

## EXPLORING CONCURRENT AUDITORY ICON RECOGNITION

Eoin Brazil, Mikael Fernström

John Bowers

Interaction Design Centre  
University of Limerick  
Limerick, Ireland

Department of Design  
Goldsmiths, University of London  
London SE14 6NW, England

eoin.brazil@ul.ie, mikael.fernstrom@ul.ie

john.m.bowers@gmail.com

### ABSTRACT

This poster explores and deepens existing studies into the identification of concurrently presented auditory icons. The motivation for this work was to gain a better understanding of auditory icons where several are played together simultaneously. A set of descriptors for everyday sounds were collected from participants and classified into action and object categories. The exploration considered the hypothesis that when auditory icons did not have the same object or action descriptors that the identifications of the auditory icons would improve. This was studied in conditions where three, six, and nine sounds were simultaneously presented. These conditions had two distinct sub-categories, the first category used a prior classification of sounds to ensure no sound pair in the condition had the same action or object properties. The second sub-category used random selection of the sounds meaning that similar sound could exist within the particular condition. An onset-to-onset gap of 300 ms between sounds being presented was used. The results supports earlier findings and showed that distinguishing between object and action properties of auditory icons did improve their identification accuracy. It was found that prior classification allows listeners to achieve close to a 10% identification improvement in accuracy.

### 1. INTRODUCTION

Auditory icons were defined by [1] as "everyday sounds mapped to computer events by analogy with everyday sound producing events". The sounds used in this exploration are auditory icons based on everyday sound, and do not typically fit into the categories of music or speech. The definition of an everyday sound is taken from Vanderveer [2] who defined it as:

"Any possible audible acoustic event which is caused by motions in the ordinary human environment. ... Besides having real events as their sources ... [everyday sounds] are usually more complex than laboratory sinusoids, ... [everyday sounds] are meaningful, in the sense that they specify events in the environment. ... The sounds to be considered are not part of a communication system, or communication sounds, they are taken in their literal rather than signal or symbolic interpretation." [2, p. 16-17]

Prior work in auditory display has often concentrated on situations where a single event or message, sometimes complex, is being conveyed but with increasingly more complex interoperating systems there is a growing need for the ability to convey multiple events or messages simultaneously. The lack of guidelines and research in the area of conveying multiple events or messages using auditory icons led us to our investigation of concurrently presented auditory icons. This work uses the definition of concurrent auditory icons as the playing of several auditory icons together and

simultaneously, to build more complex and compound auditory icons. This is inspired by how real sounds work in the world and by previous research on concurrent audio presentation [3, 4, 5, 6].

Concurrent audio presentation offers advantages such as increased bandwidth and faster presentation but can suffer from disadvantages such as the sounds can interfere with each other, confused interpretations, or perceptual masking. This can occur even where sounds are played in a sequence as Moore et al. [7, p. 331] points out that the segregation of the sounds is based on the "degree of perceptual difference" between them. He remarks that even a small perceptual difference "may be sufficient to improve performance" for tasks requiring some degree of sound segregation. They [7, p. 331] further suggests a larger perceptual differences could produce "obligatory stream segregation". If auditory icons are to be used in auditory displays as the communication mechanism it needs to be able to keep pace with the interactions occurring. This can be problematic if only sequential sounds are used as the interface may have to wait for a sound to finish before playing the next sound or may end up playing the sound for a just completed interaction, either of the situations will not provide any advantages to a user. This work investigates if object and action properties can be used to classify sounds prior to their use and if this is effective in clearly communicating auditory icons in the numbers necessary for auditory displays.

The previous work in earcons [3, 4, 5] found the limit of identification for earcons was most positively influenced by using staggered onsets between the earcons and designing multi-timbre earcons to enhance their identification or ability to 'stand out'. McGookin et al. [5] explored up to four concurrent earcons and found in many conditions only two of the four could be easily identified, however for certain mappings, three of the four earcons were identifiable. These results suggest at most three concurrent earcons can be used and opens the question about how many can be used for concurrent presentation if instead of earcons, auditory icons are used. The relative low number of concurrent earcons that can be uniquely identified is somewhat in contrast to the numbers of everyday sounds that we as listeners can identify throughout our daily activities. This was a motivating factor for conducting this exploration. The second motivation for this research was for designers of auditory displays, especially novice designers, who may make naïve psychological assumptions about what sounds were being listened to in their auditory display. Experienced designers often make educated guesses based upon their skill but even they may be wrong. These wrong choices can result in an auditory display communicating information in a manner that is demanding or confusing for listeners. The aim of this exploration was to present a method to select and validated auditory icons when used for concurrent presentation.

