A METHOD FOR COMPARATIVE EVALUATION OF LISTENING TO AUDITORY DISPLAYS BY DESIGNERS AND USERS

Milena Droumeva

Faculty of Education,
Simon Fraser University
`mvdroume@sfu.ca`

Iain McGregor

School of Computing,
Edinburgh Napier University
`i.mcgregor@napier.ac.uk`

ABSTRACT

The process of designing and testing auditory displays often includes evaluations only by experts, and where non-experts are involved, training is commonly required. This paper presents a method of evaluating sound designs that does not require listener training, thus promoting more ecological practices in auditory display design. Complex sound designs can be broken down into discrete sound events, which can then be rated using a set of sound attributes that are meaningful to both designers and listeners. The two examples discussed in this paper include an auditory display for a commercial vehicle, and a set of sound effects for a video game. Both are tested using a repertory grid approach. The paper shows that the method can highlight similarities and differences between designer and user listening experiences thus informing design decisions and subsequently reception.

1. INTRODUCTION

One of the concerns that designers have regarding the reception of auditory displays has to do with sounds being informative rather than uninformative [1, 2]. As well, the aesthetics of an auditory display are thought to affect its usefulness. If a design is too pleasing it becomes musical and listeners are distracted [3]. However, if an auditory display is displeasing it can become annoying [4]. Clarity is an important issue for video game sound design as it can allow a player to identify important game events and react accordingly [5]. Affect (emotion) is an increasingly important feature for the design of auditory displays, as positive sounds affective sounds are also responded to more quickly and attended to for longer [6]. Audio taxonomies are methods of describing sounds using readily identifiable concepts and terms [7]. To a limited extent, the taxonomies of auditory experiences have been explored for sound design purposes [8, 9, 10]. The intent has mostly been towards communication between auditory professionals, rather than as a mechanism for comparing listener and designer experiences [11, 12]. Audio professionals spend a considerable amount of time learning to shift between analytical [13] and 'everyday' listening [8], and Coleman [14] highlights the distrust that designers have for non-experts' descriptions of auditory environments. This mistrust might be due to the fact that non-experts normally listen differently than experts – employing more 'everyday' modes of listening – that is, listening for sound source, context and event types [15]. Analytical listening, in contrast, requires attending to sound properties, character, spatial and timbre qualities akin to Pièrre Schaeffer's 'reduced listening' [16]. As such, non-experts often require training in order to describe how they listen in terms meaningful to designers and researchers. With the present study we suggest a way of promoting more ecological design practices with regard to auditory display design that take into account end-user listening experiences in a manner that is conducive to design work. This paper addresses the use of repertory grids in comparing designers' and listeners' experiences of sound designs regardless of the differences in their *typical* modes of listening. Two forms of interactive media have been chosen for this study: a vehicle auditory display, and video game sound effects design. The next two sections provide background of past work and sound design issues surrounding the design of auditory icons and sound for video games. Following that, we discuss the method of repertory grids and introduce the research study design. Next we introduce the results and discussion for each test condition respectively, and conclude with implications for researchers and designers.

2. AUDITORY DISPLAY SOUND DESIGN

Auditory displays have been defined by Kramer [17] as an interface between users and computer systems using sound, and are considered a natural extension of the way in which sound is used in the physical world. User interfaces often include earcons, auditory icons, sound enhanced word processors (text to speech), or other applications. Cohen [18] highlighted the need to use sound professionals rather than computer scientists, in order to ensure an aesthetically pleasing blend of sounds and appropriateness to the information being conveyed. Concerns have long now been raised about end users not being considered sufficiently in the field of auditory display design, given they ultimately experience these sound designs. Barrass and Frauenberger [19] emphasize that designers need to consider the context of use as auditory displays might be used in a wide variety of environments and by a range of users.

Earcons can be defined as nonverbal audio messages directly related to icons [20]. Short, discernable, musical phrases, or motives, allow numerous alarms to be understood concurrently [21]. Earcons have to be memorised by the listener in order to successfully map audio sequences to specific functions, and the level of difficulty varies with each method of creation [22]. Representational earcons such as the recognisable sound of a piano 'catch phrase' are the simplest to learn and understand, as compared to abstract earcons such as musical tones or sound timbres [21]. Thematic earcons provide an easier way of remembering sound events if the first level is already understood. Hierarchical abstract earcons can be very difficult to learn, both because of the sheer number of possible combinations, but also due to the complex nature of the alterations [22]. The arbitrary nature of mapping earcons prevents users applying their own previous everyday experiences, which means that each set must be learned anew. There is also a tendency for earcons to sound like musical phrases, which may not suit workplace environments, and can quickly become annoying through repetition. Earcons are often long, in order to optimise identification, however, the reliance upon an end user's memory that is inherent in the design of earcons, limits their potential. There are a number of guidelines for the optimal design of earcons that prescribe approaches to using parameters such as pitch, rhythm, timbre, spatial orientation, sound intensity and tonality in order to best convey desired information [21, 23].

3. SOUND DESIGN FOR VIDEO GAMES

Audio is indispensible in video games, its active nature aiding immersion and aiding gameplay along with the visual imagery. Jorgensen [24] argues that sound can aid usability as well as affect a player's performance. Sound effects can therefore be thought of as signals that accurately portray sound events, or referents that symbolise actions. Unlike film, games rely heavily on adaptive-interactive design or 'mixing on the fly'. Sound is typically divided into three distinct categories: dialogue, musical score and sound effects, all of which are triggered individually according to the player's interaction [25]. In contrast to a static interface system, adaptive-interactive mixing poses additional problems for auditioning individual sound events in order to ascertain their effectiveness as part of the game's soundscape.

