Joint International IMEKO TC1+ TC7+ TC13 Symposium
August 31st – September 2nd, 2011, Jena, Germany
urn:nbn:de:gbv:ilm1-2011imeko:2

LOUDNESS MEASUREMENT BY ROBUST MAGNITUDE ESTIMATION

Francesco Crenna, Giovanni Battista Rossi, Luca Bovio
University of Genova, DIMEC, Measurement Laboratory, Genova, Italy

Abstract – A new procedure for the measurement of loudness (the perceived intensity of a sound) by persons is presented. It is based on the combination of two procedures, one aiming at quantifying differences, the other at assessing ratios. The result is expressed in a ratio scale, and inter-individual differences may be evaluated and accounted for. The application to sounds obtained in a measurement campaign in a port environment are presented and discussed. Their importance in ergonomic studies is also mentioned.

Keywords Loudness, perceptual measurement, robustness.

1. INTRODUCTION

The measurement of perceived features such as the quality of a device or the intensity of sound or noise (loudness) is of great importance in the user-oriented development of products and services. Such measurements are critical from a theoretical and metrological standpoints and research in this area is needed [1-2]. Particularly critical is the achievement of ratio scale, when an empirical addition operation is not present, as often happens for perceptual features [3-6]. On the experimental side, it is particular important to achieve reliable results through robust procedures [7].

In this regard, we have recently proposed a new measurement method, that have called robust magnitude estimation, based on combining two different and somewhat complementary scaling procedures, one aimed at assessing differences, the other yielding a ratio representation [8-9]. Theory predicts that whenever there is a considerable agreement between these two results, a ratio scale may be obtained [5,6,9]. In this paper we investigate this subject on the experimental side by presenting and discussing the result of an experiment consisting in the measurement of loudness in a port environment. These measurements are needed for ensure a safe and comfortable working environment, according to ergonomic criteria [10-12].

The experiment has to provide data for the intensive structure in order to enable the construction of the ratio/interval scale and the derived measure. So we set up two distinct and independent tests: an interval estimation test and a magnitude estimation test, to be carried out by each subject. The two procedures are very different: in the former the subject evaluates at the same time all the stimuli, while in the latter he/she deals with a couple of stimuli at time.

The stimuli presented in both tests are the same and include seven pink-noise records and four real noises, recorded on a truck. The reproduction takes place in an acoustically controlled room by means of a flat loudspeakers system. Acoustic pressure has been adjusted to a comfortable level for the listener, to prevent fatigue or annoyance, and the time history of the pressure level at the listener position has been measured with a calibrated measurement system. By processing these signals it is possible to compute sound pressure level (SPL), A-weighted SPL, and loudness, according to ISO 532 [13], for each stimulus. Each subject performs both the interval and the magnitude estimation test. Results are then processed, yielding a measure of the perceived intensity of each sound for each subject. Since at the moment the jury consists of more than thirty subjects it is also possible to evaluate the average measure over the entire group of people.

This experiment has shown a good behaviour of the proposed approach that yields a measure of the perceived intensity compatible with loudness-measurement results according to the standard in use [13]. This confirms the possibility of obtaining a metrologically validated ratio/interval measurement scale for this kind of perceptual measurement.

In the paper, after a concise presentation of the proposed procedure [8,9], we present the experiment and discuss in detail both the experimental set up and the data processing. Final results are then critically presented, for single subjects and for the average group of people.

2. BASIC THEORY

A general problem in perceptual measurement is the measurement of the intensity of a sensation [2]. This is expected to be expressed on a ratio scale since sensation ratios in practice do make sense. Yet, this ratio scale cannot be attained through an empirical addition operation, as it happens with extensive structures [3-5]. So it is important to consider representations not based on an empirical operation of sum [5-6]. This may be done when we have two empirical relations, of difference and of ratio respectively, and they are “in agreement”, in a way that may be precisely specified [6]. More formally, we define an *intensive structure* as a triple $S_I = (A, \succcurlyeq_d, \succcurlyeq_r)$, where A is a set of objects and $\succcurlyeq_d$ and $\succcurlyeq_r$ are weak order relations among pairs of objects, referring, respectively, to *difference* and to *ratio*. If these two, distinct, orderings exist and if they satisfy some compatibility conditions, then it is possible to find a measure function, $m: A \to \mathbb{R}$, such that the following representations contemporarily holds true:

