Perception of Congruent Facial and Kinesthetic Expressions of Emotions

Yoren Gaffary, Jean-Claude Martin, Mehdi Ammi
Univ. of Paris-Sud
CNRS-LIMSI
Orsay, France
Email: firstname.lastname@limsi.fr

Abstract—The use of virtual avatars, through facial or gestural expressions, is considered to be a main support for affective communication. Currently, different works have studied the potential of a kinesthetic channel for conveying such information. However, they still have not investigated the complementarity between visual and kinesthetic feedback to effectively convey emotion. This paper studies the relation between some emotional dimensions and the visual and kinesthetic modalities. The experimental results show that subjects used visual and kinesthetic feedbacks to evaluate the pleasure and the arousal dimensions, respectively. We also observed a link between the recognition rate of emotions expressed with the visual modality (resp. kinesthetic modality) and the magnitude of that emotion’s pleasure dimension (resp. arousal dimension). These different results should help in the selection of feedback according to the features of the investigated emotion.

Keywords—Facial Expression; Haptic Expression; Multimodal Perception

I. INTRODUCTION

The expression of emotion requires a synchronization between different modalities [1]. The haptic expression of emotion has received less attention than other modalities, such as facial expressions [2] or gestural expressions [3].

Several studies have investigated the potential of the haptic channel for affective communication. For instance, Hertenstein et al. [4] show how haptic expressions can convey effectively several emotions through a direct contact between two humans. Olausson et al. [5] highlight specific biological systems dedicated to the expression of emotions through this modality. While the tactile expression of emotions received recently a lot of interest, the kinesthetic expression of emotions remains less explored. Bailenson et al. [6] propose a complete platform for recording and rendering affective expressions with a kinesthetic channel through motor expressions.

Other research has tried to combine these different modalities to improve the recognition and discrimination of some emotions, especially those presenting an ambiguity with other close emotions [7]. The perceptual component is an important part of the expression of emotion [8]. However, these works are based on basic combinations of visual and haptic expressions. They did not investigate how these modalities complement each other. This lack of studies is surprising since haptic and facial expressions are a natural combination with which to express emotion. For instance, Bickmore et al. [9] highlight some non-significant tendencies of complementarity between the visual and haptic modalities for a small set of expressions (e.g., three facial expressions: positive, neutral and negative).

In our previous work [10], we conducted a study where we investigated the recognition rate of emotions depending on the modality of expression. This paper presents a statistical study of the relations between some dimensions of emotion and visual-kinesthetic modalities, which were not previously investigated. It explores how the facial and kinesthetic expressions of emotions are perceived by participants when those expressions are presented separately or simultaneously. This paper also aims at highlighting the most relevant modality for an efficient recognition of some dimensions of emotion. These different results provide some cues for the choice of modalities with which to communicate these dimensions.

Section II describes the context and investigated emotions. Section III presents the objectives and hypotheses of the study. Section IV presents the experimental study and Section V, the results. Finally, Sections VI and VII discuss the results and present some open research perspectives.

II. CONTEXT AND INVESTIGATED EMOTIONS

A. Dimensional Representation Of Emotions

The dimensional representation of emotions (typically using the dimensions of pleasure, arousal, and dominance (PAD)) has been suggested as a suitable model for capturing the complexity of everyday emotions [11].

Our study is based on the PAD space [12]. This representation suggests that emotions can be defined by three independent dimensions. The pleasure dimension describes the degree of well-being. The arousal dimension describes the degree of mental or physical activity. The dominance dimension describes the degree of the control for a situation. Figure 1 presents the PAD space with three different emotions (Joy, Disgust and Rage).

B. The Investigated Emotions

The study focuses on eight emotions. We selected these emotions because they cover the regions of the PAD space with a balanced representation of the three dimensions [13]:
we will identify which modality is used for the evaluation of each dimension.

Based on these different objectives, we propose the following hypotheses:

H1 The dimensional ratings of emotions using simultaneous visual and kinesthetic stimuli provides values that fall between the ratings obtained by the Visual only and the Kinesthetic only. This means that participants combine the different information when presented simultaneously. This hypothesis is based on the fact that in the multimodal condition, the participants know that both visual and kinesthetic expressions are intended to express the same emotion.

H2 The subjective evaluation of some dimensions favour one modality (Visual or Kinesthetic). More specifically:

H2.A The evaluation of the pleasure dimension mainly uses the Visual modality.

