Sonification of directional and emotional content: Description of design challenges

Antti Pirhonen

Dpt. of Comp. Sc. & IS
P.O. Box 35, FI-40014 Univ. of Jyväskylä, Finland
pianta@jyu.fi

Henni Palomäki

SYSOPENDIGIA Plc / Telecommunications
Hämeentie 135 A, FI-00560 Helsinki, Finland
henni.palomaki@sysopendigia.com

ABSTRACT

In the construction of sound objects into an application, the designer's skills to communicate through sounds is the cornerstone of the activity. In such an expertise, the knowledge about the human way of interpreting different properties of sounds is essential.

This paper is a description of two experiments, in which the semantics of tempo change, pitch change and intensity change of sound has been studied by asking the participants of the experiments to combine sounds to visual images. The images in the first experiment were photos which had been validated in terms of their emotional content. In the second experiment, the images were arrows pointing in various directions.

The results show that studying context independent semantics of non-speech sounds with the help of photos is problematic, but some tendencies can be revealed. On the other hand, simple information units like physical directions, can be illustrated with changes in intensity and tempo, but especially with change in pitch.

1. INTRODUCTION

By definition, sonification is the conveying of information through non-speech audio [1]. To sonify information thus denotes the use of the means of non-speech audio to express certain meanings. How the properties of non-speech sounds reflect meanings, is a largely unexplored field. Agreed, there exists abundant research on the semantics of sounds, especially concerning warnings [e.g., 2, 3], but nevertheless what we know is just a fraction of what we do not know. The more we learn about the semantics of sounds, the more we have the means to design. This study is a contribution to the body of literature on the semantics of sounds, and thus has the aim of expanding the sound designer's toolbox.

In a recent study [4], the participants in an experiment were asked to link simple iconic drawings with a selection of sounds. The test session consisted of 140 very small tasks, in which the participants were given two drawings at a time to be combined with a given sound. The experimental setting was found to be problematic, because it was noticed that the subjects simplified the meanings during the test session along something like a positive-negative dimension. Thus, there was ultimately no major difference whether the pair of drawings showed a skull and a baby, or an arrow pointing down and an arrow pointing up. Both of these picture pairs were interpreted to illustrate the same dimension.

In the current study, we decided once again to use a pictorial stimulus to be combined with sounds, but photos were used instead of simple drawings. This was because we hoped to prevent the participants from over-simplifying the content of the images. In order to draw conclusions about how participants combine pictures with the sounds, we used a set of photos which had been validated across three emotional dimensions.

In another study, we used highly symbolic images, namely arrows pointing in different directions. The objective of this experiment was to find out how the selected dimensions of non-speech sounds could be mapped on to physical dimensions.

The two experiments reported here shared only one common feature: they both used the same set of sounds. Each sound was a sequence of 25 short units, with a total duration of 3 seconds. Between each unit of a sound, there was a pause of the same duration as that of the previous part, thus making different units distinct from each other. As a timbre we used General Midi sound #74 (recorder).

The sounds used in the experiments differed from each other in terms of tempo change, pitch change and intensity change. Concerning tempo, each sound was either accelerating or decelerating continuously. The second property to be varied was pitch – each sequence of sound units was continuously increasing or decreasing between C#4 and C#6, so the scale was 2 octaves. The third variant was intensity with each sound continuously either increasing or decreasing in intensity. These three dimensions resulted in 8 different variations of the sound (Table 1):

	Tempo change		Pitch change		Intensity change	
	up	down	up	down	up	down
1		x		x		x
2		x		x	x	
3		x	x			x
4		x	x		x	
5	x			x		x
6	x			x	x	
7	x		x			x
8	x		x		x	

Table 1. Properties of different sounds

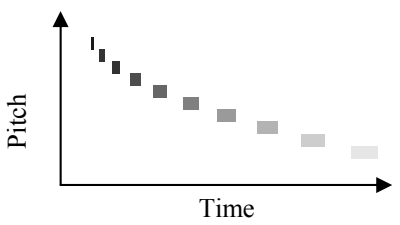

Figure 1. Illustration of sound #1

From now on, these eight different sounds are illustrated with drawings, as in Figure 1. The example in Figure 1 illustrates sound 1 (decelerating, pitch going down, intensity decreasing). Tempo is indicated by the length of the individual units, pitch

change is shown on the y-axis and intensity through black-grey shading. The illustration is simplified from the actual piano-roll view by reducing the number of units from 25 to 10, because the purpose of the illustration is only to provide a rough idea of the direction of change in each property.

