Archives of Acoustics Vol. 46, No. 3, pp. 547–559 (2021), doi: 10.24425/aoa.2021.138147

Research Paper

Indoor Sound Pressure Level From Service Equipment in Buildings: Influence of Testing Methods on Measurement Results

Elżbieta NOWICKA*, Ewa SZEWCZAK

Building Research Institute
Warsaw, Poland
*Corresponding Author e-mail: e.nowicka@itb.pl

(*received October 12, 2020; accepted April 25, 2021*)

Indoor noise can greatly affect the health and comfort of users, so the significance of the right assessment of the compliance with the requirements is obvious. But noise level testing is carried out using different methods, which may not ensure consistency in assessments

The paper presents the influence of test methods on measurement results determined based on an analysis of inter-laboratory comparative studies. The analyses presented in the paper apply to an equivalent sound pressure level determined for a permanent source of sound – an air-conditioning device. The test methods were characterised according to their precision. In order to compare them, their compatibility was analysed based on the methodology described in the literature, alongside a single-factor analysis of variance. It was determined that there were no grounds for rejecting the hypothesis about lack of statistical differences between the results obtained via different methods. Each of the methods is characterised by different precision, so consequently the same result obtained with each method carries a different risk in regards to noise assessment.

The reason for taking up this kind of research was the decision of the Polish Technical Committee in 2018 about introducing new acoustic requirements in Poland concerning the admissible indoor sound pressure levels. It was decided to implement new international methods of testing indoor sound pressure levels emanating from the service equipment in the building. It was necessary to show the differences between the current method and its new counterparts.

Keywords: test methods; compatibility; sound pressure level.

1. Introduction

The measurement of noise emitted by service equipment in a building is an acoustic requirement worldwide. This applies to the noise emanating from different pieces of service equipment, including heating and cooling devices, sanitary systems, mechanical ventilation systems, lifts, chutes, pumps, garage doors and other auxiliary devices, with the noise penetrating into the rooms both in residential and public buildings.

The real value of the noise is of stochastic character, and can be interpreted as a system randomly changing over time and in the function of space. Such a physical value is difficult to apply to any assessment so a number of indicators were defined, including e.g.:

- L_{Aeq}, A – weighted equivalent continuous sound pressure level in dB,
- L_{Ceq}, C – weighted equivalent continuous sound pressure level in dB,
- $L_{\mathrm{AS\,max}}$, maximum A – weighted equivalent continuous sound pressure level with time weighting "S" in dB,
- $L_{\mathrm{AF\,max}}$, maximum A – weighted equivalent continuous sound pressure level with time weighting "F" in dB,

which can be compared with specific criteria (ISO-16032, 2004).

The conformity assessment of compliance with the criteria always entails a risk of committing an error while making a decision. Type I error involves accepting the sound pressure level as non-compliant when it does fulfil the requirements. Type II error involves accepting the sound pressure level which does not fulfil the requirements. Type I errors are usually related to an increase in the costs due to raising the parameter subject to the assessment (noise reduction in the reference case). Type II error may affect user safety.

There is also a risk of the general type III error (Trzpiot, 2015) related to asking a wrong question (wrong zero hypothesis) which can cause both of the above-mentioned types of effects. For noise tests, it can be e.g. a wrong test method or wrong test conditions.

The role of uncertainty in the estimation of the risk concerning conformity assessment was discussed in detail in JCGM (JCGM 106:2012, 2012). Depending on the method of considering the uncertainty when determining the acceptance limits based on the tolerance limits, we deal with the rules of guarded acceptance, simple acceptance and guarded rejection. If the technical specifications do not indicate the acceptance principle, the most convenient and most common method is simple acceptance involving a risk division between the customer and the supplier, and establishing the acceptance limit equal to the tolerance limit. If the probability density function (pdf) assigned to uncertainty is known, risk can be estimated for the specific distribution based on probability, if the result is compliant or not compliant with the criterion value.

In the case of noise measurements, issues of measurement uncertainty, pdf shape and the risk assessment are not fully and uniformly solved. The most serious doubts concern the assumption about normal distribution of the noise measurement results. There are papers (e.g. Batko, Stępień, 2014; Przysucha *et al.*, 2015), which discuss the variability of the noise in the outdoor environment and the need to use different uncertainty estimation procedures other than those commonly used by laboratories (JCGM 100:2008, 2008).

Due to doubts as to the shape of the pdf, it is also questionable to use the $k = 2$ coverage factor, assuming approximately 95% coverage probability. Therefore, the authors of this article use rather standard uncertainties as comparable quantities.

For the measurement of noise, depending on the character of the sample, the accurate value of the equivalent sound pressure level can vary significantly from the approximate value, which can lead to incorrect estimation of the uncertainty of the noise indicators. The results of the measurements carried out with different methods may also vary. This poses the risk of incoherent assessments of compliance with the criteria, depending on the method, measurement laboratory or uncertainty of the measurement results.

The notion of uncertainty can be understood in different ways, even in reference to well-defined measurements (Walker *et al.*, 2003). Also, the evaluation of the results of comparative studies can be carried out in various ways (Flores *et al.*, 2018; Jagan, Forbes, 2019; Molenaar *et al.*, 2018).

The issue of the noise measurement uncertainty can follow a multi-level approach. If we take into account the difference between measurement and testing, resulting from the definition in the documents of Joint Committee for Guides in Metrology (JCGM) (JCGM 200:2012 (2008)) and European co-operation for Accreditation (EA) (EA-4/16 G:2003), it can be concluded that the measurement of point (in time and space) "real value" of the sound pressure meets the definition of measurement. Determining the values of the indicators mentioned above using the test procedures contained in various standards should be considered a test, as according to EA: "a test result typically depends on the method and on the specific procedure used to determine the characteristic..." and "In general, different test methods may yield different results...". Measurement methods of noise emitted by service equipment in a building differ in sampling, number of measurements, location of measurement points, etc. The sampling method can be of particular importance as sound pressure varies over time and space.

When estimating the uncertainty of defined above point measurement of the sound pressure, the uncertainty constituents are taken into consideration, resulting from calibration, adjustments and corrections of the equipment used and the variability related to the repeatability of the measurements (Seddeq, Medhat, 2011), usually expressed as a standard deviation for n-measurements. Of course, it is physically impossible to determine the repeatability of a point measurement, because there is only one such measurement. Therefore, an approximation must be applied using measurement results obtained with stable sound sources. Such a standard deviation does not express the repeatability of the measurement *sensu stricto*, because it also includes the variability of the sound source.

