XVII IMEKO World Congress
Metrology in the 3rd Millennium
June 22–27, 2003, Dubrovnik, Croatia

EVALUATION OF THE SOUND EMITTED BY POST-SORTING MACHINES WITH JURY TEST METHODS

Francesco Crenna, Matteo Panero, Giovanni B. Rossi

University of Genova, DIMEC, Dept. of Mechanics and Machine Design, Genova, Italy

Abstract – Ergonomics aspects in working environment have grown the awareness of the need of improving sound quality of machinery sound as well as to reduce its sound. Psychoacoustics supplies important tools (with more sophisticated methods employing ear–related metrics) in order to evaluate the sound quality. It is well recognized that jury (or panel) evaluation provides a reasonable guide to value the quality of machinery noise in working environment. The following work belongs to a project dealing with the studies and the applications of psychoacoustic methods in order to evaluate the sound emitted by post–sorting machines. After a measurement campaign, two kinds of jury tests have been performed, one based on pair comparison forced-choice trials, the other on magnitude estimation. The results allowed us to define a robust scale of pleasantness of typical sounds generated in a wide variety of post–sorting machines, in different environments and under different operating conditions.

Keywords: human perception, jury test, sound quality.

1. INTRODUCTION

The sound produced by engineering and production equipment has gained attention fairly recently due to the increasing emphasis on the reduction of noise pollution to improve the quality of life in the working environment, and on the other hand due to improve products acceptability and desirability for the users. It is nearly well-founded that working environment noise impacts task performance. In particular in high-technology automation devices, such as post–sorting machines, leading companies pay close attention to ergonomic aspects, such as quality of sound. Hence the working environment comfort represents a main purpose for the same factory enterprise itself. The noise field around a machine varies in general with direction, time, acoustic environment and the operating and mounting conditions. Although the A–weighted sound level is much frequently used in practice (because of Standards requirements), it is not a satisfactory scale for assessing the subjective response of human beings. For example, as regarding the pleasantness of a sound, it is difficult to predict it merely by physical properties since the pleasant factor is related to cognitive as well as cultural aspects.

Sound produced by post–sorting machines is an inevitable consequence of the motion of the single components with which it is assembled. Hence it is very important to find a physical metrics which show a good correlation with subjective impressions in order to predict the sound characteristic and find countermeasures to improve sound quality. Since it is based on perception, it is to be determined by people, not by instrumentation. For this reason psychoacoustic methods are widely used in Sound Quality: on one hand they drive product improvements to lead the market, on the other hand they can improve ergonomic aspects for a better working environment.

In view of ergonomic aspects, it seems useful to define a set of procedures whose results are able to lead to a pleasantness/annoyance measurement judgement related to overall acoustic sensations. The principal tool of experimental psychoacoustic is the jury (or listening panel) test. Jury tests involve presenting a series of sounds to a panel of listeners and asking them to make judgements about the sound. Tests may be performed in several ways according to the features of the stimuli, the number of stimuli to be compared, their representations, the kind of judgement required and so on.

At present, no universally accepted procedure is available, so in this work we designed and validated a test procedure with attention not only to the metrological approach but also to meet the demands of the factory enterprise designing a series of tests as fast and slim as possible.

2. POSTAL EQUIPMENT

Even if post–sorting machines present different performances according to their final application, they can be described according to a sequence of standard sections. In this paragraph a general post sorting machine configuration is described. After the manual or automatic infeed the input post is transformed in a train of single envelopes in order to read the address. Then the envelope runs over a delay line, to have enough time to process the address image. Then it goes into the proper output tray which, when full, can be downloaded manually by an operator (figure 1), or automatically by a robot. In this case post packets can be wrapped out automatically. The overall length of such a plant varies according to the number of download trays, a common length being around 30m; such a plant can sort more than 40000 letters per hour.

