

© IEEE, 2015. This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in Proceeding of International Conference on Affective Computing and Intelligent Interaction (ACII '15), <https://ieeexplore.ieee.org/document/7344643>
doi: 10.1109/ACII.2015.7344643

Perception of Intensity Incongruence in Synthesized Multimodal Expressions of Laughter

Radoslaw Niewiadomski*, Yu Ding[†], Maurizio Mancini*, Catherine Pelachaud[†], Gualtiero Volpe* and Antonio Camurri*

*Casa Paganini-InfoMus, DIBRIS

University of Genoa, Genoa, Italy

Email: radoslaw.niewiadomski@dibris.unige.it,

{maurizio.mancini;gualtiero.volpe;antonio.camurri}@unige.it

[†]CNRS-LTCI

Telecom-ParisTech, Paris, France

Email: {yu.ding;catherine.pelachaud}@telecom-paristech.fr

Abstract—In this paper, we study perception of intensity incongruence between auditory and visual modalities of synthesized expressions of laughter. In particular, we investigate whether incongruent expressions are perceived as 1) regulated, and 2) unsuccessful in terms of animation synthesis. For this purpose, we conducted a perceptive study with the use of a virtual agent. Congruent and incongruent multimodal expressions of laughter were synthesized from natural audiovisual laughter episodes, using machine learning algorithms. Next, the intensity of facial expressions and body movements were systematically manipulated to check whether the resulting incongruent expressions are perceived differently compared to the corresponding congruent expressions. Results show that 1) intensity incongruence lowers the perception of believability and plausibility, and 2) the incongruent laughter expressions displaying high intensity in the audio modality and low intensity in the body movement and facial expression are perceived as more fake than the corresponding congruent expressions. Such results have implications for both animation synthesis as well as expression regulation research.

Keywords—laughter; multimodal expressions; virtual agents; incongruence

I. INTRODUCTION

Regulation of emotion expression is an important part of human social life. It consists of “processes by which individuals influence (· · ·) how they experience and express (· · ·) emotions” [1]. Ekman and Friesen [2] proposed different types of expression regulation such as simulation (i.e., displaying a fake expression), inhibition, down regulation (suppression, deamplification) or up regulation (overacting, amplification) of an emotion. Expression regulation was studied mainly in the context of basic emotions (e.g., [3]). Within expression regulation only some elements of facial expression [3] and only some expressive modalities between audio, face, and body [4] can be voluntarily controlled, while others would leak the felt emotion. As a consequence, some form of expressive incongruence between the different body parts involved in expressing an emotion may emerge. For example, in case of down regulation some elements of the expression could convey the emotion at a higher intensity than the others. The intensity is controlled (in this case down regulated) by the displayer. In particular, in case of multimodal expressions the intensity incongruence between modalities may be a sign of emotion regulation (see [4]).

In this paper, we address the above-mentioned hypothesis on expression regulation taking laughter as a case of study. We propose a perceptive study, in which we exploit a virtual agent (VA) to evaluate congruent and incongruent multimodal laughter expressions in terms of emotion regulation. The goal of the study is twofold: 1) in order to shed light on the mechanisms of laughter expression, we study the relationship between perception of expression regulation and intensity incongruence, 2) with respect to the synthesis of expression, we check how sensitive users are to intensity incongruence between the visual and auditory modalities. In particular, we check how intensity incongruence influences human perception of believability, plausibility, and naturalness of the VA. This is an important question for synthesizing agents effectively displaying multimodal emotional expressions (especially when the different modalities are synthesized separately, e.g., see [5], [6]), because even a small intensity incongruence between the involved modalities may have very negative consequences on the perception of the virtual agent. Laughter was selected as a case of study since it is a strong signal of the emotional state of amusement¹. Consequently, it is quite often regulated [7] to avoid inappropriate laughter (e.g., in certain situational contexts, such as funerals), or on the contrary, to focus on a humorous aspect of a negative situation (e.g., to reduce stress). Being highly multimodal [8] laughter is an appropriate expression to study perception of incongruences between modalities.

This paper is organized as follows: the next section contains a survey of works on laughter regulation and on incongruence between modalities in expression synthesis; Section III describes our study on perception of laughter intensity incongruence; Section IV presents the results of this study. We present a general discussion in Section V and we conclude the paper in Section VI.

II. BACKGROUND

Not much is known about the relation between perception of expression regulation and intensity incongruence in the different modalities of laughter. Within one single modality, Lalot and colleagues [9] observed a lowering in the intensity of facial action units in explicitly down regulated expressions

¹Laughter may also have other meanings but in this paper we focus on amusement only.

