THE FLOPS GLASS: A DEVICE TO STUDY EMOTIONAL REACTIONS ARISING FROM SONIC INTERACTIONS

G. Lemaitre
IRCAM, Paris, France
lemaitre@ircam.fr

O. Houix
IRCAM, Paris, France

K. Franinović
Zürcher Hoschule der Künste, Zürich, Switzerland

Y. Visell
McGill University, Montréal, Canada

P. Susini
IRCAM, Paris, France

ABSTRACT

This article reports on an experimental study of emotional reactions felt by users manipulating an interactive object augmented with sounds: the Flops glass. The Flops interface consists of a glass embedded with sensors, which produces impact sounds after it is tilted, implementing the metaphor of the falling of objects out of the glass. The sonic and behavioural design of the glass was conceived specifically for the purpose of studying emotional reactions in sonic interactions. This study is the first of a series. It aims at testing the assumption that emotional reactions are influenced by three parameters of the sounds: spectral centroid, tonality and naturalness. The experimental results reported here confirm the significant influence of perceptual centroid and naturalness, but fail to show an effect of the tonality.

1 INTRODUCTION

New technologies make it possible for designers to consider sonic augmentations of a wide array of everyday objects that incorporate electronic sensing and computational capabilities. Sound-mediated interactions raise several interesting issues related to the functionality and the aesthetic of a design [8]. Another interesting question is that of the emotional reactions induced by such sonically augmented interactions. This question is of further importance when considering that a user's preferences and other evaluations of a product are influenced by her emotional reactions when using this product [9].

To address this issue, an interactive object was designed, called the Flops. It is a glass embedded with a tilt sensor allowing it to control the generation of impact sounds when tilted. It implements the metaphor of a glass full of virtual items that may be poured out of it. The sounds can be easily modified, in order to assess the influence of the sound parameters on the emotional reactions of the users.

In the experiment reported upon here, 25 participants were required to watch a set of videos displaying a user pouring virtual items out of the Flops glass. They had to report their emotional reactions by providing judgements on three scales: valence, arousal and dominance. The images in the videos were all the same, and only the sounds changed. The sounds were created on the basis of conclusions of other experimental studies that have suggested that the affective reactions to aircraft noises were influenced by several aspects of the sounds: sharpness, tonality and naturalness [20]. The goal of our study is to explore whether these conclusions are valid for the sounds used in the Flops glass, which consisted of very short impact sounds sequenced in various temporal patterns. This study is intended to be the first of a series. It investigates how the sounds only might influence the emotional reactions of the users.

2 EMOTIONAL REACTIONS TO SOUNDS

Emotions Emotions have been studied by philosophers and scientists for centuries. Yet the question of what are emotions is still a matter of debate. Most modern emotion theorists have adopted a componential approach to emotions, suggesting that an emotion episode consists of coordinated changes in several components: physiological arousal, motor expression, subjective feelings, behavior preparation, cognitive processes [13, 14, 15, 5]. Feelings (or core affects [13]) are considered as the conscious reflection of changes occurring in these components. There exists several approaches for the assessment of emotions, the most widespread being the physiological measures (heart rate, skin conductance, facial EMG, startle reflex, etc.) [2], the basic emotions approach [4] and the dimensional approach, on which we choose to focus in this study. From more than 50 years indeed, studies have suggested that the emotional reactions observed in, or reported by subjects can be accounted by a two- or three-dimensional framework [16]. For instance, Osgood suggested the following dimensions as primary referents of facial expressions of emotions: pleasantness, control, and activation [10]. This approach has been formalized in the circumplex model of affects proposed by Russell [12]. Three dimensions are generally considered: the va-
ence dimension, describing unpleasant to pleasant feelings, the arousal dimension, describing the degree of arousal (from calm to excited) felt by the subject, and the dominance dimension, describing how dominated or dominant feel the subjects.

Emotional reactions in sound design Assessing emotional reactions of a user using a product is of major importance for designers [18]. It has particularly been reported that attractive products are perceived easier to use [9]. Emotional reactions to the sounds of everyday products have been primary studied in terms of unpleasantness or annoyance (see for instance [6]), or preference (see for instance [19]). Specifically, Västfjäll et al. [20] have found significant correlations between valence and arousal ratings and several psychoacoustical descriptors of aircraft sounds: they found valence to be correlated with loudness and naturalness (naturalness was rated by listeners), and activation with sharpness and tonal content.

3 INTERACTION AND SOUND DESIGNS OF THE FLOPS GLASS

The Flops glass is an interface similar to a glass containing virtual objects. When tilted, virtual objects drop out of the Flops glass, producing impact sounds when hitting the surface above the glass. The sounds were created in order to test the results reported above: they were made so as to vary along their spectral centroids (similar to the sharpness descriptor used in [20]), and tonality indexes. Natural and synthetic sounds were used.

