5000 Hz. In the case of equalization by the method of adaptive filtering, this non-uniformity was reduced to 2–3 dB, i.e., by 6–7 dB.

Although the first method of correction is somewhat inferior to the second method, but this loss is less than 4–5 dB, which cannot be significant in terms of speech intelligibility criterion.

V. Conclusions

Two methods of equalization the "loudspeaker-microphone" subsystem contained audio equipment of non-professional quality level were compared according to the criterion of speech intelligibility. It was shown that a dividing the frequency response of the “loudspeaker-room-microphone” system into the frequency response of the “loudspeaker-microphone” subsystem provides almost the same equalization quality as a more complex technique of adaptive filtering. At the same time, studies have shown that such equalization is not necessary, provided that the frequency response unevenness of the “loudspeaker-microphone” subsystem does not exceed 8–10 dB in the frequency range from 100 Hz to 11 kHz. The maximum difference between the intelligibility estimates does not exceed 1% for cases when the frequency response is equalized and in the absence of such equalization.

References

[1] R. Contreras, A. Ayala, and F. Cruz, "Unmanned Aerial Vehicle Control Through Domain-based Automatic Speech Recognition," *Computers*, 9(3), 75, September 2020. https://doi.org/10.3390/computers9030075

[2] J.-S. Park, and H.-J. Na, "Front-End of Vehicle-Embedded Speech Recognition for Voice-Driven Multi-UAVs Control," *Appl. Sci.*, 10(19), 6876, September 2020. https://doi.org/10.3390/app10196876

[3] W. Yang, and J. Bradley, “Effects of room acoustics on the intelligibility of speech in classrooms,” *J. of the Acoust. Soc. of Am.*, 125 (2), pp. 922–933, March 2009. https://doi.org/10.1121/1.3058900

[4] A. Waibel, and K.-F. Lee, *Readings in Speech Recognition*. Elsevier: 1990.

[5] A. Prodeus, and M. Didkovska, "Assessment of speech intelligibility in university lecture rooms of different sizes using objective and subjective methods," *Eastern-European Journal of Enterprise Technologies*, 3(5(111), pp. 47–56, 2021. https://doi.org/10.15587/1729-4061.2021.228405

[6] J. Bradley, H. Sato, and M. Picard, “On the importance of early reflections for speech in rooms,” *J. of the Acousti. Soc. of Am.*, 113 (6), pp. 3233–3244, 2003 https://doi.org/10.1121/1.1570439

[7] H. Steeneken, “Forty years of speech intelligibility assessment (and some history),” *Proc. of the Institute of Acoustics*, 36, Pt.3, 2014.

[8] M. Schroeder, “Modulation Transfer Functions: Definition and Measurement,” Acta Acust. united with Acust., vol. 49, no. 3, pp. 79–182(4), 1981.

[9] A. Tikhonov, “O nekorrektnykh zadachakh lineynoy algebry i ustoychivom metode ikh resheniya,” DAN USSR, 163(3), pp. 591–594, 1965.

[10] L. Morales (Ed), Adaptive filtering applications. In Tech, Croatia: 2011

[11] H. Steeneken, and T. Houtgast, “Validation of the revised STIr method,” Elsevier Speech Communication, vol. 38, pp. 26–37, 2002. https://doi.org/10.1016/S0167-6393(02)00010-9

[12] O. Dvornyk, A. Prodeus, M. Didkovska, and D. Motorniuk, "Artificial Software Complex "Artificial Head," Part 1. Adjusting the Frequency Response of the Path," *Microsystems, Electronics and Acoustics*, vol. 22, no. 1, pp. 56–64, 2020. https://doi.org/10.20535/2523-4455.mea.198431

[13] O. Dvornyk, A. Prodeus, D. Motorniuk, M. Didkovska, "Hardware and Software System "Artificial Head," Part 2. Evaluation of Speech Intelligibility in Classrooms," *Microsystems, Electronics and Acoustics*, vol. 22, no. 3, pp. 48–55, 2020. https://doi.org/10.20535/2523-4455.mea.209928

Received September 06, 2021

Prodeus Arkadiy. ORCID 0000-0001-7640-0850. Doctor of Engineering Sciences. Professor.
Department of Acoustic and Multimedia Electronic Systems, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute,” Kyiv, Ukraine.
Education: Kyiv Polytechnic Institute, Kyiv, Ukraine (1972).
Research interests: digital signal processing.
Publications: 178.
E-mail: aprodeus@gmail.com

