Title: The acoustic summary as a tool for representing urban sound environments

Authors:

Damiano OLDONI, damiano.oldoni@ugent.be, Department of Information Technology, Ghent University, Belgium

Bert DE COENSEL, bert.decoensel@intec.ugent.be, Department of Information Technology, Ghent University, Belgium

Annelies BOCKSTAEL, annelies.bockstael@intec.ugent.be, Department of Information Technology, Ghent University, Belgium

Michiel BOES, michiel.boes@intec.ugent.be, Department of Information Technology, Ghent University, Belgium

Bernard DE BAETS, bernard.debaets@ugent.be, Department of Mathematical Modelling, Statistics and Bioinformatics, Ghent University, Belgium

Dick BOTTELDOOREN, dick.botteldooren@intec.ugent.be, Department of Information Technology, Ghent University, Belgium

Details corresponding author

Bert DE COENSEL,

Ghent University

Department of Information Technology

Sint-Pietersnieuwstraat 41

9000 Gent

Belgium

phone: +32-9-264-9996

email: bert.decoensel@intec.ugent.be

Highlights:

- The acoustic summary of a place is a collection of representative sounds
- Acoustic summaries of several urban and quiet area locations are constructed using an automated procedure
- A validation test with local residents assesses the quality of the acoustic summaries
- Local residents can easily identify the acoustic summary extracted at the location of their own dwelling
- A group of sounds describes the uniqueness of a place, rather than single sounds by themselves

1 Introduction

Livability of the urban environment has always been a compelling issue for urban planners. Citizen well-being is related to the quality of the urban environment in different ways. Person-environment mismatch at the dwelling may lead to stress and related health impacts (Lazarus, 1991) but also the quality of the public space is of utmost importance. High quality public spaces stimulate social cohesion, recreation, and physical activity (Bedimo-Rung, Mowen, & Cohen, 2005). The role of urban green areas in particular has been investigated extensively in this respect. Several studies from the last decades indicate that people's psychological restoration and well-being is enhanced by direct access to nature and restorative areas (Hartig, Böök, Garvill, Olsson, & Gärling, 1996; Kaplan, 1983, 1985; Ulrich, 1981; Ulrich et al., 1991), by visual access to such areas from the dwelling (Kaplan, 1993, 2001; Ulrich, 1984) and by their perceived availability (Gidlöf-Gunnarsson & Öhrström, 2007). The positive role played by such areas has mainly been studied from the perspective of visual diversity, naturalness and aesthetics. However, the role of the soundscape and in particular quietness and tranquility is increasingly being stressed (Gidlöf-Gunnarsson & Öhrström, 2007). Therefore, there is an increasing awareness of the fact that the sonic environment forms an essential component of the urban environment that requires as careful planning as the landscape (Carles, Barrio, & de Lucio, 1999; Liu, Kang, Behm, & Luo, 2014; Liu, Kang, Luo, Behm, & Coppack, 2013; Zhang & Kang, 2007). However, it is also shown that landscape and soundscape planning should not be tackled independently, as landscape indicators have a non-negligible impact on the soundscape (Liu et al., 2013, 2014).

Classically, urban sound has been treated as a waste product to be tackled with suitable noise control policies, for which the most popular and visible tool has been extensive noise mapping. However, the final goal of planning and designing urban environments is not only noise abatement, but the creation of spaces with matching positive acoustic qualities (Botteldooren, De Coensel, Van Renterghem, Dekoninck, & Gillis, 2008). This approach, typically referred to as the *soundscape approach*, is getting increasing multidisciplinary attention and is the subject of several projects and studies (Adams et al., 2006; Brown, Kang & Gjestland, 2011; Pijanowski et al., 2011a; Pijanowski et al., 2011b; Zhang & Kang, 2007). As the soundscape concept extends beyond the sonic or acoustic environment and includes the way it is perceived and understood by a typical user of the space and within a particular context, the tools at the disposal of the urban sound planner and soundscape designer should account for human auditory perception (Oldoni et al., 2013).

Today, physical registration of relevant acoustical parameters is commonly accepted as a first soundscape analysis step (Schulte-Fortkamp, Brooks, & Bray, 2008), followed by an evaluation of the perceptual effects by techniques such as targeted interviews and questionnaires, preferably involving community members who live at the location under study (Brooks, 2006; Axelsson, Nilsson, Hellström, & Lundén, 2014). The combination of these two approaches is called *combined soundscape analysis* (Adams et al., 2006; Schulte-Fortkamp et al., 2008) and it is often deployed by means of soundwalks, in which sound measurements and perception interviews are conducted simultaneously. In a research perspective, the results are combined in order to find quantitative relationships between physical sound indicators and perceptual attributes (Berglund & Nilsson, 2006; Liu et al., 2014). Soundwalks are a popular methodology for understanding outdoor soundscapes (Adams et al., 2008), but they are inherently short-term and typically include only daytime. For this reason, several long-term strategies have been developed, mainly based on mobile sound measurements and community involvement, e.g. with public workers such as local police officers (Schulte-Fortkamp et al., 2008). This approach is surely more detailed and complete, but requires a considerable organizational effort and regular and constant participation, resulting in feasibility and reproducibility issues. In both short and long term approaches, a methodology for systematically selecting and recording a comprehensive collection of sounds that is representative for the sonic environment in the way that it is perceived and understood by so-called “local experts” – inhabitants and visitors – would mean a significant step forward in soundscape methodology.

In this paper a neural-network-based model is proposed that automatically constructs an *acoustic summary*, i.e. a collection of sounds that are likely to be noticed at a particular location and together represent the sonic environment at that location. The acoustic summary can provide a quick overview of the sounds present at a specific location, thus being a useful tool for the urban planner and the soundscape designer. In contrast to most of the computational auditory scene analysis (CASA) models (see Wang & Brown (2006) for an overview), the major interest here does not lie in extracting as clean as possible sound samples for all components of the auditory scene. On the contrary, the intention is to summarize the sonic environment using only those sounds that a human observer, not particularly focusing its attention to environmental sound, would notice. Note this explicit limitation of the acoustic summary to holistic listening only. Listening is a process that can develop at different cognitive levels, and it could be attentive and analytic rather than holistic. However, within attentive and analytic listening, top-down information is taken into account, which is much harder to implement in a computational model.

The proposed model partly takes inspiration from specific CASA techniques for extracting salient fragments of the auditory scene but it is also inspired by mechanisms underlying human bottom-up attention (Duangudom & Anderson, 2007; Kalinli & Narayanan, 2007; Kayser, Petkov, Lippert, & Logothetis, 2005). Moreover, most CASA techniques are not context dependent. Distinguishing between frequently occurring sounds and out-of-context or rarely occurring sounds is a crucial aspect in constructing an acoustic summary. For this reason, besides a biologically inspired auditory processing model, learning is a very important aspect in the presented model. It is implemented by means of a neural network called *Self-Organizing Map* (SOM) or Kohonen Map (Kohonen, 2001) and a specifically tailored learning technique. Furthermore, the model attempts to create a compromise between biological accuracy and computational efficiency as the model is to be integrated in equipment for long-term outdoor measurement and the data processing underlying the decision whether or not to record particular sound events has to be performed in real-time.

The structure of this paper is as follows: Section 2 describes the neural-network based model to construct the acoustic summary. Section 3 is dedicated to the results of a validation test performed by local residents in order to assess how accurately the acoustic summary is representing the sound environment in their neighborhood. Section 4 discusses the results and future developments. Finally, in Section 5 conclusions are presented.

2 Methods

2.1 Overview

Constructing the acoustic summary requires a computational analysis of the auditory scene that mimics how a human observer would split this auditory scene in its relevant components. Considering the application of the model in long-term outdoor measurement stations, computational efficiency has to be considered. For this reason, existing detailed auditory processing models for loudness (Glasberg & Moore, 2002), masking (Glasberg & Moore, 2005) and auditory saliency (Kayser et al., 2005) are replaced by simplified versions. The proposed model is comprised of two main stages, illustrated in Fig. 1: (I) during the learning phase, a self-organizing map (SOM) is tuned to the typical sounds at the given location based on the sound level and its spectrum, and (II) during the acoustic summary formation phase, for each class of sounds thus obtained, prototypes are recorded to compile the acoustic summary. Real-time operation is required in the second stage due to the limited sound buffer of typical outdoor measurement stations. In both stages, the sound signal recorded by the microphone is first treated in a similar way as in the human peripheral auditory system (I.a and II.a), whereby both a set of acoustical features is extracted and a measure of auditory saliency is calculated. The learning stage classifies the acoustical features based on co-occurrence (I.b) using the incremental SOM algorithm and a training technique called *Continuous Selective Learning* (CSL) that was developed specifically for this purpose. Once the learning has ended, the trained SOM can be used for automatically triggering the recording of typical and salient sounds, and in this way incrementally forming a library of prototypical sounds (II.b). The acoustic summary is then compiled by selecting a small number of sounds from this sound library, based on a ranking method (II.c). In this paper three different ranking methods will be presented and validated.

2.2 Sound feature extraction

The sound feature extraction stage of the proposed model is highly inspired by a model for auditory attention that was developed earlier by the authors (Oldoni et al., 2013). A 1/3-octave band spectrum with a temporal resolution of 0.125 s is calculated starting from the raw audio signal. This temporal resolution was chosen based on the typical temporal envelope of urban environmental sounds (De Coensel & Botteldooren, 2006; De Coensel, Botteldooren, & De Muer, 2003), and allows to capture the temporal dynamics of most of the typical urban environmental sound sources. To account for energetic masking, a simplified cochleogram $s(f,t)$ is then calculated based on the Zwicker loudness model (Zwicker & Fastl, 1999) covering the complete audible frequency range (0 to 24 Bark) with a spectral resolution of 0.5 Bark, resulting in 48 spectral values at each time step. The auditory system is, in addition to absolute intensity, also sensitive to spectro-temporal irregularities (Alain, Arnott, & Picton, 2001; Bregman, 1994; Houtgast, 1989; Yost, 1992). The proposed model therefore calculates measures for intensity, spectral and temporal modulation using a center-surround mechanism (Schreiner, Read, & Sutter, 2000), based on auditory saliency models (Duangudom & Anderson, 2007; Kalinli & Narayanan, 2007; Kayser et al., 2005). More in detail, a convolution of the cochleogram with 16 2D Gaussian and difference-of-Gaussian filters is performed in parallel at each time step, resulting in a set of multi-scale features called the *sound feature vector*, consisting of 16 x 48 = 768 values. This set of values characterizes the loudness, spectral and temporal structure of the sound at each time step. The corresponding 768-dimensional vector space will be referred to as the *sound feature space*. More technical details about the sound feature extraction can be found in Oldoni et al. (2010). Finally, a scalar value called the *overall auditory saliency* is calculated from the sound feature vector, according to the algorithm developed by De Coensel and Botteldooren (2010).

2.3 Learning

The feature vector provides extensive information about the sonic environment at a given time step. Analysis of the sonic environment should usually last for a long period ranging from a few days to several weeks, depending on the richness in sounds of the sonic environment at the given location. The crucial point is how to use such a large amount of data to construct a concise but exhaustive acoustic summary. In this paper a neural-network-based approach is proposed, which makes use of a self-organizing map. Several topographic maps have been observed in the visual and auditory cortex (Heil, Rajan, & Irvine, 1994; Kayser, Petkov, Augath, & Logothetis, 2007; Morel & Kaas, 1992; Yin, 2008) and the SOM has been originally conceived as an abstract mathematical model of such topographic mapping. Moreover, the SOM is typically described as an unsupervised learning-based method for clustering and visualizing high-dimensional data (Kohonen, 1998), another important aspect to take into account due to the high-dimensionality of the sound feature space. In the framework of the present model, the SOM should eventually learn which features belong to the same auditory object based on co-occurrence. Furthermore, the size of a representational area of a sound in the primary auditory cortex is closely related to its importance (Rutkowski & Weinberger, 2005) and the strength of the memory effect (Bieszczad & Weinberger, 2010), an aspect of auditory learning that is very well modeled by a SOM and the CSL algorithm which will be described later in this section. As mentioned in Section 1, context dependency should be considered while selecting sounds for constructing an acoustic summary. Knowing the context can entail familiarity with the sonic environment and it has been shown that familiarity with the sound to be detected makes the detection easier (Lewis, Talkington, Puce, Engel, & Frum, 2011). The extensive training on sound feature vectors at the microphone location tunes the SOM to the typical sounds composing the local sound environment and thus makes the system "familiar" with them.

The SOM used in our model is a 2D network of 3750 equal-spaced units in a regular hexagonal lattice. Each unit has an associated reference vector in the high-dimensional sound feature space. The initial values of the reference vectors are calculated by means of principal component analysis on an input data subset as in Kohonen (1998). After initialization, reference vector coordinates are modified during a first training phase which is based on the Original Incremental SOM Algorithm (Kohonen, 2001). For this, sound feature vectors stemming from a particular recording location are presented to the SOM. At each time step, the unit with reference vector that most closely matches the current sound feature vector is selected (commonly called the *best-matching unit* or BMU). The reference vector of the BMU, and to a lesser extent the reference vectors of the neighboring units in the 2D lattice, are then moved closer to the input feature vector. After this initial training phase, the reference vectors of the SOM units can be seen as a non-linear discrete 2D mapping of the probability density function of the sound feature vectors used for training. In particular, some regions of the sound feature space contain more reference vectors than others, thus preserving the high-dimensional relationships underlying the input feature vectors (Kohonen, 2001). When positioning a new sound feature vector with respect to the trained SOM, the distance to the BMU gives an indication of the similarity of the current sound to earlier encountered sounds. When the distance to the BMU is small, a very similar sound was encountered before, during the training phase.