$$\Delta_{ab} \succcurlyeq_d \Delta_{cd} \Leftrightarrow m(a)-m(b) \geq m(c)-m(d), \quad (1)$$

$$a/b \succcurlyeq_r c/d \Leftrightarrow \frac{m(a)}{m(b)} \geq \frac{m(c)}{m(d)}, \quad (2)$$

where Δ_{uv} denotes the empirical difference between u and v, and u/v denotes their empirical ratio (not to be confused with the numerical ratio of the related measures, here denoted by the horizontal line). It is possible to prove that such a measure is on a ratio scale, viz. it safely undergoes similarity transformations, $m' = \alpha m$, with $\alpha > 0$. A probabilistic representation may be also developed for finite structures [9]. Concerning the practical application of these ideas to perceptual measurement, suppose that a group of subjects are asked to rate sounds in term of both loudness differences and loudness ratios and that $m', M: A \to \mathbb{R}$ are the corresponding resulting measure functions. If furthermore it is possible to fit data in such a way that, for each $a \in A$,

$$m(a) \cong \alpha_1 (m'(a) + \beta), \quad m(a) \cong \alpha_2 M(a)^\gamma, \quad (3)$$

then $m: A \to \mathbb{R}$ constitutes a measure function for loudness on a ratio scale. We call *robust magnitude estimation* (RME) such a procedure [8, 9].

3. TEST METHOD

The method consists in proposing to each subject two separate and independent tests, aimed at representing perceived differences and ratios respectively. They are based on the same set of stimuli and differ in the way subjects express their perceived sensations. For example, the perceptive space available to the subject is different in the two cases. In the magnitude-estimation test the subject has to evaluate a stimulus as compared with a fixed one, so that two stimuli are available at a time. In the interval-estimation test the overall set of reference stimuli is available to the subject, so it is possible to have a complete idea of the intensities to be evaluated, before starting the evaluation procedure. This main difference has an impact on subject's coherence also, since in the former case coherence in the evaluation of different stimuli is due mainly to subject capability to remind the previous stimuli evaluation, while in the latter all the stimuli are available for perception, so the experiment set up is more favourable to a coherent evaluation.

3.1 Test procedure

The tests were proposed to each subject, one for a time, and discussions about the tests among subjects waiting and the subject who has just undergone the tests were avoided. Instead, a brief discussion was possible with the experimenter to give a feedback satisfaction to the subject and to gather the impressions from the test users.

Before starting with the first test, some time was dedicated to the reading of a standard instruction sheet, in order to give the same basilar information to each subject. In general such an instruction was sufficient for the subject to proceed with the test but the experimenter is available, during the test execution, to give all the help needed in the management of the test interface.

3.2. Magnitude estimation

The first test consist in the evaluation of the perceived sound intensity of a single stimulus, according to the perception of a reference stimulus, or anchor stimulus, that corresponds to a fixed reference point, set for convention to 10. So the subject has the possibility to listen to the anchor, by knowing that its evaluation is set to 10, and to the stimulus under test, all the required times, before entering the evaluation for the stimulus under test, according to the perception of the anchor. For example, if the anchor is perceived twice as intense as the test stimulus, its evaluation will correspond to 5, while on the contrary if the anchor is perceived as a half intense as the test, its evaluation will be 20. The procedure is carried out by a user friendly interface as depicted in Figure 1, where the subject can click on the buttons to listen to the stimuli, and insert the evaluation in the proper field.

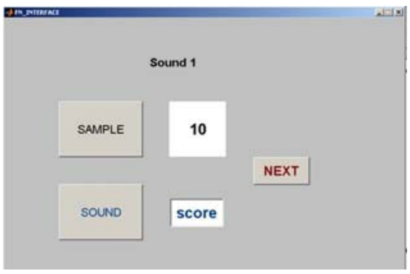


Fig. 1. Magnitude-estimation user interface

The stimuli sequence is deterministic and it is the same for all the jury. It has been designed by alternating real and synthesised stimuli (see section 4.1) for an overall of 11 evaluations.

