PERCEPTUAL INTERPOLATION AND OPEN-ENDED EXPLORATION OF AUDITORY ICONS AND EARCONS

Adam Csapo

MTA SZTAKI
Cognitive Informatics Laboratory
Kende u. 13-17, Budapest, Hungary
csapo.adam@gmail.com

Peter Baranyi

MTA SZTAKI
Cognitive Informatics Laboratory
Kende u. 13-17, Budapest, Hungary
baranyi@sztaki.hu

ABSTRACT

The design of cognitive artifacts for systematic exploration of sound spaces poses a significant challenge in auditory display research. This is especially the case if the artifact is to be based on the user's individual perceptual capabilities rather than a predefined perceptual model or a purely numerical relationship between the various sounds in the sound space. In this paper, we propose a tensor algebraic tuning model that allows for the systematic tuning and interpolation of parametric auditory icons and earcons in a way that satisfies this challenge. The proposed approach views earcons as being composed of any number of auditory icons at a lower level, thus allowing for the systematic modification of both auditory icons and earcons via local and global interactions. Due to the granular representation used by the model, it becomes possible to approximate the user's perceptual sensitivity locally in the parametric space used to generate the auditory icons and earcons.

1. INTRODUCTION

Cognitive artifacts were first defined by Norman as "artificial devices that maintain, display, or operate upon information in order to serve a representational function and that affect human cognitive performance" [1]. Such devices are crucial in any system which is to engage human users in interaction because they may be used to enhance the effectiveness of interaction by altering the task at hand [2].

Many examples can be cited of cognitive artifacts which are used every day to select specific items within large perceptual spaces, such as the various color spaces for color selection (e.g. RGB, HSL), speech-enabled GPS systems for finding information on maps, or Gmail search to recall the dates of past events. Cognitive artifacts also exist for the design of auditory displays (e.g. Barrass's Information Sound Space [3], Bezzi, dePoli and Rocchesso's Sound Authoring Tool [4], Hermann's Model-Based Sonification [5], and Larsson's EarconSampler and SoundMoulder [6]).

Despite the existence of cognitive artifacts for the design of auditory displays, there are several difficulties when it comes to using them in realistic applications. One difficulty is that unlike colors, sounds cannot be described by a multi-dimensional space that defines a unique position for every possible sound. The second difficulty stems directly from the first one, namely that due to a lack of categorization principle for sounds, there is no evident way to interpolate between them and still preserve a sense of perceptual continuity. Still a third difficulty is that perceptual continuity is a concept that may be hard to define unequivocally across all users (thus, if an earcon carries abstract or emotional information – as is the case with e.g. different kinds of warning sounds used in critical situations [6] or emotional earcons [7] – different users may attribute different meaning to the same earcon). For these reasons, any cognitive artifact used for the design of auditory displays must limit the possible scope of sounds and at the same time provide the user with rich opportunities for exploration if it aspires to provide a useful model of perceptual relationships.

In this paper, we propose a systematic model for the interpolation and open-ended exploration of auditory icons and earcons. The model rests on the single assumption that the sound space of auditory icons and earcons is parametric (i.e., it assumes that all of the sounds are created using a generative model). Apart from this condition, the model is open-ended (there is no limit to the number of sounds that can be explored and their location within the parameter space). Further, the model enables users to interpolate between sounds based on their own perception. Instead of using a set of generic perceptual rules to achieve this, the model facilitates perception-based exploration of sounds by providing a systematically configurable interface that helps users interactively discover the directions in the high-dimensional parameter space which represent perceptual continuity. In some respects this is similar to the explorative approach used in interactive evolutionary computation [8].

The paper is structured as follows. In section 2, the generative model which serves as a basis for the the proposed cognitive artifact, and its relationship with auditory icons and earcons is described. In section 3, a tuning model is described which will allow for the tuning of auditory icons and earcons. The proposed interpolation model, described in section 4 is directly based on this tuning model. Section 5 describes our implementation of the model. In section 6, an application example is given in which earcons are used to represent 10 gradations of percepts along several different tactile dimensions. Conclusions are given in section 7.

2. GENERATIVE MODEL FOR EARCONS

In this section, we describe the generative model we use for the creation of auditory icons, and how they are related to earcons. The purpose of this brief discussion will be to show that although general agreement holds that auditory icons and earcons are completely different in nature, a viable – and natural – alternative to this view would be to regard earcons as being composed of auditory icons. In this way, by providing a model for interpolation of earcons, the interpolation of auditory icons emerges as a special case in which the earcon is composed of a single auditory icon. Conversely, if there is a useful method to interpolate auditory icons, then the interpolation of earcons can focus on the interpolation of those aspects of earcons which are not embodied by auditory icons.

