12th IMEKO TC1 & TC7 Joint Symposium on
Man Science & Measurement
September, 3 – 5, 2008, Annecy, France

MEASUREMENT OF PERCEIVED SOUND INTENSITY ACCOUNTING FOR INDIVIDUAL DIFFERENCES

Francesco Crenna, *Matteo Panero*, *Giovanni Battista Rossi*

University of Genova, DIMEC, Genova, Italy. crenna@dimec.unige.it

Abstract: The measurement of quantities expressing the human response to some external stimuli is of great interest for metrology. In this paper some aspects of scale construction and measurement procedures are discussed, presenting different approaches, both direct and indirect, and focusing on the design of the panel test in order to obtain useful information for setting up the measurement scale and for the evaluation of its uncertainty.

The measurement of sound intensity is considered and a case study concerning the internal noise of container trucks operating in a port is presented. The following methods are considered:

- magnitude estimation;
- interval estimation;
- master scaling;
- indirect measurements according to A–weighted sound pressure level and Zwicker's loudness.

Preliminary experimental results show the differences and criticalities among the proposed approaches, mainly due to the inherent *inter* and *intra* individual variability related to the perception itself and to its dependence on contextual factors.

Keywords: perceived quantity, sound intensity, panel testing.

1. INTRODUCTION

The main methods for the measurement of a perceived quantity rely on jury or panel testing [1-3]. Historically panel testing was developed by the food and flavour industries, which are concerned with the relationship between the physical and perceptual domains [4]. Jury test involves the presentation of a set of stimuli (which assume the form of sound samples in our application) to groups of subjects[1] according to a strictly defined test protocol, then their responses are recorded and processed. Data so produced provide an essential empirical basis for the construction of a measurement scale for the quantity of interest. On the basis of such scale a measurement procedure may be defined in order to evaluate other quantities of the same class of those considered for scale construction, with reference to the previously defined scale.

The measurement of a perceived quantity requires several procedural steps, from the construction of the measurement scale by using some reference samples, to the measurement procedure based on the former scale. These steps can be considered direct measurement procedures, since the measurement is based on the judgments by a set of subjects. Once the measurement scale is set up, in some cases it is possible to find good correlation with a metric based on some physical quantity that can be measured by properly analysing the quantity considered. This may be configured as an indirect measurement procedure.

Differences in individual sensitivity and scaling behaviour may affect the magnitude scale. The master scaling method proposed by Berglund [5] allows the comparison with similar perceptual measurements made by other persons in other occasions [6]. So it can be particularly useful for overcoming problems related to *inter* and *intra* subjective differences.

In the present paper the authors study different approaches to the measurement of perceived sound intensity. The sounds considered have been recorded on board of container trucks operating in a port container terminal. The methods investigated include:

- magnitude estimation by free numbers,
- interval estimation,
- master scaling,
- A–weighted sound pressure level (SPL) and Zwicker's loudness.

Experimental results enable a comparison of the different approaches.

2. METHODS

Several methods have been used for the direct measurement of perceived quantities and in particular of sound intensity. All these methods involve the set up of a panel of judges who will express their response about the quantity under investigation according to a specific method. Every method shows particular characteristics, which can be related either to the metrological properties of the resulting measurement scale and to the perception process which can affect the test results. In any case the task has to be clearly defined in order to set up the experiment and to carry it out properly by describing it to the observer. In our case the task

[1] **Jury test can be performed with a single person or many people at a time; both cases have their own set of benefits and shortcomings [11].**

is to judge the perceived intensity evoked by each sound under investigation.

The methods investigated in this research activity are presented in the following paragraphs.

2.1. Magnitude estimation by free numbers

In magnitude estimation, as proposed by Stevens [7], observers are presented with a series of stimuli and asked to assign a number to some attribute of the stimulus that is appropriated to describe their perception. The only restriction imposed on the numbers assigned is that they have to be positive. For example, if a value of 10 is assigned to the first stimulus and a subsequent stimulus seems half as intense as the first, the participant would assign the second stimulus a value of 5. In some cases the range of the possible numbers is specified and/or an anchor is defined, by assigning a reference number to a default stimulus. In magnitude estimation with a standard, the experimenter presents a stimulus called standard and assigns it a number called the *modulus*. For subsequent stimuli, subjects report numerically their perceived intensity relative to the standard so as to preserve the ratio between the sensations and the numerical estimates (e.g., a sound perceived twice as loud as the standard should be given a number twice the *modulus*).

