EMOTIONAL AND BEHAVIORAL RESPONSE TO AUDITORY ICONS AND EARCONS IN DRIVER-VEHICLE INTERFACES

Pontus Larsson
Anders Opperud
Krister Fredriksson
Humans Systems and Structures, Volvo Technology Corporation
Daniel Västfjäll
Applied Acoustics, Chalmers University of Technology
Sweden
Paper Number 09-0104

ABSTRACT

Adequately designed, auditory displays in Driver-Vehicle Interfaces (DVIs) may give shorter reaction times, improved attention direction, and an increased quality impression. In this paper, we argue that emotional reactions may guide the design of such auditory displays since emotion is central in our everyday life and have strong consequences for behavior and information processing. A simulator study with 30 participants (20 of which were professional drivers) was conducted to investigate the connection between emotional and behavioral responses to auditory DVIs as well as to evaluate various sound design parameters in realistic driving situations. Auditory icons were contrasted to abstract earcon sounds in more or less imminent collision scenarios and 3D sounds were tested against monophonic sounds in different lane change scenarios. Self-report measures (Self-Assessment Manikins, SAM) and physiological measures (Galvanic Skin Response, GSR and facial Electromyogram, EMG) of emotional response as well as behavioral measures (e.g. brake response time) were used.
It was found that auditory icons were more efficient and gave up to 600ms faster brake response times than abstract sounds in imminent collision scenarios and that 3D sound gave a stronger emotional response in lane change scenarios. Moreover, the results show that emotion can predict behavior, e.g. sounds rated as being more activating and negative also gave quicker response times. Contrary to our expectations however, the findings from the SAM ratings were not reflected in the physiological measurements. An explanation to this may be that the scenario itself caused a dominant stress reaction which overrode the physiological response to the warning sounds. Our findings nonetheless strengthen the importance of auditory displays as a means to enhance vehicle safety, and that emotions may be an efficient way of predicting behavioral response to auditory DVIs. Measurements of emotion may therefore facilitate the process of designing auditory DVIs.

BACKGROUND

Sound may be a very efficient mean of providing warnings and information in vehicles, especially in situations where the visual modality is loaded with information, but can also be used to increase the quality impression of the vehicle.
A central characteristic of most auditory Driver-Vehicle Interfaces (DVIs) is that it should convey the appropriate level of urgency. Urgency can be defined as "…an indication from the sound itself as to how rapidly one should react to it." (Edworthy & Hellier, 2006). Too urgent sounds may cause annoyance, unwanted startle effects and even lead to the wrong behavior. On the other hand, if the sound is not urgent enough, reaction may be unnecessarily slow or result in that the warning is neglected.
Parameters which have been found to influence the perceived urgency of a sound include repetition speed, number of repeating units, fundamental frequency, and inharmonicity (Hellier & Edworthy, 1996). Moreover, loudness appears to be one of the stronger cues for urgency (Haas and Casali, 1995). However, the range within which loudness can be varied before the sound becomes un-ergonomic is in practice rather small; the sound should of course be loud enough to be heard over the background noise in the operator's environment and quiet enough not to cause annoyance or hearing impairment. This matter may seem trivial, but is often a central issue which is crucial for the acceptance of the sound.
Although thorough research has been conducted on the correlation between basic psychoacoustic parameters and urgency and similar perceptual aspects of sound design, the cognitive response linked to the sound is much less well understood (Edworthy & Hellier, 2006). It is therefore important to systematically investigate and be able to measure how the sound is comprehended. Edworthy and Hellier (2006) suggest that abstract sounds can be interpreted very differently depending on the many possible meanings that can be linked to a sound, in large dependent on the

surrounding environment and the listener. Designing sounds with unambiguous and appropriate meaning is perhaps the most important task in auditory warning design (Edworthy & Hellier, 2006).
A possible solution to the meaning problem is to use auditory icons. Auditory icons are representational, real world sound events that are used to signal events in Human-Machine Interfaces (HMIs). The advantage of auditory icons is that they have inherent meaning and therefore require no or little learning. Still, this meaning may not the same to all persons. As an example, the sound of a drain clearing may to some mean "Wet road" while others may interpret it the way the designer intended, namely "Low on fuel" (Winters, 1998). In general it may be difficult to find a suitable match between the function/event to be represented and the sound.

Emotional reactions

While perceptual and cognitive aspects of sound design such as urgency and meaning of sounds are important to consider in the development of efficient DVIs, we propose here to take one step back and instead look at the emotional response to sound.
Why would we consider emotion? From an evolutionary perspective emotion can be seen as the human alarm system. Positive emotions signal that everything is safe and no specific action is needed to be undertaken to survive, while negative emotions signal a potential threat and need to take quick action. Emotions thus have strong consequences for behavior and information processing.
In our everyday life, sound often elicits emotional reactions in the listener. People can be startled by the sudden sound of a door slamming or a thunder in a storm, annoyed by the noise of cars in the street, pleased by the sound of a water stream in the forest, tired after a full day of work in a noisy environment, etc. Thus, understanding the role of sound in evoking human emotional responses might improve our quality of life by helping to design objects, spaces and media applications which are emotionally optimized (Tajadura & Västfjäll, 2008). Following this, it may be argued that designing sounds that elicit an emotion is also a way of designing sounds that will elicit a reaction and may therefore be particularly suited for design of sounds for DVIs. Another advantage is that rather than focusing solely on behavioral responses (which often are difficult and time-consuming to assess) as a measure of performance, emotion psychology has a rich flora of instruments to measure emotion that may be used as a proxy measure of behavior.
In our previous research, we have devised the Emotion Reaction Model (ERM) framework which builds on neuropsychological research showing that the human brain automatically reacts to certain sound properties, either in a very fundamental way (approach/avoidance reactions linked to survival) or by activating associative networks (primes/memories from previous exposures to situations where the sound was experienced) (see Västfjäll, 2007; Västfjäll et al., 2007). Most importantly, the ERM framework suggests that emotional or affective reactions are the driving force of behavior or action (Damasio, 1994, LeDoux, 2000). Therefore, for a warning or info sound to elicit the "correct" action, it needs to induce an emotional reaction (Västfjäll et al., 2003). A central question within this research is thus what in a sound that induces an emotional reaction.

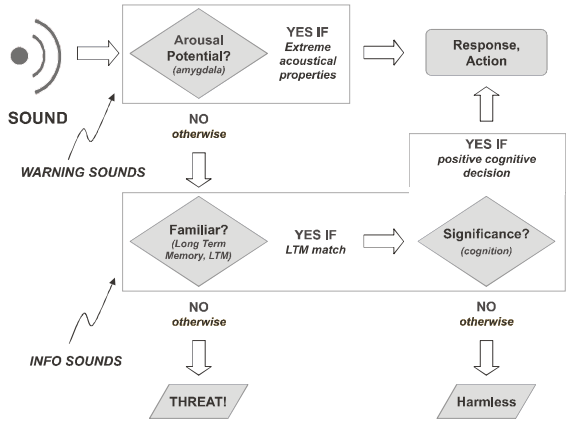

Figure 1. Schematic of the ERM framework.