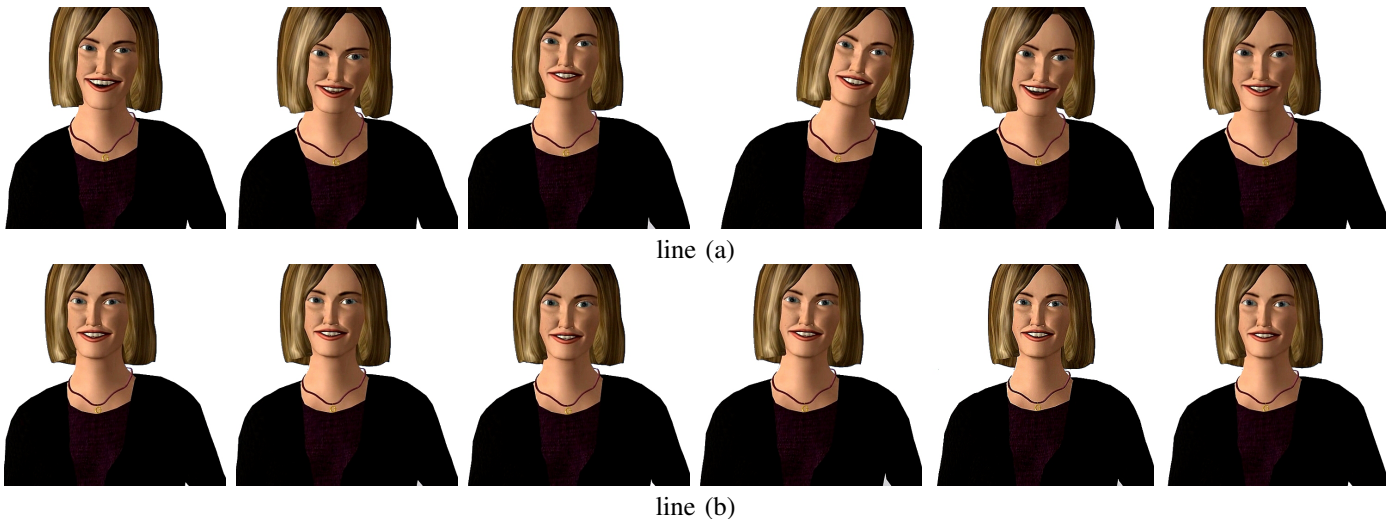


Fig. 1. Example of the generated animation from the high intensity audio episode: line (a), some frames extracted from high intensity face and body animations (i.e. face and body intensity congruent with the audio of laughter); line (b), the frames from low intensity face and body animation (i.e. face and body intensity incongruent with the audio of laughter).

of amusement. It was also shown that the sound of a fake laughter can be distinguished from the sound of an amused one [10].

Incongruence between visual modalities in synthesized expression of emotion was already a matter of research. Clavel and colleagues [11] studied the role of face and posture in the recognition of VAs' emotional expressions. Their results show that emotion recognition improves when facial and postural changes are congruent. The authors also observed that judgments were mainly based on the information displayed by the face, although adding congruent postures improved recognition. Gong and Nass [12] evaluated the trustiness and attitude towards multimodal stimuli, which are composed of real data displayed through one modality (respectively face or audio) and human-like (but artificial) data displayed through the other modality (respectively audio or face). The authors found that inconsistency between modalities caused stronger negative attitudes and less trust.

Regarding synthesis of regulated emotion expression, Niewiadomski and Pelachaud [13] proposed a model for emotion regulation based on fuzzy methods and the Ekman theory. They applied it to synthesize regulated emotion expressions appropriate to interpersonal relations. In a study on deceptive agents, Rehm and André [14] showed that users were able to differentiate between agents displaying an expression of felt emotion and an expression of fake emotion. The latter were synthesized according to Ekman's description.

Regarding the laughter synthesis, several models were proposed recently (e.g., [15], [16], [6], [17]). The role of wrinkles on a perceived meaning of synthesized laughter was showed in [18]. The multimodal expressions of laughter were also introduced to robots (e.g. [19], [20]).

III. EXPERIMENT

To test perception of intensity incongruence we conducted a study using a virtual agent (VA) to display different multimodal congruent and incongruent expressions of laughter. To

prepare the stimuli for our experiment we used the Greta VA [21], [6]. The VA allowed us to control precisely the conditions of the experiment - we modified, for example, the intensity of only one single modality. The evaluation was carried out with an online perceptive study. Our hypotheses targeted the perception of the intensity for congruent and incongruent synthesized expressions as well as their meaning. In particular, we were interested to find out whether incongruent expressions can be perceived as 1) regulated, 2) unsuccessful (in terms of animation synthesis).

A. Stimuli

We used a data-driven approach to create 8 experimental stimuli showing multimodal laughter expressions. We considered three different modalities: auditory, facial expressions, and body movements. The VA animations were generated from real audiovisual laughter episodes [6]. We chose the laughter episodes that correspond to the extreme ends of the intensity scale in the auditory modality i.e., the episodes whose audio intensity is perceived as definitely low or definitely high. It is important to notice that we were interested in the laughter sounds expressing a low/high intensity amusement, which are different from the laughter sounds having a low/high volume. In more details, the audio recordings of 19 episodes of one female subject were annotated by three experts who, independently, gave each episode an intensity score using a 5-points Likert scale. We selected the two episodes that received the lowest overall score and the two episodes that received the highest overall score. Such score was computed as the sum of the three independent scores. These four episodes were used to generate the animations used in the study. We used the original audio recording and its phonetic description generated with the algorithm proposed in [22]. The facial and body expressions were created with the machine learning models proposed by Ding et al. [6] (see Section III-B for more details). Starting from the phonetic description of laughter and further acoustic features, facial and upper body expressions were generated. The phonetic transcription was used to synchronize facial

and body expressions with laughter sound. An example of generated animation is shown in Figure 1.

