

Audio-Visual Interactions in Product Sound Design

Elif Özcan*

René van Egmond

Faculty of Industrial Design Engineering, Delft University of Technology, Landbergstraat 15,
2628 CE Delft, The Netherlands

ABSTRACT

Consistent product experience requires congruity between product properties such as visual appearance and sound. Therefore, for designing appropriate product sounds by manipulating their spectral-temporal structure, product sounds should preferably not be considered in isolation but as an integral part of the main product concept. Because visual aspects of a product are considered to dominate the communication of the desired product concept, sound is usually expected to fit the visual character of a product. We argue that this can be accomplished successfully only on basis of a thorough understanding of the impact of audio-visual interactions on product sounds. Two experimental studies are reviewed to show audio-visual interactions on both perceptual and cognitive levels influencing the way people encode, recall, and attribute meaning to product sounds. Implications for sound design are discussed defying the natural tendency of product designers to analyze the “sound problem” in isolation from the other product properties.

Keywords: Product sounds, audio-visual interactions, auditory memory, auditory identification, product design

1. INTRODUCTION

When designing sounds for products, a design team’s main tendency is to consider sound in isolation from other product properties focusing mainly on the (psycho)acoustic quality of the sound¹. However, when evaluating the designed sounds with users, it is necessary to consider the fittingness of the sound to the product². For example, when designing sounds for a sports car (see Figure 1), designers have to ensure that the engine sound is loud, dynamic, and exciting enough in order for users to be satisfied with the ‘sporty’ experience provided by the car as a whole. Consistent product experience requires congruity between product properties such as visual appearance and sound.³ Therefore, for designing appropriate product sounds, they should preferably not be considered in isolation but as an integral part of the main product concept. Because visual aspects of a product are considered to dominate the communication of the desired product concept, sound is usually expected to fit the visual character of a product.



Figure 1. A sports car by BMW.

*e.ozcan@tudelft.nl, +31152789678

We argue that simultaneous consideration of both sound and product appearance during a design process can be accomplished successfully only on the basis of a thorough understanding of the impact of audio-visual interactions on product sounds. For that, it is essential to point out at which stages of design audio-visual interactions might occur and which designerly activities are involved in the process. Main activities concerning sound design are the creative process during which the concept for a future sound is created, verbal communication of sound among designers, and the evaluation of the sound concepts⁴. All these activities rely heavily on perceptual and cognitive processing of sound. Therefore, audio-visual interactions will be studied from a psychological perspective.

1.1. Audio-Visual Interactions in Recognition and Identification

In our interaction with products, we often not only rely on one sensory modality but use the combined information of these senses in our experience and usage of these products. In addition, in product design a lot of verbal information is used to design a specific product experience. Users describe products in certain terminologies as do designers. It is therefore important to know how perceptual information from different perceptual systems are connected to each other and how these systems are influenced by the semantic system. Many studies investigated these interrelationships between systems and how they would affect memory performance.^{5, 6, 7, 8, 9} If semantic (i.e., verbal) information is congruent with perceptual (i.e., visual) information than this information has an additive effect on memory (i.e., dual-coding⁶). However, it has also been found that if people verbalize the semantic information obtained from a complex stimulus recognition may be reduced (i.e., verbal overshadowing⁸). It has been shown that a verbal overshadowing effect mostly occurs when perceptual expertise is high but lacks support from verbal expertise.¹⁰ Furthermore, it has been shown that conceptual training improves our verbal expertise, so that it can support the perceptual information. Studies have shown that with conceptual training, verbal expertise can be improved to support the perceptual information.

