Auditory Emoticons: Iterative Design and Acoustic Characteristics of Emotional Auditory Icons and Earcons

Jason Sterkenburg[1], Myounghoon Jeon[1], and Christopher Plummer[2]

[1] Cognitive & Learning Sciences,
[2] Visual and Performing Arts
Michigan Technological University
Houghton, MI, USA
{jtsterke,mjeon,cplummer}@mtu.edu

Abstract. In recent decades there has been an increased interest in sonification research. Two commonly used sonification techniques, auditory icons and earcons, have been the subject of a lot of study. However, despite this there has been relatively little research investigating the relationship between these sonification techniques and emotions and affect. Additionally, despite their popularity, auditory icons and earcons are often treated separately and are rarely compared directly in studies. The current paper shows iterative design procedures to create emotional auditory icons and earcons. The ultimate goal of the study is to compare auditory icons and earcons in their ability to represent emotional states. The results show that there are some strong user preferences both within sonification categories and between sonfication categories. The implications and extensions of this work are discussed.

Keywords: auditory icons, earcons, auditory emoticons, non-speech sounds, sonification.

1 Introduction

Since the first International Conference on Auditory Display (ICAD) in 1992, research on sonification, the use non-speech sounds [1], has proliferated. As one of the simplest sonification techniques, auditory icons [2] (representative part of sounds of objects) and earcons [3] (ear + icons, short musical motives as symbolic representations of objects) have been successfully applied to electronic devices as auditory feedback for user activity [e.g., 4, 5]. Following those precursors, spearcons [6] (compressed speech) and spindex [7] (speech + index) have also shown improved performance and reduced workload with menu navigation tasks in diverse contexts. Fairly recently, musicons [16] (music + earcons) and lyricons [17] (lyrics + earcons) have also been introduced to enhance aesthetic aspects as well as functional mappings of the non-speech sound cues. However, despite successful improvement in performance measures, relatively little research has focused on emotional or affective aspects of those auditory cues. If any, research treated with either auditory icons [8] or earcons [9] only, but few studies compared affective effects of both auditory cues in a

M. Kurosu (Ed.): Human-Computer Interaction, Part II, HCII 2014, LNCS 8511, pp. 633–640, 2014.

single study [exception, 10]. The other research gap includes that affect research has depended merely on the simple valence dimension [positive – negative, e.g., 11]. Moreover, there was little research to identify the relationship between acoustic parameters of the sounds and diverse affective dimensions for a design guideline. To take a more systematic approach to affect-related auditory cue design research, the present paper describes iterative design processes of auditory emoticons (i.e., emotional auditory icons and earcons) and evaluation results of both auditory cues. Additionally, we provide an analysis of their acoustical characteristics for future design guidelines.

2 Iterative Design Processes

Sixteen college students, who major (or minor) in sound design or audio technology at Michigan Tech, created in total 640 auditory icons and earcons for 30 affective adjectives (calm, cold, comfortable, delicate, depressed, dreamy, surprising, fancy, free, fresh/cool, impressive, intimate, magnificent, modern, plain, pleasant, simple, soft, strong, warm, harsh, boring, confused, dark, dynamic, scared, uneasy, angry, disgusting, lively) based on multi-phase design panel discussions [12] under the two sound design experts' supervision. Affective adjectives were selected from previous research using the statistical reduction processes (factor analysis and multi-dimensional scaling) [13, 14] and a couple of adjectives were added to include basic six emotions [15]. After completing iterative design panel sessions (3 times) and removing acoustically similar sounds, we selected (112 auditory icons and 115 earcons) for further evaluations.

