Smile and Laughter in Human-Machine Interaction: a study of engagement

Mariette Soury [1,2], Laurence Devillers [1,3]

1 LIMSI-CNRS, BP133, 91403 Orsay cedex, France
2 University Paris 11, 91400 Orsay, France
3 University Paris Sorbonne 4, 28 rue serpente, 75005 Paris, France

E-Mail: soury@limsi.fr, devil@limsi.fr

Abstract

This article presents a corpus featuring adults playing games in interaction with machine trying to induce laugh. This corpus was collected during Interspeech 2013 in Lyon to study behavioral differences correlated to different personalities and cultures. We first present the collection protocol, then the corpus obtained and finally different quantitative and qualitative measures. Smiles and laughs are types of affect bursts which are defined as short emotional "non-speech" expressions. Here we correlate smile and laugh with personality traits and cultural background. Our final objective is to propose a measure of engagement deduced from those affect bursts.

Keywords: affect bursts, personality traits, sense of humor

1 Introduction

While spoken language constitutes a strong communication channel in human-machine interaction, previous works demonstrated that relevant information is simultaneously and spontaneously conveyed nonverbally (Campbell, 2007).

The experiment presented in this paper was designed to evaluate the impact of a user's individual characteristics (cultural background, sense of humor, personality traits) on his propensity to express amusement in front of a machine. Specifically, we focused on affect bursts as the expression of users' emotional state. The notion of *affect burst* has been introduced by Scherer (Scherer, 1994) as v*ery brief, discrete, nonverbal expressions of affect in both face and voice as triggered by clearly identifiable events"*. Affect bursts are defined as short emotional "non-speech" expressions interrupting speech.

Examples of expressions exclusively due to push effects are *affect bursts* (i.e., Krumhuber & Scherer, 2011) or infant grunts. Pursuing the ideas of Johnstone *et al.* (Johnstone *et al.*, 2000*)* about prototypes in emotion expression, it may be that some emotions are expressed through prototypical affect bursts, while others are not. A simple criterion for prototypes could be that they are spontaneously produced by speakers and easily identified by listeners. (Schroeder, 2003) shows that affect bursts, presented without context, can convey a clearly identifiable emotional meaning. This subject, although theoretically described in detail, does not seem to have been extensively studied experimentally.

In this study, we refer to smiles and laughs as affect bursts. Our purpose is to evaluate how these affect bursts can reveal the subject's engagement in the interaction[1]. Previous work by Tanaka (Tanaka et. al., 2010) investigated the impact of cultural differences on the perception of emotions; our assumption is that cultural differences impact the expression of emotions as well. In order to account for users' individuality in the expressive behavior, we needed large amount of multicultural data. Few existing resources are multicultural. In our previous experiments, we only used French real-life corpora recorded in call center (Devillers & Vidrascu, 2007) to study negative and positive laughter. In (Schroeder, 2003) the corpus used is in German language. We opted to collect a new audio and video corpus during the Interspeech 2013 conference.

The second section of this paper presents the protocol designed to acquire multimodal data in a man-machine interaction, and the resulting corpus is detailed in the third section. The correlations between participants' profile and the collected affect bursts are described in the fourth section.

2 Collection protocol

In this study, we used two tasks for inducing smiles and laughs:

- Passive induction tasks such as watching funny videos or listening to jokes

- Active induction tasks such as Tongue-Twisters game reading on a computer screen or repeating with the Nao robot (Aldebaran Robotics).

In the passive task, we intended to observe both reflex contagious laughter and cultural laughter (i.e. funny scenes with various cultural backgrounds ("cognitive" amusement, irony). Laugher is a powerful social signal. Laugher is "contagious": seeing or hearing someone laugh can trigger our own amusement in a reflexive manner. Laughter can also be the cognitive result of the interpretation of a ridiculous or funny situation. The sense of humor is both personal and cultural.

We also used tongue-twisters game with the robot NAO in several languages as an active task inducing laughter. Tongue-twisters are sentences which exist in most languages, specifically designed to have a difficult pronunciation, by alternating similar phonemes (e.g. *"Peter Piper picked a peck of pickled peppers"*). They can be used as a diction exercise which difficulty is linked to performance that can provoke embarrassment laughter or satisfaction laughter linked to erroneous or good pronunciations. To increase the possibility of erroneous pronunciation, we included sentences in different languages including English, French, Spanish, Italian, German, and Mandarin Chinese.

