COMPARISON OF PSYCHOACOUSTIC PARAMETERS AND DESCRIPTIVE WORDS OF ENVIRONMENTAL SOUND PERCEPTION

R. Mariconte[1*] **G. Brambilla**[2] **D. Annesi**[3]
M. Diano[4] **F. Lo Castro**[5] **C. Giliberti**[1]

[1] INAIL, DIT, Roma, Italy
[2] University of Milano-Bicocca, DISAT, Milano, Italy
[3] INAIL, DiMEILA, Monte Porzio Catone (RM), Italy
[4] INAIL, Regional Directorates Calabria, Catanzaro, Italy
[5] CNR-INM, Section of Acoustic and Sensors, Roma, Italy

ABSTRACT

Psychoacoustic parameters, being closely related to sound perception, are usually applied in product sound quality and, recently, also in environmental soundscape analysis or at workplace, to investigate its potential in describing acoustic comfort.

Lexicons of descriptive words of perceptual sound attributes are available in literature, but the language is often a crucial issue.

This paper describes two experiments dealing with such words in Italian and the evaluation of their association with psychoacoustic parameters. For these experiments, 12 sounds recorded in three different environments (at workplace, in nature and in the community) were selected and processed to determine some psychoacoustic parameters. These sounds were randomly played in a quiet room at the same equivalent level L_{eq} (dB) by headphone to a group of subjects. Multidimensional scaling and correlation have been applied to compare their responses with some acoustic and psychoacoustic descriptors.

Keywords: *psychoacoustics, sound quality, environmental noise, sound perception, descriptive words.*

[*]***Corresponding author***: *r.mariconte@inail.it*

1. INTRODUCTION

Descriptive words for sound perception are largely context-dependent, and most of them have no clear relationship to properties that acousticians know how to measure. Furthermore, the variability in sound expertise among the participants is typically large, from sound experts (acousticians, sound engineers and sound designers) to non-experts (consumers, naive participants). The application field involved is also different, from the assessment of sound reproduction quality to product sound quality. Another important issue is the language, since the translation from other languages does not always keep the same concept and meaning.

Lexicons of descriptive words of perceptual sound attributes are available in literature (e.g., [1]), but only one has been retrieved in Italian, with words selected from Italian Web Corpus 2016®, and concerning the perception of sounds in areas surrounding ports [2].

The present study is a preliminary investigation on Italian descriptive words of perceptual attributes of environmental sounds and on the evaluation of their association with acoustic and psychoacoustic parameters. Two listening experiments have been carried out using 12 sounds, recorded in three different environments, namely at workplace, in nature and in the community. The subjective responses collected at the listening tests have been compared with some acoustic and psychoacoustic descriptors by means of multidimensional scaling and correlation.

2. MATERIALS AND METHODS

The 12 sound stimuli used in both listening experiments were recorded in three different environments, namely at workplace, in nature and in the community (Tab. 1). Some of them were binaural, others monoaural and, therefore, without any spatial cues. For homogeneity purpose, they were processed into monaural track and normalized at the same -30 dB rms. Afterwards, they were imported in ArtemiS SUITE v14.1 with a 0.1 s time resolution and to determine various acoustic descriptors and six psychoacoustic parameters [3], namely loudness [4], sharpness [5], roughness [6], fluctuation strength [7], tonality [8] and impulsivity [9].

Table 1. Sound stimuli and acoustic descriptors.

Environment	Sound	Descriptors
Work (W)	1 Keyboard typing	L_{eq} [dB]
	2 Weaving loom	L_{Aeq} [dBA]
	3 Big diesel engine	Dev. st. s_{LA} [dBA]
	4 Fan	N5 [sone GF]
Nature (N)	1 Seagulls	S_{avg} [acum]
	2 River	R^{*}_{avg} [asper]
	3 Rain	F_{avg} [Vacil]
	4 Sea waves	T^{*}_{avg} [tu]
Community (C)	1 Outdoor market	I^{*}_{avg} [iu]
	2 Indoor metro	1/3 oct. spectrum
	3 Urban square	centre gravity G
	4 Urban street	[Hz]

* Sottek Hearing Model HMS [10]

The sounds were randomly played at the same equivalent level L_{eq} (dB) by binaural headphone to participants, tested one at a time.
A lightweight and portable sound reproduction system has been calibrated (Fig. 1) and used to perform the listening test in any quiet room in order to get a reasonable number of participants in short time. To provide a good listening quality, the system was formed by a digital audio player Creative Zen connected to a semi-closed circumaural headphone AKG K 44.

