Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

Audiovisual analysis of relations between laughter types and laughter motions

Carlos Ishi [1], Hiroaki Hatano [1], Hiroshi Ishiguro [1]

[1] ATR Hiroshi Ishiguro Labs.

carlos@atr.jp, hatano.hiroaki@atr.jp, ishiguro@sys.es.osaka-u.ac.jp

Abstract

Laughter commonly occurs in daily interactions, and is not only simply related to funny situations, but also for expressing some type of attitude, having important social functions in communication. The background of the present work is generation of natural motions in a humanoid robot, so that miscommunication might be caused if there is mismatch between audio and visual modalities, especially in laughter intervals. In the present work, we analyzed a multimodal dialogue database, and investigated the relations between different types of laughter (including production type, vowel quality, laughing style, intensity and laughter functions) and different types of motion during laughter (including facial expressions, head and body motion).

Index Terms: laughter, facial expression, laughter motion, non-verbal information, natural conversation

1. Introduction

Laughter commonly occurs in daily interactions, and is not only simply related to funny situations, but also for expressing some type of attitude, having important social functions in human-human communication. Therefore, it is important to account for laughter in robot-mediated communication as well. The authors have been working on improving human-robot communication, by implementing humanlike motions in several types of humanoid robots. Natural (humanlike) behaviors by a robot are required as the appearance of the robot approaches the one of a human, such as in android robots. Several methods for automatically generating lip and head motions from the speech signal of a tele-operator have been proposed in the past [1-4]. Recently we also started to tackle the problem of generating natural motion during laughter [5]. However, we are still not able to generate motions according to different laughter types or different laughter functions.

Several works have investigated the functions of laughter and the relationship with acoustic features. For example, it is reported that duration, energy and voicing/unvoicing features change between positive and negative laughter, in a French hospital call center telephone speech [6]. In [7], it is reported that the first formant is raised and vowels are centralized (schwa), by analyzing English acted laughter data of several speakers. In [8-9], it is reported that mirthful laughter and polite laughter differ in terms of duration, the number of calls (syllables), pitch and spectral shapes, in Japanese telephone conversational dialogue speech. In our previous work [10], we have analyzed laughter events of students in a science classroom of a Japanese elementary school, and found relations between laughter types (production, vowel quality, and style), functions and situations.

Regarding the relationship between audio and visual features in laughter, several works have been conducted in the computer graphics animation field [11-13]. However, most of them dealt with symbolic facial expressions, so that dynamic features and differences in smiling face due to different types of laughter are not expressed. As described above, different types of laughter may require different types of smiling faces.

Thus, it is important to clarify how different motions are related to different types of laughter. In the present work, we analyzed laughter events in face-to-face human interactions in a multimodal dialogue database, and investigated the relations between different types of laughter (such as production type, laughing style, and laughter functions) and the visual features (facial expressions, head and body motions) during laughter.

2. Analysis data

2.1. Description of the data

For analysis, we use the multimodal conversational speech database recorded at ATR/IRC labs [2]. The database contains face-to-face dialogues between several pairs of speakers, including audio, video and (head) motion capture data for each of the dialogue partners. Each dialogue is about 10 ~ 15 minutes of free-topic conversations. The database contains segmentation and text transcriptions, and also includes information about presence of laughter. For the present analysis, data of 12 speakers (8 female and 4 male speakers) were used, from where about 1000 laughing speech segments were extracted.

2.2. Annotation data

The following label sets were used to annotate the laughter types and laughter functions. These are based on past works. (The terms in parenthesis are the original Japanese terms used in the annotation.)

- Laughter production type: {breathiness over the whole laughter segment (“kisoku”), alternated pattern of breathy and non-breathy parts (“iki ari to iki nashi no kougo”), relaxed (“shikan”: vocal folds relaxed, absence of breathiness:), laughter during inhalation (“hikiwarai”)}
- Laughter style: {secretly (“hisohiso”), giggle/chuckle (“kusukusu”), guffaw (“geragera”), sneer (“hanawarai”)}
- Vowel-quality of the laughter: {“hahaha”, “hehehe”, “hihihi”, “hohoho”, “huhuhu”, schwa (central vowel)}
- Laughter intensity level: {1 (“shouwarai”), 2 (“chuuwarai”), 3 (“oowarai”), 4 (“bakushou”)}
- Laughter function: {funny/amused/joy/mirthful laugh (“omoshiroi”, “okashii”, “tanoshii”), social/polite laugh (“aisowarai”), bitter/embarrassed laugh (“nigawarai”), self-conscious laugh (“terewarai”), inviting laugh (“sasoiwarai”), contagious laugh (“tsurarewarai”, “moraiwarai”), depreciatory/derision laugh (“mikudashiwarai”), dumbfounded laugh (“akirewarai”), untrue laugh (“usowarai”), softening laugh (“kanwa/ba o yawarageru”: soften/relax a strained situation)}