Audio spatial cues (i.e. environmental or other ambient sounds) contribute to immersion within games in a manner similar to cinema, however in games sound aligns to the perspective of the virtual camera towards more realistic navigation [9]. Unlike music or speech, which tend to be single-stream, sound effects can convey information concurrently about the game play, the environment, and discrete objects. Sound can be triggered by a gamer's actions, or by a game event in order to provide a sense of the world the character inhabits. In order to ensure that repeating sounds, such as pistol reloading and firing or footsteps do not bore the listener too quickly, randomised elements are used for all of the signature sounds [26]. Sounds are constructed in a manner similar to film sound (a palette of raw sounds augmented with filters and modulations), however a greater variety is provided to avoid repetitiveness. Within a game soundscape, *ambience* denotes environmental sounds, which consist of two types of elements: continuous and periodic. Continuous sounds are normally longer audio loops with varying frequency and dynamics. Periodic elements are typically environment-specific randomized one-shot sounds designed to be perceived as background sounds. Ambient sounds are played continuously throughout the game in order to help keep the player immersed within the game world. There are a number of parameters pertinent to designing the spatial dimensions of sound events. Just as graphics are seen from the position of the virtual camera, audio is experienced from a virtual microphone. Through the technique of acoustical modelling, direct-path audio is augmented with echo and reverberation. Environmental geometry and material composition are calculated in real time in order to create early and late reflections, diffusion, occlusion or transmission along with their material related frequency colourations [27]. Within games, unlike other forms of media, sound effects have priority over music and dialogue and provide valuable information to the player about what is happening in their immediate environment, and beyond what is immediately visible on the screen in front of her or him.

4. LISTENING TESTS AND REPERTORY GRIDS

In order to design either a static system of earcons (an audio interface) or an adaptive-interactive game soundscape a designer would want to ensure the sound design/auditory display is functional and effective for the end user, i.e. that it is being perceived and interpreted by listeners in the way intended. Listening tests are (and have been since 1956) commonplace in the field of product design where experienced listeners (previous experience with listening tests) are preferred [28, 27]. However, listener testing has so far been limited to products such as audio reproduction equipment, audio codecs and vacuum cleaners, and has not migrated into mainstream media, and only partially into computing [30]. In addition, consumers are not typically 'expert' listeners – therefore, there is a need to develop more ecological approaches to conducting listening tests. The method that we present here uses repertory grids in order to compare designers' and end users' listening experiences without the need for specialized training.

The repertory grid technique (RGT) is a proven method of information elicitation based on Personal Construct Theory (PCT). Fransella and Bannister [31] are the first to formalise the repertory grid technique. The RGT has been used for a number of sound studies purposes such as establishing audio quality attributes, auditory display design, sound design for video, as well as generating a common terminology for describing sounds. Grill, Flexer and Cunningham [32] found that existing audio descriptors were mostly timbre related, and suggested that the RGT would be suitable for establishing constructs for a broader range of attributes such as temporal parameters and dynamics. A common approach for repertory grid analyses involves four stages: *element elicitation, construct elicitation, rating and analysis*. All of the stages except for the analysis are normally conducted during a repertory grid interview. Elements are exemplars of the chosen subject of study: in this case, audio samples or recorded soundscape files. Elements are used in the rating of sound by way of constructs, which are polar opposite descriptions of the way in which individuals compare elements: for instance, rating a sound sample as pleasant or unpleasant in terms of aesthetic experience. Typically 10 to 13 elements (samples) are used for subjective evaluation by participants using a set of constructs that are provided [39]. Elements are rated using the constructs typically on a 3, 5 or 7 point scale [33]. Two of the more common forms of analyses of data of this type are hierarchical cluster analysis (dendogram/focus graph) and a non-hierarchical cluster analysis (pringrid) [31].

5. METHOD

5.1. Participants

Two designers and 40 listeners took part in this study. The first designer is a researcher specializing in auditory display design for heavy goods vehicles. The second designer is a sound designer for video games. The 40 listeners were a sample of convenience made up from staff and students at Edinburgh Napier University. The participants all considered themselves to be without hearing difficulties, and ranged in age from early twenties through to early fifties. Both male and female participants took part with a ratio of approximately 2:1. All of the participants were able to complete all tasks without prompting.

5.2. Materials

For the auditory display case the designer made an 11:41 min stereo recording of the auditory display within a moving Heavy Goods Vehicle (HGV). A professional driver was driving the truck with a co-driver, the designer was sitting in the centre on the back seat/bunk bed. The recording was made with a pair of electret microphones attached to the designers' spectacles. This near-ear microphone technique creates a partial binaural effect, improving distance perception and reducing *inside-head-locatedness* for listeners. The designer identified 20 different sound events within the recording (see Table 1). Seven of the sound events were part of the auditory display (AD). The 13 remaining ambient sound events where either vehicle related (10) or people related (3).