2.1. Relationship between auditory icons and earcons

Auditory icons were first proposed by Gaver [9, 10] as natural, everyday sounds that can be used to represent actions or objects within an interface. Auditory icons are natural sounds that occur as byproducts of physical processes in real-world environments. In contrast, earcons - developed by Blattner, Sumikawa and Greenberg [11] - were defined as non-verbal audio messages that do not necessarily create an intuitive set of associations with the objects or events that they represent. Earcons are therefore abstract sequences of sounds which are usually temporal in nature.

Although the overwhelming majority of publications describe auditory icons and earcons as contradictory and mutually exclusive categories [11, 12, 13, 14], the notion that earcons are built up of auditory icons is not entirely new (a compositional relationship is briefly mentioned in [15]). From our point of view – when describing a set of generic sounds for use in generic applications of auditory display – it makes sense to regard earcons as messages constructed using auditory icons, because this view only broadens the scope of earcons instead of restricting it. More specifically, earcons can be composed of any kinds of building blocks – or motives, as Blattner refers to them – (e.g., sinusoidal notes, violin notes, flute notes, sounds of different kinds of doors being slammed, etc.) since every one of these building blocks can become auditory icons if heard often within the same context in a natural environment. Thus, for our purposes, we regard auditory icons as perceptually unique entities which can be juxtaposed and / or overlapped through time to create earcons.

2.2. Compact representation for families of auditory icons

Blattner et al defined earcon families as larger groupings of earcons in which all earcons are constructed using related motives [11]. By analogy, we may define families of auditory icons as sets of auditory icons which are related (i.e., can be compared to each other) in a perceptual sense. Our model assumes that such a family of auditory icons is available to the user, and that the auditory icons are generated based on a set of generation parameters contained in vector **f** (if such a family is not available to the user, then one may be created by generating a relatively large set of auditory icons at random and asking the user to select an appropriate subset of those icons and order them perceptually, as in interactive evolutionary computation [8]). The family of auditory icons will serve as gridpoints between which interpolation can be performed. The larger the family of auditory icons, the greater the possibility that the user can perform truly local manipulations in the perceptual space.

In the general case, the family of auditory icons can be multidimensional, such that the auditory icons are ordered along multiple perceptual dimensions (e.g., an auditory icon may be loud and rough, or soft and contain no quality of auditory roughness at the same time). Thus, the parameters used to generate the auditory icons can be considered to be a function F of values along different perceptual dimensions (we refer to function F as a parameter-generating function). If there are N perceptual dimensions, H different generation parameters for each auditory icon, P_n unique

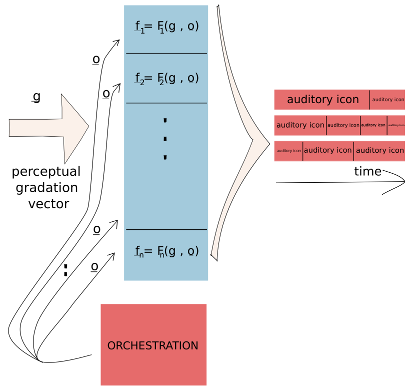


Figure 1: Structure of an earcon. The perceptual parameters denoted by vector **g** and a set of temporal parameters are used to generate the auditory icons of which the earcon is comprised. Although the temporal parameters are conceptually generated by a separate *orchestration module*, they are considered to be part of the input to the parameter generation functions for auditory icon parameters. In this way, the problem of interpolating between earcons is reduced to the problem of interpolating between pairs of auditory icons.

gradations along each dimension (e.g., there are P_1 degrees of loudness and P_2 degrees of roughness), and the grid of possible input value combinations is denoted by $G = \{g_{p_1,p_2,\ldots,p_N} \in \mathbb{N}^{[a_1,b_1]\times[a_2,b_2],\ldots,\times[a_N,b_N]}\}_{p_n=1}^{P_n}$ $(n = 1..N)$, then the function can be represented by an $(N+1)$-dimensional tensor $\mathcal{F}$ of size $P_1 \times \ldots \times P_N \times H$ such that

$$\mathcal{F}^{D(G)}_{p_1,..,p_N,h} = \mathbf{f}_{g_{p_1,p_2,\ldots,p_N}}(h) \tag{1}$$

where $\mathbf{f}_{g_{p_1,p_2,\ldots,p_N}}$ is the parameter vector used to generate the auditory icon corresponding to a given point on grid G.