In our case we have defined an anchor sound without limiting the range of the possible numbers. During a test organized in such a way the observer has the possibility to perceive two stimuli: one is the anchor and the other is the target stimulus. Then the target changes and the process goes on until all the stimuli are evaluated. In such a way a clear relationship is defined between each target and the anchor, and since an extensive quantity is involved, the perceived sound intensity, a ratio scale may be defined.

2.2. Interval estimation

In this method the observer interacts with the overall set of stimuli, with the possibility to perceive them all, one at a time, before assigning his/her judgement. In general the judgement can take place by assigning a number to each stimulus or by positioning them on a ruler [2-3]. In our experimentation we have decided to proceed graphically, so we have developed a graphical interface tool to assist the observer in judgment. The interface allows the judges to listen to the desired stimuli and consequently position them on a ruler according to his perception.

This method presents two main differences as compared with the previous one. First of all there are no anchor and no numbers. Then the whole perception space is available to the observer who can listen to all the stimuli before giving the judgement by positioning them on the ruler.

On these premises the method enables the construction of an interval measurement scale, since each observer gives as a result the relative position of each sound with respect to the others.

2.3. Master scaling

A disadvantage common to the two previous methods is related to the way the observer uses the available range of the scale. In the latter method this is evident considering that the observer can spread the stimuli across all the ruler from the minimum to the maximum limits, or he/she can start from the minimum and then proceed in filling it considering one interval at a time, finishing by positioning the maximum one in a point somewhere below the maximum of the ruler.

Moreover, since the observer is considered as the measuring instrument, his/her sensitivity may change from one observer to the other of course, and from time to time for the same observer also, giving rise to a high spread of the measurement results even if the measurand is properly controlled. In order to overcome these problems several methods can be used. All of them are based on a normalization of each subject results according to a common rule. If this way of doing on one hand provides a normalization of all the observers constituting the panel on a common scale, on the other it distorts the perception of each observer by adjusting the results in order to normalize them.

The master scaling method presents the advantage to reduce the *inter* and *intra* subject variability by normalization of each subject on a reference (*master*) scale, without distortion, but respecting the psychophisiological law governing the intensity perception of a perceived quantity. This method requires the introduction of a reference stimuli standard set (a standard context) as well as the target stimuli (the truck noises in this case), which are all evaluated at the same time by each subject. On the hypothesis that the perception behaviour follows the Stevens's perception power law [7], according to the results obtained from the observer with the reference stimuli, a specific scale for the intensive quantity under investigation can be produced. During the test properly mixed with the reference stimuli, the observer provides also results for the stimuli under investigation. So it is possible to relate each result to the scale for the specific observer at the specific time when the test was carried out, thus reducing *inter*–subject variability. Moreover it is possible to consider the specific scale of each observer in the panel, obtaining a mean perception scale called master scale. When relating both reference samples or the results for the stimuli under investigation to the master scale the *intra* and *inter* subject variabilities are reduced and results are related to a common mean reference: the master scale.

The potentialities of such a method may vary from the reduction of variability to the possibility of comparison among different individuals subjective scales, obtained in different occasions and contextual settings, through the calibration of each subjective response curve on the master scale.

A more comprehensive, formal and detailed description of the master scaling fundamentals may be found in the original papers by Berglund [5-6].

2.4. Indirect measurement

In addition to experimental direct measurements, several indirect methods have been developed to measure the same quantity through a proper stimulus processing. In the particular case of sound intensity these processing are based on sensory process physiological modelling, on the sound time evolution and on its spectral content. Even if Zwicker loudness is, at the moment, the first and only one standardised psychoacoustical quantity [10], in the scientific community the discussion about indirect loudness measurements is still open [12], that is why it is interesting to compare direct measurements with the standardised Zwicker loudness and the most widespread and commonly used in the reference standards, A–weighted sound pressure level (A–SPL).

Zwicker loudness [11] is an indirect method for measuring sound intensity taking in account the average human response. In Zwicker's method, the sound pressure of a complex sound is frequency analysed and its components are weighted taking into account the auditory sensitivity considering isophonic curves, the mutual masking due to critical bands, and inhibition due to different stimuli in a short temporal sequence. Although Zwicker's model has been developed for steady, continuous sounds, it is widely employed in sound quality evaluations and it has been used in this study also.