The uncertainty resulting from variability in the reproducibility conditions represents another uncertainty level. It results to a great extent from the measurement model, in particular from the number of factors affecting the result, which are beyond the model limits (they are not controlled). For the sound pressure measurement, the variance can be further increased by the variability of the "real" value in time.

A different kind of reasoning applies when estimating the uncertainty of test results. As already mentioned, the noise measurement can be considered as a test (as defined in EA-4/16) consisting of a number of sound pressure measurements. Their number, place of measurement, time of measurement, etc. are defined by the test model. (Later in the article, due to the common use, the authors will use the term "measurement" to refer to noise).

Besides the variability aspect of the "real" noise value one should also take into consideration aspects of uncertainty related to e.g. the background noise or reverberation. Quite high variabilities are related to the

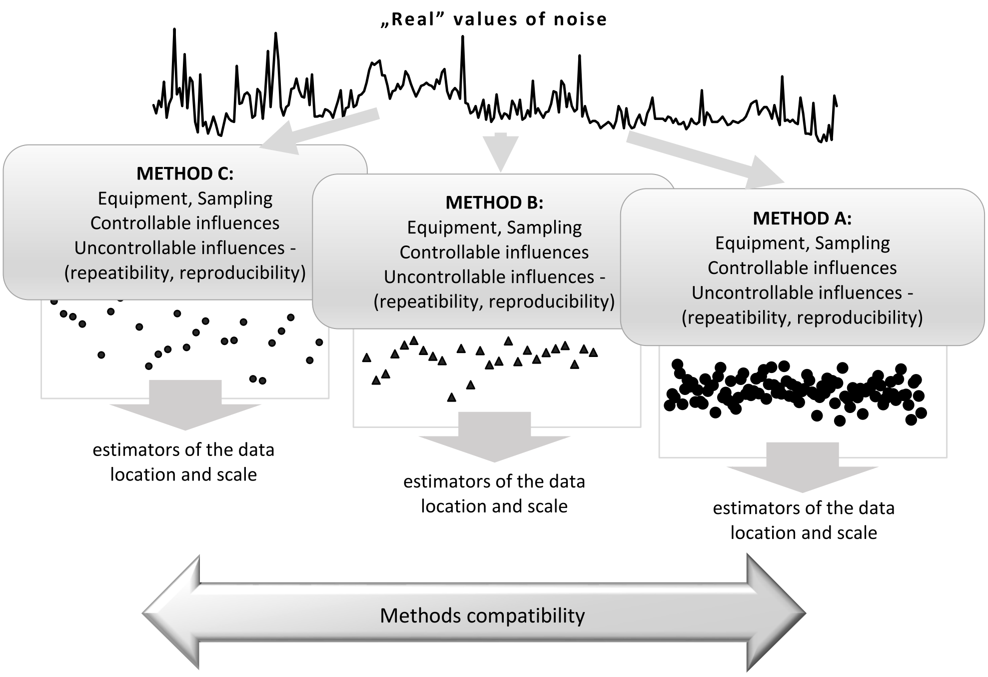

Fig. 1. Illustration of relationships concerning the variability of the results and methods compatibility.

definition of the measured value of noise resulting from the measurement model (method). The definition consists of such statements as: "the measurement time for a specified noise should be between 3 and 5 minutes" or "the measurements shall be performed for three cycles", and also a method of converting the measurement results into the output values. Such differences in the definition of the measured value result in different test results when performed with different methods. A measurement model which ignores many important factors which can affect the result, generates high variability of the results, especially in reproducibility conditions (Walker *et al.*, 2003). The results obtained via different methods and models vary despite the fact that they apply to the same "real" value (Fig. 1). In the case of noise, this is related e.g. to a different sampling method in space and time.

The differences in the measurements carried out by different laboratories as part of the noise tests can apply to e.g. the microphone location, the selection of measurement time, sampling duration and other influences, which are not defined precisely enough as part of the model (Prezelj, Murovec, 2017).

Separating the variability due to the reproducibility of tests within one method from the variability due to the use of different methods is difficult, especially since other components of uncertainty are also involved.

The comparison of the test methods is in practice limited to two aspects: comparability and consistency. According to JCGM 200:2012 (2008), metrological comparability is the *comparability of measurement results, for quantities of a given kind, that are metrologically traceable to the same reference.* If the comparability condition is met, compatibility can be inferred, determining the conditions that have to be met in order to consider the methods as compatible.

Each method is characterised by specific arrangements concerning the equipment used, the sampling method (for noise measurements: time and place), the number of measurements, the method of converting the measurement results into the final result, and non-controllable factors, namely factors which were not precisely determined in the model and may cause variabilities taking the form of differences between the results under repeatability conditions. When measurements are carried out by different laboratories, i.e. under reproducibility conditions, the differences are usually bigger (standard deviation of reproducibility). The minimisation of variability under repeatability and reproducibility conditions should be ensured as part of the measurement method validation, and completed by the organisations which develop the method. Accredited laboratories should have in place the procedures to monitor the quality of their results. This way, the risk of inaccurate assessments is minimised. One of the methods to monitor the repeatability and reproducibility of the results is to organise inter-laboratory comparisons. They are among the key activities aimed at confirming the measurements performed by the laboratories. They also play a major role in the development and validation of the measurement and test methods, the estimation of the levels of repeatability, the reproducibility or uncertainty of the test results, the characterisation of the test method, the assignment of values to the reference materials or the determination of the comparability of different methods in reference to the specific measured value (Czichos *et al.*, 2011).

The paper deals with the issue of the influence of the measurement method of the noise from indoor service equipment in the building on the measurement results based on inter-laboratory comparative studies. Previous inter-laboratory studies mentioned in the literature were mainly devoted to the measurements of sound insulation (Berardi, 2012; Pozzer *et al.*, 2019; Scamoni *et al.*, 2009; Scrosati *et al.*, 2015) or sound absorption (Scrosati *et al.*, 2020). Only a handful of authors have dealt with the issue of comparative studies related to the measurement of noise generated by the service equipment in buildings. Seddeq and Medhat (2011) analysed inter-laboratory studies carried out by five laboratories for the assessment of measurement uncertainty, repeatability and reproducibility. Di Bella *et al.* (2013) and Seddeq and Medhat (2011) described tests carried out using three methods: ISO 16032, ISO 10052 and UNI 11367 with nine participants. They emphasised that the procedure of the corner position selection for assessing the noise emitted by a water system described in ISO 16032 can be the cause of the discrepancy between the results, because it causes differences in the measurement point selection by different laboratories. The authors determined the repeatability and reproducibility resulting from the conducted comparisons and, in the conclusions, they indicated the need for further analysis of the accuracy of the survey methods (like ISO 10052), which contrary to the engineering method (ISO 16032) can reduce the duration of the measurements and results processing considerably.