Code	Description
AA	Windshield wiper
AB	Engine
AC	Tapping sound, "tick tick… tick tick" (non-imminent message, e.g. new sms message)
AD	Warbling warning (p-brake)
AE	Mech. of sound handbrake release or similar
AF	Continuous ticking (tachograph)
AG	Female speech (driver)
AH	Male speech (co-driver 1)
AI	Male speech (co-driver 2)(laughter)
AJ	Four fast beeps (telling driver that they are not attending to the driving task appropriately)
AK	Windshield wiper loud scraping
AL	"Beep beep … Beep beep" (go to workshop within x km, or fix something with the vehicle)
AM	Turn signal
AN	Turn signal off
AO	Car passing
AP	Four sharp, fast beeps (lane keeping support, the vehicle is drifting out of lane)
AQ	Fast turn signal sound 3x 2 ticks (is it broken?)
AR	Four rough beeps, slow tempo (highest urgency, stop the vehicle - oil leak or similar)
AS	Beep beep-beep beep (driver is not attending to driving task appropriately)
AT	Seatbelt fastening

Table 2: Auditory display sound events by code (bold/underlined codes denote designed auditory display sounds)

The second design utilized sound effects for a commercially released console video game. All of the sound events were part of a typical game company's sound library that designers use in the construction of game soundscapes. Eight separate audio files were included; the shortest was less than 1 second long and the longest was 1 minute and 19 seconds. Half of the files were single sound events (recordings of a female voice speaking single words) and the remaining four were atmospheric soundscapes containing three to five individual sound events each (see Table 2).

Code	Description	Cod	Description
AA	Female voice 'Bye'	AJ	Water
AB	Female voice 'Hello'	AK	Kiss
AC	Female voice 'Tomorrow'	AL	Hit
AD	Female voice 'Tonight'	AM	Birds
AE	Dog growl	AN	Waterfall
AF	Barking dog	AO	Voice
AG	Water (long)	AP	Birds soft
AH	Dog growl (long)	AQ	Birds high loud
AI	Dog barking (long)	AR	Waterfall (long)

Table 2: Sound effects design sound events by code

5.3. Design

The repertory grid technique used in this study used fixed elements and fixed constructs. The elements are the individual sound events (e.g. AA: windshield wiper), which made up the respective sound design and were provided by the designers. The categories or *constructs* used in this study were user and designer generated categories validated in two earlier studies [34, 35], as follows: pan *(left/right);* depth *(front/back);* type *(speech/sound effect);* material *(gas/sol)d);* interaction *(impulsive/continuous)s;* temporal *(short/long);* spectral *(high/low);* dynamics *(loud/soft);* content *(informative /uninformative);* aesthetics *(pleasing/displeasing);* clarity *(clear/unclear)*.

The constructs were derived through a questionnaire completed by 75 audio professionals, and a think-aloud experiment with 40 end users who were asked to describe audio stimuli. This set of categories provided a consistent indication of key dimensions for the perception of soundscapes and their relative importance. For instance, both audio professionals and listeners were concerned with the spatial orientation of a sound (Left/Right and Front/Back). Speech was differentiated from other types of sounds by both listeners and professional sound designers (Speech/Sound effect). Material (Gas/Solid) and interaction (Impulsive/Continuous) provides a method of communicating the onomatopoeic descriptions of sound events in a similar manner to Gaver's [23] *interacting materials*. Both listeners and audio professionals used temporal attributes to describe sound events (Short/Long) as well as Spectral attributes (High/Low). A loud/soft distinction was consistently used for describing dynamics. As well, both groups highlighted aesthetic attributes, simplified here to pleasing/ displeasing. The category of informative/uninformative was added in order to account for the functionality of the design. Clarity applies to the perceptual intelligibility of a sound, where both professionals and listeners used positive and negative terms to describe clarity. In the present study the set of constructs was used as a fixed schema for the rating and evaluation of sound. Listeners and designers were asked to rate sound elements using the provided categories (constructs).

5.4. Procedure

Each designer supplied the sound events in the design to be tested, and classified each sound (element) according to the rating system of constructs. Listener tests for both the auditory display and the game sound effects were conducted in an auralisation suite using fully enclosed stereo headphones. Listeners were asked to listen to an audio recording and verbally rate the elements using the supplied constructs. Each construct allowed three choices for rating, e.g. pan (left/right) could have a value of 1 (left), 2 (neither left nor right) or 3 (right). Listeners could replay the files as often as they wished, and were made fully aware of the context of use for the two designs (HGV and video game). As suggested by Fransella *et al.* [33] participant ratings were entered into the grid by the researcher thus preventing participants from comparing ratings for previous elements during the study.

According to Fransella *et al.* [33], the number of points on the rating scale only have a limited impact upon the results, except for the number of 0 ratings which increase in an evaluative 3 point scale. It is also suggested that the order in which ratings are made does not affect the results, so listeners were asked to classify an element (sound file) using all of the constructs, rather than to rate all of the elements using a single construct before moving onto the next one. Working in this direction allowed listeners to concentrate on a single sound event (element) rather than have to repeat elements (sound events) in order to become familiar with them again. A non-evaluative scale (1 - 3) was chosen over an evaluative scale (+1 0 -1) as indicating positive and negative polarity might bias results. The results were translated into tabulated information into data plots and charts. The designers' responses were inputted exactly as reported, and RepSocio (part of Repgrid) was used to compare the designers' and listeners' grids.