2.3. Structure of earcons in terms of auditory icons

As described in section 2.1, for our purposes we view earcons as being comprised of one or more auditory icons. The advantage of this conceptual approach is that an earcon can be considered as a set of auditory icons along with temporal information which determines the starting and ending time of each auditory icon. Hence, if we can perceptually interpolate between pairs of auditory icons, then we may just as easily interpolate between pairs of earcons if we interpolate between the auditory icons they contain and the corresponding temporal information separately.

The problem of interpolating earcons can be transformed entirely into the problem of interpolating auditory icons if the necessary temporal information is considered to be part of the perceptual vector that serves as input to parameter-generating function F. Fig. 1 demonstrates the structure of earcons in terms of auditory icons.

3. TUNING MODEL FOR PARAMETRIC AUDITORY ICONS

Based on the previous section, every parametric earcon is defined by a set of parametric auditory icons if the parameters used to generate the auditory icons include the starting and ending times between which each auditory icon is played. Further, we have established that the parameters for the auditory icons can be stored in an $(N+1)$-dimensional tensor, which we denote by $\mathcal{F}^{D(G)}$. Using higher-order singular value decomposition (HOSVD), it is possible to decompose the tensor in the following form [16]:

$$\mathcal{F}^{D(G)} = \mathcal{S} \overset{N+1}{\underset{n=1}{\boxtimes}} \mathbf{X}_n \tag{2}$$

where $\boxtimes$ refers to the tensor product operation defined in [16], and:

1. $\mathbf{X}_n = (\mathbf{x}_1^{(n)}, ..., \mathbf{x}_{I_n}^{(n)})$, $n = 1..N$ is a unitary matrix of size $(P_n \times I_n)$
2. $\mathbf{X}_{N+1} = (\mathbf{x}_1^{(N+1)}, ..., \mathbf{x}_{I_{N+1}}^{(N+1)})$ is a unitary matrix of size $(H \times I_{N+1})$
3. $\mathcal{S}$ is a real tensor of size $I_1 \times ... \times I_N \times I_{N+1}$, the subtensors $\mathcal{S}_{i_n=\alpha}$ of which have the following properties:
   - all-orthogonality: any pair of the subtensors of $\mathcal{S}$ are orthogonal, i.e. for all possible values of n, α and β subject to $\alpha \neq \beta$:

   $$< \mathcal{S}_{i_n} = \alpha, \mathcal{S}_{i_n} = \beta >= 0 \tag{3}$$

   - ordering: All of the subtensors of $\mathcal{S}$ along any given dimension n are ordered according to their Frobenius norm, i.e. $\forall n = 1..N+1$:

   $$||\mathcal{S}_{i_n=1}|| \geq ||\mathcal{S}_{i_n=2}|| \geq ... \geq ||\mathcal{S}_{i_n=I_n}|| \geq 0 \tag{4}$$

The values $I_1..I_{N+1}$ are the ranks of the system along each of the input dimensions. Tensor $\mathcal{S}$ is referred to as the core tensor, and matrices $\mathbf{X}_i$ are referred to as weighting matrices. If we modify the values in just the p_k-th row of any $\mathbf{X}_k$ in Eq. (2) (that is, the p_k-th value of $\mathbf{x}_i^{(k)}$, where $i = 1..I_k$), then only those output values of function $\mathcal{F}^{D(G)}$ will be changed which belong to the p_k-th perceptual gradation along the k-th dimension of hyper-rectangular grid G. This can be easily seen if we express a single element of $\mathcal{F}^{D(G)}$ as follows:

$$\mathcal{F}^{D(G)}(g_{p_1,...,p_N}) = \mathcal{S} \overset{N+1}{\underset{\substack{n=1,\\ n\neq k}}{\boxtimes}} \mathbf{x}_{n,p_n} \times_k \mathbf{x}_{k,p_k} \tag{5}$$

where the p_k-th row of matrix $\mathbf{X}_k$ is denoted by $\mathbf{x}_{k,p_k}$. It is obvious that if any point on hyper-rectangular grid G is chosen in which the value of the k-th dimension is not the p_k-th discretization point, then the output value of the function will be unchanged. For this reason, we refer to the manipulation of vector $\mathbf{x}_{k,p_k}$ as the *local tuning of the auditory icon along the* k-th *input dimension*. The values in vector $\mathbf{x}_{k,p_k}$ in turn are referred to as *tuning weights*.