1		[s0.wav]
2		[s10.wav]
3		[s20.wav]
4		[s30.wav]
5		[s40.wav]
6		[s50.wav]
7		[s60.wav]
8		[s70.wav]

Table 2. Illustration of each sound sample, with links to audio files

We now report the two different experiments separately, referring to the numbering of different sounds as indicated in Tables 1 and 2.

2. EXPERIMENT 1: TRACING THE SEMANTICS OF SOUNDS WITH THE HELP OF PHOTOS

In the first experiment, the objective was to chart the relationships between the variables of the sounds (Table 1) and the emotional content of a set of photos. The photos were selected from the International Affective Picture System (IAPS) collection. IAPS, devised by the Center of the Study of Emotion and Attention in the University of Florida, consists of more than 900 photos which have been empirically validated across three dimensions: Valence (extremes happy – sad), arousal (excited – sleepy), and dominance (in control – dominated). We had available all the validation data. From the vast number of photos we chose nine which all presented extreme cases on some dimension or dimensions. However, we excluded the most violent ones as well as those having strong sexual references to overcome ethical problems in our empirical study, and substituted those photos with photos which had the second highest scores on each dimension. In addition, we changed two photos for cultural reasons: Validation has been carried out in the U.S., and we only used Finnish participants in our study. In a small pilot study it turned out that the participants did not understand what a photo of a revolver pointing towards them presented – someone asked whether it was some musical instrument. That particular photo was replaced by a photo of a snake attacking. Another photo, which had high scores in valence and arousal, was about a ski jumping competition. It showed a skier who had just started his descent from the tower. However, the photo appeared to be quite old, in that the equipment, style and the clothing made it quite comical in the eyes of Finns, who are familiar with ski-jumping. So we found it more appropriate to replace it with a photo of a sky diving group.

The final selection of photos is presented in Figure 2. In the experiment, the participant took a seat in front of a table, on which the prints of the photos were taped in the order they are in Figure 2. Behind the table there were two loudspeakers, which were connected to a computer sound card. Each of the sounds was played in turn, in randomised order. After having heard a sound, the subject was supposed to choose which one of the photos had the best match with the sound heard. In addition, the participant was asked to estimate on a scale of 1-4 how sure she or he was about the choice. This value was supposed to work as a weighting factor when making a synthesis of the results. The participant was encouraged to think aloud, giving reasons for each choice. In the laboratory there were, in addition to a participant, two researchers; one controlling a computer and the other giving instructions and making notes.

The total number of participants was 22 (13 males, 10 females), all university students. 10 of them had some level of formal education in music (in addition to the music education given in the Finnish comprehensive school). Each session took about 10 minutes and the participants received a movie ticket as a reward.

Figure 2. Photos used in the experiment

Valence: 2.06 Arousal: 4.00 Dominance: 3.13	1	Valence: 4.31 Arousal: 2.85 Dominance: 4.85	2	Valence: 7.25 Arousal: 4.23 Dominance: 5.83	3
Valence: 3.26 Arousal: 3.93 Dominance: 4.31	4	Valence: 7.63 Arousal: 3.74 Dominance: 7.69	5	Valence: 3.46 Arousal: 6.87 Dominance: 3.08	6
Valence: 4.95 Arousal: 1.55 Dominance: 7.25	7	Valence: 1.68 Arousal: 6.14 Dominance: 3.37	8	Valence: 7.32 Arousal: 7.06 Dominance: 5.59	9

Table 3. The scores in IAPS for the selected photos

Table 3 above is the list of scores on each dimension in the IASP validation with regard to the selected photos. The order of cells in the table refers to the order of photos in Figure 2. While IAPS used a 9-point rating scale on each dimension, it can be seen that, for instance, the photo of a cemetery (#1) got high scores in arousal and low scores in valence and dominance.

2.1. Results and conclusions of experiment 1

After the experiment the overall impression of the researchers was that the choices of the participants were extremely evenly distributed over all photos. It was therefore hard to find any

trends in their selections. However, we summed up the results and classified the arguments, and finally something emerged from the data.