3 User Evaluation

3.1 Method

Thirty three undergraduate students were recruited using the online recruitment system (SONA) at Michigan Technological University. Auditory stimuli were presented via computer and headphones (Sennheiser HD 380 Pro headset). The auditory stimuli used fell into two categories: 1) auditory icons and 2) earcons. Each participant listened to several (2 – 7: $M = 3.73$ for auditory icons; $M = 3.83$ for earcons) sound clips from one of the categories. They could listen to the same sound repeatedly as much as they wanted. After listening, participants were asked to record which of the sound clips best conveyed a specific affective adjective (e.g., angry, fearful, etc.). In total, thirty adjectives were used. Upon completion of the task for one category (e.g., auditory icons), participants did the same for the other category (e.g., earcons). The order of affective adjectives, the order of category (auditory icons and earcons), and the order of sound clip presentation were randomized. Finally, participants were asked to decide between their favorite for each category which better conveyed the specific emotion.

3.2 Results

Table 1. Each row shows an affective adjective, a description of each auditory cue type and the percentage of participants who preferred the sound. * indicates p-values < 0.05.

Affective Adjective	Description of Preferred Auditory Icon	Percentage Preferred	Description of Preferred Earcon	Percentage Preferred
Angry	Traffic Jam	52%	Distorted percussive guitar chords	48%
Boring	Sigh	55%	Descending base (plucked)	45%
Calm	Breeze through trees and birds chirping	52%	Dreamy pad	48%
Cold	Wind and shivering	67%	Wind and descending piano notes	33%
Comfortable	Sigh of relief and creaking of chair as sinking in	61%	Woodwind chords	39%
Confused	Quizzical grunt	55%	Pitch bent tuning fork	45%
Dark	1) Thunder clap, 2) Distant ominous sound, 3) Owl hooting	58%	Ominous descending strings	42%
Delicate	Glass breaking	45%	High-pitched Oscillating piano notes	55%
Depressed (sad)	Dog whimpering	39%	Sad piano song	61%
Disgusting	Man Vomiting	64%	Descending deep synthesized tones	36%
Dreamy	Synthetic Pulsing	6% *	Whole tone scale	**94% ***
Dynamic	Crowd Cheering	39%	2 high pitched trumpet sounds	61%
Fancy	Spoon tapping Champaign glass	30% *	Baroque style harpsichord	70% *
Free	Wings flapping and bird chirping	64%	Synthesized choir and chime	36%
Fresh/cool	Water pouring into an ice-filled glass	**70% ***	Funk music baseline	30% *
Harsh	Grating metal	42%	Combination of high pitched keyboard notes	58%
Impressive	Amazed "woah"	55%	Trumpet fanfare	45%
Intimate	Girl pleased "ooh"	18% *	Aura (pad) and bass plus snare	**82% ***
Lively	Cheering and applauding crowd	**70% ***	Ascending synthetic violin with percussion	30% *
Magnificent	1)Trumpet fanfare, 2) Thunder clap	45%	Synthesized choir	55%
Modern	Typing, and cacophony of beeping	24% *	Fuzzy pad and staccato melody	**76% ***
Plain	Typing on keyboard	36%	Single flute note	64%
Pleasant (happy)	Child laughing	70%	3 ascending piano notes	30%

Table 1. (*Continued*)

Scared (fearful)	Woman blood curdling screaming	30% *	Tremolo string sound	**70% ***
Simple	Single tick of clock	48%	Xylophone	52%
Soft	Wobbly bell	42%	Descending piano (with reverb)	58%
Strong	Loud bang	42%	Synthetic bass drum	58%
Surprising	Man short gasp	52%	Ascending Fuzzy Keyboard	48%
Uneasy	Scraping fingernails on chalkboard	36%	Tremolo Keyboard	64%
Warm	Fire crackling	67%	Acoustic guitar chords	33%

Clear trends appeared in preference within sound categories. There were preferences shown for many affective adjectives, as determined by chi-square goodness of fit tests. Further, clear trends in categorical preference arose. To illustrate, there was a strong preference for auditory icon representation of words, such as cool (water poured into ice-filled glass) ($p = .024$), and happy (laughing child) ($p = .024$). Meanwhile, earcons were preferred to represent words, such as dreamy (whole tone scale) ($p < .001$), fancy (Baroque style harpsichord sound) ($p = .024$), intimate (pad, bass, & snare) ($p < .001$), and scared (tremolo string sound) ($p = .024$).