Figure 1 shows a simplified schematic of the proposed framework where sounds (upper left corner) can be processed along two principal routes; either it will elicit immediate responses (arousal potential) or sound properties will be compared to previously stored memories of sounds (familiarity/significance check). In both cases the sound will elicit both a reaction and an evaluation of the possible danger or threat of the situation (as signaled by the sounds).
The upper route in Figure 1, arousal potential, is a preattentive warning system that has evolved to make a quick assessment of the potential threat of the situation (Jacobsen et al., 2003; LeDoux, 2000; Schröger, 1997; Tiitinen et al., 1994). This system needs to be extremely fast and elicit immediate and correct responses to ensure survival. Evidence points to that humans are hardwired to react to certain extreme acoustical characteristics (loud, sharp noises with quick rise time) (Bradley & Lang, 2000). Thus, when a sound exceeds a certain arousal threshold it will activate fight or flight tendencies (Giard et al., 1995; Lang, 1995; Rauscheker, 1998).

If the sound does not have an arousal potential that exceeds the threshold, it will be processed by other parts of the brain (Belin & Zatorre, 2000; Peretz & Zatorre, 2005); this is visualized by the lower route in Figure 1. First, the secondary auditory cortex will process the sound followed by associative and motor cortex areas (Edeline & Weinberger, 1992). Here the incoming sound will be compared to sound representations stored in long-term memory (Saarinen et al., 1992). If the sound has been encountered before (or resembles a sound that has been encountered before) it will elicit on of two possible reactions (Jääskelinen et al., 2004): If it is an unfamiliar sound, it will immediately signal the same alarm system as a sound with high arousal potential. The same holds true if the sound matches a sound representation that is associated with a previous negative experience (Damasio, 1994; Damasio et al., 2000). If the sound matches a previously stored representation it will be evaluated on basis of it significance for survival (Blood & Zatorre, 2001; Todd, 2001). If our previous experience has coded the sound as something potentially threatening or coexisting with something else that may be a threat, the system will call for an action. If, on the other hand, the sound is evaluated as non-harmful, no action will be required.

Implications for sound design - This simplified framework has several implications for design of warning and information sounds. First, it postulates that warning sounds should have a certain degree of arousal potential so that it evokes an immediate correct response. The main task for future research here is to map the arousal potential of various sounds and create sounds that have just the right amount of potential as too much may result in freezing behavior and incorrect responses (Panksepp & Bernatsky, 2002).
Second it suggests that both information and warning sounds would benefit from having sound elements that are familiar (such as auditory icons that rely on naturally occurring sounds to convey information). Third, the proposed framework suggest that emotional reactions to sounds is a common currency with which the urgency, behavioral significance and needed action will be evaluated against (Damasio et al., 2000). This also suggests that when assessing the effectiveness of warning and information sounds, affective reactions should be measured along with process measures (reaction time and decision) of the action.

The affect circumplex – It has been shown that the emotional reaction to natural and product sounds can be efficiently be described by two bipolar dimensions, activation and pleasantness-unpleasantness (valence), (Bisping, 1995, 1997; Bradley & Lang, 2000; Västfjäll et al., 2003). Taking this approach, it is assumed that any emotional state can be described by the *combination* of these two orthogonal dimensions (Russell, 1980; Russell & Feldman-Barrett, 1999). The so-called affect circumplex (Russell, 1980), shown in Figure 2, visualizes the two dimensional approach to describing emotional reactions. As an example, an emotional state such as excitement is the combination of pleasantness and high activation (upper left quadrant in Figure 2). An emotional reaction such as calmness is similar in pleasantness, but low in activation (lower left quadrant). Boredom is the combination of unpleasantness and low activation (lower right quadrant) and distress is the combination of unpleasantness and high activation (upper right quadrant).

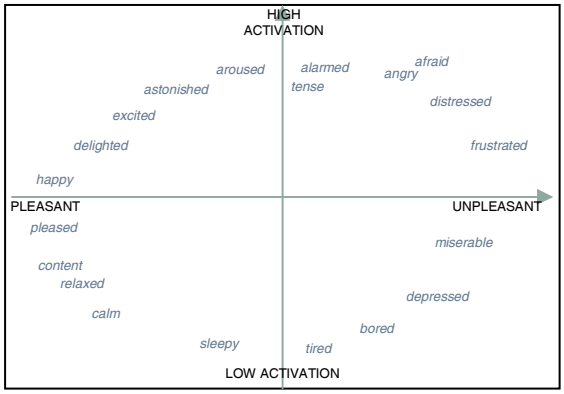

Figure 2. The affect circumplex (Adapted from Russell, 1980).

Measurement of emotional reactions to sound

There is a number of different ways to measure emotional reactions, including self report, physiological measures such as EEG and behavioral measures.
Self-report measures rely on that participants accurately can report their felt emotion. The main self-report measure used in the ERM framework is the Self Assessment Manikin (SAM) scales (Bradley & Lang, 1994, see Figure 3) which aims at capturing the activation/valence dimensions described in the previous section. The advantage of the SAM measure is that it can be understood by different populations in different cultures and that it is easy to administrate.
Many different physiological processes indicate emotional experiences. For instance, video recordings of the face can obtain measures of facially expressed emotions (Sebe et al., 2002). However, emotional reactions can also be captured via physiological measures of activation and valence. The method preferred within the ERM framework to measure valence are Electromyographical (EMG) measures of facial muscle contractions (Bradley & Lang 2000).

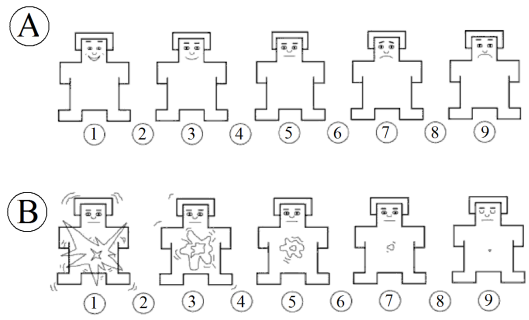

Figure 3. The Self Assessment Manikin (SAM) scales for valence (top) and activation (bottom) (Bradley & Lang, 1994).

This is typically measured by attaching electrodes in the facial region and measuring muscle micro-movements. Activity in the *Corrugator supercilii* (which controls eyebrow contraction) can be linked to unpleasant emotions (negative valence) whereas activity in the *Zygomaticus major* (the "smile muscle") may be linked to pleasant emotions (positive valence). The activation dimension is preferably measured physiologically using EDA (Electro- Dermal Activity) which can be obtained by measuring the galvanic skin resistance on subjects' fingers or palms (Bradley & Lang 2000).

EXPERIMENT

The simulator experiment described in this paper aimed at testing the following hypotheses:

H1. Sounds that contain ecological components (i.e. sounds that represent naturally occurring events – auditory icons) are more efficient / urgent than earcons (entirely abstract/synthetic sounds such as the ones used in the majority of all HMI systems in trucks). According to the ERM framework, familiar information contained within the sound should facilitate the emotion response process by activating associative networks and lead to more correct and rapid action.