3.3. Interval estimation

Once completed the magnitude-estimation test, the subject has about one minute pause, necessary to memorise some personal information such as age and music and sound capabilities, before proceeding to the interval estimation test.

As already depicted, in this test the subject has the possibility to listen to a set of 7 synthesized sounds, by clicking on the corresponding buttons. Then, by right clicking on them, it is possible to move the buttons on a line, positioning them from the one perceived as the least intense to the most intense, according not only to the rank, but to the difference between them also. So that sounds perceived with a similar intensity will be nearer than sound perceived with a larger difference. An idea of the user interface is given in Figure 2.

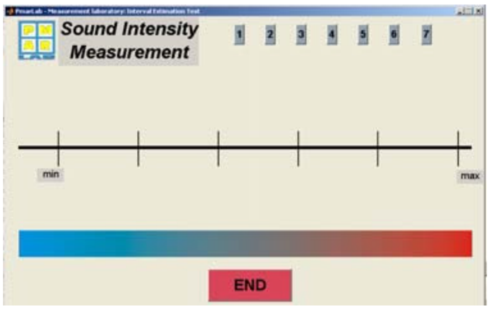


Fig. 2. Interval-estimation user interface

It is interesting to note that different subjects may implement different procedures to carry out this test. Of course it is possible to start from the least intense stimulus and to increase step by step following the perceived intensity ranking. But it is possible also to start fixing the perceived least and most intense and then to set the intermediate stimuli. The two procedures can give different results since the latter starts from the two extreme positions of the line, while the former starts from the minimum (or symmetrically from the maximum) and ends in a unknown point due to the differences in the perceived intensities. The different use of the reference line, will create problems in the processing phase, since, in order to create an average result, the evaluations of perceived intensity obtained by different subjects have to be normalised, for example by setting the minimum and maximum values for each subject to 0 and 100. Of course this way of doing is absolutely necessary to process the results from all the jury, never the less it presents an amount of arbitrariness since subject evaluations are modified by the experimenter. We will come back on this point in section 5.2, when discussing the proposed procedure.

Once the evaluation of the first set of seven synthesised stimuli is completed their position along the line is frozen and the buttons cannot be moved anymore. Now the other 4 stimuli are evaluated one at a time, with respect to the positions of the first seven stimuli. This second phase of the test can be compared to a direct measurement procedure on a standard reference scale constituted by the previous seven stimuli which now act as reference samples of perceived intensity. At the end of this second phase the positions of all 11 stimuli are recorded and the test is completed.

4. EXPERIMENTAL SET UP

Once defined the specific test methods, it is possible to proceed with the experiment design, by selecting the proper stimuli and the hardware equipment necessary for their presentation to the subjects.

4.1. Stimuli selection

In order to select a general test case for the proposed measurement procedure, a set of mixed synthesised and real stimuli was selected.

Real sounds are four noises recorded on board of trucks at the terminal container. One noise is a truck stopped with the engine at minimum regime, while the other are different trucks models running in the harbour to execute the same duty for containers transport.

Synthesised sounds are seven pink noises designed with A-weighted sound pressure levels to cover the full interval among the four real sounds. Table I gives some figures about stimuli standard acoustic pressure level, A-SPL and a perception focused parameter, Loudness as defined in [13].

TABLE I. Stimuli characteristics

Stimuli		**Parameter**	
		A-SPL [dBA]	Loudness[1] [sone]
Pink Noises	1	50	7
	2	53	9
	3	56	11
	4	59	13
	5	62	16
	6	65	20
	7	68	24
Real Sounds	8	55	10
	9	64	18
	10	66	18
	11	67	19

Real and synthesized stimuli were trimmed to a standard duration of 3s, suitable to give the proper intensity perception and short enough to reduce annoyance.

As regards the magnitude estimation test the synthesised stimulus with mean A-SPL was selected as anchor reference, while the pseudo-random presentation sequence has been established to propose real sounds alternated with pink noises.