The A–weighted sound pressure level is the most common standard metric used to perform acoustical measurements correlated to the human response. It is based on the 40–phon Fletcher–Munson equal loudness contour for typical human hearing [13]. A–weighting, although originally intended only for the measurement of low level sounds (around 40 phon[2]), is now commonly used for the measurement of environmental noise and industrial noise, as well as for the assessment of potential hearing damage and other noise health effects at all sound levels. Its widespread use is due mainly to the easiness and cheapness possibility of realisation in a filter without needing many processing components.

3. TEST DESIGN

As already described the basic technique for the measurement of quantities related to human perception is panel or jury testing. Generally this kind of tests requires a carefully selected panel, a set of stimuli to be proposed to each subject[3], an appropriate equipment and a procedure for presenting them and gathering the judgements. The design of the test procedure is fundamental to obtain useful information for the construction of a robust measurement scale, and consequently, for the reliability of the measurement results. In the following paragraph some of the salient issues involved in the subjective testing will be shown. Metrological aspects regarding jury testing methods and how they may be kept under metrological control may be found in [14]. Specific topics to be covered include:

- samples (the set of manifestations),
- reproduction system and
- panel test design.

3.1. Sound samples (the class of manifestations).

In our experiments the stimuli[4], whose perceived intensity we are interested to evaluate, consist of four recordings of the sounds emitted by trucks in the port environment. They are recordings of the sound inside the driver cabin of tractors used in the maritime container terminals for the movement of containers, under the following conditions:

1. vehicle stopped with engine running;
2. vehicle running:
    - without load,
    - with 20 TEU[5] load and
    - with 40 TEU load.

As already described in section 2.3., in order to apply the master scale method, in addition to the sounds under investigation a set of reference sounds is required to construct the specific psychophysical function of each observer. For these purposes we have synthetically produced seven recordings of pink noises. The set of reference stimuli was designed to cover a range of loudness and A–weighted SPL, wider than the target sounds, whose intensity has to be measured.

3.2. Sound reproduction

Sound reproduction plays a very important role in sound evaluation [14, 16]. Even if sound reproduction through headphones may be much easier to control [17], in order to give the listener a sensation as similar to the real environment as possible, the reproduction is carried out in an acoustically controlled environment, to limit the effects of the acoustic reflections and spectral modifications, using a sound diffusion system. The sound reproduction system has been realised exploiting a digital audio interface, an audio power amplifier matched with high quality speakers and subwoofer, together giving an overall almost flat frequency response [18]. The sound intensity was verified by positioning a microphone at the same place of the listener and recording the sound as emitted by the diffusion system. Then the level was adjusted to match the SPL of the original sound. Since the levels obtained were rather high, we decided to reduce them by a default amount of dB to give to the panel a more comfortable acoustic environment.

Although this reduction affects the experimental results, due to the fact that the human hearing system is more sensitive to some frequencies than others, and furthermore, its frequency response varies with level, we are interested in comparing the measurement methods and not in the absolute evaluation of the sound loudness.

3.3. Panel test design

Perceived loudness was assessed with the direct methods previously presented; results from the free number test were then processed according to the master scaling scheme.

In the free number test, the procedure requires subjects to assign a number according to their perception of the sound intensity, to each of the seven reference stimuli and to the four target stimuli, presented one at a time, in a random order. Although there is no limit to the range of numbers that a subject may use, it has been decided to assign a given

[2] **The *phon* was proposed as a unit of perceived loudness level L_N for pure tones by S. S. Stevens. By definition, 1 phon is equal to a sound pressure level of 1 dB at a frequency of 1 kHz.**

[3] **In this article, the terms *subject* and *observer* are used to refer to any person taking part to the evaluation of sounds in a listening study.**

[4] **A *stimulus* is defined as an event that may evoke a response [15].**

[5] ***Twenty-foot equivalent unit* (TEU) is often used to express the container capacity. It is a measure of containerized cargo capacity equal to one standard 20 ft (length) × 8 ft (width) container. A unit of measurement equal to the space occupied by a standard twenty foot container.**

number (*modulus*) to one particular reference stimulus. In this case a rating of 10 has been assigned to the pink noise sound at the half of the A–weighted SPL range. Numbers were gathered through a user interface as depicted in figure 1a. Note that the observer has the possibility to listen to the anchor sound as well as the sound to be evaluated. When the number is expressed, the interface proposes a similar situation with the same anchor and another sound to be evaluated. The process is repeated for all the eleven sounds.