The learning algorithm described above is purely based on frequency of occurrence and does not take into account the fact that human perception and retrospective assessment of a sonic environment also depends on the saliency of the sounds. Salient sound events would be better noticed and remembered than less salient ones (Ranganath & Rainer, 2003), even if they do not occur that often. Therefore, the SOM trained with the original incremental SOM algorithm is used as a starting point for a second much longer training phase which implements (continuous) selective learning (Oldoni, 2015; Oldoni et al. 2013). The instantaneous overall auditory saliency, scaled as a number between zero and one, is used for modulating the learning rate parameter during the selective learning phase (Oldoni, 2015): the learning based on sound feature vectors whose related saliency values are higher than 0.5 is enhanced (by moving the reference vector of the BMU and neighboring units closer to the input feature vector by a greater amount), while learning based on feature vectors corresponding to sounds with lower saliency is somewhat suppressed (by moving the reference vector of the BMU and neighboring units closer to the input feature vector by a lesser amount). The second goal of using saliency in selective learning is to reduce the number of SOM units whose reference vectors are related to often occurring but non-relevant sounds, such as the urban background hum, and to increase the number of SOM units that are related to sound events. At each time step, the BMU is found as before. However, not all input sound feature vectors are used as inputs during the selective learning: a learning phase is triggered only if the distance to the BMU is higher than an activation threshold T_{up} (indicating the presence of a sound that has not been encountered before). All subsequent input vectors are then selected as inputs for training, until the distance to the BMU drops below a deactivation threshold T_{down}. Furthermore, sound feature vectors occurring a few seconds before the triggered learning period are included. In this paper, a 2-second pre-trigger period is used, corresponding to 16 time steps. The thresholds T_{up} and T_{down} are chosen in such a way that less than 10% of all sound feature vectors are used as input for selective learning. After some weeks of running the CSL, it is observed that the SOM can identify – in terms of distance to the BMU – most of the sounds occurring in the acoustic environment for which the SOM was trained (Oldoni, 2013).

In order to visualize the effects of training on the SOM reference vectors, the so-called U-matrix (Ultsch, 1993) is used. This matrix shows the distances between the reference vectors related to each pair of neighboring SOM units. The effects of the CSL on the clustering of SOM units can be seen in Figure 2 where the U-matrix after the first training using the original incremental SOM algorithm is shown next to the U-matrix of the final SOM after the CSL phase. By means of a color coding, the U-matrix allows to distinguish groups of SOM units with similar reference vectors (small distances between neurons, in white) form areas with high variability (large distances between neurons, in black). After the first initial training, the SOM is generally still characterized by large distances between all neurons. The contours of only one “valley” are visible at the left side, related to background hum. In contrast, after the CSL phase, the SOM shows much more structure, various valleys are visible, corresponding to different categories of sounds.

2.4 Sound sample retrieval and selection

The reference vectors associated to the trained SOM units can be seen as representative abstract sound prototypes, encoded by their sound feature vectors. Once a SOM is trained, it can be used for constructing a library of sounds, whereby sound samples that are most similar in the sound feature space to the sound prototypes within the SOM are recorded. As shown in the schematic overview in Figure 1, the first step in constructing the acoustic summary is calculating feature vectors for the sound observed at each time step as explained in Section 2.2. The BMU is then selected, and the distance between its reference vector and the current sound feature vector is calculated. Based on this distance, sound recording is triggered if the selected SOM unit has not been the BMU before (meaning that the encountered sound has not occurred before during the sound sample retrieval phase), or if the distance to the BMU is smaller than any earlier distance for this BMU (meaning that a better matching sound sample is encountered). These steps have to be taken with low latency due to the limited audio recording buffer of typical acoustical measurement equipment. Sound samples are recorded from 3 seconds before to 2 seconds after the recording trigger, for a total sound sample duration of 5s. This duration has been heuristically found to be sufficient for producing an overall impression of the sound at a particular instant in time. It turns out that, for typical urban soundscapes, for the bulk of the SOM units a representative audio sample is found after a few days of sound sample retrieval. This set of sounds can be seen as a sound library describing the sound environment at the measurement location.

The large number of audio samples that is gathered through the procedure described above is unpractical for easy exploration of the given sound environment by listening. For this reason, three ranking criteria are presented, which can be used to select a subset of sounds that is most representative for the given sound environment; this subset is then called the acoustic summary. The first proposed ranking criterion is based on saliency: the higher the saliency, the more likely the sound sample will be representative and the higher its ranking. As explained in Section 2.2, a measured overall saliency value can be calculated at each time step from the sound feature vector. The SOM reference vectors lie in the sound feature space, therefore saliency values can be calculated for each of the units, resulting in a saliency overlay on the SOM. A second criterion is based on how often each of the SOM units was selected as the BMU during a given time interval, typically one day or more, resulting in a frequency of occurrence overlay on the SOM. As mentioned in Section 2.3, the frequency of occurrence of sounds is not likely to be a sufficient criterion to represent the sounds that will be noticed and remembered. For this reason, a third intermediate method is proposed, in which a linear combination between both saliency and frequency of occurrence of each SOM unit is performed:

$$c_i = \beta_{occ} \cdot \frac{\log(o_i + 1)}{\log N} + \beta_{sal} \cdot s_i,$$

where c_i is the combined ranking value of the SOM unit (and thus the associated sound), o_i is the number of time steps for which the SOM unit i is the BMU, N is the total number of samples used for calculating the frequency of occurrence, s_i is the saliency of unit i and β_{occ} and β_{sal} are two positive weighting coefficients between 0 and 1 so that $\beta_{occ}+\beta_{sal}=1$. In case $\beta_{occ}=1$ is chosen, selection is performed purely on the basis of frequency of occurrence; in case $\beta_{sal}=1$ is chosen, selection is performed purely on the basis of saliency. Any intermediate value represents a trade-off between both extremes.

The number of sounds to be selected depends on the envisaged use of the acoustic summary. In the validation test that will be discussed in Section 3, 32 sounds for each criterion have been selected based on their ranking. An a posteriori justification for selecting exactly this number of sounds is given in Section 4.

3 Validation test

3.1 Overview

A validation test has been designed to check the representativeness of the automatically generated acoustic summaries for an urban sound environment. Sound recording devices were installed at 6 locations in and around the Belgian city of Ghent, that will be referred to as Bi, Ko, Bu, Sp, Be, and Dr. In Table 1 the day-evening-night equivalent sound level, L_{den}, and a qualitative description of the sonic environment for each location is given. Four locations Bi, Ko, Bu and Sp are situated in urbanized areas, Be is located in the very heart of the city, while Dr is in the suburbs. Sound recording devices were installed on a windowsill along the front façade of dwellings. Such a configuration is not standard for environmental noise level measurements, where microphones are usually placed at 1m from the façade, in order to remove the influence of façade reflection on the sound level. However, for the purpose of audio recording, this is a less important issue, and simply placing the devices on the windowsill is much more cost-effective. Sixteen people living in the surroundings of the sound recording devices placed in Bi, Ko, Bu and Sp were contacted for participating in the test as local residents, four per location, based on the proximity of their dwelling to the microphone positions. Recruitment was carried out by putting flyers with an invitation to participate in a listening experiment in the mailbox; the reward was one movie ticket. In Table 2 the gender and age of the participants is listed. Very few people were living in the direct surroundings of the devices placed in Be and Dr, so nobody was contacted from these two locations. The acoustic summaries from these two locations were therefore exclusively used as confounders and their quality was not assessed by the validation test. For this reason, Bi, Ko, Bu and Sp will be referred to as group 1 in the remainder of the paper, while locations Be and Dr will be referred to as group 2.

For each participant, three locations were selected at the beginning of the test. The first selected location was always the location from group 1 where the participant lived. The two other locations were randomly selected: one location was chosen among the others of group 1, and one among the two locations of group 2. The validation test itself was composed of four consecutive experiments, followed by a small questionnaire in which comments could be formulated. The test duration was not fixed and varied among the participants from 30 minutes up to one hour, as participants could listen to all sounds presented as much as wanted or needed. A portable computer with high quality sound card and a closed-type Sennheiser HD-280 PRO headphone were used for the experiment. The complete experiment, including display of instructions, audio presentation, data collection and timing, was automated using a graphical user interface in Matlab. A preliminary test was performed in order to select the correct sound level for the experiment and to ascertain the absence of hearing loss with each participant. The experiment took place either at the home of the participants or in a listening test room at the university laboratory, depending on the availability of the participants. In case the test was performed at the participant's home, quietness and the absence of distracters were considered a prerequisite. Before starting the experiment, the participants were informed about the general aim of the study; a verbal informed consent was provided by the participants.

3.2 Experiment 1

In the first experiment, the participants explored the sounds of the acoustic summaries of the three selected locations and had to select the one that they thought corresponded to the direct surroundings of their home (see Appendix A for a snapshot of the experiment). This experiment was repeated three times, with acoustic summaries constructed using each of the three criteria – saliency, frequency of occurrence and combined criterion – in randomized order. Each acoustic summary was visualized as a panel of 32 buttons, each corresponding to a different sound sample. A color map spanning from yellow to red was used to color the different buttons. Depending on the three different ranking criteria, the color encoded (1) the saliency value s_i, (2) the frequency of occurrence o_i, or (3) the combined value c_i of the corresponding SOM unit. To stress color differences, yellow was assigned to the smallest value and red to the highest value among the 32 values for s_i, o_i and c_i. Participants could listen to each of the sound samples as much as they wanted, by clicking the respective button, before selecting an acoustic summary from the three candidates shown in randomized order.

In Figure 3 the results of the first experiment are shown. In total 13 participants out of 16 correctly selected the acoustic summary that corresponded to the direct surroundings of their home for summaries constructed on the basis of saliency and on the basis of the combined criterion. Only 11 participants selected the correct acoustic summary in case it was constructed on the basis of frequency of occurrence. The few errors are not equally divided among the four locations included in this test. All participants at the locations Bi and Sp correctly recognized the acoustic summaries; at the location Bu only one error for both saliency and occurrence criteria occurred. The acoustic summaries from Ko were hardly recognized. The comments left by the participants suggest an overall lack of representativeness of the summaries for this location. This may be due to a combination of both site characteristics (e.g. the soundscape at that location may be more diverse than at the other locations) as well as model and recording characteristics (e.g. soundmarks were missed at that location). The overall representativeness of the summaries will be further discussed in Section 3.5 and Section 4. Overall, most errors were made for the acoustic summary formed by frequency of occurrence, followed by the combined criterion and then the saliency criterion.

In general, the high and similar number of correct answers for all three ranking-selecting criteria indicates that the sound library from which the sounds are selected is composed of typical and representative sounds for the given location. To further explore possible differences between the three criteria, the number of sounds to which each participant listened before making a choice is analyzed. From Figure 4 it is clear that participants decided faster in case of acoustic summaries based on saliency, while on average they needed to listen to the highest number of sounds for frequency of occurrence-based acoustic summaries. This could be due to the on average higher information content within the sounds, when they are selected based on saliency. In order to check if the differences in number of sounds played between the three selection criteria are statistically significant, a linear regression model $Y = ax + b$ was constructed, with Y the number of played sounds, $a = (a_1,a_2)$ the coefficients of the regression model, b the constant term of the regression and x a two-dimensional dummy variable encoding the different selection criteria, such that $x = (0,0)$ for the acoustic summary based on saliency, and $x = (1,0)$ and $x = (0,1)$ for the frequency of occurrence and the combined criterion respectively. After excluding the outliers in Figure 4, the null hypothesis H_0: $a_1 = a_2 = 0$ is rejected based on an overall F-test for regression: $F(2,40) = 3.42$, $p = 0.04$. This means that the selection criterion has a significant influence on the number of sounds played ($\alpha < 0.05$). In this regard, it should be noted that, although randomized, the order in which the summaries based on each of the three criteria were presented could have influenced the number of played sounds, even given that the acoustic summaries constructed using the different selection criteria contained different sounds. The order, also coded as a two-dimensional dummy variable, is thus added to the above regression model, and the null hypothesis H_0: $a_1 = a_2 = b_1 = b_2 = 0$ cannot be rejected this time, with $F(4,38) = 1.82$, $p = 0.14$. This implies that the order of presentation does not have a significant influence on the number of played sounds. Moreover, the adjusted $\bar{R}^2$ is the highest when the criterion is the only explanatory variable ($\bar{R}^2 = 0.10$) and it decreases if the order of presenting the three criteria is added to the regression model ($\bar{R}^2 = 0.07$). The same holds if such order is included in the regression equation as the only explanatory variable ($\bar{R}^2 = 0.02$). A further indication that the number of sounds is only influenced by the acoustic summary criterion and not by the order of presentation is given by an F-test comparing the two regression models. The extended regression model with the order added does not provide a significantly better fit: $F(2,38) = 0.34$, $p = 0.72$.

