# The use of semantic differential scales in listening tests: a comparison between context and laboratory test conditions for the rolling sounds of office chairs

Doriana Dal Palù[a] *, Eleonora Buiatti[a], Giuseppina Emma Puglisi[b], Olivier Houix[c], Patrick Susini[c], Claudia De Giorgi[a], Arianna Astolfi[b]

[a] DAD – Department of Architecture and Design, Politecnico di Torino, Viale Mattioli 39, 10125 Torino, Italy.
[b] DENERG – Department of Energy, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy.
[c] IRCAM – Équipe Perception et Design Sonores, IRCAM – Institut de Recherche et Coordination Acoustique/Musique, Place Igor Stravinsky 1, 75004 Paris, France.

**Abstract**
Semantic differentials are frequently used to investigate sounds from a subjective point of view. The application of semantic differentials to the case of the rolling sound of office chairs is dealt with in this study. After a preliminary selection of the semantic differentials by fifty-two participants, another ninety participants took part in a listening test and described the acoustic stimuli of two office chairs, of high and low quality, respectively, rolling over polyvinyl chloride (PVC), ceramic and wood floorings, under context and laboratory test conditions. Under laboratory condition, recorded real stimuli were presented to the listeners via headphones, or under SounBe condition. SounBe is a new tool that has recently been conceived to explore sound at an early design stage. With this method, interactions between a chair and the floor are simplified, a mechanical sound is produced of a wheel moving across a flooring tile, and the recorded stimuli are then presented to the listeners through headphones. Four 7-point Likert scale semantic differentials, related to calmness, roughness, pleasantness and annoyance, were used to collect subjective data. Objective data were instead obtained from psychoacoustic indexes. Factors such as gender and weight were found to have no effect on the subjective and objective data. The flooring factor instead resulted to have much more influence than the chair factor. No statistically significant difference was observed between the test conditions on the semantic differential scales, thus proving the compatibility between SounBe and real sounds.

**Keywords**
Listening test; semantic differentials; context condition; laboratory condition; rolling sound; sound design

**Highlights**

- The flooring factor has a greater influence on rolling sounds than the chair factor.
- Semantic differentials are comparable in context and laboratory listening tests.
- SounBe can be used with semantic differentials to investigate sound stimuli.
- SounBe allows a product sound to be assessed and the prototyping phase to be avoided.

* *corresponding author*: Dr. Doriana DAL PALÙ
*postal address*: MATto - Materioteca del Politecnico di Torino
c/o Cittadella Politecnica del Design e della Mobilità
Corso Settembrini 178
10135 Torino, Italy
*e-mail address*: doriana.dalpalu@polito.it

## 1. Introduction[1]

Nowadays, the overall sound quality of a product is a key-factor in its acceptance by consumers. Sound quality is assessed considering several factors, connected to both objective and subjective aspects. Apart from the traditional sound measurements based on the physical description of sounds (e.g. sound wave, frequency, amplitude parameters), objective aspects, such as the psychoacoustic indexes [Fastl & Zwicker, 2007], are often involved in perceptive investigations. However, subjective aspects are linked above all to the experience and the expectations of the users towards the products [Schifferstein & Hekkert, 2008], to the identification and recognition of the sound [Cahen, 2015], to emotional factors, and to many others. The semantic differential technique [Osgood, 1952] applied to sound perception [Solomon, 1958] still seems to be one of the most frequently used methods to investigate sounds from a subjective point of view in different situations [Schütte, Müller, Sandrock, Griefahn, Lavandier & Barbot, 2009; [illegible], Han & Uchida, 2016]. Thus, both context [Gaver, 1993] and laboratory listening test conditions are generally considered [Jeon, You & Chang, 2007; Steffens, Schulte-Fortkamp & Becker-Schweitzer, 2011]. Nevertheless, the debate on the validity and plausibility of listening tests conducted in laboratory conditions, compared with those performed in context conditions, is still open [Steffens, 2013; Lindau & Weinzierl, 2011].

### 1.1 Testing the perception of a product sound

Several methods [Lyon, 2003; Özcan & Van Egmond, 2004] and tools have been developed over the last few decades to assess and predict the human perception, acceptance and emotion towards a product sound, as well as to support the design phase. A tool with an abstract shape was developed and tested to assess sonic feedback in tangible interfaces (i.e. a digital environment) [Lemaitre, Houix, Visell, Franinović, Misdariis & Susini, 2009]. Environmental sound categorization was found to be reliable as a sound design tool [Houix, Lemaitre, Misdariis, Susini & Urdapilleta, 2012], and voice is also currently used by sound designers to simulate (i.e. reproduce) a product sound before, and then to design it later [Lemaitre & Rocchesso, 2014].
In 2011, a new toolandmethod, called SounBe, was conceived and patented to support architects and designers in the delicate meta-projectual phase of choosing the best material for an object, taking into account the sound aspect as a fundamental project requirement [De Giorgi, Astolfi, Buiatti, Lerma, Arato & Dal Palù, 2011], as well as to assess a product sound quality. SounBe is a physical toolkit that is kept in a suitcase. It consists of a variety of instruments that are used to “sound” material samples and products (i.e. sticks and resting planes in different materials, a measuring cup, some support bars, etc.), and allows collecting and resubmitting sounds by a microphone and headphones; the method, i.e. a protocol, is conceived to analyse and design the sound of an object that interacts with another object, by a simplified procedure that splits the sound in its three generating variables: the material, the configuration form and the exciting mode interaction. Since the interactions are simplified, no prototype of the final product is required. It can therefore be applied in the early design phase of different product design contexts (e.g. for furniture design, packaging design, clothes design, etc.). In order to fit to different design contexts, the method has to be adapted case by case to the different sound sources to be reproduced and assessed. As an example, it considers the interaction between a wheel and a tile in order to investigate the sound of a trolley being dragged over a ceramic floor. Once a sound coming from the material-configuration and exciting mode interaction has been acquired and repeatably reproduced, the sound profile can be defined through a standardized descriptive procedure. It is well known that a specific and shared vocabulary is necessary to verbalise the characteristics of sounds [Houix, Lemaitre, Misdariis, Susini & Urdapilleta, 2012; Carron, Dubois, Misdariis, Talotte & Susini, 2014]. Semantic descriptors that define the sensorial recall produced by the sounds themselves are attributed by a testing panel (also called acoustic “tasters”, i.e. a group of experts, trained in acoustic sensorial analyses, who become the real judges of the perceptive characteristics of a stimulus). Following SounBe method, each sound is matched to the descriptor that has been judged the most suitable, and each sound-descriptor matching can be used by architects and designers as starting information on sound perception, and can be collected in a sound

[1] The following non-standard abbreviations are used in this paper:
C1 = context test condition; C2 = laboratory test condition; C3 = SounBe test condition; HLC = high level chair; LLC = low level chair; SD1 = calm-strident; SD2 = pleasant-unpleasant; SD3 = smooth-rough; SD4 = not annoying at all-very annoying.

database. By means of a keyword search in the database, it will then be possible for anyone to be able to forecast and consciously design the product sound.
Since these evaluations are possible, thanks to the new tool, and the prototyping phase can be avoided, the method represents a low cost and effective data collection opportunity [Dal Palù, De Giorgi, Astolfi, Lerma & Buiatti, 2014]. An experiment on the semantic differential technique applied to sounds obtained by this tool was considered appropriate, considering the widespread and easy use of the semantic differential tool.

### 1.2 A comparison between context and laboratory listening tests

This work considered two main issues. The first issue pertains to the comparability of the results of listening tests, with semantic differential scales, in context conditions and laboratory conditions. In fact, the correspondence between conditions represents a crucial point for a large number of studies [Steffens, 2013; Lindau & Weinzierl, 2011]. Many researchers have raised the problem of the "ecological validity" of laboratory experiments, questioning whether the perception of a reproduced sound or complex acoustic environment is the same or different from what might be expected on site [Guastavino, Katz, Polack, Levitin, & Dubois, 2005; Raimbault, 2006; Davies, Bruce, & Murphy, 2014; Turchet & Serafin, 2014]. Moreover, the possibility of reducing the investigations to laboratory listening tests with semantic differentials would simplify the product sound testing process to a great extent. The second issue pertains to the validity of the semantic differential scales to sounds generated with the new tool. The proved comparability of alternative methods for the sound quality investigation could represent another opportunity of forecasting the perception of a product sound, and of avoiding the prototyping phase. In fact, firms could reproduce the future sound with the new tool, and collect subjective data on perception by means of semantic differential listening tests.
The rolling sound of office chairs was selected as the stimulus for this experiment, because of its non-stationary and unpredictable nature [Astolfi & Pellerey, 2008], which has proved to negatively affect workers' comfort [Bodin Danielsson & Bodin, 2008; Saeki, Fujii, Yamaguchi & Harima, 2004]. Non-stationary noises have been investigated much less than stationary noises in the workspace [Kjellberg & Landström, 1994], and there is still a gap in literature concerning the rolling sound of office chairs. On the contrary, an extensive amount of literature already exists on the perception of outdoor rolling sounds, i.e. for vehicles [Soeta & Shimokura, 2013; Kasess, Noll, Majdak & Waubke, 2013; Ohiduzzaman, Sirin, Kassem & Rochat, 2016; Sirin, 2016]. Finally, increasing interest in furniture sound design in different living environments has been observed [De Rouvray, Bassereau, Duchamp, Schneider & Charbonneau, 2008; Alves, Filho, Silva & Câmara, 2012; Xie & Kang, 2012].
The experiment was carried out in three different test conditions: the context test condition (C1), the laboratory test condition (C2) and the SounBe test condition (C3). Subjective (qualitative) and objective (quantitative) data were collected during the experiment.

## 2. Materials and methods

The study was designed in two phases. A preliminary phase was carried out to select the most suitable semantic differential scales from scientific literature, in order to evaluate a rolling chair sound. Subsequently, in the main phase, the rolling chair sounds were evaluated by means of the selected semantic differentials scales.

