SIP (2019), vol. 8, e6, page 1 of 10 © The Authors, 2019.

doi:10.1017/ATSIP.2018.32

ORIGINAL PAPER

Analysis and generation of laughter motions, and evaluation in an android robot

CARLOS TOSHINORI ISHI, TAKASHI MINATO AND HIROSHI ISHIGURO

Laughter commonly occurs in daily interactions, and is not only simply related to funny situations, but also to expressing some type of attitudes, having important social functions in communication. The background of the present work is to generate natural motions in a humanoid robot, so that miscommunication might be caused if there is mismatching between audio and visual modalities, especially in laughter events. In the present work, we used a multimodal dialogue database, and analyzed facial, head, and body motion during laughing speech. Based on the analysis results of human behaviors during laughing speech, we proposed a motion generation method given the speech signal and the laughing speech intervals. Subjective experiments were conducted using our android robot by generating five different motion types, considering several modalities. Evaluation results showed the effectiveness of controlling different parts of the face, head, and upper body (eyelid narrowing, lip corner/cheek raising, eye blinking, head motion, and upper body motion control).

Keywords: Emotion expression, Laughter, Motion generation, Human–robot interaction, Non-verbal information

Received 29 June 2018; Revised 10 December 2018

I. INTRODUCTION

Laughter commonly occurs in daily interactions, not only simply in relation to funny situations, but also to some type of attitudes (like friendliness or interest), having an important social function in human–human communication [1, 2], as well as positive influences on human's health [3, 4]. Laughter also provides social signals that allow our conversations to flow smoothly among topics; to help us repairing conversations that are breaking down; and to finish our conversations on a positive note [5, 6]. Thus, it is important to account for laughter in robot-mediated communication as well.

We have been working on improving human–robot communication, by implementing humanlike motions in several types of humanoid robots. So far, we proposed and evaluated several methods for automatically generating lip and head motions of a humanoid robot in synchrony with the speech signal [7–10]. Throughout the evaluation experiments, we have observed that more natural (humanlike) behaviors by a robot are expected, as the appearance of the robot becomes closer to the one of a human, such as in android robots. Further, it has been observed that unnaturalness occurs when there is a mismatch between voice and motion, especially during short-term emotional expressions, like in laughter and surprise. To achieve a smooth human–robot interaction, it is essential that natural (humanlike) behaviors are expressed by the robot. Thus, in the present work, we focused on natural motion generation during the laughter events of humanoid robots (i.e. when the robot produces laughing speech.) To design the laughter motion of a humanoid robot, two issues need to be considered: (1) the modalities related to laughter have to be clarified, for generating motion in synchrony with the laughing speech intervals; (2) the generated motion suffers from the limitation of robot hardware system (i.e. motion range and controllable parts are limited).

In this study, in order to account for these two issues, we first analyzed how different modalities (lip corners, eyelids, cheeks, head, and body movements) appear in synchrony with the laughing speech intervals, in face-to-face dialogue interactions. Then, we proposed a method for motion generation in our android robot, based on the main trends from the analysis results. As a first step for a complete automation of laughter generation from the speech signal, we assume that the laughter interval is given, and investigate if natural motion can be generated in the android during laughing speech. We then conducted subjective experiments to evaluate the effects of controlling different modalities in the proposed laughter motion generation method, under the android hardware limitation.

This manuscript is organized as follows. In Section II, we present related works on laughter motion generation. In Section III, we report the analysis results of motion during the laughter events. In Section IV, we describe the proposed

ATR Hiroshi Ishiguro Laboratories, 2-2-2 Hikaridai, Keihanna Science City, Kyoto, Japan

Corresponding author:
C. T. Ishi
Email: carlos@atr.jp

method for generating laughter motion from the speech signal, and present evaluation results of motion generation in an android. In Section V, we discuss the interpretation of the evaluation results, and Section VI concludes the paper. This paper is an extended version of our previous studies reported in [11, 12].