The game was carried out in public (with a small group of persons). This experiment consisted of 4 tasks further referred as: video, tongue-twister read, tongue-twister repeat and jokes.

3   Contents of the Inter-LAUGH corpus

3.1 Participants

We collected audio data with a lapel microphone at 16 KHz, and captured the "computer" part of the experiment using a webcam at 25 frames per second in 640x480 resolution. The participants also filled psychological evaluations regarding their OCEAN personality traits (John et. al, 1991) and sense of humor (McGhee, 1996). We recorded 45 participants (31 male and 14 female; between 25 and 60 years old).

Interspeech is an international conference and attracts researchers of several nationalities: our participants came from 24 countries. The ratio of the sample size to the number of distinct country of origin is low (1.88). It is therefore necessary to gather the participants in larger groups in order to correlate our observations to their "community". The belonging to a community could be defined by a common language, place of birth, or place of residence. For the purpose of this paper we decided to gather the participants in 3 meta-regions (America, Europe, Asia), depending on their cultural background and defined by the country where they spent their childhood.

This resulted in the following repartition: 6 participants for America (from the USA, Canada, and Mexico), 27 participants for Europe (from France, the UK, Germany, Italy, Spain, Russia, Eastern and Northern Europe) and 12 participants for Asia (from China, Japan, Thailand, and India).

3.2 Description of the audio corpus

In order to perform an audio analysis of the collected laughs, we manually segmented the recorded sessions a posteriori, and annotated positive (amused) and negative (embarrassed) laugh segments. All the laughter of the subjects were segmented and annotated by two expert coders following an adapted protocol of annotation (Devillers & Vidrascu, 2007). The annotation took the interaction context in consideration, to evaluate whether laughs were embarrassed or amused.

We decided to observe both the number of laughs generated by each task, and the relative duration of laughs over the duration of the task. Those metrics give us insight on phenomenon such as multiple chuckles, or a single but hearty laugh. This analysis, displayed in Table 1, gives us a more accurate evaluation of the reactions of the participants.

45-person sample	*All*	*Videos*	*Tongue-twisters (read on screen)*	*Tongue-twisters (repeat after the robot)*	*Jokes*
Number of embarrassed laughs	197	4	68	**93**	32
Relative duration of embarrassed laughs	-	0.09%	2.15%	**4.79%**	1.80%
Number of amused laughs	**339**	**199**	23	32	85
Relative duration of amused laughs	-	4.14%	1.09%	2.07%	**5.86%**

Table 1 : Type of laugh elicited by task - number of laughs generated by each task, and relative duration of laughs over the duration of the task

3.3 Description of the video corpus

In order to perform a video analysis of the collected smiles, we manually segmented the recorded sessions a posteriori, and annotated smiles in the obtained segments. The setting of the experiment only allowed us to capture video while the participants were facing the computer (i.e. for the "videos" and "tongue-twisters read" tasks).

We discriminated between "open smile" (where the teeth are visible between the lips) and "closed smiles". Positive laugh usually corresponds to open smile. The video were annotated without sound, to focus only on visual cues. As for the laughs, we extracted both the number of smiles, and their relative length to the task duration. Those results are visible in Table 2.

45-person sample	All	Videos	Tongue-twisters (read on screen)
Number of closed smiles	164	131	33
Relative duration of closed smiles	-	9.48%	2.70%
Number of open smiles	429	251	178
Relative duration of open smiles	-	32.78%	13.83%

Table 2 : Type of smile elicited by task - number of smiles, and relative length to the task duration

3.4 Co-occurring events in audio and video

Our work hypothesis is that while laugh is an open expression of emotion, smile is a more subtle one and might occur more easily in a non-familiar environment.

We temporally aligned annotations for audio and video channels for the "videos" and "tongue-twisters read" task, and observed the co-occurrence of events. We distinguished 4 cases: no event in either modality, event only in audio (laugh), event only in video (smile), and bimodal event. The latter represents 12.44% of all events observed on the complete session. Results are visible in Table 3.