This paper presents the results of inter-laboratory studies performed with the participation of twelve laboratories. The studies were carried out according to three test standards: the test standard currently valid in Poland (PN-B-02156, 1987), which can be classified as a survey method, and two international standards (ISO-10052, 2004, ISO-16032, 2004). The test results were used for the assessment of variability under reproducibility conditions and for the assessment of the compatibility of the applied test methods. Issues related to the risk of committing an error while making decisions concerning the assessment of the compliance of the results with the noise requirements were also evaluated.

The main purpose of the study was to assess the variability of the results, which can be attributed to the tests carried out in reproducibility conditions, and the variability resulting from the application of different measurement methods. The issue was important for testing the possibility of using survey methods (like ISO 10052), which are less time-consuming than engineering methods (ISO 16032).

2. Standard noise measurements in protected rooms

The Polish standard dating back to the 1980s is still obligatory in Poland in reference to noise measurement (PN-B-02156, 1987). The standard was established for different measurements and building techniques. Polish Technical Committee and polish authorities made a decision to update the requirements concerning noise because they do not correspond either to modern construction standards or to contemporary measurement methods, included in the international standards effective in Poland.

Two European standards (ISO-10052, 2004; ISO-16032, 2004) concerning the measurement of noise penetrating indoors from service equipment installed permanently in buildings, such as heating and cooling devices, sanitary systems, mechanical ventilation systems, lifts, rubbish chutes, pumps, garage doors and other auxiliary equipment should replace the Polish standard (PN-B-02156, 1987) in accordance with which the noise measurements were carried out.

A simplified method of noise measurement given in ISO-10052 (2004) is actually similar to the measurement method according to the standard effective in Poland (PN-B-02156, 1987). According to PN, only the A-weighted mean sound pressure level L_{Am} for a specified noise or the maximum sound pressure level $L_{\mathrm{A,max}}$ (with time weighting “S”) for a noise with unspecified levels are measured indoors. The essential differences in measurement method procedure are summarised in Table 1.

The methods differ in the way they determine the final measurement result. According to PN-B-02156 (1987), a correction resulting from the influence of the background sound (from 0 to −3 dB, depending on the difference between the noise level and the background sound pressure level) should be subtracted from the measurement result (the highest), and for non-furnished rooms a correction resulting from the influence of the sound absorption of the room should also be subtracted. For the assessment of a specific noise, a level not equivalent to the device’s working cycle (as in European standards) should be determined for a time section equal to half an hour at night and eight hours in the day, including any possible breaks in the device’s operation. According to the Polish standard, the sound pressure level can be assessed only when the difference between the measured A-weighted sound pressure level and the level of the background sound is greater than 3 dB. According to ISO standards, if the difference is less than 6 dB, it should only be recorded in the report.

The noise measurement method according to ISO-16032 (2004) is considered to be more accurate than the method according to ISO-10052 (2004). This method is very time-consuming. Tests are carried out

Table 1. Basic differences between the sound level measurement methods specified in presented test methods.

	PN-87/B-02156	ISO 10052:2004	ISO 16032:2004
Frequency range	63–8000 Hz	125–2000 Hz	31.5/63–8000 Hz
General requirements	1) The doors and windows should be closed during the measurement. If air exchange is required through the openings, the measurements should be carried out with open doors and windows. 2) Only two persons operating the sound analyser can stay in the room during the measurement. 3) Any noise sources have to be turned off in the test room. 4) Only the equipment and objects which are the part of the test room equipment can remain in the room.	Carry out the measurements with closed doors and windows and open roller shutters.	Carry out the measurements with closed doors and windows. The person doing the measurements should stay outside the room.
Number and location of measurement points	No less than 3, located as follows: • at the height of 1.2 ± 0.1 m, • 1 m above the service equipment, • 1.5 m from the windows, • 0.5 m from the operator. (see Fig. 2a)	Two set positions: 1) Close to a visible corner, with the hardest acoustic surface. 2) In the diffusion field of the room. Each point at least 1.5 m from the noise source. (see Fig. 2b)	Three set positions: 1) One in the corner of the room, established based on the highest value of the C-weighted sound pressure level, 0.5 m away from the walls and above the floor. 2) Two in the diffusion field of the room. Minimum distance: 1) Between points 1 and 2, and 2 and 3 should be 1.5 m. 2) Between points 2 and 3 should be 0.75 m. Height above the floor: 0.5–1.5 m. (see Figs 2c and 2d)
Measurement time	It should be assumed depending on the character of the noise and it can be shorter than the assessment duration. The measurement time (duration) for an unspecified noise should be selected so that it covers all changes in the sound level characteristic of the situation or noise source. The measurement time for a specified noise should be between 3 and 5 min.	The measurement time in each point should correspond to at least one working cycle of the device. The measurements shall be performed for three cycles.	A single measurement time should correspond to at least one working cycle of the device. The number of the measurements should comply with the methodology given in the standard.

for a higher number of measurement cycles, and the reverberation time in the room under assessment is also measured, while the A- or C-weighted sound pressure levels determined as the final result based on the results of measurements in octave bandwidths may vary for a limited frequency range (63–8000 Hz) from the values determined with a simplified method. Still, the method helps to determine noise occurrence and its onerousness, even for noise with very low A-weighted sound pressure levels (20–25 dB), which cannot be assessed with single-number indicators and have very small differences between the noise and background

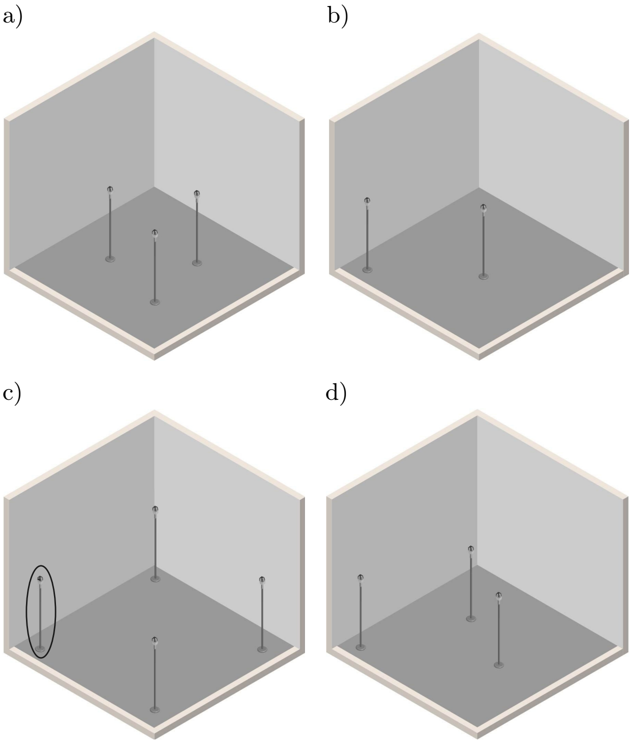

Fig. 2. Location of the measurement points characteristic of each of the presented test methods: a) PN-B-02156, b) ISO 10052, c) and d) ISO 16032 (the point with the highest L_C is marked in Fig. 2c).

sound. It is also possible to determine the character of the noise (e.g. tonal or low-frequency) and to indicate the most onerous constituents.

