

Lyricon (Lyrics + Earcons) Improves Identification of Auditory Cues

Yuanjing Sun¹ and Myounghoon Jeon^{1,2(✉)}

¹ Mind Music Machine Lab, Cognitive & Learning Sciences,
Michigan Technological University, Houghton, MI, USA
{ysun4, mjeon}@mtu.edu

² Mind Music Machine Lab, Computer Science,
Michigan Technological University, Houghton, MI, USA

Abstract. Auditory researchers have developed various non-speech cues in designing auditory user interfaces. A preliminary study of “lyricons” (lyrics + earcons [1]) has provided a novel approach to devising auditory cues in electronic products, by combining the concurrent two layers of musical speech and earcons (short musical motives). An experiment on sound-function meaning mapping was conducted between earcons and lyricons. It demonstrated that lyricons significantly more enhanced the relevance between the sound and the meaning compared to earcons. Further analyses on error type and confusion matrix show that lyricons showed a higher identification rate and a shorter mapping time than earcons. Factors affecting auditory cue identification and application directions of lyricons are discussed.

Keywords: Auditory display · Auditory icons · Auditory user interface · Cognitive mapping · Earcons · Lyricons · Sonification

1 Introduction

Auditory researchers have developed various non-speech (e.g., auditory icons [2], earcons) and tweaked speech cues (e.g., spearcons [3], spindex [4]) in user interfaces. Auditory icons [1] use part of analogic sounds of the object or item. The sound can be thought of as the result from the interaction of real-world. Therefore, they are commonly used as feedback of an operation to enhance the realistic feeling of a virtual interface.

Earcons can represent more abstract operations or processes in user interfaces, by using well-structured musical motives. However, their indirect link to the referent has some limitations [e.g., 5] and requires users’ learning. Stevens and her colleagues [5] claimed that users might not recognize up to 40 % of the earcons when there are more than seven earcons at the same time. However, with the application of some musical principles in designing earcons, people could recall up to 25 distinct and uninitiated earcons.

While speech is clearer than other auditory cues, speech might be more intrusive and less aesthetic. The most common application of speech, Telephone-based interface

(TBI) reveals three limits when speech is used alone. First, speech is not only slow and serial, but also makes users difficult to retrieve and scroll the target item in TBI. Also, it is hard for speech to concurrently represent both the content itself and the structure or the hierarchy of content (e.g., menu or function) [6]. Third, speech requires high quality of signal and an undistributed background as well. Researchers have also tried to tweak speech in designing auditory cues. Spearcons and spindex cues have shown successful cases in auditory menus, but each of them requires a specific context (e.g., spearcon: multi-dimensional menu, spindex: one-dimensional menu) [4] or optimal applications. Auditory cognition pathway [e.g., 7] illustrates speech interpretation procedure on a phonological level. Echoic memory is a type of the auditory short-term memory (STM), which functions when auditory stimuli are received by the ear until the lexical selection occurs. Spearcons and spindex cues can function as non-speech cues even before the lexical selection occurs.

From this background, “lyricons” (lyrics + earcons) [8] have provided a novel approach to combining the two layers of musical speech sounds (lyrics) and non-speech sounds (earcons) concurrently. This combination is expected to improve both semantics and aesthetics of auditory user interfaces. Such redundant displays might enhance user’s recognition and interpretation of auditory cues, while improving the learnability for first-time users to operate auditory user interfaces.

The present study aims to: (1) briefly present the results of focus groups conducted to obtain users’ attitude towards their awareness of auditory user interfaces and comments on the initial design of lyricons, (2) validate the effectiveness of lyricons compared to traditional earcons, and (3) introduce a framework to evaluate the recognition and identification of auditory cues.