Figure 1. Calibration of the lightweight and portable sound reproduction system used for the listening tests.

The 12 sounds were presented diotically by binaural headphone in a random order to minimize the bias on responses due to the presentation order [11]. The test room was quiet without any significant sound interfering with the listening. The listening level was fixed and previously calibrated on a head and torso simulator (Fig. 1). Each sound was one shot played for 10 s and, on participant request, loop listening until the questionnaire was filled in was available.
In the selection of participants, those with high education were preferred, taking into account the task to be performed, requiring a wide knowledge of the language suitable to run the session without too much assistance of the experimenter.

2.1 Experiment 1

In the questionnaire the participant was asked to select in a list of 22 words those considered most appropriate to describe her/his perception of the sound just heard. The option to indicate other words not present in the list was also available. In choosing the 22 descriptive words of sound perception, the outcome of the study in [2] was taken into account.
Twenty four subjects (average age 40 ± 14 years) participated to this experiment, 50% male and 54% with degree or PhD education. They self-reported an average noise sensitivity of 6.5 (± 1.8) on a scale from 1 (not at all) to 10 (very much). The average duration of the listen sessions was 9 minutes and 30 s, with an average response time of 30 s for the choice of each attribute. Notwithstanding the possibility to indicate other attributes not in the proposed list, this option was chosen only by 8 (33.3%) participants for a total of 9 descriptive words.

2.2 Experiment 2

For each sound stimulus the participant was asked to rate on seven semantic bipolar scales her/his sound perception. The perceptual attributes were chosen taking into account the outcome of experiment 1 as follows:

- Scale S1: *Unknown* vs. *Known*;
- Scale S2: *Dull* vs. *Hissing*;
- Scale S3: *Ugly* vs. *Nice*;
- Scale S4: *Steady* vs. *Fluctuating*;
- Scale S5: *Boring* vs. *Lively*;
- Scale S6: *Pleasant* vs. *Annoying* (reverse coding);
- Scale S7: *Blur* vs. *Clear*

The above attributes are here reported in English and efforts have been made to select those having meanings as close as

possible to that they have in Italian, the language used in the experiment.
Each bipolar scale had 7 points, with middle point corresponding to the neutral "*Neither/Nor*" rating (Fig. 2).

Attribute 1	Much	Fairly	A little	Neither /Nor	A little	Fairly	Much	Attribute 2

Figure 2. Semantic bipolar scale used in experiment 2.

Twelve subjects (average age 42 ± 15 years) participated to this experiment, 50% male and 58% with degree or PhD education. None of these participants took part in the experiment 1. They self-reported an average noise sensitivity of 7.3 (± 2.2) on a scale from 1 (not at all) to 10 (very much).
The average duration of the listen sessions was 9 minutes and 45 s with an average response time of 1 minute and 25 s for each sound.

3. RESULTS AND DISCUSSION

The similarity of sounds in terms of their acoustic descriptors (Tab. 1), with standardized values (zero mean and unit variance) because measured on different scales, has been visualized by the Multidimensional Scaling (MDS). The pairwise euclidean distances among the data set, plotted into a cartesian space in Fig. 3, show a good discrimination among the 12 sound stimuli not only for the environment where they have been recorded and their semantic meaning, but also for their acoustic descriptors.

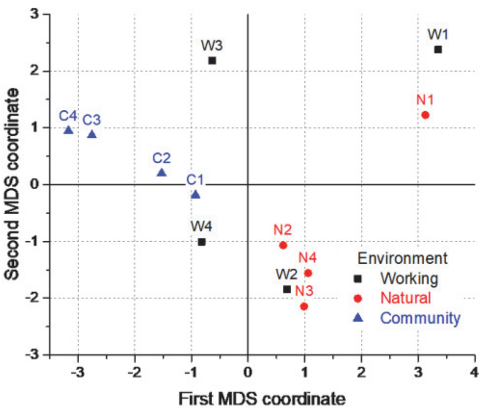

Figure 3. Multidimensional Scaling of the 12 sound stimuli based on their acoustic descriptors (Tab. 1).

