SEMANTICS OF SOUNDS AND IMAGES – CAN THEY BE PARALLELED?

Antti Pirhonen

Department of Computer Science & Information Systems
P.O.Box 35, 40014 University of Jyväskylä, Finland
pianta@jyu.fi

ABSTRACT

This paper presents a practical and simple web based test to survey the semantics of non-speech sounds in relation to simple images with a wide variety of users. The main findings from the data colleted are presented. A case study of how changes in pitch are related to the interpretation of short non-speech sounds is discussed based on the results of the semantic tests. Finally the experiment method itself is discussed in terms of its appropriateness for the analysis of the semantics of non- speech sound.

[Keywords: Non-speech sound, semantics]

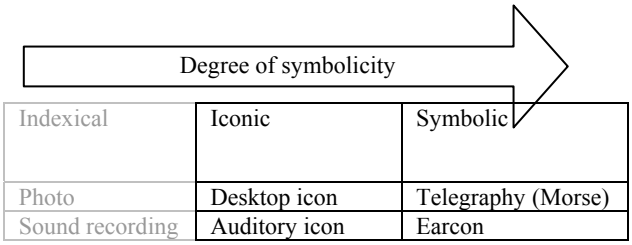
1. INTRODUCTION

The semantics of non-speech sounds is an important research topic for a number of reasons. Although it can be argued that sounds can be used without the intention to deliver a certain message, in most applications the crucial criterion for the success of sound design is the evoked meaning.

A meaning of a non-speech sound, i.e., its semantics, can be approached from two directions:

1) sound as an icon
2) sound as a symbol

When analysing a single sound as an icon, the focus of interest is on what it represents. If the connection between the sound and what it represents is obvious, it can be classified as an auditory icon [1]. When this relationship seems arbitrary, the more the sound is symbolic and gets its semantic value through a certain semantic system. In linguistics, this system is a grammar of certain language. In sounds, this could be an acquired hierarchy of earcons [2]. Similar to natural language, earcons have a syntax, which needs to be learned in order to communicate and understand them. Auditory icons, in turn, are designed to be interpretable without prior knowledge. However, the division of sounds into iconic and symbolic categories is far from clear (see the conceptual analysis in e.g. [3]). It would be more appropriate to define purely symbolic and purely iconic sounds as two extreme ends of a continuum[1] (Table 1). One reason for defining a continuum rather than separate categories is that there are features in symbolic sounds which are universally interpreted in the same way. For instance, there are studies that provide empirically derived guidelines for the design of warning signals (see, e.g., [4, 5]). In other words, even if a warning signal is arbitrary and therefore highly symbolic in nature, it is important that it can be intuitively interpreted as a warning without learning.



Indexical	Iconic	Symbolic
Photo	Desktop icon	Telegraphy (Morse)
Sound recording	Auditory icon	Earcon

Table 1. Degree of symbolicity with examples.

The majority of literature concerning the semantics of non-speech sounds deals with warnings. A possible reason for this is that situations that necessitate a warning usually require a strong and rapid reaction. However if we knew more about the intuitive interpretation of highly symbolic sounds, they could be applied in much wider variety of current use contexts.

Symbolic sounds are too complicated to be designed in purely analytic manner. We will never have a complete set of rules to determine how specific meanings should be expressed as non-speech sound. However this does not make empirical studies of semantics of sounds useless. On the contrary, the more we know about the semantics of symbolic sounds, the more likely it will be to achieve effective designs.

This paper is one step in the body of research concerning the semantics of non-speech sounds. The aim of this study was to

- identify similarities in the interpretation of pitch changes of short sounds and to
- investigate the applicability of an Internet survey to implement a simple test concerning the interpretation of sounds.

In the following section of this paper we present the organisation of the study. The third section presents the core of the results. In the fourth subsection the results are discussed and finally the method is analysed in terms of this study.

[1] Semioticians would define this as an indexical relationship, instead of iconic, at the end of the continuum, iconic being an intermediate stage. We restrict the scope of this discussion to icon-symbol part of the axis because it is clearer in terms of auditory icons and earcons.