### 2.1 Preliminary phase: The selection of the semantic differential scales

The Von Bismarck semantic differential scales were chosen as a reference [Von Bismarck, 1974]. Since it has been proved that the translation process can affect the connotation of a word [Namba, Kuwano, Hashimoto, Berglund, Zheng, Schick, Hoege & Florentine, 1991], each descriptor was first translated into Italian using general English-Italian and Italian-English bilingual dictionaries [Martignon-Burgholte & Cyffka, 2007; Ragazzini, 2015] and specific bilingual ones [Nicolao & Noselli, 2007]. Furthermore, each term was validated by a group of 3 English-Italian bilingual subjects, who picked the most suitable translation in Italian from among those proposed in the dictionaries. A summary of the original semantic differential pairs by Von Bismarck and the Italian translations is shown in Appendix B.

In the same way as in the case of Von Bismarck [Von Bismarck, 1974], a selection of semantic differentials pairs was necessary to reduce the number of pairs proposed in phase 2.2, and to avoid cognitive overloading and test annoyance [Brinkman, 2009]. Since the pre-selection made by the experimenter could have affected the results of the test to a great extent, and the descriptors may not have necessarily conformed with those a participant would spontaneously use [Susini, Lemaitre & McAdams, 2011], a first objective pre-selection was necessary.

#### 2.1.1 Participants

A group of 52 participants (28 women and 24 men, $\overline{x}$ = 37.7 years, $\sigma$ = 17.2 years) took part in this preliminary selection. All the participants were Italian, but from different regions. The group included both "experts", e.g. people who declared they had had physical acoustics and/or a formal musical education (No. = 20), and some "laymen" in these topics (No. = 32), in order to verify whether background knowledge affected the choice of vocabulary and the selection of the pairs.

#### 2.1.2 Questionnaire

The semantic differential pair selection was performed by means of a multiple-choice questionnaire. The participants were asked to select the Von Bismarck semantic descriptor pairs they considered to be the most appropriate to describe the rolling sounds of office chairs. The number of possible answers was not fixed, in order to verify how many couples of descriptors were chosen spontaneously. The pairs were presented randomly, and each pair was also presented in reverse order (e.g. *high-low* or *low-high*), to avoid order effects and possible sequence bias. Furthermore, each descriptor was provided with a short sentence taken from an Italian dictionary [Zingarelli, 2015] in which the descriptor was presented (e.g. "the *high* scream of the child woke up his brother" / "lo strillo *acuto* del bambino svegliò suo fratello") in order to have a better understanding of its meaning [Parizet & Nosulenko, 1999]. A summary of the short sentences presented for the semantic differential pairs selection is shown in Appendix B.

### 2.2 Main phase: The listening test in different test conditions

The rolling chair sounds were evaluated by means of the previously selected semantic differentials scales in the second, main experimental phase.

#### 2.2.1 Test conditions

The experiment was carried out in a dead room (reverberation time at 0.5-1 kHz equal to 0.1 s), under three different test conditions: the context test condition (C1), the laboratory test condition (C2) and the SounBe test condition (C3):

- In C1, the participants were asked to sit on real office chairs and move across paved platforms. The listening test was conducted here in an active listening condition. The participants were overall able to experience the chair and the flooring (not only from a listening point of view, but also visually, haptically, etc.);
- In C2, the participants were asked to listen to office chair rolling sounds that had previously been produced and were now delivered to them through headphones. In C2, the listening test was conducted in a passive listening condition. Furthermore, no visual information on the chair or on the flooring was provided to the participants;
- In C3, the participants were asked to give similar evaluations to sounds that had previously been produced, with a simplified procedure, through the use of the new tool [Dal Palù, De Giorgi, Astolfi, Lerma & Buiatti, 2014], and then delivered to them through headphones. In C3, as well as in C2, the listening test was conducted in a passive listening condition. No visual information on the chair or on the flooring was provided to the participants in this case either.

#### 2.2.2 Stimuli

The acoustic stimuli produced by office chairs rolling on different floorings were used for the present study. Two operative chairs, which are representative of the typical office chairs used in Italian offices [Centro Studi

Cosmit/FederlegnoArredo, 2015] and that comply with the UNI EN 1335 norm, were chosen. The chairs were chosen to represent a high-quality office chair (here called HLC) and a low quality one (here called LLC), respectively, on the basis of their construction characteristics and their selling price (Figure 1). Three walkable platforms (2500 x 1250 mm of walkable surface), representing some of the most common paving systems in the Italian workspace landscape, that is polyvinyl chloride (PVC), ceramics and wood, were prepared *ad hoc* (Figure 1). Just one wheel per chair and three flooring items for each material were used for the stimuli considered in C3 (Figure 1).

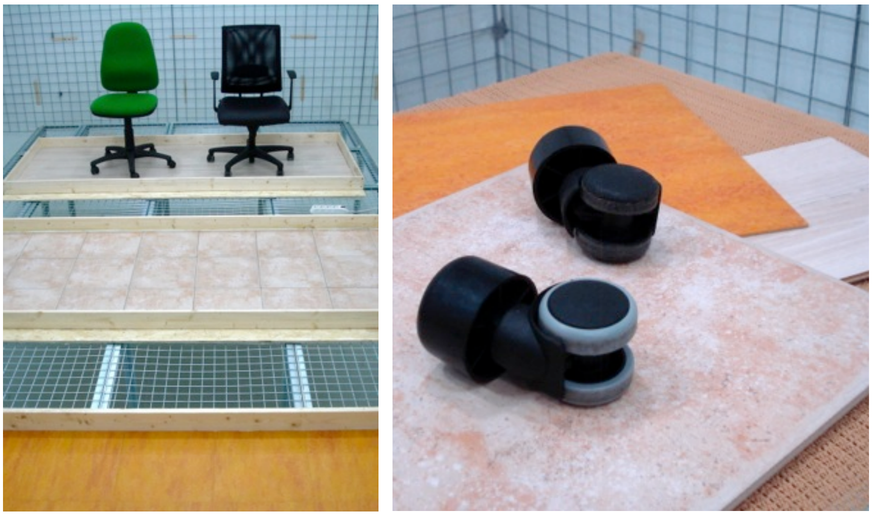
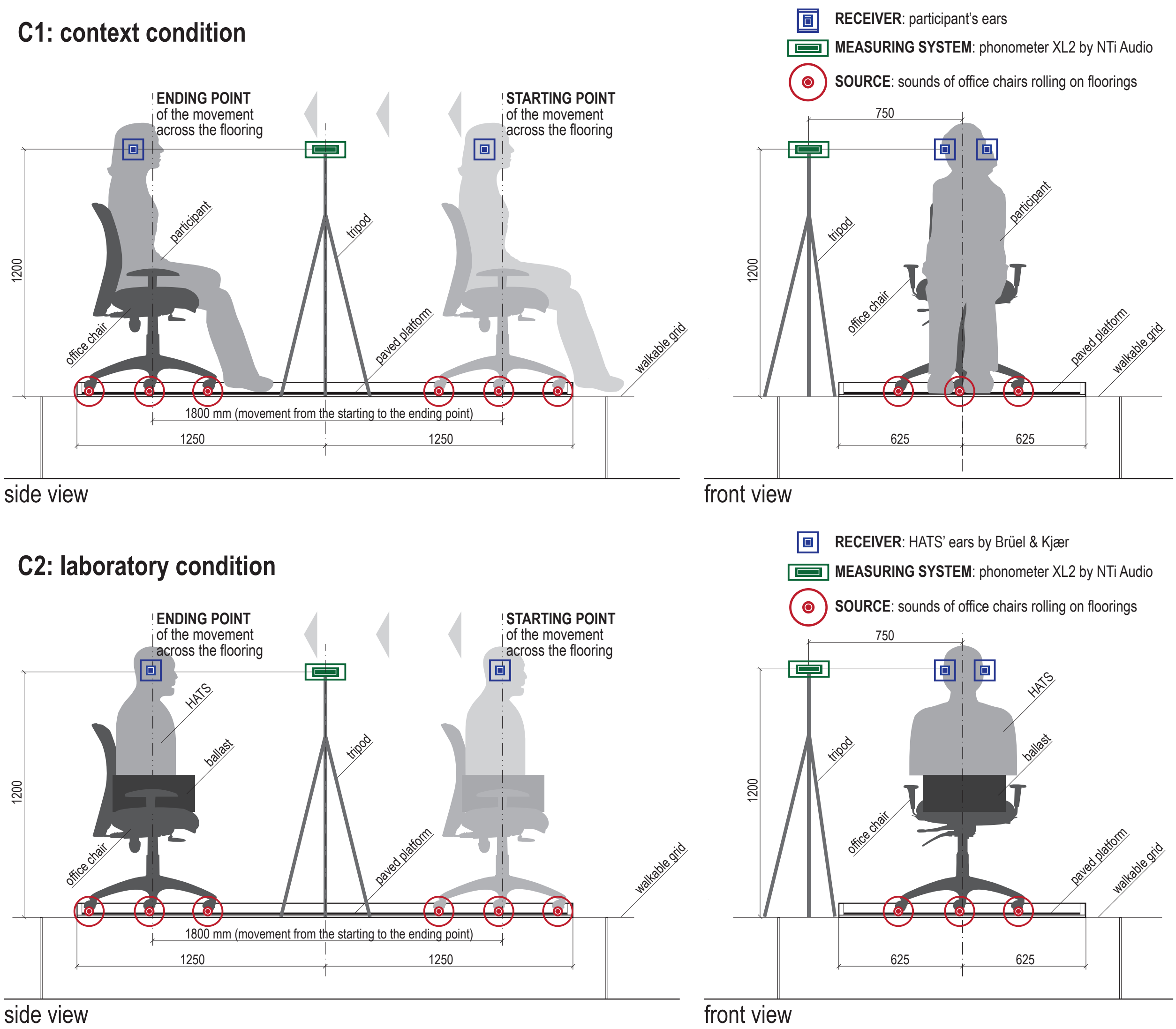


**Figure 1**. The materials tested in the dead room in the C1 and C2 tests are shown in the picture on the left: LLC and HLC on the left and right, respectively, in the upper part of the image; the three paved platforms, provided with the wooden flooring (at the top), the ceramic flooring (in the middle) and the PVC flooring (at the bottom), were placed on top of the floating grid. The materials (ceramic, PVC and wood floorings) tested in the dead room in C3 are shown in the picture on the right: both the LLC wheel and the HLC wheel were provided with handles.