Fig. 1. Manual tray download

3. THE RECORDING SET–UP

The stimuli consist of recording of the acoustic emissions produced by post–sorting machines. Recording measurements have been performed both on operators' positions and on additional positions which may be occasionally reached, for instance for surveillance or maintenance reasons. Actually, the noise recording was carried out considering practical limitations involving machine accessibility and non ideality of measurement conditions. The recorded noises were inserted in a database structure, which has been defined according to the machine type, working condition and recording position.

As regards measurement system, it consists of two parallel and independent measurement devices. The first consists of a capacitive microphone connected to a digital acquisition system in order to perform classical noise emitted measurements. For these kind of measurements fail to take the directional dependent amplifications of the human head into account, and may thus be very inaccurate for psychoacoustic purposes; so the latter is realized by a dummy head connected to a DAT recorder. This solution offers superior spatial performance compared to other recording techniques.

3.1. Sound pressure equalisation

For the purpose of the present research, which intends to evaluate noise features, regarding pleasantness, more than the sound pressure level, the recorded signals have been normalised. Several options are available in this sense: we decided to equalise signals at the same sound pressure level (SPL), without any weighting for the human frequency response. This seems to be the more objective normalisation leaving to the subject the judgement of the perceived sound. On the other hand it is possible that the effect of very different perceived SPL masks other perception effects which might be of interest. Actually this was not the case since noise sources are really similar and recording condition are as normalised as possible.

3.2. Reproduction system

The sound reproduction plays a very important role in Sound Quality. In this sense headphone reproduction is perhaps the most controlled method. In order to recreate the same sound scene, and then reproduce the same sensation, of the working environment all the jury tests have been performed in an acoustically controlled environment to limit the effects of the environment reflections and spectral modifications. The signals reproduction has been performed by using a flat response audio reproduction system. This equipment consists of a PC, a high quality audio interface, and a diffusion system with integrated power amplifier, consisting of 4 speakers and a subwoofer.

4. TEST DESIGN

Jury studies involve presenting a series of sounds to a panel of listeners and asking them to make judgements about the sounds. Such experiments have to be carefully constructed in order for the results to have statistical validity. One of the earliest examples was judgements of the loudness tones which produced the famous Fletcher–Munson curves.

Several jury test methods have been studied for the perception measurement, due to our particular test case, we can consider only constant stimuli tests, since during the test the physical stimulus can not be varied but it can only be selected from a fixed set of previously recorded sounds.

So two kinds of jury tests have been performed, one based on paired comparison forced–choice trials and the other on magnitude estimation. General and formal aspects regarding the test procedure design and the processing of the results are presented in other papers [1, 2]. In this paper we will address some problems related to this particular application and how we have decide to overcome them. During the design of a particular jury test two aspects play a fundamental role:

1. the achieved results must be as robust as possible;
2. the temporal length of the signals must to be shortest as possible.

Experimental evidences have showed how the attention of the judge diminishes, during a time consuming test. With regard to the last point we have decided to limit the time history of each recorded signal to 3s: this time is a good compromise between signal characterisation and annoyance due to the test length. Regarding both two points we have decided to limit our investigation to 6 different noises which were selected from a rich recording database, to have a set of signals representative of the possible noises emitted by a general post sorting machine.

Before starting with the tests, a formal description of his task (what he has to judge) the subject data are recorded and a short description of the test is presented.

4.1. Paired comparison

In this case the response required to each judge is a preference judgement related to the question: «*choose which one is less annoying*». Since two noises were presented to the subject, and since the equal condition was not admitted, the result is simply an «*one or the other*» answer. In order to avoid biasing of the results a great importance is given to the order of stimuli's presentation. For this reason a random extraction of signals in not satisfying, so we preferred to use an optimized sequence [3].