3.3 Experiment 2

In the second experiment, three acoustic summaries, all calculated for the location where the participant lives, but either formed by the saliency, the frequency of occurrence, or the mixed criterion were presented. The participants were asked to rank the presented fragments based on perceived accuracy in representing the surroundings of the participant's own home (see Appendix B for a snapshot of the experiment). The results of this experiment are shown in Figure 5 where frequency of the given ranks (1, 2, or 3) is depicted per acoustic summary. The acoustic summary based on frequency of occurrence is clearly considered the least representative: its cumulative distribution, shown in Figure 5 (b), lies under the cumulative distributions related to the other two criteria. Moreover, the cumulative distribution related to the combined criterion shows that the acoustic summary related to this criterion is ranked first or second by 15 out of 16 participants. In order to reject the null hypothesis of a discrete uniform distribution over the ranking, a Pearson's χ^2 test has been performed for each criterion, rejecting this hypothesis for both the frequency of occurrence ($\chi^2 = 6.13$, p = 0.95) and the combined criterion ($\chi^2 = 6.13$, p = 0.95). The same cannot be said about the ranking distribution related to the saliency-based criterion ($\chi^2 = 0.88$, p = 0.35), due to the non-negligible group of people considering it the least appropriate. A possible reason for it will be discussed in Section 4.

3.4 Experiment 3

In the third experiment, participants were asked to construct their own collection of sounds that represented the direct surroundings of their home, by selecting sounds from a set of 64 sounds (see Appendix C for a snapshot of the experiment). Half of the sounds from which the participants could choose were recorded at their home location, the other half were recorded at two other randomly chosen locations: 16 sounds at a location of group 1 and 16 sounds at a location of group 2. The participants were not told about such subdivision. All sounds belonged to acoustic summaries based on the combined criterion. This inclusion/exclusion of sounds in the final sound collection can be seen as a binary classification task; therefore it makes sense to define true and false positives or negatives. The sounds coming from the participant's location that were rightly selected by the participant are called true positives (TPs), while selected sounds recorded at other locations are called false positives (FPs). The true negatives (TNs) are the sounds from other locations correctly not selected and the false negatives (FNs) are the sounds from the surroundings of the participant's home that were not selected. The higher the number of TPs and TNs, the better the acoustic summary model has captured the peculiarities of the sound environment at each location.

An overview of the results for all participants is shown in Figure 6. The high variability among participants was to be expected. Nevertheless, 10 of the 16 participants scored TPs and TNs both greater than 16, with 16 being the expected result of a random guess. The *False Positive Ratio* (FPR) and the *True Positive Ratio* (TPR) are calculated and shown in Figure 7. The FPR is defined as the ratio between the FPs and the number of sounds from other locations, i.e. 32, while the TPR is the ratio between the TPs and the number of sounds from the participant's location, again 32. The higher the TPR and the lower the FPR are, the more convincing the acoustic summary. In Figure 7 one can see that all participants except one score better than a random guess (which would give a point along the diagonal line, the so-called line of no-discrimination). Moreover, the participant called Ko2 in Figure 6 is very far from this line too, showing that this participant was completely misled by the proposed sounds. In fact, from

Figure 6 it can be seen that he/she only selected sounds from the two other locations. The results of the third experiment support the findings from the first experiment. Participants from Bi and Sp –not making any mistake in the first experiment– scored on average better than participants from Bu, who, in turn, scored better than participants in Ko, as shown in Figure 8 where the accuracy, defined as (TPs+TNs)/64, is plotted. In addition, the participants from Ko show the highest variability: the first and second participant respectively have the best and the worst accuracy among all participants.

It can be noted that the accuracy of the participants from Ko follow the results they obtained during the first experiment: the first participant got the best score in the first experiment, making only one mistake, the third participant made two mistakes out of three, while the other two participants could never select their own acoustic summary. It is also worthwhile checking whether accuracy was influenced by the number of sounds played in the second experiment. Participants listened exclusively to sounds coming from their own surroundings just before performing the third experiment. So it could have been possible that correct selection in the third experiment was enhanced if more sounds had been listened to in the second experiment. An F-test on the simple linear regression model between accuracy and number of played sounds in the second experiment does not reject the null hypothesis of unrelated variables, i.e. a slope equal to zero ($F = 2.18$, $p = 0.16$). The same conclusion holds if precision, defined as TPs/(TPs+FPs), instead of accuracy is considered ($F = 1.13$, $p = 0.31$).

3.5 Experiment 4

In the last experiment, participants were asked to label 20 sounds that were randomly selected from the 32 sounds composing the saliency-based acoustic summary from their dwelling location (see Appendix D for a snapshot of the experiment). This experiment was followed by a small questionnaire in which each participant was asked to leave free comments about the experiment (see Appendix D). In an open question, it was asked whether there were sounds not heard in the labeling experiment that should have been included in order to better represent the surroundings of the participant's home. The comments, summarized in Table 3, are important hints to better understand the obtained results. For example, the comments written by the participants from Ko can explain their errors in the first experiment: three out of four were expecting the typical sounds of the market held each Sunday morning in their neighborhood. Those sounds were not present in the acoustic summaries because the sound sample retrieval was not running during any Sunday, thus missing the very specific so-called *soundmarks* of that location (Schafer, 1977). The same could be said about the comment of participant Ko2: the construction works referred to were a very recent activity, which started only after the sound sample retrieval stage was completed. In addition, the participants from Bu missed the typical sound of the elementary school located at the backside of their dwelling. These soundmarks were not recorded because the microphone was placed at the front façade of the dwelling. It is worth noting that the main remarks came from the participants living in Ko and Bu, which were the only ones making errors during the first experiment.

4 Discussion

The main rationale behind this work was to introduce a new way of investigating the acoustic environment at a particular location based on sounds instead of visual maps or other visually-based methods. The first idea emerging from this study is the importance of soundmarks in describing a soundscape: any acoustic summary which lacks soundmarks would be considered to be less representative, as occurred in Ko or, to a lesser extent, in Bu. Typically, soundmarks have a very specific temporal pattern and occurrence, thus sound sample retrieval needs to run continuously in order to include also these potentially less frequent, but highly relevant soundmarks. In Ko, for example, the sounds produced on Sunday by the music bands and by visitors of the flower market are important soundmarks, not captured by sound sample retrieval and therefore not included in the acoustic summary. This lack is the principal cause of the wrong answers for experiment 1.

Together with soundmarks, spatiality also plays an important role in defining the soundscape. The present research focused on the front façade, where one would have assumed to find the majority of characteristic sounds, but it can happen that soundmarks can only be observed at the other side of the dwelling, as occurred in Bu. Participants appear to be capable of taking these spatial differences into account when judging the acoustic summaries; despite the lack of typical school sounds, participants from Bu scored quite well thanks to typical sounds from the front façade. The results from the third experiment demonstrate that, in general, participants can identify “their” sounds better than random guessing. Moreover, the results from the first and the third experiment suggest that the representativeness of an acoustic summary is a direct consequence of the representativeness of each sound composing it: the summaries that were composed of non-representative sounds were also not recognized. Nevertheless, the number of false negatives and false positives cannot be neglected in general: the sound samples composing an acoustic summary can, most of the time, be associated to more than one location, if they are considered separately from the other sounds. Therefore, results of this experiment confirm the validity of using an acoustic summary for representing or evoking a soundscape. Considered as a whole, such a collection of sounds can be much more representative of the uniqueness of a sonic environment than each single sound on itself that is part of the acoustic summary.

The finding that most participants were able to answer correctly given the limited number of sounds played, suggests that 32 is a sufficient number of sounds for an acoustic summary to characterize a location. Thus, selecting such a limited set of sounds is as crucial as the sound sample retrieval itself: it would make no sense to continuously retrieve sound samples if the soundmarks and other typical sounds would not be selected for the acoustic summary afterwards. In this work, the number of sounds composing the acoustic summary was heuristically determined and was the same for all locations. However, the richness of a soundscape depends intrinsically on the considered location. Our model could therefore be improved in future, considering acoustic summaries composed by a variable number of sounds. For example, a measure of the overall similarity among the SOM reference vectors could be used to determine the richness of the sonic environment at a given location, and consequently the number of sound samples that should be selected.

The second experiment confirms that frequency of occurrence is not the best criterion for selecting the sounds composing the acoustic summary. In many locations the sounds selected based on this criterion are typically very quiet, especially in residential areas or parks, thus missing the less often occurring but much more salient sounds. Hence, saliency is a better criterion for constructing the acoustic summary, but there is still a non-negligible group of people considering it the least appropriate. Selecting only highly salient sounds typically comes down to selecting loud sounds, and an excessive number of such fragments is no longer representative of the sound environment in urban residential areas. Therefore, a combination of frequency of occurrence and saliency was conceived and tested. The second experiment demonstrates that such a combination is a simple and valid strategy for representing a soundscape in the way a human would. Based on these results, more advanced processing models could be tested in the future, for example, adding human-like top-down attention mechanisms in the model as in Boes, Oldoni, De Coensel, and Botteldooren (2013, 2014). In the present work, a fixed sound sample duration of 5s was used; however, every sound event has its own typical duration and it should be preserved in order to better represent the sound events composing the acoustic summary. The model presented by Boes et al. (2013, 2014) could help to solve this issue.

5 Conclusions

This work presents a computational model for constructing a comprehensive and representative collection of sounds that are present at a given location. Such a collection, called an acoustic summary, can be a useful tool for quickly presenting and analyzing the sound environment at a given location. The model consists of two stages: in a first stage, a Self-Organizing Map is tuned to the typical sounds at the given location, and, in a second stage, an acoustic summary is constructed by first collecting and then selecting specific sound samples based on the trained map. The model takes into account aspects of human auditory perception, such as bottom-up selective attention and learning.

A listening test involving local residents has been performed to evaluate the ability of the model to produce acoustic summaries representative of the sound environment at a number of urban locations. The test demonstrated that the model can construct representative acoustic summaries. In particular, the model produces broad and satisfactory sound libraries from which the acoustic summary can be extracted. In general, satisfactory results are obtained from all the three tested criteria used for selecting representative audio samples from the sound library to compose the acoustic summary. However, the acoustic summary criterion combining saliency and frequency of occurrence of the sound events generally produces the best acoustic summary. The saliency-based criterion produces good acoustic summaries as well but risks overweighing highly informative and salient sounds. In addition, participants judged the acoustic summaries based on frequency of occurrence alone to be the least representative due to the prevalence of quiet sounds, which are much less informative of the given soundscape, even though they occur very often in residential areas. Finally, the test demonstrated that only a few sounds are needed to represent the sound environment of an urban area, confirming the choice of 32 sounds for each location.

The procedure for calculating acoustic summaries introduced in this work has already been automated and implemented in low-cost sound measurement hardware (Botteldooren et al., 2013), such that a plug-and-measure device can be put outside, and after a few weeks the set of sounds comprising the acoustic summary at that location is available online. Nevertheless, the potential of the acoustic summary tool for representing and analyzing an existing sound environment would still be sensibly improved by wrapping it in a user-friendly application at the disposal of urban planners or any other interested end users. Furthermore, the approach outlined in this work allows to compile an acoustic summary for a virtual acoustic environment in the same way as it would for any existing one. Although the challenge of acoustic design of urban space has attracted sporadic attention since long, during the past decade, research interest has risen considerably, partly driven by the advent of realistic environmental simulation models, such as auralization. Substantial progress in this field can be expected during the coming years; increasingly efficient and accurate physics-based methods may soon make it possible to render virtual acoustic scenes that cannot be distinguished from real auditory environments. Combining computational models of auditory perception of environmental sound, such as the acoustic summary presented in this paper, with state-of-the-art auralization would put the results of this work on the cutting edge of this field, promoting a multisensory approach in creating the soundscape of future cities.

References

Adams, M., Cox, T., Moore, G., Croxford, B., Refaee, M., & Sharples, S. (2006). Sustainable soundscapes: Noise policy and the urban experience.*Urban Studies*, 43(13), 2385-2398. doi - 10.1080/00420980600972504

Adams, M., Bruce, N. S., Davies, W. J., Cain, R., Jennings, P., Carlyle, A., … Plack, C. (2008). Soundwalking as a methodology for understanding soundscapes. Institute of Acoustics (IOA). *Spring Conference of the Institute of Acoustics 2008: Widening Horizons in Acoustics: Proceedings of a meeting held in Reading, UK, 10-11 April 2008* (p 708). Salford, UK: Acoustics Research Centre, University of Salford.

Alain, C., Arnott, S. R., & Picton, T. W. (2001). Bottom-up and top-down influences on auditory scene analysis: Evidence from event related brain potentials. *Journal of Experimental Psychology: Human Perception and Performance*, 27(5), 1072-1089. doi - 10.1037/0096-1523.27.5.1072

Axelsson, Ö., Nilsson, M. E., Hellström, B., & Lundén, P. (2014). A field experiment on the impact of sounds from a jet-and-basin fountain on soundscape quality in an urban park. *Landscape and Urban Planning*, 123, 49-60. doi - 10.1016/j.landurbplan.2013.12.005

Bedimo-Rung, A. L., Mowen, A. J., & Cohen, D. A. (2005). The significance of parks to physical activity and public health: a conceptual model. *American journal of preventive medicine*, *28*(2), 159-168. doi - 10.1016/j.amepre.2004.10.024

Berglund, B., & Nilsson, M. E. (2006). On a tool for measuring soundscape quality in urban residential areas. *Acta Acustica united with Acustica*, 92(6), 938-944.