All the stimuli were produced and the tests were performed in an dead room, where the background noise level was lower than $L_{Aeq}$ of 20 dB(A). The same set of stimuli was used for the test. Nevertheless, the acquisition and submission of the sounds varied according to the test conditions:

- In C1, two office chairs were placed at the same height on three paved platforms arranged on the walkable grid in the dead room. The stimuli were produced by the participants actively moving their chairs across the floorings. Each stimulus was measured by means of a previously calibrated phonometer (XL2 handheld audio and acoustic analyser made by NTi Audio), arranged at a height of 120 cm from the paved platform surface (corresponding to the ear height of an office seated worker, referring to the 95th percentile of the European population [Pheasant & Haslegrave, 2005]), at the halfway point of the length of each walkable platform (Figure 2);
- In C2, the submitted stimuli were produced by a chair ballasted with the total reference weight of 75 kg (corresponding to the standard weight required by the EN 1335-3 norm). They were recorded binaurally using an artificial head (Head and Torso Simulator, HATS, 4128C by Brüel & Kjær) "sitting" on the chair, i.e. simulating the body of the participant, and measured through a calibrated phonometer. The recording and measuring procedures were performed with the same scheme, as far as the reciprocal position between the sound source, the receivers and the phonometer is concerned, as in C1 (Figure 2). Since these stimuli were submitted by means of headphones (HD600 headphones by Sennheiser), each acoustic signal recorded with HATS (both for the left and the right channels) was convolved with a Kirkeby inverse filter, in order to eliminate any eardrum or headphones effects. The filters were built by applying the Kirkeby reverse filter method [Farina, Martignon, Azzali & Capra, 2004] to the recorded impulse responses obtained by generating sweeps at both HATS channels through the headphones. An Adobe Audition (version 3.0) software package was used with an Aurora (version 4.4) Alfa plug-in;

- In C3, the previous chair-flooring interaction was simplified by the new method [Dal Palù, De Giorgi, Astolfi, Lerma & Buiatti, 2014]. In this case, the rolling sound source evaluated in C1 was simplified by the material-configuration form-exciting mode interaction as follows: the material of the paved platform in C1 was the same of each flooring tile adopted in C3; the configuration form was the paving scheme in C1 and it was reproduced with three pieces of floorings disposed with the same paving scheme in C3; the exciting mode interaction of moving with a chair across a paved platform in C1 corresponded to “sliding an object (i.e. one wheel) on one surface (i.e. the tiles)” in C3. Three pieces of flooring (e.g. PVC, ceramic and wood tiles) were then arranged on the floor of the dead room on SounBe tool, according to the same paving scheme as that of the paved platforms (checkered and deck schemes); two sets of office chair wheels, which had previously been removed from the chairs, were equipped with a handle to facilitate hand grip. Each type of flooring was repeatably stressed by rolling each wheel over it, following SounBe protocol (Figure 2) [Dal Palù, De Giorgi, Astolfi, Lerma & Buiatti, 2014]. From the original toolandmethod, the supporting frame, the simplified procedure method and the data collection and analysis protocol were adopted. The sounds produced by this action were recorded binaurally with HATS and measured by the phonometer. Again in this case, the procedure was performed with the same reciprocal position scheme between the sound source, the receiver, and the phonometer (Figure 2). The sounds recorded with HATS were convolved with the previously described Kirkeby inverse filter.



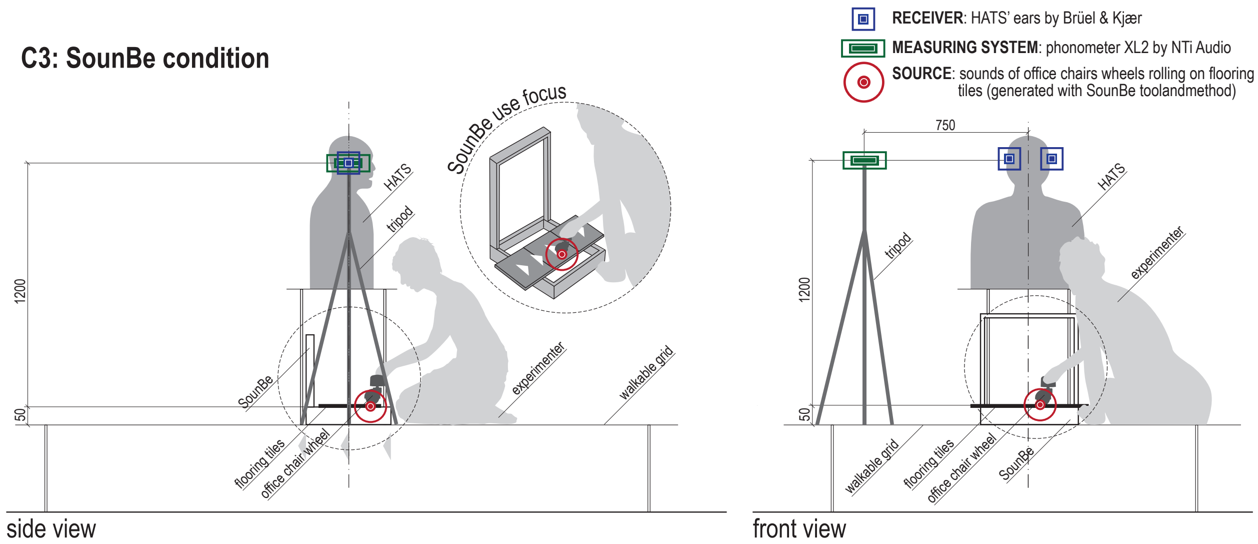



**Figure 2**. A scheme describing the method adopted to collect the stimuli and all the reciprocal positions of receiver, measuring system and source in all the three conditions (C1, C2 and C3). A specific focus on SounBe toolandmethod use clarifies how the SounBe device was used to produce the sounds in C3.

### 2.2.3 Participants

A group of 90 participants (46 women and 44 men, $\bar{x}$ = 27.1 years, $\sigma$ = 6.5 years) was recruited to take part in the experiment. All the participants reported normal hearing and no motor impairments. The group included both "experts", e.g. people who declared they had received physical acoustics and/or a formal musical education (No. = 36), and "laymen" on these topics (No. = 54), in order to verify whether background knowledge could affect the subjective data. In order to avoid fatigue, boredom, or to make participants become accomplished through practice and experience to the test, a between-subject design was set up [Charness, Gneezy, Kuhn, 2012]. Three sub-groups of 30 randomly assigned subjects were created, for the C1 test (16 women and 14 men, $\bar{x}$ = 28.5 years, $\sigma$ = 8.3 years, 12 experts and 18 laymen), the C2 test (16 women and 14 men, $\bar{x}$ = 28.2 years, $\sigma$ = 6.2 years, 11 experts and 19 laymen) and the C3 test (14 women and 16 men, $\bar{x}$ = 24.7 years, $\sigma$ = 3.5 years, 13 experts and 17 laymen), respectively. In C1, participants were asked their weight (women: $\bar{x}$ = 59 kg, $\sigma$ = 6.3 kg; men: $\bar{x}$ = 70.4 kg, $\sigma$ = 10.5 kg) in order to verify whether the weight factor could affect the descriptive process.

### 2.2.4 Subjective data: semantic differentials

The task was the same for each condition. The participant was asked to rate the sounds heard during the listening test on the three previously selected semantic differentials scales (section 2.1). In this questionnaire, the source of the sound was declared [Susini, Houix, Misdariis, Smith, Langlois, 2009]. A further semantic differential pair was used to assess the sound annoyance (*not annoying at all-very annoying*, reported as SD4) to obtain an overall rating. The rationale behind the addition of this further question was to be able to investigate the correlation with other subjective and objective data. In fact, it has been proved [Park, Jeon, Choi & Park, 2015] and discussed [Lyon, 2003] that annoyance perception is not always related to the psychoacoustic index of Loudness, especially as far as unsteady and temporary sounds are concerned.
The questionnaire was based on a 1 to 7 point Likert scale for each semantic differential pair. The statistic data (mean values, variance and median) obtained in the qualitative questionnaire are shown in Appendix A. These data were compared with the objective data obtained with the phonometer in each test condition, as well as for the different conditions (C1, C2 and C3) (section 3.2.2).

### 2.2.5 Objective data: acoustic and psychoacoustic measurements

The submitted stimuli were measured using a previously calibrated phonometer (XL2 handheld audio and acoustic analyser made by NTi Audio). Each sound was analysed using the PULSE Reflex software package (version 17.1.0) by Brüel & Kjær. Acoustic and psychoacoustic measurements of the A-weighted equivalent sound pressure level ($L_{Aeq}$), the maximum level with A-weighted frequency response and Fast time constant ($L_{AFmax}$), Loudness (L), Sharpness (S), Roughness (R), Fluctuation Strength (FS) and Tonality (Ton) were calculated [Fastl & Zwicker, 2007]. The objective data were compared with the subjective data, and the statistic data (mean values and variance) resulting from the tests are shown in Appendix A.

## 3. Results

The results obtained in the preliminary phase and in the main phase of the study are presented separately in the following sections.

### 3.1 Semantic differential pairs selected in the preliminary phase

Figure 3 shows the results obtained with the descriptive statistics on the semantic differential pairs selection conducted during the preliminary phase of the study (section 2.1). Among the 13 Von Bismarck semantic descriptor pairs, the participants spontaneously selected a mean value of 3 pairs. The 3 most frequently rated pairs were *calm-strident* (SD1), *pleasant-unpleasant* (SD2) and *smooth-rough* (SD3). A correlation was found between the judgements given by the experts and laymen (Pearson Chi-Square test $\chi^2$ = 0.30, significance level *p*-value = 0.05) for these pairs. The three most frequently rated descriptor pairs were used in the main phase of the experiment, in order to assess the stimuli submitted during the test.