We always use products in a certain context. This may, for example, be our living room or kitchen. It is therefore important to investigate how context can influence our identification and with that our experience of products. It has been shown that object identification is not isolated from contextual scenes.^{11, 12, 13, 14, 15} Several stages can be discerned in the object identification. They all relate to the extraction and integration of featural aspects of objects. In general, three types of contextual effects can be distinguished that interact with the feature extraction and integration processes.¹⁶ First, a context will provide knowledge that interacts with the perceptual analysis of an object (i.e., feature extraction and integration).¹² The contextual activation enhances the most likely perceptual reconstruction of the object if it and its features are sufficiently analyzed.^{17, 18} Second, sometimes in the matching stage a rough perceptual analysis of the object in combination with context information is already sufficient to map it long-term memory representations.^{15, 18} Third, if both the object and the scene are equally identifiable object identification (including the matching stage) occurs independent of context information.^{16, 19, 20}

1.2. This Study

Product sounds are a sub-category of environmental sounds that are emitted by domestic appliances (e.g., shavers, alarm clock, kitchen extractors). These appliances have mechanical and electrical parts that produce sounds when functioning. In an earlier study,²¹ we distinguished six product sound categories on the bases of their perceptual similarities (i) air sounds (e.g., sounds caused by hairdryers, vacuum cleaners); (ii) alarm sounds (e.g., sounds mostly digitally produced by alarm clocks, microwave oven, and conventional ovens); (iii) cyclic sounds (e.g., sound caused by microwave ovens, washings machines); (iv) impact sounds (e.g., sounds caused by door opening-closing of washing machines or microwave ovens, buttons and switches); (v) liquid sounds (e.g., sounds caused by coffeemakers and water boilers); and (vi) mechanical sounds (e.g., sounds caused by shavers, toothbrushes, and hairclippers). These categories are also based on the mechanisms that cause the sounds. For example, air sounds are caused by moving air due to the rotating fans used to blow or suck air, alarm sounds are mostly digitally produced and designed especially to provide feedback and to warn listeners, cyclic sounds are caused by rotating parts which result in a cyclic event and a periodicity in the sound, impact sounds are caused by a short impact between product parts, liquid sounds are caused by moving or heating up liquid, and mechanical sounds are caused by engines at high rpm and small rotating, rubbing mechanical parts of products.

Product sounds are meaningful entities and are able to evoke a wide range of conceptual associations. Özcan and van Egmond²² determined 11 basic concepts by which product sounds can be described. These basic concepts are action (opening door, air blowing), emotion (obtrusive, pleasant), location (bathroom, kitchen), materials (metal, plastic), abstract meaning (sophisticated), onomatopoeia (brrr, buzz), psycho-acoustics (sharp, loud), sound type (air, digital), source (microwave oven, shaver), source properties (large), and temporal descriptions (repetitive, constant). The

distribution of these concepts over product sound categories differs. Some concepts are more representative for certain sound groups (e.g., action descriptions occur very frequently for impact sounds and location descriptions for air sounds).

Furthermore, consistent product experience requires that meaning attribution evoked by auditory and visual product properties is consistent. However, not much is known about how meaning attribution takes place for product sounds. Therefore, understanding the basics of auditory memory and of sound identification with regard to product sounds is vital. Such cognitive functions underlie not only designer's creative processes when designing product sounds, but also the processes users undergo when they experience a product and, in particular, a product sound.

Accordingly, this paper reviews two experimental studies^{23, 24} that tackled audio-visual interactions on both perceptual and cognitive levels influencing the way people learn product sounds and attribute meaning to them. The results of these studies are discussed with regard to their implications for sound design defying the natural tendency of product designers to analyze the "sound problem" in isolation from the other product properties.