II. RELATED WORK

Regarding the issue on motion generation synchronized with laughter, several works have been reported in the computer graphics (CG) animation field [13–16]. For example, the relations between laughter intensity and facial motion were investigated, and a model which generates facial motion position only from laughter intensity was proposed in [13]. In [14], the model of laughter synthesis was extended by adding laughter duration as input, and selecting pre-recorded facial motion sequences. In [15], a multimodal laughter animation synthesis method, including head, torso, and shoulder motions, is proposed. Methods to synthesize rhythmic body movements (torso leaning and shoulder vibration) of laughter and to integrate them with other synthesized expressions are proposed in [16]. The torso leaning and shoulder vibrations are reconstructed from human motion capture data through synthesis of two harmonics.

However, an issue regarding robotics application is that, different from CG agents, androids have limitations in the motion degrees of freedom (DOF) and motion range. Those studies on CG agents assume rich 3D models for facial motions, which cannot be directly applied to the android robot control. So, it is important to clarify what motion generation strategies are effective to give natural impressions with laughter, under limited DOFs. In the social robotics field, some studies have implemented a facial expression of smile or laughing in humanoid robots [17, 18]. However, these studies only cover symbolic facial expressions, so that dynamic features of laughter are not dealt with. Many studies have shown that dynamic features of facial motion influence the perception of facial expression (surveyed in [19]). For example, the experimental results using CG agents indicated that different meanings of laugh (amused, embarrassed, and polite) can be expressed depending on the movement size and duration of facial parts and those timings [20]. In other words, if the motion timing of the facial parts is not appropriately designed, the meaning of the robot's laugh may not be correctly interpreted by people, which is a crucial issue for natural human–robot interaction. So, it is important to clarify which motion generation strategies are effective to give natural impressions during laughter, under limited DOFs.

Thus, in this study, we focused on the analysis of the dynamic features of different modalities relative to the laughing speech intervals, and evaluated the contribution of each modality control on the naturalness of laughter motion in our android robot.

III. ANALYSIS DATA

We firstly analyzed audio-visual data of laughing speech segments appearing in human–human dialogue interactions.

A) Description of the data

For analysis, we used the multimodal conversational speech database recorded at ATR/IRC laboratories. The database contains face-to-face dialogues between pairs of speakers, including audio, video, and (head) motion capture data of each dialogue partners. Each dialogue is about 10 ∼ 15 min of free-topic conversations. The conversations include topics about past events, future plans for trips, self-introductions, topics about a person they know in common, topics regarding family and work, and past experiences. The database contains segmentation and text transcriptions, and also includes information about presence of laughter. All laughter events in the database were naturally occurred within the conversations, i.e. the participants were not elicited to laugh. The laughter events were manually segmented, by listening to the audio signals and looking at the spectrogram displays. Data of 12 speakers (eight female and four male speakers) were used in the present analysis, from where about 1000 laughing speech segments were extracted.

B) Annotation data

The laughter intensity and motions were annotated for analyses. The laughter intensity was categorized in four levels.

- Laughter intensity: {level 1: small laughter, level 2: medium laughter, level 3: loud laughter, level 4: burst out laughter}

The laughter intensity levels were annotated by listening to each laughing speech segments.

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 laughter event, a research assistant annotated the above labels (related to motion and facial expressions), by observing the video and the motion data displays.

The overall distributions of the motions during laughter are shown in Fig. 1, for each motion type. Firstly, as the most representative feature of all facial expressions in laughter, it was observed that the lip corners are raised (moved up) in more than 90% of the laughter events. Cheeks were