3.3 Discussion

Why are auditory icons preferred sometimes while earcons are preferred other times? To answer these questions it is necessary to review the history of auditory icons [2] and earcons [3]. The seminal works of Gaver and Blattner et al. established the defining characteristics of auditory icons and earcons, respectively. These characteristic features can be used to differentiate the two categories. A closer examination of these may be useful in identifying and understanding user preferences in various contexts.

Auditory Icons. Auditory Icons are often considered analogous to visual icons. Of course, the major difference is the sensory modality, with visual icons utilizing the visual system and auditory icons dependent upon auditory channels. But the analogy is important for understanding the key characteristics of auditory icons.

Similar to visual icons which contain features that human visual systems can detect in parallel (i.e., size, contour, color, etc.) auditory icons can be said to possess analogous features (pitch, tone, volume). Visual icons contain information which represents "real world" actions or events (e.g., an image of a camera represents a camera function in a software program). Auditory icons are the sounds which are paired with those same events (e.g., camera shutter sound on a phone indicating a picture was taken). Auditory icons then, are simply the "naturally-occurring" sounds coinciding with actions and events. This is in contrast to other forms of auditory stimulation like alarms, music, and earcons (to be introduced in the next section). In this view auditory icons can be thought of as the sounds which result from the interaction of real-world objects.

The nature of auditory icons makes them better suited for conveying different types of information. Auditory icons take advantage of the natural mode of listening which is “to identify the events that caused them” [2, p. 169]. Insofar as this the meaning of an auditory icon is universal so too can the effects of auditory icons be considered universal. That is, auditory icons represent interactions between objects in the environment (e.g., sound of a camera shutter, represents real-world actions occurring inside a phone) and as long as the sound of the camera shutter has meaning to the listener it has the potential to convey the same information to that listener. This fact highlights one of the potential benefits of auditory icons, namely, it can transcend language and cultural barriers.

Often auditory icons can easily convey a lot of information. This is because auditory icons map the sounds to their respective sources in a way that takes advantage of premade knowledge structures. In other words, mappings between auditory icons and their sources have already been learned, unlike synthetic mappings (e.g., earcons) which have not been learned.

Gaver [2] also suggests that the auditory icons should be useful for representing dimensional data. For each change in dimension (size, weight, speed, etc.) which is to be represented there is a corresponding change in the sound in the physical world. For example, if an icon is supposed to represent an increase in size of some value, then it can sound heavier (louder, deeper, longer lasting, etc.) This still takes advantages of the preexisting knowledge structures of the listener. This makes the mapping seamless and easy relative to other synthetic sounds and can reduce or eliminate the effort necessary to learn the mapping of the auditory icon to the function it represents.

Earcons. Earcons, in contrast to auditory icons, are not naturally occurring. Blattner et al. defines earcons as “nonverbal audio messages used in user-computer interface to provide information to the user about some computer object, operation, or interaction.” [3 p. 13]. In Blattner’s work she describes earcons in a more inclusive way than is intended in this paper. Earcons, here, are better described as synthetic nonverbal auditory messages. Earcons are called synthetic to mark a distinction between them and auditory icons which are either naturally occurring sounds or caricatures of naturally occurring sounds.

Earcons have specific advantages over other auditory categories, including auditory icons. First, because they are synthetic, earcons can be organized more easily. For instance, earcons can be simple sounds (motives) or they can be more complex, involving multiple layers of simple sounds. In this case, each layer can represent a different detail about the real world it is intended to represent. Earcons can be grouped together into families based on how many features they share. These relationships can be shown hierarchically.