Bieszczad, K. M., & Weinberger, N. M. (2010). Representational gain in cortical area underlies increase of memory strength, *Proceedings of the National Academy of Sciences*, 107(8), 3793-3798. doi - 10.1073/pnas.1000159107

Boes, M., Oldoni, D., De Coensel, B., & Botteldooren, D. (2013). A biologically inspired recurrent neural network for sound source recognition incorporating auditory attention. In *Proceedings of the International Joint Conference on Neural Networks* (pp. 596–603), Dallas, TX, USA. doi - 10.1109/IJCNN.2013.6706791

Boes, M., Oldoni, D., De Coensel, B., & Botteldooren, D. (2014). Long-term learning behavior in a recurrent neural network for sound recognition. In *Proceedings of the International Joint Conference on Neural Networks* (pp. 3116–3123), Beijing, China. doi - 10.1109/IJCNN.2014.6889658

Botteldooren, D., De Coensel, B., Van Renterghem, T., Dekoninck, L., & Gillis, D. (2008). The urban soundscape: a different perspective. In G. Allaert & F. Witlox (Eds.), *Duurzame mobiliteit Vlaanderen: de leefbare stad* (pp. 177–204). Presented at the Duurzame mobiliteit Vlaanderen: de leefbare stad, Gent, België: Universiteit Gent. Instituut voor Duurzame Mobiliteit.

Botteldooren, D., Van Renterghem, T., Oldoni, D., Dauwe, S., Dekoninck, L., Thomas, P., Wei, W., Boes, M., De Coensel, B., De Baets, B., & Dhoedt, B. (2013). The internet of sound observatories. *Proceedings of Meetings on Acoustics*, 19(1), 040140. doi – 10.1121/1.4799869

Bregman, A. S. (1994). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: The MIT Press.

Brooks, B. M. (2006). Traditional measurement methods for characterizing soundscapes. *The Journal of the Acoustical Society of America*, 119(5), 3260. doi - 10.1121/1.4786099

Brown, A. L., Kang, J., & Gjestland, T. (2011). Towards standardization in soundscape preference assessment. *Applied Acoustics*, 72(6), 387-392. doi - 10.1016/j.apacoust.2011.01.001

Carles, J. L., Barrio, I. L., & de Lucio, J. V. (1999). Sound influence on landscape values. *Landscape and Urban Planning*, 43(4), 191-200. doi - 10.1016/S0169-2046(98)00112-1

De Coensel, B., & Botteldooren, D. (2006). The quiet rural soundscape and how to characterize it. *Acta Acustica united with Acustica*, 92(6), 887-897.

De Coensel, B., & Botteldooren, D. (2010). A model of saliency-based auditory attention to environmental sound. In Burgess, M., Don, C., Davey, J., & McMinn, T., *Proceedings of 20th International Congress on Acoustics, ICA*, Sidney, Autralia, 23–27 August 2010 (pp. 3480-3487). Red Hook, NY: Curran Associates, Inc.

De Coensel, B., Botteldooren D., & De Muer, T. (2003). 1/f Noise in Rural and Urban Soundscapes. *Acta Acustica united with Acustica*, 89(2), 287-295.

Duangudom, V., & Anderson, D. V. (2007). Using auditory saliency to understand complex auditory scenes. In Domański, M., Stasiński, R., & Bartkowiak, M. (Eds.), *Proceedings of the 15th European Signal Processing Conference* (EUSIPCO 2007), Poznań, Poland, 3-7 September 2007 (pp. 1206-1210).

Gidlöf-Gunnarsson, A., & Öhrström, E. (2007). Noise and well-being in urban residential environments: The potential role of perceived availability to nearby green areas. *Landscape and Urban Planning*, 83(2), 115-126. doi - 10.1016/j.landurbplan.2007.03.003

Glasberg, B. R., & Moore, B. C. J. (2002). A model of loudness applicable to time-varying sounds. *Journal of the Audio Engineering Society*, 50(5), 331-342.

Glasberg, B. R., & Moore, B. C. J. (2005). Development and evaluation of a model for predicting the audibility of time-varying sounds in the presence of background sounds. *Journal of the Audio Engineering Society*, 53(10), 906-918.

Grahn, P., & Stigsdotter, U. K. (2010). The relation between perceived sensory dimensions of urban green space and stress restoration. *Landscape and Urban Planning*, 94(3), 264-275. doi - 10.1016/j.landurbplan.2009.10.012

Hartig, T., Böök, A., Garvill, J., Olsson, T., & Gärling, T. (1996). Environmental influences on psychological restoration. *Scandinavian Journal of Psychology*, 37(4), 378-393. doi - 10.1111/j.1467-9450.1996.tb00670.x

Heil, P., Rajan, R. & Irvine, D. R. F. (1994). Topographic representation of tone intensity along the isofrequency axis of cat primary auditory cortex. *Hearing Research*, 76(1), 188-202. doi - 10.1016/0378-5955(94)90099-X

Houtgast, T. (1989). Frequency selectivity in amplitude-modulation detection. *The Journal of the Acoustical Society of America*, 85, 1676. doi - 10.1121/1.397956

Kalinli, O., & Narayanan, S. (2007). A saliency-based auditory attention model with applications to unsupervised prominent syllable detection in speech. In International Speech Communication Association (ISCA) (Eds.), *8th Annual Conference of the International Speech Communication Association (Interspeech 2007)*, Antwerp, Belgium, 27-31 August 2007 (pp. 1941-1944). Red Hook, NY: Curran Associates, Inc.

Kaplan, R. (1983). The role of nature in the urban context. In: Altman, I., Wohlwill, J. F. (Eds.), *Behavior and the natural environment* (pp. 127-161). New York – New York City. doi - 10.1007/978-1-4613-3539-9_5

Kaplan, R. (1985). Nature at the doorstep: Residential satisfaction and the nearby environment. *Journal of Architectural and Planning Research*, 2(2), 115-127.

Kaplan, R. (1993). The role of nature in the context of the workplace. *Landscape and urban planning*, *26*(1), 193-201. doi - 10.1016/0169-2046(93)90016-7

Kaplan, R. (2001). The nature of the view from home psychological benefits. *Environment and Behavior*, *33*(4), 507-542. doi - 10.1177/00139160121973115

Kayser, C., Petkov, C. I., Augath, M., & Logothetis, N. K. (2007). Functional imaging reveals visual modulation of specific fields in auditory cortex. *The Journal of Neuroscience*, 27(8), 1824-1835. doi - 10.1523/JNEUROSCI.4737-06.2007

Kayser, C., Petkov, C. I., Lippert M., & Logothetis, N. K. (2005). Mechanisms for Allocating Auditory Attention: An Auditory Saliency Map. *Current Biology*, 21(15), 1943-1947. doi - 10.1016/j.cub.2005.09.040Kohonen, T. (1998). The self-organizing map. *Neurocomputing*, 21(1-3), 1-6. doi - 10.1016/S0925-2312(98)00030-7

Kohonen, T. (Ed.). (2001). *Self-Organizing Maps* (3rd ed.). Heidelberg.

Lazarus, R. S. (Ed.). (1991). *Emotion and adaptation*. London: Oxford University Press.

Lewis, J. W., Talkington, W. J., Puce, A., Engel, L. R., & Frum, C. (2011). Cortical networks representing object categories and high-level attributes of familiar real-world action sounds. *Journal of Cognitive Neuroscience*, 23(8), 2079-2101. doi - 10.1162/jocn.2010.21570

Liu, J., Kang, J., Behm, H., & Luo, T. (2014). Effects of landscape on soundscape perception: Soundwalks in city parks. *Landscape and Urban Planning*, 123, 30-40. doi - 10.1016/j.landurbplan.2013.12.003

Liu, J., Kang, J., Luo, T., Behm, H., & Coppack, T. (2013). Spatiotemporal variability of soundscapes in a multiple functional urban area. *Landscape and Urban Planning*, 115, 1-9. doi - 10.1016/j.landurbplan.2013.03.008

Morel, A., & Kaas, J. H. (1992). Subdivisions and connections of auditory cortex in owl monkeys. *Journal of Comparative Neurology*, 318(1), 27-63. doi - 10.1002/cne.903180104

Oldoni, D. (2015). *Computational modelling for soundscape analysis inspired by human auditory perception and its application in monitoring networks*. Doctoral dissertation. Ghent University. Faculty of Engineering and Architecture, Ghent, Belgium.

Oldoni, D., De Coensel, B., Boes, M., Rademaker, M., De Baets, B., Van Renterghem, T., & Botteldooren, D. (2013). A computational model of auditory attention for use in soundscape research. *The Journal of the Acoustical Society of America*, 134(1), 852-861. doi - 10.1121/1.4807798

Oldoni, D., De Coensel, B., Rademaker, M., Van Renterghem, T., De Baets, B., & Botteldooren, D. (2010). Context-dependent environmental sound monitoring using SOM coupled with LEGION. In Proceedings of the *IEEE International Joint Conference on Neural Networks (IJCNN)* (pp. 1413–1420), New York, NY, USA: IEEE. doi - 10.1109/IJCNN.2010.5596977

Pijanowski, B. C., Farina, A., Gage, S. H., Dumyahn, S. L., & Krause, B. L. (2011a). What is soundscape ecology? An introduction and overview of an emerging new science. *Landscape ecology*, 26(9), 1213-1232. doi - 10.1007/s10980-011-9600-8

Pijanowski, B. C., Villanueva-Rivera, L. J., Dumyahn, S. L., Farina, A., Krause, B. L., Napoletano, B. M., ... & Pieretti, N. (2011b). Soundscape ecology: the science of sound in the landscape. *BioScience*, 61(3), 203-216. doi - 10.1525/bio.2011.61.3.6

Ranganath, C., & Rainer, G. (2003). Neural mechanisms for detecting and remembering novel events. *Nature Reviews Neuroscience*, 4(3), 193-202. doi - 10.1016/j.cub.2005.09.040

Rutkowski, R. G., & Weinberger, N. M. (2005). Encoding of learned importance of sound by magnitude of representational area in primary auditory cortex, *Proceedings of the National Academy of Sciences of the United States of America*, 102(38), 13664-13669. doi - 10.1073/pnas.0506838102

Schafer, R. M. (Ed.). (1977). *The Tuning of the World*. New York, New York.

Schreiner, C. E., Read, H. L., Sutter, M. L. (2000). Modular organization of frequency integration in primary auditory cortex. *Annual Review of Neuroscience*, 23(1) 501-529. doi - 10.1146/annurev.neuro.23.1.501

Schulte-Fortkamp, B., Brooks, B. M., & Bray, W. R. (2008). Soundscape: An Approach to Rely on Human Perception and Expertise in the Post-Modern Community Noise Era. *Acoustics Today*, 3(1), 7-15. doi - 10.1121/1.2961148

Ulrich, R. S. (1981). Natural versus urban scenes some psychophysiological effects. *Environment and behavior*, 13(5), 523-556. doi - 10.1177/0013916581135001.

Ulrich, R. S. (1984). View through a window may influence recovery from surgery. *Science*, 224(4647), 224-225. doi - 10.1126/science.6143402

Ulrich, R. S., Simons, R. F., Losito, B. D., Fiorito, E., Miles, M. A., & Zelson, M. (1991). Stress recovery during exposure to natural and urban environments. *Journal of environmental psychology*, 11(3), 201-230. doi - 10.1016/S0272-4944(05)80184-7

Ultsch, A. (1993). Self-organizing neural networks for visualisation and classification. In *Information and classification Information and classification* (pp. 307-313). Springer Berlin Heidelberg. doi - 10.1007/978-3-642-50974-2_31

Valero, X., Alías, F., Oldoni, D., & Botteldooren, D. (2012). Support vector machines and self-organizing maps for the recognition of sound events in urban soundscapes. In C. Burroughs, C., & Conlon, S. (Eds.), *Proceedings of the 41st International Congress and Exposition on Noise Control Engineering (Internoise)*, New York City, NY: Curran Associates, Inc.

Wang, DL., & Brown, G. J. (Eds.). (2006). *Auditory Scene Analysis: Principles, Algorithms, and Applications*. New Jersey–Hoboken.

Yin, H. (2008). The Self-Organizing Maps: Background, Theories, Extensions and Applications. In Fulcher, J., & Jain, L. C. (Eds.), *Computational Intelligence: A Compendium* (pp. 715-762). Berlin, Springer-Verlag.

Yost, W. A. (1992). Auditory perception and sound source determination, *Current Directions in Psychological Science*, 1(6), 179-184. doi - 10.1111/1467-8721.ep10770385

Zhang, M., & Kang, J. (2007). Towards the evaluation, description, and creation of soundscapes in urban open spaces. *Environment and Planning B: Planning and Design*, 34(1), 68. doi - 10.1068/b31162

Zwicker, E., & Fastl, H. (Eds.). (1999). *Psychoacoustics. Facts and Models* (2nd ed.). Berlin, Springer-Verlag.

List of tables

Table 1

Coordinates, L_{den} (dBA) and qualitative description of the sonic environment at the six locations where the acoustic summary model has been tested. All the locations are situated in the Ghent municipality, five of them in the city, one in a suburban area a few kilometers from the city center. A KML file with the locations is available with the online version of the paper.

Location	Coordinates	L_{den} (dBA)	Description
Ko	51° 2' 59.6142" N, 3° 43' 26.0544" E	71.4	Urban square in the city center. Road traffic noise due to private and public transportation, noise from pedestrians and a music fanfare on Sunday. Microphone placed on a windowsill at the 3rd floor.
Bi	51°3'26.7588" N, 3°43'44.6880" E	61.3	Urban no-through street in the center of Ghent, mainly used for parking. Limited road traffic noise due to private transportation, noise from pedestrians and children playing from a recreational area in the neighborhood. Microphone placed on a windowsill at the 1st floor.
Sp	51°2'30.5262" N, 3°42'26.4852" E	65.5	Urban street in a residential area. Road traffic noise due to private and public transportation. Microphone placed on a windowsill at the 2nd floor.
Bu	51°1'54.7176" N, 3°43'38.0064" E	73.3	Urban street parallel to a railway. Road traffic noise due to private and public transportation, train noise. Microphone placed on a windowsill at the 3rd floor.
Be	51°3'15.6384" N, 3°43'31.0080" E	65.2	Urban street in a restricted traffic zone in the very heart of Ghent. Limited road traffic noise due to the transit of taxi and trucks for restaurant and shop delivery, noise from pedestrians due to the presence of the most important tourist attractions of the city and very distinct bell melodies from the nearby belfry.
Dr	51°3'14.4216" N, 3°38'37.4640" E	56.4	Quiet rural place, about 500 meters from a railway. Microphone placed in the backyard of a house in a countryside village.