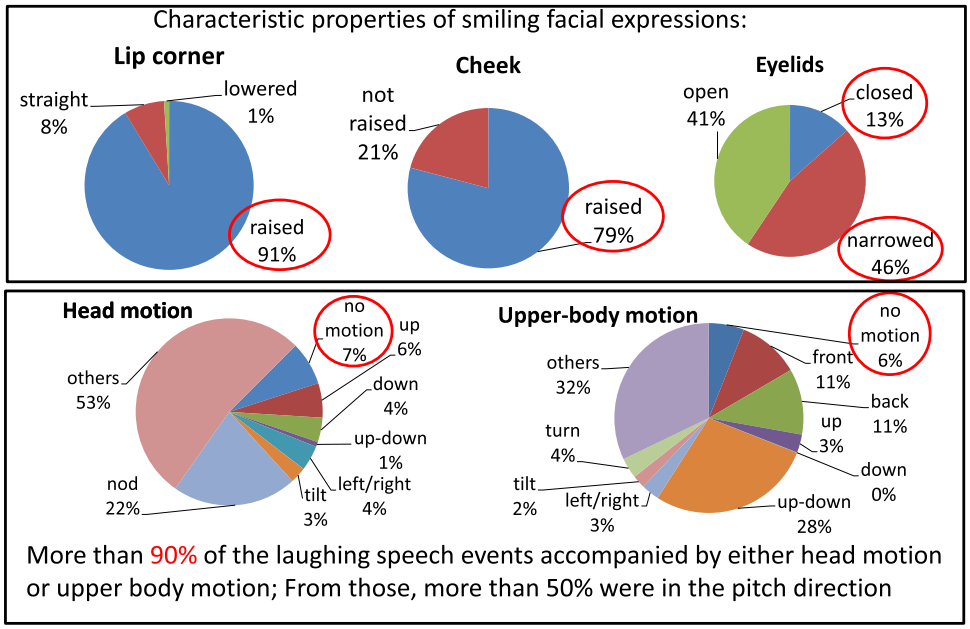

Fig. 1. Distributions of face (lip corners, cheek, and eyelids), head, and upper body motions during laughing speech.

raised in 79%, and eyes were narrowed or closed in 59% of the laughter events. Most 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 (head pitch or upper body pitch motion).

C) Analysis of laughter motions and laughter intensity

Figure 2 shows the distributions of the laughter motions (eyelids, cheeks, lip corners, head motion, and body motion) according to the different laughter intensity-level categories ("1" to "4"). The symbols shown on the bars indicate statistical significance test results after conducting χ^2 tests: "$\wedge$" means significantly higher occurrences ($\wedge p < 0.05$, $\wedge\wedge p < 0.01$), while "$\vee$" means significantly lower occurrences ($\vee p < 0.05$, $\vee\vee p < 0.01$). For example, in the "Eyelids" panel, eyelids are open with significantly higher occurrences ($\wedge\wedge$) in low laughter intensity (level "1"), and significantly lower occurrences ($\vee\vee$) in high laughter intensities (levels "3" and "4").

From the results shown for eyelids, cheeks and lip corners, it can be said that the degree of smiling face increased in proportion to the intensity of the laughter, that is, eyelids are narrowed or closed, and both cheeks and lip corners are raised (Duchenne smile faces [21]).

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 (<20, for eight categories).

From the results of 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.

D) Analysis of motion timing during laughter events

For investigating the timing of the motions during laughing speech, we conducted detailed analysis for five of the female speakers (in her 20s).

The time instants of eye blinking and the start and end points of eye narrowing and lip corner raising were manually segmented. The eye narrowing and lip corner raising were categorized in two levels. For example, for lip corners, slightly raised (level 1) and clearly raised (level 2) were distinguished.

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 relatively to the end time of the laughing speech by 1.2 ± 0.5 s.

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. It was observed that an eye blinking occurred within an interval of 0.1 s from the offset time of a smiling face in 70% of the laughter events. In contrast, it occurred within an interval of 0.1 s from the onset time of a smiling face in only 23% of the laughter events.

Regarding lip corner raising, it was observed that the lip corners were clearly raised (level 2) at the laughter segments by expressing smiling faces, while they were slightly raised (level 1) during a longer period in non-laughing intervals by expressing slightly smiling faces. The percentage in time of smiling faces (level 2) was 20%, while by including slight smiling faces (levels 1 and 2), the percentage in time was 81%