As regards the interval estimation test, a graphical interface tool was developed to assist the subject in evaluation. In this case the test is organised in two phases: first of all the seven reference samples are available for listening as shown in figure 1b. The seven reference sounds are represented by buttons, so the observer can either play sounds and position them on a ruler according to their perceived intensity. Once the end button has been pressed the intervals are gathered by reading the buttons positions on the screen along the ruler in the horizontal direction. In a second phase the four target sounds are proposed one at a time, for evaluation on the same ruler, where there are still the seven reference stimuli as positioned previously by the same subject. The observer can listen to either the target or the reference sounds of course, but it is not possible to move anymore the reference sounds, which now constitute the observer measurement interval scale.

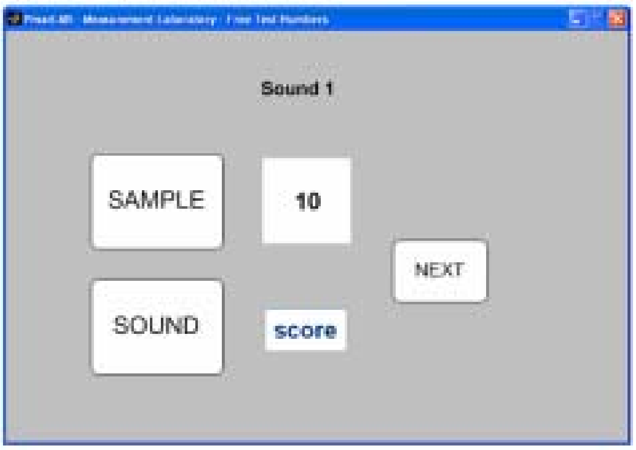

Fig. 1a. User interface for the free number test

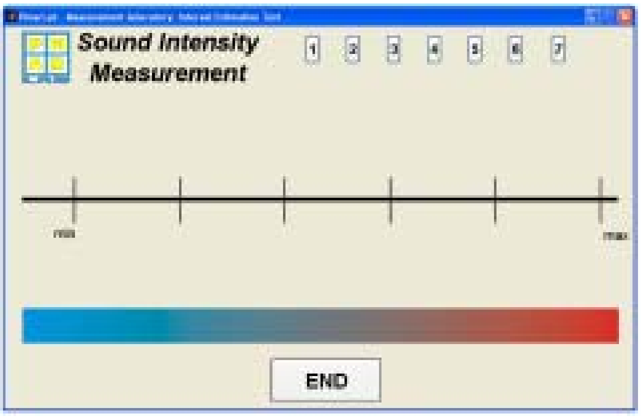

Fig. 1b. Graphical interface tool to assist the observer during interval estimation test

The two procedures differs from many point of views. The perception space is different: in the first case only the anchor and a target or a reference sound are available, while in the second the overall set of reference sounds is available for listening (as shown in figure 1a and 1b). In the first case the free number expression is leaded by the *modulus* and no other reference is available, while in the second case the subjects have the possibility to directly evaluate the differences between sounds and put them on the ruler accordingly.

3.4. Panel composition

The results we are going to present were obtained processing the test carried out by a panel of 18 people aged from 54 to 24 years old. All of them have carried out both tests. Some people have performed the tests several times in a time span of about 5 month to verify the behaviour of their judgement according to the situation at the moment. All the tests were conducted by one subject at a time after that an instructor, in the reproduction room, has introduced them to the test interface and given some directions regarding the perceived quantity under investigation. So during the test there is no possibility to have influence between subjects. Interactions between the subjects of the panel are avoided both before and after the tests.