3. Statistical methods used in the analysis of the results

The majority of statistical analyses referring to random dispersion of the test results use an assumption that random variable has a normal distribution. The literature data concerning outdoor noise reveal that long-term noise indicators do not have a normal distribution (Batko, Stępień, 2014; Przysucha *et al.*, 2020; Wszolek, 2006). Conversely, it must be accepted that the random variable distribution in the presented cases is related to the special character of the traffic noise variability. In studies of indoor noise, the variability of the sound pressure level in time is much lower than for traffic noise. The conditions in this study ensured that the noise was coming from a relatively stable source (noise established according to PN-B-02151-02, 1987) in reference to the sources typical of outdoor traffic noise. The variability of the "real" sound pressure level is then lower and the other variabilities related to the measurement uncertainty, e.g. from the equipment, repeatability and reproducibility due to factors beyond the measurement model limits will start to play a greater role than in outdoor noise measurements. It can be roughly assumed that a combination of several pdfs can finally render a distribution similar to the normal one pursuant to the central limit theorem. Despite the fact that the determination of the results distributions was not the purpose of this paper, they were checked for normal distribution with a Shapiro-Wilk test, which is characterised by a relatively high power.

In order to detect a single deviating value in a data set obtained by the laboratories using the same test method, a Grubbs test was carried out. The values of the reproducibility standard deviations were obtained according to the methodology presented in ISO 5725-2 (1994).

The recommended assessment methods of the results obtained by each laboratory as part of the interlaboratory studies (ISO 13528, 2015) can be divided into three groups. The first includes comparisons in which the result value is the assigned value and its uncertainty is known to the organisers from other sources (e.g. when the sample includes certified reference materials). The second group covers tests in which the assigned value is determined based on the measurement results from reference laboratories with better measurement capabilities. The third includes a situation in which the results from all comparison participants are used to determine the assigned value. In the case of the presented comparisons, we clearly lack knowledge about the assigned value, and reference laboratories. All participating laboratories are accredited and their equipment complies with the requirements so there are no grounds to discriminate for or against any of them. In this case, the determination of the assigned value should cover many laboratories. With twelve participating laboratories the parameters used for the assessment based on the assigned value and its uncertainty, determined by the participants' results, can be of low reliability and may render false positive results (ISO 13528, 2015; Szewczak, Bondarzewski, 2016). Greater certainty can be ensured by employing robust statistics to determine the estimators of data location and data scale for the assigned value. The following statistics were used in this paper: MED_i is median of the results for the i method as a scale estimator for the i method, and $\sigma_{\mathrm{MAD}i}$ is median absolute deviation for the i method as a scale estimator for the i method (Daszykowski *et al.*, 2007). The results in each method were also assessed using the algorithm A with iterated scale (ISO 13528, 2015).

The measurements performed by all laboratories as part of all methods apply to the same physical value: sound pressure level. They were carried out in each laboratory with the same its own equipment (metrologically traceable to the same reference) so the methods used in the presented tests were considered to have met the condition of comparability of the measurement results.

The authors applied the concept of metrological compatibility presented by Kacker *et al.* (2010) to a situation in which different methods are used for measuring the same measurands (Kessel *et al.*, 2011).

For the results of tests obtained with different i and j methods to be regarded as compatible, the following condition has to be met for every pair of results:

$$\zeta\left(x_i - x_j\right) = \frac{\left|x_i - x_j\right|}{a^*} < \kappa, \tag{1}$$

where

$$a^* = \sqrt{u^2\left(x_i\right) + u^2\left(x_j\right) - 2r(x_i, x_j)u(x_i)u(x_i)}$$

and x_i, x_j are results of measurement's obtained by i and j methods, with their standard uncertainties; $u(x_i)u(x_j)$ are standard uncertainty for i, j method; $r(x_i, x_j)$ is symbol of correlation coefficient $R(X_i, X_j)$ between the variables X_i and X_j represented by results x_i, x_j; κ is resolution. $\kappa = 2$ is typically assumed for normal distribution of the results and a consequent confidence level of ca. 0.05

The variables X_i and X_j can be considered as uncorrelated. The measurements were made independently for each method. The results obtained with the method i were not taken into account in the measurements with the method j. The selection of measurement points was made again at each test. Equation (1) takes the following approximate form:

$$\zeta\left(x_i - x_j\right) = \frac{\left|x_i - x_j\right|}{\sqrt{u^2(x_i) + u^2(x_j)}} < \kappa. \tag{2}$$

The standard uncertainty $u(x)$ of the results obtained with a given method has many components. In further considerations on the compatibility of test results obtained with different methods, some simplifications regarding the value of standard uncertainty can be accepted. Usually, the greatest role in the uncertainty value of the tests performed by different laboratories is played by the dispersion of results due to reproducibility conditions, expressed as the reproducibility standard deviation. In the discussed tests, the uncertainty component resulting from reproducibility was for example over 10 times greater than the component resulting from the measurement uncertainty of the devices used. The differences in the test results obtained with different methods reflect the differences between the methods regarding the influences under the control of the method. The differences in the uncertainty component due to the reproducibility conditions illustrate the differences between the methods regarding the influences beyond the control of the method, so taking only this predominant component into account when comparing test methods may be justified. Equation (2) can then be converted to the form:

$$\zeta\left(x_i - x_j\right) = \frac{\left|x_i - x_j\right|}{\sqrt{s_{Ri}^2 + s_{Rj}^2}} < \kappa, \tag{3}$$

where s_{Ri}, s_{Rj} are standard deviations of the reproducibility for i and for j method.

Reducing the value of the real standard uncertainty to one component increases the value of $\zeta\left(x_i - x_j\right)$ hence it is more difficult to meet the condition $\zeta\left(x_i - x_j\right) < \kappa$.

The estimators of data location (like mean or median) and of data scale (like standard deviation or median absolute deviation) used in the assessment have mathematically the same value, regardless of pdf. Only on the stage of the confidence level determination and the related coverage factor, does the character of the distribution begin to play a role, and that is why the criterion value κ was not assumed in this paper.

