A tool to elicit and collect multicultural and multimodal laughter

Mariette SOURY[1,2], Clément GOSSART[1], Martine ADDA[1,3], Laurence DEVILLERS[1,4]

[1] LIMSI-CNRS
[2] Université Paris XI Orsay
[3] Université Sorbonne Nouvelle - *Paris 3*
[4]Université Paris Sorbonne - *Paris 4*
France

soury@limsi.fr, madda@limsi.fr, devil@limsi.fr

Abstract

We present the implementation of a data collection tool of multicultural and multi-modal laughter for the 14th Interspeech conference. The application will automatically record and analyze audio and video stream to provide real-time feedback. Using this tool, we expect to collect multimodal cues of different kind of laughers elicited in participants with funny videos, as well as jokes and tongue-twisters games with the Nao robot. The collected corpus will be used for paralinguistic challenges.

Index Terms: laughter, smile, data collection, feedback

1. Introduction

Laughter helps humans to clarify their intentions in social interaction and provides an emotional context to conversations. Laughter is used as a signal of acceptance and positive interactions with others. However, it can also be caused by emotional states such as embarrassment or stress [2]. Companion machines such as robots should detect and use it. Researchers investigate the detection of laughter [1, 4, 7, 8] and its elicitation [6]. For building such systems and use machine learning techniques, we need large amount of data. Few existing resources are multicultural. The Interspeech Conference gives us the opportunity to study the cultural aspects of laugher.

In social interactions, affective cues are an integral part of the discourse: a smile, crossed arms or an embarrassed laughter are all information that impacts the spoken message beyond words. Affective computing field has developed for more than a decade on this postulate: a machine able to interpret a non-neutral discourse will be able to adapt its reactions in a much more subtle manner. In human-machine interactions, emotions are notably carried by non verbal cues such as multimodal affect bursts [5, 9]: laughter, sights, etc. The notion of affect burst refers to sudden, full-blown displays of emotion in which facial, vocal, and gestural components are highly synchronized. Affect bursts (such as laughter/smile) are defined as short emotional non-speech expressions that convey a clearly identifiable emotional meaning in response to highly affective charged and often unexpected events.

In order to design social and affective interactive systems with robots, experimental grounding is required to study expressions of affect bursts during interactions. During the 2013 Interspeech Conference, we would like to collect multimodal laughter through an experiment with multicultural participants

2. Elicitation of Laugher

Two main techniques have been used to collect spontaneous laughter: real-life data collection [2], and laughter-induction tasks [3]. In this experiment, we use two tasks for eliciting laughter: passive induction tasks such as watching funny videos or listening to jokes, and active induction tasks such as Tongue-Twisters. In the video task, we intend to observe both reflex contagious laughter and cultural laughter.

2.1. Funny Videos

In our experiment, we intend to study 2 types of laughter, by displaying videos of laughing babies (“reflex” amusement) and of funny scenes with various cultural background (“cognitive” amusement).

Laugher is a universal human reaction, happening in the first months of life. A powerful social signal, laugher is "contagious": seeing or hearing someone laugh can trigger our own amusement in a reflexive manner. We will study this contagion loop. Laughter can also be the cognitive result of the interpretation of a ridiculous or funny situation. Such interpretation is unique to each individual: the sense of humor is both personal and cultural, and not everyone laughs of the same things. We will study those differences.

2.2. Jokes

Humor is a complex mechanism, building on a shared cultural referential between the teller and the listener. One crucial element in effective jokes is the building and breaking of tension: a story becomes funny when it takes an unexpected turn. The timing in the story telling is important, to allow the tension to build before relieving it.

In our experiment, we will use the Nao robot to tell jokes and study the synchronization between story-telling phases (premise, punch line), and amusement expressed by the participants.

2.3. Tongue-Twisters

In our experiment, we would like to benefit of the international venue of the Interspeech Conference, and use tongue-twisters 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. They can be used as a diction exercise which difficulty is linked to performance that can provoke amusement linked to erroneous pronunciations.

Examples:

- **English:** “Peter Piper picked a peck of pickled peppers”
- **French:** “Les chaussettes de l’archiduchesse sont-elles sèches ou archi-sèches ?” (Are the Archiduchesse socks dry or extra-dry ?)
- **Mandarin Chinese:** *“四十四隻死石獅子”* [si shi si zhi si shi shi zi] (44 dead stone lions.)