2. MEMORY FOR PRODUCT SOUNDS

Memory is important in our interaction with every day objects. People with memory failure have often difficulty in functioning properly. But even for people with a good memory performance it is important that objects that are perceived can be stored in memory. Storage in memory will depend on how well we can code featural aspects in the perceptual domains (e.g., auditory and visual). If sounds or images are difficult to encode then this may result in a decay of performance in memory. In Chapter 1.2 it has been described that product sounds are inherently ambiguous because of a lack of structure in the temporal-spectral domain. In addition, it was discussed that this lack of structure differed between the six product sounds categories (see also, Özcan, et al.²¹). It was suggested that the mechanisms in products (e.g., domestic appliances) producing sounds will yield in a high level of randomness in the sounds (noise). As a result, only part of the sound can be encoded. In addition, the operating mechanisms between products are similar and therefore sounds will be produced that are also judged very similar.²¹ For example, shavers and toothbrushes both contain engines and gears that produce a similar sound. These two aspects, that is, the ambiguity of product sounds caused by noise and the similarity between product sounds resulting from the operating mechanisms make product sounds difficult to store uniquely in memory.

During conceptual design or other creative activities, designers are often forced to reproduce the sound related information from memory by retrieving the previously stored auditory information which can be on a perceptual, conceptual, or semantic level. For example, if the sound of a vacuum cleaner is in question, the designers may be expected to retrieve the spectral-temporal composition of the sound, the appearance of the product causing the sound and perhaps the product's name, and other conceptual associations that are somehow linked to vacuum cleaner sounds in memory (e.g., similar products, the location, etc.). Therefore, we investigated if labels at encoding improve the memory of product sounds (see Özcan & van Egmond,²³ for more details).

In this study, by using different types of labels and different memory tasks it was tested which labels were most effective in enhancing memory performance. Three different memory tasks have been used: a recall, recognition, and matching task. In this review, only the first two tasks are relevant for our discussion. The recall and recognition tasks relate to different memory processes. In a recall task people just have to reproduce by heart what they just have learned. Therefore, the recall task does not require the processing of new information and is only dependent on an internal search in memory. A recognition task, in which people are asked if they heard this sound before, requires the processing of a presented stimulus (i.e., sound) and involves a comparison with prior encoded information. In addition, recognition will depend on how well the previous information was coded and the present information can be coded. Thus, both memory tasks may reveal different aspects that could be informative to designers in different stages of the design process. Three different label types have been employed: self-generated text; provided text; and an image. In Figure 2 an example of micro-wave oven is depicted for the three label types. The left picture is an example of a self-generated text, the middle image of a provided text and the right image a picture of a microwave oven. People (in three different groups: each label condition was employed for a different group) were asked to remember the sound label combination.

Three main effects have been found. First, the amount of structural features differed for each product sound category. The memory performance of sounds in a category with very few structural properties (e.g., impact sounds) was lower than the memory performance for sounds with relatively high structural properties (e.g., alarm and mechanical sounds). Second, the self-generated labels to a sound result in higher memory performance than the other labels. Third, the effect

of visual information on memory performance was different for the recall task and the recognition task. The memory performance for the recall task increased when visual information was presented at the time of encoding. This effect is known as the dual-coding effect (c.f., Paivio⁶). The visual information caused the memory performance in the recognition task to decrease. It has been suggested that this is caused by the complexity of the visual information that yielded in a decay of attention for the auditory domain in favor of the visual domain. This effect has been coined as visual-overshadowing (c.f., Schooler & Engstler-Schooler⁸). In conclusion, we have shown that memory performance of product sounds can be improved by labeling. However, the lack of structural information in these sounds makes it still difficult to encode these.