Two sets (*CONG* and *INCONG*) of stimuli were built, each of them containing 4 animations created from the 4 audio stimuli (2 min and 2 max intensity): *CONG* contains animations of multimodal expressions where all three modalities are congruent, whereas *INCONG* contains animations of multimodal expressions where face and body express a different intensity than the audio. Using data-driven animation synthesis had an important advantage comparing with simple retargeting of the motion capture data: this approach allowed us to precisely control the intensity of the expression. At the same time the models were trained on a large number of laughter episodes that ensures a realistic synthesis.

B. Multimodal Laughter Synthesis

The synthesis animation model of laughter used for synthesizing the stimuli for the experiment is displayed in Figure 2. It includes three modules: 1) lip and jaw motions synthesis, 2) head motion and upper facial expression synthesis and 3) torso motion synthesis. More details on multimodal motion synthesis can be found in [6]. Here we only briefly review them. Lips motion is synthesized using a contextual Gaussian model (CGM). A CGM is learned for each pseudo-phoneme using Maximum Likelihood Estimation (MLE). CGM allows capturing the relation between a pseudo-phonemes sequence and the corresponding lip motions. Head and facial expressions are obtained by concatenated motion capture data. Motions are selected and concatenated to correspond to an input sequence of pseudo-phonemes. A cost function is used to find the best motions sequence. The torso motion synthesis model is based on a PD (proportional derivate) controller. The parameters used in these models are trained from a human laughter dataset. Head, facial expressions and lip motions are generated from laughter acoustic features, phoneme label, phoneme intensity and its duration. Although the information regarding laughter audio is not taken as input to the third model, the synthesized head motion is used to drive torso motion thus ensuring obtaining a synchronized multimodal animation of laughter.

C. Participants

The sample consisted of 84 adult participants from 22 countries (31 females, age 18-68; mean = 34.11, SD = 11, most frequent countries of origin: Italy 24%, France 13%, Poland 10%, Germany 8%).

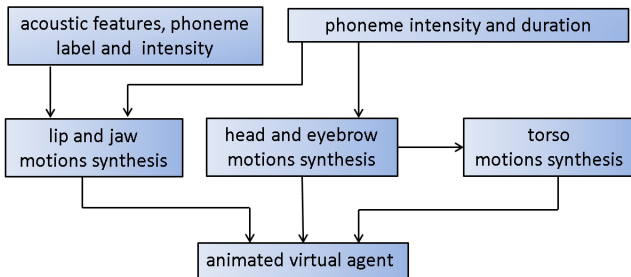


Fig. 2. Overall architecture of multimodal laughter synthesis.

D. Procedure

The perceptive study was carried out online, along a set of web pages enabling participants to evaluate animations. Each web page displayed one animation. At the beginning of the test, each participant was asked to provide her gender, age, and nationality. Each participant had to evaluate the whole set of stimuli. Participants could see the animations any times they wanted and they had to answer all questions before being able to see the next animation. The duration of the animations lasted from 7 to 12 seconds. The animations were displayed randomly. Participation was anonymous.

E. Hypotheses and Evaluation Questionnaire

Participants were asked to evaluate three characteristics of the laughter expression they had seen, using a 7-points Likert scale from “Very low” to “Very high”:

- (Q1) Naturalness,
- (Q2) Plausibility,
- (Q3) Believability.

They were also asked to answer questions on the perception of laughter regulation, using a 7-points Likert scale from “Definitely not” to “Definitely yes”:

- (Q4) Is the avatar freely expressing its amusement?
- (Q5) Is the laughter fake?
- (Q6) Does the avatar seem to restrain the laughter?

Our hypotheses were:

- H1. Congruent expressions (*CONG*) are more natural (Q1), more plausible (Q2), and more believable (Q3) than incongruent ones.
- H2. Incongruent expressions (*INCONG*) increase the perception of laughter regulation.

We expected that there is a significant difference in the perception of plausibility, believability, and/or naturalness between *CONG* and *INCONG* expressions. The intensity incongruence between modalities may lower the perception of naturalness, plausibility, and believability. Incongruent expressions can be seen as 1) expressions that were the result of unsuccessful animation synthesis, 2) expressions that emerged from the emotion regulation. Consequently, they might be considered less believable or plausible.