А. М. Продеус. Вирівнювання частотної характеристики вимірювальної системи при об'єктивному оцінюванні розбірливості мовлення

Голосове керування безпілотним літальним апаратом має ряд переваг, якщо оператор знаходиться у приміщенні. В цьому випадку спотворення мовленнєвих команд, обумовлені впливом шумових перешкод, можна значно зменшити. Однак недоліком такого управління є негативний вплив реверберації на розбірливість мовлення. Тому перед сеансом управління безпілотним літальним апаратом доцільно провести попередню оцінку розбірливості мовлення в приміщенні. Цю оцінку можна виконати модуляційним методом, використовуючи оцінку імпульсної характеристики кімнати. Якщо для оцінки імпульсної характеристики кімнати використовуються гучномовець і мікрофон непрофесійної якості, помилки в оцінці імпульсної характеристики кімнати можуть вплинути на результати оцінки розбірливості мовлення. У цій роботі порівнюються два способи вирівнювання частотної характеристики аудіоапаратури непрофесійного рівня якості, що використовується для оцінки імпульсної характеристики приміщення. Показано, що ділення частотної характеристики системи «гучномовець-кімната-мікрофон» на амплітудну частотну характеристику підсистеми «гучномовець-мікрофон» забезпечує майже таку ж якість вирівнювання, як і більш складний спосіб адаптивної фільтрації. У той же час дослідження показали, що таке вирівнювання не є необхідним за умови, що нерівномірність частотної характеристики підсистеми «гучномовець-мікрофон» не перевищує 8–10 дБ в діапазоні частот від 100 Гц до 11 кГц.

Ключові слова: розбірливість мовлення; імпульсна характеристика приміщення; вирівнювання частотної характеристики; аудіоапаратура непрофесійного рівня якості.

Продеус Аркадій Миколайович. ORCID 0000-0001-7640-0850. Доктор технічних наук. Професор.
Кафедра акустичних та мультимедійних електронних систем, Національний технічний університет України «Київський політехнічний інститут ім. І. Сікорського», Київ, Україна.
Освіта: Київський політехнічний інститут, Київ, Україна (1972).
Напрямок наукової діяльності: цифрова обробка сигналів.
Кількість публікацій: 178.
E-mail: aprodeus@gmail.com

А. Н. Продеус. Выравнивание частотной характеристики измерительной системы при объективной оценке разборчивости речи

Голосовое управление беспилотным летательным аппаратом имеет ряд преимуществ, если оператор находится в помещении. В этом случае искажения речевых команд, обусловленные влиянием шумовых помех, можно существенно уменьшить. Однако недостатком такого управления является негативное влияние реверберации на разборчивость речи. Поэтому перед сеансом управления беспилотным летательным аппаратом целесообразно провести предварительную оценку разборчивости речи в помещении. Эту оценку можно выполнить модуляционным способом, используя оценку импульсной характеристики помещения. Если для оценки импульсной характеристики помещения используются громкоговоритель и микрофон непрофессионального качества, ошибки при оценке импульсной характеристики помещения могут повлиять на результаты оценки разборчивости речи. В данной работе сравниваются два метода выравнивания частотной характеристики аудиоаппаратуры непрофессионального уровня, используемой для оценки импульсной характеристики помещения. Показано, что деление частотной характеристики системы «громкоговоритель-комната-микрофон» на амплитудную частотную характеристику подсистемы «громкоговоритель-микрофон» обеспечивает почти такое же качество выравнивания, как и более сложный способ адаптивной фильтрации. В то же время исследования показали, что без такого выравнивания можно обойтись при условии, что неравномерность частотной характеристики подсистемы «громкоговоритель-микрофон» не превышает 8–10 дБ в диапазоне частот от 100 Гц до 11 кГц.

Ключевые слова: разборчивость речи; импульсная характеристика помещения; выравнивание частотной характеристики; аудиоаппаратура непрофессионального уровня качества.

Продеус Аркадий Николаевич. ORCID 0000-0001-7640-0850. Доктор технических наук. Профессор.
Кафедра акустических и мультимедийных электронных систем, Национальный технический университет Украины «Киевский политехнический институт им. И. Сикорского», Киев, Украина.
Образование: Киевский политехнический институт, Киев, Украина, (1972).
Направление научной деятельности: цифровая обработка сигналов.
Количество публикаций: 178.
E-mail: aprodeus@gmail.com