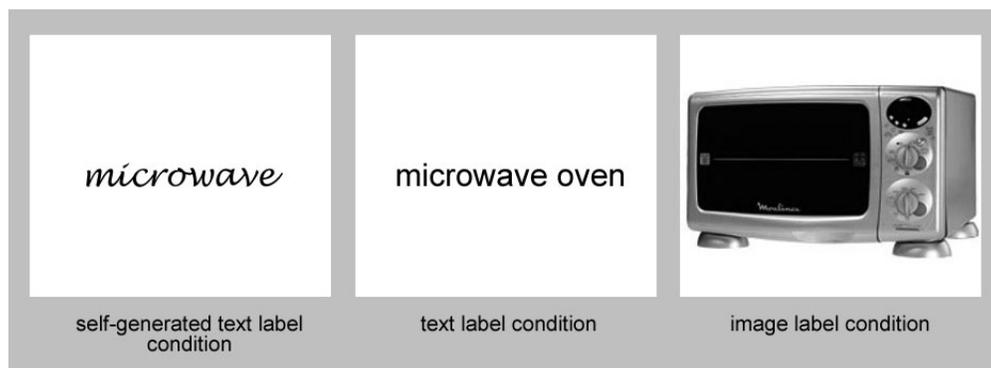


Figure 2. Examples of labels presented as self-generated text, provided text, and images.

3. IDENTIFICATION OF PRODUCT SOUNDS WITHIN CONTEXT

Everyday objects are presented to us in relation to other objects and scenes. Even when partially seeing an object, people are able to identify it correctly because the scene in which the object is presented and other objects in that scene confirm the plausibility of the result of identification. Products (e.g., hairdryers, mixers) are part of daily scenes and very often define the meaning of the scenes. For example, a bathroom scene consists of a shower, a sink, a towel, a shaver, and a toothbrush. More often than not, products are designed to fit a particular scene or particular objects within that scene. For example, if the bathroom scene imposes ‘hygiene, personal care, and privacy’, then products should be designed to carry some of the characteristics of the scene.

With regard to product sounds, there is not much knowledge of how people attribute meaning to product sounds within particular scenes. However, people do relate product sounds to specific locations.²² For example, the sound of a coffeemaker belongs to a kitchen scene. Previous research has also shown that product sounds are ambiguous and therefore difficult to correctly identify.²⁵ For example, in the absence of context (any particular scene or related objects), a continuous droning air sound can be identified in several different ways; such a sound can be a hairdryer, a dust-buster, or a mixer. However, our daily interactions with such sounds can dictate that we are able to differentiate the sound of a vacuum cleaner from the sound of a hairdryer if such sounds are presented to us in proper contexts.

Designers, when designing product sounds, very often discard the contextual information and focus only on the sound and the capacity of the sound to convey certain concepts such as sportiness and elegance (as in the example of the BMW sports car). One reason for such focus on sound is the field of product sound design lacks empirical evidence that states that there is a strong link between the sound and the scene in which the product is used. Therefore, we have investigated in another study the effect of context in the identification of product sounds (see Özcan & van Egmond,²⁴ for details).

In this study, using a contextual priming paradigm, we tested listeners’ identification responses to product sounds within a congruent visual context (scene vs. object). We also tested these sounds within a control condition in which no contextual cues were provided. As shown in Figure 3, we used two types of images as visual context: scenes and single objects. For example, if participants were asked to identify a citrus-press sound, then the image of a kitchen was presented to the participants of ‘scene context’ condition and the image of half an orange was presented to the participants of ‘object context’ condition. The participants of the control condition only received an abstract image that contained a meaningless pattern during sound identification. By providing different types of contexts, we aimed to see whether sound identification would improve as a function of context specificity. That is, we hypothesized that object

context would provide more specific and direct information for correct sound identification and scene context would provide less specific and more vague information. The sounds used represented the six categories of product sounds which were described in Paragraph 1.2.

The results show that the congruent visual context significantly improves the identification accuracy of product sounds. As hypothesized, the specificity of the context (object and scene) also significantly improves the identification accuracy. That is, the identification accuracy for the sounds presented in object context are higher than the identification accuracy for the sounds presented in the scene context. Furthermore, a significant interaction effect was found between sound type and context type. Product sounds that are location-specific (e.g., air, cyclic, and alarm sounds) are identified equally good in scene context and object context conditions and product sounds that are event-specific (impact, liquid, and mechanical sounds) are identified better in object context condition.



Figure 3. Examples of images presented as control condition, scene, and object context.