The arguments were first classified into 8 categories according to their content. The categories and the number of comments in the category are presented in Table 4. From the table it can be seen that more than one third (approx. 35%) of the comments concerned physical vertical direction.

Category	Proportion of comments (%)
Physical vertical direction	35
Property of sound	2
Positive (in general)	12
Negative (in general)	8
Surprise	14
Chaos	3
Iconic meaning	5
Cannot say / unclassified	21

Table 4. Classification of arguments

The interesting feature in this classification, which has been done solely on the basis of the content of the comments without any underlying framework, is that there are classes which reflect the dimensions of IASP. "Valence" in IASP clearly denotes positive and negative comments (in Table 4); surprise can be seen to indicate "Arousal", and chaos, at least to some extent, a low value in "Dominance". These categories, which effortlessly map on to IASP, cover 37% of all the instances. We will now analyse these 37% separately.

Arguments including a general statement of "positive feeling" were amazingly evenly distributed across all sound samples (1-4 comments of this kind per sound). Two sounds, namely #8 and #3, got the most positive comments (4 each). Both of these (see Tables 1 and 2) were ascending, but differed in terms of tempo and intensity. The result thus does not provide a basis for strong conclusions on how to evoke positive emotions using the means examined in this paper.

In contrast to positive comments, negative ones were strongly biased towards descending sounds: all except one of the negative comments including negative expressions were given to sounds #1, #2, #5 and #6. Nearly 70% of these were decelerating. It seems that a descending sound with decelerating tempo is much more likely to evoke negative or sad emotions than the opposite.

The category which is labelled as "surprise" in Table 4 includes comments containing expressions of surprise, attack, mystery and emptiness. We interpreted this to indicate similar qualities with IASP's *Arousal*-dimension. Again, the comments in this category were quite evenly distributed across the sounds, though there was some bias towards sound #7: all other sounds got on an average of 2.5 votes, with an SD of 1.13 (excluding #7), but #7 got 6 votes. Interestingly, four of them were connected to photo #6 (attacking snake). So when striving for high arousal and aggressive effect, a sound with accelerating tempo, increasing pitch and decreasing intensity would be optimal when using these three parameters of sound.

The largest individual category was what we call in Table 4 "Physical vertical direction". This refers to comments including a reference to movement in an up-down direction. Our interpretation was that this did not have anything to do with emotional dimensions of IASP. Rather, these comments indicated direct mapping of the physical dimension on to a property of sound. Therefore all the arguments concerning physical direction are irrelevant for analysis in terms of IASP. Because 35 percent of the comments were related to physical direction, it has to be concluded that the primary associations evoked by the set of sounds used, did not match the emotional dimensions which IASP is based on. Rather than trying to find more connections between the selected photos and sounds, we shall continue in experiment 2 by exploring in more detail the semantics concerning the observed connection between the sounds and physical dimensions.

As mentioned above, the use of the IASP collection to trace evoked meanings proved problematic in this case. However, the only observed problem was not the mismatch between the emotional dimensions which IASP is concerned with and the sound set used, but there were problems in the interpretation of the photos as well. Even if IASP has been validated with respectable data, it turned out that many of the photos we used had contradictory interpretations among the participants. A typical example of the problems concerned photo #4 (Figure 3). The photo is of an elderly woman and a nurse (or at least someone taking care of her). Some of the participants saw love and care, very warm feelings, in this photo,. However, some felt that the nurse in the photo is irritated and tries to prevent the old woman from touching the cup. This example shows that while a photo has numerous possible interpretations, it is difficult to use it to reveal something essential about the interpretation of the sound to which it was connected. In future work, it might be wise not to blindly rely on ready-made validations, but to carry out a concise validation of photos as part of the experiment. In this way the impact, for example, of culture, age and society, on interpretations could be taken into account.