This kind of test naturally leads to an ordinal scale. So the problem arises of how to verify that the assumption of existing an underlying order. According to the theory of measurement [4], the empirical relation to be tested should be order. Order implies transitivities, so if *a*, *b* and *c* are three sounds, and "✳" is the empirical weak order relation, i.e. means "*sound a is not preferred to sound b*", transitivity implies:

$$a \precsim b \;\;\&\;\; b \precsim c \rightarrow a \precsim c \qquad (1)$$

Transitivity is usually the most critical one. It means that every judge in every moment will judge the triple *a*, *b* and *c* in accordance with it. Surely it can not be interpreted in a deterministic sense, i.e. as a property applying to each couple of objects, in any observation, by any subject. It may be rather assumed to hold in mean, which may be formally expressed by the weak probabilistic transitivity assumption. That is, if $\mathrm{P}(\mathrm{a} \precsim \mathrm{b})$ is the probability of *a* being not preferred to *b*, it may be shown that the assumption of an order scale is justified whenever the following weak probabilistic transitivity condition holds:

$$P(a \precsim b) \geq 0{,}5 \;\;\&\;\; \mathrm{P}(b \precsim c) \geq 0{,}5 \rightarrow P(a \precsim c) \geq 0{,}5 \quad (2)$$

So this condition may actually be checked on the preference matrix and provides a very powerful consistency test. The so–called preference matrix contains the symbol '✳' if the signals (labelled by letters) in the row *i* is preferred to the signal in the column *j*.

TABLE I. Preference matrix

	A	B	C	D	E	F
A	–	✳	✳	✳	✳	✳
B	✳	–	✳	✳	✳	✳
C	✳	✳	–	✳	✳	✳
D	✳	✳	✳	–	=	✳
E	✳	✳	✳	=	–	✳
F	✳	✳	✳	✳	✳	–

The intransitivity condition is graphically depicted by a matrix in which the upper and lower triangles are filled with coherent symbols, as shown in particular in table II. Since the matrix in table II has a type of antisymmetry, the portion below the main diagonal is superfluous and has been suppressed. In this example there is a transitive violation due to a equivalent relation (indifference condition) between the signal *E* and *D*.

TABLE II. Ordered preference matrix

	D	E	F	A	B	C
D	–	=	✳	✳	✳	✳
E		–	✳	✳	✳	✳
F			–	✳	✳	✳
A				–	✳	✳
B					–	✳
C						–

4.2. Magnitude estimation

It is an experimental technique used to quickly and easily determine how much of a given sensation a person is having. A set of stimuli (in this case six noise signals) are presented to the judges and are asked them to assign each of the stimuli a number relative to perceived quantity. This test, according to the way it is actually performed, may lead to an interval or to a ratio scale. A Matlab® program manages the audio reproduction and the users responses are gathered through a graphical user interface, as depicted in figure. This gives the possibility to the subject to interact with a graphical user interface on the screen, on which several buttons correspond to the different stimuli to be investigated. By clicking on the button it is possible to listen to the stimulus and then each button can be moved from left to right, on the panel, according to the given judgement.

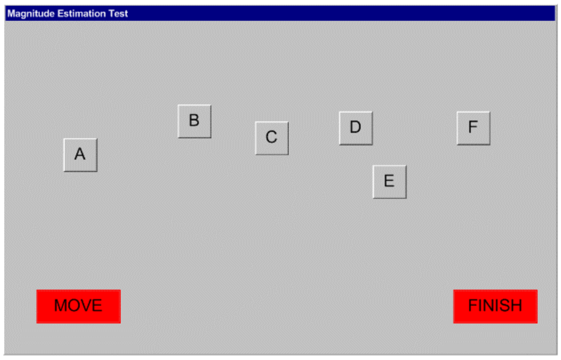

Figure 2. Graphical interface for the magnitude estimation test

This method gives the possibility to process the results according to an ordinal scale (only position is considered) or to an interval scale (position and score are considered).

The test results are arranged as a set of scores for each signal, as large as the number of subjects constituting the jury. Since each subject has a different approach to the scale it is necessary to establish a common reference in order to process the results all together. We decided to proceed in the simplest way by normalising the set of six scores of each subject in a 0–100 scale. Then it is possible to establish the mean score over the jury and the corresponding standard deviation for each signal.

Table III shows the scores (and relative measurement scale) obtained by the test.