H2. 3D sounds are more efficient/urgent than mono sounds. By combining sound icons and 3D information it is likely that one can rapidly and efficiently convey the sensation of that something dangerous is e.g. approaching from a certain direction (using 3D directional cues). Hence, such 3D sounds should be perceived as being more urgent and lead to more rapid and correct action.

Participants

20 professional truck drivers (19 male, age M= 42.3 SD= 9.2 years) and 10 Volvo employees with truck driving license (9 male, age M= 40.7 SD= 8.6 years) participated. The professional truck drivers received vouchers worth SEK 200 as compensation for their participation.

Measures

Participants' reactions to sounds were measured using self-reports, physiological measurement and behavioral methods. The Self-Assessment Manikin (SAM) – scales (shown in Figure 3) were used to collect self-ratings of Activation (high to low) and Valence (positive to negative). Participants were instructed to verbally report their responses using the scales (by saying e.g."A1, B5") after hearing each sound. A sheet with the scales was placed in the middle of the steering wheel, see Figure 5. Activation was measured physiologically by Galvanic Skin Response (GSR) measurements on participants' index and middle fingers on their non-dominant hands (see Figure 4). GSR was sampled continuously through each driving session at a rate of 391 Hz. To obtain a physiological correspondence to valence, electromyogram (EMG) responses of the Corrugator and Zygomaticus muscles (the "frown and smile" muscles in the face) were measured (see Figure 4). As with the GSR, the EMG responses were sampled throughout each driving session but at a sample rate of 3125 Hz. In addition, several driving parameters such as brake, wheel, and throttle response were logged for all sessions.

Instrumentation

Simulator - The simulator used in the experiment is shown in Figure 5. It consists of a stationary truck compartment and a 130-degree cylindrical display onto which 3 BARCO CRT projectors project the image. Side rear view mirrors views were simulated with LCD monitors. A Linux cluster consisting of one master and five slave computers were used to run the driving simulation and render the graphics. In-house developed software, "DriveSim", based on SGI Performer was used as the main simulation application. Moreover, a Windows computer was used for controlling communication between dashboard instruments and the master computer and one XPC computer was used for receiving and passing on CAN information from driver (throttle and steering wheel) to the main application. For the distraction task in part 2 (see "Scenarios" below), a Windows laptop with PowerPoint which presented numbers on a 19" LCD monitor placed on the floor to the right of the participants inside the truck compartment was used (see Figure 5).

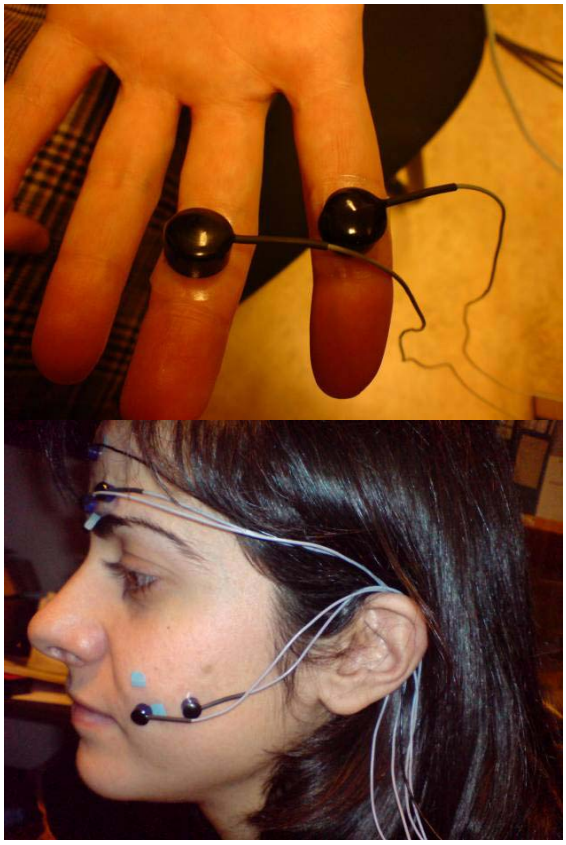

Figure 4. GSR (top) and EMG electrodes mounted on a participant.

Sound - Sound icons were presented using loudspeakers, located as shown in Figure 6, and dedicated amplifier (Creative Inspire T7700). Since only two-channel sound was available from the simulator master computer, a Dolby Pro-logic™ preamplifier (Proton AS-2620) was used to distribute the sound to the three loudspeakers. The levels of the signals sent to the amplifier/loudspeakers were adjusted in the preamplifier so that the sound was perceived equally loud from all three loudspeakers. Communication between the participant and the experiment leader (who sat at a desk approximately 6.5 m behind the simulator cab) was enabled by a talkback system consisting of two Genelec 1029A active monitors and two microphones (a Shure Prologue el. dynamic microphone at the experiment leader's desk and a Panasonic electret microphone mounted to the driver's seat) and microphone preamps. The sound inside the compartment was also recorded using a Shure BG4.1 condenser microphone and the BIOPAC system described in the next paragraph.

Physiology - To measure physiological responses, a BIOPAC MP150 system with Acqknowledge™ 3.8.1 running on a Windows laptop was used. Facial electromyogram (EMG) responses of the Corrugator and Zygomaticus muscles were acquired via Ag-AgCl shielded lead

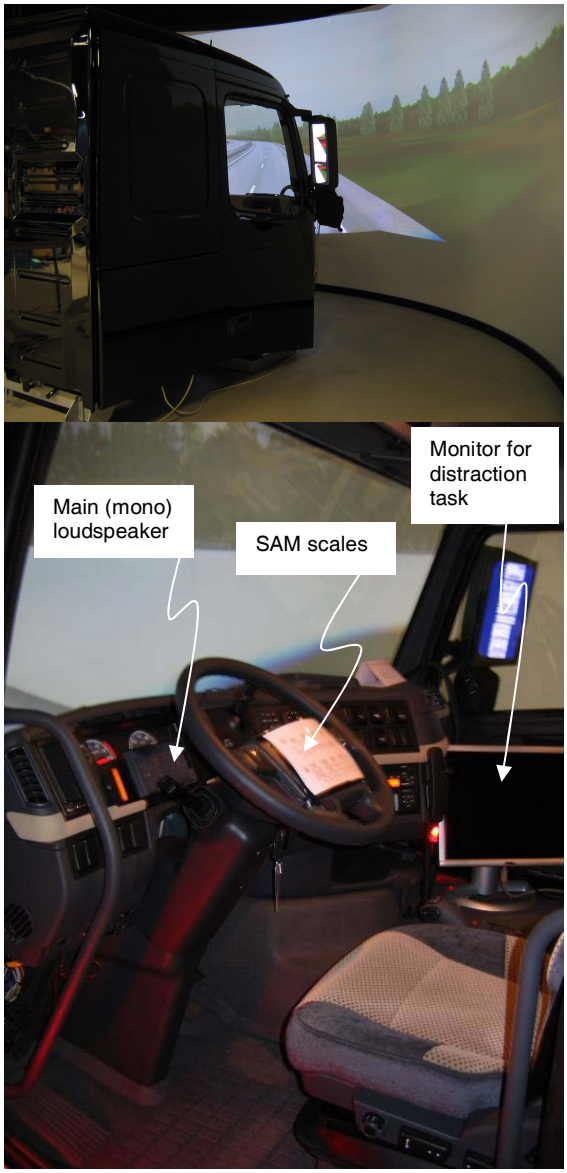

Figure 5. The simulator (top) and inside the simulator compartment, showing the main loudspeaker, SAM scales and the distraction monitor.