4.2. Stimuli presentation

The test were carried out in an acoustically controlled room, with limited acoustical reflection, by using a set of flat response loudspeakers. The loudspeakers and subject's positions in the room have been optimised according to previous studies [7, 8].

The reproduction was characterised by recording the stimuli with a calibrated reference microphone and

[1] Zwicker's Loudness according to ISO 532

a binaural dummy head, in the same subject position. The volume was settled in order to obtain the values as presented in Table I.

5. RESULTS

The results we are presenting refer to 33 tests carried out by a jury of 31 subjects. Two subjects have conducted the tests two times in different moments.

5.1. Processing

Let's start with the pink noises' results. Consider a specific subject, since we have results for the same i^{th} stimuli, from the two tests carried out in the same test session, we will indicate with M_i the magnitude test result and with m'_i the one from interval estimation, for the same i^{th} stimulus.

According to the theory presented in section 2, and in particular equation (3), we will have:

$$\alpha\left(m'_i+\beta\right)=\gamma M_i^{\delta}, \qquad (4)$$

with four parameters to be determined for each subject. Applying logarithms to the equation we can write:

$$\frac{1}{\delta}\log\frac{\alpha}{\gamma}+\frac{1}{\delta}\cdot\log\left(m'_i+\beta\right)=\log M_i, \qquad (5)$$

where now we have 3 independent parameters plus γ: a scale parameter. Unfortunately the relationship among the parameters is non linear, so we need a non linear procedure to estimate the parameters values. In this first phase we have used a non linear least squares method.

According to equation (3) we have 2 measures of perceived loudness, that in principle should be equivalent:

$$Si_i \triangleq \alpha\left(m'_i+\beta\right) \quad \text{and} \quad S_i \triangleq \gamma M_i^{\delta}, \qquad (6)$$

where the new symbols *Si* and *S,* have been introduced, the former denoting the results obtained from the interval, the latter those from the magnitude estimation. The scale parameter γ may be evaluated by normalising a specific result to a reference value. As an example in the following we have evaluated the γ parameter by considering that for a specific pink noise, let's indicate it as stimulus *0*, the perceived intensity measure obtained in the magnitude estimation test has to be equal to the total Zwicker's Loudness measure:

$$S_0=L_0=\gamma M_0^{\delta}, \qquad (7)$$

from which we can evaluate the scale parameter:

$$\gamma=\frac{L_0}{M_0^{\delta}}, \qquad (8)$$

Once the complete set of parameters has been estimated for a specific subject, it is possible to verify experimentally if the two perceived Loudness measurements, effectively correspond each other, and eventually if they correspond with the Zwicker's Loudness previously computed on the basis of the recorded signal.

In a second moment, once the subject's parameters have been fully established, it will be possible to consider the results for the real sound, obtaining as for the reference pink noises, two perceived Loudness measures, that can be compared between themselves and with the Zwicker's Loudness, taking into account that they were established after a sort of 'subject calibration' based on reference pink noises.

5.2 Some experimental results

We will now present some experimental results. First of all we will have a look to the different behaviour of magnitude and interval estimation results, as presented in figure 3. This difference for the same perceived quantity, by using the same set of stimuli, may be due only to the different test methods, and it was one of the motivations for this study.

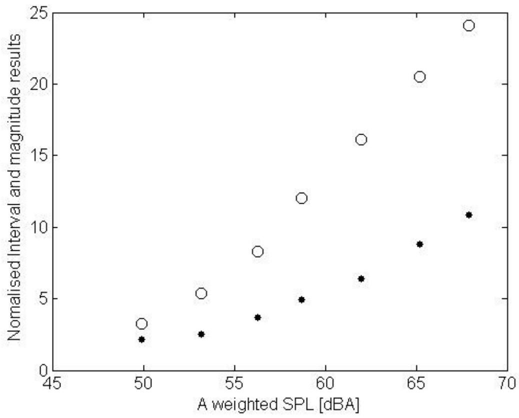


Figure 3. Mean results for the perceived loudness, obtained with magnitude (dots) and interval estimation tests.