4. INTERPOLATION OF PARAMETRIC AUDITORY ICONS

The local tuning model provided in Eq. 5 operates on a minimal-rank orthogonal system within the parameter space of auditory icons. The minimal-rank tuning weights could in theory be used for interpolation between pairs of auditory icons, but the cognitive load experienced by the user in keeping track of the perceptual effects of each of the weights may be significant depending on the rank of the system. We refer to the ease with which the user can keep track of these perceptual effects as the *interpretability* of the tuning model. Using rank reduction techniques it would be possible to increase the interpretability of the tuning model, but in exchange the system would suffer loss of complexity (i.e., only a subspace of the original parameter space would be available for tuning using a rank-reduced system of tuning weights).

It is important to relax this tradeoff between interpretability and complexity, because the user can only concentrate on the perceptual effects of changes made to tuning weights if the number of these weights is small, but it is equally important that the user be able to explore the parameter space of $\mathbb{R}^H$ to the fullest extent possible. To this end, we propose an approach that combines rank reduction and the adaptive modification of the basis vectors in the rank-reduced space so that the full-rank space can be explored in a restricted, but systematic way. Such an approach can be achieved through the following steps:

1. Assuming that the user would like to tune the system in the k-th dimension, the first step is to reduce its rank from I_k to 1. This can be achieved using Higher-Order Orthogonal Iteration (HOOI) – a method which is proven to yield optimal rank-reduction [17] – to obtain:

   $$\underset{\hat{\mathcal{S}}, \{\hat{\mathbf{X}}_n\}_{n=1}^{N+1}}{\operatorname{argmin}} \{(\mathcal{F}^{D(G)} - \hat{\mathcal{F}}^{D(G)}) \tag{6}$$

   where

   $$\hat{\mathcal{F}}^{D(G)} = \hat{\mathcal{S}} \overset{N+1}{\underset{n=1}{\boxtimes}} \hat{\mathbf{X}}_n \tag{7}$$

   such that $\hat{I}_k = 1, \hat{I}_n = I_n, \forall n \neq k$, and $\hat{I}_k$ is the rank along the k-th dimension of the new system.
2. Having obtained this approximation, our goal is to increase the rank of the vector space which can be controlled by the single tuning parameter which remains in the k-th dimension. In order to achieve this, we expand weighting matrices $\hat{\mathbf{X}}_n, n = 1..(N+1)$, as well as the core tensor, $\hat{\mathcal{S}}$ so that:

   $$\mathcal{F}^{D(G)} = \tilde{\mathcal{S}} \overset{N+1}{\underset{n=1}{\boxtimes}} \tilde{\mathbf{X}}_n \tag{8}$$

   where each $\tilde{\mathbf{X}}_n, n = 1..(N+1)$ has a rank of P_n (the number of discretizaton points along the given dimension, which is equivalent to saying that there are at least P_n columns in the n-th matrix), and $\tilde{\mathcal{S}}$ is augmented appropriately. After this step, the single tuning parameters (i.e., in the first column of $\tilde{\mathbf{X}}_k$) can still only be used to control a single dimension within the system, but the space that can be reconstructed by the system as a whole is $\mathbb{R}^H$ once again.

3. By systematically modifying certain elements in the new core tensor, $\tilde{\mathcal{S}}$, as the first column of $\tilde{\mathbf{X}}_k$ is modified by the user, it becomes possible for the user to traverse a subspace $\mathcal{V} \subseteq \mathbb{R}^H$ such that $\mathcal{V}$ is also H-dimensional.

In order to demonstrate how steps 2 and 3 can be achieved (the fact that the original function can be reconstructed in Eq. 8, and that if certain elements of $\tilde{\mathcal{S}}$ are systematically modified, then a single tuning weight can be used to control an H-dimensional subspace of $\mathbb{R}^H$), we provide the following lemma and theorem.