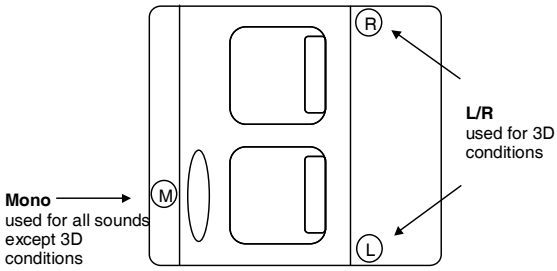

Figure 6. Position of loudspeakers in the simulator compartment.

electrodes and two BIOPAC EMG100C amplifiers. Galvanic Skin Responses (GSR) were acquired via Ag-AgCl electrodes and a BIOPAC GSR100C amplifier.

Sounds

The thirteen different warning and information sounds included in the experiment are described below. Both entirely synthetic sounds (earcons) and auditory icon-type, ecological sounds (i.e. sounds representing real events) were used.

- FCW_earcon: 4 sharp pulses of 0.1s duration and 0.01s silence between pulses. Fundamental around 207 Hz
- FCW_aicon: Auditory icon - car horn sound, continuous, 1.24 s duration. Fundamental around 417 Hz
- ACC_earcon1: 4 high-pitched pulses of 0.1s duration and 0.01s silence between pulses. Tone cluster 2482 Hz and 2631 Hz
- ACC_earcon2: 4 sharp, low-pitched pulses of 0.29s duration and 0.21s silence between pulses. Fundamental around 95 Hz
- ACC_aicon: Auditory icon - car horn sound, low pass filtered, continuous, 1.24 s duration. Fundamental around 417 Hz
- Caution_earcon1: 2 pulses of 0.1s duration and 0.2s silence between pulses – repeated once, 1.8s silence between. Fundamental around 980 Hz.
- Caution_earcon2: 2 brief tones of 0.03s duration and 0.06s silence between tones. Echo effect. Repeated once, 0.68s silence between.
- Caution_aicon: Earcon/Auditory icon hybrid: Two chime tones, 0.45s duration and the sound of a ratchet handle (symbolizing the need to contact service). Chime fundamental around 260 Hz
- LCS_mono: Auditory icon - 2 car horn honks, 0.27s duration, fundamental around 417 Hz
- LCS_3Dl: Same as LCS_mono but played in left loudspeaker
- LCS_3Dr: Same as LCS_mono but played in right loudspeaker
- LDW_earcon1: 4 very sharp pulses of 0.23s duration and 0.24s silence between pulses intended to symbolize the sound of a rumble strip. Fundamental about 74 Hz
- LDW_earcon2: 19 rapid dull pulses of 0.08s duration and 0.02s silence between pulses intended to symbolize the sound of a rumble strip. Fundamental somewhere around 68 Hz.

Design

The experiment was divided into two parts: The first part consisted of Forward Collision Warning (FCW), Automatic and Caution scenarios/sounds, and the second part consisted of Lane Change Support (LCS) and Lane Departure Warning (LDW) scenarios/sounds. Three different groups (orders) were used for the first part and two groups were used for the second part in order to randomize the presentation of different sounds for each scenario. The main design type was thus a within-group design (if one considers the scenarios for a certain event type, e.g. FCW, to be comparable). For practical reasons, Caution sounds could not be played more than once why this type of sound was a between-groups variable.

Scenarios

Thirteen different scenarios were created to test the different sounds. All roads used in the scenarios were modeled according to Swedish standards. The scenarios are described below.

1. FCW, approaching car in the wrong lane.
2. ACC, car turns into the lane in front.
3. Caution, nothing particular happens on the road but the engine warning lamp in the dashboard starts to flash for a few seconds.
4. FCW, approaching car in the wrong lane.
5. ACC, car in front brakes suddenly before intersection.
6. FCW, meeting car in intersection suddenly turns left.
7. ACC, Car turns quickly into the lane in front.
8. LCS, participant is instructed to take right to the departure lane, when a car suddenly appears in departure lane.
9. LDW, the participant is instructed to read the numbers that appears on the screen on the floor inside the simulator compartment. When the participant starts to read, the experiment leader momentarily steers the truck of the road.
10. LCS, fast bicycle crosses lane after roundabout.
11. LDW, the participant is instructed to read the numbers that appears on the screen on the floor inside the simulator compartment. When the participant starts to read, the experiment leader momentarily steers the truck of the road.
12. LCS, the participant is instructed to turn to left lane where a car suddenly appears.
13. LCS, the participant is instructed to turn to right lane where a car suddenly appears.

Procedure

Participants arrived individually to the simulator lab. A female or male test leader first briefed the participant generally about the experiment and the conditions for their participation. Participants were then seated in the simulator and physiological equipment (finger and facial electrodes) was attached to the participant. The use and control of the simulator was then introduced. Participants were instructed to try to drive as they normally drive and follow all instructions, road signs, speed limits etc. They were also specifically instructed that they could abort the test if they were feeling nauseous or uncomfortable in any way. Instructions on how to use the SAM scales were given and participants then commenced a short test drive during which they also rated a test sound (a sound which was not included in the main experiment) on the SAM scales. The main test then started with the first part (mean duration= 18 min) followed by a short break to let some air into the truck compartment and the second part (mean duration= 11 min). Participants were then debriefed and thanked for their participation.

RESULTS, PART 1

Self reports (SAM)

Ratings of activation were submitted to a 2 (earcon / auditory icon) x 2 (ACC / FCW) ANOVA to determine the influence of type of sound design and warning level on self reported activation. For simplicity of reading, the activation ratings were inverted from the original ratings so that high ratings indicate high activation. Bonferroni's method was used to adjust for multiple comparisons. As expected, it was found that participants rated FCW sounds as being more activating than the ACC sounds (M= 6.692 vs. M= 5.769, $p<.01$). The mean rated activation of the auditory icon sounds was slightly higher than for the earcon sounds (6.325 vs. 6.135), however this effect was not significant ($p= .547$).

Separate ANOVAs were performed to reveal any differences between individual sounds. A statistically significant difference in activation was found between FCW_aicon and FCW_earcon, where the former was rated as being more activating (M= 7.161 vs. 6.452, $p<.05$). No statistically significant differences were found between the ratings of the three ACC sounds (earcon1, earcon2 and aicon).

In a similar fashion, ratings of valence were submitted to a 2 (earcon / auditory icon) x 2 (ACC / FCW) ANOVA to determine the influence of type of sound design and warning level on self reported valence. The mean ratings indicate that participants rated the auditory icons as being more negative than the earcons and that the FCW sounds were more negative than the ACC sounds; however, the analysis showed that these differences were not statistically significant ($p= 0.365$ and $p= 0.524$ respectively).