2 Initial Design of Lyricons and Focus Group

2.1 Initial Design of Lyricons and Earcons

An experienced sound designer (> 15 years) created nine lyricons for nine basic functions of home appliances (Table 1). Earcon design follows literature and industry standards [9]. Lyrics came from previous research [8]. Ballas [10] once proposed four key factors affecting sound identification in the sensory transduction process, which include acoustic properties (e.g., intonation and stress), ecological frequency (how often the signal occurs in the environment), causal uncertainty (whether the signal is easily confusing with other signals), and sound typicality (how typical the signal is of a particular source). This framework lays the foundation of our design and evaluation method of lyricons. Based on that, the sound designer intended to instill a hierarchical implication in speech cues by systematically manipulating the sound variables. In this way, the number of musical notes, the range of frequency, and the total duration of the sound represent the ecological frequency [10] of the function, that is, “how often the function is used in daily life.” For example, POWER-ON /OFF stays on the top level of the function hierarchy because it would only occur once at the very beginning or end of the use.

Table 1. Function names and denotation as well as corresponding musical lyrics components.

No.	Function name and denotation	Music notes	Lyric
20	Power On (boot up system)	C4G4E4G4C5	DaDiDaDiDang
16	Power Off (turn off system)	C5G4E4G4C4	DiDaDiDaDing
5	Function On (start laundry/wind)	C4D4E4	Func/Tion/On
2	Function Off (turn off or pause heat, wind)	E4D4C4	Func/Tion/Off
11	Magnitudeup (speed up, each step min to max)	G4A4	Up
12	Magnitude down (slow down, each step max to min)	A4G4	Down
1	Cancel (no, reject)	G5G5	Back
21	Touch (yes, entrance)	G5	Tink
22	Unavailable (Unavailable area, cannot work)	G1	Bang

In sum, the use of ecological frequency made the function hierarchy, with POWER-ON /OFF on top, FUNCTION ON/OFF below that, followed by MAGNITUDE CHANGE and CANCEL/TOUCH/UNAVAILABLE on the bottom. Such a combination is expected to enhance user’s recognition and interpretation of auditory cues, while improving the learnability for first-time users to operate auditory interface.

2.2 Focus Group

To obtain users’ comments on our initial lyricon designs, we conducted focus groups with twelve undergraduate students (mean age = 23, female = 5). They provided general comments on the issues of auditory user interfaces (e.g., annoyance, controllability, indexibility, emotional mapping, etc.) and recommendations about the next lyricon design (e.g., serial combination of speech and sound, using more than one instruments, etc.).

None of our participants has hearing impairments or professional music background. After a consent form procedure and introduction to the study, participants (3–5 in one session) discussed with a moderator their personal experience of the use of auditory user interfaces in electronic devices, and their advantages and disadvantages. Then, the moderator showed initial lyricon designs and participants provided comments on lyricons.

A majority of participants emphasized that auditory cues should convey a straightforward meaning, which is not necessarily the case in reality. They stated that they can easily fall into confusion when the meaning of the sound is uncertain, “Sometimes, I heard the sound but still don’t know which part goes wrong, especially when I am driving. It’s really annoying because neither can I stop the sound nor can I understand what the problem is.” In addition to functional interpretation, some participants were likely to associate auditory cues with their memory or affect in their daily lives.

Sound could evoke emotions and the related context. Recall of memory could help to intuitively grasp the meaning of the designated function. For example, participant G mentioned, “I like the sound from vacuum when I just wake up. It links my memory

with my mom.” As long as the sound from products was used as a trigger of behavioral shift or attentional shift, participants allowed for an appropriate level of interference, “I like the prompt tone of SKYPE when someone is talking to me. I think it is OK for me if it’s not too loud to be a noise.” However, simultaneously, they wanted to have control over the auditory cue. Once they lost control of it, they tended to regard it as a noise. Some participants favored speech sounds, “I like natural voice to tell me what’s wrong with my car,” and “It will be even better if the oven can talk to me. I mean I like to pretend all equipment at home is a human,” which supports the application of lyricons.