3.1 Physical design

The physical interface of the Flops glass is shown in Figure 1. Its shell is modeled in 3D software and is extruded in ABS plastic. The interface contains an accelerometer (Analog Devices ADXL 320 3-axis MEMS accelerometer), sensing the gesture performed with the glass. The sensor is wired to an Arduino BT board, sending the tilting data through Bluetooth connection to a remote computer. The sensor data processing and the playback of the sounds are real time processed in Cycling’74 Max/MSP 5.0.6.

3.2 Interaction design

The interaction model, transforming tilt angle to a flow of objects falling, is based on the model of items sliding without friction on a tilted rod, and falling when reaching the extremity of the rod (see Figure 2). A reservoir of virtual items is situated aft inside the Flops glass (at a distant of the mouth of the Flops glass), where virtual items are stored (regularly separated by a distance $d_0$). Assuming no friction, when the Flops glass is tilted with a constant angle $\alpha$, the position of each item is:

$$x_n(t) = -nd_0 + k \sin(\alpha) t^2$$

where $k$ is a constant. Assuming that $d \gg d_0$, the time between two successive dropped items is $\Delta t = K / \sqrt{\sin \alpha}$, where $K$ is a constant. The rate of impacts is therefore constant for $\alpha$ constant.

3.3 Sound design

The model described above is used to drive the generation of impact sounds: when the Flops glass is tilted, a series of impact cues is generated, with a rate computed as described above. Each cue triggers the playback of a sample of an impact sounds.

Thirty-two samples were created. Sixteen sounds (“natural sounds”) were created by recording different impact sounds, from collision of everyday objects to musical percussions. Sixteen (“synthetic sounds”) were samples of sounds synthesized by various kinds of algorithms (mostly additive and subtractive synthesis). The creation and the selection of these sounds were made such as homogeneously sampling across two psychoacoustical descriptors: spectral centroid
and tonality index (computed according the IrcamDescrip-
tor toolbox [11]). The spectral centroid of the sounds vary
from 410 Hz to 1890 Hz, and the tonality of the sounds vary
from 0.07 to 0.96 (the index of tonality can theoretically
vary from 0 to 1, with 0 corresponding to a white noise, and
1 to a pure tone). All the sounds were created to have a
rather short attack time (from 44 ms to 90 ms). All samples
have the same duration and last approximately 350 ms.

4 EXPERIMENTAL STUDY

The experimental study reported here aims at testing the as-
sumption that the parameters used to create the Flops glass
sounds (spectral centroid, tonality, naturalness) influence the
emotional reactions of participants watching a set of videos
displaying a user manipulating the Flops.

4.1 Method

Participants Twenty-five participants (14 women and 11
men) volunteered as listeners and were paid for their partic-
ipation. They were aged from 19 to 45 years old (median:
28 years old). They were selected on the basis of the Spiel-
berger trait anxiety inventory [3]. They had to have a score
lower than 39 (indicated low trait anxiety).

Stimuli Thirty-two videos were generated, correspond-
ing to the 32 sounds described above. They showed a user
manipulating the Flops glass (see Figure 1). All the videos
were the same, except for the sounds. Each video was 8 s
long. All the soundtracks had been equalized for loudness in
a preliminary study (for we are not interested in this param-
eter). The levels of the sounds varied from 52 dB(A) to 79.9
dB(A) (median 71.5 dB(A)). The video showed a user tilting
three times the Flops glass: first, he drops slowly three
items out of the glass, then tilts the Flops glass more quickly
to increase the rate of times dropping out of the Flops glass.
A total of 28 items are dropped.

Apparatus The stimuli were amplified over a pair Yamaha
MSP5 loudspeakers. Participants were seated in a double-
walled IAC sound-isolation booth. The experiment was run
using the PsiExp v3.4 experimentation environment includ-
ing stimulus control, data recording, and graphical user in-
terface [17]. The sounds were played with Cycling’74 Max/
MSP version 5.0.6, with Jitter displaying the videos. The
scales were presented on an Elo Touch Screen. This allowed
the participants to interact with the interface by only touch-
ing the screen.

Procedure The participants were first presented with a text
explaining the procedure, and explaining the meaning of the
3 scales of valence, arousal and dominance. Then they were
presented with a selection of pictures from the IAPS set of
images [7], and required to report their emotional reactions.
Then, they were presented with all the videos played one
after the other. Finally, they watched again each video, and
had to report for each video how their emotional reactions.

The participants had to indicate their emotional reactions
by selecting an item on each of the three 9-point scales, us-
ing the Self-Assessment Manikins (SAM) [1]. The SAM is
non-verbal pictorial assessment technique that directly mea-
sures the pleasure, arousal, and dominance associated with
a person’s affective reaction. The SAM scales used in this
study are reported on Figure 4.