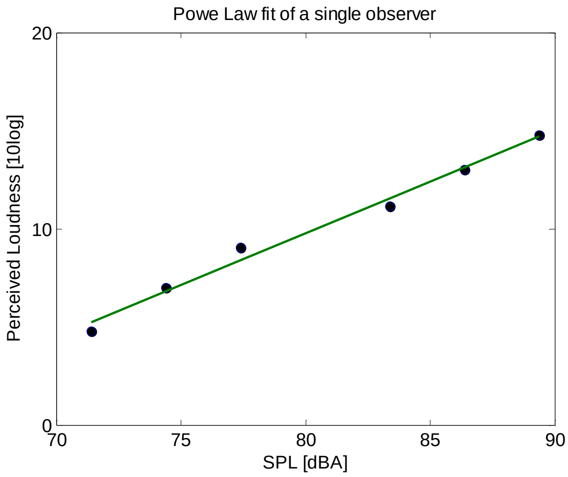

Fig. 2. Steven's power law fit for an observer perceived loudness

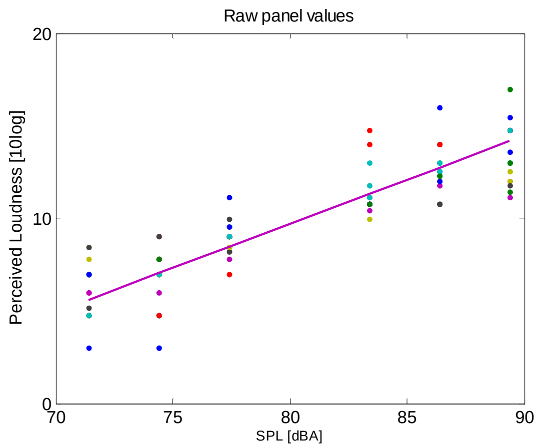

Fig. 3. Raw free number test results for reference samples

4. RESULTS OF THE EXPERIMENTATION

This paragraph deals with some experimental results obtained using the methods previously described with the panel presented in section 3.4.

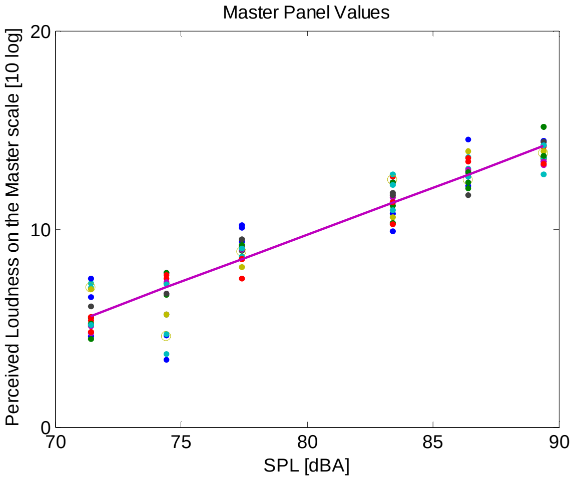

Fig. 4. Free number test results for reference samples, normalized on master scale

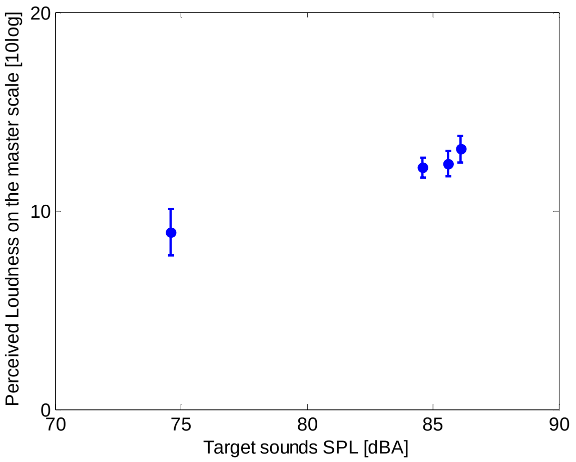

Fig. 5a. Free number test mean results for target sounds, normalized on master scale, vs SPL

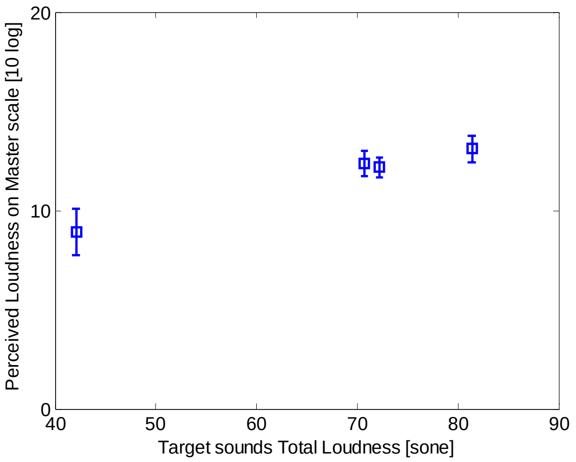

Fig. 5b. Free number test mean results for target sounds, normalized on master scale, vs Zwicker's total loudness

4.1. Free number test

As already depicted, the free number test results are processed according to the master scale principles. The first step involves the hypothesis of a power law intensity perception. In figure 2 the results of one observer for the six reference sounds (the anchor sound is omitted) are shown. The scale for both the A–weighted sound pressure level and the perceived intensity are logarithmic. The good fit obtained with the straight line confirms the validity of such an hypothesis.

Figure 3 displays the results for the overall panel as regards the reference stimuli. The line is the mean power law for the panel or in other word the reference master scale. Then according to the master scale procedure it is possible to normalize the result from each subject on the master scale. This is shown in figure 4. It is worth noting that figures 3 and 4 show free number tests results before and after normalization: the reduction in the results variability is evident. The following figures can give a better idea in this sense.