6. RESULTS

The results are presented below by test condition: vehicle auditory display followed by video game sound design. For each condition we discuss designer-listener comparisons with regard to both individual sound events and application of constructs. Statistical significance was not calculated at this stage since only a single designer represented the 'designer' perspective. With multiple designers scores can be compared using a 'permutation test' or similar 'bootstrap' methods. In the absence of that, we employed a convention that if the match between elements or constructs is 75% or above then they are of interest, and below this figure the results are too dissimilar to be considered effective [33]. For each condition we consider and discuss (where applicable) both between-participant matches for sound events and categories (constructs), as well as the modal listener response. In fact, it was established that the modal listener response (the most typical participant rating for each sound event according to each construct) represented the between-participants agreement more accurately than both the median and mean of individual responses.

6.1. Auditory display

The tabulated results from the AD test are presented in Figure 1. The matrix at the top left of the figure represents a listener-designer perspective by rating match. White (blank) spaces represent a match, and the numbers denote by how much the responses differ between the designer and the listeners. The figure makes possible to identify the match for each construct and each sound event. Construct matches are denoted in blue at the top right of the matrix and we can see that 45.5% of the constructs had a match of 75% or greater. Sound events or *elements* are denoted in red at the bottom right of the figure. Over half of the sound events had a match of 77.3% or greater, with the lowest match being 63.6% for four of the sound events (AI, AR, AS, AG). The auditory display test had an overall agreement between listeners and the designer of 75%. Eleven out of 20 sound events had a match of 75% or greater (see Figure 1).

Constructs with a relatively high level of agreement between the listeners and designer included: sound type as *sound effect*; duration as *short*; dynamics as *neither loud nor soft*; content as *informative*; and sound aesthetics as *neither pleasing nor displeasing*. These findings suggest that the intention for the sound design was successfully accomplished for these parameters. In terms of construct agreement for the AD test condition, listener-designer agreement for pan (left-right), sound type (speech/sound effect), and aesthetics (pleasing/displeasing) was over 95%, while the agreement for dynamics (loud/soft) and temporal nature (short/long) was at 75%.Agreement for the rest of the constructs scored below 70%. Ratings for depth (front/back) differed both between participants (agreement of 61%) as well as between listeners' and designer's perspective (67%) suggesting a wide variation in perception for both user groups. This could be due to the context of listening with an ambient soundscape characteristic of a commercial vehicle. When it came to rating the material qualities of sounds (gas/solid) while agreement between participants was significant at 77.75%, listener-designer agreement was at 57.5%. This, in conjunction with anecdotal evidence from the study, suggests that participants on the whole did not understand the concept of material properties of sound.

Results are similar, however not quite as drastic, with regard to the sound interaction construct (*impulsive/ continuous*). Again, between-participant agreement was significant (78%) while listener-designer agreement was at 70%, indicating that participants had trouble with the nature of the construct itself. Most participants rated sounds as *neither* impulsive nor continuous or *continuous*, whereas designers rated most sounds as *impulsive.* This could be due to the designer's attending to the temporal structure of sounds (attack/sustain/decay) whereas participants are listening more holistically for type of event and duration. Agreement was higher between listeners' and designer's perspective on the sounds' duration (*short/long*) indicating that participants did not necessarily correlate the construct of interaction with duration. Since this aspect of perception is important to the design of auditory displays it is worth noting that if we take the ratings of agreement for only the auditory display sounds, while designers characterized all ADs as *impulsive* and *short*, listeners experienced them consistently as mostly *short* but *neither* impulsive nor continuous. In terms of spectral characteristics (*high/low*) and dynamics of sound (*loud/soft*) between-participant agreement was significant (78% and 82% respectively) in contrast to listener-designer agreement, which was less so (65% and 75% respectively). Basically, even if listeners' interpretations diverge from the designer's, they are nevertheless consistent among listeners themselves.

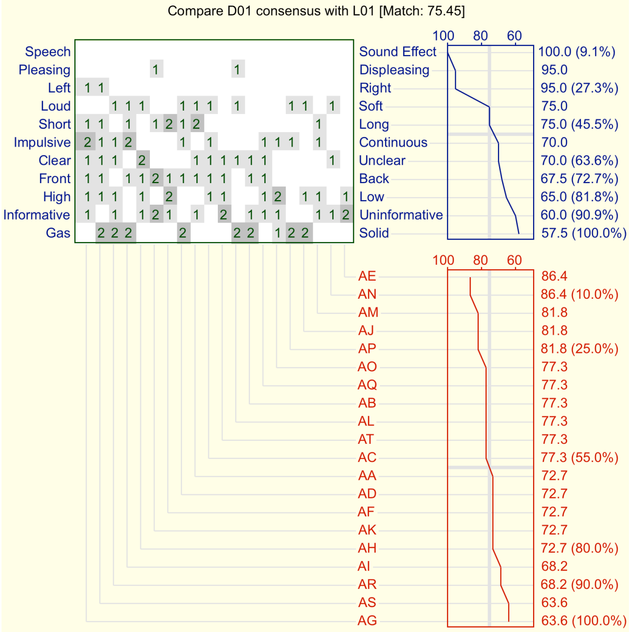

Figure 1: Comparison of designer's and listeners' application of constructs (Auditory Display test condition)