45-person sample	no event	audio event	video event	bimodal event
complete "screen" session	36,88%	11,82%	38,87%	12,44%
Videos	33,53%	10,69%	**40,17%**	**15,61%**
tongue-twisters read	42,98%	13,86%	36,49%	6,67%

Table 3 : co-occurrence of laughs and smiles by task

If we observe tasks separately, bimodal events represent 15.61% and 6.67% of the observed events in the "videos" and "tongue-twisters read" tasks, respectively. The tongue-twisters task implied the active participation of the subjects, whereas the videos task allowed for more passive reactions, which could explain the disparity.

We can see that events occurred more frequently in video than in audio: on the "screen" part of the interaction alone, we induced more smiles than laughs.

4 Correlations between smiles, laughs and user profile

4.1 With cultural background

In order to evaluate the impact of the cultural background of the participants on their propensity to laugh and smile, we separated the data in 3 categories. We observed the ratio of amused (resp. embarrassed) laughs

produced by each group on the total number of amused (resp. embarrassed) laughs (presented in Figure 1). We proceeded in the same manner for open and closed smiles (presented in Figure 2). The imbalanced representation of the 3 populations was taken into account to correct the skewness of the data.
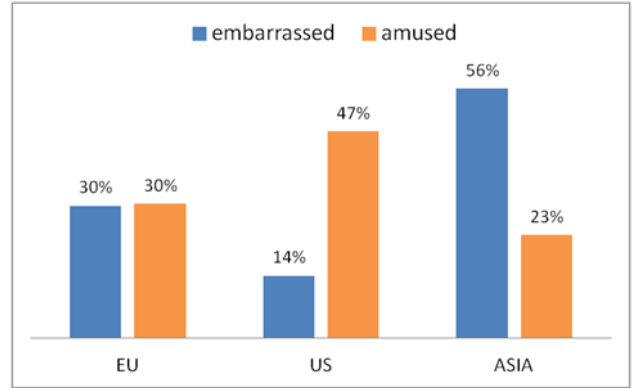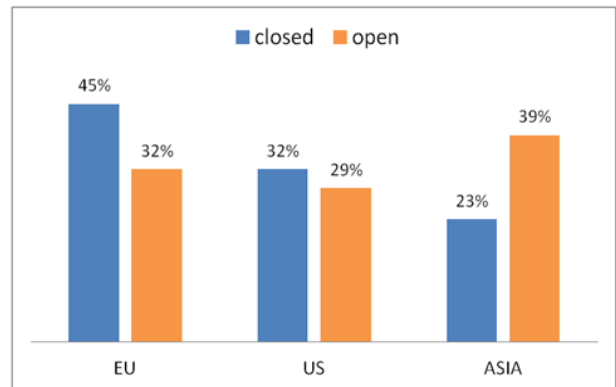

Figure 1: Laugh type by group

Figure 2: Smile type by group

On the observed sample, the participants of the "Asia" group produced the majority of the embarrassed laughs (56%), while the participants of the "US" group produced the majority of the amused laughs (47%). For those two groups, the number of amused and embarrassed laughs are imbalanced (resp. 23% and 56% for Asia, resp. 47% and 14% for US), contrary to the "EU" group who contributed to both categories equally.

Regarding the smiles, the "EU" group produced the majority of the closed smiles (45%) while the "Asia" group produced the majority of the open smiles (39%). The "US" group contributed to both categories in almost equal measures.

4.2 With OCEAN personality traits

During the design of the experiment, we expected the personality of each participant to impact on the expression of their amusement. We have been working with the OCEAN personality trait inventory on previous experiment to asset personality traits; we made the assumption that Openness, Extraversion and Neuroticism traits would have the most impact on affect bursts

production. We asked participants to fill a sub-version of the OCEAN test for those 3 characteristics.

We correlated the scores for each trait with the production of laughs and smiles, in number and in duration (absolute and relative to the task duration), for each participant, using R's implementation of Pearson correlations.

Contrary to our initial expectations, the Neuroticism trait did not significantly correlate [2] with any of our measurements. While they did not correlate with any measurements regarding smiles, the Openness and Extraversion traits did significantly correlate[2] with laughs.

45-person sample	Amused Laugh	Embarrassed Laugh
Number of events	0.1816	-0.2800
Duration of event	0.2439	**-0.2980**
Relative Duration of Event	**0.2966**	-0.2686

Table 4 : Person correlations between Openness and laugh measurements – significant correlations are marked in bold[2]

45-person sample	Amused Laugh	Embarrassed Laugh
Number of events	**0.3457**	-0.0500
Duration of event	**0.3454**	-0.0319
Relative Duration of Event	**0.3823**	-0.0167

Table 5 : Person correlations between Extraversion and laugh measurements – significant correlations are marked in bold[2]

From Table 4 and Table 5 we can see that the opened and extraverted participants produced more and longer amused laughs, and shorter embarrassed laughs.