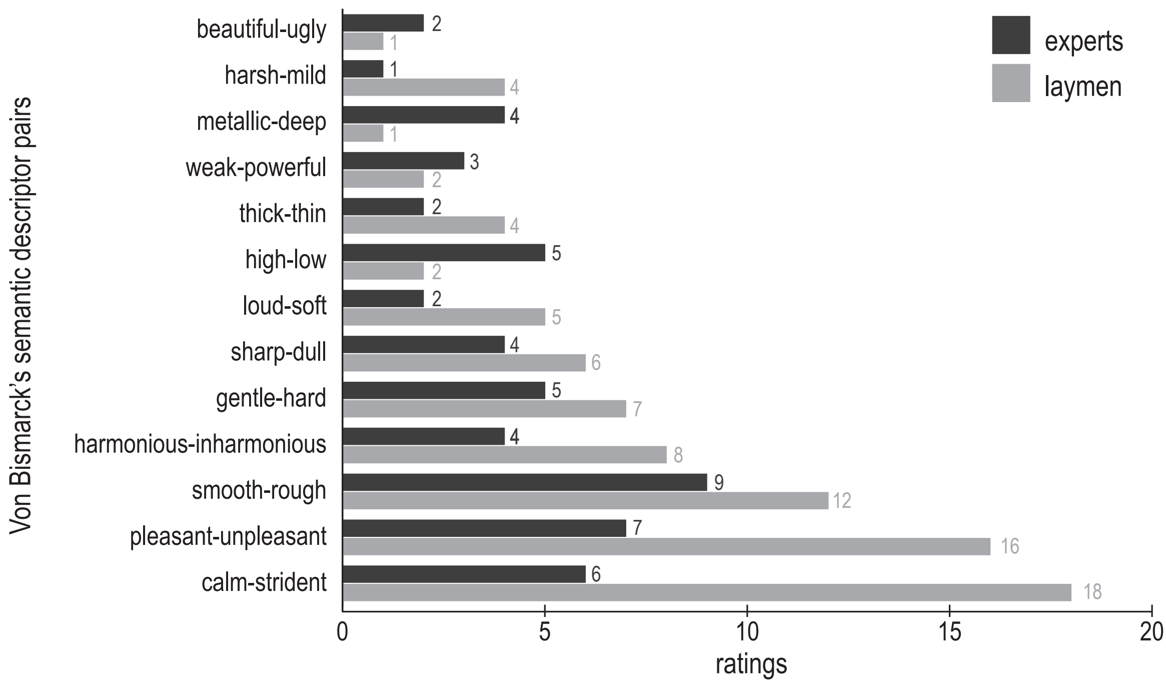



**Figure 3**: Distribution of the ratings for each descriptor pair, considering the responses of both the experts and laymen.

### 3.2 Comparison of the subjective and objective data collected in the main phase

A summary of the subjective and objective data collected in the C1, C2 and C3 listening tests in the main phase is presented in Appendix A. Statistical analyses were carried out with the IBM SPSS Statistics software package (version 22, Armonk, NY).

#### 3.2.1 Kolmogrov-Smirnov normality distribution test

A first test on the normality distribution of the data was necessary to define the subsequent type of tests (parametric or non-parametric tests) that had to be used. The Kolmogrov-Smirnov (K-S) normality distribution test was used to verify the normality of the distribution of the subjective (SD1; SD2; SD3; SD4) and objective (LAeq; LAFmax; L; S; R; FS; Ton) data [Siegel & Castellan, 1988]. The Lilliefors significance correction was then applied [Dallal & Wilkinson, 1986]. A significance level of 0.05 was assumed for this calculation. The two-tailed *p*-value obtained from the K-S normality distribution test was lower than the significance level of 0.05 (for all the variables *p*-value = 0.00) for each variable. The K-S normality distribution test demonstrated that the data were not normally distributed. As a result, the subsequent analyses were calculated by means of non-parametric statistics.

### 3.2.2 Mann-Whitney U and Kruskal-Wallis H tests

In the subsequent statistical analyses, non-parametric tests, such as the Mann-Whitney U (M-W U) and the Kruskal-Wallis H (K-W H) tests, were applied in order to find any statistical differences between the different groups of data. The choice of adopting the M-W U or the K-W H tests was determined by the number of independent groups that had to be compared. These tests were used for comparisons between two or more than two groups, respectively [Siegel & Castellan, 1988]. A significance level of 0.05 was assumed for the subsequent data analyses.

*General analysis on the effect of the participant-related variables*

A general analysis of the effect due to the participant-related variables on the data collected in all the experimental conditions was carried out. The main goal of this analysis was to establish any possible influence of the participants on the subjective and objective data. The effect of variables such as gender (woman; man) and expertise (layman; expert) on subjective data was computed. The effect of weight (lighter than 62.5 kg; heavier than or equal to 62.5 kg) on the subjective and objective data was also computed. The reference weight of 62.5 kg was chosen, because it represented the median value of the weight variable.

Table 1 shows the two-tailed *p*-values obtained from the M-W U test for each variable. No significant difference was found for the gender or weight variables. Nevertheless, a significant difference was found for the expertise variable on the pleasant-unpleasant and not annoying at all-very annoying semantic differential scales. Considering the median values and the interquartile ranges, Figure 4 shows that the experts overall judged rolling sounds as a little more pleasant and less annoying than laymen. However, since a correlation between the judgements made by the experts and laymen had emerged in the preliminary experimental phase (see paragraph 3.1), the data of these two subgroups were grouped again in the subsequent analyses.

**TABLE 1.** The Mann-Whitney U (M-W U) test was calculated on the subjective and objective data collected in all the experimental conditions (C1; C2; C3). Two-tailed *p*-values of the significance of the difference in the distributions of data in relation to the participant-related variables are shown. The *p*-values lower than or equal to the significance level of 0.05 are in bold.

| | | subjective data | | | | objective data | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Variable** | **compared groups** | **SD1** | **SD2** | **SD3** | **SD4** | **$L_{Aeq}$ [dB]** | **$L_{AF max}$ [dB]** | **L [sone]** | **S [acum]** | **R [asper]** | **FS [vacil]** | **Ton** |
| Gender | (woman; man) | 0.92 | 0.11 | 0.46 | 0.16 | | | | | | | |
| Expertise | (layman; expert) | 0.21 | **0.01** | 0.11 | **0.03** | | | | | | | |
| Weight | (< 62.5 kg; ≥ 62.5 kg) | 0.91 | 0.70 | 0.89 | 0.60 | 0.44 | 0.32 | 0.23 | 0.94 | 0.67 | 0.13 | 0.84 |

Note: SD1 = calm-strident; SD2 = pleasant-unpleasant; SD3 = smooth-rough; SD4 = not annoying at all-very annoying; $L_{Aeq}$ = A-weighted equivalent sound pressure level; $L_{AFmax}$ = maximum level with A-weighted frequency response and Fast time constant; L = Loudness; S = Sharpness; R = Roughness; FS = Fluctuation Strength; Ton = Tonality.


**Figure 4**: Boxplots of the pleasantness and annoyance semantic differential scale ratings for data grouped according to the expertise variable. The grey boxes represent the interquartile (IQ) range, which contains the middle 50% of the records. The whisker lines that extend from the upper and lower edges of the box refer to the highest and lowest values, which are no greater than 1.5 times the IQ range. The thick line across the box indicates the median.

*"Intra-condition" statistical analysis*

A deeper analysis of each test condition was carried out, through a general-to-specific approach. The aim of this analysis was to find any differences and similarities between chair-flooring matching in each test condition. The effects of the chair (HLC; LLC) and flooring (PVC; ceramics; wood) variables on the subjective and objective data were calculated for each condition.

Table 2 shows the two-tailed *p*-values obtained from the M-W U and K-W H tests on the data collected in C1, the data set having been split into smaller groups. The following results can be pointed out:

- When the data are grouped according to the chair variable (i.e. taking into account both chairs together), a significant difference in the objective data can be found for $L_{AFmax}$, L and R. Nevertheless, this difference is not outlined by the subjective data for this grouping;
- When the data are grouped considering each chair separately (i.e. taking into account just HLC first, and just LLC later), in a more detailed analysis of the previous significant objective data, a similar trend can be found when ceramic flooring is included in the comparison. On the contrary, only a few significant differences can be found for $L_{AFmax}$ and L when PVC and wood are compared;
- When the data are grouped according to the flooring variable, a significant difference in the subjective data can generally be found. The difference is perceived much more when the ceramic flooring is included in the comparison. On the contrary, this difference is barely perceived when PVC and wood are compared. This trend is confirmed for the objective data. In fact, in this case, no significant difference can be found between $L_{AFmax}$ and L when PVC and wood are compared.

**TABLE 2.** The Mann-Whitney U (M-W U) and the Kruskal-Wallis H (K-W H) tests were calculated on the subjective and objective data collected in C1. Two-tailed *p*-values of the significance of the difference in the distributions of data, in relation to the chair and flooring variables, are shown. The *p*-values lower than or equal to the significance level of 0.05 are in bold.

| | | subjective data | | | | objective data | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **variable** | **compared groups** | **SD1** | **SD2** | **SD3** | **SD4** | **$L_{Aeq}$ [dB]** | **$L_{AFmax}$ [dB]** | **L [sone]** | **S [acum]** | **R [asper]** | **FS [vacil]** | **Ton** |
| chair | (HLC; LLC) | 0.31 | 0.15 | 0.90 | 0.10 | 0.07 | **0.01** | **0.00** | 0.57 | **0.00** | 0.95 | 0.06 |
| HLC | (PVC; ceramics; wood) | | | | | | **0.00** | **0.00** | | **0.00** | | |
| | (PVC; ceramics) | | | | | | **0.00** | **0.00** | | **0.00** | | |
| | (PVC; wood) | | | | | | 0.71 | 0.95 | | **0.03** | | |
| | (ceramics; wood) | | | | | | **0.00** | **0.00** | | **0.00** | | |
| LLC | (PVC; ceramics; wood) | | | | | | **0.00** | **0.00** | | **0.00** | | |
| | (PVC; ceramics) | | | | | | **0.00** | **0.00** | | **0.00** | | |
| | (PVC; wood) | | | | | | 0.92 | **0.01** | | **0.02** | | |
| | (ceramics; wood) | | | | | | **0.00** | **0.00** | | **0.00** | | |
| flooring | (PVC; ceramics; wood) | **0.00** | **0.00** | **0.00** | **0.00** | **0.00** | **0.00** | **0.00** | **0.00** | **0.00** | **0.00** | **0.00** |
| | (PVC; ceramics) | **0.00** | **0.00** | **0.00** | **0.00** | **0.02** | **0.00** | **0.00** | **0.00** | **0.00** | **0.00** | **0.01** |
| | (PVC; wood) | 0.39 | **0.04** | 0.12 | 0.25 | **0.01** | 0.74 | 0.10 | **0.00** | **0.01** | **0.02** | **0.02** |
| | (ceramics; wood) | **0.00** | **0.00** | **0.00** | **0.00** | **0.00** | **0.00** | **0.00** | 0.65 | **0.00** | **0.00** | **0.00** |

Note: HLC = high level chair; LLC = low level chair; SD1 = calm-strident; SD2 = pleasant-unpleasant; SD3 = smooth-rough; SD4 = not annoying at all-very annoying; $L_{Aeq}$ = A-weighted equivalent sound pressure level; $L_{AFmax}$ = maximum level with A-weighted frequency response and Fast time constant; L = Loudness; S = Sharpness; R = Roughness; FS = Fluctuation Strength; Ton = Tonality.