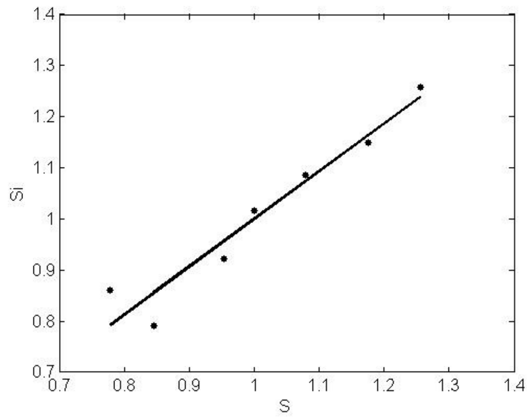


Figure 4. Perceived Loudness measures *S*, from magnitude, vs *Si*, from interval estimation tests.

Now we will have a look to some results for a single subject and then we will introduce some grouped data for the overall jury of 31 subjects and 33 tests.

Figure 4 presents the behaviour of the perceived Loudness measure obtained from interval data, *Si*, versus the same measure obtained from the magnitude estimation results, *S*. The graph shows a high correlation between the two measures.

Note that in the proposed procedure no arbitrary normalisation is required while previously it was necessary when dealing with different scale used in the interval estimation test. So in this case we can analyse grouped data for all the jury by using the specific set of parameters determined for each subject in the specific moment in which the test has been carried out. It may happen that at different times, the same subject will have a different behaviour, but the proposed procedure takes this into account by considering the results of the two tests carried out at the same time.

Figure 5 presents the histogram for the correlation coefficients between perceived loudness measures Si_i and S_i. Good correlation levels are most frequent, while the minimum, that in any case is about 0,8 , is due to a single subject.

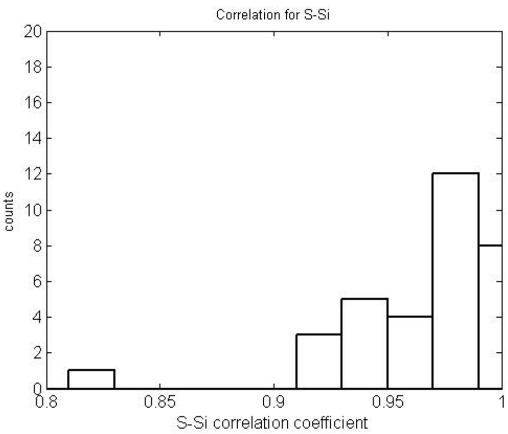
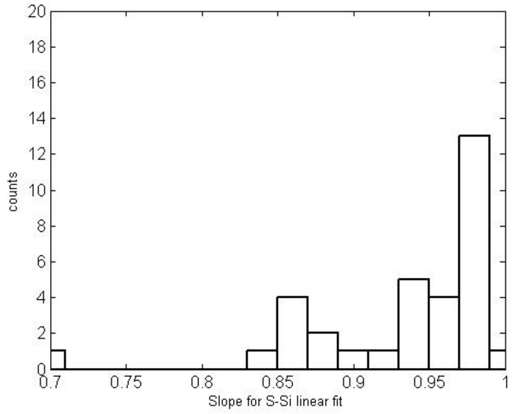


Figure 5. Histogram of correlation coefficients for S and Si

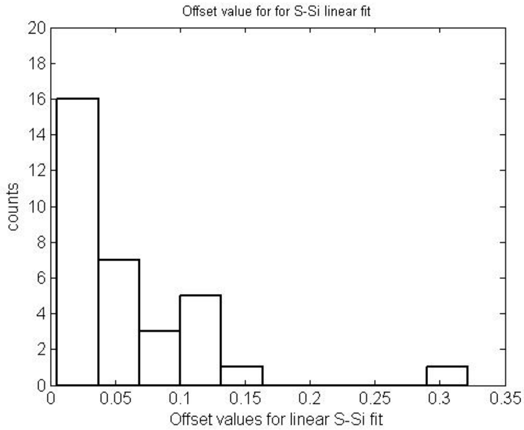
A deeper analysis can be done considering the linear relationship between the two measures:

$$Si_i = S_i \cdot k + v \,, \tag{9}$$

A linear dependence is evident from the correlation coefficient, but the fact that they refer to the same scale becomes clear by analysing the coefficients of the linear relationship. Figure 6 presents the histograms for the parameters *k* and *v* for each of the tests considered. Now it is clear that the two measures refers to the same scale since the constant parameter *v* is zero with a little dispersion and *k* behaves similarly around one.