H2.B The evaluation of the arousal dimension mainly uses the Kinesthetic modality.

Studies on the tactile expression of emotions claim that this modality enables to convey the pleasure dimension of emotions [15]. However, previous studies on the kinesthetic expression of emotions highlighted the potential of facial expressions for effectively conveying the pleasure dimension, and the role of physical expressions, such as kinesthetic, to effectively convey the arousal dimension [16].

IV. EXPERIMENTAL STUDY

A. Experimental Platform

Our experimental platform features a Geomagic Touch X device. This haptic device enables the playing of 3D kinesthetic expressions. Participants held the device as if they were holding someone’s wrist. The device moved their wrist in correspondence with a previously recorded expression [13]. A desktop screen displays the facial expressions. The MARC framework was used to generate the static facial expressions [17].

The haptic device was connected to a dedicated computer (server node) in order to prevent haptic instabilities due to the calculation latency of the manager module. The haptic device is controlled with a low-level module (haptic module) based on the OpenHaptics library. The client–server configuration exploits an UDP connection in a local network. The average delay time is approximately 32 ms.

B. Method

1) Participants: Forty-one people (10 women and 31 men), between 20 and 62 years old (28 years old on average, SD = 9), participated in the experiment. Thirty-six of them were right-handed. Thirty-two of them had received a European education.
2) Conditions: Three conditions have been investigated:

- In the Visual only condition (V), a facial expression of emotion is presented (no kinesthetic expression).
- In the Kinesthetic only condition (K), a kinesthetic expression of emotion is presented (no visual expression).
- In the Visual-Kinesthetic condition (VK), congruent facial and kinesthetic expressions of the same emotion are presented.

3) Measures: To investigate the different hypotheses, we analyzed the following measurements:

- M1 Reported PAD dimension: this concerns the perception of the different PAD dimensions (pleasure, arousal and dominance). The participants rated the presented expressions according to the PAD dimensions using a five point Likert scale (1: very low value and 5: very high value). This measure concerns H1 and H2.

- M2 Recognized emotion for each expression: the participants reported the emotion category perceived in the expression(s). They could choose one category from among eight possible categories. This measure aims to investigate the evolution of recognition rate in function of the dimensions of emotions.

4) Procedure: The participants started by reading a document describing 1) the context and objective of the experiment, 2) the investigated emotions, and 3) the dimensional description of the emotions (PAD space).

After reading the document, the participants were seated in front of a desk on top of which were a screen and a haptic device. They were left alone in the room. The GUI interface displayed a series of instructions during the experiment. At the beginning, a text explained that a series of expressions corresponding to different emotions would be displayed with three types of stimuli (V, K and VK). The order of the conditions (V, K and VK) was counterbalanced across participants.

After a short training session with the three conditions, a series of eight expressions (V, K or VK) were presented to participants. The participants were asked to hold the haptic device as if they were holding someone’s else wrist (see Figure 3a). The screen displayed facial expressions. The haptic device reproduces previously recorded kinesthetic expressions [13]. For the multimodal condition, both facial expressions and congruent kinesthetic expressions were rendered simultaneously. After each stimulus, the participant completed a form with the GUI asking for the recognized emotion(s) from a list of the eight possible emotions (M2) and PAD perception (M1; each axis rated on a five-point Likert scale).

At the end of the experiment, the score for each condition was displayed.

V. RESULTS

A. Comparing The Visual-Kinesthetic Condition With The Visual-only And Kinesthetic-only Conditions

Table I shows the average value of evaluation (M1) of the three dimensions of emotions (pleasure, arousal, dominance)
according to three conditions (V, K, VK). We observed that the majority of values of the VK condition are between the values of the V and K conditions.