Table 1 displays the relative variability reduction obtained by master scale normalization. When dealing with mean results from the panel, the variability on the master scale has to be reduced considering the number of observers in the panel, so it assumes values around and below 10%, which allows reliable measurements for common applications.

	Reference sounds	Target sounds
Variability reduction (Δσ/σ)	-20 ÷ -70 %	-20 ÷ -50 %
Variability on Master scale	4 ÷ 23 %	8 ÷ 25 %

Table 1. Relative variability on the master scale and variability reduction obtained by master scale normalization

The master scale numbers obtained for the target sounds are shown in figure 5a and 5b. The figures present the behaviour of the direct measurement in respect to the indirect one: A–SPL and Zwicker's loudness. Mean values for the panel are identified by squares, while the bar lengths correspond to two standard deviations.

It is possible to have a look at the *inter*–subject variability, by referring to the results of several test sessions carried out by the same observer at different times.

Figure 6a and 6b shows raw and master scale results for a single observer who has carried out four times the test, in a time span of 5 months. In this case it is possible to have the view of how much the perception power law may vary for the same subject at different time, even if the context of the test is not varying. Master scale takes care of these variations considering as a reference a mean (Master) power law for the subject.

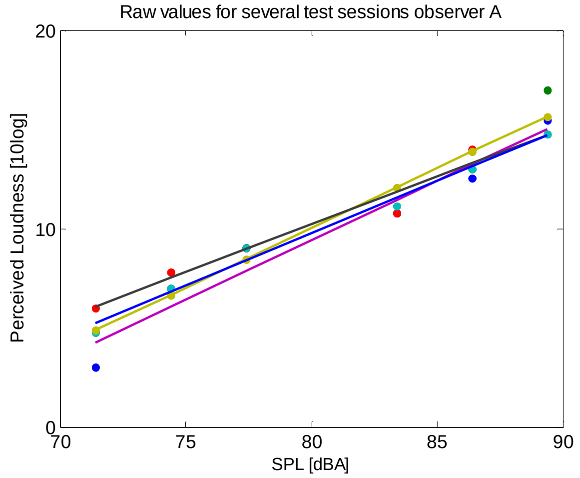

Fig. 6a. Free number test results for reference samples , for a specific observer .

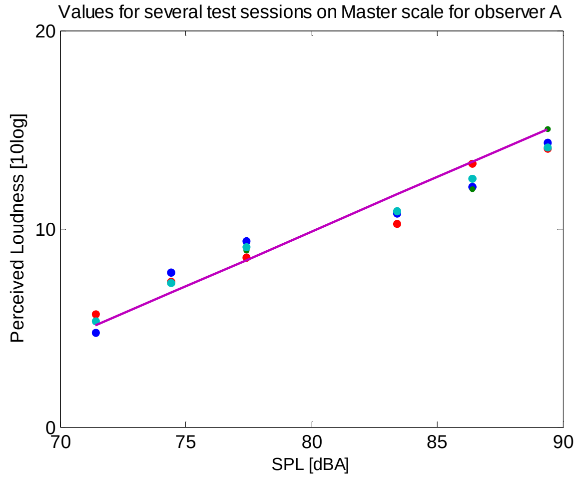

Fig. 6b. Free number test results for reference samples , for a specific observer, normalized on his own master scale

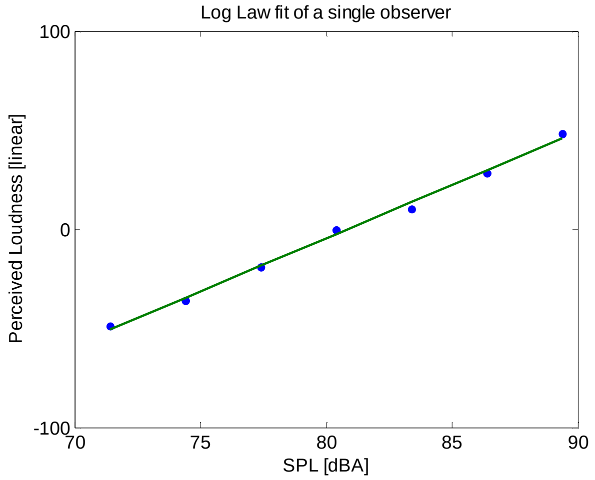

Fig. 7. Interval estimation test, results for one observer for the reference sounds.