We suggest that a scene or an object is able to activate a frame of conceptual associations and can build a semantic link between an ambiguous sound and its source/label. Furthermore, we also suggest that objects, which co-exist with the product causing the sound, have a stronger impact on correct sound identification. However, it is interesting to see that products that often stay in the background (e.g., washing machines, kitchen extractors, even warning sounds) are actually part of the scene in which they are presented. Therefore, it is easier to access their lexical representation via the activation by the scene information. Similarly, products that are salient in certain daily events (e.g., making orange juice, closing microwave doors, preparing coffee) are part of a group of objects that are conceptually related by a certain event. Therefore, it is easier to access their lexical representation (i.e., citrus-press, door closing of a microwave oven, and coffee-maker) via objects. Furthermore, the audio-visual interaction studied in the context experiments are very likely to occur due to a semantic interaction between the representations activated by sounds and by visual context. However, the location-specific sounds also benefited from the object context. This might provide some evidence that the audio-visual interaction may occur on a perceptual level. It is plausible that the visual scene of a location may have activated all the objects and sounds that belong to that scene. The activation of auditory information may have guided the perception of the sound that was being heard, resulting in a correct identification of the sound which is congruent with a scene. In conclusion, a context can be a scene or an object as long as the context frame builds a strong link between the ambiguous sounds and their plausible representations and the strong conceptual activation provided by the scene or object.

4. DISCUSSION

Figure 4 summarizes the audio-visual interactions that occur during the identification of a product sound. The figure suggests that an ambiguous product sound, regardless of the context in which it is presented, will activate memory representations that are directly related either to the sound. These memory representations may be too general in the beginning and get specified as more and more conceptual representations are activated. Finally, the activated semantic network for the sound heard may refer only to the quality of the sound, locations in which the sound is heard, and also the plausible labels of products that may cause the sound. Context frame can influence (i) the activation of the auditory representations by guiding the actual perceptual process, (ii) confirmation of the plausibility of the product concept, and finally (iii) the lexical activation.