For instance, Figure 4 shows the spread of the L_{Aeq} levels of the sound stimuli versus their centre of gravity G of the 1/3 octave band spectrum.

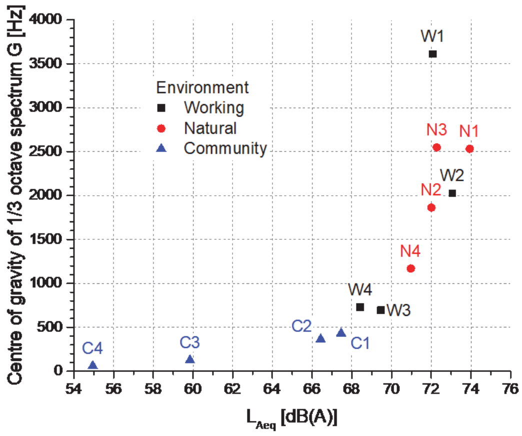

Figure 4. Spread of L_{Aeq} levels versus the centre of gravity G of the 1/3 octave band spectrum.

3.1 Experiment 1

As shown in Fig. 5, the sounds were rather familiar or known to the participants (19.4% of all the 1159 selections) and, therefore, it is likely that the chosen attributes were the outcome of an aware selection rather than a random one. The sounds recorded in the working environment were mainly described as *Intrusive* (82.4% of all the 51 responses on this attribute), *Ugly* (73.7% of all the 57 responses on this attribute) and *Annoying* (70.4% of all the 71 responses on this attribute). Natural sounds were described as *Nice* (73.8% of all the 65 selections reported for this attribute) and *Pleasant* (71.9%). Community sounds were described mainly *Blur* (56.7% of all the 67 selections reported for this attribute) and *Fluctuating* (54.0%), as well as more *Lively* (16.1%) than *Annoying* (14.1%).

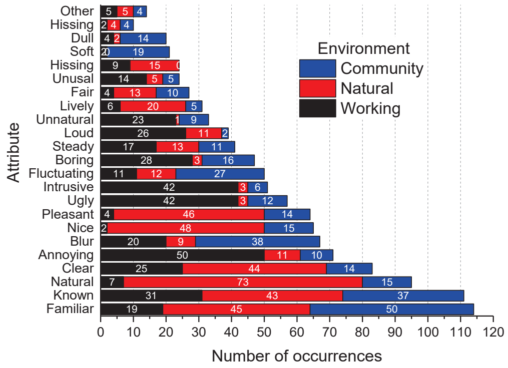

Figure 5. Number of occurrences of the selected attributes.

The Spearman rank correlation matrix (Fig. 6) shows that a sound described as *Natural* is positively correlated with *Pleasant* (ρ = 0.9) and *Lively* (ρ = 0.8), whereas ρ is negative for *Ugly* (-0.8) and *Boring* (-0.7). The correlation points out also the correspondence among synonyms, such as *Nice* and *Pleasant* (ρ = 0.9).

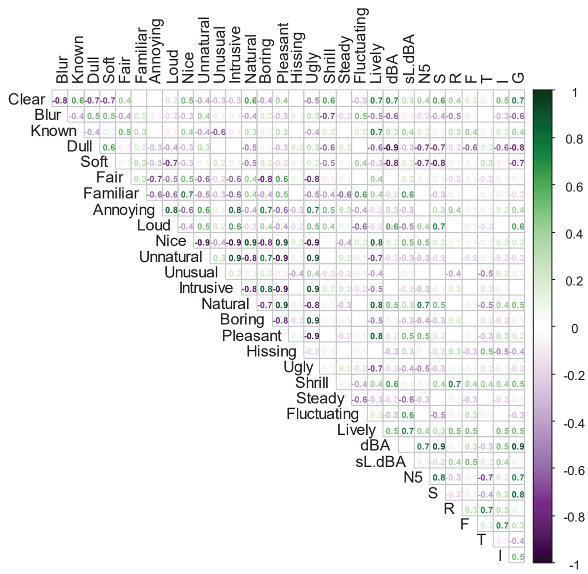

Figure 6. Spearman rank correlation matrix between attributes and sound descriptors.