A research assistant (native speaker of Japanese) annotated the labels above, by listening to the segmented intervals

 doi: 10.21437/SpeechProsody.2016-165

(including five seconds before and after the laughter portions.) For the label items in "laughter style" and "laughter functions" (items 2 and 4 in Table 1), annotators were allowed to select more than one item per laughter event. No specific constraints were imposed for the number of times for listening, or the order for annotating all items in Table 1.

The number of laughter calls (individual syllables in an /h/-vowel sequence) was also annotated for each laughter event, by looking at the spectrogram displays.

The following label sets were used to annotate the visual features related to motions and facial expressions during laughter.

- eyelids: {closed, narrowed, open}
- cheeks: {raised, not raised}
- lip corners: {raised, straightly stretched, lowered}
- head: {no motion, up, down, left or right up-down, tilted, nod, others (including motions synchronized with motions like upper-body)}
- upper body: {no motion, front, back, up, down, left or right, tilted, turn, others (including motions synchronized with other motions like head and arms)}

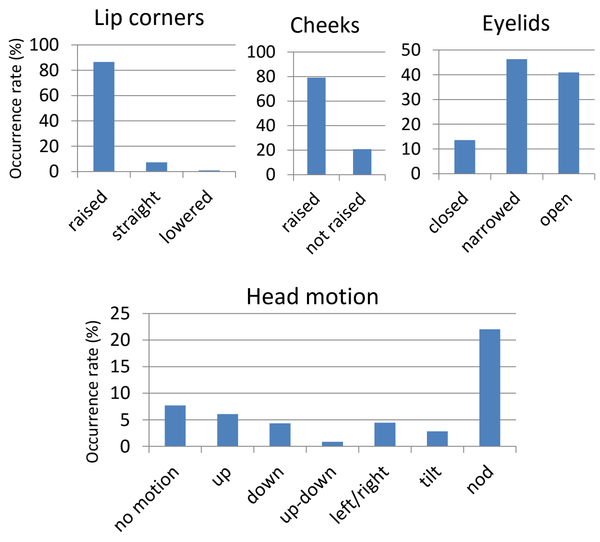
For each laughing speech event, another research assistant annotated the labels related to motion and facial expressions, by looking at the video and the motion data displays.

For all annotations above, it was allowed to select multiple labels, if multiple items are perceived.

3. Analysis of the laughter events

3.1. Analysis of laughter motions

The overall distributions of the motions during laughter were first analyzed. Fig. 1 shows the distributions for each motion type. Firstly, as a most representative feature for facial expression in laughter, it was observed that lip corner is raised in more than 80% of the laughter events. Cheeks were raised in 79%, and eyes were narrowed or closed in 59% of the laughter events. More than 90% of the laughter events were accompanied either by a head or upper body motion, from which the majority of the motions were in the vertical axis (up-down or front/back body motion, and nods for head motion).



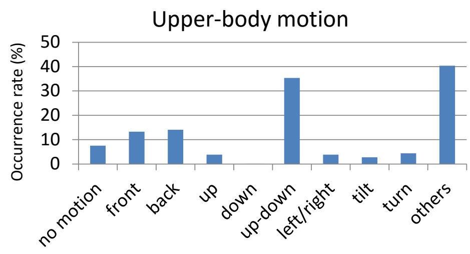


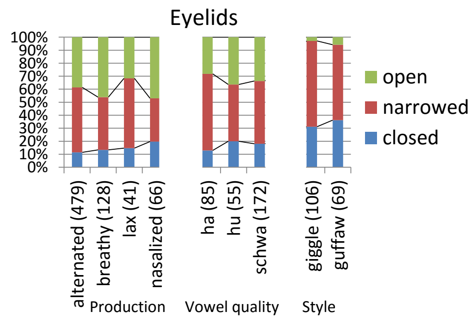
Figure 1. Distributions of face (lip corners, cheek and eyelids), head and upper-doby motions during laughter speech.