TABLE III. Magnitude estimation scores

	A	B	C	D	E	F
Mean score	49	72	81	17	21	46
Mean std dev	4	4	5	4	4	4

5. EXPERIMENTAL RESULTS AND VALIDATION

The two test were performed by a 40 people jury, corresponding to the listening of more than 700 couples (or listening about 240 times to a single sound), and to 40 scores for each sound. As seen in the previous paragraphs, the results of the pair-comparison test, if the measurability conditions are fulfilled, provide a reference ordinal scale, while the magnitude estimation test, in the way it was implemented in the present experiment, may lead to the construction of an interval scale.

In an ordinal scale only the order is represented, while in the interval scale also the "*distance*" between two different samples is represented. In this way the interval scale is said to be "stronger" than the ordinal one and consequently, if feasible, it is preferable, since it provides more information.

5.1 Validation

When a such kind of tests are performed a very important aspect is given to the validation. In the present paper a cross validation is presented: if both tests (paired comparison and magnitude estimation) are performed over the same set of samples, it may be checked whether the order outcoming from the paired comparison test is in agreement with the order outcoming from the magnitude estimation test (figure 3).

It should be noted that although the order coming from the pair comparison test may be validated through the probabilistic consistency condition (2), such validation is not possible for the order outcoming from the magnitude estimation test. So cross validation of these two tests actually carries an added value in gaining confidence on the results.

Moreover there are important differences in the cognitive processes underlying the judgement: in one case there is a direct comparison of two stimuli, which is a rather easy duty for a common subject. In the second there is the assessment of all the available judgement space (the complete set of stimuli to be judged is available to the subject in every moment) and then the expression of a score for each stimuli or sound. In general this is felt as a more complex duty by the common subject. For these reasons the two performed methods can be identified as independent

In this sense an ordinal scale for such a quality will be a robust result as obtained following two different and independent methods.

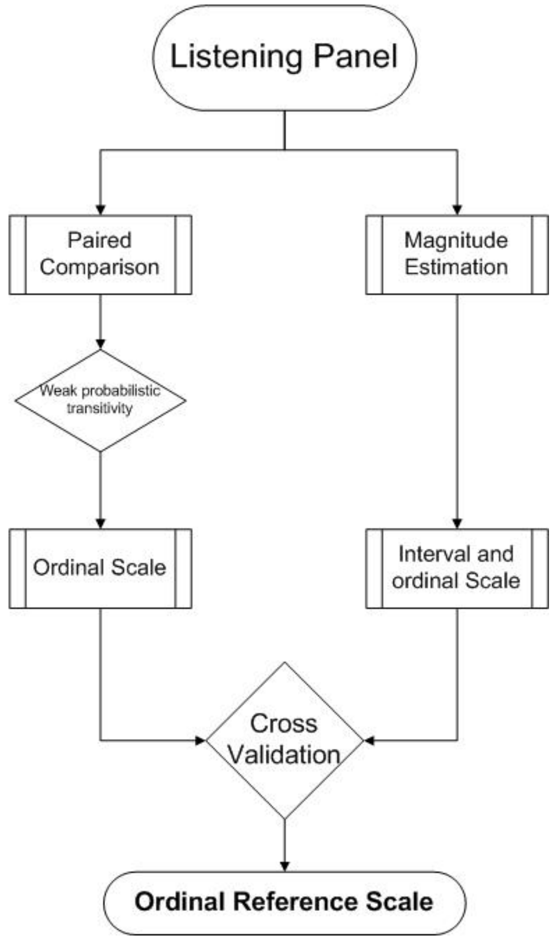

Figure 3. Cross validation method for ordinal scales

5.2. Results

After proper processing of the results from both tests, the scales obtained were compared. A graphical comparison is possible by plotting the scores from the magnitude estimation versus the order obtained from the paired comparison: if the scores are monotonically increasing, by increasing in the order, then there is correspondence in the order obtained in the two tests. This graph is presented in figure 4. By it is also possible to verify that the equivalence between signals *E* and *D* obtained from the paired comparison, is confirmed by two almost equal scores and by the overlap of the corresponding standard deviations. Although the disagreement is not so critical (Spearman rank–order correlation coefficient equal to 0.94), this part of the scale needs a deeper investigation (a border effect may have had some influence).