Valence ratings were also analyzed separately for each sound. The results from this analysis showed that there was a marginally significant difference ($p=0.114$) in valence between FCW_earcon and FCW_aicon with the latter being slightly more negative (M= 5.161 and M=5.871 respectively). There were no such trends for the three ACC sounds.

Finally, activation and valence ratings of the "caution" sounds were submitted to independent samples t-tests as these sounds were a between-group variable in the current design. However, no significant group differences were found in either activation or valence for these sounds. It is likely however that significance would have been reached with more participants in each group.

The SAM ratings from the first part are summarized in Figure 7 where means are plotted in the activation-valence plane.

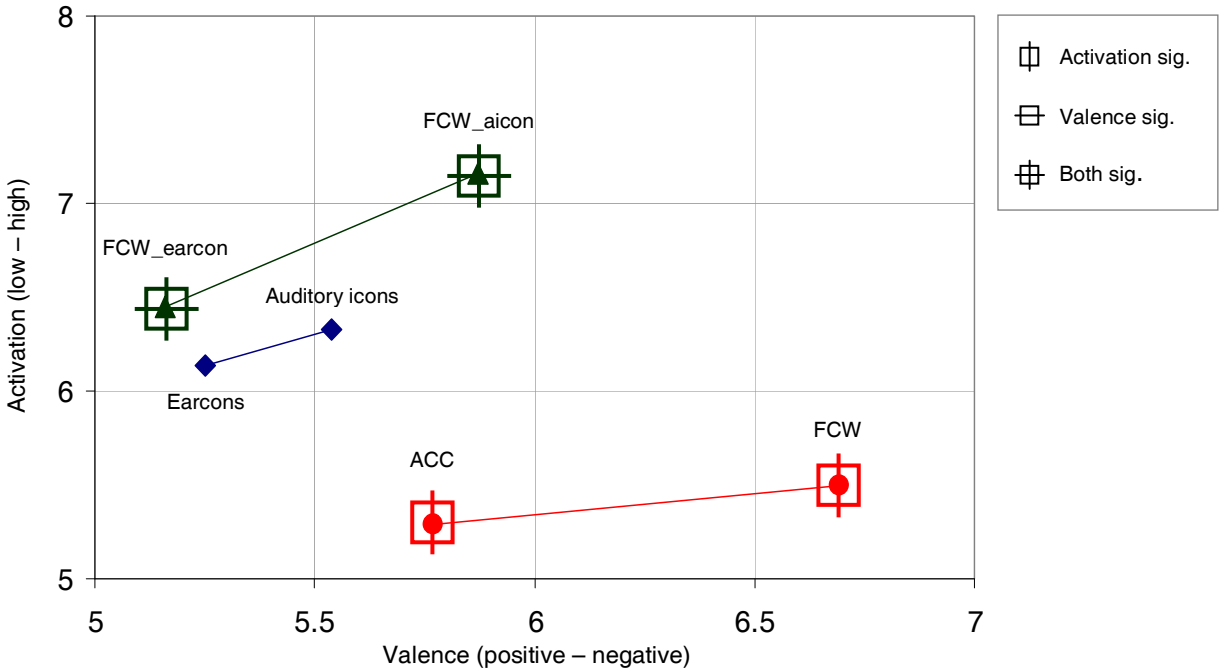

Figure 7. SAM ratings from the first part of the experiment plotted in the activation/valence plane.

Physiological responses

As data recorded from GSR electrodes was recorded continuously during each driving session, the data files first had to be cut in segments around the points in time when each sound was triggered. Next, the GSR segments were downsampled to 20 Hz. Each segment was then visually inspected and the point where a steep increase GSR curve could be noted was stored as a cut point. The GSR score was then calculated as the difference between the maximum derivative up to 2 seconds after this cut point and the maximum derivative in the segment from 2 seconds before the cut point up to the cut point. An example of a GSR segment from one of the sounds and one of the participants is shown in Figure 8.

After these pre-processing steps, GSR scores were first submitted to a 3 (earcon1/earcon2/aicon) x 2 (ACC/FCW) ANOVA to determine the overall effect of type of sound design and warning level on physiological activation. No significant differences between these factors were found. Scores were also submitted to a 1x7 ANOVA to reveal differences in activation between separate sounds. The only possible effect found was between FCW_earcon and the Caution sounds, with p=0.054. (Note that the score from three different caution sounds have been grouped together in this case). For reference, means of the GSR scores are shown in Figure 9 (whiskers show standard error).

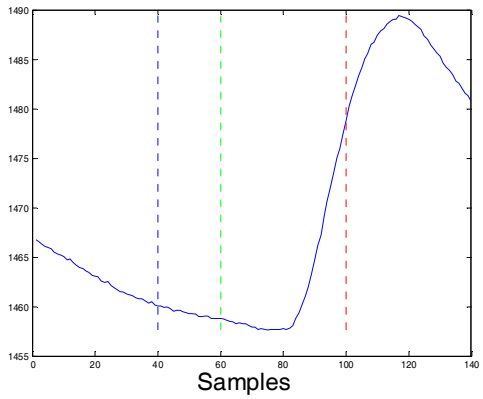

Figure 8. Example of one of the GSR segments cut from one of the data files. The leftmost dashed line marks the start of the sound, and the two other dashed lines mark the region where it is most likely to see a response to the stimulus (1-3 seconds after stimulus start). In this case, the cut point was selected at about 80 samples (i.e. 2 seconds after stimulus start).

Data from facial EMG (Corrigator / Zygomaticus) were first cut into segments around the points in time when each sound was triggered (in a similar way as was done with the GSR recordings). The segments were then downsampled to 1000 Hz and highpass filtered at 90 Hz to remove unwanted high- and low-frequency noise. Next, the envelope of the segments was extracted by taking the Hilbert transform of the segments. Finally, scores were calculated as the difference in means 2 seconds

after each sound was played and 2 seconds before each sound was played.

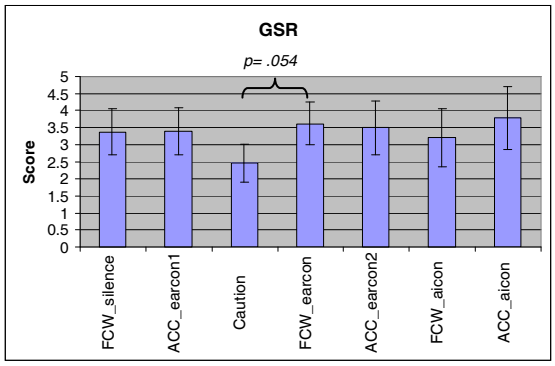

Figure 9. Means of the GSR scores. The only difference which is close to being statistically significant is the one marked with p-value (Caution vs. FCW_concept). Note however that the Caution score is a score obtained from three different sounds.

The Corrugator / Zygomaticus scores were then submitted to separate 3 (earcon1/earcon2/auditory icon) x 2 (ACC/FCW) ANOVAs to determine the overall effect of type of sound design and warning level on negative and positive valence. No significant differences between these factors were found. Scores were also submitted to a 1x7 ANOVA to reveal differences in physiological valence between separate sounds. An effect was found in the Zygomaticus score between FCW_aicon and the Caution sounds, with M=0.02 for the FCW_aicon and M= -0.02 for the caution sounds ($p<.05$) (note that the score from three different caution sounds have been grouped together in this case).