Physical directions dominated not only the interpretation of

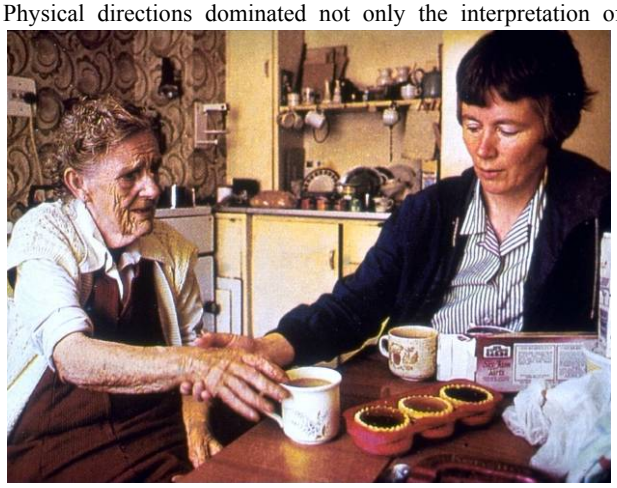

Figure 3. Photo #4 in experiment 1

sounds, but because of the nature of task, directions were sought in the photos as well. This tendency, in turn, caused clear contradictions among the results (which photo was connected to each sound), the IASP validation data and the results of a previous study [4]. In the previous study it was found that positive things tend to be connected to increasing pitch, and negative ones to decreasing. In the set of photos which we used, photo #6 (snake) had low scores in valence, i.e. it had been interpreted as 'bad' in the validation. Photo #9 (sky divers), in turn, had very high valence, i.e. evoked positive connotations. According to our previous study, negative connotations are likely to be associated with decreasing pitch, and positive to increasing, respectively. However, in the

experiment, the photo with negative connotations was strongly associated with a sound with increasing pitch while the photo with high valence (positive connotations according to IASP-validation) was connected to sounds with decreasing pitch. The reasons for this contradiction can be found in the comments of the participants: The participants felt that the snake in the photo was rising up, while sky divers were obviously falling. The conclusion is that the contradiction was in the primary emotional content of these photos and the physical movement they represent. This contradiction does not have anything to do with the sounds but is purely a picture interpretation issue. However, since the contradiction had a major effect on the sound-picture association task, we conclude that the characteristics of the sounds used should be taken into account when choosing pictorial material for this kind of experiment.

3. EXPERIMENT 2: THE POWER OF CHANGE IN PITCH, TEMPO AND INTENSITY IN THE ILLUSTRATION OF PHYSICAL DIRECTIONS

Having concluded in the first experiment, that physical directions dominate the interpretations of sounds, we carried out another experiment. In it, the research question was how changes in tempo, pitch and intensity could be applied to evoke associations with physical directions.

The second experiment resembled the first one in that the participants were supposed to connect sounds and pictures. We also used the same sound set as in experiment 1 (Tables 1 and 2). In other respects, the experimental setting was quite different. Instead of photos, we used simple arrow figures as the pictorial stimulus. All the arrows used in the experiment were identical in terms of their direction. The possible directions were up, down, left, right and the directions midway between these, resulting in 8 directions altogether (Figure 4).

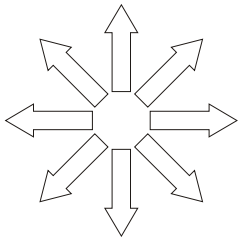

Figure 4. The range of directions of arrows

The arrows were presented to subjects in pairs. The task of the subject was to look at the pair of arrows, while listening to a sound, and choose which one of the pairs had a better match with the sound (forced two-choice test).

The pairs of arrows were arranged so that each pair included some semantically interesting contradiction, providing a basis for drawing conclusions about the semantics of the sounds used. The pairs can be seen in the first column of Table 5.

The test was implemented over the Internet with a customised application. There were 23 participants (10 females, 13 males), from all age groups between 20 and 70. Six checked the alternative *musician* on the background information form. All of the participants were either current or former staff of the Department of Computer Science and Information Systems at the University of Jyväskylä. There was no reward, but the test took only a couple of minutes per participant.

In the test, participants were asked to place their index fingers on the 'F' and 'J' keys of their computer keyboard. They were then presented with one pair of arrows and one sound at a time. The participant was asked to press 'F' if the leftmost arrow better matched the sound, and 'J' to choose the rightmost arrow. While there were 7 pairs of arrows to be combined with 8 sounds, the total number of tasks was 56. The order of tasks was randomised, and the data were collected by the application.

The collected data have been summarised in Table 5. They indicate how many percent of the participants chose a certain arrow from a presented pair. For instance, cell A1 shows that when right and left pointing arrows were presented with descending pitch, decreasing tempo and lowering intensity, 43% of the participants chose the right pointing arrow, and the remaining 57% the left pointing arrow (the sum of the two numbers in each cell is always 100).