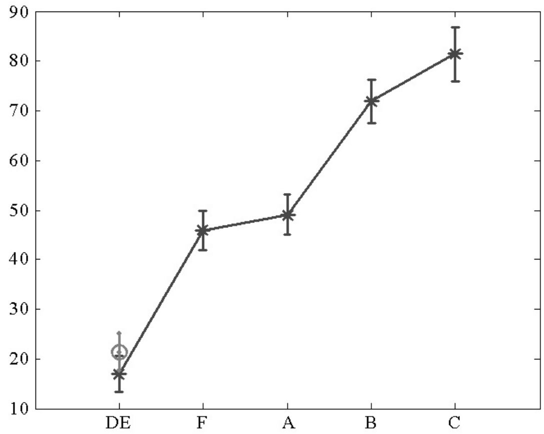

Figure 4. Comparison of the results from the two test methods

6. CONCLUSIONS

A sound quality evaluation by jury testing has started in the field of post sorting machines with a particular attention toward the operator acoustic working conditions. A test procedure for test execution and processing of the results has been designed and applied to post sorting machine sound, previously recorded. At present, no universally accepted procedure is available, so for the construction of a scale of reference, two fundamental kinds of tests have been considered, paired comparisons and magnitude estimation, pointing out their complementary characteristics. A validation technique has, also, been proposed based on the checking of assumptions and a cross validation of the two tests measured (at the moment on a merely ordinal scale level). The results seems to be encouraging, since the validation procedure has proved to be sensitive and informative and the final measurement procedure seems to be acceptable in terms of efficiency and uncertainty, for the kind of measurement considered. Preliminary results confirm the effectiveness of the test procedure and identify a precise machine component as the most annoyance giving the possibility to improve the ergonomics of the working environment.

Future work will regard further investigation on the general noises and evaluation of their psychoacoustic parameters in order to attempt a characterization of the investigated annoyance on the basis of a physical description of the noises.

REFERENCE

[1] Belotti V., Rossi G.B, Crenna F., “Measurement of perceived noise quality for the ergonomics of post-sorting machines”, *IMEKO TC 18 Congress*, Sapporo, 2001.

[2] Crenna F., Panero M., Rossi G.B., Weber R., “Measurement of the noise quality of post sorting machines by jury testing”, *ICA Congress*, Rome 2001.

[3] Wherry R.J., “Orders for presentation of pairs in the method of paired comparisons”, *J. of Experimental Psychology*, vol. 23, 651–660 (1938).

[4] Roberts F. S., “Measurement theory”, *Addison Wesley*, 1979.

[5] Luce R. D., Bush R. R., Galanter E., “Handbook of Mathematical Psychology”, *John Wiley and Sons*, 1963

[6] Purghé F., “Methods of psycho-physics and uni–dimensional scaling” (in Italian), *Bollati Boringhieri*, 1999.

[7] Lyon R. H., “Designing for product sound quality”, *M. Dekker*, 2000.

[8] Zwicker E., Fastl H., “Psycho-acoustics“, *Springer Verlag*, 1999.

[9] Yang S. J., Ellison A. J., “Machinery noise measurement”, *Oxford University Press*, 1985.

[10] Codda M., Crenna F., Rossi G.B., “Jury test methods for the measurement of perceived quantities”, *IMEKO TC 7 Congress*, Cracow, 2002.

Authors: Francesco Crenna, Matteo Panero, Giovanni B Rossi University of Genova, DIMEC, Dept. of Mechanics and Machine Design, Via all’Opera Pia 15A, I-16145, Genova, ITALY, phone: +39 010 353 2232, fax: +39 010 353 2834, E-mail: gb.rossi@dimec.unige.it