4.3 With Sense of Humor

To evaluate the impact of individual differences in laughing matters, we asked the participants to fill a questionnaire after the experiment, to assert their sense of humor. We used McGhee Sense of Humor Scale (SHS) (McGhee, 1996) which proposes 24 questions to evaluate 6 aspects of humor: *Enjoyment of Humor, Laughter, Verbal Humor, Finding Humor in Everyday Life, Laughing at Yourself, and Humor Under Stress*. Each aspect is rated between 4 and 28, and the global sense of humor is rated by the sum of the 6 sub-categories.

We correlated this global Sense of Humor Score (SHS) with the same method used for the OCEAN traits previously mentioned. The SHS significantly correlated[2]

with several of our measurements (marked in bold in Table 6).

45-person sample	Amused Laugh	Embarrassed Laugh	Open Smile	Closed Smile
Number of events	**0.4474**	0.2635	0.2613	-0.2785
Duration of event	**0.3907**	**0.3015**	0.1952	**-0.3032**
Relative Duration of Event	**0.3403**	**0.3157**	0.2248	**-0.3070**

Table 6 : Person correlations between SHS and smile and laugh measurements – significant correlations are marked in bold[2]

From Table 6 we can see that participants with a higher SHS laughed more and longer, and produced shorter closed smiles.

5   Results

This experiment provided data regarding two aspects of man-machine interaction: the multimodal expression of human behavior, and the influence of individual "profile" on this expression.

The results of the video analysis demonstrate that a non-negligible quantity of information regarding users' appreciation of the interaction was not perceivable in the audio channel. This reinforces Scherer's definition of affects burst as multimodal events. Moreover, the noisy conditions of the experiment rendered the audio channel less reliable for an automatic analysis, but the addition of a clean video channel can compensate this deficiency.

The wide variety of participants was an opportunity to observe the impact of individual preferences on the enjoyment of a man-machine interaction. We took into account OCEAN personality traits, Sense of Humor and cultural influences in our measurements. While the observed sample is too small for strong conclusions, it encourages further investigations in that regard.

6   References

[1] Campbell, N. *"On the use of nonverbal speech sounds in human communication"*, in proc. of the COST 2102 Workshop on Verbal and Nonverbal Communication Behaviours, 2007

[2] Devillers, L., Vidrascu L., *"Positive and negative emotional states behind the laughs in spontaneous spoken dialogs"*, In Proceedings of ICPH'07 Workshop on Laughter, 2007

---

[2] Person p-coefficient below 0.05

[3] Krumhuber E. & Scherer, K. *"Affect bursts: Dynamic patterns of facial expressions"*, journal of Emotion. vol. 11, No 4 (pp 742), 2011

[4] John O. P., Donahue E. M., Kentle R. L., *"The Big Five Inventory--Versions 4a and 54."*, Berkeley: University of California, Berkeley, Institute of Personality and Social Research, 1991

[5] Johnstone T & Scherer, K. *"Vocal communication of emotion"*. In M. Lewis & J.Haviland (Eds). Handbook of emotions (pp. 200-235) New-York - Guilford Press, 2000

[6] McGhee, P. E., *"Humor, healing, and the amuse system: Humor as survival training."*, Dubuque, Iowa: Kendall/Hunt Publishing, 1996

[7] Scherer, K. R.. *"Affect Bursts, in Emotions: Essays on Emotion Theory"* (S. H.M. van Goozen, N. E. van de Poll, & J. A. Sergeant, eds.),. Hillsdale, NJ: Lawrence Erlbaum, (pp 161-193), 1994

[8] Schröder, M. *"Experimental study of affect bursts"*, Speech Communication – Special session on speech and emotion, vol. 40, No 1-2, 2003

[9] Tanaka A., Koizumi A., Imai H., Hiramatsu S., Hiramoto E., de Gelder B., *"I Feel Your Voice Cultural Differences in the Multisensory Perception of Emotion."* Psychological science vol. 21, No 9 (pp 1259-1262), 2010