2. INTERPRETING SOUNDS WITH IMAGE CHOICES

Interpreting a sign in a different modality can be difficult or even impossible. The phrase 'one picture is worth one thousand words' reflects this difficulty. Therefore, investigating possible interpretations of non-speech sounds is a challenging process. We could have asked the participants of our experiment to verbalise how they understood sample sounds (like, e.g., [6]). However, we wanted to use a large number of participants and a method, which would not be sensitive to cultural differences. Therefore we chose to use a rapid forced dual choice test. In the test, the participants were presented one sound and two images at a time. The task of the participant was to choose which one of the images best matched the sound. There were 14 different sounds and 10 different pairs of images, resulting in 140 tasks. The order of tasks was randomised.

1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

Table 2. The image pairs used as pictorial stimulus.

The images used as a stimulus were arranged in fixed pairs, and the task of the participant was to choose one of the images in certain pair. The ten image pairs are presented in Table 2. The order of the two images in each pair was always the same, i.e., if the image pairs in a task contained a bouquet and a tank (see Table 2), the bouquet appeared consistently on the left hand side with the tank on the right hand side.

The image pairs were of two types. Firstly, there were five pairs of simple drawings of real life entities. In each pair, one image represented the opposite of the other (Table 2). For instance, there was a bouquet and a tank, suggesting love and hate. A pair consisting of images of new born child and a skull was intended to represent life and death. Secondly, there were five pairs of simple arrows. All the arrows were identical except for their direction. The directions of the arrows differed by 45 degrees resolution, e.g., there were 8 different possible directions.

The audio stimulus consisted of 14 different sounds. The duration of each sound was between 600 and 1200 ms depending on the complexity of the sound. The sounds were originally designed with a sequencer, using a GM2 patch 74 (recorder) as the timbre. In each sound, the pitch changed continuously. The difference between each sound was the form of pitch change. The range of pitch change for each sound was approx. from F#5 to A5. The form of pitch change in each sound is illustrated in Table 3.

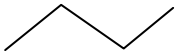
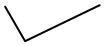
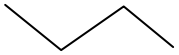
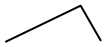
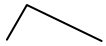
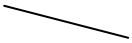
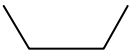
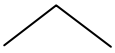
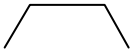
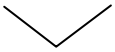
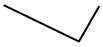
#1		#8	
#2		#9	
#3		#10	
#4		#11	
#5		#12	
#6		#13	
#7		#14	

Table 3. The form of pitch changes in each sound.

The survey was implemented with a web-application. The invitation to participate was delivered through random selection of thematically relevant email lists and individual addresses in different continents (personal contacts). The invitation encouraged the recipient to forward the message. However the final set of participants illustrated that people are not too willing to forward such messages, perhaps due to the excessive amount of junk mail which fills our inboxes nowadays.

The invitation consisted of a polite explanation of the experiment and the system requirements. The essential information was the URL of the test application itself. The

detailed instructions about the test were included at the beginning of the test web site. The test site was opened with a technical test and adjustment:

"*First test the sound properties of your system by pressing f. You can repeat the test sound as many times as you wish to adjust the volume in a convenient level. If you cannot hear the sound, you probably have to check the audio system of your computer and restart this application. Javascript support is also required from your browser.*

I can hear the sound.

I cannot hear the sound."

Clicking the later alternative ("I cannot hear the sound") lead to a page in which the user could choose to simply halt the test or send a report of the problem before possibly trying again. The first alternative resulted in continuing the start-up procedure and lead to personal information page. In it, age, sex, nationality, whether the participant was a musician or not, and optional email address were asked. Filling in the email field was told to indicate that the participant wished to receive a summary of the study.

Next page contained instructions for the test session:

"*You are supposed to associate the sound you hear with either of 2 presented pictures. I.e., you will be presented one sound and a pair of pictures at a time. Once you have selected either of the pictures, you will get a new set of one sound and 2 pictures. The whole experiment contains 140 such sets. There is no time limit.*

Make your selection using f- and j-keys. F-key refers to the left hand side picture, j-key to the right hand side. Place your hands conveniently on the keys and concentrate on the sound. You are supposed to choose either of the pictures in any case. If you think that neither is good, just choose the one that might be closer to what you wished. Please do not use your browser's navigation functions (like back-button) during the session.