In order to verify this observation, we perform an analysis based on the comparison of distances (dist) between the values of the three conditions. For each emotion, three distances are computed. Each one corresponds to the distance between two conditions (C1, C2); V–K, V–VK, and K–VK (see Figure 4a). The calculation of the distance is based on the following equation:

\[
dist_{ei}(C1, C2) = \sqrt{(C1_{e1} - C2_{e1})^2 + (C1_{e2} - C2_{e2})^2 + (C1_{e3} - C2_{e3})^2}
\]

where CI \_i corresponds to the average evaluation value of the emotion e (e={Joy, Elation, Disgust, Contempt, Anxiety, Fear, Irritation, Rage}) for the dimension i (i={pleasure, arousal, dominance}) according to the condition I (I \_i \in \{1, 2\})

Figure 4a shows an example of the distances computed for Joy, and Figure 4b shows the distances for all emotions. Based on these distances, we perform a Wilcoxon signed-rank test in order to compare the three sets of distances (dist_{e1}(V, K), dist_{e2}(V, VK), dist_{e3}(K, VK), e={Joy, Elation, Disgust, Contempt, Anxiety, Fear, Irritation, Rage}).

In order to check if the VK condition gives intermediate results and the VK are at similar distances from the V and K conditions.

These criteria mean that the V and K conditions present intermediate and close to the highest value of distance observed, see Figure 4b).

The results of the statistical test show:

- A significant difference between dist(V, K) and dist(V, VK) (p = 0.0499, W = 51.). Moreover, Figure 4b shows that the dist(V, K) presents a greater value than dist(V, VK) (\_i = 0.50 and \_i = 0.30 respectively).

- A significant difference between dist(V, K) and dist(K, VK) (p = 0.0499, W = 51.). Moreover, Figure 4b shows that dist(V, K) presents a greater value than dist(V, VK) (\_i = 0.50 and \_i = 0.31 respectively).

- A non-significant difference between dist(V, VK) and dist(K, VK) (p = 0.72, W = 28.) around \_i = 0.30.

As the two first criteria are statistically satisfied, and there is no significant difference between dist(V, VK) and dist(K, VK), the hypothesis H1 is validated. The dimensional rating of emotions using simultaneous visual and kinesthetic stimuli provides for five emotions values that fall between the ratings obtained by the Visual only and the Kinesthetic only stimuli (except for elation, contempt and rage, which remain close to the highest value of distance observed, see Figure 4b). This is above chance level.

978-1-4799-9953-8/15/$31.00 ©2015 IEEE
of the arousal dimension in the Visual-Kinesthetic condition presents similar results to the Kinesthetic only condition but a significant difference from the Visual only condition.

Table II shows that for four emotions (Elation, Disgust, Anxiety, Irritation), the visual modality is dominant for the evaluation of the pleasure dimension. This result supports H2.A. For other emotions (Joy, Disgust, Contempt, Rage), the kinesthetic modality is predominant for the evaluation of the arousal dimension. This result supports H2.B. No result was highlighted for the dominance dimension. These statistical results simultaneously observed with different emotions highlighted a global tendency for the link between some modalities and the evaluation of dimensions, which supports H2. The subjective evaluation of some dimensions favours one modality (Visual or Kinesthetic).

These results highlighted relations between the modality of expression and the evaluation of the dimensions of emotions. The next section aims to identify possible correlations between the recognition rate, the modality of expression and the dimensions of emotions.

C. Relations Between Recognition Rates And Emotional Dimensions

The previous section showed that the participants rely on different modalities for the perception and evaluation of different emotional dimensions. The current section discusses how the objective recognition rate of an emotion depends on the modality of its expression and its dimensional values.

Figure 5a provides the recognition rate (M2) obtained for visual expressions of emotions according to the magnitude of the pleasure. Figure 5b provides the recognition rate (M2) obtained for kinesthetic expressions of emotions according to the magnitude of the arousal. These recognition rate correspond to the number of people who successfully recognized the emotion, normalized between 0 and 1.

Spearman’s test highlighted the better recognition rate of emotions conveyed with visual expressions and presenting important magnitudes of pleasure ($p = 0.046, \rho = 0.74$). Similarly, we observed a tendency for a better recognition rate of

### TABLE I: Averages and standard deviations of PAD dimension perceptions by participants. Results are presented for each condition and each emotion. The high standard deviations observed are mainly due to the five point Likert scale. Gray cells in the multimodal condition represent values inside the visual and kinesthetic results (71% of the multimodal results).