We analysed the table by reflecting on it with the hypotheses we had. Hypotheses #1-#3 and #5 were based on our previous study [4]:

- H1: Increasing pitch is associated with the up-direction, decreasing with the down-direction.
- H2: Increasing pitch is associated with a right pointing arrow, decreasing with a left pointing arrow.
- H3: In the case of conflict between directional associations, the vertical direction dominates the horizontal. For instance, if an arrow is pointing up-left, the up-direction dominates the interpretation.
- H4: The hypotheses concerning changes in pitch hold for tempo and intensity as well.
- H5: In the case of conflict among change of different properties, pitch change dominates the interpretation.

	1	2	3	4	5	6	7	8
A ⇨ ⇦	43 / 57	35 / 65	52 / 48	61 / 39	30 / 70	57 / 43	64 / 36	70 / 30
B ⇩ ⇧	83 / 17	87 / 13	17 / 83	9 / 91	83 / 17	77 / 23	17 / 83	0 / 100
C ⇖ ⇘	13 / 87	9 / 91	61 / 39	74 / 26	9 / 91	9 / 91	65 / 35	65 / 35
D ⇨ ⇧	65 / 35	83 / 17	17 / 83	13 / 87	78 / 22	61 / 39	26 / 74	13 / 87
E ⇩ ⇦	87 / 13	78 / 22	48 / 52	30 / 70	83 / 17	83 / 17	35 / 65	26 / 74
F ⇨ ⇖	70 / 30	61 / 39	48 / 52	22 / 78	61 / 39	70 / 30	39 / 61	17 / 83
G ⇦ ⇘	30 / 70	22 / 78	45 / 55	52 / 48	13 / 87	22 / 78	45 / 55	50 / 50

Table 5. Frequencies in percents

3.1. Results and conclusions of experiment 2

We now analyse the table by assessing the data against the hypotheses.

3.1.1. Pitch change

Of the whole table, only one single value does not support H1-H3 and H5. Namely, according to H2, a descending sound should have been associated with a left pointing arrow if the alternative is a right pointing one. In cell A6 we can see that slightly more participants chose against the hypothesis. However, the difference is very small, mainly indicating that horizontal cues are not as strong as vertical ones; when the same sound was combined with up and down pointing arrows (cell B6), the support for H1 is clear.

A good example of the dominance of the vertical direction is row F. Even if the right hand side arrow points to the right, or "forward" in Western writing systems, it is systematically chosen to match a decreasing pitch better than the up-left pointing arrow. This is probably because it is the least bad of the alternatives when the other arrow is pointing up (and "back"). So the up-direction of the second arrow could not be overridden, and the first arrow had to be chosen.

3.1.2. Tempo change

According to H4, the changes in tempo should have been reflected in the associations with physical directions as well as pitch changes. As a rule, the hypothesis holds. However, in Table 5, a few exceptions can be found.

Given that H1-H3 and H5 hold, pitch change determines which one of the two arrows is chosen. To trace the impact of tempo change, we compared pairs of cells in Table 5 by choosing pairs in which other properties (pitch and intensity change) are identical and the only difference is tempo change. We then interpreted H4 in such a way that if other properties are identical, increasing or decreasing tempo should have an effect on the proportion of choices. For instance, sounds #2 and #6 are such a pair. When comparing B2 and B6, for example, it can be seen that acceleration (B6) strengthens the "up" connotation (upward arrows chosen in B2 by 13% while in B6 23% chose it).

Having analysed each pair of sounds in the described way, we found out that in all but one pairing H4 holds in 71% of the cases. The exception was the pair of sounds #1 and #5. With these, H4 only held in 43% of the cases. What is then so different in sound pair #1 and #5 that it gave such a different result? At least it can be seen that in these sounds, both pitch change and intensity change were decreasing, while tempo change alone was increasing. It can be concluded that probably tempo change is such a weak cue that it loses its power if both pitch change and intensity change indicate a different direction than it does.

3.1.3. Intensity change

The analysis of tempo change and reflection on H4 was carried out in a similar way as the analysis of tempo change; the focus was on pairs in which tempo change was the only discriminating factor. H4 would have meant that intensity change could have been seen in these comparisons.