Behavioral responses

Brake reaction times (BRT) were extracted from simulator log files as the time from sound start to 30% of maximum brake pressure for each participant and sound. BRTs were then analyzed in a between-groups fashion using t-test for each scenario since the scenarios were fairly different and hence probably provoked different behaviors. First, it was found that BRT was significantly lower for scenario 1 (FCW situation) for the group which received the auditory icon sound compared to the group which did not receive any sound at all (N=10, M=1.9s vs. M=2.5s, $p<.05$). BRT was also significantly lower in group 1 (auditory icon) compared to group 3 (earcon) in scenario 6 (also FCW situation) with N=10, M= 0.4 vs. M=0.7, $p<.01$. In other words, the auditory icon sound gave a brake response in the range 300-600 ms faster compared to the concept sound and when no sound was presented. These results are visualized in Figures 10-11 below.

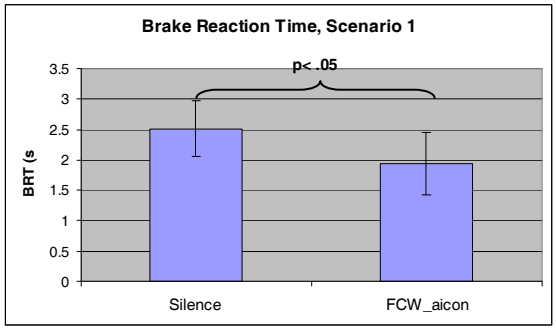

Figure 10. Brake reaction times, no sound vs. FCW_aicon. Whiskers show standard deviation. Bold p-value indicates statistically significant difference.

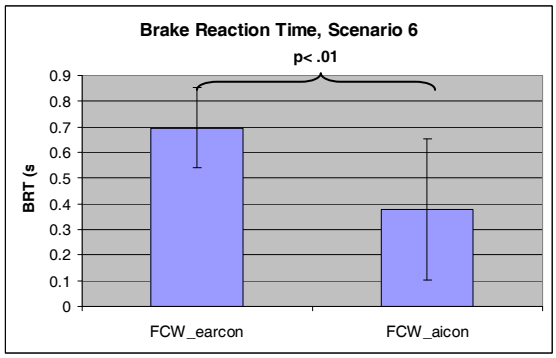

Figure 11. Brake reaction time, FCW_earcon vs. FCW_aicon. Whiskers show standard deviation. Bold p-value indicates statistically significant difference.

RESULTS, PART 2

Self reports (SAM)

Ratings of activation for the LCS sounds were submitted to a 2 (mono / 3D) x 2 (scenarios 8 and 10 / scenarios 12 and 13) ANOVA primarily to determine the influence of spatialization on self reported activation. To complete the factor analysis but also to reveal learning effects, scenario (8 and 10: highway departure & bicycle in roundabout, 12 and 13: lane change) was included as a factor. For simplicity of reading, the activation ratings were inverted from the original ratings so that high ratings indicate high activation.

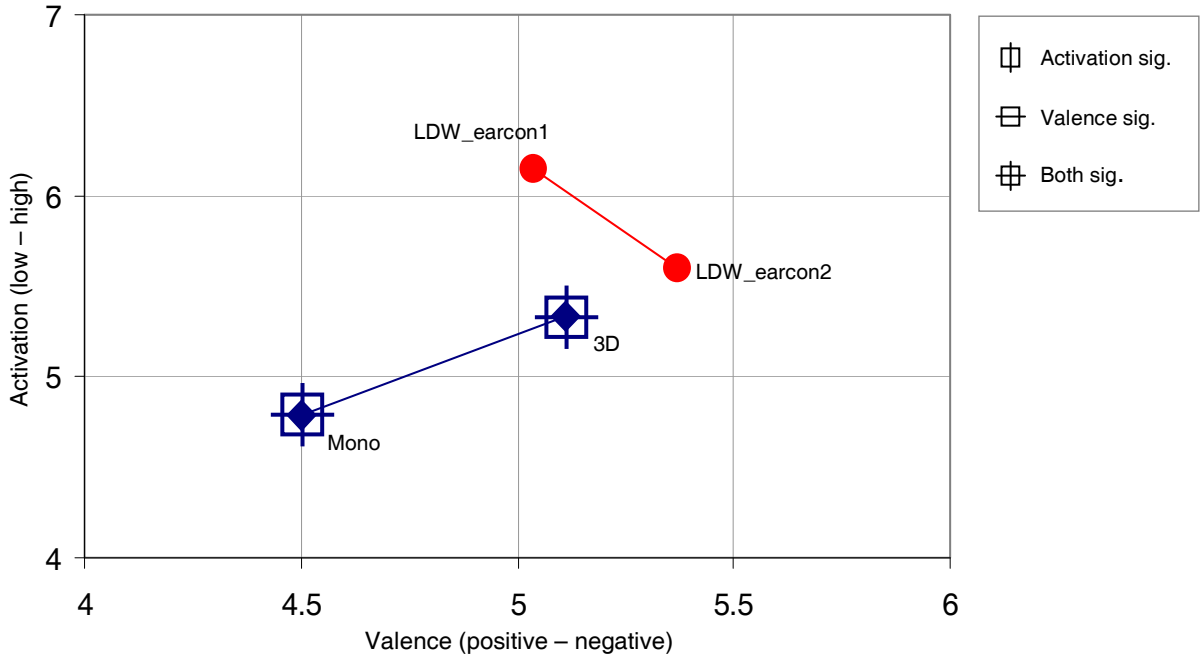

Figure 12. SAM ratings from the second part of the experiment plotted in the activation/valence plane.

Bonferroni's method was used to adjust for multiple comparisons. As expected, it was found that participants rated 3D sounds as being more activating than the mono sounds (M= 5.339 vs. M= 4.790, p<.01). The mean rated activation was also higher in the first two scenarios (8 and 10) compared to the second two (12 and 13) (5.435 vs. 4.694, p<.05), indicating that participants felt that the latter two situations were less serious or that they were more relaxed (or bored) towards the end of the test.

Furthermore, ratings of activation for the LWD sounds were submitted to an ANOVA to determine the influence of sound design (earcon1 / earcon2) on self reported activation. As before, the activation ratings were inverted from the original ratings so that high ratings indicate high activation. Bonferroni's method was used to adjust for multiple comparisons. It was found that the mean rating of the LDW_earcon1 sound was slightly higher than the LDW_earcon2 (M= 6.148 vs. M= 5.593; however, this effect was not significant (p= .134).

As with the activation ratings, ratings of valence for the LCS sounds were submitted to a 2 (mono / 3D) x 2 (scenarios 8 and 10 / scenarios 12 and 13) ANOVA primarily to determine the influence of spatialization on self reported valence. Bonferroni's method was used to adjust for multiple comparisons. It was found that participants rated 3D sounds as being more negative than the mono sounds (M= 5.113 vs. M= 4.500, p<.01). The mean rated valence was also slightly higher (i.e. more negative sensations) in the first two scenarios (8 and 10) compared to the second two (12 and 13) (4.984 vs. 4.629); however, this effect did not reach significance (p= 0.150).