3.2 Experiment 2

Figure 7 reports the percentages of occurrences observed for each bipolar scale pooling all the responses to all sounds. As in experiment 1, most of the stimuli were much (59.7%) and fairly (20.1%) known for the participants and they reported the lowest percentage (2.8%) of neutral response *Neither/Nor*. This result suggests that their ratings on the other scales were likely the outcome of an aware choice rather than a random one.

The highest percentage of neutral response *Neither/Nor* (20.8%) was observed for the scale *Boring/Lively*, whereas the sounds were, overall, more often *Annoying* (52.8%) than *Pleasant* (36.8%).

More hints may be drawn considering the mean scores on the seven scales S versus the three different sound recording environment (Fig. 8). It is clear that the perception of natural sounds was rated more positively than those in the working environment, whereas the sounds in the community is rated in-between.

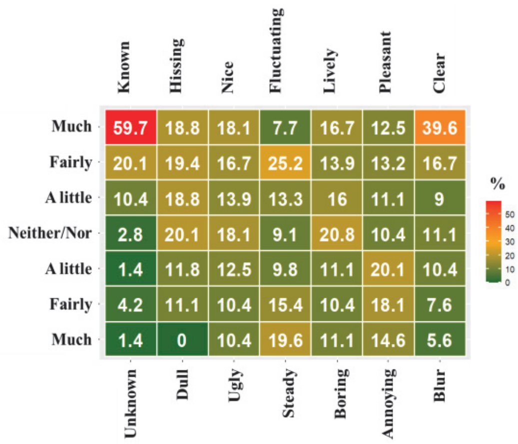

Figure 7. Percentages of occurrences for each bipolar scale pooling all the responses to all sounds.

A more detailed analysis has been performed for the sounds recorded in the same environment, such as that reported in Figure 9 for the natural sounds. The highest overall mean of the scores (1.8) was observed for the seagull sound and the lowest one (0.9) for rain. Only 11% of all the mean scores corresponded to slightly negative attributes, confirming that people enjoy the perception of natural sounds.

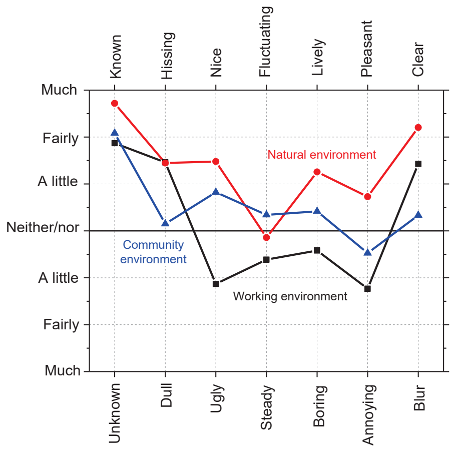

Figure 8. Mean score for each bipolar scale and the sound recording environment.

Moreover, the Spearman rank correlation matrix (Fig. 10) was computed considering the mean scores obtained on the bipolar scales (S1,…, S7) for each sound and the corresponding acoustic parameters listed in Table 1.

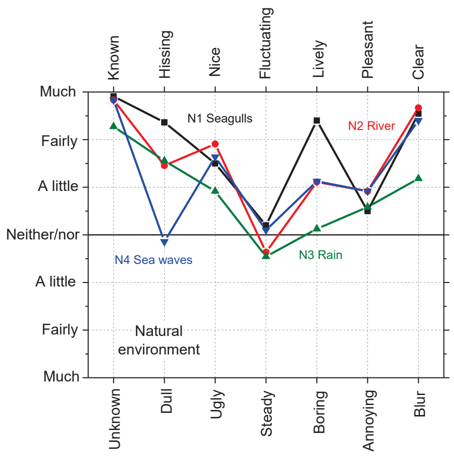

Figure 8. Mean score for each bipolar scale for natural sounds.

The variables are displayed according to their hierarchical clustering order (Ward agglomeration method) considering two groups, highlighted by the two rectangles with blue borders.