Table 2

Gender and age of the participants in the experiment. The participants are identified by their location and a progressive number.

Participant	Gender	Age
Ko1	M	33
Ko2	M	31
Ko3	F	31
Ko4	M	44
Bi1	F	27
Bi2	M	39
Bi3	M	42
Bi4	M	34
Sp1	M	28
Sp2	M	30
Sp3	F	20
Sp4	F	21
Bu1	M	34
Bu2	M	22
Bu3	F	51
Bu4	M	23

Table 3

Main concepts expressed in the comments written by the participants after listening to and labeling 20 sounds randomly selected from the 32 sounds composing the acoustic summary based on saliency. In particular, the participants were asked whether there were sounds not heard in the labeling experiment that should have been included in order to better represent the surroundings of their home. The concepts are linked to the participants who wrote them.

Participant	Comment
Ko1, Ko3, Ko4	It would be nice to include sounds of the music bands playing on Sunday morning and during flower market on Sunday.
Ko2	I didn't hear noise samples of the construction works going on in the square where we live. Otherwise it was very representative. Ninety-five percent of the audio samples were traffic noises: it corresponds well to the amount of traffic we have in front of our apartment.
Bi1, Bi2, Bi3	No comment or positive remarks as “good representation, typical sounds and ambience”
Bi4	I would include some sounds from the music school at the other side of the street
Sp1, Sp3, Sp4	The sounds represent our street, especially the buses.
Sp2	More calm situations are needed.
Bu1, Bu2, Bu3	I miss the sounds of the back of the house, e.g. the children playing in the playground.
Bu4	Most of the sounds are present.

List of figures

ranking (a) and its cumulative distribution (b) are shown. Rank 1 means that the acoustic summary is considered “the most representative”, while rank 3 means “the least representative”.

Figure 6. Overview of the results of the third experiment. Participants were asked to make their own acoustic summary that represented the direct surroundings of their home, by selecting appropriate sounds among 64 sounds. The participants are denoted by a location acronym and a progressive number. The sounds from the participant’s location correctly selected, called true positives (TP), are shown in black; the sounds from a different location wrongly selected, called false positives (FP), are shown in dark grey; the sounds from the participant’s location not selected, called false negatives (FN), are shown in light grey; the sounds from other locations correctly not selected, called true negatives (TN), are shown in white.

Figure 7. Scatter plot of the True Positive Rate versus the False Positive Rate, calculated on the basis of the results shown in Figure 6. Different markers are chosen for the four locations from which the participants were recruited. The line of no-discrimination is also shown; a random guess would give on average a point on this line.

Figure 8. Accuracy in selecting one’s own acoustic summary, for all participants, subdivided by location.

Figure A1. Snapshot of the first experiment.

Figure B1. Snapshot of the second experiment.

Figure C1. Snapshot of the third experiment.

Figure D1. Snapshot of the fourth experiment.

Figure D2. Snapshot of the comment page.

List of Appendices

Appendix A. Title: Snapshot of the first experiment

In the first experiment the participants were asked to perform the following task:

In the pictures below you will discover a collection of sounds by clicking on different areas of these pictures. Each picture corresponds to a particular place in Ghent. The intensity of red color indicates how frequently each sound would be noticed at this place. One of the pictures corresponds to the direct surroundings of your home. Select the button below the one you think it is.

In figure A1 a snapshot of the first experiment is shown.

Appendix B. Title: Snapshot of the second experiment

In the second experiment the participants were asked to perform the following task:

In the pictures below you will discover a collection of sounds by clicking on different areas of these pictures representing the direct surroundings of your home. The intensity of red color indicates how frequently each sound would be noticed. Now please rank these pictures according how appropriate they are to the direct surroundings of your home. Type 1 for the most appropriate one, 3 for the least appropriate one.

In figure B1 a snapshot of the second experiment is shown.

Appendix C. Title: Snapshot of the third experiment

In the third experiment the participants were asked to perform the following task:

Now we would like you to make your own collection of sounds that represents the direct surroundings of your home. For this, select the appropriate sounds in the table below and indicate how frequently you hear them using the color scale.

In figure C1 a snapshot of the third experiment is shown.

Appendix D. Title: Snapshot of the fourth experiment

In the fourth experiment the participants were asked to perform the following task:

Finally, could you please name in your own language the following sounds recorded in the surroundings of your home?

In figure D1 a snapshot of the fourth experiment is shown. Afterwards, the participants were asked to leave free comments:

Thanks for your participation. Would you like to leave any comment about the experiment? In particular, are there sounds not heard in the last experiment which should have been included in order to represent the surroundings of your home?

In figure D2 a snapshot of the final comment page is shown.