With regards to a fairly limited number of results, in order to evaluate the compatibility of the methods, the robust estimators of data location and scale were also taken into consideration. Hence, the equation becomes:

$$\zeta\left(y_i - y_j\right) = \frac{\left|y_i - y_j\right|}{\sqrt{s_{ci}^2 + s_{cj}^2}} < \kappa, \tag{4}$$

where y_i, y_j are location estimators which are MED_i; MED_j is median of the results obtained by all laboratories for the i and j method; s_{ci}, s_{cj} are scale estimators which are $\sigma_{\mathrm{MAD}i}$; $\sigma_{\mathrm{MAD}j}$ is median absolute deviation for the i and j method.

An analysis of the influence of the factors on the measurement results can be performed by means of the methods of the analysis of variance (ANOVA). Their application depends on the conditions concerning normal distribution, independence of the tests and equality of the variance in the tests. ANOVA should be then preceded e.g. with a Shapiro-Wilk test and a Levene test (NIST/SEMATECH, 2013), as performed in this paper.

The selection of the measurement points by the laboratory can be among the factors affecting the results. The influence of the organiser's indication of the measurement points on the results was evaluated in the experiment, for which a single-factor analysis of variance was also used.

4. Experimental work

The noise measurements were carried out in an office of ca. 47 m^2 floor area (a conference room) in a multi-storey building. The air-conditioning system installed in the room was the subject of the tests. The noise was emitted by an air-handling unit (50 kW) and ice-water generator installed on the roof of the building.

Fig. 3. View of the office where the tests were carried out.

Fig. 4. View of the air-conditioning outlet.

Each research laboratory taking part in the comparative studies carried out four stages of tests summarised in Table 2. The measurements were repeated three times. Each laboratory was measuring the A-weighted sound pressure level:

- L_{Am} according to PN-87/B-02156, or
- L_{AeqT} according to PN-EN ISO 10052 and PN-EN ISO 16032.

Table 2. Description of the testing cycle divided into four measurement stages.

Method	Measurements	According to
1	The sound pressure level is measured three times according to laboratory practice. *Note:* the positions are determined independently for each measurement.	PN-EN ISO 10052
2	The sound pressure level parameter is measured three times at the points indicated by the organiser.	PN-EN ISO 10052
3	The sound pressure level parameter is measured according to the standard, in compliance with laboratory practice.	PN-87/B-02156
4		PN-EN ISO 16032

Stages 1 and 2 were intended to assess variability due to the selection of the measurement points. In order to achieve this goal, it was important to follow the test organiser's instructions. Stages 3 and 4 were the actual inter-laboratory studies meant for the assessment of the expertise of each measurement team, the correct implementation of the test methods and analyses of the influence of the test method on the obtained results of indoor sound pressure level.

Table 3 presents the results obtained by each laboratory.

Table 3. Summary of the sound pressure level measurement results obtained by laboratories using four test methods.

Laboratory	Method 1	Method 2	Method 3	Method 4
	L_{Aeq} [dB]			
Lab 1	38.80	39.13	36.10	38.90
Lab 2	38.90	40.47	40.90	40.90
Lab 3	37.07	36.77	33.50	30.70**
Lab 4	36.50	36.27	36.40	36.90
Lab 5	38.55	38.30	38.00	36.72
Lab 6	41.47*	39.93	40.70	39.10
Lab 7	37.00	36.33	35.80	38.81
Lab 8	36.60	36.20	37.70	35.70
Lab 9	37.58	37.30	36.40	35.80
Lab 10	37.53	37.20	41.50*	–
Lab 11	39.46	39.17	36.50	–
Lab 12	–	42.40*	35.00	–
Mean value (x_i)	38.13	38.29	37.38	37.06

* Warning signal according to criteria referring to parameter z (Eq. (8)).

** Action signal according to criteria referring to parameter z (Eq. (8)).

5. Results and discussion

The Shapiro-Wilk test of normality was carried out for the given methods to confirm or reject the hypothesis about the normal distribution of the results. The test results are summarised in Table 4.

Table 4. The Shapiro-Wilk test results ("+" means that $W > W_{kr}$ (there are no grounds to reject the hypothesis on normal distribution); "–" means that $W < W_{kr}$ (the hypothesis about normal distribution can be rejected); W is value of the Shapiro-Wilk statistics; W_{kr} is critical value of the distribution for $\alpha = 0.05$).

Measurement method	Shapiro-Wilk test result at the confidence level $\alpha = 0.05$
Method 1	+
Method 2	+
Method 3	+
Method 4, final result	+
63 Hz	–
125 Hz	–
250 Hz	+
500 Hz	–
1000 Hz	+
2000 Hz	+
4000 Hz	+
8000 Hz	+

The results of the Shapiro-Wilk test proved that there were no grounds to reject the hypothesis on the normal distribution of the results of sound pressure level L_{Aeq} obtained with each of the methods 1–4. Only the results for the frequency values of 63, 125 and 500 Hz did not confirm the zero hypothesis. Further deliberations did not include using the tests for the results in each bandwidth in method 4.

The Grubbs test, applied to the results shown in Table 3 in order to detect a single deviating value, did not reveal any deviating values.

The values of the statistics were determined for each method: MED_i is median of the results for the i method, and $\sigma_{\mathrm{MAD}i}$ is median absolute deviation for the i method. Using the A-weighted iterative algorithm (ISO 13528, 2015), the values of the location estimator of the results, x^*, and scale estimator s^*, for each method results were determined. The results are summarised in Table 5. In order to assess the convergence of the results obtained by the tests participants for each method, a typical factor-based criterion was applied, but using robust location and scale estimators obtained according to algorithm A:

$$z_{i,k} = \frac{|x_{i,k} - x_i^*|}{s_i^*}, \tag{5}$$

where $x_{i,k}$ is result obtained with the i method by the k participant; x_i^* is location estimator for the i method; s_i^* is scale estimator for the i method.

The typical criteria referring to parameter z, used in inter-laboratory studies, include:

- $z \leq 2.0$ – results of laboratory acceptable,
- $2.0 < z < 3.0$ – warning signal for laboratory,
- $z \geq 3$ – action signal for laboratory.

The criteria above result from the assumption of normal distribution of the population of the results. Warning signals (*) and action signals (**) are marked in Table 3 for illustrative purposes.

The standard deviations of repeatability s_r and reproducibility s_R were estimated for methods 1, 2 and 4 in each bandwidth (ISO 5725-2, 1994), taking into account the series of three results for each laboratory. Single results were obtained in method 3 and in reference to the final result of method 4, while the standard deviation for a set of single results from all laboratories was adopted as a standard deviation of reproducibility. Table 5 shows the results for repeatability and reproducibility and the coefficient of variation for each method.