The sounds used in this exploration had been classified with particular focus on two categories, the object category of the sound and the action category of the sound. Fernström et al [8] have

previously described this classification. Previous research [8, 6] found that actions of sounds are better identified than the objects involved in a sound. The choice of these categories was also informed by soundscape classification studies [9], which found participants included both source (object) and actions in 100% of their descriptions of a soundscape and were the most salient items used by listeners to describe a soundscape. Research from Ballas and Howard [10] found the semantic context in sound interpretation is an important factor and listener's auditory perception is directed towards awareness of the sources of sounds i.e. the events producing sounds. They also stated the function of auditory perception is to recognise events rather than processing acoustic patterns. Events are defined for this study to consist of actions, objects and context. Our study explored action and object properties in the context of the identification accuracy of concurrently presented auditory icons.

## 2. EXPLORING THREE, SIX, AND NINE CONCURRENT AUDITORY ICONS

In a similar fashion to previous studies [2, 6], a listening test approach was used to study the recorded sounds used in the exploration. The sound were presented to participants using headphones, who responded in free-text format to what they thought each sound was. These text descriptions were often highly descriptive and accurate. In this particular study, the number of concurrent auditory icons being presented varied from three to six to nine in two conditions, one where the stimuli was limited to no overlapping action or object descriptors (constrained) and the second condition allowed random selection of stimuli from a pool with no conditions. This study used three stimuli pools where each pool consisting of 63 stimuli.

### 2.1. Participants

26 participants were recruited from the wider community of post-graduates and staff at the University of Limerick. All participants reported normal hearing and had normal or corrected to normal vision. Written consent was obtained prior to the study from all participants.

### 2.2. Stimuli

Three different sets (pools A,B, and C) of 63 high-quality monophonic sounds (44.1 Kilohertz 16-bit) everyday sounds (durations between 1.0 and 5.0 seconds) were used. These sounds were selected from the BBC sound effects CD collection [11], the online creative-commons Freesound [12], a local sound collection<sup>1</sup>, and a commercial sound effects CD collection [13]. The sound were chosen to represent complex, dynamic and informational events with different temporal patterns [14] and edited to a duration allowing for the "sound event" or "sound object" [15] to appear to occur naturally [16].

#### 2.2.0.1. Stimuli Classification

The sounds selected were classified with particular focus on two categories, the object category of the sound and the action category of the sound. The analysis concentrated on the properties of the events, the actions, the objects, and the context of the events. The set of descriptions for the sounds provided by the participants using their free text responses were also collected. Full descriptions of the sound are available online as described in Section 4.

<sup>1</sup>Everyday sound Collection - <http://www.idc.ul.ie/mikael/sounds/ecosound.zip>

### 2.3. Training

The focus of this training phase was to familiarise the participants with the concurrent presentation of auditory icons. A number of sets of concurrent auditory icons were presented to the listener using a standard media player. The stimuli used in the training phase were not used in the later tests. The participants spent five minutes listening to the stimuli and were then introduced to the platform used for the study. Users had headphones to listen to the sounds (in mono) while interacting with the system

### 2.4. Design of Study

Using a within-subjects design, the stimuli were presented in random order within each condition and the task order was counter-balanced for the conditions (three, six, and nine concurrent auditory icons). Each set of stimuli whilst presented randomly were presented as a single block for the particular condition. Stimuli were selected from two of the three pools randomly selected for each participant. One of the pools was used for the constrained sounds (non-overlapping action or object descriptors) and the second pool was used for the randomly selected sounds.

#### 2.4.1. Procedure

The participants listened to the recorded sounds (mono) in random order using headphones, responding in free-text format to what each sound was. The conditions varied from three to six to nine sounds being concurrently presented. Non-directive instructional guidelines [2] were used and participants were asked to describe the sounds in their own words. A 300 ms onset-to-onset gap [5] was used to improve identifiability. This start-to-start gap between two concurrent or almost concurrent sounds can prevent merging of the two sounds into a single stream.