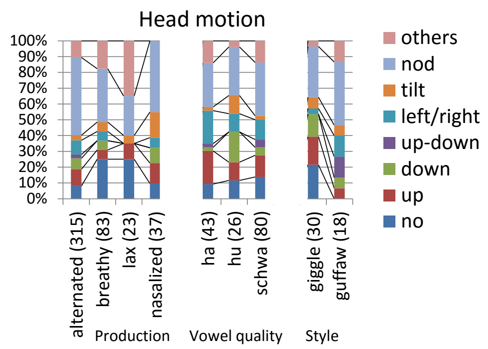
For investigating the timing of the motions during laughter speech, we conducted detailed analysis for two of the speakers (female speakers in her 20s). The instants of eye blinking and the start and end points of eye narrowing and lip corner raising were segmented.

As a result, it was observed that the start time of the smiling facial expression (eye narrowing and lip corner raising) usually matched with the start time of the laughing speech, while the end time of the smiling face (i.e., the instant the face turns back to the normal face) was delayed relative to the end time of the laughing speech by 0.8 ± 0.5 seconds for one of the speakers, and 1.0 ± 0.7 seconds for the other speaker. Furthermore, it was observed that an eye blinking is usually accompanied at the instant the face turns back from the smiling face to the normal face.

3.2. Analysis of laughter motions and laughter types

Fig. 2 shows the distributions of the laughter motions according to different laughter types (production, vowel quality, and style). The number of occurrences for each item is shown within parenthesis. The items with low number of occurrences are omitted. The results for lip corner and cheek motions are also omitted, since most of laughter events are accompanied by lip corner raising and cheek raising.





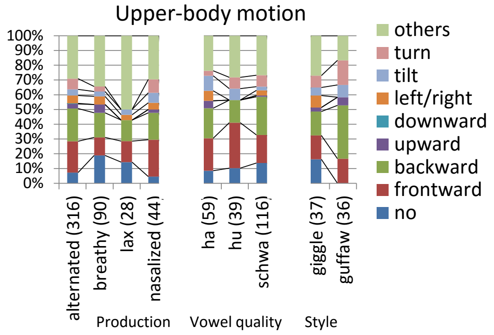


Figure 2. *Distributions of eyelids, head motion and body motion, for different categories of production type (left), vowel-quality (mid) and laughter style (right). The total number of utterances is shown within brackets.*

From the results in Fig.2, it can be observed that almost all laughter events are accompanied by eyelid narrowing and closing in giggle and guffaw laughter styles. In guffaw laughter, all laughter events were accompanied by some body motion, from where the occurrence rate of backward motion was relatively higher.

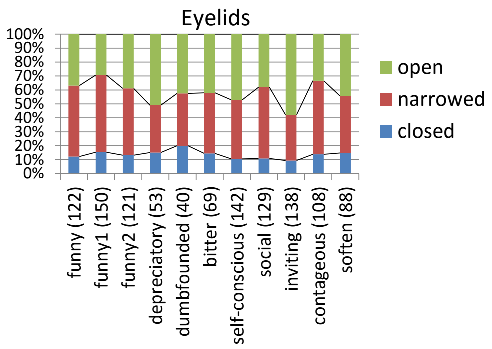
Regarding the vowel quality, by comparing the distributions of "ha" and "hu", it can be observed that in "hu" the occurrence rate of head down and body frontward motion, while in "ha", head up motion occurs with relatively high rate.

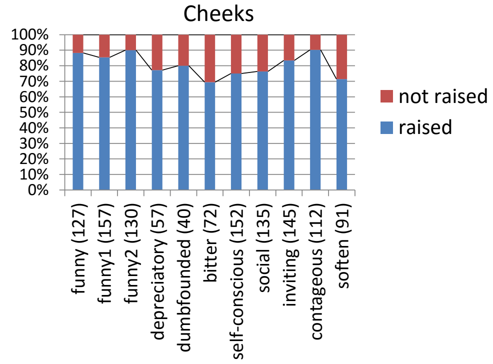
Regarding the production type, "breathy" and "lax" production types show higher occurrence of "no" motion for both head and body motion, compared to the "alternated" pattern.

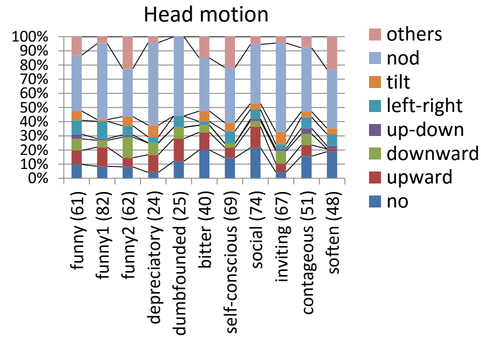
3.3. Analysis of laughter motions and laughter functions

Fig. 3 shows the distributions of the laughter motions (eyelids, head motion, and body motion) according to different laughter functions. The number of occurrences for each item is shown within parenthesis. The items with low number of occurrences are omitted.