Figure 5 shows the distribution of the subjective data on the submitted stimuli, grouped according to the flooring variable. The difference in perception of the ceramic flooring compared with the PVC and wood flooring is confirmed graphically.

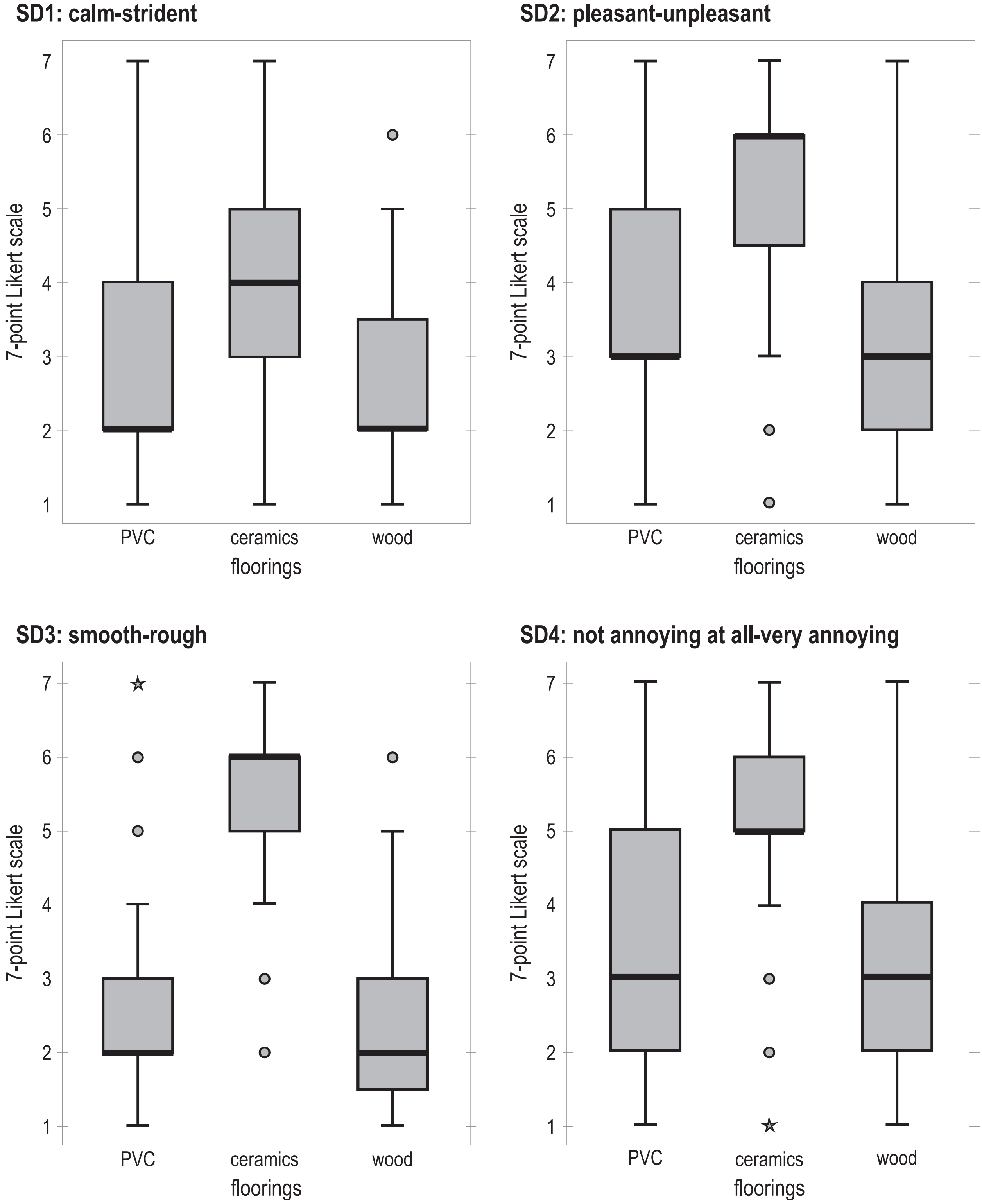



**Figure 5**: Boxplots of the semantic differential scale ratings for data grouped according to the flooring variable. The grey boxes represent the interquartile (IQ) range, which contains the middle 50% of the records. The whisker lines that extend from the upper and lower edges of the box refer to the highest and lowest values, which are no greater than 1.5 times the IQ range. The thick line across the box indicates the median. The grey dots represent the outliers, i.e. the cases with values between 1.5 and 3 times the IQ range. The asterisks are the extremes, i.e. cases with values more than 3 times the IQ range.

Table 3 and Table 4 show the two-tailed *p*-values of the same tests (M-W U and K-W H tests) obtained from the subjective data collected in C2 and C3, respectively. The objective data were not taken into consideration here since, in these conditions, only one reference stimulus was submitted for each chair-flooring matching during the listening test. As a result, the objective data did not vary according to the test participants in these conditions.

The following results can be pointed out:

- In both conditions, when the data are grouped according to the chair variable, no significant difference can be found, thus confirming the trend outlined in C1.
- When the data are grouped according to the flooring variable, almost all the semantic differential scales in each flooring comparison in C2 outline a significant difference in perception (Table 3), even for the PVC-wood comparison. On the contrary, the trend in C3 is more similar to that in C1: the difference is perceived much more when the ceramic flooring is included in the comparison, and it is barely perceived when PVC and wood are compared (Table 4).

**TABLE 3.** The Mann-Whitney U (M-W U) and the Kruskal-Wallis H (K-W H) tests were calculated on the subjective data collected in C2. Two-tailed *p*-values of the significance of the difference in the distributions of data, in relation to the chair and flooring variables, are shown. The *p*-values lower than or equal to the significance level of 0.05 are in bold.

| | | subjective data | | | |
|---|---|---|---|---|---|
| **Variable** | **compared groups** | **SD1** | **SD2** | **SD3** | **SD4** |
| Chair | (HLC; LLC) | 0.43 | 0.78 | 0.72 | 0.90 |
| Flooring | (PVC; ceramics; wood) | **0.00** | **0.00** | **0.00** | **0.00** |
| | (PVC; ceramics) | **0.00** | **0.00** | 0.13 | **0.00** |
| | (PVC; wood) | **0.03** | **0.00** | **0.00** | **0.00** |
| | (ceramics; wood) | **0.00** | **0.00** | **0.00** | **0.00** |

Note: HLC = high level chair; LLC = low level chair; SD1 = calm-strident; SD2 = pleasant-unpleasant; SD3 = smooth-rough; SD4 = not annoying at all-very annoying.

**TABLE 4.** The Mann-Whitney U (M-W U) and the Kruskal-Wallis H (K-W H) tests were calculated on the subjective data collected in C3. Two-tailed *p*-values of the significance of the difference in the distributions of data, in relation to the chair and flooring variables, are shown. The *p*-values lower than or equal to the significance level of 0.05 are in bold.

| | | subjective data | | | |
|---|---|---|---|---|---|
| **variable** | **compared groups** | **SD1** | **SD2** | **SD3** | **SD4** |
| chair | (HLC; LLC) | 0.77 | 0.94 | 0.20 | 0.60 |
| flooring | (PVC; ceramics; wood) | **0.00** | **0.00** | **0.00** | **0.00** |
| | (PVC; ceramics) | **0.00** | **0.00** | **0.00** | **0.00** |
| | (PVC; wood) | 0.15 | 0.18 | 0.69 | 0.28 |
| | (ceramics; wood) | **0.00** | **0.00** | **0.00** | **0.00** |

Note: HLC = high level chair; LLC = low level chair; SD1 = calm-strident; SD2 = pleasant-unpleasant; SD3 = smooth-rough; SD4 = not annoying at all-very annoying.

### *"Inter-condition" statistical analysis*

A further analysis of the test conditions was carried out. The aim of this analysis was to find any differences and/or similarities of the test conditions. Table 5 shows the two-tailed *p*-values obtained from the K-W H test on the total amount of subjective data collected during the experiment. No significant difference arises from the comparison between C1, C2 and C3 for any of the semantic differential scales.

**TABLE 5.** The Kruskal-Wallis H (K-W H) test was calculated on the subjective data collected in C1, C2 and C3. Two-tailed *p*-values of the significance of the difference in the distributions of data, in relation to the different conditions, are shown. The *p*-values lower than or equal to the significance level of 0.05 are in bold.

| | | subjective data | | | |
|---|---|---|---|---|---|
| **Variable** | **compared groups** | **SD1** | **SD2** | **SD3** | **SD4** |
| Condition | (C1; C2; C3) | 0.09 | 0.15 | 0.05 | 0.09 |

Note: C1 = context condition; C2 = laboratory condition; C3 = SounBe condition; SD1 = calm-strident; SD2 = pleasant-unpleasant; SD3 = smooth-rough; SD4 = not annoying at all-very annoying.

## 4. Discussion

The results of the previous section offer some relevant considerations for industries, researchers and designers who plan to investigate the perception of a product sound by means of listening tests. Some considerations about the effect of participant-related variables are pointed out. Furthermore, the comparability of the collected data, in different test conditions, is discussed. Finally, the industrial implications of these findings are presented.

### 4.1 The effect of participant-related variables on the semantic differential listening tests

This experiment, in the preliminary and main phases, involved participants with various personal features, who differed according to gender, expertise level and weight.

Unlike the Von Bismarck approach [Von Bismarck, 1974], a mixed group of experts and laymen took part in the preliminary semantic differential pair selection. This preliminary study outlined a difference in the selection strategy, which could probably be attributed to the well-known difference between *musical* and *everyday* listening [Gaver, 1993]. In fact, laymen showed a more focused judgement strategy than the experts. It is possible to hypothesize that the experts judged the sounds mostly on the basis of acoustic measures [Schinkel-Bielefeld, Lotze, & Nagel, 2013], thus their judgement strategy was more diffused than that of the laymen. Nevertheless, their background knowledge did not affect the pair selection; in fact, both groups selected the same three pairs.