Lemma 1. *Let us consider the HOSVD of a discretized multivariate function* $\mathcal{F}^{D(G)} = \mathcal{S} \boxtimes_{n=1}^{N} \mathbf{X}_n$*, and let* $\hat{\mathcal{F}}^{D(G)} = \hat{\mathcal{S}} \boxtimes_{n=1}^{N} \hat{\mathbf{X}}_n$ *denote the rank-reduced instance of the same function obtained using HOOI. The original tensor,* $\mathcal{F}^{(D(G))}$ *can be reconstructed if:*

1. *each* $\hat{\mathbf{X}}_n, n = 1..(N+1)$ *is augmented so that* $rank(\tilde{\mathbf{X}}_n) = P_n$
2. *each dimension of* $\hat{\mathcal{S}}$ *is augmented so that the length of the* n*-th dimension is the same as the number of columns in* $\tilde{\mathbf{X}}_n$

Proof. It is trivial that if the lemma holds true, then the new core tensor, $\tilde{\mathcal{S}}$ can be expressed from Eq. 8 as:

$$\tilde{\mathcal{S}} = \mathcal{F}^{D(G)} \boxtimes_{n=1}^{N+1} \tilde{\mathbf{X}}_n^{+} \tag{9}$$

The pseudoinverse of the augmented weighting matrices always exists, and because $rank(\tilde{\mathbf{X}}_n) = P_n$, and the rank along the n-th dimension of the original tensor, $rank(\mathcal{F}^{D(G)})_{(n)} = I_n \leq P_n$, the original tensor can be reconstructed.

□

Theorem 1. *The weights in the first column of* $\tilde{\mathbf{X}}_k$ *(denoted by* $w_i, i = 1..P_k$*) define a hyperline in the* H*-dimensional space* $\mathbb{R}^H$*:*

$$HL = \{\tilde{\mathcal{S}} \boxtimes_{\substack{n=1 \\ n \neq k}}^{N} \tilde{\mathbf{x}}_{n,p_n} \times_k w_{p_k} \times_{N+1} \tilde{\mathbf{X}}_{N+1} | w_{p_k} \in \mathbb{R}\} \tag{10}$$

The projection of this hyperline onto output dimensions $q_1, q_2, \leq H, q_1 \neq q_2$ *has a slope which depends only on the values in the first subtensor of* $\tilde{\mathcal{S}}$ *along the* k*-th dimension,* $\tilde{\mathcal{B}} = \tilde{\mathcal{S}}^{p_k=1}$*:*

$$slope = \frac{(\tilde{\mathcal{B}} \boxtimes_{\substack{n=1 \\ n \neq k}}^{N} \tilde{\mathbf{x}}_{n,p_n}) \times_{N+1} \tilde{\mathbf{x}}_{(N+1),q_2}}{(\tilde{\mathcal{B}} \boxtimes_{\substack{n=1 \\ n \neq k}}^{N} \tilde{\mathbf{x}}_{n,p_n}) \times_{N+1} \tilde{\mathbf{x}}_{(N+1),q_1}} \tag{11}$$

Proof. Expanding equation 8, we obtain:

$$\begin{aligned} \mathcal{F}^{D(G)}(g_{p_1,\ldots,p_N}) = \\ w_{p_k}((\tilde{\mathcal{B}} \boxtimes_{\substack{n=1 \\ n \neq k}}^{N} \tilde{\mathbf{x}}_{n,p_n}) \times_{N+1} \tilde{\mathbf{X}}_{N+1}) + \\ \mathbf{r}_{p_k}((\tilde{\mathcal{B}}^c \boxtimes_{\substack{n=1 \\ n \neq k}}^{N} \tilde{\mathbf{x}}_{n,p_n}) \times_{N+1} \tilde{\mathbf{X}}_{N+1}) \end{aligned} \tag{12}$$

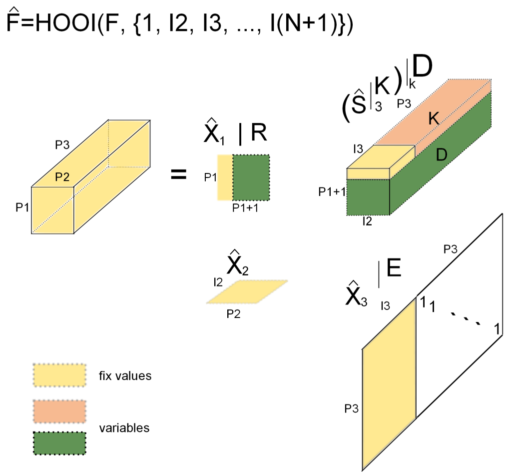


Figure 2: Proposed augmentation method used to compensate for reduced interpretability after the tuning model is rank-reduced, assuming that $k = 1$ and $N = 2$. The light (yellow) shades contain fix values, and the dark (mauve and green) shades contain variables which can be chosen so as to compensate for the rank-reduction as well as to alter the slope of the hyperline that is traversed within $\mathbb{R}^H$ during tuning.

where the complement of $\tilde{\mathcal{B}}$, $\tilde{\mathcal{B}}^c$ contains all subtensors of $\tilde{\mathcal{S}}$ except the first along the k-th dimension, $\tilde{\mathbf{X}}_k = (\mathbf{w}|\mathbf{R})$ ($\mathbf{w}$ is the first column of $\tilde{\mathbf{X}}_k$), w_{p_k} is the p_k-th element of $\mathbf{w}$, and $\mathbf{r}_{p_k}$ is the p_k-th row of $\mathbf{R}$.