Moreover, ratings of valence for the LWD sounds were submitted to an ANOVA to determine the influence of sound design (earcon1 / earcon2) on self reported valence. Bonferroni's method was used to adjust for multiple comparisons. It was found that the mean rating of the earcon2 LDW sound was slightly higher than the earcon1 sound (M= 5.370 vs. M= 5.037; however, this effect was not significant (p= .320).

The results from the analysis of the SAM ratings from the second part of the experiments are summarized in Figure 12.

Physiological responses

GSR data was preprocessed in the same fashion as in part 1 and the resulting GSR scores were then first submitted to a 2 (mono / 3D) x 2 (scenarios 8 and 10 / scenarios 12 and 13) ANOVA primarily to determine the influence of spatialization on physiological activation. To complete the factor analysis but also to reveal learning effects, scenario (8 and 10: highway departure/bicycle in roundabout, 12 and 13: lane change) was included as a factor. As with the self-reports, the last two scenarios were in mean less activating than the first scenarios (M= 1.747 vs. M= 0.403), however this difference was not significant (p= 0.139). Similarly, the 3D sounds were in mean more activating than the mono sounds (M= 2.845 vs. M=-.695, shown in Figure 13), but again this difference was not statistically significant (p= 0.378). Scores resulting from the LDW sounds were then submitted to a 1x2 ANOVA to reveal

differences between sound designs (earcon1 and earcon2). It was found that the earcon1 sound was significantly more activating than the earcon2 sound (M= 3.628 vs. 0.038, p< .05), see Figure 14.

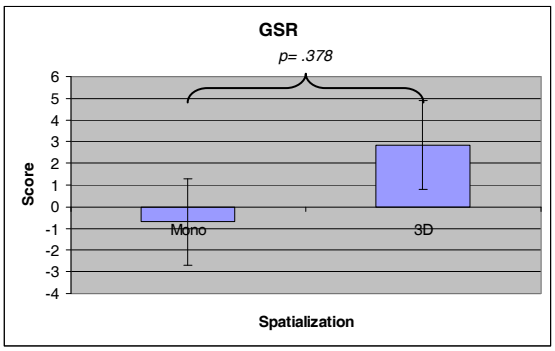

Figure 13. Means of GSR scores, mono vs. 3D sound. P-value in italics indicates non-significant difference.

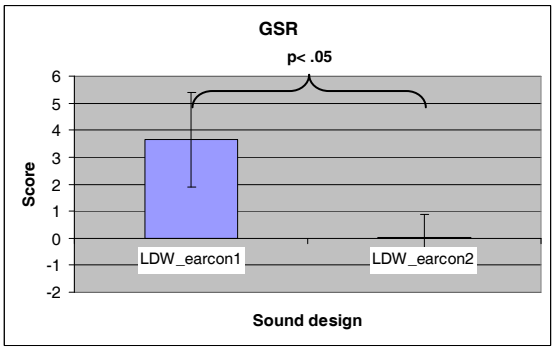

Figure 14. Means of GSR scores, LDW_earcon1 vs. LDW_earcon2. Bold p-value indicates statistically significant difference.

Data from facial EMG (Corrugator / Zygomaticus) were first preprocessed in a similar manner as in part 1. The Corrugator / Zygomaticus scores were the submitted to a 2 (mono / 3D) x 2 (scenarios 8 and 10 / scenarios 12 and 13) ANOVA to determine the influence of spatialization on physiological activation. As before, scenario (8 and 10: highway departure / bicycle in roundabout, 12 and 13: lane change) was included as a factor. No significant differences between these factors were found. Corrugator and Zygomaticus scores resulting from the LDW sounds were then submitted to separate 1x2 ANOVAs to reveal differences between sound designs (existing and concept). No significant effects were found in this case either.

Behavioral responses

Brake reaction times (BRT) were extracted for the LCS scenarios only (as BRT was considered to be an inappropriate measure for the LDW scenarios) from simulator log files in a similar manner as in the first part. Due to a technical error, only responses from Scenarios 8 and 12 could be analyzed. BRTs over 4s were excluded from the dataset. These BRTs were then analyzed in a between-groups fashion using t-test for each scenario. No statistically significant differences were found. The BRTs are shown in Table 1 below. As can be seen, the number of valid sample points is low for all scenarios/conditions, about 50% for Scenario 8 and about 25-30% for Scenario 12, which simply means that the participants' response was not always to depress the brake when hearing the sound. In other words, BRT may not be the most suitable measure of behavioral response in this case.

Table 1.
BRTs for the LCS scenarios 8 and 12

	Group	N	Mean	Std. Dev.
Scenario 8	3D	8	1.039888	0.676731
	mono	8	1.797263	1.125288
Scenario 12	3D	4	1.9008	1.366652
	mono	5	0.85034	0.455207

DISCUSSION

In part 1, it was found that FCW sounds were more activating than ACC sounds, which supports the overall sound design goal for these two types of sounds. Auditory icons in general were not more activating than the earcons, but specifically for the FCW sounds is seems that the auditory icon was more efficient in activating the driver. There were no significant differences in valence for the part 1 sounds, although a trend indicating that the auditory icon FCW was more negative than the earcon FCW sound could be seen. The findings from the SAM measurements were not reflected in the physiological measurements. It is likely that it is the situation/scenario itself, and not the sound, which causes the dominant stress reaction and any physiological response differences due to difference in sounds are overridden. This is supported by the fact that the only close-to-significant case in the GSR recordings was between the Caution sound (where nothing actually happened) and FCW_earcon (a collision scenario). The same effect was found also in the Zygomaticus scores between Caution and FCW_aicon (the effect was statistically significant in that case). To investigate this matter, more controlled studies with repetitions of each sound and/or a between-groups design with more participants would be required. The current GSR scores were actually analyzed also in a between-groups fashion for each scenario, but probably due to the low number of participants, this analysis did not show any statistically significant results. Brake reaction times however

confirmed the SAM ratings and hypotheses in terms of auditory icons being more urgent than earcons (i.e. more activating and more negative). Both compared to no sound at all and compared to the FCW_earcon sound, the FCW_aicon sound resulted in a much quicker brake reaction; mean reaction times were in the range 300ms – 600ms shorter for the auditory icon sound.

In part 2, strong evidence for 3D sounds being more urgent (i.e. more activating and more negative) was found in the SAM ratings. It should be again noted that the ”3D sounds” and the ”mono sounds” were exactly the same car horn sounds, but played through either the mono loudspeaker in front of the participant or through one of the rear loudspeakers. Interestingly, during the debriefing sessions with the participants, not many reported having heard any difference between the conditions; still, the ratings indicate a difference in emotional response and urgency. An explanation to the effect is that participants to greater extent associated the 3D sound with something outside the car – an approaching and potentially dangerous situation – while the mono sound was associated with their own car horn i.e. something less urgent. An effect of scenario could also be seen indicating that the latter scenarios were either less activating and less negative or that participants simply became less alert / more relaxed as the test progressed. The GSR scores partly confirmed the SAM ratings, although no significant effects were found, but as in part 1 it is likely that one has to employ a different design to reveal any physiological response differences between the types of spatialization. The behavioral measure, brake reaction time, seemed to be inappropriate for the scenarios used in part 2; few instances were participants actually pressed the brake in response to the sound occurred and no statistically significant differences could be found in the data which did pass the pre processing stage. It is likely that other behavioral measures such as gaze would be more suitable for these scenarios or in general when spatial properties of sound are to be investigated. Concerning the LDW sounds, no strong results were found in SAM ratings, although a trend was found indicating that LDW_earcon1 sound is more activating than the LDW_earcon2 sound. This finding was supported by GSR scores, where a significantly lower score was found for the LDW_earcon2 sound. One should however treat this result with some caution since large wheel deflections, which may have influenced the GSR score, often were the response to the sound in these scenarios.