Figure 1
Click here to download high resolution image

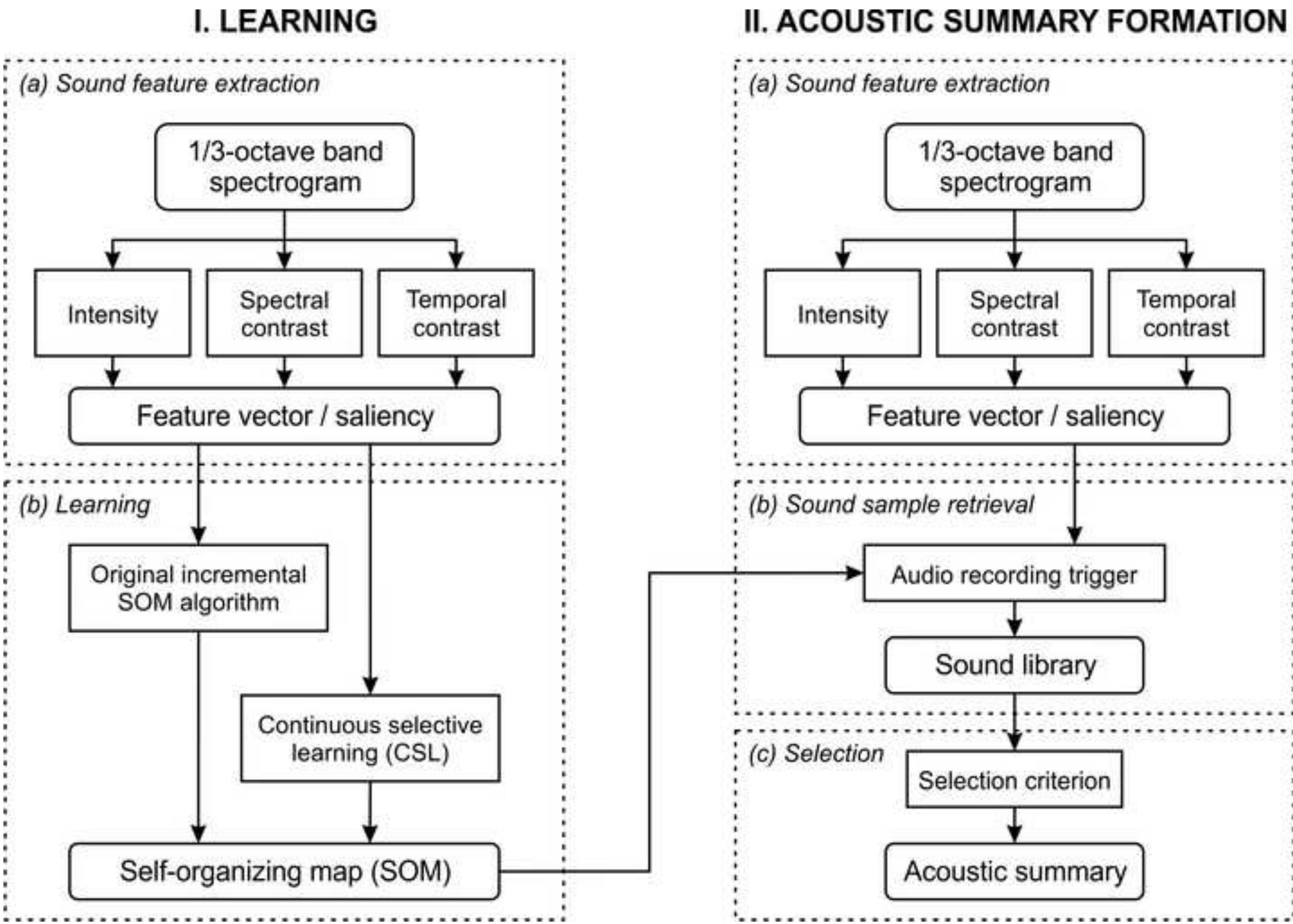


Figure 2
Click here to download high resolution image

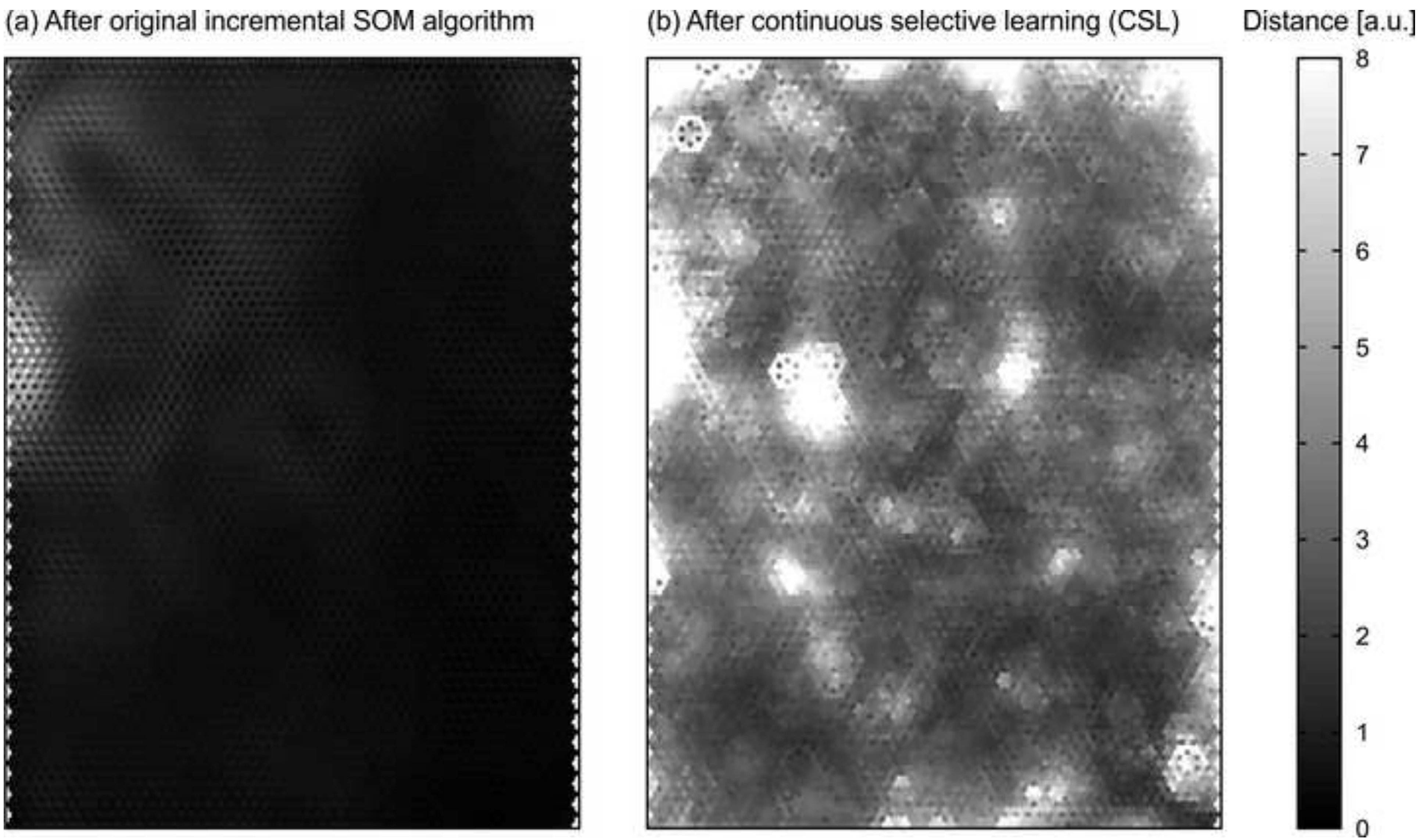


Figure 3

Click here to download high resolution image

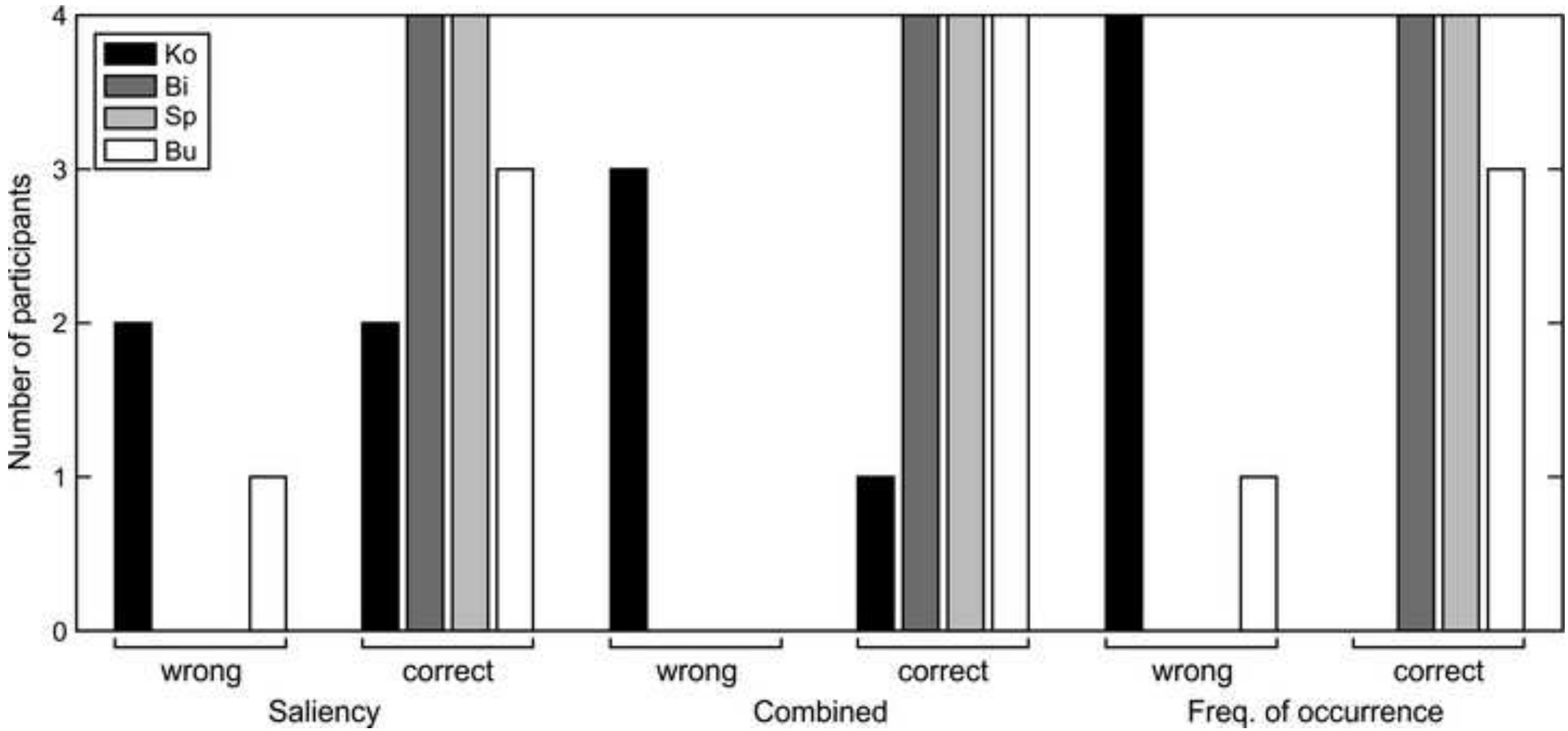


Figure 4
Click here to download high resolution image

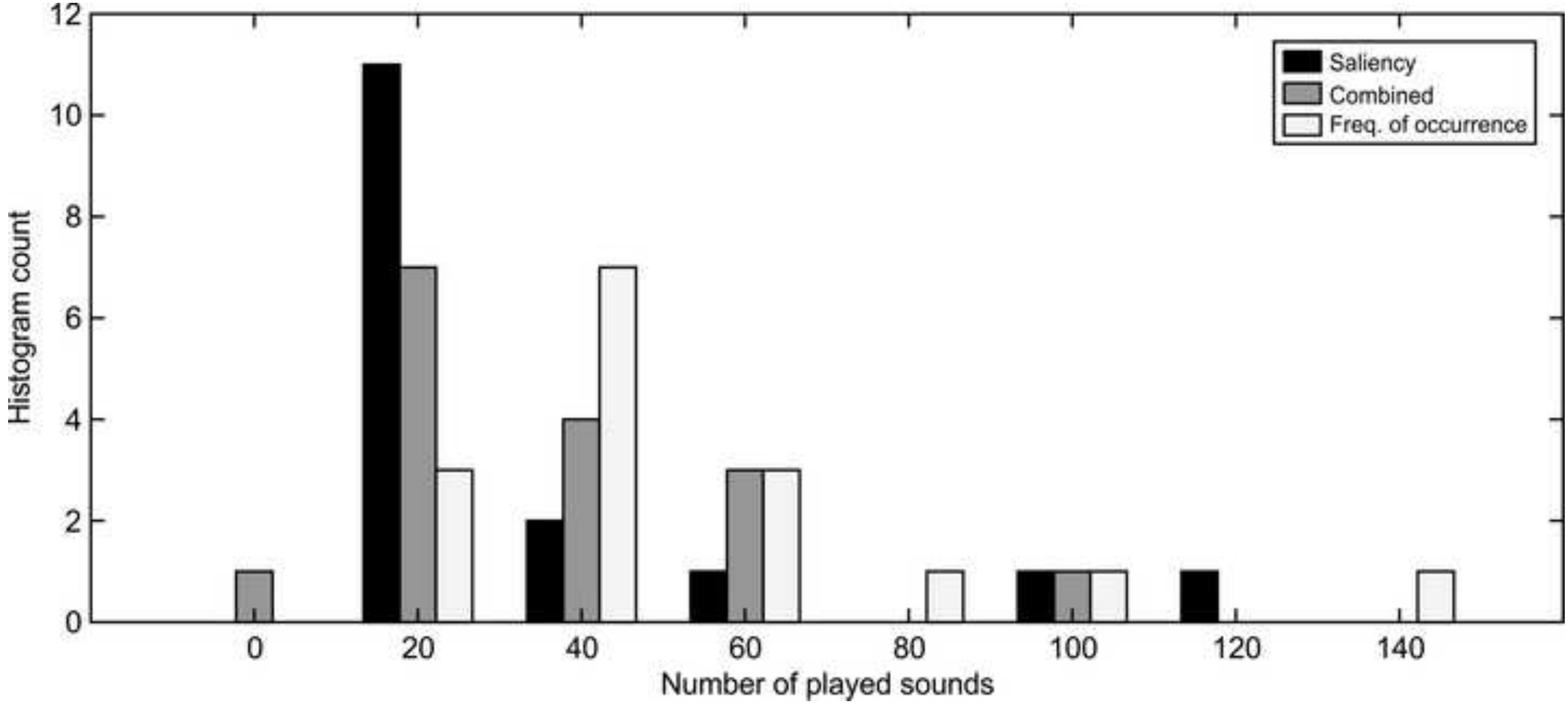


Figure 5
Click here to download high resolution image

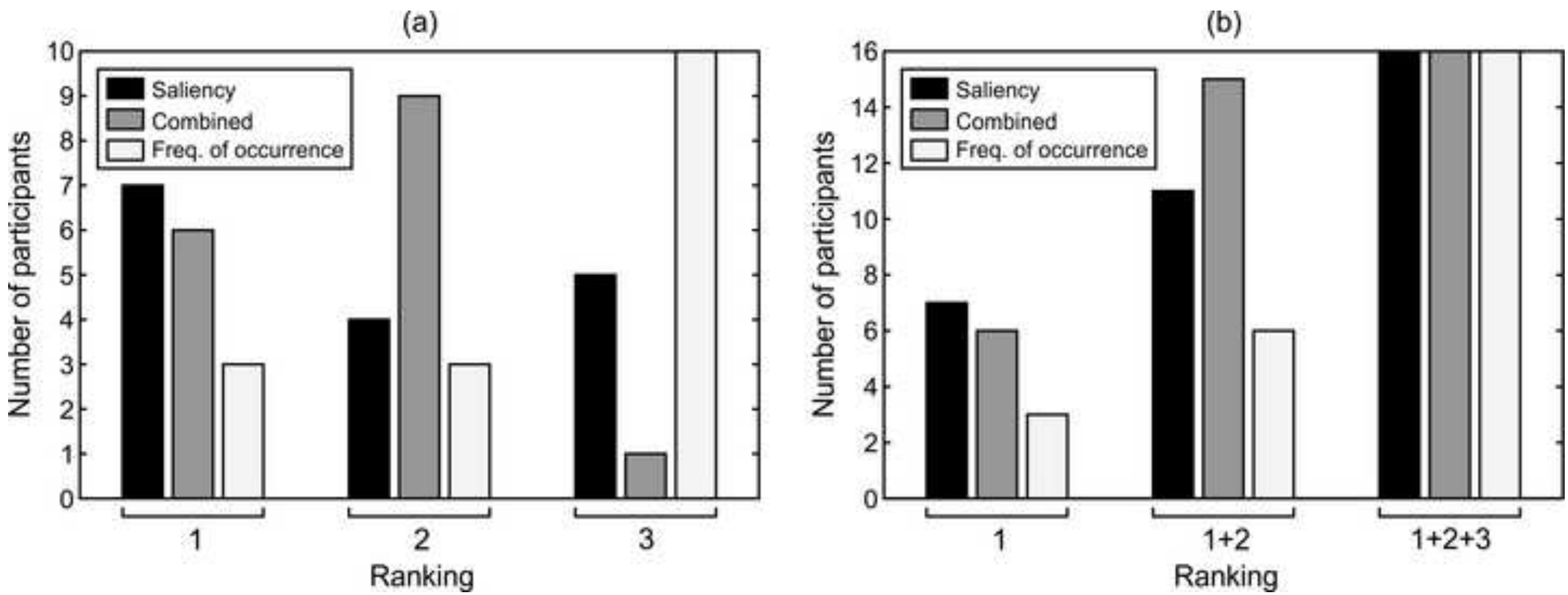


Figure 6
Click here to download high resolution image