Now adjust the volume of your system to a convenient level with the help of this test sound (press f to play).

Play the testsound by pressing f. Start the test by pressing j."

The final number of participants was about 70. However, it is hard to say on the basis of the registered data the exact figure since the data indicates that some people tried, but due to technical problems gave up and possibly tried again with another computer or with new browser settings. We rejected all cases in which the amount of undone tasks was greater than 2 (out of 140). The amount of valid performances was then 41, out of which only 4 had one or two undone tasks.

39% of the participants (with valid performance) were 20-29 years old, 32% between 30 and 39, 22% between 40-49 and 7% 50-59 years old. Finally only 5 participants outside Europe managed to perform the experiment (Australia 3, Canada 1, China 1). Most were Finns (37%) or Britons (34%), and the rest of the participants came from Greece (3), Ireland (2), Bulgaria (1), France (1), and the Netherlands (1). In other words, cultural coverage was not very good. 15 of the participants were female and 26 male.

3. DATA ANALYSIS

In the experiment, as described above, each participant had a sequence of 140 simple tasks. Figure 1 illustrates the screen design of one task.

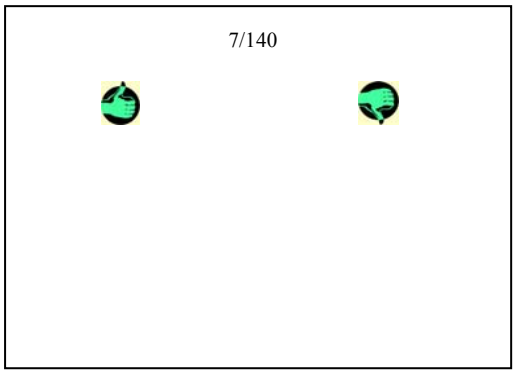


Figure 1. The screen design of one task.

The screen design was as simple as possible to help the participants to concentrate on the essence. The numbers on the top of the screen illustrated the progress; e.g., the sample task of Figure 1 is the seventh out of 140.

Since each task was to choose between two alternatives, we extracted the frequencies of choices from the raw data. In other words, we calculated the percentage of participants (with valid performance) that chose a certain image with a certain sound. Although this could be considered simple analysis, it revealed a lot about

- the semantics of pitch changes
- the interpretation of simple images and sounds in an intensive experiment context
- the appropriateness of this kind of survey.

Since the task was to listen to the sound and choose which one of the two presented images matched better to the sound, it was not only a sound interpretation task. The choice of image also informed about the user's interpretation of each image. The underlying assumption in the design of the experiment was that the images in the experiment were quite clear. This was particularly true concerning the arrows (see Table 1). However, the data indicated that the rest of the images were more ambiguous. The results will now be discussed in detail, extracting the most salient details.

3.1. Key findings

The first impression of the data was that there were a number of tasks in which there was surprisingly high agreement among participants. The clearest cases were related to a sound with an increasing pitch. Table 4 contains percentages of choices in some interesting cases. (The order of images in pairs is not identical to that in actual test. In Table 4, the images have been arranged so that the image with the higher percentage in the case of increasing pitch is presented first. The order of image pairs in the actual test is presented in Table 2).

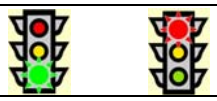
	Sound #3		Sound #4	
1 (pair #7)				
	89.5%	10.5%	17.1%	82.9%
2 (#4)				
	81.6%	18.4%	12.2%	87.8%
3 (#3)				
	81.6%	18.4%	43.9%	56.1%
4 (#2)				
	76.3%	23.7%	22.0%	78.0%
5 (#5)				
	76.3%	23.7%	70.7%	29.3%
6 (#8)				
	73.7%	26.3%	14.6%	85.4%
7 (#9)				
	84.2%	15.8%	46.3%	53.7%
8 (#6)				
	63.2%	36.8%	58.5%	41.5%

Table 4. Strongly agreed choices for sounds with increasing and decreasing pitch.