#### 2.4.2. Results

The set of participant responses to the presented sounds were sorted and categorized, as well as evaluated for correctness. The sorting and categorisation was similar to Marcell's [17] equivalence judgement method and to Vanderveer's [2] heuristic for correct response method, both of which were used in classification of free text descriptors from listeners in a similar type of a task. The number of auditory icons correctly identified by participants was determined by analysis the participant's responses. This used a two stage classification process [8], in the first stage the criteria did not have to be kept consistent. In the second stage, the criteria were generalised and then grouped to reduce the number of criteria while maintaining consistency to the internal measures used. Examples of the second stage categories for objects include vehicle, metal, fabric/cloth and for actions include hitting, rubbing, opening/closing. Synonyms were related to the most relevant category or used to create a new category. This was done in a similar manner to previous researchers [?]. For each set of (three, six, or nine) concurrently presented auditory icons, the set of auditory icons presented and the set of participant responses to those auditory icons were compared. If the description of an auditory icon from a participant's response matched an auditory icon in the set presented, and if the auditory icon has not already been identified and matched with a previous description, the number of correctly identified auditory icon was increased by one, and the auditory icon description was marked as allocated.

The overall identification results are shown in Figure 1 and showed the sounds in conditions with prior classification had on average, a 9 – 10% identification accuracy performance improvement when compared to those with no prior classification. These

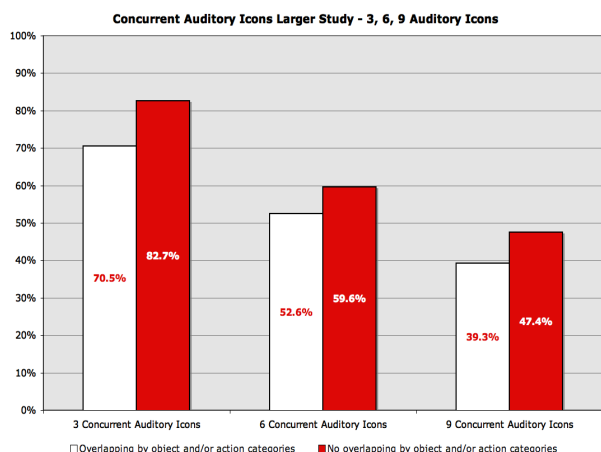


Figure 1: Graph showing the average proportion of auditory icons identified for the three, six, and nine auditory icon conditions.

results have shown in situations of nine concurrent auditory icons without classification the soundscape is becoming a cacophony as listeners can only identify 39% for the sounds. The prior classification condition with nine concurrent auditory icons showed a slightly better identification accuracy performance from listeners with an average identification rate of approximately 47%. The results for the identification accuracy with 9 (48%), 6 (60%), and 3 (83%) prior classified concurrent auditory icons indicates that this type of classification is effective technique for sound selection. These results indicate that designers should limit the number of concurrent auditory icons to range of between three and six concurrent auditory icons with prior classification. A possible explanation for the difference between the results and earlier studies may be due to the lower numbers of stimuli overlapping in the conditions as shown in Table 1. The random selection of stimuli from the three larger pools did not seek to increase the degree of overlap in either the action or object categories. Further large scale studies with highly overlapping action and object categories is necessary to clarify this confusion.

2.4.2.1. Average proportion of correctly identified auditory icons

Examining the results for correct identification with prior and with no prior classification shows auditory icons identification growing more difficult as the number presented is increased when using prior classification however this was not as severe as the difficulties encountered in the conditions where no classification of auditory icons was used. The distribution was not normally distributed and required the use of Kruskal Wallis and Dunn multiple comparison tests [18] to determine if any of the differences were statistically significant. The results matched with common sense finding with more sounds presented, each individual sound is more difficult to identify. The findings did not support the hypothesis that prior classification would improve identification.

2.4.2.2. Post study participant debriefing

As part of this study, a post study debriefing was used to collect data from the participants. Participants highlighted the particular difficulties they found in the more dense 6 and 9 auditory icon conditions. One participant stated that they were “starting to question their sanity” due to the odd combinations of sounds. This highlights the need to further consider the ecological combinations of auditory icons, in the case of the particular participant a number