From Fig. 3, it can be observed that in funny laughter (funny/amused/joy/mirthful) and contagious laughter, the occurrence rates of cheek raising are higher (above 90%). This is because such types of laughter are thought to be spontaneous laughter, so that Duchenne smiles [14] occur and the cheek is usually raised. Similar trends were observed for eyelid narrowing or closing.







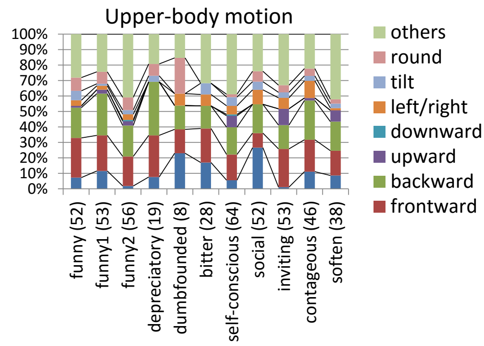


Figure 3. *Distributions of eyelids, cheeks, head motion and body motion categories, for different categories of laughter functions. The total number of utterances is shown within brackets.*

Regarding head motion and body motion, relatively high occurrence of "no" motion (about 20%) are observed in bitter, social, dumbfounded, and softening laughter. It can be interpreted that the occurrence of head and body motion decreases in these laughter types, since they are not spontaneous, but "artificially" produced.

3.4. Analysis of laughter motions and laughter intensity

Fig. 4 shows the distributions of the laughter motions (eyelids, cheeks, lip corners, head motion, and body motion) according to different laughter intensity categories. The correlations between laughter intensity and different types of motions are much clearer than the laughter styles or laughter functions shown in Sections 3.2 and 3.3. From the results shown for eyelids, cheeks and lip corner, it can be said that the degree of smiling face increased according to the intensity of the laughter, that is, eyelids are narrowed or closed, and both cheeks and lip corners are raised (Duchenne smile faces).

Regarding the body motion categories, it can be observed that the occurrence rates of front, back and up-down motions increase, as the laughter intensity increases. The results for intensity level 4 shows slightly different results, but this is probably because of the small number of occurrences (around 20, for 8 categories).

From the results for head motion, it can be observed that the occurrence rates of nods decrease, as the laughter intensity increases. Since nods usually appear for expressing agreement, consent or sympathy, they are thought to be easier to appear in low intensity laughter.

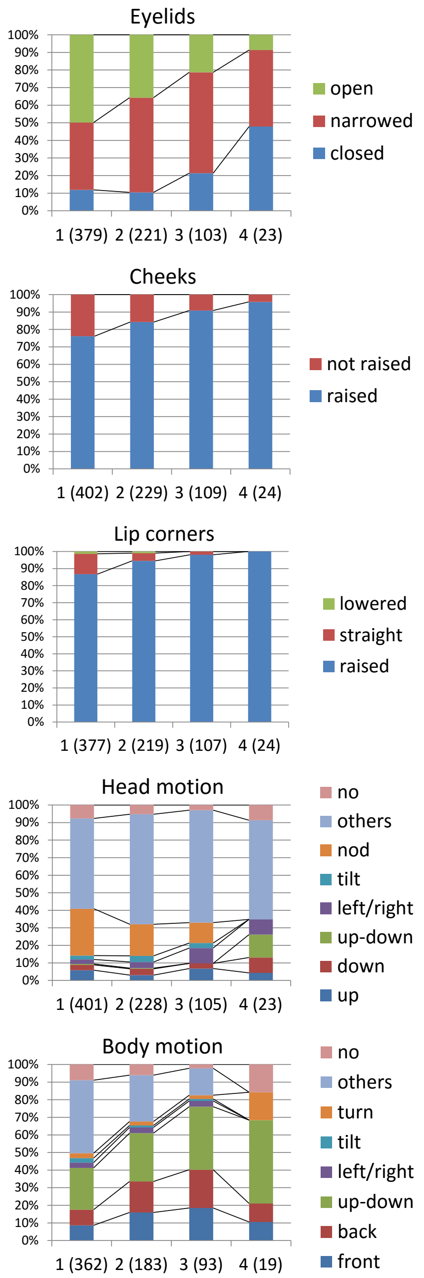


Figure 4. *Distributions of eyelid, cheek, lip corner, head and body motion categories, for different categories of laughter intensity (1 to 4). The total number of utterances is shown within brackets.*

4. Conclusions

In the present work, we analyzed audiovisual properties of laughter events in face-to-face dialogue interactions.