Starting from row 1 in Table 4, the first frequencies appear more than obvious: increasing pitch and arrow up are strongly related (almost 90% of the participants). The opposite, decreasing pitch and arrow pointing down, was almost as clear. Thumbs pointing up and down (row 2) were interpreted in as obvious a manner – there was hardly any difference to arrows (row 1). On row 3, the images are more ambiguous, but the life-death image pair still was interpreted clearly in the same way as up and down arrows, life representing 'up' and death 'down'. However, a decreasing sound was not always commonly agreed with this image pair. Row 4 in table 4 illustrates that a laughing character was associated with increasing pitch and the angry character with decreasing pitch. In another clear example the green traffic light was assigned a rising pitch, even if the green light is the lowest and therefore could have been associated with the decreasing sound.

The results for the rest of the arrows (rows 6-8) are interesting to interpret, since there are two dimensions which both could be related to increasing and decreasing pitch: up-down and left-right. Since almost all participants were from countries with a writing system which proceeds from left to right, it could be hypothesised that a common association of left to right would be 'forward' or 'upwards', thus relating to increasing pitch. However, the data of this experiment indicates that left-right dimension in the sense of going forward or backwards is much weaker than up-down direction. Arrows pointing left and right were still usually interpreted as hypothesised (from left to right -> increasing pitch and vice versa), but when combining up-down –dimension and left-right –dimension, the up-down one was clearly dominating. On row 6 can be seen that when the arrow pointed up and backwards, it was associated with increasing pitch by ¾ of participants. On the same row, we can see that the forward-down pointing arrow and decreasing pitch was even clearer proof of the same phenomenon.

The argued weakness of a left-right dimension as a metaphor of moving on forward-backward axis (or increasing-decreasing pitch) seems to be especially true concerning going backwards: on row 7 the two later cases (decreasing pitch) show that there was hardly any difference between up-forward and up-backward arrows. On row 8, we can see that when participants were to choose between forward and backward arrows when hearing a decreasing pitch, the majority actually chose the forward pointing arrow.

The reason for the weak association between decreasing pitch and right-left direction is not possible to see from this data. It would be worth clarifying with other methods, probably more qualitative ones in which the participants had an opportunity to explain the criterion for their choices. Perhaps sound signs in general are conceptualised as dynamic, forward moving entities, which makes it hard to associate them with an arrow pointing backwards (row 8).

The sounds with increasing and decreasing pitch (sounds 3 and 4 in Table 3) were the simplest ones, and illustrate clearly the way that participants reacted to given stimulus. The results concerning most of the other sounds can be interpreted best by comparing them to the results of sounds 3 and 4. In Table 5 we compare the choices between up and down arrows with different sounds.

Table 5 illustrates that when pitch splits into two directions as part of the same sound, the dominating direction (portion of duration) usually determines the interpretation. Thus sounds 8 and 9 were associated analogously to sound 3; sound 10, respectively, was associated with an arrow pointing down, just like a sound with decreasing pitch (#4). The only exception is the interpretation of sound 7, which was associated, though weakly, with an upward arrow even though the decrease of pitch was dominating. In sounds 5 and 6, in which the duration of increase and decrease of pitch was equal, participants slightly preferred the upward arrow.

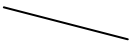
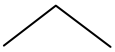
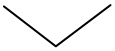
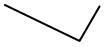
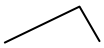
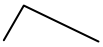
			
Sound #3		10.5%	89.5%
#4		82.9%	17.1%
#5		36.6%	63.4%
#6		32.5%	67.5%
#7		37.5%	62.5%
#8		14.6%	85.4%
#9		24.4%	75.6%
#10		70.7%	29.3%

Table 5. A comparison of choice rates of up and down arrows

In the interpretation of Tables 4 & 5 it can be concluded that sonifying directions backward and down is more difficult than illustrating upward and forward directions with sounds.

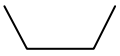
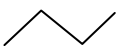
					
Sound #11		78.0%	22.0%	58.5%	41.5%
#12		34.1%	65.9%	29.3%	70.7%
#1		80.5%	19.5%	68.3%	31.7%

Table 6 Interpretation of sounds with two turningpoints in pitch change.