The results obtained with methods 1 and 2 reveal a lower value of the reproducibility variance than those obtained with methods 3 and 4. The measurement uncertainty level of the values of the repeatability standard deviation for method 1 result from equipment uncertainty (the typical value of the standard uncertainty assigned to testing equipment amounts to ca. 0.2 dB according to the authors' experience).

Table 6 summarises the repeatability and reproducibility standard deviations for each bandwidth (method 4).

The character of repeatability and reproducibility depends on the frequency, which can be observed in

Table 5. Statistical parameters obtained for the value of the equivalent sound pressure level L_{Aeq} for each test method.

Estimator	Method 1	Method 2	Method 3	Method 4
	[dB]			
MED_i – median	37.58	37.80	36.45	36.90
$\sigma_{\mathrm{MAD}i}$ – median absolute deviation	1.45	2.10	2.00	2.83
x_i^* – location estimator for i method, (algorithm A)	37.98	38.17	37.02	37.38
s_i^* – scale estimator for i method (algorithm A)	1.32	1.95	2.05	2.55
s_{ri} – repeatability standard deviation	0.20	0.73	–	–
s_{Ri} – reproducibility standard deviation	1.53	2.06	2.49	2.94
Coefficient of variation s_{Ri}/x_i	4.0%	4.3%	5.6%	9.0%

Table 6. Repeatability and reproducibility of the sound pressure measurements for each bandwidth obtained with method 4.

	Frequency [Hz]							
	63	125	250	500	1000	2000	4000	8000
Scale estimator [dB]								
s_r (x_4) repeatability standard deviation	4.46	1.72	1.27	0.96	2.12	1.62	2.94	2.81
$s_R(x_4)$ reproducibility standard deviation	4.85	2.85	2.62	2.85	3.29	3.27	4.33	3.86
Scale estimator [%]								
Coefficient of variation $s_R(x_4)/x_4$	9.04	6.45	6.72	8.13	10.16	13.44	24.76	25.09

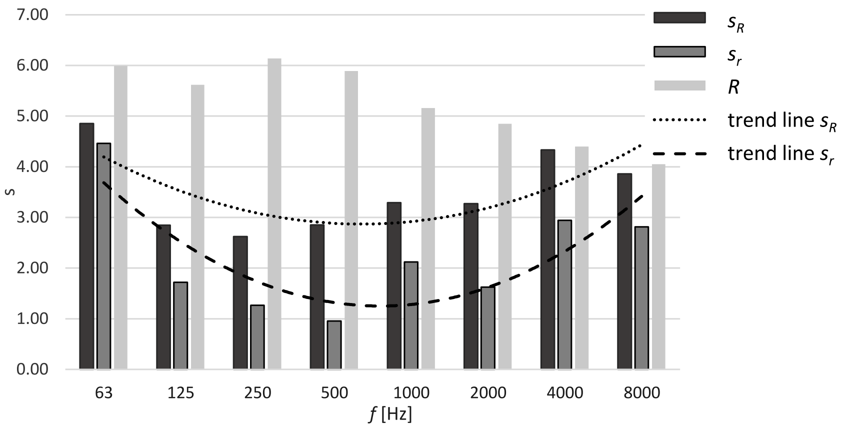

Fig. 5. Relationship of the repeatability (s_r) and reproducibility (s_R) standard deviation and frequency. For comparison, the figure also includes the values of R – reproducibility standard deviation for each bandwidth presented in the literature (Seddeq, Medhat, 2011). Dashed lines plot the trend lines for the best matching $s_r(f)$ and $s_R(f)$ results. They are the square curves.

the studies carried out by the authors of this paper, and does not fully correspond to the observations presented by Seddeq and Medhat (2011), Fig. 5. With five laboratories using the same test method, Seddeq and Medhat obtained higher values of the reproducibility standard deviation. The trend of the changes depending on the frequency also seems different. The best matching (the highest value of Pearson's coefficient of correlation amounting to 0.81 for s_r and 0.74 for s_R) was achieved for the trend line with the square polynomial equation and the minimum in the 500–1000 Hz area. The results presented by Seddeq and Medhat (2011) did not reveal such a trend.

In order to assess the influence of the method for testing indoor noise from service equipment in buildings, the analyses of the test method compatibility were carried out using formulas (3) and (4). The results of the methods compatibility assessment are shown below. The values of ζ for each method pair are given in the cells.

All methods rendered compatible results at the level of threshold $\kappa \ll 1$, according to the definition given in the literature (Kacker *et al.*, 2010), both when the standard mean of the results (Eq. (3)) and robust statistics (Eq. (4)) were used for the assessment.

The results of the Levene test and single-factor analysis of variance are shown in Table 8. The Levene

Table 7. Assessment of the compatibility of the methods.

Type of statistics	Method	Method 2	Method 3	Method 4
$\zeta\,(x_i - x_j)$ (Eq. (3))	Method 1	0.07	0.29	0.32
$\zeta\,(Y_i - Y_j)$ (Eq. (4))		0.09	0.46	0.21
$\zeta\,(x_i - x_j)$ (Eq. (3))	Method 2		0.34	0.36
$\zeta\,(Y_i - Y_j)$ (Eq. (4))			0.47	0.26
$\zeta\,(x_i - x_j)$ (Eq. (3))	Method 3			0.09
$\zeta\,(Y_i - Y_j)$ (Eq. (4))				0.13

Table 8. Results of the Levene test (NIST/SEMATECH, 2013) and single-factor ANOVA (Microsoft Excel 2013/Data Analysis ToolPack). W – value of Levene statistics, F – value of F-statistics in ANOVA test, F_{kr} (α, $k-1$, $N-k$) – critical value of F-statistics at the level of significance α, $k-1$ and $N-k$ degrees of freedom, k – number of groups, N – total number of measurements taken into account.

Considered results	Levene test	Single-factor ANOVA
Methods 1–4	$W = 1.22 < F_{kr}$ (0.05, 3, 40)	$F = 0.73 < F_{kr}$ (0.05, 3, 40)
Methods 1–4 with rejected results considered as warning signals and action signals	$W = 1.76 < F_{kr}$ (0.05, 3, 36)	$F = 0.66 < F_{kr}$ (0.05, 3, 36)
Methods 1–2	$W = 2.79 < F_{kr}$ (0.05, 1, 21)	$F = 0.05 < F_{kr}$ (0.05, 1, 21)
Methods 1–2 with rejected results considered as warning signals and action signals	$W = 4.06 < F_{kr}$ (0.05, 1, 19)	$F = 0.04 < F_{kr}$ (0.05, 1, 19)

test for all series of results presented in Table 3 and for methods 1 and 2, revealed that the hypothesis assuming that the variances are homogeneous cannot be rejected. A single-factor ANOVA revealed that a hypothesis about lack of test method influence on the obtained results cannot be rejected.