Participant L provided recommendations of the next lyricon designs, “To a new user, it will be better to have the speech part first and then, the sound, so he or she knows the specific function of the sound clearly. After a while, they can choose to skip the speech, but keep using the sound. If more instruments in different ranges were used, it would be easy to distinguish from each other.”

3 Sound-Function Mapping Experiment

3.1 Method

We expect that lyricons outperform earcons, but a parallel combination of speech and earcons in lyricons might confuse users more [9]. Therefore, we conducted an empirical sound-function mapping experiment to compare the identification accuracy between lyricons and earcons.

Thirty-three undergraduate students (mean age = 21, 10 female) participated in the auditory cue-function mapping experiment. None of them has participated in the previous focus group session. They were randomly allocated to two groups: lyricon group or earcon group. After a consent form procedure, participants conducted a sound card sorting task [8]. Nine function-index cards were placed on the desk. Each card contained a definition and specific examples of the function. The sound stimuli consisted of nine lyricons and nine earcons (same as those used in lyricons). Participants listened to sound stimuli generated from a SONY sr16 computer using a Sennheiser HD380 pro headphone.

First, an experimenter explained the meaning of each function to participants. Then, participants paired each sound stimulus with the function that they believe the sound best represents. Participants were allowed to have as much time as they wanted to complete the sorting task.

3.2 Results

The results showed that the average of accuracy rate of the lyricon group (82.35 %) was almost double than that of the earcon group (46.53 %). An independent samples t-test showed a significant difference between the lyricon and the earcon groups, $t(30.4) = 3.60$, $p < 0.001$. Moreover, the sorting time of the lyricons ($M = 5.26$ min, $SD = 1.17$) was much shorter than that of the earcons ($M = 6.34$ min, $SD = 3.60$). We

Table 2. Confusion matrix of lyricon mapping results.

Intended answer \	Power-on	Power-off	Function-on	Function-off	Magnitude up	Magnitude down	Cancel	Touch	Unavailable
Power-on	94.12%		5.88%						5.88%
Power-off		94.12%		5.88%	5.88%				
Function-on	5.88%		76.47%		5.88%	5.88%		17.65%	
Function-off		5.88%	11.76%	82.35%					
Magnitude up				11.76%	76.47%	5.88%			
Magnitude down						88.23%			11.76%
Cancel							88.23%	11.76%	
Touch			5.88%		11.76%		5.88%	58.82%	
Unavailable							5.88%	11.76%	82.35%

Table 3. Confusion matrix of earcon mapping results.

Intended answer \	Power-on	Power-off	Function-on	Function-off	Magnitude up	Magnitude down	Cancel	Touch	Unavailable
Power-on	56.25%	12.50%	12.50%		18.75%				
Power-off	6.25%	50.00%		18.75%		18.75%	6.25%		
Function-on	31.25%	18.75%	25.00%		18.75%	6.25%			
Function-off		18.75%	6.25%	56.25%	6.25%	12.50%	6.25%		
Magnitude up	6.25%		25.00%	6.25%	31.25%	12.50%	12.50%		
Magnitude down			18.75%	12.50%	12.50%	25.00%	12.50%	6.25%	12.50%
Cancel			6.25%	6.25%	6.25%	12.50%	31.25%	31.25%	
Touch			6.25%		6.25%	6.25%	18.75%	62.50%	
Unavailable						6.25%	12.50%		81.26%

also plotted the confusion matrix to identify which functions confused the participants most (Tables 2 and 3).