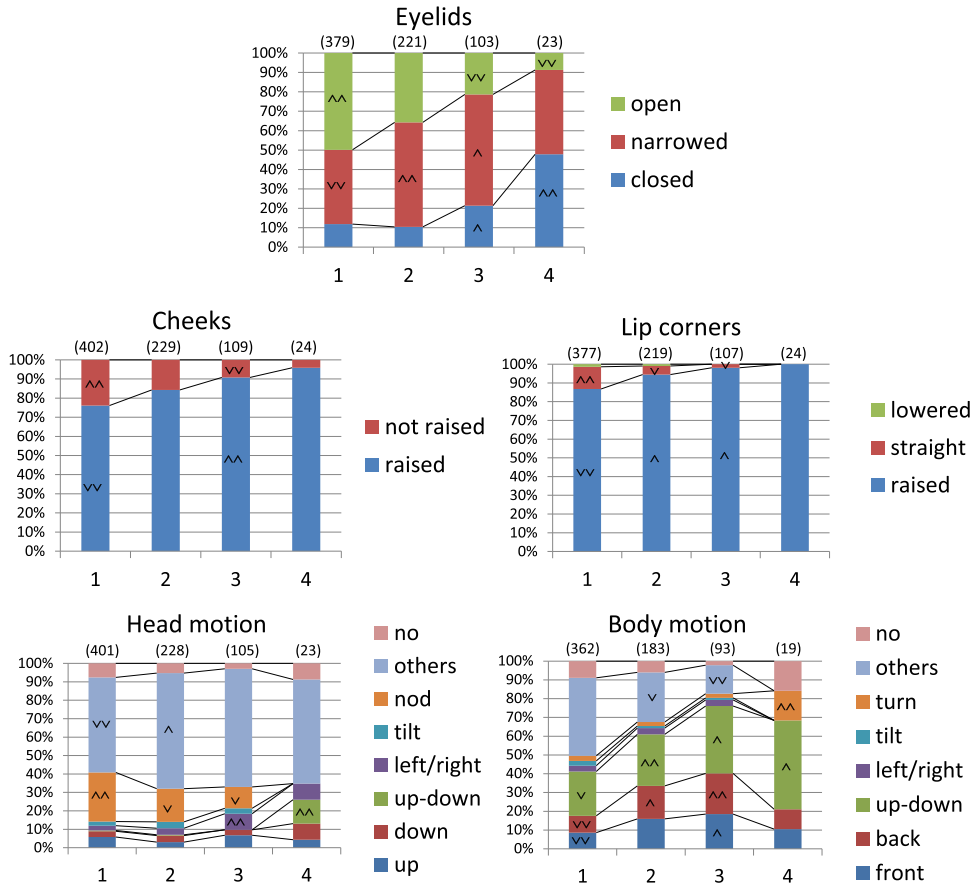

Fig. 2. Distributions of eyelid, cheek, lip corners, head, and body motion categories, for different categories of laughter intensity levels ("1" to "4"). The total number of occurrences for each laughter intensity level is shown within brackets. The symbols within the bars mean: "$\wedge$" for significantly higher occurrences ($\wedge p < 0.05$, $\wedge\wedge p < 0.01$), and "$\vee$" for significantly lower occurrences ($\vee p < 0.05$, $\vee\vee p < 0.01$), after χ^2 tests.

on average, ranging from 65 to 100% (i.e. one of the speakers showed slight smiling facial expressions over the whole dialogue). Obviously, these percentages are dependent on the person and the dialogue context. In the present data, most of the conversations were in joyful context.

The amplitudes of the pitch axis of upper body motion were also analyzed. It was observed that in both forward and backward motions, the pitch angle rotation velocities were similar ($10 \pm 5^\circ$/s for forward, and $-10 \pm 4^\circ$/s for backward directions).

IV. PROPOSED MOTION GENERATION IN AN ANDROID

Based on the analysis results, we proposed a motion generation method during laughing speech, accounting for the following four factors: facial expression control (eyelid narrowing and lip corner raising), head motion control (head pitch direction), eye blinking control at the transition from smiling face to neutral face, slight smiling facial expression (additional lip corner control) in the intervals other than laughing speech, and body motion control (torso pitch direction).

A) Android actuators and control methods

A female android robot was used to evaluate the proposed motion generation method. Figure 3 shows the external appearance and the actuators of the robot.

The current version of the android robot has 13 DOF for the face, 3 DOF for the head motion, and 2 DOF for the upper body motion. From those, the following were controlled in the present work: upper eyelid control (actuator 1), lower eyelid control (actuator 5), lip corner raising control (actuator 8, cheek is also raised), jaw lowering control (actuator 13), head pitch control (actuator 15), and upper body pitch control (actuator 18). All actuator commands range from 0 to 255. The numbers in red in Fig. 3 indicate default actuator values for the neutral position.