4.2. Interval estimation test

Figure 7, shows interval estimation results for the seven reference sounds. It is worth noting that in this case the perception behaviour does not follow the power law but a logarithmic one, according to what could be expected for an interval scale [1, 20, 21]. This is evident in the graph noting that the perceived intensity is presented in linear units instead of logarithmic as when dealing with the free number test.

In this case there is no possibility to apply a Master scale processing, since the basic hypothesis is not valid anymore.

A normalization may be proposed considering the central reference sample and setting it to the same value for each observer[6]. In such a way the values of each sample are translated without influencing the perception behaviour. At the same time, the reference sample used as an anchor in the free number test, has now the same value for each observer, so the comparison of the results is more straightforward. Proceeding in such a way a slight variability reduction is possible and results for the reference samples and for the target sounds are displayed in figure 8a and 8b. Squares and bars are as in the previous figure mean values and double standard deviations.

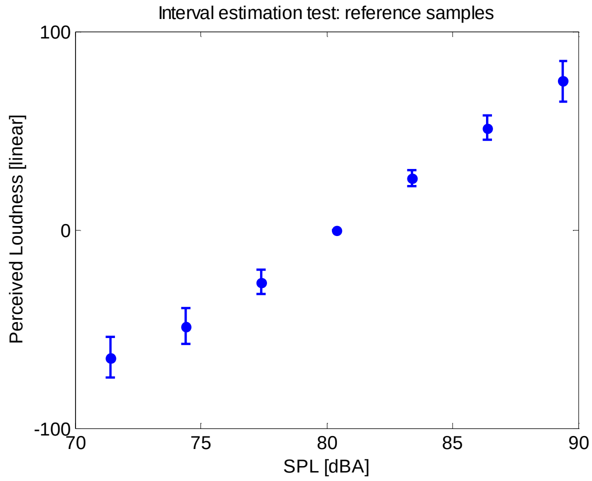

Fig. 8a. Interval estimation test, overall panel results for the reference samples.

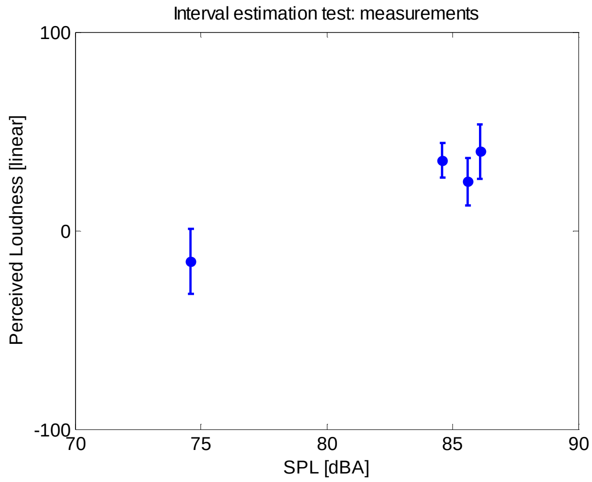

Fig. 8b. Interval estimation test, overall panel results for the target sounds.

6 **This kind of normalization may be considered as an equal intercept transformation [6].**

5. CONCLUSION

The paper dealt with an investigation on different methods for measuring a perceived intensity, with application to the perceived sound intensity. It is shown that when using direct methods based on jury testing, conditions and methods are fundamental since they affect the perception and consequently the feedback from each subject. Moreover the test method considered might affect the properties of the measurement scale which is under construction influencing the basic hypothesis such as the perception law. Experimental data presented have shown that the hypothesis of perception according to Steven's power law is valid when considering a free number test method, while it is not consistent, as expected, when evaluating intervals, such as in the second test modality described. In our application the more suitable perception law in this second case seems to be a Fechner law, considering the stimulus logarithm.

The method based on master scaling shows its main advantage in the reduction of the *inter* and *intra* subject variability, with the possibility to normalise the results obtained for different subjects and/or in different contexts. In our application the variability reduction is important and thanks to it a measurement of the perceived intensity for the considered sound was possible.

Moreover, the paper analysed indirect intensity measurement methods, revealing a general agreement with the perceived quantity, even if in detail the two methods considered, A–weighted SPL and Zwicker's loudness show some differences.

Even if a larger panel is required to consolidate the experimental results, the test case considered shows the potentialities of these techniques in industrial applications to improve the ergonomy at the workplace.

ACKNOWLEDGMENTS

This work is carried out with the contribution of the European CA MINET, Measuring the Impossible NETwork (Contract No.043297-NEST).