Participant	Random Act 3	Random Obj 3	Random Act 6	Random Obj 6	Random Act 9	Random Obj 9	Average overlap
1	0%	0%	33%	33%	22%	44%	22%
2	0%	0%	33%	33%	89%	67%	37%
3	0%	67%	50%	67%	0%	67%	42%
4	0%	0%	50%	50%	44%	44%	31%
5	0%	67%	50%	67%	0%	67%	42%
6	0%	0%	67%	67%	22%	44%	33%
7	67%	0%	0%	33%	44%	56%	33%
8	67%	0%	0%	33%	22%	44%	28%
9	67%	0%	33%	0%	33%	67%	33%
10	67%	0%	83%	33%	44%	56%	47%
11	100%	0%	67%	33%	22%	67%	48%
12	67%	67%	67%	83%	89%	67%	73%
13	0%	67%	67%	50%	78%	44%	51%
14	0%	0%	0%	33%	44%	56%	22%
15	0%	0%	67%	67%	56%	78%	44%
16	0%	0%	67%	67%	22%	44%	33%
17	67%	0%	33%	33%	22%	44%	33%
18	67%	67%	67%	83%	44%	78%	68%
19	0%	67%	67%	33%	67%	44%	46%
20	67%	0%	0%	0%	22%	22%	19%
21	0%	67%	83%	67%	22%	44%	47%
22	67%	0%	0%	33%	33%	56%	31%
23	100%	0%	83%	67%	56%	78%	64%
24	0%	0%	83%	33%	44%	22%	31%
25	0%	0%	50%	33%	67%	78%	38%
26	100%	0%	83%	67%	56%	78%	64%
Average overlap	35%	18%	49%	46%	41%	56%	41%

Table 1: Percent of overlapping sound categories within the 3, 6, and 9 concurrent auditory presentation conditions in the third study.

Conditions	Mean Rank Difference	P-value
NonOverlap3 vs. NonOverlap6	-50.58	p<0.001
NonOverlap3 vs. NonOverlap9	-73.46	p<0.001
Over3 vs. Over6	-39.69	p<0.05
Over3 vs. Over9	-55.15	p<0.001

Table 2: Kruskal-Wallis and Dunn multiple comparison test results for the three concurrent auditory icons when compared to the six and nine auditory icons conditions

of unusual sounds were combined that would rarely if ever occur in the real world. The majority of participants felt similar to comments like “some had a lot going on but they weren’t really distinguishable, certain ones had a lot but couldn’t describe them all” or “much harder in some, did I hear it or not, was it in the set?”. Several felt that “I know there are way more sounds, I’m just not getting them!”. Participants did feel that upon hearing a “certain related sound, (they were) creating a script where each feed into the other”. This indicates how the participants find it easier to associate sounds to a particular sequence of events. In the case of conditions where unusual or non-natural sequences of sounds occurred, the difficulty of identification increased.

2.4.2.3. Future directions

Work in previous studies [6] have shown it can be useful to consider Ballas’s method of causal uncertainty [19] as a means of further refining the sound selections. This method could be used to determine the confused sounds. These sounds can be removed and replaced with less confused and more perceptually distinct sounds. The comments and free text responses from participants do not give detailed insight into how they associate sounds to a particular sequence of events. The repertory grid [20] is one method that can be used to provide a more detailed understanding of such mappings and categorisations.

### 3. CONCLUSIONS

In this exploration concurrent auditory icons were studied with and without prior classification. The results of this exploration show a smaller difference when compared to the identification results of the earlier two studies [6], however the number of participants (26) and the wider selection of stimuli from three distinct pools give this result a much firmer grounding and indicates the use of prior classification should be used to help in distinguishing auditory icons in concurrent presentation. The study does indicate the concurrent presentation of auditory icons may be equally damaging whether or not prior classification is used. A side note from this result is that the degree of overlap between either object or action categories in conditions with no prior classification may have contributed to improvement in identification results when compared to the earlier studies.

The results from this exploration are somewhat contradictory as it would indicate carefully designed auditory icons for use in concurrent presentations only provide a marginal identification accuracy improvement. It suggests any random set of everyday sounds are almost ( $\approx -9-10\%$ ) as effective as a selected set classified using object or action properties when used to communicating multiple messages. The selection of the stimuli in the random conditions had a lower degree of overlap than occurred in previous research [6]. This may have been a factor in the apparent success of the random select conditions when compared to those that had prior classification. These confusing results require further explorations where there are higher or near total overlap in the random condition to determine what effect this has on the identification of the sounds.

The work has demonstrated that it is possible to recognise and identify complex concurrent sounds. The approach used minimal training and single testing sessions and would fit into a 'worst case'. The fact these sounds proved effective would indicate the sounds would be even more usable in auditory displays where more training and longer use would occur.