4.2 Analysis

For the 32 sounds, the standard deviation of the judgements
made by the participants varies from 1.27 to 2.2 for the va-
lence scale, from 1.38 to 2.13 for the arousal scale, and from
1.36 to 1.91 for the dominance scale, which is consistent
with the data gathered for the IAPS image sets [7]. This
indicates that the emotional reactions caused by the videos
tend to be rather consistent across the participants. It also
indicates that the videos have elicited emotional reactions in
the subjects. In the following, the judgements will therefore
be averaged across participants.

The distributions of the judgements averaged over the
25 participants are represented on Figure 5 for the three
scales, and for the 2 groups of sounds. The judgements on
the valence scale vary from 2.08 to 5.8 (the range of the
judgements is therefore 3.1 on a scale of 9) with an average
of 4.39, indicating that the participants have mainly used
the center of the scale for all the videos. The judgements
are rather concentrated, and skipped toward the “unpleasant
part” of the scale. When considering separately the natu-
ral and synthetic sounds, it appears that the natural sounds

Figure 3. Example of the soundtrack of one of the videos
used in the experiment, together with the tilt angle of the
Flops glass.
Figure 4. Interface used to report the emotional reactions. The SAM is non-verbal pictorial assessment technique that directly measures the pleasure, arousal, and dominance associated with a person’s affective reaction.

have caused a rather neutral judgement on the valence scale among the participants (average = 4.9), whereas the synthetic sounds have caused a slightly more unpleasant feeling in the participants (average = 3.9). A Student t-test indeed reveals that the averages of these two distributions of judgements are significantly different ($t(30)=3.81, p<0.01$).

The judgements of arousal are also concentrated on the middle of the scale. Across all the 32 videos, the judgements of arousal vary from 4 to 7.04 (the range of the judgements is therefore 3.0 on a scale of 9), with an average of 5.06: the participants have not used the whole range of the scale. The averages of the distributions of judgements for the two sets of 16 sounds are not statistically significant: $t(30)=-0.93, p=0.37$. Overall, the sounds have caused a rather medium arousal in the participants.

The judgements of dominance vary from 3.96 to 6.6 (the range of the judgements is therefore 2.7 on a scale of 9) for all the 32 sounds, with an average value of 5.60. For this scale also, the participants have used a narrow range of the scale, slightly skipped toward the “in control” part of the scale. The averages of the distributions of judgements for the two sets of 16 sounds (natural vs. synthetic) are significantly different ($t(30)=3.80, p<0.01$). The average dominance judgement is 5.91 for the natural sounds, and 5.23 for the synthetic sounds. The participants have therefore felt slightly more in control when watching the videos with the natural sounds than the videos with the synthetic sounds, yet this difference is small.

The judgements on three scales are significantly correlated: Valence vs. Arousal, $r(30)=-0.78, p<0.01$ Valence vs. Dominance $r(30)=0.92, , p<0.01$, Arousal vs. Domi-

Figure 5. Bar plots of the distributions of the judgements on the three scales (valence, arousal, dominance), averaged across the participants, for the two groups of 16 sounds (natural vs. synthetic). A Gaussian curve with the same average and standard deviation as the distributions of judgements is also represented on top of each bar plot.

nance, $r(30)=-0.76, p<0.01$. This indicates that the participants have not used the three scales independently. There are systematic patterns in the judgements: the videos causing emotional reactions judged as pleasant tended to cause at the same time calm and dominant feelings, whereas the
videos causing emotional reactions judged as unpleasant tended
to cause systematically feelings judged as excited and domi-
nated. This was confirmed by the informal post experimen-
tal interviews with the participants: many participants had
indeed spontaneously indicated that “shrill” sounds tended
to irritate them (i.e. unpleasant, excited and dominated judg-
ements), whereas “soft and low” sounds tended to be felt as
more relaxing (pleasant, calm, and in-control).

These observations are further confirmed when studying
the correlations between the judgements on the scales and
several acoustic descriptors. We tested not only the spec-
tral centroid measure and tonality index (which were used
to create and select the sounds), but also all the descriptors
contained in the IcamDescriptor toolbox [11]. Overall, the
judgements on the three scales are correlated with the many
variants of the spectral centroid, the best correlations be-
ing obtained by the “perceptual” spectral centroid. This de-
scriptor is computed as the centroid computed using the spe-
cific loudness of the Bark scale [11]. The correlations with
the three scales is statistically significant: Valence: \( r(30)=
0.63, \ p<0.01 \); Arousal: \( r(30)=0.64, \ p<0.01 \); Dominance,
\( r(30)=0.48, \ p<0.01 \). Interestingly, these correlations are
different for the two groups of sounds: whereas the correla-
tions are very good for the natural sounds (Valence: \( r(30)=
0.83 \); Arousal: \( r(30)=0.81 \); Dominance: \( r(30)=-0.72 \); each:
\( p<0.01 \)), the judgements are much more spread for the syn-
thetic sounds (Valence: \( r(30)=0.64 \); Arousal: \( r(30)=0.57 \);
\( p<0.01 \); Dominance: \( r(30)=-0.43 \); \( p<0.05 \)).