Most participants rated sounds as *neither* high nor low whereas designers rated more sounds as *high* or *neither*. Rather than misunderstanding of the construct, this finding might indicate a difference in the habitual approaches to contextual listening that designers and listeners engage in. While designers are more attuned to spectral characteristics of sound and thus more focused on accurate identification (*reduced* or *analytical* listening), end users likely hear most familiar sounds as *neither* high nor low, especially without a reference tone (everyday listening). As far as individual sound events, the lowest agreement between listeners and designer concerned two of the voice sound events (AG, AI) and two of the auditory displays (AR, AS). AR (four rough beeps) in particular was also rated by listeners as *displeasing* yet the predominant rating for both AR and AS was as *clear* and *informative*. A number of the ADs in the recording were in fact rated as *displeasing* by the participants, and one of the specialized sounds (AR: tachograph ticking sound) was the only one rated as *unclear*. These findings suggest two issues – sounds that are too specialized to the context of the activity such as the tachograph may not be as effective in coming across to an average listener, in contrast to sounds that are otherwise familiar (breaking sound, door latch). On the other hand, even familiar sounds such as windshield wiper (AA) or voices (AG, AI) can have a low match between the designer's and listeners' perspective if the recording is unclear due to soundscape density or the duration of the sample.

With regard to ADs specifically, while listeners considered most to be *neither high nor low* in frequency, they rated most of them as *loud* in terms of dynamics. Designers, on the other hand, rated all auditory displays as *neither* loud nor soft. Once again, if we consider that designers are concerned with the optimal utility and effectiveness of ADs while participants are listening holistically to all sonic elements, most vehicle beeps are likely to be heard as *loud* in an everyday context. Finally, with regard to content (*informative/ uninformative*), while both listeners and designers rated the AD portion of the soundscape as *informative*, listeners also considered the majority of other sounds to be *informative*, whereas designers considered those sounds to be *neither* informative nor uninformative. Similarly, while listeners considered most sounds including ADs to be on the whole *clear*, designers rated many sounds as *neither clear nor unclear*. This suggests that designers employ a more specific standard for clarity, likely attending to sound quality in conjunction with semantic meaning, while end users attend primarily to context and meaning.

6.2. Video game sound events

The video game sound effects test had an overall match between listeners and designer of 79%, slightly higher than the first test condition. Fourteen out of the 18 sound events had a match of 75% or greater (see Figure 2). Ratings with a relatively high level of agreement for sound events by both the listeners and designer included sound type as *sound effect*; duration as *short*; content as *informative*; aesthetics as neither *pleasing* nor *displeasing;* and pan as neither *left* nor *right*. The sound events with the lowest level of match (64%) were "birds soft" and "kiss" (AP, AK) followed closely by one of the "waterfall" sound events and a female voice sample (AR, AD). The most prominent difference between the designer's and listeners' ratings for AP ("birds soft") were that the designer rated the sound event as *left*, *front*, and *gas*, whereas the listeners rated AP as *right, back* and *solid*. For AK ("kiss") the designer considered the sound event to be *speech* and *solid*; in contrast, the listeners considered AK to be a *sound effect* and *gas*. Sounds that are timbrally ambiguous likely contribute to differences in listeners' and designer's perception as well as their subjective evaluation. For instance, designers would attend to the mechanism of sound production (lips, vocal cords) while end-users interpret the sound as a discrete event (kiss).

Similarly to the auditory display condition, it might be argued that the *informative/ uninformative* and *clear/unclear* constructs are two of the most important dimensions of sound design. These two constructs had a match of 100% and 81% respectively, suggesting that the design could be considered successful in terms of content and clarity. The constructs with the highest listener-designer match with above 80% agreement were pan (*left/right*), type (*speech/sound effect*), interaction (*impulsive/continuous*), dynamics (*loud/soft*), clarity (*clear/unclear*), content (*informative/uninformative*) and aesthetics (*pleasing/ displeasing*). Similarly to the vehicle auditory display, depth (*front/back*) and material (*gas/solid*) had a significantly lower agreement (58% and 50% respectively). While listeners rated all sounds as *solid*, designers rated sounds predominantly as *gas*. The sounds of liquid such as "water" and "waterfall" were rated by listeners as *neither gas nor solid* and the rest of the sounds were considered *solids* given that they correspond to a material event, object or sound-making body (e.g. a dog). In contrast, designers rated most vocalizations including dog vocalizations as *gas*, and the remaining as *neither*. In terms of interactional character, listeners rated sound events that are naturally continuous such as "water" and "waterfall" as *continuous*, while a naturally short event such as a kiss was rated as *short* and the rest of the non-vocal sound effects such as dog barking and birds as *neither* impulsive nor continuous.

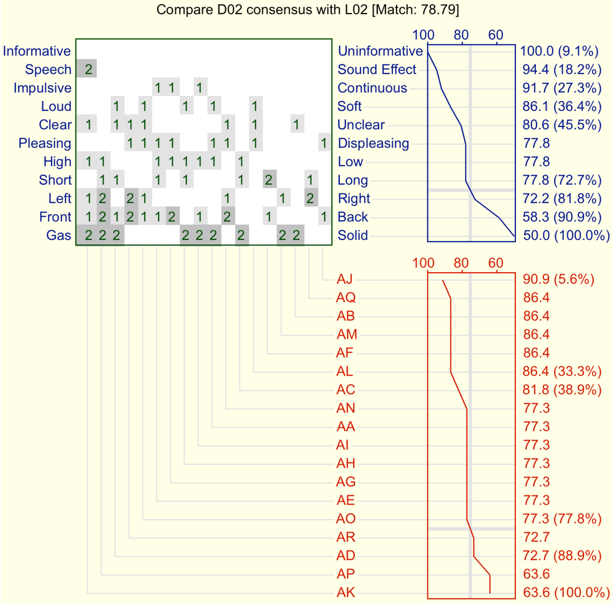

Figure 2: Comparison of designer's and listeners' application of constructs (Video game sound test condition)