Analysis results of laughter events revealed relationships between laughter motions (facial expressions, head and body motions) and laughter type, laughter function and laughter intensity. Firstly, it was found that giggle and guffaw laughing styles are almost always accompanied by smiling facial expressions and head or body motion. Artificially produced laughter (such as social, bitter, dumbfounded and softening laughter) tends to be accompanied by less motion compared to spontaneous laughter (such as funny and contagious laughter). Finally, it was found that the occurrence of smiling faces (Duchenne smiles) and body motion increase, and the occurrence of nods decrease, as the laughter intensity increases.

Future works include evaluation of acoustic features for automatic detection and classification of laughter events, and applications to laughter motion generation in humanoid robots.

5. Acknowledgements

This study was supported by JST/ERATO. We thank Mika Morita, Kyoko Nakanishi and Megumi Taniguchi for contributions in the annotations and data analyses.

6. References

[1] Ishi, C., Liu, C., Ishiguro, H. and Hagita, N. (2012). “Evaluation of a formant-based speech-driven lip motion generation,” In 13th Annual Conference of the International Speech Communication Association (Interspeech 2012), Portland, Oregon, pp. P1a.04, September, 2012.

[2] C.T. Ishi, C. Liu, H. Ishiguro, and N. Hagita. “Head motion during dialogue speech and nod timing control in humanoid robots,” Proc. of 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2010), pp. 293-300, 2010.

[3] C. Liu, C. Ishi, H. Ishiguro, and N. Hagita. Generation of nodding, head tilting and gazing for human-robot speech interaction. International Journal of Humanoid Robotics (IJHR), vol. 10, no. 1, January 2013.

[4] S. Kurima, C. Ishi, T. Minato, and H. Ishiguro. Online Speech-Driven Head Motion Generating System and Evaluation on a Tele-Operated Robot, IEEE International Symposium on Robot and Human Interactive Communication (ROMAN 2015), pp. 529-534, 2015.

[5] Ishi, C., Minato, T., Ishiguro, H. (2015) "Investigation of motion generation in android robots during laughing speech," Intl. Workshop on Speech Robotics (IWSR 2015), Sep. 2015.

[6] Devillers, L. & Vidrascu, L., “Positive and negative emotional states behind the laughs in spontaneous spoken dialogs”, Proc. of Interdisciplinary Workshop on The Phonetics of Laughter, 37-40, 2007.

[7] Szameitat, D. P., Darwin, C. J., Szameitat, A. J., Wildgruber, D., & Alter, K. Formant characteristics of human laughter. J Voice, 25, 32-37, 2011.

[8] Campbell, N., “Whom we laugh with affects how we laugh”, Proc. of Interdisciplinary Workshop on The Phonetics of Laughter, 61-65, 2007.

[9] Tanaka, H. & Campbell, N., “Acoustic features of four types of laughter in natural conversational speech”, Proc. of ICPhS XVII, 1958-1961, 2011.

[10] Ishi, C., Hatano, H., Hagita, N. (2014) "Analysis of laughter events in real science classes by using multiple environment sensor data," Proc. of 15th Annual Conference of the International Speech Communication Association (Interspeech 2014), pp. 1043-1047, Sep. 2014.

[11] H. Yehia, T. Kuratate, and E. Vatikiotis-Bateson, Using speech acoustics to drive facial motion, Proc. of *the 14th International Congress of Phonetic Sciences* (*ICPhS99*), 1, pp. 631-634, 1999.

[12] R. Niewiadomski, M. Mancini, Y. Ding, C. Pelachaud, and G. Volpe (2014). “Rhythmic Body Movements of Laughter,” In Proc. of the 16th International Conference on Multimodal Interaction (ICMI '14). ACM, New York, NY, USA, 299-306.

[13] Niewiadomski, R, Ding, Y., Mancini, M., Pelachaud, C., Volpe, G., Camurri, A. “Perception of intensity incongruence in synthesized multimodal expressions of laughter,” The sixth International Conference on Affective Computing and Intelligent Interaction (ACII2015), 2015

[14] P. Ekman, R.J. Davidson, W.V. Friesen. The Duchenne smile: Emotional expression and brain physiology II. Journal of Personality and Social Psychology, Vol. 58(2), 342-353, 1990.