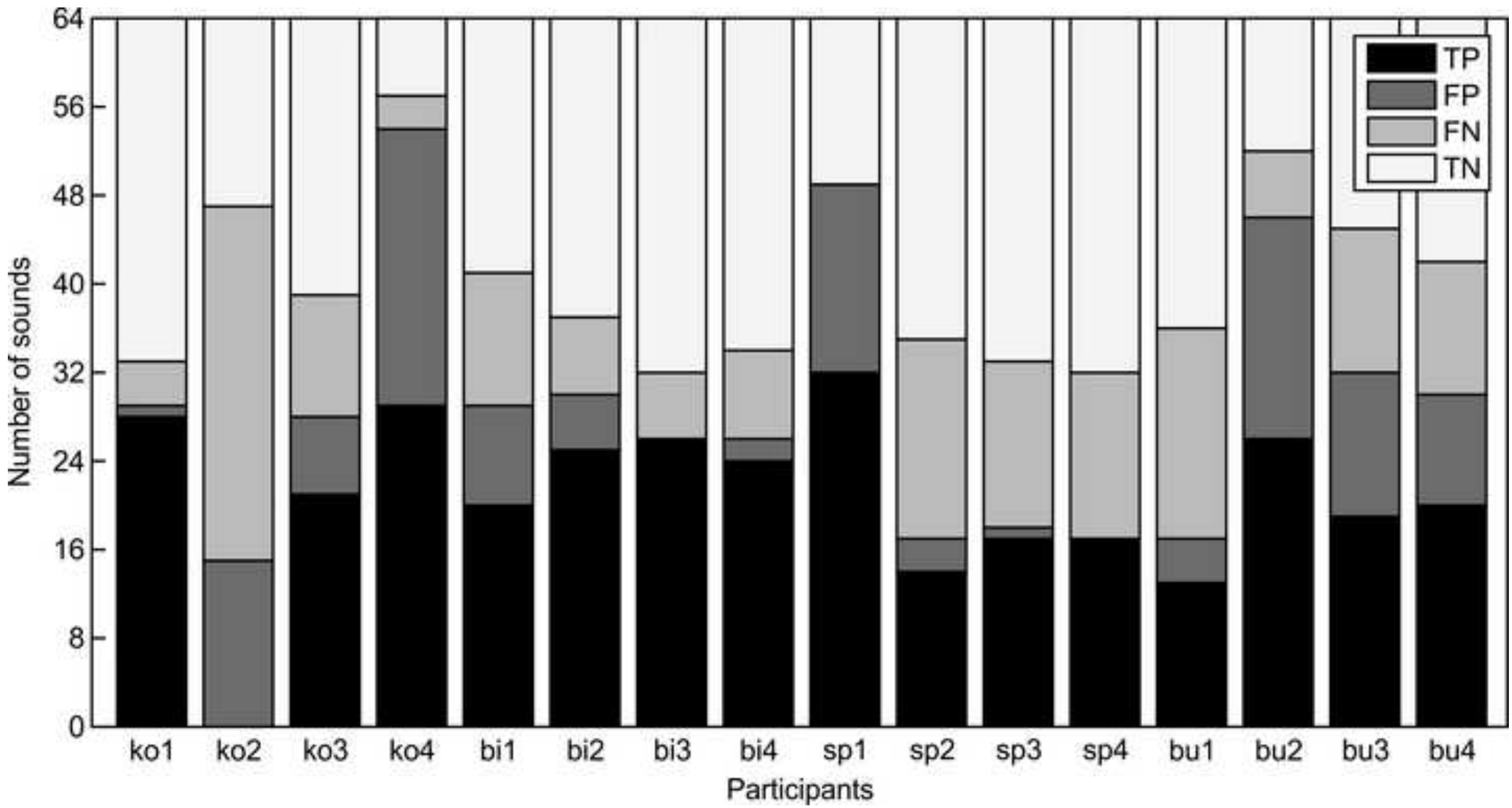


Figure 7
Click here to download high resolution image

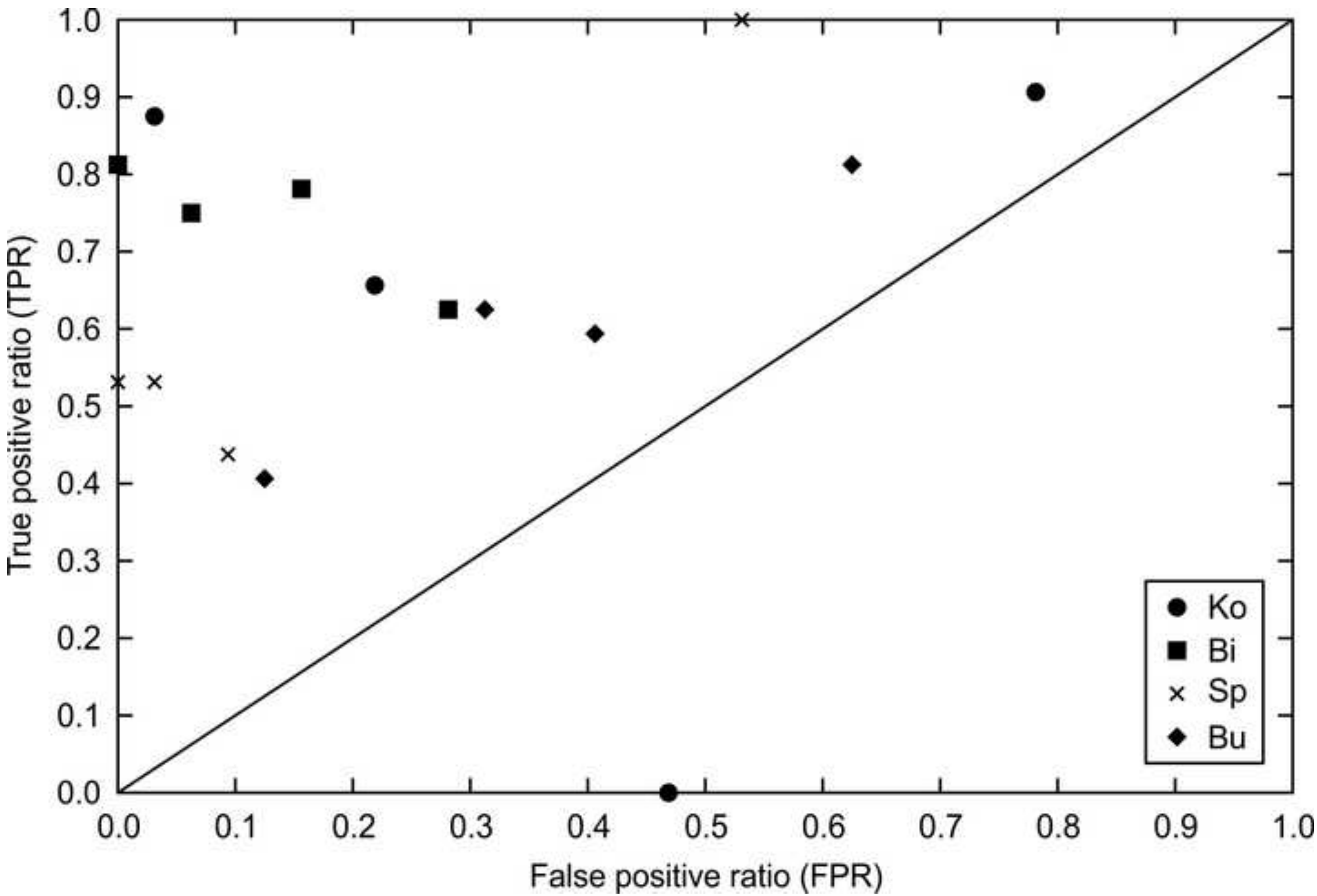


Figure 8
Click here to download high resolution image

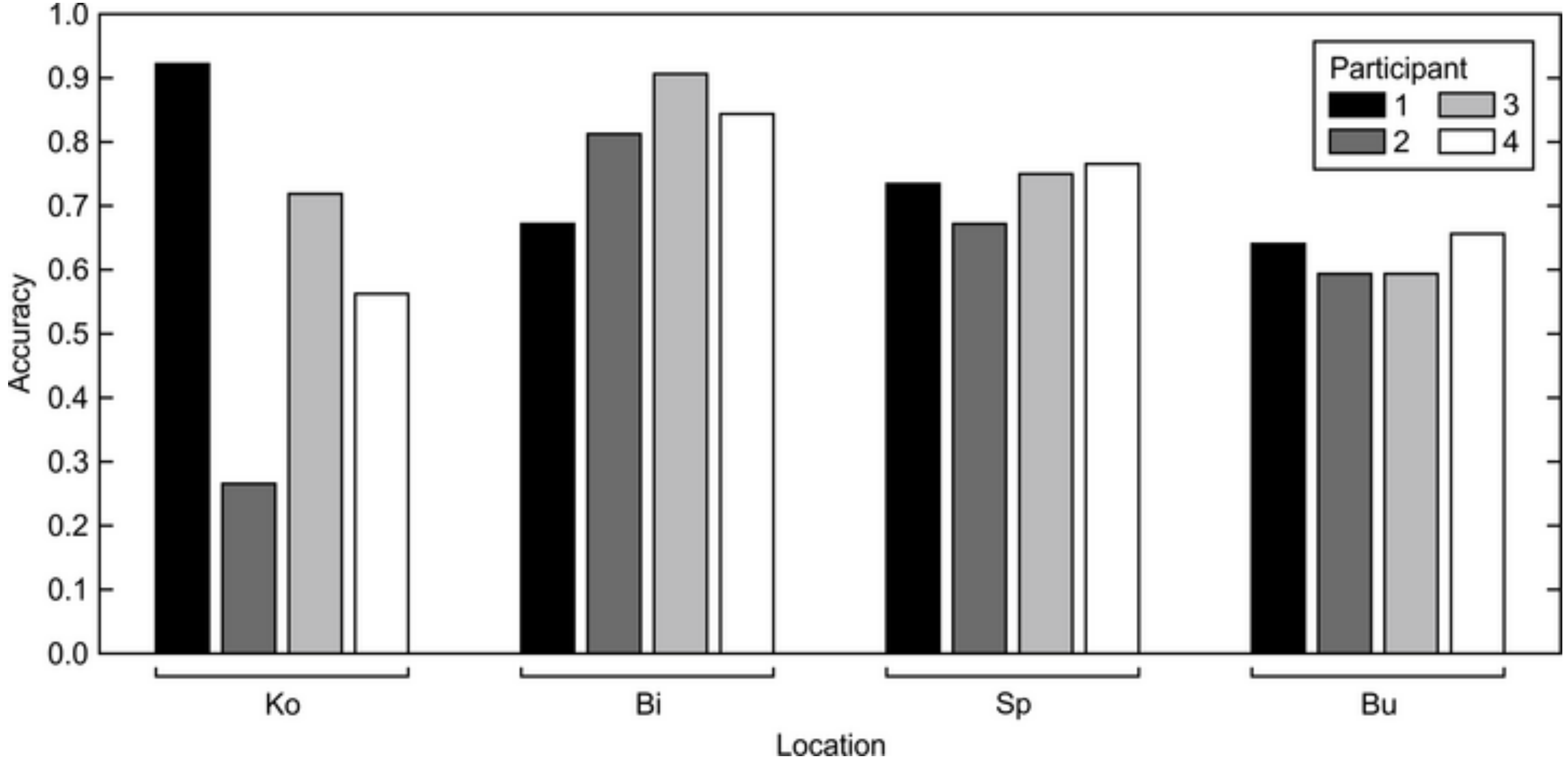


Figure A1

Click here to download high resolution image

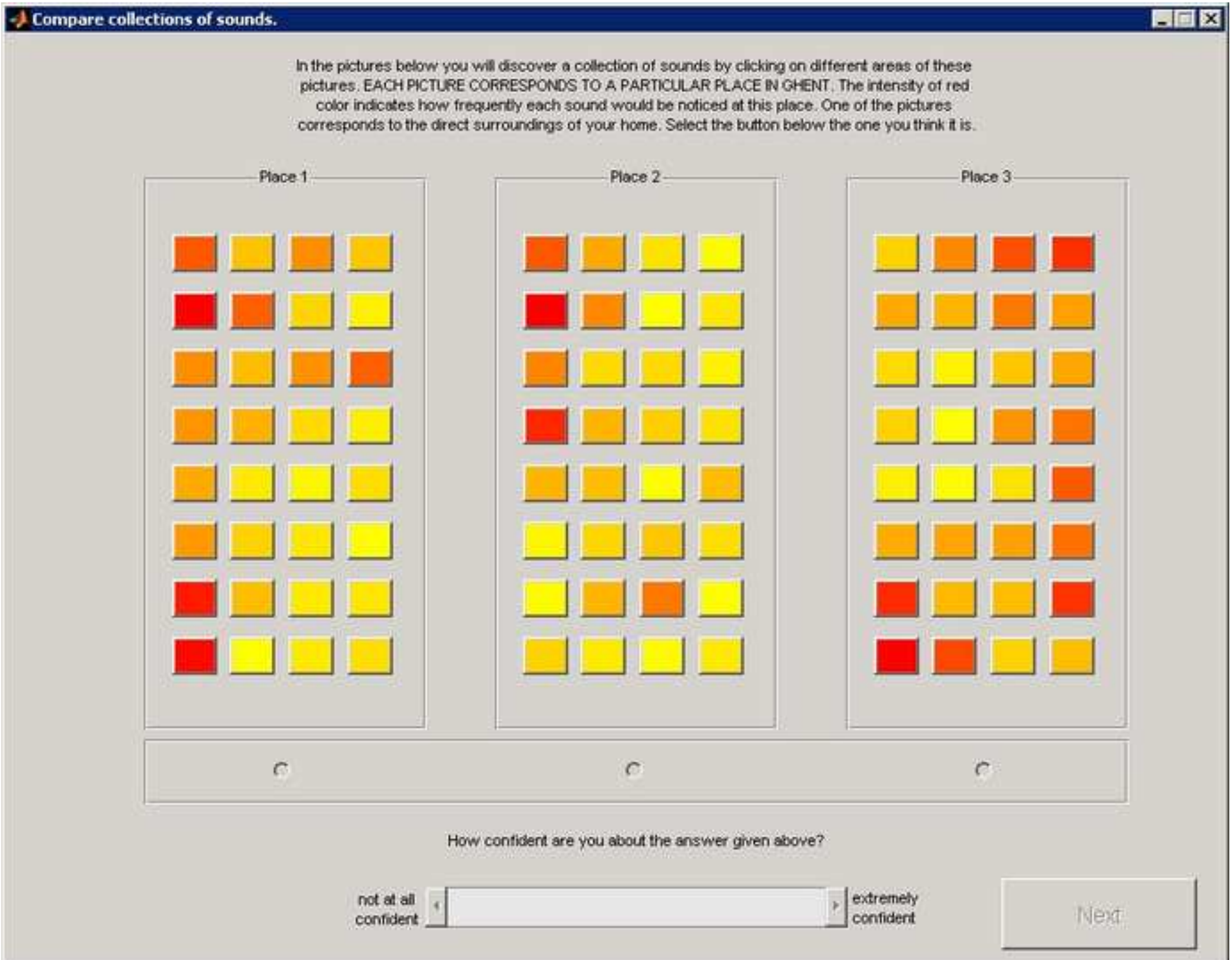


Figure B1

Click here to download high resolution image

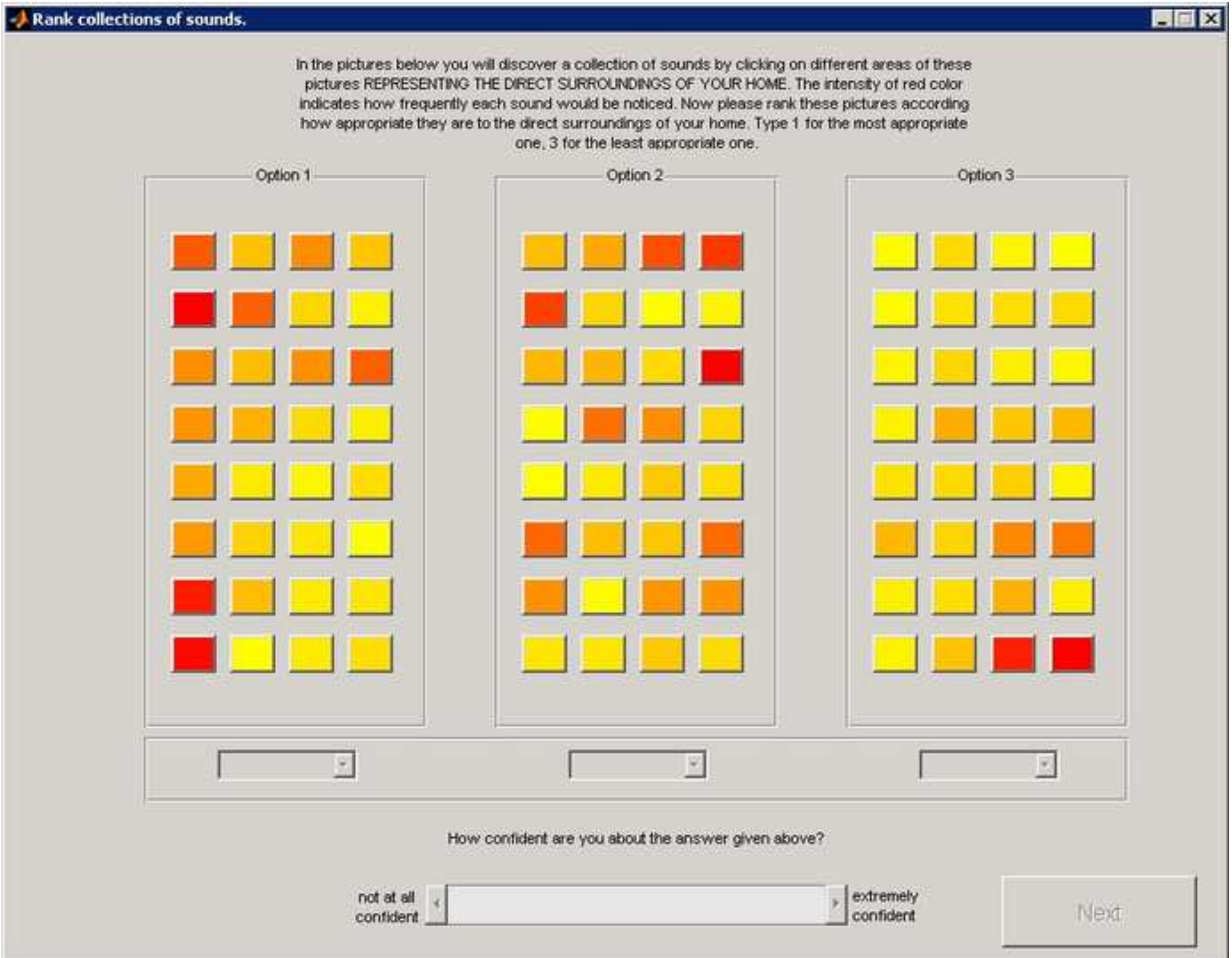


Figure C1
Click here to download high resolution image

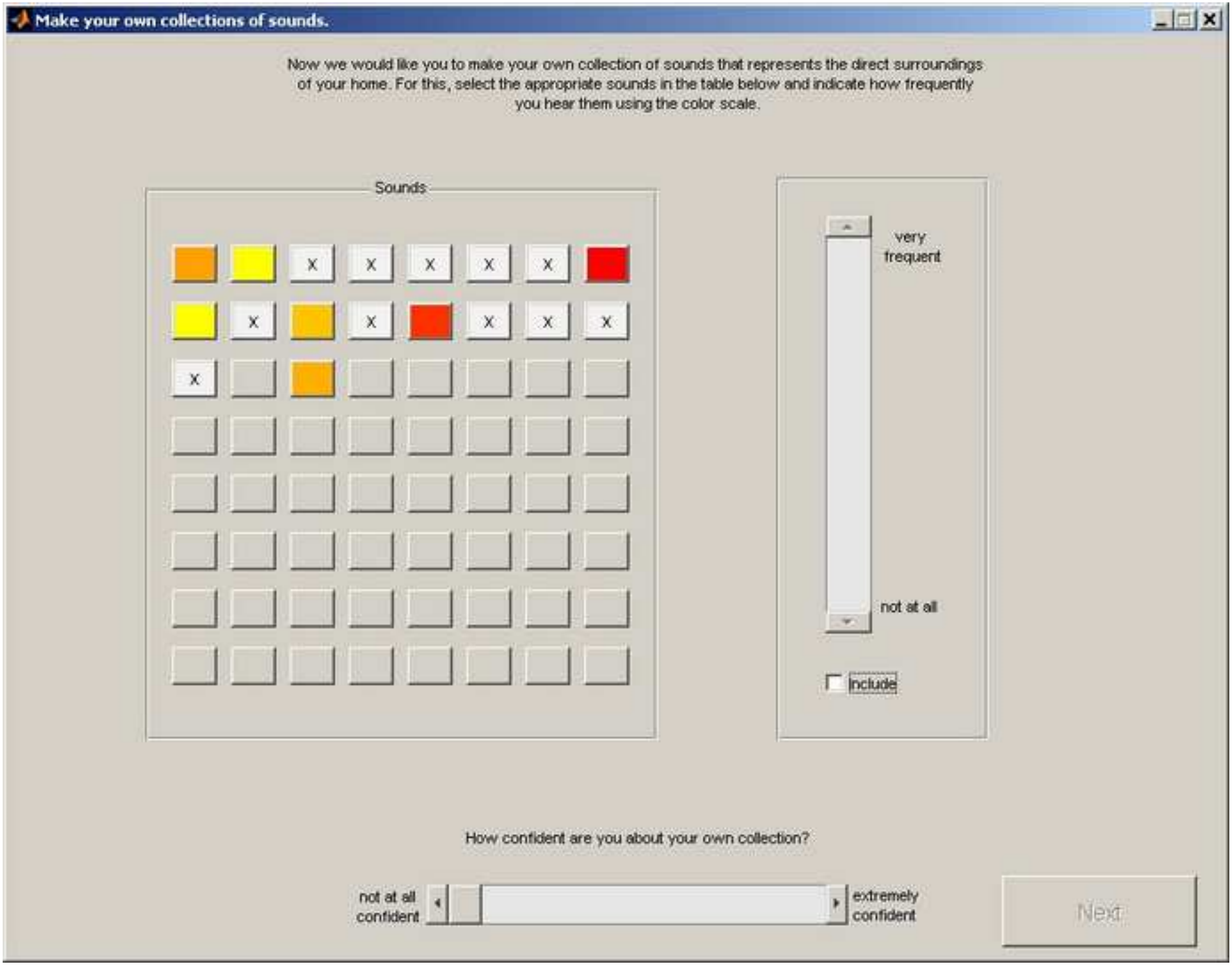