If we consider just a single weight, w_{p_k}, the right-hand side of the equation is a vector of length H, and each element of this vector is broken up into a sum of two values: the first of these is a variable that is scaled by the weight and is also dependent on the values of $\tilde{\mathcal{B}}$, while the second can be regarded as a constant. Hence, the first term of the sum represents the slope of an H-dimensional hyperline with respect to w_{p_k}. The projection of the hyperline onto the output coordinates defined by dimension q_1, q_2 has the slope stated in the theorem. □

5. IMPLEMENTATION OF THE TUNING MODEL

Our implementation of step 2 in the previous section is shown in Fig. 2. In our implementation, only two weighting matrices are augmented (and accordingly, the core tensor, $\hat{\mathcal{S}}$ is only augmented along two of its dimensions). These two dimensions are the k-th and the $(N+1)$-th dimensions. Augmenting just the k-th dimension would have been sufficient due to the fact that rank reduction was only performed in this dimension, but augmenting the $(N+1)$-th weighting matrix with an identity matrix is useful for controlling the slope of the hyperline that is traversed when the tuning weights are manipulated. More specifically, when calculating the i-th parameter in parameter vector $\mathbf{f}$, the core tensor is only multiplied by the i-th row of $\mathbf{X}_{N+1}$ along its $(N+1)$-th dimension, and due to the fact that this weighting matrix is augmented by an identity matrix, only a single and unique subtensor of $\mathcal{K}$ will affect the slope of the hyperline.

We can see that after augmenting $\hat{\mathcal{S}}$ in dimensions other than

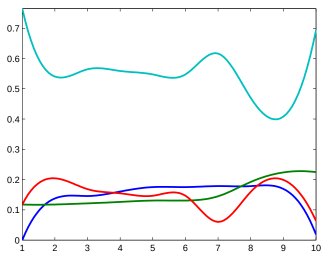


Figure 3: Weighting functions contained in $\mathbf{X}_1$ for the dimension of roughness. The horizontal axis contains 10 gradations along the first input dimension (roughness).

k, we augment the number of values that are scaled by each w_{p_k}. If a certain part of these values is systematically and periodically altered (namely, the values of $\mathcal{K}$), the slope of the hyperline controlled by the tuning model can be periodically changed. By changing these values back and forth, between two extremes, the path controlled by the a single tuning weight will resemble a hyper-spiral that extends within a hyper-cone in $\mathbb{R}^H$.

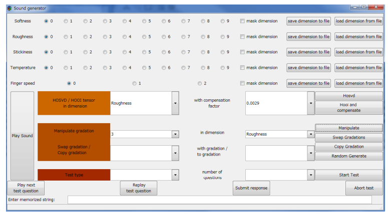
6. APPLICATION EXAMPLE: SURFACE SOUNDS

Our goal in one of our applications was to design earcons in order to represent tactile sensations in virtual reality. The application can be useful because tactile feedback devices are more expensive and can also be more cumbersome to wear and / or carry than audio feedback equipment.

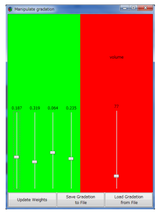
According to psychophysicists, the four most salient features of tactile perception are softness (compliance), roughness, stickiness and temperature [18, 19]. In our model, we represent each of these dimensions along 10 distinct gradations, which theoretically yields a combination of 10^4 different auditory icons that are used to construct earcons. In reality, the number of auditory icons used is less, because the dimension of stickiness is transferred to the temporal domain. Our configuration of auditory icons takes the following form:

- An auditory icon for softness is played for a certain time x. This is equivalent to saying that the user is pressing down on the surface to see how soft it is.
- After time x, the sound for roughness is played instead of the sound for softness. This is equivalent to saying that the user is dragging his/her finger along the surface in order to perceive how rough it is.
- Time x depends on how sticky the surface is. The interactive rationale behind this is that the stickier the surface, the longer the time it will take for the user to be able to drag his/her finger across the surface and perceive its roughness.
- The sound for temperature can be heard throughout, because it is only a function of contact, and as long as the user's finger is touching the surface, its temperature can be heard.