We also expected to find a significant difference in the responses to questions Q4 - Q6 between *CONG* and *INCONG* animations. Questions Q4 - Q6 focus on different forms of emotion regulation. According to Ekman and colleagues (e.g., [23], [24], [4], [3]) people who mask or display fake emotions can control only some parts of the expression. Although they did not address explicitly multimodal expression of laughter, we supposed that multimodal incongruent expressions of laughter i.e., with lowered or increased facial and body intensity (compared to the laughter audio) are perceived as more restrained, more fake, and less freely expressing amusement. We expected that participants may try to explain to themselves the observed intensity incongruence by assuming that one modality is voluntarily controlled and, thus, the VA is trying to regulate its expressions.

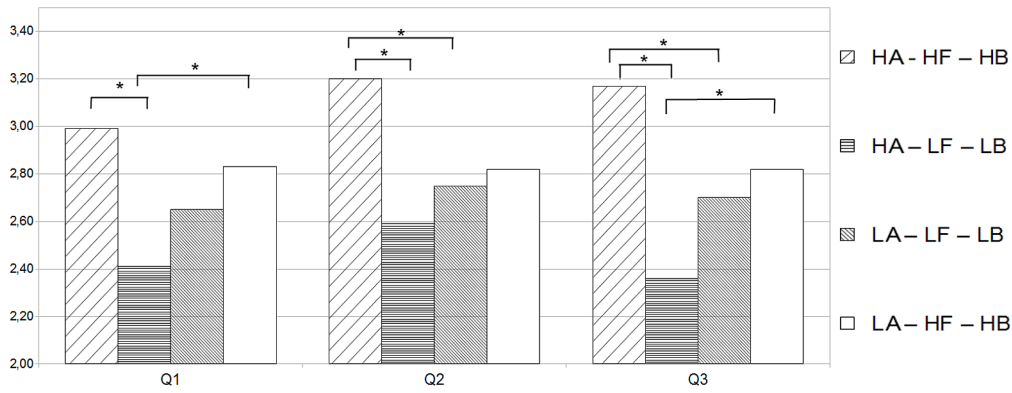


Fig. 3. Results of perceived naturalness (Q1), plausibility (Q2), believability (Q3). Notation: HA - high intensity audio, LA - low intensity audio, HF - high intensity face, LF - low intensity face, HB - high intensity body, LB - low intensity body. The significant differences are signed with “*”.

IV. RESULTS

First, we checked the effect of participant’s gender on the animations evaluation. For this purpose, an ANOVA was conducted with one independent between-subjects variable Gender and 12 dependent variables (results Q1-Q6 for *CONG* and *INCONG* animations). Between-subjects analysis did not show a significant main effect of Gender ($F(12, 71) = 1.001; p = .457, Wilk's \lambda = .855$). Consequently, the data of female and male participants were considered altogether in all the remaining analyses.

The remaining results reported in this section were computed with two-way (2×2) ANOVAs with Modality (congruent vs. incongruent), and Intensity (high vs. low) as independent variables (see Table I and Figures 3 and 4 for detailed results).

Naturalness. The results nearly showed a significant effect of Modality ($F(1, 83) = 3.927, p = .051$) but no effect of Intensity ($F(1, 83) = .066, p = .798$). The two-way Modality \times Intensity interaction ($F(1, 83) = 10.410, p < .01$) was significant. Next, the Modality \times Intensity interaction effect was analyzed using a post hoc tests with LSD adjustment. The effect of Modality was significant for the high intensity expressions ($F(1, 83) = 14.108, p < .001$), but not for the low intensity ones ($F(1, 83) = 1.325, p = .253$). The effect of Intensity was significant for the incongruent expressions ($F(1, 83) = 5.014, p < .05$), but not for the congruent ones ($F(1, 83) = 3.005, p = .087$).

For the high intensity expressions, lowering the intensity of the body and face displays lowered their perceived naturalness. The opposite situation, i.e., increasing the intensity of face

and body displays in the low intensity stimuli did not have any effect on the perception of naturalness. The low intensity stimuli both congruent and incongruent were perceived equally natural. Additionally, the low intensity incongruent stimuli were considered more natural than the high intensity incongruent ones. Between the congruent expressions both the high and low intensity ones were perceived equally natural. Thus, perception of naturalness is strongly influenced by lowering the intensity of the visual modalities.

Plausibility. The results of ANOVA showed an effect of Modality ($F(1, 83) = 8.682, p < .01$) and no effect of Intensity ($F(1, 83) = .592, p = .444$). The two-way Modality \times Intensity interaction ($F(1, 83) = 9.580, p < .001$) was significant. Next, the Modality \times Intensity interaction effect was analyzed using a post hoc test with LSD adjustment. The effect of Modality was significant for the high intensity expressions ($F(1, 83) = 16.960, p < .001$), but not for the low intensity ones ($F(1, 83) = .228, p = .634$). The effect of Intensity was significant for the congruent expressions ($F(1, 83) = 6.041, p < .05$), but not for the incongruent ones ($F(1, 83) = 1.620, p = .207$).