In sum, strong support was found for the hypothesis that auditory icons are more efficient than conventional earcons in urgent situations. It was shown that both emotional response and brake reaction time could be significantly improved by using an auditory icon sound. Moreover, spatialized sounds were found to be more activating and more negative, i.e. more urgent, in a lateral warning situation (e.g. lane change). These findings should be considered in future DVI development, but it is recommended that more controlled studies are carried out to establish the optimal parameters of these sound design dimensions.

ACKNOWLEDGEMENTS

This research was supported by the Swedish Foundation for Strategic Research (SSF) and Vinnova.

REFERENCES

Belin, P. & Zatorre, R. J. (2000). What, where, and how in auditory cortex. Nat Neurosci, 3, 965-966.

Bisping, R. (1995). Emotional effect of car interior sounds: Pleasantness and power and their relation to acoustic key features. SAE paper 951284, 1203-1209.

Bisping, R. (1997). Car interior sound quality: Experimental analysis by synthesis. Acta Acustica, 83, 813-818.

Blood, A.J. & Zatorre, R.J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated with reward and emotion. Proceedings of the National Academy of Sciences, 98, 11818-11823.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion:The self-assessment manikin and the semantic differential Journal of Behavioral Therapy and ExperimentalPsychiatry, 25, 49–59.

Bradley, M. M., & Lang, P. J. (2000). Affective reactions to acoustic stimuli. Psychophysiology, 37, 204-215.

Damasio, A. R. (1994). Descartes' Error: Emotion, Reason, and the Human Brain. Avon Books: New York.

Damasio, A. R., Grabowski, T. J., Bechara, A., Damasio, H., Ponto, L. L. B., Parvizi, J. & Hichwa, R. D. (2000). Subcortical and cortical brain activity during the feeling of self-generated emotions. Nature Neuroscience, 3, 1049-1056.

Edeline, J. M. & Weinberger, N. M. (1992). Associative retuning in the thalamic source of input to the amygdala and auditory cortex: Receptive field plasticity in the medial division of the medial geniculate body. Behavioral Neuroscience, 106, 81-105.

Edworthy, J., Hellier, E. (2006). Complex nonverbal auditory signals and speech warnings. – In Handbook of warnings, M. S. Wogalter (ed.). Lawrence Erlbaum Associates, 199-220.

Giard, M. H., Lavikainen, J., Reinikainen, K., Perrin, F., Bertrand, O. & Thévenet, M. et al. (1995). Separate representation of stimulus frequency, intensity and duration in auditory sensory memory: An event-related potential and dipole-model analysis. Journal of Cognitive Neuroscience, 7, 133-143.

Haas, E. C., Casali, J. G. (1995). Percieved urgency and response time to multi-tone and frequency-modulated warning signals in broadband noise. Ergonomics 38 (11) 1995 2313-2326.

Hellier, E. J., Edworthy, J. (1996). On using psychophysical techniques to achieve urgency mappings in auditory warnings. Applied Ergonomics 30, 167-171.

Jacobsen, T., Horenkamp, T. & Schröger, E. (2003). Pre-attentive memory-based comparison of sound intensity. Audiology & Neuro-Otology, 8, 338-346.

Jääskeläinen, I. P., Ahveninen, J., Bonmassar, G., Dale, A. M., Ilmoniemi, R. J. & Levänen, S. et al. (2004). Human posterior auditory cortex gates novel sounds to consciousness. Proceedings of the National Academy of Sciences of the USA, 17, 6809-6814.

Lang, P. J. (1995). The emotion probe: Studies of motivation and attention. American Psychologist, 50, 372-385.

Lang, P. J. Lang, M. K. Greenwald, M. M. Bradley, A. O. Hamm. (1993). Looking at pictures: affective, facial, visceral and behavioral reactions. Psychophysiology 30, 261-273.

LeDoux J.E. (2000). Emotion circuits in the brain. Annual Review of Neuroscience, 23, 155-184.

Panksepp, J. & Bernatzky, G. (2002). Emotional sounds and the brain: The neuro-affective foundations of musical appreciation. Behavioural Processes, 60, 133-155.

Peretz, I., & Zatorre, R. (2005). Brain organization for music processing. Annual Review of Psychology, 56, 89-11.

Rauschecker, J. (1998). Cortical processing of complex sounds. Current Opinion in Neurobiology., 8, 516-521.

Russell, J. A. (1980). The circumplex model of affect. Journal of Personality and Social Psychology, 39, 1161-1178.

Saarinen, J., Paavilainen, P., Schröger, E., Tervaniemi, M. & Näätänen, R. (1992). Representation of abstract attributes of auditory stimuli in the human brain. NeuroReport, 3, 1149-1151.

Schröger, E. (1997). On the detection of auditory deviants: A pre-attentive activation model. Psychophysioiogy, 34, 245-257.

Sebe, N., Cohen, A., Garg, M. S., Huang, T.S. (2002). Emotion Recognition Using a Cauchy Naive Bayes Classifier. International Conference on Pattern Recognition (ICPR'02), vol I, 17-20, Quebec City, Canada, August 2002.

Tajadura, A. & Västfjäll, D (2008). Auditory induced-emotion: A neglected channel for communication in HCI. In Affect and Emotion in HCI (Eds- C, Peter and B. Russell). Springer Verlag: Germany.

Tiitinen, H., May, P., Reinikainen, K. & Näätänen, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. Nature, 370, 90-92

Todd, N. (2001). Evidence for a behavioral significance of saccular acoustic sensitivity in humans. Journal of the Acoustical Society of America, 110, 380-390.

Västfjäll, D., Kleiner, M & Gärling, T. (2003). Affective reactions to interior aircraft sound quality. Acta Acustica, 89, 693-701.

Västfjäll, D., Friman, M., Gärling, T & Kleiner, M. (2002). The measurement of core affect: A Swedish self-report measure. Scandinavian Journal of Psychology, 43, 19-31.

Västfjäll, D. (2007). Affective reactions to sounds without meaning, in press. Cognition and Emotion.

Västfjäll, D., Sköld, A., Genell, A., Larsson, P., Tajadura, A., Väljamäe, A., & Kleiner, M. (2007). Emotional acoustics: A new framework for designing and measuring the effects of information and warning sounds. Manuscript submitted for publication.

Winters, J. J. (1998). An investigation of auditory icons and brake response times in a commercial truck-cab environment. Virginia Polytechnic Institute and State University (M. Sc. Thesis).