With regard to the context of listening (video gameplay), it is interesting to note that participants rated all sounds as *clear* and *informative*, whereas the designer rated most sounds as *neither* informative nor uninformative. This finding might indicate that to designers 'informative' implies a sound of high importance and value; in contrast, listeners' rating of all sounds as *informative* suggests that recognition and understanding of the source already provides 'information' similarly to the way ambient sounds provide information in the context of everyday life. Once again, comparing results can signal differences between everyday and specialized listening and interpretation. While designers are much better versed in the intricacies of sound production and propagation, listening critically and analytically to sound properties, listeners potentially perceive discrete sounds events as timbrally and materially whole, in essence ignoring the nuances of sound properties and focusing on the relationship between sound source and meaning – an important consideration towards designing auditory displays.

6.3. Constructs

Looking at the application of the sound categories (constructs) across listeners and designer groups allows us to evaluate their use a sustainable tool for comparative listener evaluation. Matches that are above 75% for both groups can be considered as consistently applied across the two listening contexts and sound designs. Figures 3 and 4 provide a webbed overlay of responses in three layers (1 to 3) where 1 refers to the first rating of the construct (e.g. in the case of *clear/unclear* 1 means *clear*), 2 refers to the *neither* rating and 3 refers to the polar opposite rating (e.g. 3 means *unclear*). Only 4 constructs out of 11 had a match of 75% or higher for both the auditory display and the video game sound conditions: *speech/sound effect*, *short/long*, *loud/soft*, and *pleasing/displeasing*. A further 5 constructs had a match in the region of 65 – 75%: *left/right*, *impulsive/ continuous*, *high/low* and *clear/unclear*.

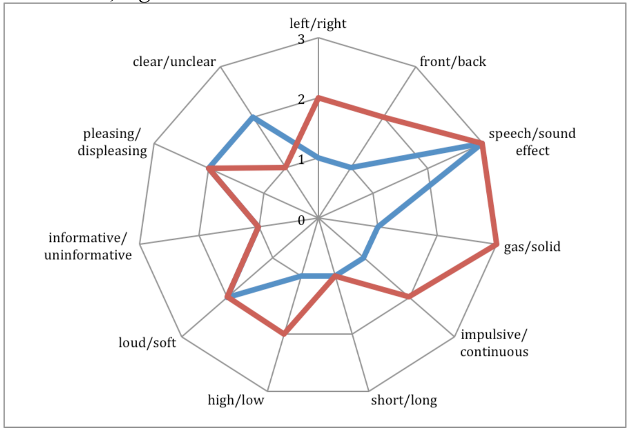

Figure 3: Comparison of designer's (blue) and modal listener's (red) application of constructs (Auditory Display test condition)

The remaining 3 constructs: *front/back*, *gas/solid* and *informative/ uninformative* all had a match below 65% and therefore were not being rated consistently. However, only *gas/solid* falls below 65% for both the auditory display and the video game sound effects. In the web chart in Figure 3 it is possible to see that the designer and the listeners broadly agreed with regard to 5 out of 11 constructs for the auditory display condition. In the video game sound effects there was a similar level of agreement with regards to 6 out of 11 constructs (see Figure 4). In both tests there were also mismatches with the spectral and dynamics ratings indicating an inconsistent listening experience or interpretation. In some cases such as *clarity* and *dynamics*, one of the conditions had a high agreement while the other did not, indicating specific differences in the perception of sound designs depending on context and sound types.

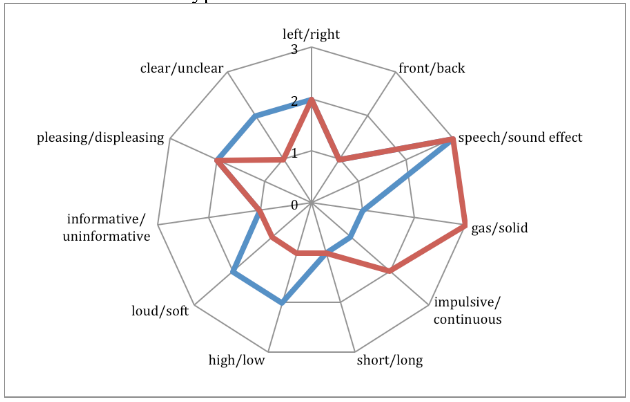

Figure 4: Comparison of designer's (blue) and modal listener's (red) application of constructs (Video game sound test condition)

7. DISCUSSION

Adopting this method of comparative evaluation highlights important differences between the design and intention of ADs in relation to how they are perceived and interpreted by end users. Areas of convergence and divergence in rating sounds using the provided constructs bring light to the way designers and end users might attend to sound with different habitual orientations. There is of course a critical difference between divergence in perception and confusion over applying semantic categories and ratings to sounds. This method of evaluative comparison allows us to explore and stipulate about both, towards improving auditory display designs and listening test procedures. Based on the data gathered from the two tests, we identify two main factors that impact designer-listener differences in experiencing auditory display designs: contextual differences in auditory displays reception and differences in everyday vs. specialized listening.