Therefore, earcons have a generative syntax which allows participants learn the meaning of specific earcons without ever having heard them before. In many ways it can be considered similar to be a language of nonverbal sounds. However, learning the relationships between each of the motives and families and the syntax of the earcons can be arduous.

User Preferences. When participants' indicated the sound which they thought best captured each emotion, what factors influenced their decisions? Mapping is likely one of those factors. Sounds which do a poor job of representing the event or action which they were intended to describe should likely do poorly. Conversely, sounds with good mapping should do relatively well. What constitutes good mapping? What follows is an interpretation of what factors might be influencing user preferences in auditory emoticons.

It is plausible to suggest that auditory icons and earcons differ in their mapping ability. Further, these differences might influence user's perception of auditory emoticon effectiveness in representing an emotion. The ability of a sound to represent an emotion is dependent upon its connection to a mental representation of an emotion in the user's memory. Thus, in part, these user preferences can be considered a reflection of each user's past experience and memory. For example, the results show the word "scared" (violin tremolo) was best represented by an earcon. This is likely because scary things are often paired with a violin tremolo sounds in popular media and entertainment. Additionally, the affective state happiness is much more saliently linked to the sound of laughing (auditory icon) than any earcon, at least in the minds' of a significant majority of the participants in this study. The link between the auditory signal and a mental representation residing in memory determines how well each emoticon represents an emotion, and insomuch it informs their preferences. An auditory emoticon which is strongly linked to a mental representation of an affective state will be preferred to a weakly linked emoticon. So, user preferences are subject to the variation in the structure of the users' knowledge and memory. Of course, these preferences are also subject to influence from other variables, such as personality and external factors like stress, mood, and other forms of affect.

Additionally it appears that there could be a large cultural component, especially in the case of earcons. Many of the preferred earcons are similar to sonifications used in entertainment (i.e., movies, TV shows, news media). For example, the violin tremolo is often used to convey fear in movies, the harpsichord (affective adjective: fancy) is often used to represent aristocracy in movies. In fact, in this experiment, a case could be made that all earcons which were chosen are similar to those commonly used sonifications in entertainment.

A further interesting observation is apparent by looking at Ekman's basic emotion set [15] and close emotions to the set. A cursory analysis shows that preferences are almost evenly split between the auditory icons and the earcons. However, a closer look reveals that the earcons are more often chosen for the emotions which lie on the negative (or avoidance) dimension (e.g., sad, scared). Conversely, the auditory icons are chosen for the positive valence emotions (e.g., happy, lively). Even though auditory icons are more salient, if they remind users of unpleasant events or experiences, they could be avoided by users. In those cases an earcon might be preferable, as in the case of "scared" where participants preferred a tremolo violin sound to a woman screaming sound. This could be because of an avoidance of unpleasant sounds like woman screaming in the electronic products.

Consequently, it is expected that auditory emoticons which are strongly linked to mental representations of affective adjectives will outperform auditory emoticons which are weakly linked or unrelated in user's minds. This expectation does not predict that auditory icons or earcons will be better at representing any specific affect, but it does imply that the preferences (overall winners) will be the auditory emoticons which were most commonly linked to affective mental representations. Further, it has been speculated that there could be a relationship between the preference of earcons used in this study and the popularity of those sonifications in media and entertainment. Additionally, it was postulated that Ekman's basic emotions may be treated differentially, with preferences for positive valence emotions to be represented by auditory icons and negative emotions to be represented by earcons. This information suggests that auditory display designers should take into consideration the culture, experience, and memories of their target audience when creating affective auditory displays. Further research is required to investigate the complex nature of the relationship between auditory emoticons and user preferences in auditory displays.

4 Conclusion and Future Work

Emotional auditory cue sets were created and refined by iterative design processes and validated by user evaluation. We are collaborating with international researchers to replicate and extend this study to generalize its implications across different cultures. Moreover, we will construct affective dimensions of emotional auditory cues and compare them with affective dimensions of emotional visual cues to see the commonalities and differences between modalities. As a practical application of the auditory emoticons, we plan to test those cues in various contexts, ranging from mobile devices, telecommunication applications, to in-vehicle infotainment.