4.3 Discussion

The emotional reactions to the 32 videos extend over a rather
small portion of the valence-arousal-dominance space, and
are centered around the neutral positions of each scale. This
is not really surprising, because these sounds only vary along
basic acoustical properties. It could therefore not be ex-
pected that a set of videos displaying a user droppping virtual
items out of a glass would cause emotional reactions com-
parable with those caused by sounds or images with a strong
semantic content (e.g. violent images, etc.).

More problematic however are the correlations of the three
scales. It is obvious from the experimental results that we
have not succeeded in creating sounds that cause emotions
varying independently along the three dimensions of emo-
tions used here. On the contrary, systematic patterns of
judgements appear in the judgements: videos causing pleas-
ant emotions caused at the same time dominant reactions,
and conversely. This is further confirmed when considering
that no scale is correlated with any metric of ity. Indeed,
following the results of Västfjäll et al. [20], the sounds were
created along three aspects: naturalness, spectral centroid,
and tonality. This last parameter was assumed to be corre-
lated to the arousal judgements, which is not the case here.
A possible explanation is that tonality is probably a relevant
parameter for long and continuous sounds such as aircraft
noise, but not for short impact sounds. Note that other sys-
tematic patterns of variations in the arousal valence space
(i.e. “boomerang-shaped”) have also been reported for other
acoustic stimuli [2]. The variations of spectral centroids of the sounds are
fairly correlated with the judgements on the three scales. The two groups of sounds (natural vs. synthetic) have pro-
duced slightly different emotional reactions: natural sounds
are judged as causing more pleasant and more dominant
feelings than synthetic sounds. These two aspects of the
sounds are therefore here the predictors of the emotional re-
actions.

5 CONCLUSION

This article is the first in a series aiming at studying users’
emotional reactions when manipulating sound-mediated in-
teractive objects. A experimental object (the Flops glass)
was designed. It is a plastic glass capable of capturing the
extent to which it is tilted. The object implements the metaphor
of virtual items stored in it that are dropped when it is tilted.
Each item virtually dropped out of the Flops glass produces
sound when impacting the surface below.

The study reported here aimed at assessing the influence
of the sounds solely on users’ emotional reactions. Thirty-
two sounds were created, with the purpose of testing the va-
 Validity of the conclusions found in [20] for aircraft sounds,
when extended to the case of the impact sounds used in
the Flops glass. The sounds were therefore created so as
to vary along two acoustical parameters: spectral centroid,
and tonality index. Two kinds of sounds were used: records
of “natural” impacts, and samples of sounds synthesized by

 various additive-substractive algorithms. The three parame-
ters investigated had previously been found to influence the
emotional reactions of participants listening to the aircraft
sounds.

In the experimental study conducted here, 25 participants
watched videos of a user manipulating the Flops glass. These
videos were all the same, except for the sounds. They had
to report their emotional reactions on three scales: valence,
arousal and dominance. They used the Self-Assessment Manikin
proposed in [1]. The results show that the two types of
sounds influenced the emotional reactions: natural sounds
were found to be slightly more pleasant, and they caused
participants to feel more in control than synthetic sounds.
These conclusions are consistent with those found in [20].
However, no scale appeared to be correlated with any de-
scriptor related to the tonality of the sounds. Furthermore,
the judgements on the three scales were correlated, indicat-
ing that the sounds caused emotional reactions that varied
along a single axis: from pleasant, calm, and in-control feel-
ings, to unpleasant, exciting and dominated feelings. This
suggests a single positive-negative dimension. However, it
can be noted that whereas natural sounds caused more pleasant feelings than synthetic sounds, they did not cause calmer feelings, indicating that the participants were able to distinguish the three scales.

These conclusions offer interesting results for followup studies on emotional reactions to sound-mediated interactive objects. They allow, for instance, to select sounds causing negative, neutral or positive reactions. The next study that is planned will address the influence of the usability of the interface on users’ emotional reactions. A interesting related issue is the interaction between emotional reactions caused by the sounds, and those caused by the usability.

Acknowledgments

This work was funded by the EU project CLOSED FP6-NEST-PATH no. 29085. The glass was firstly designed as a part of a Flops installation where three glasses were used. The development of this project was also supported by the Interstices Research Group. The glass was produced by Kar men Franinovic and Yon Visell, with the support of Martin Peach, Fabienne Meyer and Thomas Tobler.

6 REFERENCES