In the main phase of the experiment, gender, expertise and weight participant-related variables were considered. The results show that these variables had no effect on the subjective and objective data. This could mean for example that, since tests on the acceptance of a product need to be predictive [Samli, 1996], a subject's weight factor is irrelevant in the case of an office chair assessment. Nevertheless, gender, expertise and weight are just some of the obvious factors to be verified, but it cannot be excluded that other participant-related variables or personal factors, e.g. intelligence, emotive quotient, culture and every other personality constructs, could be significant or could influence the results [Charness, Gneezy, Kuhn, 2012]. This additional uncertainty intrinsic of a between-subject design should always be taken into account.

### 4.2 Correspondence between the ratings given to semantic differentials in the context and laboratory test conditions

According to the categorization of Özcan & Van Egmond [Özcan & Van Egmond, 2012], the semantic differential pairs selected in the preliminary phase of the experiment pertained to perceptual (*smooth-rough)*, cognitive (*calm-strident),* and emotional (*pleasant-unpleasant* and *not annoying at all-very annoying)* factors.

In all the considered conditions, the flooring factor was found to have more influence on the subjective data than the chair factor, a result that is consistent with recent findings on in-vehicle rolling sound perception [Li, Qiao & Yu, 2016]. Specifically, the ceramic flooring made the difference with other floorings being more clearly perceivable, probably because of the greater coarseness of its surface and the presence of deeper joints.

There seem to be an overall agreement in the scientific community that simulated sound environments (i.e. laboratory conditions) allow a better control of the sounds presented to individuals. This means researchers can investigate the relationship between cause and effect, which is useful for theory advancement, but results obtained in a laboratory should always be validated in context [Aletta, Kang, & Axelsson, 2016]. As presented in Table 5, a comparison of all the conditions (i.e. the context, C1, the laboratory, C2, and the SounBe condition, C3) outlined that no significant differences among medians were found overall for the ratings given to these semantic differential scales on these stimuli. However, differences in variability of the ratings can be observed (see the variances appearing in Appendix A). This finding is consistent with evidence that has shown that the linguistic analysis of verbal data is a reliable measure of the ecological validity of reproduction systems in experimental settings [Guastavino, Katz, Polack, Levitin, & Dubois, 2005]. Furthermore, recent studies suggest that even though it is not expected that a sample of participants would assess sounds in laboratory conditions exactly as they would in context (i.e. the same scores), it is reasonable to assume that the relative preference (i.e. the 'ranking') would be consistent [Aletta, Kang, Fuda, & Astolfi, 2016]. Similarly, the finding of the present study suggest that even passive and abstract listening tests could be performed in order to reliably assess the qualities of an office chair sound. Nevertheless, the subjective data, i.e. the given ratings, collected in C3 and presented in Table 4 proved that these stimuli were judged to be more comparable with the real office chair sounds, i.e. those assessed in C1 and presented in Table 2, than those judged in C2 and presented in Table 3, especially for the PVC-wood comparison. This means that the simplified procedure, based on the new tool, the material-configuration and the exciting mode interaction, did not entail any loss of information in this case. Therefore, this experiment highlighted the convenience of the use of the SounBe tool to generate rolling chair office sounds for listening tests using headphones.

### 4.3 Comparison between the subjective and objective data

Several approaches that prevalently involve objective data have been adopted to deal with product sound acceptance, but subjective data have rarely been considered. A previous study on office chair sounds [Alves, Filho, Silva & Câmara, 2012] assumed the Sound Pressure Level (SPL) as the meter to define the noise acceptability for workers. The present study has shown that parameters such as $L_{AFmax}$, L and R are consistent in outlining the objective differences between chairs and floorings in office chair rolling sounds, especially when ceramic flooring is included in the comparison. Nevertheless, these parameters cannot be considered alone: in fact, the study has proved that the participants in C1 were able to perceive a difference in pleasantness between chairs rolling on PVC and on wood, even though they found no differences between $L_{AFmax}$ and L. This finding can be explained by considering the well-known involvement of multisensory perception in everyday life [Stein & Meredith, 1993]. However, this fact points out that human perception cannot be predicted just by means of acoustic measures, and that a comparison between subjective and objective data provides useful extra information for industries and researchers.

### 4.4 Industrial implications

The semantic differentials adopted in listening tests on industrial products have seldom shown any consistency between laboratory and context testing conditions [Steffens, Schulte-Fortkamp & Becker-Schweitzer, 2011]. It is well known that when listening, ones attention is first focused on identifying the sound-producing event, i.e. the source of the sound [Gaver, 1993]. Recent studies have proved that subjects who are well aware of what to judge generally yield more consistent results in laboratory listening tests [Nykänen, Lennström, & Johnsson, 2015]. In other words, experiencing a representative product (e.g. driving a car) before being submitted to the listening test related to the characteristics of the product itself (e.g. the interior car sounds) positively increases the reliability of the answers on semantic differential scales. Other studies have proved that even just providing the information related to the sound source of a product increases the validity of laboratory listening tests [Susini, Houix, Misdariis, Smith, Langlois, 2009]. Therefore, it is possible to hypothesise performing listening tests on sound acceptance before having the final product, or an advanced prototype. In this context, the new tool with semantic differential scales, related to pleasantness and annoyance, can be considered a useful support for industries and sound design stakeholders.

## 5. Conclusion

The evidence obtained from this experiment highlights the efficacy of the considered laboratory conditions in performing listening tests with semantic differential scales. In other words, it is possible to conclude that:

- The subject-related gender and weight variables have been shown to have no effect on the assessment of the rolling sounds of office chairs with semantic differential scales in this study;
- From a perceptive point of view, no difference has emerged between the context test condition (C1), the laboratory test conditions (C2) and the SounBe test condition (C3) on the semantic differentials scales ratings related to calmness, roughness, pleasantness and annoyance for office chair rolling sounds;
- The efficacy of the new tool, in consideration of its forecasting attitude, offers the possibility of generating a product sound, without the necessity of the product prototyping phase. Furthermore, the

use of this tool to create the sound stimulus can be coupled with a semantic differential technique to investigate the future product sound perception.

In conclusion, the adaptability and cost-effectiveness of the laboratory test conditions, as well as the forecasting approach of the SounBe tool to generate the stimuli suggest the possible interest of industries, researchers and designers in using these techniques in order to perform reliable listening tests on product sounds.

## Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors would like to thank the Ares Line and Scamet Italia companies for having provided the office chairs used in this study. Special thanks are due to Louena Shtrepi and the technical team of the Department of Energy at the Politecnico di Torino for their collaboration during the experimental phase, as well as to the Perception and Sound Design team at IRCAM for the suggestions and the support during the development of the main part of the experiment. Finally, many thanks are due to all the students and participants who took part in the test, as well as to Francesca Arato, co-inventor of the SounBe toolandmethod in the preliminary stage of this research.

## References


Aletta F, Kang J, Axelsson Ö. Soundscape descriptors and a conceptual framework for developing predictive soundscape models. Landscape and Urban Planning 2016;149:65-74.

Aletta F, Kang J, Fuda S, Astolfi A. The effect of walking sounds from different walked-on materials on the soundscape of urban parks. Journal of Environmental Engineering and Landscape Management 2016;24(3):165-75.

Alves EJW, Filho JN, Silva SJ, Câmara JJD. Noise analysis in professional office chairs. Work 2012;41:1664-9.

Astolfi A, Pellerey F. Subjective and objective assessment of acoustical and overall environmental quality in secondary school classrooms. The Journal of the Acoustical Society of America 2008;123:163-73.

Bodin Danielsson C, Bodin L. Office type in relation to health, well-being, and job satisfaction among employees. Environment and Behavior 2008;40(5):636-68.

Brinkman WP. Design of a questionnaire instrument. In: Love S, editor. Handbook of mobile technology research methods. Hauppauge, NY: Nova Publisher; 2009, p. 31-57.

Cahen R. In the circle of Education and Innovation. Teaching Sound-Design @ENSCI Les Ateliers. In: Proceedings of Cumulus Conference 2015, Milano, Italy; 2015.

Carron M, Dubois F, Misdariis N, Talotte C, Susini P. Designing Sound Identity: Providing new communication tools for building brands "corporate sounds". In: Audio Mostly 2014, Aalborg, Denmark; 2014.

Centro Studi Cosmit/FederlegnoArredo. Rapporto di settore 2015 – sistema ufficio con consuntivi annuali 2014. Document updated on the 28th March 2015. Milano, Italy, 2015.

Charness G, Gneezy U, Kuhn MA. Experimental methods: Between-subject and within-subject design. Journal of Economic Behavior & Organization 2012;81:1-8.

Dallal GE, Wilkinson L. An analytic approximation to the distribution of Lilliefors' test for normality. The American Statistician 1986;40:294-6.

Dal Palù D, De Giorgi C, Astolfi A, Lerma B, Buiatti E. SounBe, a toolkit for designers dealing with sound projects. In: Proceedings of the International Design Conference 2014, Dubrovnik, Croatia; 2014.

Davies WJ, Bruce NS, Murphy JE. Soundscape reproduction and synthesis. Acta Acustica United with Acustica 2014;100(2):285-92.

De Giorgi C, Astolfi A, Buiatti E, Lerma B, Arato F, Dal Palù D. SounBe: Method and device for acoustic sensorial analysis of materials. Patent number: IT/TO20110089, Politecnico di Torino, Torino; 2011.

De Rouvray A, Bassereau JF, Duchamp R, Schneider JS & Charbonneau S. Perception and deception: How quantity and quality of sensory information affect users' perception of office chairs. The Design Journal 2008;11(1):29-59.

Farina A, Martignon P, Azzali A, Capra A. Listening Tests Performed Inside a Virtual Room Acoustic Simulator. In: Proceedings of the Seminar Música Ciênca e Tecnologia "Acústica Musical" 2004, São Paolo do Brasil, Brasil; 2004.

Fastl H, Zwicker E. Psychoacoustics. Facts and Models. Berlin: Springer-Verlag; 2007.

Gaver WW. What in the world do we hear? An ecological approach to auditory source perception. Ecological Psychology 1993;5(1):1-29.