<table>
<thead>
<tr>
<th>Emo</th>
<th>VP</th>
<th>VA</th>
<th>VD</th>
<th>KVP</th>
<th>KVA</th>
<th>KVD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joy</td>
<td>-</td>
<td>0.31</td>
<td>0.35</td>
<td>0.28</td>
<td>0.29</td>
<td>0.23</td>
</tr>
<tr>
<td>Anx</td>
<td>-</td>
<td>0.40</td>
<td>0.55</td>
<td>0.45</td>
<td>0.44</td>
<td>0.49</td>
</tr>
<tr>
<td>Irr</td>
<td>-</td>
<td>0.46</td>
<td>0.50</td>
<td>0.41</td>
<td>0.41</td>
<td>0.40</td>
</tr>
<tr>
<td>Fea</td>
<td>-</td>
<td>0.004</td>
<td>0.01</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>Cont</td>
<td>-</td>
<td>0.001</td>
<td>0.01</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>Dis</td>
<td>-</td>
<td>0.001</td>
<td>0.01</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>Conept</td>
<td>-</td>
<td>0.001</td>
<td>0.01</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>Elation</td>
<td>-</td>
<td>0.001</td>
<td>0.01</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>Rage</td>
<td>-</td>
<td>0.046</td>
<td>0.05</td>
<td>0.041</td>
<td>0.041</td>
<td>0.041</td>
</tr>
</tbody>
</table>

### TABLE II: Comparison of the perception of emotions between each monomodal condition (V or K) and the VK condition. The p-value is indicated in case of significant difference. When a modality (V or K) is different from the VK perception, and at the same time the other one is similar to the VK perception, this last modality is dominant for the evaluation of the corresponding dimension in the multimodal expressions (gray cells).

(a) Average recognition rate in the visual only condition, as a function of the magnitude of pleasure.

(b) Average recognition rate in the kinesthetic only condition, as a function of the magnitude of arousal.

Fig. 5: Average recognition rate in the V and K conditions depending on the dimensions of the PAD space. We observe a correlation in both cases.
emotions conveyed with kinesthetic expressions and presenting important magnitudes of arousal ($p = 0.062, \rho = 0.68$). Thus, the recognition rate of emotions expressed with a given modality (visual or kinesthetic respectively) appears to be linked with the magnitude of a specific dimension (pleasure or arousal, respectively). However, no correlation was highlighted for the dominance dimension.

VI. DISCUSSION

The comparison of the distances of the perceived dimensions between the three conditions (V, K, VK) (See Section V-A and H1) revealed that the VK condition gives results intermediate between those of the V and K conditions. This means that if an emotion is perceived with a given PAD evaluation for the kinesthetic modality and with another PAD evaluation for the visual modality, the multimodal configuration (VK) provides something intermediate between the monomodal evaluations. This suggests that the visual and the kinesthetic modalities might not be equal in terms of their ability to convey the different emotions.

More precisely, the second result (See Section V-B and H2) shows that participants use differently the visual and kinesthetic modalities. The participants seem to rely on the visual modality for the evaluation of the pleasure dimension and on the kinesthetic modality for the evaluation of the arousal dimension. This result is in line with previous work [9] and suggests that modalities are interpreted in complementary.

The last result (See Section V-C) highlights a link between the recognition rate of emotions expressed with the visual modality (resp. kinesthetic modality) and the magnitude of the pleasure dimension (resp. the arousal dimension). The higher the magnitude, the better the recognition rate. This result highlighted an approach for improving the discrimination between some close emotions according to the pleasure dimension. In fact, if these emotions present an important arousal, the addition of a suitable kinesthetic expression should improve the level of recognition rate compared to facial expression alone.

VII. CONCLUSION

This paper investigated the complementarity of the visual and kinesthetic modalities for affective communication. It is important to understand this coupling for an efficient communication of emotions in mediated systems.

The experimental results showed that congruent multimodal combination provides an intermediate subjective evaluation between monomodal configurations. This result suggests that subjects tend to merge the different expressions to evaluate emotions. The individual study of the emotions’ dimensions showed that subjects used the visual modality to evaluate the dimension of pleasure while the kinesthetic modality was used to evaluate the dimension of arousal. Finally, the analysis of the objective results highlighted a link between the magnitude of an emotion’s dimensions and the level of its recognition rate. Emotions presenting important magnitudes of pleasure (respectively arousal) are better recognized with the visual (respectively with kinesthetic) modality. Thus, the choice of the modality or multimodality depends on the expressed emotion to improve the correctness of the perception.

Future research will explore how to use these results to improve the recognition of close emotions according to the pleasure dimension and presenting similar visual expressions. This might help to support the integration of the kinesthetic modality in multimodal affective interfaces.

REFERENCES