Taking into account the fact that too high variances inside groups may result in type II errors, both in the Levene test and in the ANOVA, tests were also carried out for data rejecting the results considered as warning signals and action signals. In this case, the results also confirmed that the hypothesis about lack of test method influence cannot be rejected.

The compatibility of the results obtained with all four methods shows that all seem adequate for the assessment. Still, each method has a different level of precision. Consequently, the same result obtained with each method poses a different risk of making a wrong decision in relation to noise assessment. If a working assumption is made, according to the previous analysis, that the results dispersions are characterised by a normal distribution, the probability of wrong assessment can be estimated based on the knowledge concerning uncertainty. The value of 40 dB was adopted as the upper limit (T_u) in the analyses, which is the maximum acceptable sound pressure level value in offices according to the requirements (PN-B-02151-02, 1987).

The results of the analyses revealed that the methods were compatible, which means that they render statistically equivalent results but the specificity of each method affects the assessment risk level, as shown in Table 9. Each method is characterised by different precision, so the same result obtained with each method can pose a different risk of making a wrong decision related to noise assessment.

Table 9. Probability values of right assessment p_c and wrong assessment $\overline{p}_c$, based on the results shown in Table 3.

Method	$\frac{T_U - x_i}{s_{Ri}}$	p_c	$\overline{p}_c$
1	1.22	0.89	0.11
2	1.04	0.85	0.15
3	1.05	0.85	0.15
4	1.00	0.84	0.16

The differences between methods related to the risk of incorrect evaluation can be better illustrated if we hypothetically assume that all laboratories obtained the same result (in this case 38 dB was assumed), as shown in Table 10.

The selection of the measurement point can contribute to the variability of the test result. The sound pressure level in the chosen point is determined mainly by the noise source. However, it shall be noted, that the sound pressure level is influenced also by reverberation time and background noise.

Table 10. Probability values of the right assessment p_c and wrong assessment $\bar{p}_c$, based on the same sample sound pressure level result of x = 38 dB assumed for all methods and s_{Ri} values presented in Table 5.

Method	$\frac{T_U - x}{s_{Ri}}$, x = 38 dB	p_c	$\bar{p}_c$
1	1.31	0.90	0.10
2	1.22	0.89	0.11
3	0.80	0.79	0.21
4	0.68	0.75	0.25

In method 1, the participant chose the measurement points himself, in method 2 (according to the same standard), the organizer set the measurement points. It turns out that this does not cause a significant change both in the result and in the coefficient of variation. In method 4, the researcher selects two points, and in method 3 three points in the diffusion field. The precision of these methods varies significantly, which also affects variability and the risk of wrong assessment.

Based on the estimated precision of the methods, ISO 16032 (method 4) turned out to be the one with the lowest precision under relatively stable noise conditions. The coefficient of variation for this method amounted to 9% in the inter-laboratory experiment. Consequently, it entails a greater risk of sound pressure level assessment. The coefficients for other methods ranged from 4 to 5.6%. It seems that it might be related to the procedure of the measurement points selection.

One should take into account that this experiment was carried out in conditions of relative stability of measurand in time and space. High variability of noise in time and space (different points in the room) could demonstrate that method 4 is more relevant, as it provides measurements for a greater number of points and in all bandwidths. On the other hand, the large coefficient of variation based on the reproducibility standard deviation indicates that the method still has too many influences beyond the control such as the procedure of selecting the measurement points which allows different interpretations by laboratories.

The survey method defines quite precisely the location of the measurement points: close to a visible corner with a hardest acoustic surface and in the diffusion field of the room, but not closer than 1.5 m from the noise source. Even if the description is not very clear and precise, in fact most measuring teams has chosen similar locations during the measuring process. It might influence the lower variation of the test results. The engineering method is much more difficult and require larger amount of measuring points. The measuring points selection method seems to be more accurate, but in fact, measuring teams had a tendency to choose different points in the diffuse field. It might cause the higher variation of the test results. Further studies on the subject will be carried out.

6. Summary and conclusions

The paper presents the results of noise tests obtained by twelve laboratories with the use of four tests methods including survey methods (like ISO 10052), and engineering methods (ISO 16032). Statistical analyses of the results were carried out, including: Shapiro-Wilk test of normality, analysis of reproducibility variance for all methods, assessment of compatibility of the methods, Levene test of homogeneity of variances and ANOVA analysis concerning influence of the test method on the obtained results of indoor noise.

The results of the Shapiro-Wilk test proved that there were no grounds to reject the hypothesis on the normal distribution of the sound pressure level L_{Aeq} results obtained with each of the four methods. The assumption regarding the normality of distributions allowed for the evaluation of the laboratories results using the modified "z" parameter (standard score). The test showed that the criteria for an interlaboratory comparison were met. Two results differed from the others at the warning signal level.

All methods rendered compatible results at the level of threshold $\kappa \ll 1$, according to the definition given in the literature.

The results of Levene's test showed that the homogeneity of variance hypothesis could not be rejected. A single-factor ANOVA revealed that a hypothesis about lack of test method influence on the obtained results cannot be rejected.

The compatibility of the results obtained with all four methods shows that all seem adequate for the assessment. But based on the estimated precision of the methods, ISO 16032 (method 4) turned out to be the one with the lowest precision under relatively stable noise conditions. The coefficient of variation for this method amounted to 9% in the inter-laboratory experiment. Consequently, it entails a greater risk of noise assessment. The coefficients for other methods ranged from 4 to 5.6%. This is most probably related to the procedure of location of the measurement point. Further studies on the subject will be carried out.