References

1. Kramer, G.: An introduction to auditory display. In: Kramer, G. (ed.) Auditory Display: Sonificaiton, Audification, and Auditory Interfaces. Addison-Wesley, MA (1994)
2. Gaver, W.W.: Auditory icons: Using sound in computer interfaces. Human-Computer Interaction 2, 167–177 (1986)
3. Blattner, M.M., Sumikawa, D.A., Greenberg, R.M.: Earcons and icons: Their structure and common design principles. Human-Computer Interaction 4, 11–44 (1989)
4. Gaver, W.W.: The SonicFinder, a prototype interface that uses auditory icons. Human-Computer Interaction 4, 67–94 (1989)
5. Brewster, S.A.: The design of sonically-enhanced widgets. Interacting with Computers 11(2), 211–235 (1998)
6. Walker, B.N., Lindsay, J., Nance, A., Nakano, Y., Palladino, D.K., Dingler, T., Jeon, M.: Spearcons (speech-based earcons) improve navigation performance in advanced auditory menus. Human Factors, Online First Version (2012)
7. Jeon, M., Walker, N.B.: Spindex (Speech Index) improves acceptance and performance in auditory menu navigation for visually impaired and sighted users. ACM Transactions on Accessible Computing 3(3), 10:11-26 (2011)
8. Schleicher, R., Sundaram, S., Seebode, J.: Assessing audio clips on affective and semantic level to improve general applicability. In: Proceedings of the DAGA (2010)

9. Lemmens, P.M.C., De Haan, A., van Galen, G.P., Meulenbroek, R.G.J.: Emotionally charged earcons reveal affective congruency effects. Ergonomics 50(12), 2017–2025 (2007)
10. Larsson, P., Opperud, A., Fredriksson, K., Västfjäll, D.: Emotional and behavioural response to auditory icons and earcons in driver-vehicle interfaces. In: Proceedings of the 21st International Technical Conference on the Enhanced Safety of Vehicles, Stuttgart, Germany, June 15-18 (2009)
11. Lemmens, P.M.C.: Using the major and minor mode to create affectively-charged earcons. In: Proceedings of the 7th International Conference on Auditory Display, Limerick, Ireland (2005)
12. Pirhonen, A., Tuuri, K., Mustonen, M.-S., Murphy, E.: Beyond clicks and beeps: In pursuit of an effective sound design methodology. In: Oakley, I., Brewster, S. (eds.) HAID 2007. LNCS, vol. 4813, pp. 133–144. Springer, Heidelberg (2007)
13. Lee, J.-H., Jeon, M., Han, K.H.: The analysis of sound attributes on sensibility dimensions. In: Proceedings of the 18th International Congress on Acoustics (ICA 2004) , vol. II, Kyoto, Japan (April 2004)
14. Jeon, M., Lee, J.-H., Kim, Y.E., Han, K.H.: Analysis of musical features and affective words for affection-based music search system. In: Proceedings of the 2004 Korean Conference on Cognitive Science (KCCS 2004), Seoul, Korea (June 2004)
15. Ekman, P.: An argument for basic emotions. Cognition and Emotion 6, 169–200 (1992)
16. McGee-Lennon, M., Wolters, M.K., McLachlan, R., Brewster, S.: Hall. C.: Name that tune: Musicons as reminders in the home. In: Proceedings of the SIGCHI Conference on Human Factors in Computing System, Vancouver, BC, Canada (2011)
17. Jeon, M.: Lyricons (Lyrics + Earcons): Designing a new auditory cue combining speech and sounds. In: Stephanidis, C. (ed.) HCII 2013, Posters, Part I. CCIS, vol. 373, pp. 342–346. Springer, Heidelberg (2013)