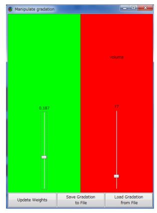
From the orchestration (i.e., the temporal relationship between the auditory icons), we can deduce that the sounds for softness and roughness depend not only on the gradation of softness and roughness, but also depend on the stickiness of the surface (this will have an effect on the ending time of the sound for softness and the starting time of the sound for roughness). In addition, we also view the duration of the complete sound as an input parameter. Hence, the tensor representations of the parameter-generating functions will be 4, 4, and 3-dimensional for softness, roughness and temperature, respectively.



(a) Main window



(b) Tuning weights after HOSVD



(c) Tuning weights after HOOI

Figure 4: Interface developed for creating earcons representing tactile percepts.

By way of example, we describe the parameter tuning model for the dimension of roughness. The original tensor representation of F is denoted by $\mathcal{F}^{(10\times10\times3\times5)}$, because the parameter vector for the synthesis algorithm has 5 elements (the first three dimensions stand for degrees of roughness, softness, and total length of the earcon). The weighting functions along the first input dimension (for *degree of roughness*) are shown in Fig. 3. The figure shows 4 different weighting functions, which is equivalent to saying that $\mathbf{X}_1$ for roughness contains 4 columns. Using HOOI, it is possible to reduce the rank of the whole system, and obtain just a single column in $\hat{\mathbf{X}}_1$. Thus we obtain $\hat{\mathcal{S}}^{(1\times2\times2\times5)}$, instead of the original $\mathcal{S}^{(4\times2\times2\times5)}$.

According to the proposed interpolation method, we re-augment matrices $\hat{\mathbf{X}}_1$ and $\hat{\mathbf{X}}_4$ to obtain $\tilde{\mathbf{X}}_1^{(10\times11)}$ and $\tilde{\mathbf{X}}_4^{(5\times10)}$, as well as the core tensor to obtain $\tilde{\mathcal{S}}^{(11\times10\times3\times10)}$. In the new core tensor, the values for which the first index is 1 and the last index is between 6 and 10 are free parameters. The choice of these parameters affects the slope of the hyperline that can be explored by modifying the values in the first column of $\tilde{\mathbf{X}}_1$. Hence, a structured subset of the original parameter space can be explored by using just these weights and modifying the free parameters in the core tensor systematically, between two extremes.

In our implementation, the absolute value of the two extreme values in each position of tensor $\mathcal{K}$ is the same, but this absolute value is chosen randomly per position in $\mathcal{K}$. The speed at which the distance between the two extreme values is traversed (with respect to changes in the tuning weight w_{p_k}) can be tuned by the user. Hence, the user has two parameters to modify during the open-ended tuning process: the tuning weight itself (vector $\mathbf{w}$) and the sensitivity of the slope of the traversed hyperline with respect to vector $\mathbf{w}$.

The interface developed for the system can be seen in Fig. 4. Subfigure (a) shows the main window of the application, while subfigures (b) and (c) show the tuning weights for the fourth degree of roughness in the original tuning model and the rank-reduced and interpretability-optimized case. After rank reduction and compensation, only a single tuning weight remains for each auditory icon. The sensitivity of the the basis functions with respect to the tuning weight can be specified in the main windo. A set of pre-defined parameters (in this case, the volume of the earcon) can be tuned separately and directly so that the generation of unpleasant earcons can be evaded.

7. CONCLUSION

In this paper, we proposed a systematic model for the interpolation and open-ended exploration of auditory icons and earcons. The model was based on the assumptions that the sound space of auditory icons and earcons is parametric, and that earcons can be regarded as being composed of a set of auditory icons and temporal instructions which dictate the temporal relationships between the auditory icons. The proposed model does not impose any limitation on the number of sounds that can be explored and their location within the parameter space. Further, the model enables users to interpolate between sounds based on their own perception. Instead of using a set of generic perceptual rules to achieve this, the model facilitiates perception-based exploration of sounds by providing a systematically configurable interface that helps users interactively discover the directions in the high-dimensional parameter space which represent perceptual continuity.