7.1. The role of context

Context refers to the general characteristics of a listening situation – what is the nature of the surrounding soundscape; what kinds of sounds are typically present; how dense is the soundscape; what is the nature of activity taking place; what are the subjective properties of sound in that context; how would a typical (vs. an expert) listener attend to sounds in that context. All of these elements form a situation that listeners approach with habitual ways of attending to sound, including interpreting the meaning of sounds and evaluating their subjective properties [8]. Specifically, with one context being a 'work' environment of commercial vehicle, and the other an 'entertainment' context of video game play, there are some salient and interesting differences in the level of designer-listener agreement. If we consider that a video game is a more highly designed and 'virtual' listening experience consisting entirely of sound effects, this is reflected in the higher level of construct agreement as both listeners and designers overwhelmingly rated sound events as *clear*, *informative*, *neither* pleasing nor displeasing, *short*, and correctly identified the *sound effects* and *voices* in the recordings. Interestingly, listeners identified a number of sounds as *pleasing* – perhaps related to the nature of the activity of gameplay, which is generally associated with leisure rather than work. In contrast, in the auditory display condition, listeners rated all of the actual ADs as *unpleasant* albeit *clear* and *informative* – possibly due to the context of driving with traffic noise and the association with work.

7.2. Everyday vs specialized listening

Many of the divergences in listener and designer experience of listening to the two sound designs definitely signal differences between a specialized and an everyday approach to listening and interpretation of sound events. There are a number of theories and classifications of listening modes [8, 13] that involve distinctions between active and passive listening, or aesthetic versus informational listening. Using Schaeffer's ontology, Chion [16] distinguishes between causal (listening for source), semantic (listening for meaning) and reduced (listening to sound's character) modes of attention. In this study, we borrow from these categories to construct an 'everyday' and a 'specialized' or expert listening attention towards a comparison of end user and designer ratings of sound events by category. For instance, designers apply a more discerning and highly trained listening that attends to material qualities of sound production and propagation, relative amplitude, spectral interaction with other sonic elements in the soundscape and semantic content towards the intended function of the sound design. On the other hand, end users likely apply a more 'everyday' [8, 15] approach to the experience of ADs and sound effects in context. An everyday listening approach considers most if not all sounds inherently informative, with ADs being particularly clear since they are timbrally and spectrally unique; qualities like sound interaction (*impulsive/continuous*) and material (*gas/solid*) might be experienced by everyday listeners more holistically and timbrally rather than empirically and functionally.

To a designer, ADs serve an informational function, while the rest of the sounds serve a less important, ambient function. To an end user, most sounds are experienced as *informative* and *clear* and in the context of everyday listening each sound event gives information about a number of dimensions: general ambience, the events taking place and the course of action needed. Listeners do not necessarily expect ADs to be excessively clear, informative or pleasant, as long as they are identified and understood in the context of the larger soundscape. Therefore the divergence in agreement with respect to sound design functions seem to point to a difference in values and interpretation on behalf of designers and listeners respectively – something that needs to be taken into account when designing auditory displays for average 'everyday' listeners. That is, in some cases, designers may not be the best judge of which auditory displays 'get the job done' and are perceived easily and clearly by end users.

8. CONCLUSION

In this paper, we have presented the results and discussion from two listening tests conducted as comparative evaluation between 40 listeners and 2 designers in two listening conditions – an auditory display environment for a transport vehicle, and a set of video game sound effects. The study was conducted using a repertory grid approach [39] and the tabulated results are presented and discussed. We offer this approach as a model for conducting listening tests without prior training and as a more ecological way of involving the end users in the design process of auditory displays. Comparing agreement allows us to rate the suitability of the constructs in relation to evaluating a wide variety of sound designs including auditory displays, sound effects, interface sounds and complete soundtracks. Essentially, such a comparison works to highlight where the designer's intention is perceived by the listeners accurately and where there is a misalignment. Differences in rating of sound events, in turn point to possible flaws in the auditory design, or a mismatch between designer and listener expectations and habituation to listening-in-context. The findings in this study offer a comparison between the values that end users and designers place on sounds, the degree of precision in terms of identifying and interpreting the sound designs; and the confusion over sound constructs that might require training in order to be articulated by listeners. We hope to have demonstrated that using a repertory grid approach in comparative evaluations of sound design can be a unique and valuable tool when conducting listening tests for the design of a wide variety of auditory displays and contexts of use.

9. REFERENCES

[1] Buxton, W. (1992) The three mirrors of interaction: a holistic approach to user interfaces. *Industrial DESIGN*, Japan Industrial Designer's Association, pp.6 - 11.

[2] Brewster, S. A. (2008). Nonspeech auditory output. In A. Sears & J. Jacko (Eds.), *The Human Computer Interaction Handbook* (2nd ed., pp. 247-264).

[3] Vickers, P., & Hogg, B. (2006). Sonification abstraite/sonification concrete: An'aesthetic perspective space'for classifying auditory displays in the ars musica domain. *In Proc. 12th International Conference on Auditory Display*, London.

[4] Henkelmann, C. (2007). *Improving the Aesthetic Quality of Realtime Motion Data Sonification.* University of Bonn.

[5] Ekman, I., & Lankoski, P. (2009). Hair-Raising Entertainment: Emotions, Sound, and Structure in Silent Hill 2 and Fatal Frame. In B. Perron (Ed.), *Horror video games: essays on the fusion of fear and play* (pp. 181-199). Jefferson, NC: McFarland.