References

1. Batko W.M., Stępień B. (2014), Type a standard uncertainty of long-term noise indicators, *Archives of Acoustics*, **39**(1): 25–36, doi: 10.2478/aoa-2014-0004.
2. Berardi U. (2012), A comparison of measurement standard methods for the sound insulation of building façades, *Building Acoustics*, **19**: 267–282, doi: 10.1260/1351-010X.19.4.267.
3. Czichos H., Saito T., Smith L. (2011), *Springer Handbook of Metrology and Testing*, Springer Berlin–Heidelberg, doi: 10.1007/978-3-642-16641-9.
4. Daszykowski M., Kaczmarek K., Vander Heyden Y., Walczak B. (2007), Robust statistics in data analysis – A review: basic concepts, *Chemometrics and Intelligent Laboratory Systems*, **85**: 203–219, doi: 10.1016/J.CHEMOLAB.2006.06.016.
5. Di Bella A., Pontarollo C.M., Granzotto N., Remigi F. (2013), Interlaboratory test for field evaluation of noise from equipment in residential buildings, [in:] *AIA-DAGA 2013 Merano*, Merano, pp. 1880–1883.
6. EA-4/16 G:2003 (2003), EA guidelines on the expression of uncertainty in quantitative testing, EA, https://european-accreditation.org/publications/ea-4-16-g/ (retrieved 18.01.2021).
7. Flores M., Fernández-Casal R., Naya S., Tarrío-Saavedra J., Bossano R. (2018), ILS: An R package for statistical analysis in interlaboratory studies, *Chemometrics and Intelligent Laboratory Systems*, **181**: 11–20, doi: 10.1016/j.chemolab.2018.07.013.
8. ISO-10052 (2004), *Acoustics – Field measurements of airborne and impact sound insulation and of service equipment sound – Survey method.*
9. ISO-16032 (2004), *Acoustics – Measurement of sound pressure level from service equipment in buildings – Engineering method.*
10. ISO 13528 (2015), *Statistical methods for use in proficiency testing by interlaboratory comparison.*
11. ISO 5725-2 (1994), *Accuracy (trueness and precision) of measurement methods and results – Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method.*
12. Jagan K., Forbes A.B. (2019), Assessing interlaboratory comparison data adjustment procedures, *International Journal of Metrology and Quality Engineering*, **10**: 1–8, doi: 10.1051/ijmqe/2019003.
13. JCGM 100:2008 (2008), *Evaluation of measurement data – Guide to the expression of uncertainty in measurement*, JCGM.
14. JCGM 106:2012 (2012), *Evaluation of measurement data: The role of measurement uncertainty in conformity assessment*, JCGM.
15. JCGM 200:2012 (2008), *International vocabulary of metrology – Basic and general concepts and associated terms (VIM)*, 3rd ed., JCGM.
16. Kacker R.N., Kessel R., Sommer K.D. (2010), Assessing differences between results determined according to the guide to the expression of uncertainty in measurement, *Journal of Resarch of the National Institute of Standards and Technology*, **115**: 453–459, doi: 10.6028/jres.115.031
17. Kessel R., Kacker R.N., Sommer K.D. (2011), Combining results from multiple evaluations of the same measurand, *Journal of Resarch of the National Institute of Standards and Technology*, **116**: 809–820, doi: 10.6028/jres.116.023
18. Molenaar J., Cofino W.P., Torfs P.J.J.F. (2018), Efficient and robust analysis of interlaboratory studies, *Chemometrics and Intelligent Laboratory Systems*, **175**: 65–73, doi: 10.1016/j.chemolab.2018.01.003

19. NIST/SEMATECH (2013), e-Handbook of Statistical Methods, Ch. 1.3.5.10, http://www.itl.nist.gov/div898/handbook/ (retrieved 12.08.2020).
20. PN-B-02151-02 (1987), *Building acoustics – Noise protection of apartments in buildings – Permissible values of sound level in apartments* [in Polish: *Akustyka budowlana – Ochrona przed hałasem pomieszczeń w budynkach – Dopuszczalne wartości poziomu dźwięku w pomieszczeniach*].
21. PN-B-02156 (1987), *Building acoustics – Methods for measurement of sound power of A-level in buildings* [in Polish: *Akustyka budowlana – Metody pomiaru poziomu dźwięku A w budynkach*].
22. Pozzer T., Wunderlich P., Monteneiro C., de Frias J. (2019), Interlaboratory and proficiency tests for field measurements in Brazil, [in:] *INTER-NOISE and NOISE-CON Congress and Conference Proceedings,* Vol. 259, No. 1, pp. 8120–8130, Institute of Noise Control Engineering, http://www.sea-acustica.es/fileadmin/INTERNOISE_2019/Fchrs/Proceedings/2200.pdf.
23. Prezelj J., Murovec J. (2017), Traffic noise modelling and measurement: Inter-laboratory comparison, *Applied Acoustics*, **127**: 160–168, doi: 10.1016/j.apacoust.2017.06.010.
24. Przysucha B., Batko W., Szeląg A. (2015), Analysis of the accuracy of uncertainty noise measurement, *Archives of Acoustics*, **40**(2): 183–189, doi: 10.1515/aoa-2015-0020.
25. Przysucha B., Szeląg A., Pawlik P. (2020), Probability distributions of one-day noise indicators in the process of the type A uncertainty evaluation of long-term noise indicators, *Applied Acoustics*, **161**: 107158, doi: 10.1016/j.apacoust.2019.107158.
26. Scamoni F. *et al.* (2009), Repeatability and reproducibility of field measurements in buildings, [in:] *Proceedings of 8th European Conference on Noise Control 2009, EuroNoise09*, Edinburgh, Scotland, UK, 26–28 October, 2009.
27. Scrosati C. *et al.* (2015), Uncertainty of faqade sound insulation measurements obtained by a round robin test: The influence of the low frequencies extension, [in:] *Proceedings of the 22nd International Congress on Sound and Vibration (ICSV22)*, Florence, Italy, pp. 12–16.
28. Scrosati C. *et al.* (2020), Towards more reliable measurements of sound absorption coefficient in reverberation rooms: An Inter-Laboratory Test, *Applied Acoustics*, **165**: 107298, doi: 10.1016/j.apacoust.2020.107298
29. Seddeq H.S., Medhat A.A. (2011), Indoor noise measurements evaluations for HVAC-Unit using interlaboratory comparisons, *International Journal of Metrology and Quality Engineering*, **2**(2): 75–81, doi: 10.1051/ijmqe/2011104
30. Szewczak E., Bondarzewski A. (2016), Is the assessment of interlaboratory comparison results for a small number of tests and limited number of participants reliable and rational?, *Accreditation and Quality Assurance*, **21**(2): 91–100, doi: 10.1007/s00769-016-1195-y.
31. Trzpiot G. (2015), Some remarks of type III error for directional two-tailed test, *Studia Ekonomiczne. Zeszyty NaukoweUniwersytetu Ekonomicznego w Katowicach*, **219**: 5–16.
32. Walker W.E. *et al.* (2003), Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support, *Integrated Assessment*, **4**(3): 5–17, doi: 10.1076/iaij.4.1.5.16466
33. Wszolek T. (2006), Effect of traffic noise statistical distribution on $L_{\mathrm{Aeq},T}$ measurement uncertainty, *Archives of Acoustics*, **31**(3): 311–318.