Again, for the high intensity expressions, lowering the intensity of body and of face displays lowered the perceived plausibility. The opposite situation, i.e., increasing the intensity of face and of body displays in the low intensity stimuli did not have any effect on the perception of plausibility. Both congruent and incongruent low intensity stimuli were perceived as equally plausible. Additionally, the high intensity congruent stimuli were considered more plausible than the low intensity congruent ones, while no difference was observed between the incongruent expressions.

Believability. The ANOVA results showed an effect of Modality ($F(1, 83) = 16.343, p < .001$) and no effect of Intensity ($F(1, 83) = .000, p = .984$). The two-way Modality \times Intensity interaction ($F(1, 83) = 16.872, p < .001$) was significant. Next, the Modality \times Intensity interaction effect was analyzed using a post hoc test with LSD adjustment. The effect of Modality was significant for the high intensity expressions ($F(1, 83) = 30.315, p < .001$), but not in the low intensity ones ($F(1, 83) = .776, p = .381$). The effect of Intensity was significant for the congruent expressions ($F(1, 83) = 6.278, p < .05$) and for the incongruent ones ($F(1, 83) = 5.903, p < .05$).

TABLE I. RESULTS OF PERCEIVED NATURALNESS (Q1), PLAUSIBILITY (Q2), BELIEVABILITY (Q3), “FREELY EXPRESSED AMUSEMENT” (Q4), SIMULATION (Q5), AND RESTRAINT (Q6).

Intensity	Congruent		Incongruent	
	High	Low	High	Low
Audio Face and Body	High	Low	High	High
Question	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Q1	2.99 (1.33)	2.65 (1.47)	2.41 (1.49)	2.83 (1.38)
Q2	3.20 (1.30)	2.75 (1.48)	2.59 (1.43)	2.82 (1.38)
Q3	3.17 (1.32)	2.70 (1.45)	2.36 (1.38)	2.82 (1.35)
Q4	3.70 (1.38)	1.96 (1.40)	2.71 (1.55)	2.42 (1.32)
Q5	2.52 (1.37)	2.64 (1.44)	3.04 (1.50)	2.55 (1.48)
Q6	2.01 (1.41)	3.56 (1.45)	2.79 (1.60)	3.21 (1.49)

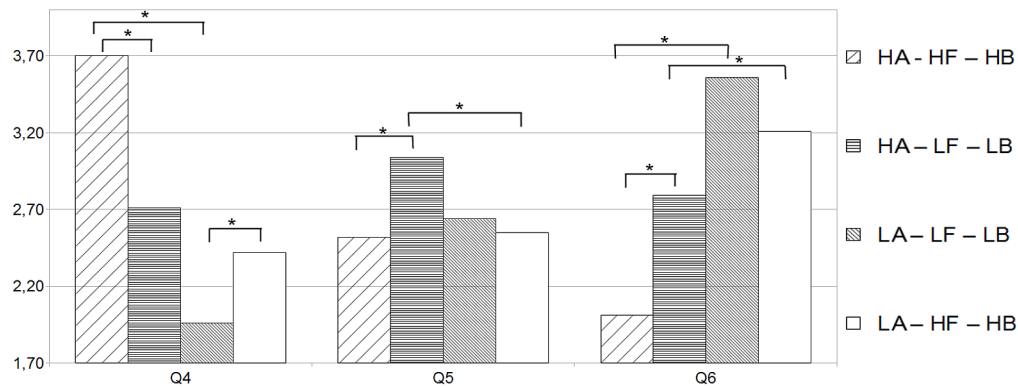


Fig. 4. Perception of “freely expressed amusement” (Q4), simulation (Q5), restraint (Q6). Notation: HA - high intensity audio, LA - low intensity audio, HF - high intensity face, LF - low intensity face, HB - high intensity body, LB - low intensity body. The significant differences are signed with “*”.

As in the case of questions Q1 and Q2, for the high intensity stimuli, lowering the intensity of body and of face displays influenced negatively the perception of believability. The opposite situation, i.e., increasing the intensity of body and of face displays in the low intensity stimuli did not have any effect on the perception of believability. Additionally, both types of stimuli with intense body and face expressions were perceived as more believable than the stimuli displaying low intensity body and face expressions.

Amusement. The results of ANOVA showed an effect of Modality ($F(1, 83) = 7.269, p < .001$) and Intensity ($F(1, 83) = 42.763, p < .001$). Also, the two-way Modality \times Intensity interaction ($F(1, 83) = 40.836, p < .001$) was significant. Next, the Modality \times Intensity interaction effect was analyzed using a post hoc test with LSD adjustment. The effect of Modality was significant for the high intensity expressions ($F(1, 83) = 36.023, p < .001$) and for the low intensity ones ($F(1, 83) = 12.033, p < .01$). The effect of Intensity was significant for the congruent expressions ($F(1, 83) = 78.861, p < .001$), but not for the incongruent ones ($F(1, 83) = 2.367, p = .128$).