Table 6 lists the most salient trends in the interpretation of two sounds where both have two changes in the direction of pitch change. As can be seen, the interpretation of sound 11 is quite similar to the interpretation of simple pitch increases (sound 3). The interpretation of sound 12 resembles the interpretation of decreasing pitch respectively. It appears that at least in these sound samples, the ending of the sound determines the interpretation. This is probably due to the relatively long steady pitch before the change in the ending, thus making the rapid change in the end the most notable detail. The results concerning sound 1 (see Table 6) are in accordance with both the assumption concerning the importance of relative duration of increase of pitch and the importance of the ending of the sound. For example sound 1 was interpreted in very much the same way as a purely increasing sound.

The interpretation of the data so far has concerned results which have a logical explanation. However, there was one pair of images which made the participants respond in an extremely unexpected manner. The image pair on row 1 in Table 2 (a bouquet and a tank) evoked choices which appear, in some cases, completely contradictory to our expectations. Table 7 shows the most unexpected results.

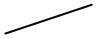
					
Sound #2		76.3%	23.7%	28.9%	71.1%
#3		42.1%	57.9%	81.6%	18.4%
#4		65.9%	34.1%	12.2%	87.8%
#12		80.5%	19.5%	29.3%	70.7%

Table 7. A comparison of frequencies concerning two different image pairs.

Since the direction of the thumb image was in most cases in accordance with positive things like up (vs. down), forward (vs. backward), birth (vs. death), it can be assumed that a thumb pointing up evoked positive connotations and a thumb pointing down negative. Therefore, it seems strange that concerning certain sounds, for example sounds 2 and 12 and to some extent concerning sounds 3 and 4, the choices were in conflict with this positive-negative trend; the same sound (#2) which was connected to 'negative' thumb direction, was strongly connected to a bouquet ('positive'). However, when discussing conflict or contradiction, the underlying assumption is of course that the image of a bouquet would evoke positive and tank negative connotations. The current data is inadequate to provide a basis to speculate on the explanation for this. However, it at least indicates that when there are two simultaneous interpretation tasks (the interpretation of images and the interpretation of sound), the results are the more unpredictable the more ambiguous the images and sound are. Given the assumption that an image pair consisting of a bouquet and a tank would be an unambiguous contrast between love and hate, the results indicate that the interpretations were more diverse.

4. CONCLUSIONS AND DISCUSSION

As the aims of the study are concerned with both the applicability of the method and certain aspects of the semantics of non-speech sounds, we discuss these issues separately below.

4.1. The semantics of pitch change

According to our experiences, pitch change is a strong means of expressing certain meanings. Although the sound samples we used were relatively short, one (Table 3, sounds 5-10) or two (sounds 1-2, 11-14), turning points in the direction of changing pitch resulted in different interpretations of the meanings. This revealed that in comparison to straight increasing or decreasing pitch, pitch changes can be an effective means of expression, even in short sounds

What do pitch changes convey then? As stated in the introduction, the semantics of sounds will never produce a complete set of guidelines, on the basis of which any meaning could be sonified. However, as part of implementing this study, we learned something about how to use sounds effectively convey a certain message. We now summarise the central findings.

1) The simpler the sound, the more universal the interpretations should be. The simplest sounds in our experiment were 600 ms sound samples with continuously increasing or decreasing pitch (sounds 3 and 4 in Table 3). The data shows that pointing up and down, forward and backward, expressing something positive vs. negative is relatively simple with these kinds of sounds: increasing sound means up and forward, decreasing the opposite. However, pointing down and especially backwards is more difficult to interpret. This is in accordance with our previous studies, in which we found left-right dimension difficult to be sonified even with spatial sound [7, 8]
2) When there are changes in the direction of pitch change, i.e., the sound contains both increasing and decreasing sections, the dominating direction of changes determines the evoked meaning. In our experiment, the domination meant the longest relative duration, or the ending of the sound. In other words, when designing a sound with pitch changes, the most important change is the one which last longest or is the last one.