3.3 Error Type Analysis

For further analysis, we divided mapping errors into three types: hierarchy error, tone polarity error, and random error. Hierarchy error means that participants can successfully recognize the pair, but mistakenly assign them in the wrong function hierarchy (e.g. Assigned MAGNITUDE UP/DOWN stimuli in FUNCTION ON/OFF labels). Tone polarity error means the right assignment to function hierarchy but put polarity upside down (Put MAGNITUDE UP in MAGNITUDE DOWN labels). The remaining errors are attributed to random error (see Fig. 1). In a diagram, the horizontal

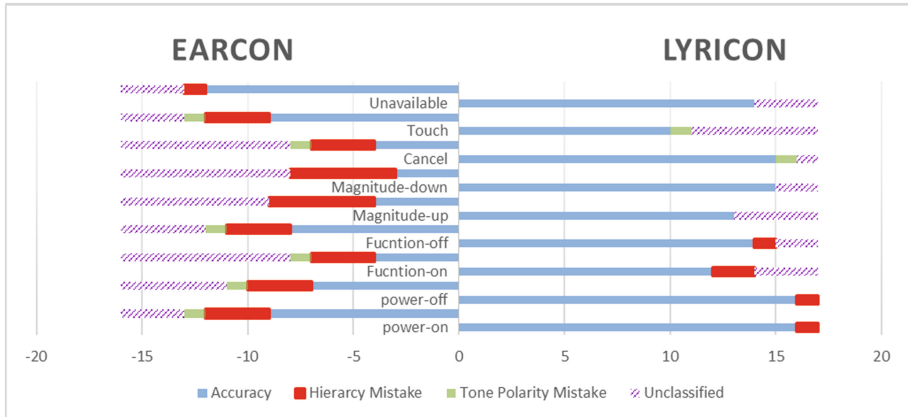


Fig. 1. Stacked bar of error type distribution diagram. Left side is the earcon group and right side is the lyricon group.

axis is the number of errors. The vertical axis represents nine functions. There are four patterns in legend to represent four conditions: blue is correct; thicker red is hierarchy error, green is tone polarity error and purple light diagonal pattern is random error. The diagram of error distribution reveals that the lyricon group made fewer hierarchy errors than the earcon group. Selective lyrics seemed to strengthen the signal-referent relationship, and eliminate the causal uncertainty in mapping.

4 Discussion

We also plotted the confusion matrix to identify which functions confused the participants most as showed in Table 2 and 3.

4.1 Confusion Matrix Analysis

When it comes to the phonological level process again, our participants in the earcon group seem to remember nine earcons only in the echoic memory before the lexical selection occurs. Lacking of direct connection with meaning might lead to low accuracy as well as longer time in the mapping task. Such processing might also require higher mental resources, and thereby, decreased the recognition performance. Without the help of lyrics, participants seem to have difficulty in identifying the relations between the sounds and the functions, and among nine earcons.

As mentioned, causal uncertainty refers to whether the signal is easily confusing with other signals [10]. It is an important calibration in measuring the confusion among a group of sound stimuli. In particular, “UNAVAILABLE” had the highest accuracy rates in both lyricon and earcon groups probably because its unique low pitch was distinctive from other stimuli. In contrast, the other single-tone function, “TOUCH”

showed the lowest accuracy in the lyricon group, which was even lower than that of the earcon group.

Similarity in acoustic profile of stimuli reduced the recognition of a specific signal. We found that in the earcon group the most confusing factor came from the function, “CANCEL” because the auditory stimulus had the same pitch as the function, “TOUCH” but with double notes. This slight difference was hard for participants to capture because none of our participants had professional musical training. It implies that consecutive notes in the same pitch may give rise to causal uncertainty. General users’ just-noticeable-difference threshold should be considered in design in order to prevent users from being confused with other signals.

In the lyricon group, “TOUCH” was mostly confusing with “FUNCTION-ON”, probably because both of them have a positive meaning. This case gives a particular example to show the importance of the lyric selection in connecting a signal with a referent. In the lyricon group, a sound designer used a word, “BACK” as lyric in the lyricon, “CANCEL”. It is an appropriate choice because “BACK” expresses a negative meaning, which is also popular in yes/no/cancel dialog boxes. It provides the context that successfully links the echoic memory with semantic meaning of the function. In contrast, the designer used “TINK” as lyric in the lyricon, “TOUCH” instead of other popular words in dialog boxes (for example: “Yes”, “OK”, “Enter”, etc.) to express a positive meaning. The ambiguous word weakened the link in the lexical-semantic stage, and thus, misled participants to the other lexical candidate, “FUNCTION-ON”.