For the high intensity stimuli, lowering the intensity of body and of face displays was perceived as less “freely expressing amusement”. Increasing the intensity of body and of face displays in the low intensity stimuli was perceived as more “freely expressing amusement”. The high intensity congruent stimuli were considered as more “freely expressing amusement” comparing to the low intensity congruent ones. Interestingly, the incongruent low and high intensity stimuli were perceived equally in this respect. Thus, the perception of “freely expressed amusement” was related to the intensity of the stimuli, with the congruent high intensity audio, body and face displays being the most “freely expressing amusement”, the incongruent expressions being the less “freely expressing amusement”, and the low intensity congruent expressions being the least “freely expressing amusement”.

Simulation. The results of ANOVA showed an effect of Modality ($F(1, 83) = 4.712, p < .033$), but not of Intensity ($F(1, 83) = 1.625, p = .206$). The two-way Modality \times Intensity interaction ($F(1, 83) = 5.491, p < .05$) was significant. Next, the Modality \times Intensity interaction effect was analyzed using a post hoc test with LSD adjustment. The effect of Modality was significant for the high intensity

expressions ($F(1, 83) = 9.370, p < .01$), but not for the low intensity ones ($F(1, 83) = .334, p = .565$). The effect of Intensity was not significant for the congruent expressions ($F(1, 83) = .343, p = .56$), but it was significant for the incongruent ones ($F(1, 83) = 6.129, p < .05$).

Within the high intensity stimuli, lowering the intensity of body and of face displays, resulted in expressions that were considered more fake than their congruent correspondents. The same stimuli were also perceived more fake than low intensity incongruent expressions (i.e., low intensity of audio, high intensity of body and of face displays). The effect of increasing the intensity of body and of face displays in the low intensity stimuli for the perception of fakeness was, however, not observed. Importantly, the intensity was not influencing the perception of fakeness of the congruent expressions. Thus, only lowering the intensity of body and of face displays provoked stronger perception of fakeness.

Restraint. The results showed the effect of Intensity ($F(1, 83) = 34.035, p < .001$) but not of Modality ($F(1, 83) = 2.657, p < .107$). The two-way Modality \times Intensity interaction ($F(1, 83) = 19.482, p < .001$) was significant. Next, this interaction effect was analyzed using a post hoc test and LSD adjustment. The effect of Modality was significant for the high intensity expressions ($F(1, 83) = 17.949, p < .001$), but not for the low intensity ones ($F(1, 83) = 3.492, p = .065$). The effect of Intensity was significant for the congruent expressions ($F(1, 83) = 52.626, p < .001$) and for the incongruent ones ($F(1, 83) = 4.060, p < .05$).

The low intense stimuli were perceived as more restrained than the high intensity ones. Additionally the high intensity stimuli with the decreased intensity of body and of face displays were perceived more restrained than their congruent high intensity correspondents. Even though the statistical significance was not reached, the low intensity stimuli with the increased intensity of body and of face displays seem to be perceived less restrained than their congruent low intensity correspondents ($p = .065$). It seems that the perception of restraint is related to the intensity of the stimuli, with the congruent high intensity expressions being perceived as the least restrained displays of amusement, the incongruent laughter expressions being perceived as more restrained, and the low intensity congruent expressions being perceived as the most restrained displays of amusement.

V. DISCUSSION

In this experiment, we evaluated the perception of intensity incongruence in synthesized multimodal laughter expressions. Hypothesis H1 was confirmed: participants evaluated congruent expressions to be more plausible and believable compared to incongruent ones. In particular, significantly lower scores were observed for expressions where both visual modalities (face and body) were displaying laughter at a lower intensity than the auditory modality. Participants were instead rather insensitive to body and facial intensity increment in the incongruent expressions. So, if laughter audio was expressing higher intensity than the other modalities, the scores for plausibility and believability were lowered. In the opposite situation, i.e., when high intensity face/body movements were accompanied by the audio expressing a lower intensity, the effect was not observed. This result can be explained if we consider the natural phenomenon of unvoiced laughter [25], i.e., when even high intensity face and body movements appear with no voice (or nearly no voice). Indeed, in the low intensity incongruent stimuli, the intensity of the audio remains low, and is accompanied by high intensity body and face movements. It is possible that participants considered such a low intensity incongruent laughs as unvoiced laughter and thus they did not consider it less believable or less plausible. We would like to further check this hypothesis in future works.

Regarding hypothesis H2, interesting differences were observed between congruent and incongruent expressions. In questions Q4 (freely expressed amusement) and Q6 (restraint) the relation to “global intensity” (understood as the sum of the intensities of all modalities) was observed. High intensity congruent expressions were considered the most “freely expressing amusement” and, at the same time, the least restrained; incongruent expressions were less “freely expressing amusement” and, at the same time, more restrained; low intensity congruent expressions were perceived as the “least freely expressing amusement” and, the most restrained among all the laughter expressions. In particular, contrary to our expectations the low intensity congruent expressions were perceived the least “freely expressing amusement” and the most restrained ones. An explanation for this result can be that, when the context is unknown (as in our experiment), people may have some expectations on the laughter intensity: i.e., the laughter is expected to be of high intensity. Consequently, the low intensity laughs would not be perceived as a reaction to a low intense stimulus (e.g., a joke that was not very funny), but as restraint of the amusement. This hypothesis should be further checked in the future works.