Figure D1

Click here to download high resolution image

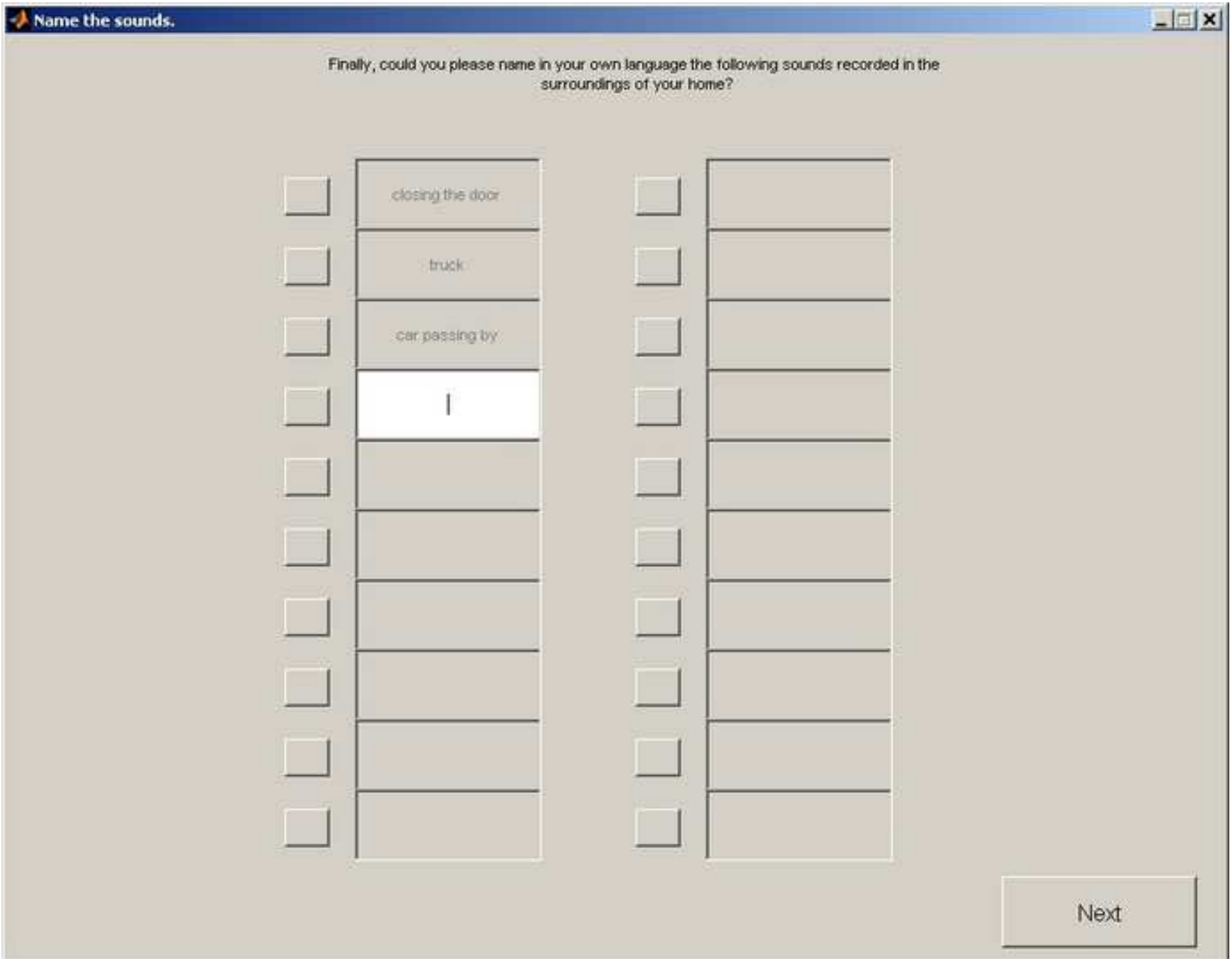
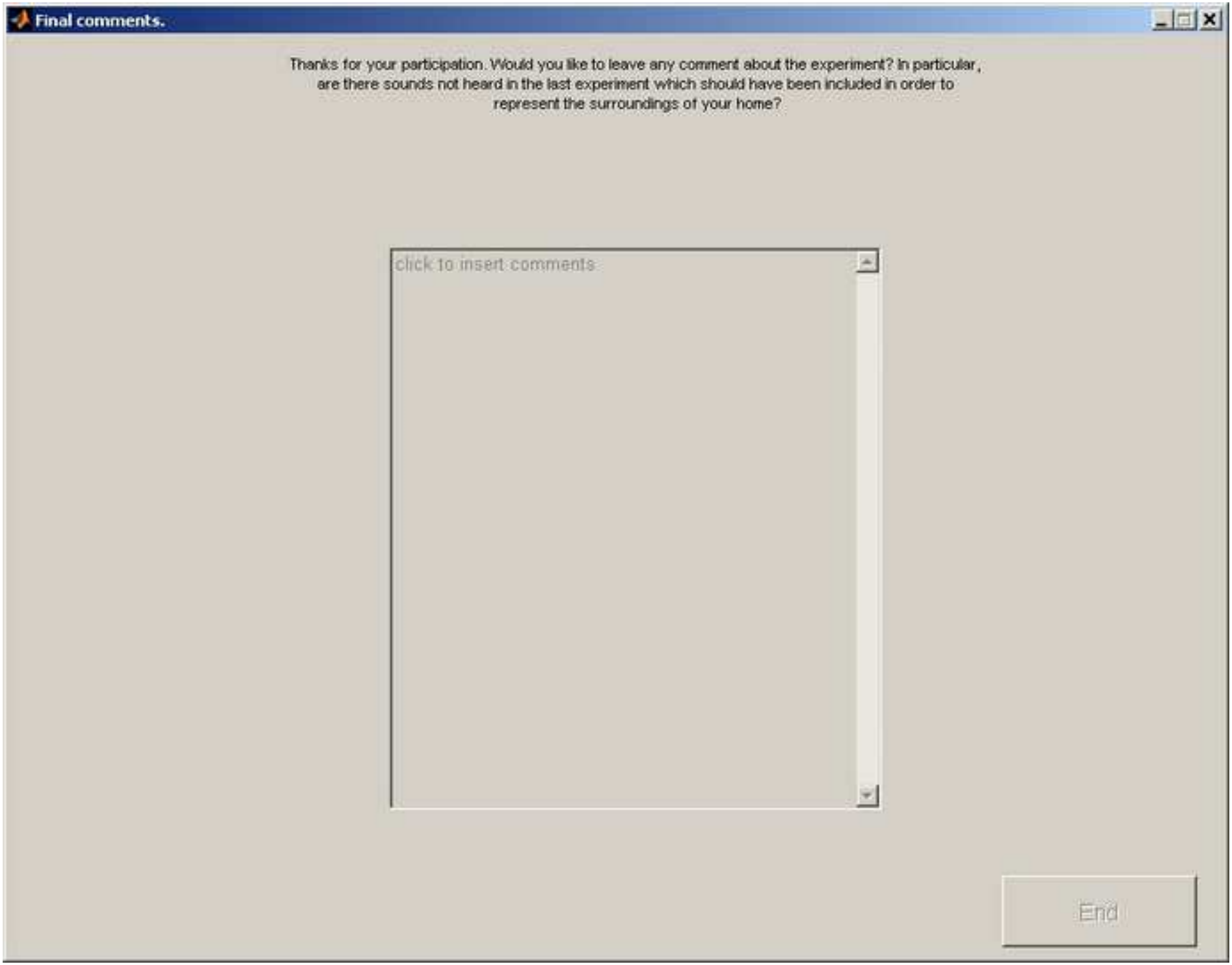


Figure D2
Click here to download high resolution image



Acknowledgements

Bert De Coensel and Annelies Bockstael are postdoctoral fellows, and Michiel Boes is a doctoral fellow of the Research Foundation-Flanders (FWO-Vlaanderen); the support of this organization is gratefully acknowledged. The authors would like to thank Samuel Dauwe for his technical contribution in collecting measurement data.