(a)

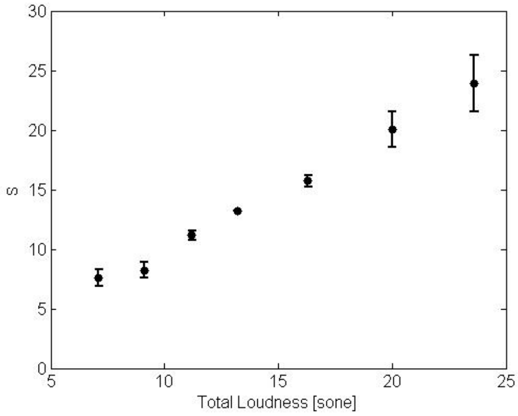


(b)

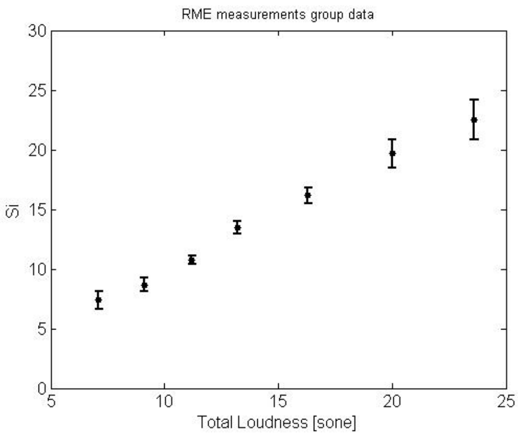
Figure 6. Histogram for slope *k* (a) and offset *v* (b) of the linear relationship between *S* and *Si*.

Now that we have established that there is a unique measurement scale for the perceived Loudness, independent from the test method, we can proceed comparing the results with the Zwicker's Loudness computed according to [13]. Figure 7 presents the results for the pink noises, for the two measures. Note that the Loudness measure computed from the Magnitude estimation results, *S*, has been used for normalisation in correspondence of the fourth stimulus, while the measure derived from the interval test, *Si*, depends on all the four parameters characterising each subject.

By considering the good agreement between computed Loudness and measured perceived Loudness the idea of a unique measurement scale emerges. This is confirmed by the results on the real sounds, that are presented in Figure 8. Here we can note the substantial agreement of the mean values, with a larger dispersion. This may be due to the large difference in perception between a pink noise, and the real sounds. In fact in the former case the power is uniformly distributed in the perception bands, while in the latter the power distribution is peculiar of the generation process.

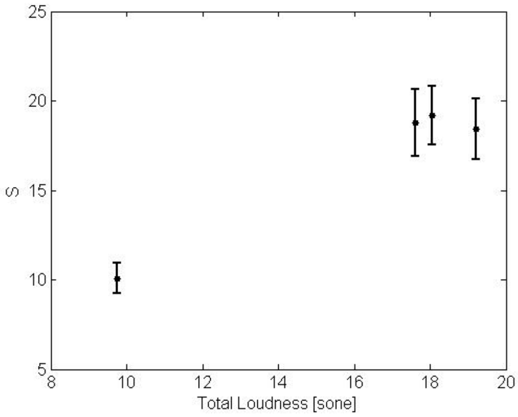


(a)

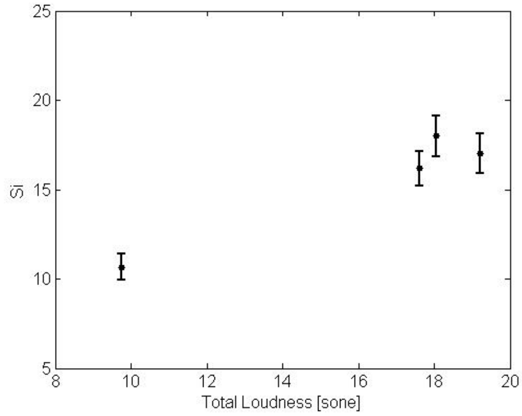


(b)

Figure 7. Perceived Loudness measure S (a) and Si (b),as a function of the Total Loudness according to ISO 532, for the seven pink noises. Bars indicate ± one sigma referred to the mean value.