Interestingly, a different relation between incongruent and congruent expressions was observed in answers to the question Q5. Only decreasing the intensity of face and of body displays in the high intensity laughter expressions increased the perception of expression fakeness. Such high intensity incongruent expressions were perceived as more fake compared to both their congruent and incongruent correspondents. Again, a similar effect for increasing the intensity of body and of face displays was not observed. This result can be also explained with the unvoiced laughter hypothesis. Interestingly, Bachorowski and Owren [25] found that the attributed level of amusement is lower when the unvoiced laughter is displayed.

More general conclusions can be drawn regarding the

intensity incongruent laughter expressions. From our results, it seems that incongruent high intensity expressions were at the same time considered more fake and less plausible. Thus, in the case of incongruence the participants may perceive it in two ways: 1) the stimulus is showing some regulated expression (e.g., it is fake) or 2) the animation synthesis of face/body was unsuccessful (i.e., it is less plausible). This duality between the perception of expression meaning and its animation correctness should be addressed when developing more subtle (e.g., regulated) synthesized expressions. It might be that human users perceive the synthesized expressions displaying emotion regulation as the results of some errors in the animation synthesis. Importantly, the latter is different from the general quality of the animation (i.e., the rendering quality, including e.g., the use of textures, lighting, or shadows). We aim to check this hypothesis in future work by asking participants to explicitly evaluate synthesis quality.

VI. CONCLUSIONS

In this paper, we evaluated the perception of the quality and the meaning of intensity congruent and incongruent multimodal laughter expressions. According to our results, 1) intensity incongruence lowers the perception of believability and plausibility of laughter animations, 2) lowering the intensity of at least one modality lowers the perception of freely expressed amusement and increases the perception of restraint, 3) the incongruent laughter expressions composed of high intensity audio and low intensity body and face displays are perceived as more fake than their congruent correspondents. Thus, this case of intensity incongruence increases the perception of expression regulation. Our results target the perception of expression synthesis, as well as the knowledge about expression regulation in laughter. Whereas incongruent synthesized expressions are perceived less plausible and believable, they can also communicate expression regulation (and more precisely laughter simulation).

It is important to notice that the results of this study focuses on a passive perception of context-free expressions of amusement laughter. The perception of the laughter incongruence may, however, vary within the interaction and its context. Moreover, in this study, only one laughter meaning was considered, namely amusement laughter.

This work addressed new research questions that need to be further studied. In future works we plan to address two other aspects of intensity incongruence in laughter expressions. First of all, we would like to check our hypothesis regarding unvoiced laughter. Secondly, we would like to evaluate the role of single modalities (body or face) in the perception of laughter incongruent intensity expressions.

ACKNOWLEDGMENT

The research leading to these results has received fundings from the EU 7th Framework Programme under grant agreement n 270780 ILHAIRE, and the EU-H2020 under grant agreement n 645553 DANCE.

REFERENCES

- [1] J. J. Gross, “The emerging field of emotion regulation: An integrative review,” *Review of General Psychology*, vol. 2, no. 3, pp. 271–299, 1998.