Through an example application, we briefly demonstrated that a compositional relationship between auditory icons and earcons does not contradict real-world experience, and that the proposed model for exploration and interpolation can be implemented in engineering systems. Although statistically rigorous test results have not yet been provided, preliminary feedback from users has been positive.

8. ACKNOWLEDGMENT

This research was supported by the ETOCOM project (TÁMOP-4.2.2-08/1/KMR-2008-0007) through the Hungarian National Development Agency.

9. REFERENCES

[1] D. Norman, *Cognitive Artifacts*. New York, NY, USA: Cambridge University Press, 1991, pp. 17–38. [Online]. Available: http://portal.acm.org/citation.cfm?id=120352.120354

[2] P. Barnard, *Bridging between basic theories and the artifacts of human-computer interaction*. Cambridge University Press, 1991, pp. 103–127.

[3] S. Barrass, "Auditory information design," Ph.D. dissertation, Australian National University, 1997.

[4] M. Bezzi, G. de Poli, and D. Rocchesso, "Sound authoring tools for future multimedia systems," in *Multimedia Computing and Systems, 1999. IEEE International Conference on*, vol. 2, 1999, pp. 512 –517.

[5] T. Hermann, "Sonification for exploratory data analysis," Ph.D. dissertation, Bielefeld University, Bielefeld, 2002.

[6] P. Larsson, "Tools for designing emotional auditory driver-vehicle interfaces," in *Auditory Display*, ser. Lecture Notes in Computer Science, S. Ystad, M. Aramaki, R. Kronland-Martinet, and K. Jensen, Eds. Springer Berlin / Heidelberg, 2010, vol. 5954, pp. 1–11.

[7] G. Wersenyi, "Auditory representations of a graphical user interface for a better human-computer interaction," in *Auditory Display*, ser. Lecture Notes in Computer Science, S. Ystad, M. Aramaki, R. Kronland-Martinet, and K. Jensen, Eds. Springer Berlin / Heidelberg, 2010, vol. 5954, pp. 80–102.

[8] H. Takagi, "Interactive evolutionary computation: fusion of the capabilities of ec optimization and human evaluation," *Proceedings of the IEEE*, vol. 89, no. 9, pp. 1275 –1296, 2001.

[9] W. Gaver, "Auditory icons: Using sound in computer interfaces," *Human Computer Interaction*, vol. 2, no. 2, pp. 167–177, 1986.

[10] ——, "The sonicfinder: An interface that uses auditory icons," *Human Computer Interaction*, vol. 4, no. 1, pp. 67–94, 1989.

[11] M. Blattner, D. Sumikawa, and R. Greenberg, "Earcons and icons: Their structure and common design principles," *Human Computer Interaction*, vol. 4, no. 1, pp. 11–44, 1989.

[12] A. Pirhonen, "Non-speech sounds as elements of a use scenario: a semiotic perspective," in *Proceedings of the 12th International Conference on Auditory Display (ICAD2006*, 2006.

[13] P. Lemmens, M. Bussemakers, and A. De Haan, "Effects of auditory icons and earcons on visual categorization: the bigger picture," in *In Proceedings of the International Conference on Auditory Display (ICAD '01)*, 2001, pp. 117–125.

[14] M. Wogalter, *Handbook of Warnings*. Lawrence Erlbaum Associates, 2006.

[15] T. Hermann and H. Ritter, "Listen to your data: Model-based sonification for data analysis," in *Advances in intelligent computing and multimedia systems*, G. E. Lasker, Ed., Int. Inst. for Advanced Studies in System research and cybernetics. Int. Inst. for Advanced Studies in System research and cybernetics, 8 1999, p. 189–194.

[16] L. D. Lathauwer, B. D. Moor, and J. Vandewalle, "A multi linear singular value decomposition," SIAM *Journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1253–1278, 2000.

[17] M. Ishteva, L. Lathauwer, P. Absil, and S. van Huffel, "Dimensionality reduction for higher-order tensors: Algorithms and applications," *International Journal of Pure and Applied Mathematics*, vol. 42, no. 3, pp. 337–343, 2008.

[18] T. Yoshioka, S. Bensmaia, J. Craig, and S. Hsiao, "Texture perception through direct and indirect touch: an analysis of perceptual space for tactile textures in two modes of exploration," *Somatosensory and Motor Research*, vol. 24, no. 1–2, pp. 53–70, 2007.

[19] M. Hollins and S. Risner, "Evidence for the duplex theory of tactile texture perception," *Perceptual Psychophysics*, vol. 62, pp. 695–705, 2000.