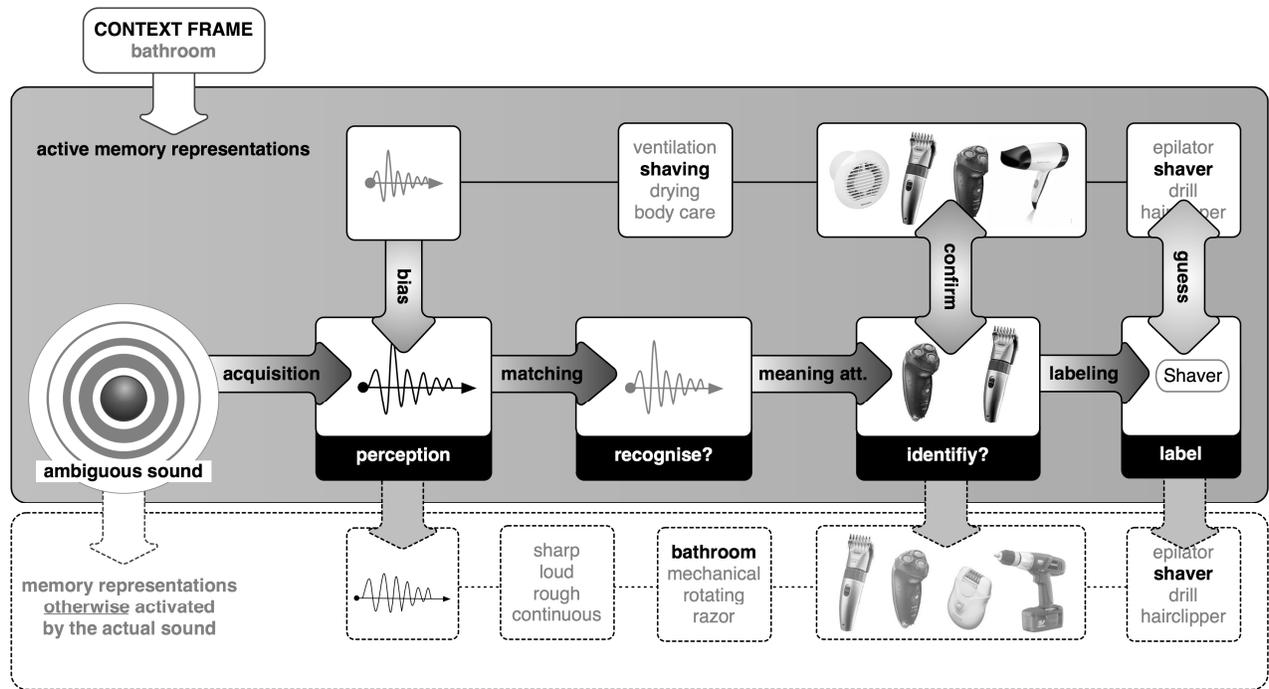


Figure 4. Audio-visual interactions during sound identification.

The studies that are reviewed in this paper have provided enough evidence that sounds are an integral property of the product. That is, the product's inherent relationship to its label, to its location where it is used and to product-specific objects determine how well a product sound can be identified. Furthermore, audio-visual interactions are shown to be both on a perceptual level and on a cognitive level. On a perceptual level, visual information interferes (diminishes or guides) the amount of auditory information that can be coded and stored. This finding is not necessarily a negative finding. The ambiguity in storing auditory information can also be understood as an advantage for sound designers. One of the aims of designing sounds is making the sound in question identifiable and fitting the function of the product. Thus, ambiguity naturally presupposes that as long as the identified sound belongs to a plausible category, designers will feel safe to manipulate sound within that category. That is, if a shaver sound is being designed, the designed sound should fit in the category of mechanical sounds. Furthermore, many of the audio-visual interactions are observed on a semantic, thus cognitive, level. This also suggests that the sound and the product both activate a specific conceptual network. Thus, there needs to be an overlap between these two different types of conceptual activations. Then, designers can guarantee the fittingness of the sound to the product.

5. IMPLICATIONS FOR PRODUCT SOUND DESIGN

A product has indispensable properties that lead to visual, auditory, and tactile experiences. This study has shown that two seemingly different product properties interact both on a perceptual and cognitive level which may alter users' responses towards products. To extrapolate the findings, such interactions may influence users' reasoning, emotional state, purchase decisions, preference, and expectations regarding product and product performance. Therefore, designers should always consider product sounds in relation to the product and the context of use. In the current practice of product sounds, designers are more concerned with the acoustic quality and the pleasantness of the sound and less with the experiential quality. Experiential quality of the product sounds may play an important role in product satisfaction and product attachment because sound as a product property emerges, more often than not, during product usage but not product purchase.

Our advice to product designers is to consider the fittingness of a sound to the product at all design stages. Product sounds are meaningful entities that are able to evoke a wide range of conceptual associations. Therefore, the sound activated associations should match the product image activated associations. Because the visual product property has a

strong influence on the sound, sound should fit the concept of the product. Consequently, sound quality evaluations should be conducted with the presence of the product or at least its image.

On the other hand, the acoustic quality of the sound is of importance for auditory aesthetics. It is the spectral-temporal composition of the product sound that gives rise to aesthetic judgments which further influence the cognitive judgments and preferences. Therefore, our advice to product designers is to work on the acoustic quality in the absence of the product or its image. Because, visual information negatively influences the perceptual processing of the sound which results in badly encoded auditory representations. However, a sound designer should be able to pay full attention on the perceptual details of the sound.

Overall, this study has shown that sound is an integral property of a product and for consistent product experience sound should be designed according to product image. However, we find it interesting to follow up this study with haptic or tactile information and see how that influences with product sounds. Another study will also investigate, how the experience of abstract concepts (e.g., a *sporty* car or an *elegant* hairdryer) is influenced by audio-visual interactions.