In an astonishingly similar way to the results of tempo change, the analysis of intensity change shows that in all except one pair the results clearly support H4. When comparing either #3 and #4, #5 and #6 or #7 and #8, 90% of the results support H4. For instance, when comparing B3 and B4, the only difference is that in B3 intensity decreases and in B4 it increases; we can see that the proportion of those choosing the upward pointing arrow grows from 83% to 91% when using increasing intensity.

However, the exception is really clear: when comparing sounds #1 and #2, only 29% of the data supports H4. Focusing on the properties of #1 and #2 we can see that both pitch and tempo are decreasing. The conclusion would be analogous to the one concerning the exception in tempo change: intensity change alone is too weak a cue to have an effect on interpretation if both pitch and tempo change pull the interpretation in a certain direction.

3.1.4. Other findings and conclusions

There is one cell in Table 5 which attracts attention when trying to analyse the data: In B8 conditions every single participant made the same choice. In that task, the participants had to choose between up and down pointing arrows while hearing a sound which was increasing in all of the studied dimensions. As all the participants chose the up pointing arrow, this can be seen as strong support for the hypotheses. Also, the clear result provides proof of the reliability of the test setting.

It was interesting to notice how strong the association between physical vertical direction and direction of pitch change was; anyway, the concepts of 'high' and 'low' pitch are highly metaphorical [5] and don't have direct physical basis.

To sum up the other findings it can be said that pitch change dominates the interpretations, but changes in tempo and intensity have some effect as well. When striving towards clear design and strong effect, contradictions among these factors should be avoided or at least used with care. However, if there is need to attenuate the strong directional cue of pitch change, that could be done with tempo and intensity change, which both are in contradiction with pitch change. One of the weaker cues (tempo and intensity) alone is not enough.

4. DISCUSSION

The two experiments described in this paper, have a very different orientation from each other. However, both aim at revealing something essential in the interpretation of sound signs which differ in pitch change, tempo change and intensity change.

Whether studied with the help of photos (as in expt. 1) or simple symbols (expt. 2), certain features became clear. Above all, in accordance with previous studies, change of pitch strongly dominates the interpretation. How much can be done with tempo and intensity change, is probably highly dependent on conditions: how much the intensity really changes and how drastic is the tempo change. The results reported in this paper obviously only concern the particular sound samples we used.

In future work it would be important to study the strength of change in these dimensions as well. This way we would have a more comprehensive idea of the relevant parameters.

Methodologically these two experiments were very different. From the second one we got clearer results, but could only handle very simple information (physical directions). Using photos has potential for going into deeper semantics, but as discussed earlier in this paper, there are problems with the interpretation of photos. Despite the problems the setting in the first experiment was encouraging in that it provoked the participants to talk about their ideas. Perhaps with some other kind of photos deeper emotional phenomena could be accessed. We thus conclude that photos with emotional content have potential in the study of the semantics of sounds, but the photo set should be locally validated, to take into account cultural and other background factors.

ACKNOWLEDGMENTS

This work is funded by the Finnish Funding Agency for Technology and Innovation (Tekes).

5. REFERENCES

[1] G. Kramer, B. Walker, T. Bonebright, P. Cook, J. Flowers, N. Miner, J. Neuhoff, R. Bargar, S. Barrass, J. Berger, G. Evreinov, W. Fitch, M. Gröhn, S. Handel, H. Kaper, H. Levkowitz, S. Lodha, B. Shinn-Cunningham, M. Simoni, and S. Tipei. *The Sonification Report: Status of the Field and Research Agenda*. Report prepared for the National Science Foundation by members of the International Community for Auditory Display. ICAD, Santa Fe, NM, 1999.

[2] J. Edworthy, E. Hellier & R. Hards. The semantic associations of acoustic parameters commonly used in the design of auditory information and warning signals. *Ergonomics 38*, 11, pp. 2341–2361, 1995.

[3] E. Hellier, J. Edworthy and I. Dennis. Improving auditory warning design: Quantifying and predicting the effects of different warning parameters on perceived urgency. *Human Factors 35*, 4, pp. 693–706, 1993.

[4] A. Pirhonen. Semantics of sounds and images - can they be paralleled? In *Proceedings of International Conference on Auditory Display ICAD 2007*, Schulich School of Music, McGill University Montreal, Canada, 26-29 June 2007, pp. 319-325.

[5] A. Pirhonen, To simulate or to stimulate? In search of the power of metaphor in design, in *Future Interaction Design*, Springer Verlag, London, pp. 105-123, 2005.