4.2. The appropriateness of the method

Using the Internet to conduct an experiment instead of laboratory-based tests enables a fast way of collecting data from all over the world. However, the disadvantages are quite as obvious. Laboratory conditions can be controlled by the researcher, while performing a test remotely via Internet contains a lot of uncertainty. First of all, the researcher can never be sure in such a test about the quality of audio conditions. Even though the participants were asked to adjust the volume before the actual test session, the audio equipment (sound card, speakers etc) as well as the acoustic environments were unique for each participant. Secondly, the technical reliability and compatibility with the experiment application of each workstation was a risk. Despite intensive tests with different browsers the data showed that almost half of the attempts to participate failed. However, assuming that this loss was not systematic but merely random, it doesn't need to be taken into account in the interpretation of data.

One rationale for distributing the experiment over the Internet was to attain wide cultural coverage. However, in this experiment, the participation was clearly concentrated in Europe. On the other hand, even if we had managed to get participants from all over the world, it wouldn't necessarily mean that we had achieved cultural coverage. In many countries Internet usage is restricted to a privileged minority, which hardly represents an average view and dominating sub-culture of that country.

In western countries, the amount of junk mail could cause problems in the delivery of requests to participate this kind of experiment. Furthermore it is an ethical issue; to ask people to randomly forward a request could quickly resemble junk mailing.

An interesting and highly relevant observation about the behaviour of the participants can be seen in the coherence of certain choices. As can be seen in Table 4, when presenting a sound with increasing pitch, the sound is connected equally frequently to an upward arrow, upward pointing thumb, laughing face or green traffic light. In general, it seems that there was no big difference in the interpretation of those images, what ever the sound was. We did not systematically interview the participants, but a couple of random discussions gave us a reason to assume that this phenomenon is related to the research method. When a participant is exposed to an intensive experiment, in which s/he spends one or two seconds per task, the participant creates a strategy. She (or he) does not want to appear unintelligent and therefore creates a simple logic for choices. In this case, the participants might have decided after a few tasks that increasing pitch means all positive things, and direction upward. In other words, the participants seemingly made a simple classification of different types of sounds and images, and tried to be consistent in their choices after that. This kind of strategy makes the performance of whole set of tasks easier and faster. However, the assumed strategy is a disadvantage for analysis of the experiment. This kind of strategy made it difficult or impossible to trace any nuances in the evoked meanings. In future experiments, we will stress in the task instructions to rely on intuition and not to try to be logical. Another solution would be to develop an application, which would pop up randomly every now and then e.g. in the middle of a work day, presenting one task at a time. Although this could be a more valid method to obtain intuitive choices, it might be more difficult to get users to participate.

Despite its weaknesses, the method is – once the application has been implemented – an effective way to test the semantics of sounds with a wide variety of people with reasonable effort. It could be used in basic research of semantics of sounds as well as in practical sound design.

ACKNOWLEDGMENTS

This work is funded by Finnish Funding Agency for Technology and Innovation (Tekes). Thanks to all participants of the

experiment, Markus Bengts for the technical implementation of the test application and Emma Murphy for her valuable help in writing process.

5. REFERENCES

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

[2] 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.

[3] A. Pirhonen, E. Murphy, G. McAllister and W. Yu, "Non-speech sounds as elements of a use-scenario: A semiotic perspective," in *Proc. 12th Int. Conf. on Auditory Display (ICAD)*, London, UK, June 2006, CD-ROM-format.

[4] J. Edworthy, S. Loxley, and I. Dennis, "Improving auditory warning design: Relationship between warning sound parameters and perceived urgency," *Human Factors*, vol 33, no. 4, pp. 693-706, 1993.

[5] R. D. Patterson, "Guidelines for auditory warning systems on civil aircraft," C.A.A. Paper 82017, Civil Aviation Authority, London, 1982.

[6] H. Palomäki, "Meanings conveyed by the simple auditory rhythm," in *Proc. 12th Int. Conf. on Auditory Display (ICAD)*, London, UK, June 2006, CD-ROM-format.

[7] A. Pirhonen, S. Brewster, and C. Holguin, "Gestural and audio metaphors as a means of control for mobile devices", in *Proc. Computer Human Interaction (CHI)*, Minneapolis, Minnesota, April 2002, pp. 291-298.

[8] 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.