- [2] P. Ekman and W. Friesen, "The repertoire of nonverbal behavior's: Categories, origins, usage and coding," *Semiotica*, no. 1, 1969.
- [3] P. Ekman, *The Face Revealed*. London: Weidenfeld & Nicolson, 2003.
- [4] P. Ekman, M. O'Sullivan, W. V. Friesen, and K. R. Scherer, "Invited article: Face, voice, and body in detecting deceit," *Journal of Nonverbal Behavior*, vol. 15, no. 2, pp. 125–135, 1991. [Online]. Available: <http://dx.doi.org/10.1007/BF00998267>
- [5] R. Niewiadomski, M. Mancini, Y. Ding, C. Pelachaud, and G. Volpe, "Rhythmic body movements of laughter," in *Proceedings of the 16th International Conference on Multimodal Interaction*, ser. ICMI '14, 2014, pp. 299–306. [Online]. Available: <http://dx.doi.org/10.1145/2663204.2663240>
- [6] Y. Ding, K. Prepin, J. Huang, C. Pelachaud, and T. Artières, "Laughter animation synthesis," in *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems*, ser. AAMAS '14, 2014, pp. 773–780.
- [7] N. R. Giuliani, K. McRae, and J. J. Gross, "The up- and down-regulation of amusement: Experiential, behavioral, and autonomic consequences," *Emotion*, vol. 8, no. 5, 2008.
- [8] W. Ruch and P. Ekman, "The expressive pattern of laughter," in *Emotion, qualia and consciousness*, A. Kaszniak, Ed. Tokyo: World Scientific Pub., 2001, pp. 426–443.
- [9] F. Lalot, S. Delplanque, and D. Sander, "Mindful regulation of positive emotions: a comparison with reappraisal and expressive suppression." *Frontiers in Psychology*, vol. 5, no. 243, 2014.
- [10] G. A. Bryant and C. A. Aktipis, "The animal nature of spontaneous human laughter," *Evolution and Human Behavior*, vol. 35, no. 4, pp. 327 – 335, 2014.
- [11] C. Clavel, J. Plessier, J.-C. Martin, L. Ach, and B. Morel, "Combining facial and postural expressions of emotions in a virtual character," in *Intelligent Virtual Agents*, ser. Lecture Notes in Computer Science, Z. Ruttkay, M. Kipp, A. Nijholt, and H. Vilhlmsson, Eds. Springer Berlin Heidelberg, 2009, vol. 5773, pp. 287–300.
- [12] L. Gong and C. Nass, "When a talking-face computer agent is half-human and half-humanoid: Human identity and consistency preference," *Human Communication Research*, vol. 33, no. 2, pp. 163–193, 2007. [Online]. Available: <http://dx.doi.org/10.1111/j.1468-2958.2007.00295.x>
- [13] R. Niewiadomski and C. Pelachaud, "Affect expression in ECAs: Application to politeness displays," *Int. J. Hum.-Comput. Stud.*, vol. 68, no. 11, pp. 851–871, Nov. 2010. [Online]. Available: <http://dx.doi.org/10.1016/j.ijhcs.2010.07.004>
- [14] M. Rehm and E. André, "Catch me if you can: Exploring lying agents in social settings," in *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*, ser. AAMAS '05. New York, NY, USA: ACM, 2005, pp. 937–944. [Online]. Available: <http://doi.acm.org/10.1145/1082473.1082615>
- [15] D. Cosker and J. Edge, "Laughing, crying, sneezing and yawning: Automatic voice driven animation of non-speech articulations," in *Proc. of Computer Animation and Social Agents*, 2009, pp. 21–24.
- [16] P. C. DiLorenzo, V. B. Zordan, and B. L. Sanders, "Laughing out loud: control for modeling anatomically inspired laughter using audio," *ACM Transactions on Graphics (TOG)*, vol. 27, no. 5, p. 125, 2008.
- [17] Y. Ding, J. Huang, N. Fourati, T. Artires, and C. Pelachaud, "Upper body animation synthesis for a laughing character," in *Intelligent Virtual Agents*, ser. Lecture Notes in Computer Science, T. Bickmore, S. Marsella, and C. Sidner, Eds. Springer International Publishing, 2014, vol. 8637, pp. 164–173.
- [18] R. Niewiadomski and C. Pelachaud, "The effect of wrinkles, presentation mode, and intensity on the perception of facial actions and full-face expressions of laughter," *ACM Transactions on Applied Perception*, vol. 12, no. 1, pp. 2:1–2:21, 2015. [Online]. Available: <http://doi.acm.org/10.1145/2699255>
- [19] C. Becker-Asano, T. Kanda, C. Ishi, and H. Ishiguro, "Studying laughter in combination with two humanoid robots," *AI & SOCIETY*, vol. 26, no. 3, pp. 291–300, 2011. [Online]. Available: <http://dx.doi.org/10.1007/s00146-010-0306-2>
- [20] T. Kishi, N. Endo, T. Nozawa, T. Otani, S. Cosentino, M. Zecca, K. Hashimoto, and A. Takanishi, "Bipedal humanoid robot that makes humans laugh with use of the method of comedy and affects their psychological state actively," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, May 2014, pp. 1965–1970.
- [21] R. Niewiadomski, M. Obaid, E. Bevacqua, J. Looser, L. Q. Anh, and C. Pelachaud, "Cross-media agent platform," in *Proceedings of the 16th International Conference on 3D Web Technology*, ser. Web3D '11. New York, NY, USA: ACM, 2011, pp. 11–19. [Online]. Available: <http://doi.acm.org/10.1145/2010425.2010428>
- [22] J. Urbain, H. Cakmak, and T. Dutoit, "Automatic phonetic transcription of laughter and its application to laughter synthesis," in *Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, ser. ACII '13. Washington, DC, USA: IEEE Computer Society, 2013, pp. 153–158. [Online]. Available: <http://dx.doi.org/10.1109/ACII.2013.32>
- [23] P. Ekman and W. Friesen, *Unmasking the Face. A guide to recognizing emotions from facial clues*. New Jersey: Prentice-Hall, Inc., Englewood Cliffs, 1975.
- [24] P. Ekman, *Telling lies: Clues to deceit in the marketplace, politics, and marriage*. W. W. Norton & Company, 1985.
- [25] J. Bachorowski and M. Owren, "Not all laughs are alike: voiced but not unvoiced laughter readily elicits positive affect," *Psychol Sci.*, vol. 12, no. 3, pp. 252–7, 2001.