Guastavino C, Katz BF, Polack J, Levitin DJ, Dubois D. Ecological validity of soundscape reproduction. Acta Acustica United with Acustica 2005;91:333-41.

Houix O, Lemaitre G, Misdariis N, Susini P, Urdapilleta I. A lexical analysis of environmental sound categories. Journal of Experimental Psychology 2012;18(1):52-80.

Kasess CH, Noll A, Majdak P, Waubke H. Effect of train type on annoyance and acoustic features of the rolling noise. Journal of the Acoustical Society of America 2013;134(2):1071-81.

Kjellberg A, Landström U. Noise in the office: Part II. - The scientific basis (knowledge base) for the guide. International Journal of Industrial Ergonomics 1994;14:93-118.

[illegible] T, Han J, Uchida H, Approach for combining physical properties and sensibility for pleasant beverage can-opening sound. Applied Acoustics 2016;103(A):64-70.

Jeon JY, You J, Chang HY. Sound radiation and sound quality characteristics of refrigerator noise in real living environment. Applied Acoustics 2007;68:1118-34.

Lemaitre G, Houix O, Visell Y, Franinović K, Misdariis N, Susini P. Toward the design and evaluation of continuous sound in tangible interfaces: The spinotron. International Journal of Human Computer Studies 2009;67(11):976-93.

Lemaitre G, Rocchesso D. On the effectiveness of vocal imitations and verbal descriptions of sounds. Journal of the Acoustical Society of America 2014;135(2):862-73.

Lindau A, Weinzierl S. Assessing the plausibility of virtual acoustic environments. In: Proceedings of Forum Acusticum 2011, Aalborg, Denmark; 2011.

Li Q, Qiao F, Yu L. Impacts of pavement types on in-vehicle noise and human health. Journal of the air & waste management association 2016;66(1):87-96.

Lyon RH. Product Sound Quality – from Perception to Design. Sound and Vibration 2003;18-22.

Martignon-Burgholte R, Cyffka A. Barron’s Italian-English dictionary. Hauppauge: Barron's Educational Series; 2007.

Namba S, Kuwano S, Hashimoto T, Berglund B, Zheng RD, Schick A, Hoege H, Florentine M. Verbal expression of emotional impression of sound: A cross-cultural study. Journal of the Acoustical Society of Japan 1991;12(1):19-29.

Nicolao U, Noselli G. Audio Dizionario. Milano: Il Rostro; 2007.

Nykänen A, Lennström D, Johnsson R. Car ride before entering the lab increases precision in listening tests. SAE International Journal of Passenger Cars - Mechanical Systems 2015;8(3):982-8.

Ohiduzzaman MD, Sirin O, Kassem E, Rochat JL. State-of-the-art review on sustainable design and construction of quieter pavements-Part 1: Traffic noise measurement and abatement techniques. Sustainability 2016;8(8):1-28.

Osgood CE. The nature and measurement of meaning. Psychological Bulletin 1952;49:197-237.

Özcan E, Van Egmond R. Basic semantics of product sounds. International Journal of Design 2012;6(2):41-54.

Özcan E, Van Egmond R. Pictograms for sound design: A language for the communication of product sounds. In: Proceedings of the International Conference on Design and Emotion 2004, Ankara, Turkey; 2004.

Parizet E, Nosulenko V. Multi-dimensional listening test: selection of sound descriptors and design of the experiment. Noise Control Engineering Journal 1999;47:1-6.

Park B, Jeon JY, Choi S, Park J. Short-term noise annoyance assessment in passenger compartments of high-speed trains under sudden variation. Applied Acoustics 2015;97:46-53.

Pheasant S, Haslegrave CM. Bodyspace: Anthropometry, Ergonomics and the Design of Work, third edition. Boca Raton: Taylor & Francis Group; 2005.

Raimbault M. Qualitative judgements of urban soundscapes: Questionning questionnaires and semantic scales. Acta Acustica United with Acustica 2006;92(6):929-37.

Ragazzini G. Dizionario Inglese-Italiano. Italian-English Dictionary. Bologna: Zanichelli; 2015.

Saeki T, Fujii T, Yamaguchi S, Harima S. Effects of acoustical noise on annoyance, performance and fatigue during mental memory task. Applied Acoustics 2004;65(9):913-21.

Samli CA. Information-Driven Marketing Decisions. Development of Strategic Information Systems. Westport: Praeger; 1996.

Schifferstein HNJ, Hekkert P, editors. Product experience. Amsterdam: Elsevier; 2008.

Schinkel-Bielefeld N, Lotze N, Nagel F. Audio quality evaluation by experienced and inexperienced listeners. In: Proceedings of Meetings on Acoustics 2013, Montreal, Canada; 2013.

Schütte M., Müller U., Sandrock S., Griefahn B., Lavandier C., Barbot B. Perceived quality features of aircraft sounds: An analysis of the measurement characteristics of a newly created semantic differential. Applied Acoustics 2009;70(7):903-14.

Siegel S, Castellan Jr NJ. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill; 1988.

Sirin O. State-of-the-art review on sustainable design and construction of quieter pavements-Part 2: Factors affecting tire-pavement noise and prediction models. Sustainability 2016;8(7):1-21.

Soeta Y, Shimokura R. Survey of interior noise characteristics in various types of trains. Applied Acoustics 2013;74(10):1160-66.

Solomon LN. Semantic approach to the perception of complex sounds. Journal of the Acoustical Society of America 1958;30(5):421-5.

Steffens J. When do we judge sounds? Relevant everyday situations for the estimation of ecological validity of indoor soundscape experiments. In: Proceedings of Meetings on Acoustics 2013, Montreal, Canada; 2013.

Steffens J, Schulte-Fortkamp B, Becker-Schweitzer J. Design of domestic Soundscapes – Evaluation of household appliances in laboratory and reality. In: Proceedings of Forum Acusticum 2011, Aalborg, Denmark; 2011.

Susini P, Houix O, Misdariis N, Smith B, Langlois S. Instruction's effect on semantic scale ratings of interior car sounds. Applied Acoustics 2009;70:389-403.

Susini P, Lemaitre G, McAdams S. Psychological measurement for sound description and evaluation. In: Berglund B, Rossi GB, Townsend JT, Pendrill LR, editors. Measurement with Persons: Theory, methods and implementation areas. New York: Psychology Press; 2011, p. 227-253.

Turchet L, Serafin S. Semantic congruence in audio–haptic simulation of footsteps. Applied Acoustics 2014;75:59-66.

Von Bismarck G. Timbre of steady sounds: factorial investigation of its verbal attributes. Acustica 1974;30:146-59.

Xie H, Kang J. Sound field of typical single-bed hospital wards. Applied Acoustics 2012;73(9):884-92.

Zingarelli N. Lo Zingarelli 2016. Bologna: Zanichelli; 2015.

## Appendix A

A summary of the subjective and objective data collected in C1, C2 and C3 during the listening tests conducted in the main phase is presented. Subjective data, i.e. semantic differentials ratings, are discrete variables and they were treated as ordinal scale; objective measures, i.e. acoustic metrics, are continuous variables and they were treated as interval scale.
Since in C2 and C3 just one stimulus for each subgroup (e.g. HLC-PVC) was submitted, variance was not calculated for the objective data.

A summary of the subjective and objective data collected in C1, C2 and C3 during the listening tests. The mean values, variance and median obtained in C1, C2 and C3 are reported for the subjective data of each chair-flooring matching. The mean values and variance in C1 are reported for each chair-flooring matching. The mean values of the acoustic metrics of the rolling sounds submitted for each chair-flooring matching are reported for C2 and C3.