[6] Schleicher, R., Sundaram, S., & Seebode, J. (2010). *Assessing audio clips on affective and semantic level to improve general applicability.* Paper presented at the Fortschritte der Akustik - DAGA 2010, Berlin.

[7] Cano, P., Koppenberger, M., Le Groux, S., Ricard, J., Herrera, P., & Wack, N. (2004). *Nearest-neighbor generic sound classification with a WordNet-based taxonomy.* In Proc. 116th AES Convention, Berlin, Germany.

[8] Gaver, W. W. (1993). What in the World do we Hear? *Ecological Psychology, 5*(1), 1-29.

[9] Grimshaw, M. (2008). The Acoustic Ecology of the First-Person Shooter: The Player Experience of Sound in the First-Person Shooter Computer Game. Saarbrucken: VDM Verlag Dr. Muller.

[10] Liljedahl, M., & Fagerlönn, J. (2010). *Methods for sound design: a review and implications for research and practice.* In Proc. at the 5th Audio Mostly Conference.

[11] Brazil, E., & Fernström, M. (2009). Subjective experience methods for early conceptual design of auditory displays.

[12] Frauenberger, C., & Stockman, T. (2009). Auditory display design—An investigation of a design pattern approach. *International Journal of Human-Computer Studies, 67*(11), 907-922.

[13] Truax, B. (2001). *Acoustic Communication*. Ablex Publishing.

[14] Coleman, G. W. (2008). *The Sonic Mapping Tool.* (PhD), University of Dundee.

[15] Ballas, J. (1992). Common Factors in the Identification of an Assortment of Brief Everyday Sounds. J*ournal of Experimental Psychology*: Human Perception and Performance, 19(2), pp. 250-267.

[16] Chion, Michel (1994). *Audio-Vision: Sound on Screen*. New York: Columbia University Press.

[17] Kramer, G. (1994). An Introduction to Auditory Display. In G. Kramer (Ed.), *Auditory Display: Sonification, Audification, and Auditory Interfaces* (pp. 1-77). Reading, MA: Addison-Wesley.

[18] Cohen, J. (1994). Out to Lunch: Further Adventures Monitoring Background Activity. In G. Kramer & S. Smith (Eds.), *Proceedings of the Second International Conference on Auditory Display* (pp. 15-20)

[19] Barrass, S., & Frauenberger, C. (2009). A coummunal map of design in auditory display *Proceedings of the 15 International Conference on Auditory Display, Copenhagen, Denmark, May 18-22, 2009*

[20] Blattner, M. M., Sumikawa, D. A., & Greenberg, R. M. (1989). Earcons and Icons: Their Structure and Common Design Principles. *Human-Computer Interaction, 4*(1), 11-44.

[21] Brewster, S. A. (1994). Providing a Structured Method for Integrating Non-Speech Audio into Human-Computer Interfaces. (PhD), University of York, York.

[22] Leplâtre, G., & Brewster, S. A. (2000). Designing non-speech sounds to support navigation in mobile phone menus *International Conference on Auditory Display*.

[23] Gaver, W. W. (1997). Auditory Interfaces. In R. M. Baecker et al. (Eds.), *Readings in Human Computer Interaction* (2nd ed., pp. 1003-1041). San Francisco: Morgan Kauffmann Publishers Inc.

[24] Jorgensen, K. (2006). On the Functional Aspects of Computer Game Audio *Audio Mostly* (pp. 48-52).

[25] Sanger, G. A. (2004). The Fat Man on Game Audio: Tasty Morsels of Sonic Goodness. Indianapolis, IN: New Riders.

[26] Lecky-Thompson, G. W. (2002). *Infinite Game Universe: Level Design, Terrain, and Sound*. Hingham, MA: Charles River Media.

[27] Brandon, A. (2005). *Audio for Games*. Berkeley, CA: New Riders.

[28] Bech, S. (1992). Selection and Training of Subjects for Listening Tests on Sound-Reproducing Equipment. *Journal of the Audio Engineering Society, 40*(7/8), 590 - 610.

[29] Engelen, H. (1998). Sounds in Consumer Products. In H. Karlsson (Ed.), *Stockholm, Hey Listen!* (pp. 65-66). Stockholm: The Royal Swedish Academy of Music.

[30] Bech, S., & Zacharov, N. (2006). *Perceptual Audio Evaluation*. Chichester, West Sussex: Wiley.

[31] Fransella, F., & Bannister, D. (1977). *A manual for repertory grid technique*. New York: Academic Press.

[32] Grill, T., Flexer, A., & Cunningham, S. (2011). *Identification of perceptual qualities in textural sounds using the repertory grid method.* Paper presented at the 6th Audio Mostly Conference, Coimbra, Portugal.

[33] Fransella, F., Bell, R., & Bannister, D. (2004). *A Manual for Repertory Grid Technique* (2nd ed.). Chichester, UK: John Wiley & Sons.

[34] McGregor, I., Leplatre, G., Crerar, A., & Benyon, D. (2006). Sound and Soundscape Classification: Establishing Key Auditory Dimensions and their Relative Importance *ICAD 2006* London: Department of Computer Science, Queen Mary, University of London

[35] McGregor, I., Crerar, A., Benyon, D., & Leplatre, G. (2007). Establishing Key Dimensions for Reifying Soundfields and Soundcapes from Auditory Professionals *ICAD 2007*.