(a)



(b)

Figure 8. Perceived Loudness measure *S* (a) nd *Si* (b) as a function of the Total Loudness according to ISO 532, for the four real noises considered in the study. Bars indicate ± one sigma referred to the mean value,

5. CONCLUSIONS

A new procedure for the measurement of loudness by persons has been presented and discussed. It has been show how to combine results from two different and complementary procedures, to obtain a measure on a ratio scale. The procedure, called Robust Magnitude Estimation, combines results on an interval scale and on a ratio scale, in a single measurement value on a ratio scale, without any arbitrary normalisation. Results are in agreement with Loudness evaluated according to current standard [13]. The procedure also allows the evaluation of intra and inter individual differences. The application to sounds obtained in a measurement campaign in a port environment has been presented and benefits in term of ergonomics have been mentioned.

REFERENCES

[1] Finkelstein, L., “Widely, strongly and weakly defined measurement”, Measurement, vol. 34, pp.39-48, 2003

[2] Rossi G.B., Berglund B., Measurement involving human perception and interpretation, Measurement, vol. 44, pp. 815-822, 2011

[3] Roberts, F. S., Measurement theory. Addison-Wesley, Reading, MA ,digitally reprinted, 2009

[4] Narens, L., “A theory of ratio magnitude estimation”. Journal of mathematical psychology, vol. 40, pp. 109-129, 1996

[5] Krantz, D. H., Luce, R. D., Suppes, P. & Tversky, A., Foundations of measurement. Dover, Mineola, 2007 (originally published by Academic Press, 1971).

[6] Miyamoto, J. M., “An axiomatization of the ratio difference representation”, Journal of mathematical psychology, vol. 27, pp. 439-455, 1983

[7] Rossi G. B., Crenna F., Panero M., “Panel or jury testing methods in a metrological perspective”, Metrologia, vol 42, pp. 97-109, 2005

[8] Crenna, F., Panero M., Rossi, G. B. “Measurement of perceived sound intensity accounting for individual differences”, Proc. of the 12th IMEKO TC7 & TC1

Joint Symposium on Man Science and Measurement, 2008, Annecy, France, pp. 81-88

[9] Rossi G.B., Crenna F., "Insight in ratio scales", in A. Bastianelli & G. Vidotto (Eds.) (2010) Fechner Day 2010. Proceedings of the 26th Annual Meeting of the International Society for Psychophysics, Padua, Italy. The International Society for Psychophysics, October 19-22, 2010, pp. 215-220

[10] Carletti E, Brambilla G , Pedrielli F, "The noise climate at the operator station of construction machines: comparison between naive and experienced subjects", Proc. 9th ICSV, Int.Conf. on Sound and Vibration, Orlando, USA, 2002

[11] Crenna F, Panero M and Rossi G B , "Evaluation of the sound emitted by post-sorting machines with jury test methods", Proc. 17th IMEKO World Congress ,Dubrovnik,2003

[12] Brambilla G. , Carletti E., Pedrielli F, "Perspective of the sound quality approach applied to noise control in earth moving machines", International Journal of Acoustics and Vibration. vol 6, nr.2, pp 90-96,2001

[13] ISO 532:1975, "Acoustics - Method for calculating loudness level"

Author(s): Dr. F. Crenna, Prof. G.B. Rossi, Dr. L. Bovio, DIMEC, Dept. of Mechanics and Machine Design, Measurement Laboratory, University of Genova, Via all'Opera Pia 15A, I-16145 GENOVA, ITALY.
crenna@dimec.unige.it, gb.rossi@dimec.unige.it, bovio@dimec.unige.it