Figure 4 shows a block diagram of the proposed method for laughter motion generation in our android. The method requires the speech signal and the laughing speech intervals as input. In future, the laughing speech intervals can be automatically detected from the speech signal. The different modalities of the face, head, and body are driven by different features extracted from the speech signal. Lip corner raising and eye narrowing, which are the main features of smiling facial expression, are synchronized with the laughing speech intervals. Eye blinking events are also coordinated

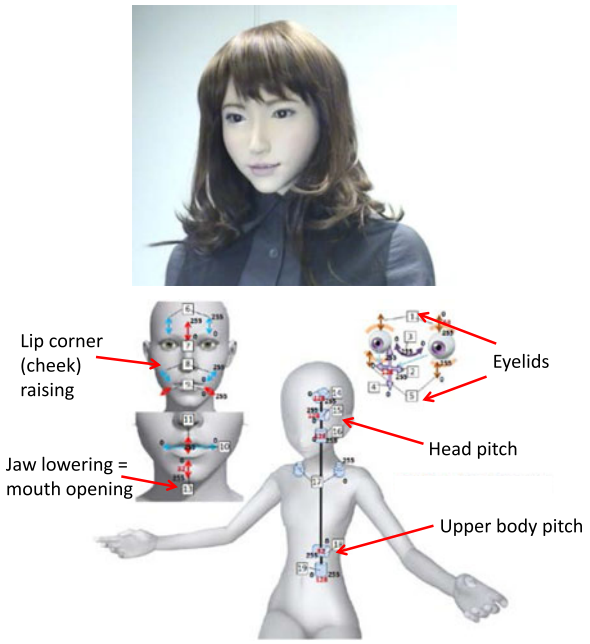

Fig. 3. External appearance of the female android and corresponding actuators.

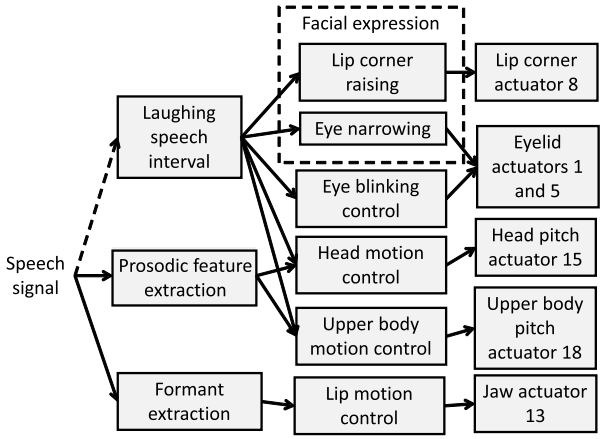

Fig. 4. Block diagram of the proposed method for motion generation during laughing speech.

with the transition from smiling to neutral facial expression. Head motion is dependent on the voice pitch, while upper body motion is dependent on laughter intensity, so these are driven by both laughter intervals and prosodic features. Lip motion (excluding lip corner raising) is dependent on the phonetic contents of the speech signals, so it is driven by formant features. Details about the motion generation methods are described as follows, for each modality.

For the facial expression during laughter, the lip corners are raised (act[8] = 200), and the eyelids are narrowed (act[1] = 128, act[5] = 128). These values were set so that a smiling face can be clearly identified. Based on the analysis results, we send the eyelid and lip corner actuator commands at the instant the laughing speech segment starts, and set the actuator commands back to the neutral position 1 s after the end of the laughing speech interval. Figure 5 shows a scheme for the synchronization of laughing speech and motion streams.

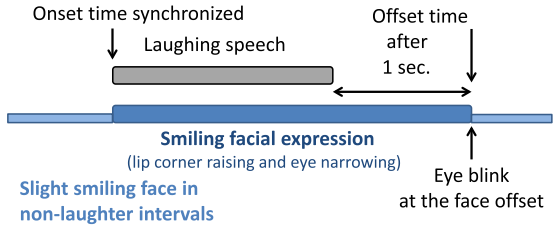

Fig. 5. Dynamic features of the facial parts during laughing speech, controlled in the laughter motion generation.

Fig. 6. Smiling face generated during laughter (left) and idle slightly smiling face generated in non-laughter intervals (right).

From the analysis results presented in Section III.D on the relationship between facial expression and eye blinking during laughter events, it has been observed that an eye blinking was usually accompanied when the facial expression turns back to the neutral face. We then decided to check the effects of controlling the eye blinking timing. The eye blinking was implemented in our android, by closing the eyes (act[1] = 255, and act[5] = 255) during a period of 100 ms, and opening the eyes back to the neutral face (act[1] = 64, act[5] = 0).