5 Conclusion and Future Work

By designing lyricons, we attempted to integrate speech and non-speech cues to overcome existing problems in auditory user interface design. Our empirical experiment on sound-event mapping demonstrated that lyricons could enhance the relevance between the sound and the meaning compared to earcons. The lyricon group showed a higher identification rate and a shorter mapping time than the earcon group, which is promising in terms of lyricon applications in auditory user interfaces. Based on this experiment, we confirmed key factors [10] affecting sound cue identification – distinguishability, ecological frequency, lyric choice, etc. In lyricons, the lyric part can improve identification of the function, while the earcon part can imply the hierarchical structure of the functions.

In the current study, we used only the piano sound for the experimental purpose. More acoustic properties and musical parameters, such as timbre, register, tempo, rhythm, have already been included in auditory display design in industry to iteratively enhance lyricons’ aesthetic quality. We plan to analyze phonetic patterns of each functional speech [e.g., speech-to-song illusion, 11] and to reflect them on the innate acoustic profiles of the earcon part in order to find an optimal combination of speech and musical clips. Such an endeavor may enhance participants’ perception and interpretation of the message conveyed by lyricons. In a practical application, once users get familiar with lyricons, they could customize the earcon-only part without the lyric part. Based on this design and evaluation effort, researchers and practitioners could create more effective and efficient auditory interactions between a user and a system.

References

1. Blattner, M.M., Sumikawa, D.A., Greenberg, R.M.: Earcons and icons: their structure and common design principles. *Hum. Comput. Interact.* **4**, 11–44 (1989)
2. Gaver, W.W.: Auditory icons: using sound in computer interfaces. *Hum. Comput. Interact.* **2**, 167–177 (1986)
3. 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. *Hum. Factors* **55**(1), 157–182 (2012)
4. Jeon, M., Walker, B.N.: Spindex (speech index) improves auditory menu acceptance and navigation performance. *ACM Trans. Accessible Comput.* **3**(3), 10:11–26 (2011)
5. Stevens, C., Brennan, D., Parker, S.: Simultaneous manipulation of parameters of auditory icons to convey direction, size, and distance: effects on recognition and interpretation. In: *Proceedings of the International Conference on Auditory Display (ICAD2004)*, Sydney, Australia (2014)
6. Brewster, S.A.: Using nonspeech sounds to provide navigation cues. *ACM Trans. Comput. Hum. Interact. (TOCHI)* **5**, 224–259 (1998)
7. Baldwin, C.L.: *Auditory Cognition and Human Performance: Research and Applications*. CRC Press, Boca Raton (2012)
8. Jeon, M.: Lyricons (lyrics + earcons): Designing a new auditory cue combining speech and sounds. In: Stephanidis, C. (ed.) *HCI 2013, Part I. CCIS*, vol. 373, pp. 342–346. Springer, Heidelberg (2013)
9. Jeon, M., Lee, J.-H.: The ecological AUI (auditory user interface) design and evaluation of user acceptance for various tasks on smartphones. In: Kurosu, M. (ed.) *HCI/HCI 2013, Part IV. LNCS*, vol. 8007, pp. 49–58. Springer, Heidelberg (2013)
10. Ballas, J.A.: Common factors in the identification of an assortment of brief everyday sounds. *J. Exp. Psychol. Hum. Percept. Perform.* **19**(2), 250–267 (1993)
11. Deutsch, D., Henthorn, T., Lapidis, R.: Illusory transformation from speech to song. *J. Acoust. Soci. Am.* **129**(4), 2245–2252 (2011)