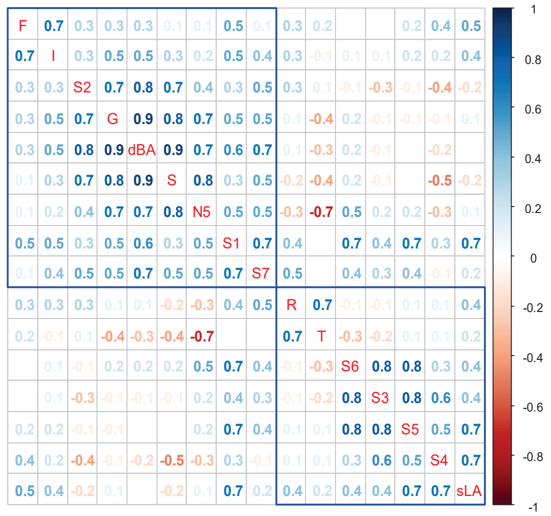

Figure 9. Spearman rank correlation matrix between mean scores on semantic scales S and sound descriptors.

In the top-left group, scale *S2 Dull/Hissing* is positively correlated with the L_{Aeq} level (ρ = 0.8), the spectrum centre of gravity G and Sharpness S (ρ = 0.7). Scale *S1 Unknown/Known* is correlated with scale *S7 Blur/Clear* (ρ = 0.7), and this scale is somewhat correlated with roughness R (ρ = 0.5). High correlation (ρ = 0.9) is observed between G and L_{Aeq} level, as also already shown in Figure 3.

In the bottom-right group, scale *S3 Ugly/Nice* is positively correlated (ρ = 0.8) with *S6 Annoying/Pleasant* and *S5 Boring/Lively*, whereas scale *S4 Steady/Fluctuating* is correlated with the sound level standard deviation sL_A (ρ = 0.7).

The above results show that the set of acoustic descriptors considered correspond satisfactorily to the descriptive words of sound perception as reported by participants in this study, and they can quantify the various perceptual dimensions (e.g., time and frequency patterns, semantic content).

4. CONCLUSIONS

This study is a preliminary effort in the view to create a lexicon of Italian descriptive words of the perception of environmental sounds.

Even though the results cannot be generalized, being the sounds limited to three different contexts only, they show that the set of acoustic descriptors considered correspond

satisfactorily to the perceptual sound attributes selected and quantified by the participants to describe the various perceptual dimensions.
Further investigations are planned to enlarge the sounds under test and the listening panel, to improve the perceptual descriptive words set and to apply further statistical analyses.

5. REFERENCES

[1] T.H. Pedersen, "Lexicon of Sound-Describing Words – Version 1", Delta Report AV 11/05 THP, 2008.

[2] A. Magrini, G. Di Feo and A. Cerniglia, "Questionnaire analysis survey for acoustic investigation-Preliminary considerations", in *Proc. 48th InterNoise 2019,* (Madrid, Spain), 2019.

[3] H. Fastl and E. Zwicker: *Psychoacoustics. Facts and Models*. Springer Berlin, Heidelberg, 2007.

[4] DIN 45631/A1: 2010-03*: Calculation of loudness level and loudness from the sound spectrum - Zwicker method - Amendment 1: Calculation of the loudness of time-variant sound*.

[5] DIN 45692: 2009-08: *Measurement technique for the simulation of the auditory sensation of sharpness*.

[6] R. Sottek, J. Becker and T. Lobato "Progress in Roughness Calculation" in *Proc. of InterNoise 2020* (Seoul, South Korea), 2020.

[7] H. Fastl and E. Zwicker: *Psychoacoustics. Facts and Models*. Chap. 10, Springer Berlin, Heidelberg, 2007.

[8] ECMA-418-2: *Psychoacoustic metrics for ITT equipment — Part 2 (models based on human perception)*, 2020.

[9] R. Sottek and T. Moll "Loudness perception and modeling of impulsive sounds" in *Proc. of Euronoise 2015* (Maastricht, Netherlands), 2015.

[10] R. Sottek and K. Genuit: "Models of signal processing in human hearing", *AEU - International Journal of Electronics and Communications,* 59(3), pp. 157-165, 2005.

[11] R. Zeelenberg and D. Pecher: "A method for simultaneously counterbalancing condition order and assignment of stimulus materials to conditions", *Behav. Res.*, 47 (127), pp. 127-133, 2015.