REFERENCES

- [1] Lyon, R. H. (2000). *Designing for Product Sound Quality*. New York: Marcel Dekker, Inc.
- [2] Blauert, J. and Jekosch, U. (1997). Sound Quality Evaluation – A multi-layered problem. *Acustica*, 83(5), 747-753.
- [3] Ludden, G. (2008). *Sensory incongruity and surprise in product design*. Doctoral dissertation, Delft University of Technology.
- [4] Özcan, E. and Van Egmond, R. (2006). *Product Sound Design and Application: An Overview*. In P.M.A. Desmet, M.A. Karlsson, and J. van Erp (Eds.), *Proceedings of the 5th International Conference on Design and Emotion*. Gothenburg: Chalmers University of Technology.
- [5] Paivio, A., Philipchalk, R. and Rowe, E. J. (1975). Free and Serial Recall of Pictures, Sounds, and Words. *Memory & Cognition* 3(6): 586-590.
- [6] Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology - Revue Canadienne De Psychologie*, 45(3), 255-287.
- [7] Thompson, V. A. and Paivio, A. (1994). Memory for pictures and sounds: independence of auditory and visual codes. *Canadian Journal of Experimental Psychology* 48(3): 380-398.
- [8] Schooler, J. W. and Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: some things are better left unsaid. *Cognitive Psychology* 22(1): 36-71.
- [9] Melcher, J. M. and Schooler, J. W. (1996). The misremembrance of wines past: Verbal and perceptual expertise differentially mediate verbal overshadowing of taste memory. *Journal of Memory and Language* 35(2): 231-245.
- [10] Melcher, J. M. and Schooler, J. W. (2004). Perceptual and conceptual training mediate the verbal overshadowing effect in an unfamiliar domain. *Memory & Cognition* 32(4): 618-631.
- [11] Bar, M. (2004). Visual objects in context. *Nature Reviews: Neuroscience*, 5(8), 617-629.
- [12] Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 213-263). New York: Lawrence Erlbaum.
- [13] Chaigneau, S. E., Barsalou, L. W., & Zamani, M. (2009). Situational information contributes to object categorization and inference. *Acta Psychologica*, 130, 81-94.

- [14] Heit, E., & Barsalou, L. W. (1996). The instantiation principle in natural categories. *Memory*, 4(4), 413-451.
- [15] Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, 3(5), 519-526.
- [16] Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, 50, 243-271.
- [17] Bar, M., & Aminoff, E. (2003). Cortical analysis of visual context. *Neuron*, 38(2), 347-358.
- [18] Bar, M., & Ullman, S. (1996). Spatial context in recognition. *Perception*, 25(3), 343-352.
- [19] Hollingworth, A., & Henderson, J. M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General*, 127(4), 398-415.
- [20] De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research*, 52(4), 317-329.
- [21] Özcan, E., Van Egmond, R., Jacobs, J., and Ballas, J. A. (2009). Bases for Categorization and Identification of Product Sounds. In review for JEP: Applied.
- [22] Özcan, E. and Van Egmond, R. (2005). *Characterizing Descriptions of Product Sounds*. In D Brock (Ed.), Proceedings of the 11th International Conference on Auditory Display. Limerick: ICAD.
- [23] Özcan, E. and Van Egmond, R. (2007). Memory for Product Sounds: The Effect of Sound and Label Type. *Acta Psychologica*, 126, 196-215.
- [24] Özcan, E. and Van Egmond, R. (2009). The Effect of Visual Context on the Identification of Ambiguous Environmental Sounds. *Acta Psychologica*, 131, 110-119.
- [25] Özcan, E. and Van Egmond, R. (2007). *How well do we identify product sounds?* Proceedings of the 13th International Conference on Auditory Display. Montreal: ICAD.& van Egmond, 2007 (icad)