### 4. ONLINE INTERNET RESOURCES

The participant log files, spreadsheets, scripts and non-commercial sounds used for this study are available for download online. Several of the sounds were taken from commercial CD's and information regarding these sounds is included but due to copyright issues, the actual sound files are not available.

### 5. ACKNOWLEDGEMENTS

The research is supported by Grant No. ILRP/202x from Enterprise Ireland under the Industry Lead Research Programme. The authors wish to thank Prof. Liam Bannon for his useful conversations and advice. We would like to express our thanks to all those who participated in this exploration.

### 6. REFERENCES

- [1] W. W. Gaver, "Auditory interfaces," in *Handbook of Human-Computer Interaction*, M. G. Helander, T. K. Landauer, and P. Prabhu, Eds., pp. 1003–1042. Elsevier Science, Amsterdam, The Netherlands, 1997.
- [2] N. J. Vanderveer, *Ecological Acoustics: Human Perception of Environmental Sounds*, Ph.D. thesis, University Of Cornell, 1979.
- [3] S. A. Brewster, *Providing a Structured Method for Integrating Non-Speech Audio into Human-Computer Interfaces*, Ph.D. thesis, University of York, 1994.
- [4] A. L. Papp III, *Presentation of Dynamically Overlapping Auditory Messages in User Interfaces*, Ph.D. thesis, University of California, 1997.
- [5] D. McGookin and S. A. Brewster, "Understanding concurrent earcons: Applying auditory scene analysis principles to concurrent earcon recognition," *ACM Transactions on Applied Perceptions*, vol. 1, no. 2, pp. 130–155, 2004.
- [6] E. Brazil and M. Fernström, "Investigating concurrent auditory icon recognition," in *Proceedings of ICAD 2006 - The 12th International Conference on Auditory Display*, London, 2006, Queen Mary.
- [7] B. C. J. Moore and G. Hedwig, "Factors influencing sequential stream segregation," *Acta Acustica united with Acustica*, vol. 88, no. 3, pp. 329–333, 2002.
- [8] M. Fernström, E. Brazil, and L. Bannon, "Hci design and interactive sonification for fingers and ears," *IEEE Multimedia*, vol. 12, no. 2, pp. 36–44, 2005.
- [9] I. McGregor, G. Leplâtre, A. Crerar, and D. Benyon, "Sound and soundscape classification: Establishing key auditory dimensions and their relative importance," in *ICAD2006*, 2006, pp. 105–112.
- [10] J. A. Ballas and J. H. Howard, "Interpreting the language of environmental sounds," *Environ. Behav.*, vol. 1, no. 9, pp. 91–114, 1987.
- [11] BBC, "Bbc sound effects library: sound effects library," 2006.
- [12] www.freesound.org, "Freesound creative commons: sound effects library," 2007.
- [13] Hollywood-Edge, "The hollywood edge [sound recording] : sound effects library," 1990.
- [14] J. J. Jenkins, "Acoustic information for objects, places, and events," in *Persistence and Change - Proc. of the First International Conference on Event Perception*, W. H. Jr. Warren and R. E. Shaw, Eds. 1985, Lawrence Erlbaum Associates.
- [15] R. F. Port, F. Cummins, and J. D. McAuley, *Mind as Motion: Exploration in the dynamics of cognition*, chapter Naive time, temporal patterns, and human audition, M.I.T. Press, 1995.
- [16] F. L. Wightman and R. Jenison, *Perception of Space and Motion*, chapter Auditory Spatial Layout, Academic Press, 1995.
- [17] M. M. Marcell, M. Malatanos, C. Leahy, and C. Comeaux, "Identifying, rating, and remembering environmental sound events," *Behaviour Research Methods*, vol. 39, no. 3, pp. 561–569, 2007.
- [18] M. J. Crawley, *Statistics: An Introduction Using R*, John Wiley And Sons, Chichester, 2005.
- [19] J. A. Ballas, M. J. Sliwinsky, and J. P. Harding, "Uncertainty and response time in identifying non-speech sounds," *Acoustical Society of America*, vol. 79, 1986.
- [20] E. Brazil and J. M. Fernström, "Investigating ambient auditory information systems," in *International Conference on Auditory Display (ICAD-07)*, Gary P. Scavone, Ed., Montreal, Canada, June 26-29 2007, pp. 326–333.