3. Implementation

3.1. Interface

We would like to invite the Interspeech public to participate in a data collection of a few minutes, in order to build a corpus of laughter. Volunteers will watch several short video clips stimulating reflex and/or cognitive amusement; then they will have to read several tongue-twisters in different languages. They will also interact with the Nao Robot, who will try to tell jokes and assert if the participants found them funny.
To realize this experiment, we are developing an application which will display pre-scripted stimuli (on-screen or through Nao), record participant's reactions and meta-data (mother-tongue, age…) and display a feedback in real time (on-screen or through Nao).
The application is conceived so as to be able to connect with existing tools from LIMSI: an audio segmentation system, an audio feature extraction system, and a classification system using audio laughter models. The interface will use those 3 blocks to realize a stimuli-perception-reaction loop. In this experiment, voice and face signals feedback will be voluntary treated independently.

3.2. Sensors

We are naturally interested in speech in the context of the Interspeech Conference; audio recordings will be captured during the experiment. Given the context of the recording (a conference room), we will use a directional microphone, to capture mostly participants' voice and limit noise. As the video stimuli also include sound, participants will be using headphones so that the stimuli sound is not captured along with their voice.
We are also considering affective facial expressions; we make the assumption that laughers are multimodal affect bursts emitted along with smiles. To validate this hypothesis, we will record the face of volunteers using an of-the-shelf web-cam.

4. Expected Results

The aim of this experiment is the collection of a social signals corpus, which could be used in future paralinguistic challenges. The international venue of the Interspeech Conference will notably give us the opportunity to study cultural aspects of laugher; we expect to collect data from enough participants from different background to observe different reactions to the same stimuli.
The collected data will help us to build a humor module to enhance man-machine interactions in future works, including a project with the Nao robot.

5. References

[1] Bachorowski, J., Smoski, M, Owren, M., “The acoustic features of human laughter.” Journal of the Acoustical Society of America, vol. 110, pp. 1581–1597, 2001.

[2] Devillers, L. Vidrascu L., “Positive and negative emotional states behind the laughs in spontaneous spoken dialogs.” Workshop Laughter 2007 of ICPHs 2007.

[3] McKeown, G. Cowie, R., Curran, W., Ruch, W., Douglas-Cowie E., “ILHAIRE Laughter Database.” Proceedings of the 4th International Workshop on Corpora for Research on EMO-TION SENTIMENT & SOCIAL SIGNALS, Satel-lite of LREC 2012, May 2012, Istanbul, Turkey, ELRA.

[4] Knox T. and Mirghafori W., “Automatic Laughter Detection Using Neural Networks”, Proceedings of INTERSPEECH, 2007

[5] Krumhuber E. & Scherer, K. (2011) “Affect bursts: Dynamic patterns of facial expressions”, journal of Emotion. 2011 Aug ; 11 (4):742.

[6] Niewiadomski, R. Urbain, J., Pelachaud, C., Dutoit, T., “Finding out the audio and visual features that influence the perception of laughter intensity and differ in inhalation and exhalation phases”, Proceedings of the 4th International Workshop on Corpora for Research on EMOTION SENTIMENT & SOCIAL SIGNALS, Satellite of LREC 2012, May 2012, Istanbul, Turkey, ELRA

[7] Scherer, S. Schwenker, F., Campbell, N. “Multi Modal Laughter Detection in Natural Discourses”. Human Centered Robot Systems systèmes cognitifs Monographies 2009, Volume 6, 111-120, DOI: 10.1007/978-3-642-10403-9_12, 2009

[8] Schuller, B., Steidl S., Batliner, A., Alessandro Vinciarelli A., Scherer K., Ringeval, F., Chetouani, M., Weninger, F., Eyben, F., Marchi, E., Mortillaro, M., Salamin, H., Polychroniou, A., Valente F., Kim, S., “The INTERSPEECH 2013 Computational Paralinguistics Challenge Social Signals, Conflict, Emotion, Autism”, Interspeech 2013.

[9] Schröder, M., “Experimental study of affect bursts”, Speech Communication – Special session on speech and emotion, vol. 40, Issue 1-2. 2003.