| | | | | subjective data | | | | objective data | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| cond. | chair | flooring | | SD1 | SD2 | SD3 | SD4 | $L_{Aeq}$ [dB] | $L_{AFmax}$ [dB] | L [sone] | S [acum] | R [asper] | FS [vacil] | Ton |
| C1 | HLC | PVC | $\bar{x}$ | 2.97 | 3.43 | 2.83 | 3.2 | 51.7 | 63.2 | 9.74 | 0.91 | 2.1 | 2.22 | 0.06 |
| | | | $\sigma^2$ | 2.79 | 2.05 | 2.97 | 2.3 | 8.4 | 44.81 | 3.74 | 0.01 | 0.16 | 0.48 | 0 |
| | | | M | 2 | 3 | 2 | 3 | | | | | | | |
| | | ceramics | $\bar{x}$ | 3.9 | 5.07 | 5.3 | 4.9 | 51.1 | 68.4 | 13.55 | 1.07 | 2.65 | 2.95 | 0.05 |
| | | | $\sigma^2$ | 2.23 | 2 | 2.01 | 1.89 | 32.66 | 34.78 | 9.59 | 0.01 | 0.18 | 0.56 | 0 |
| | | | M | 4 | 5 | 6 | 5 | | | | | | | |
| | | wood | $\bar{x}$ | 2.5 | 2.93 | 2.53 | 3.03 | 52.8 | 63.2 | 9.93 | 1.2 | 2.33 | 2.28 | 0.07 |
| | | | $\sigma^2$ | 1.84 | 1.65 | 2.53 | 2.17 | 12.75 | 55.77 | 5.45 | 0.02 | 0.13 | 0.8 | 0 |
| | | | M | 2 | 3 | 2 | 3 | | | | | | | |
| | LLC | PVC | $\bar{x}$ | 2.9 | 3.9 | 2.77 | 3.77 | 50.8 | 64.1 | 11.14 | 0.91 | 2.48 | 2.13 | 0.05 |
| | | | $\sigma^2$ | 1.89 | 2.44 | 1.98 | 3.22 | 22.86 | 2.4 | 1.15 | 0 | 0.08 | 0.05 | 0 |
| | | | M | 3 | 4 | 2 | 3 | | | | | | | |
| | | ceramics | $\bar{x}$ | 4.2 | 5.33 | 5.6 | 5.5 | 49.3 | 68.8 | 15.8 | 1.2 | 3.01 | 2.66 | 0.05 |
| | | | $\sigma^2$ | 1.89 | 1.82 | 2.25 | 1.78 | 28.39 | 2.97 | 2.33 | 0 | 0.09 | 0.22 | 0 |
| | | | M | 5 | 6 | 6 | 6 | | | | | | | |
| | | wood | $\bar{x}$ | 2.83 | 3.27 | 2.27 | 3.17 | 52.8 | 64.2 | 11.91 | 1.08 | 2.69 | 2.29 | 0.06 |
| | | | $\sigma^2$ | 1.59 | 2.2 | 1.31 | 2.42 | 10.31 | 2.56 | 1.34 | 0.01 | 0.1 | 0.08 | 0 |
| | | | M | 2 | 3 | 2 | 3 | | | | | | | |
| C2 | HLC | PVC | $\bar{x}$ | 3.2 | 3.63 | 4.23 | 3.4 | 56.4 | 74 | 12.02 | 0.92 | 2.01 | 2.07 | 0.05 |
| | | | $\sigma^2$ | 1.68 | 2.31 | 2.74 | 2.04 | | | | | | | |
| | | | M | 3 | 4 | 5 | 3 | | | | | | | |
| | | ceramics | $\bar{x}$ | 4.3 | 4.57 | 4.37 | 4.57 | 52.3 | 75.2 | 18.62 | 1.19 | 2.84 | 2.25 | 0.03 |
| | | | $\sigma^2$ | 2.15 | 3.01 | 3.83 | 3.56 | | | | | | | |
| | | | M | 4 | 5 | 5 | 5 | | | | | | | |
| | | wood | $\bar{x}$ | 2.83 | 3.1 | 3.3 | 2.63 | 53 | 78.6 | 12.48 | 1.24 | 1.86 | 2.32 | 0.09 |
| | | | $\sigma^2$ | 1.87 | 2.16 | 3.11 | 2.24 | | | | | | | |
| | | | M | 2 | 3 | 3 | 2 | | | | | | | |
| | LLC | PVC | $\bar{x}$ | 3.23 | 3.8 | 4.2 | 3.5 | 62.7 | 62.5 | 9.98 | 0.85 | 2.28 | 1.86 | 0.05 |
| | | | $\sigma^2$ | 2.19 | 1.82 | 2.72 | 2.81 | | | | | | | |
| | | | M | 4 | 4 | 5 | 4 | | | | | | | |
| | | ceramics | $\bar{x}$ | 4.07 | 4.73 | 4.87 | 4.73 | 47.9 | 69.2 | 16.7 | 1.18 | 2.92 | 3.17 | 0.05 |
| | | | $\sigma^2$ | 3.17 | 2.55 | 3.5 | 3.24 | | | | | | | |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | M | 5 | 5 | 6 | 5 | | | | | | | |
| | | **wood** | $\bar{x}$ | 2.5 | 2.5 | 3.13 | 2.37 | 46 | 63.9 | 12.5 | 1.26 | 2.54 | 3.24 | 0.06 |
| | | | $\sigma^2$ | 2.26 | 2.12 | 3.5 | 2.31 | | | | | | | |
| | | | M | 2 | 2 | 3 | 2 | | | | | | | |
| **C3** | **HLC** | **PVC** | $\bar{x}$ | 2.47 | 3.23 | 3.13 | 2.87 | 57 | 58.1 | 10.15 | 1.64 | 2.25 | 4.43 | 0.07 |
| | | | $\sigma^2$ | 1.57 | 2.05 | 1.84 | 1.77 | | | | | | | |
| | | | M | 3 | 3 | 3 | 3 | | | | | | | |
| | | **ceramics** | $\bar{x}$ | 5.03 | 4.9 | 5.13 | 4.97 | 57.9 | 72.5 | 21.61 | 1.8 | 4.08 | 5.05 | 0.02 |
| | | | $\sigma^2$ | 1.83 | 2.71 | 2.67 | 2.93 | | | | | | | |
| | | | M | 5 | 5 | 5 | 6 | | | | | | | |
| | | **wood** | $\bar{x}$ | 3.47 | 4 | 4.03 | 3.73 | 54 | 59.6 | 10.62 | 1.58 | 2.04 | 4.45 | 0.07 |
| | | | $\sigma^2$ | 2.53 | 2.62 | 2.38 | 3.03 | | | | | | | |
| | | | M | 3 | 4 | 4 | 4 | | | | | | | |
| | **LLC** | **PVC** | $\bar{x}$ | 2.97 | 3.63 | 3.83 | 3.2 | 53.7 | 60.6 | 10.39 | 1.43 | 2.22 | 4.4 | 0.04 |
| | | | $\sigma^2$ | 2.1 | 2.1 | 3.11 | 2.44 | | | | | | | |
| | | | M | 3 | 4 | 4 | 3 | | | | | | | |
| | | **ceramics** | $\bar{x}$ | 4.93 | 4.93 | 4.67 | 4.9 | 56.4 | 73 | 25.88 | 1.93 | 3.72 | 4.88 | 0.04 |
| | | | $\sigma^2$ | 2.13 | 2.34 | 2.99 | 3.13 | | | | | | | |
| | | | M | 5 | 5 | 5 | 5 | | | | | | | |
| | | **wood** | $\bar{x}$ | 2.87 | 3.6 | 2.73 | 3.03 | 54.8 | 65.1 | 12.01 | 1.52 | 2.69 | 4.52 | 0.04 |
| | | | $\sigma^2$ | 2.67 | 2.32 | 2.2 | 2.38 | | | | | | | |
| | | | M | 3 | 4 | 2 | 3 | | | | | | | |

Note: C1 = context condition; C2 = laboratory condition; C3 = SounBe condition; HLC = high level chair; LLC = low level chair; SD1 = calm-strident; SD2 = pleasant-unpleasant; SD3 = smooth-rough; SD4 = not annoying at all-very annoying; $L_{Aeq}$ = A-weighted equivalent sound pressure level; $L_{AFmax}$ = maximum level with A-weighted frequency response and Fast time constant; L = Loudness; S = Sharpness; R = Roughness; FS = Fluctuation Strength; Ton = Tonality; $\bar{x}$ = mean value; $\sigma^2$ = variance; M = median.

## Appendix B

A summary of the original semantic differential pairs by Von Bismarck, the Italian translations, and the short sentences presented for the semantic differential pairs selection is shown.

On the left columns, the Von Bismarck semantic differential pairs in English and the Italian translations (in bold the best translation selected by the bilingual subjects). On the right columns, the Italian sentence presenting the descriptor, and the English translation in brackets.

| English version | Italian version | Italian sentence presenting the *descriptor*<br>(English translation of the Italian sentence presenting the *descriptor*) |
|---|---|---|
| calm | **calmo**, tranquillo | Quel suono *calmo* mi ha tranquillizzata<br>(That *calm* sound comforted me a lot) |
| strident | acuto, **stridente**, stridulo | Quel suono *stridente* mi ha fatto venire la d'oca<br>(That *strident* sound gave me gooseflesh) |
| pleasant | **gradevole,** piacevole | Mi piace sentirlo parlare, la sua voce è *gradevole*<br>(I like to hear him talking, his voice is *pleasant)* |
| unpleasant | **sgradevole,** spiacevole | Quel suono *sgradevole* non mi ha fatto avvicinare<br>(That *unpleasant* sound didn't make me move closer) |
| smooth | liscio, piatto, **regolare** | Quel suono era talmente *regolare* che dopo un po' non lo sentivo più<br>(That sound was so *smooth* that after some times I couldn't even perceive it) |
| rough | **irregolare**, grezzo, ruvido | La fontana scrosciava con un suono *irregolare*<br>(The fountain poured down with a *rough* sound) |
| harmonious | **armonico**, musicale | Le note musicali di questa sinfonia creano un suono *armonico*<br>(The music notes of this symphony generate a *harmonious* sound) |

| | | |
|---|---|---|
| inharmonious | **disarmonico**, discordante | Quel suono *disarmonico* mi ha proprio infastidito<br>(That *inharmonious* sound annoyed me) |
| gentle | delicato, **lieve** | Durante il compito in classe il chiacchiericcio era *lieve*<br>(During the assignment, the chatting sound was *gentle*) |
| hard | **duro**, forte | Quel suono *duro* ha richiamato la mia attenzione<br>(That *hard* sound grabbed my attention) |
| sharp | **penetrante**, pungente, tagliente | Quel suono era così *penetrante* che mi ha sconvolto<br>(That sound was so *sharp* that it shocked me) |
| dull | soffocato, **sordo**, tenue | Quel suono *sordo* si sentiva appena<br>(That *dull* sound was barely audible) |
| loud | alto, **forte**, rumoroso | La TV è *forte*, non riusciamo a sentirci<br>(The TV is *loud*: we cannot hear each other) |
| soft | **delicato**, tenue, tranquillo | Mi parlava con voce *delicata* e suadente<br>(He spoke me with *soft* and mellow voice) |
| high | **acuto**, alto, forte | Lo strillo *acuto* del bambino svegliò il fratello<br>(The *high* scream of the child woke up his brother) |
| low | **basso**, grave | Parlavamo a voce *bassa* per non svegliare gli altri<br>(We were speaking with *low* voice, not to wake up the others) |
| thick | **corposo**, spesso | Quel suono *corposo* mi fece vibrare<br>(That *thick* sound made me shake) |
| thin | debole, flebile, **sottile** | Con la sua voce *sottile* potrebbe fare la solista nel coro<br>(With her *thin* voice, she could sing as a choir soloist) |
| weak | atono, **debole**, leggero | Era troppo *debole* per essere udito<br>(It was too *weak* for being audible) |
| powerful | forte, **potente** | Quell'attore ha una voce *potente*<br>(That actor has a *powerful* voice) |
| metallic | metallico, **tintinnante** | Quel suono *tintinnante* e acuto ha richiamato la mia attenzione<br>(That *metallic* sound caught my attention) |
| deep | basso, **profondo** | Cantava Blues con voce *profonda*<br>(He sang Blues with a *deep* voice) |
| harsh | **aspro**, stridente, stridulo | Mi ha rimproverato con voce *aspra*<br>(He scolded me with a *harsh* voice) |
| mild | **gentile**, mite | Con tono garbato e *gentile* mi spiegò dove stavo sbagliando<br>(With a polite and *mild* tone, she explained me where I was making a mistake) |
| beautiful | **bello**, splendido | Un così *bel* suono non lo si sente tanto spesso<br>(Such a *beautiful* sound is not so frequent) |
| ugly | **brutto**, sgradevole | Quel *brutto* suono ha fatto piangere il bambino<br>(That *ugly* sound made the child cry) |