REFERENCES

[1] G. B. Rossi, "Measurability", Measurement 40, pp. 545-562, 2007.
[2] G. B. Rossi, F. Crenna, M. Codda, "Measurement of quantities depending upon perception by jury-test methods", Measurement, 34,1pp. 57-66 , ISSN: pp. 0263-2241, 2003.
[3] G. B. Rossi, F. Crenna, M. Panero, "On the concept of measuring system in measurements based on jury testing methods", Proc. of the 2nd International Symposium on Measurement, Analysis and Modelling of Human Functions and 1st Mediterranean Conference on Measurement, pp. 77-82, ISBN 88-901344-0-2,June 14-16, Genova (Italy), 2004.
[4] Richard H. Lyon, "A Sound Guide to Product Acceptance", The Industrial Physicist 3, pp. 50-51, 1998;
[5] B. Berglund, E. L. Harju, "Master scaling of perceived intensity of touch, cold and warmth", Europ. Journal of pain 7, pp. 323-334, 2003.
[6] B. Berglund, "Quality assurance in environmental psychophysics" in S J Bolanowski, G A Gescheider (eds), "Ratio scaling of Psychological Magnitudes - in honour of the memory of S S Stevens", ch.11, pp. 140-162, Erlbaum 1991.
[7] S.S. Stevens, "On the psychophysical law", Psychological Review 64, pp. 153–181, 1957.
[8] L. Narens, "A theory of ratio magnitude estimation", Journal of Mathematical Psychology 40, pp. 109–129, 1996.
[9] W Ellermeier, G Faulhammer, "Empirical evaluation of axioms fundamental to Stevens's ratio-scaling approach: I. Loudness production", Percept Psychophysics 62-8, pp 1505-11, 2000.
[10] ISO 532:1975, "Acoustics - Method for Calculating Loudness Level", International Organization for Standardization, 1975.
[11] E. Zwicker, H. Fastl, "Psycho-Acoustics: Facts and Models", Springer, 1989.
[12] B. C. J. Moore, "An Introduction to the Psychology of Hearing", 5th Ed., Academic Press, San Diego, 2003.
[13] H Fletcher and W A Munson, "Loudness, its definition, measurement and calculation", J.Acoust. Soc Am.5, pp. 82-108, 1933.
[14] G. B. Rossi, F. Crenna, M. Panero, "Panel or jury testing methods in a metrological perspective", Metrologia, 42, pp. 97-109, 2005.
[15] ANSI S3.20–1973, "American National Standard–Psychoacoustical Terminology", 1973.
[16] O. Norm, "Guidelines for jury evaluations of automotive sounds", Sound & Vibration Magazine, pp. 1-14, April 2001.
[17] M. Bodden, "Perceptual sound quality evaluation", Proc. Internoise, Nice 2000.
[18] AES20–1996, "AES recommended practice for professional audio – Subjective evaluation of loudspeakers", Audio Engineering Society, 1996.
[19] S. Kuwano, S. Namba, "Dimensions of Sound Quality and Their Measurement", 17th International Congress on Acoustics, Rome, September 2-7, 2001.
[20] F. Purghè, "Methods of Psycho-Physics and Uni–Dimensional Scaling", Boringhieri Bollati (in Italian), 1999.
[21] G. Fechner, "Elements of Psychophysics", trans. H. E. Adler, eds., E. G. Boring and D. Howes, Holt, Rinehart and Winston, New York, 1966.
[22] S. Bech, "Training of subjects for auditory experiments", Acta Acustica 1, pp. 89-99, 1993.
[23] S. Bech, "Selection and training of subjects for listening tests on sound-reproducing equipment", J. Audio Eng. Soc., 40, 7/8, pp. 590-610, 1992.
[24] M. P. Snow, R. C. Williges, "Empirical Models Based on Free-Modulus Magnitude Estimation of Perceived Presence in Virtual Environments", Human Factors, Vol. 40, 1998.
[25] E. Borg, "On Perceived Exertion and its Measurement", Doctoral Dissertation, Stockholm University, Faculty of Social Sciences, Department of Psychology, 2007.

[26] B.C.J. Moore and B.R. Glasberg, A revised model of loudness perception applied to cochlear hearing loss, Hear. Res. 188: 70-88, 2004
[27] J C Baird, E Noma, "Fundamentals of Scaling and Psychophysics", John Wiley and Sons, New York, 1978.
[28] M. Mashhour, "Note on the validity of the power law", Scandinavian Journal of Psychology 6-2, pp. 220-224, 1965.