After preliminary analysis on facial motion generation, we have observed that the neutral facial expression (i.e. in non-laughter intervals) looked scary for the context of a joyful conversation. In fact, analysis results showed that the lip corners were slightly or clearly raised in 80% of the dialogue intervals. Thus, we proposed to keep a slight smiling face during non-laughter intervals, by controlling the eyelids and lip corner actuators to have intermediate values between the smiling face and the neutral (non-expression) face. For the facial expression during the idle smiling face, the lip corners are partially raised (act[8] = 100), and the eyelids are partially narrowed (act[1] = 90, act[5] = 80). Figure 6 shows the smiling face and the idle smiling face generated in the android (compare with the neutral face in Fig. 3).

Regarding head motion, the analysis results in Fig. 2 indicated that head motion is less dependent on the laughter events, in comparison to the other modalities. Rather, it is known that there is some correlation between head pitch and voice pitch, i.e. the head tends to be raised when the voice pitch is risen and *vice-versa* [22]. Thus, for the head motion control, we adopted a method for controlling the head pitch (vertical movements) according to the voice pitch (fundamental frequencies, *F*0). Although this control strategy is not exactly what humans do during speech,

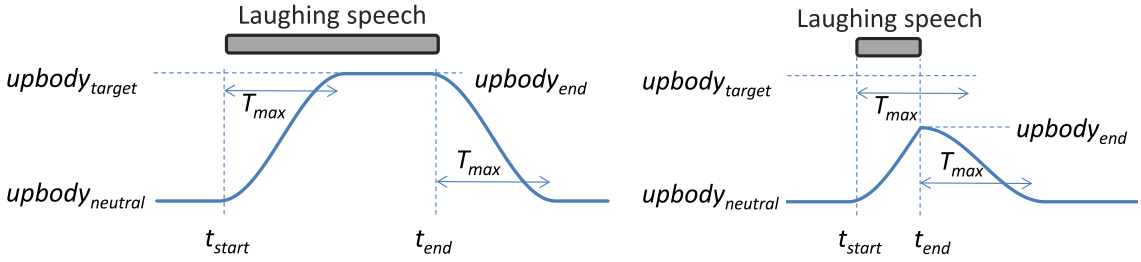

Fig. 7. Examples of upper body motion control synchronized with the laughing speech interval.

we can expect that natural head motion can be generated during laughing speech, since humans usually tend to raise their head for high *F*os especially in inhaled laughter. The following expression is used to convert *F*o values to the head pitch actuator 15, which has the effects of raising/lowering the head for high/low *F*os.

$$act[15] = head_{neutral} + (Fo - center\_Fo) \times Fo\_scale, \quad (1)$$

where $head_{neutral}$ is the actuator value for the neutral head pose (128, for our android), *center_Fo* is the speaker's average *F*o value (around 120 Hz for male, and around 240 Hz for female speakers) converted to semitone units, *F*o is the current *F*o value (in semitones), and *Fo_scale* is a scale factor for mapping the *F*o (voice pitch) changes to head pitch movements. In the current experiments, *Fo_scale* factor was set in a way that a 1 semitone change in voice pitch corresponds to approximately 1° change in head pitch rotation.

For the upper body motion control, we proposed a method for moving the upper body in the forward and backward directions, according to the expressions (2) and (3). These expressions are simply half cosine functions to smoothly move the actuators from and back to the neutral position.

$$act[18][t] = act_{neutral}32 + upbody_{target} \times \frac{1 - \cos(\pi(t/(T_{\max})))}{2}. \quad (2)$$

The upper body is moved from the start point of a laughing speech interval t_{start}, in order to achieve a maximum target angle corresponding to the actuation value $upbody_{target}$, in a time interval of T_{max}. The $upbody_{neutral}$ corresponds to the actuator value for the upper body neutral pose (32 for our android).

From the end point of the laughing speech interval, the upper body is back to the neutral position according to expression (3).

$$act[18][t] = upbody_{neutral} + (upbody_{end} - upbody_{neutral}) \times \frac{1 - \cos(\pi + \pi((t - t_{end})/(T_{max})))}{2}, \quad (3)$$

$upbody_{end}$ and t_{end} are the actuator value and the time at the end point of the laughing speech interval. Thus, if the laughter interval is shorter than T_{max}, the upper body does not achieve the maximum angle.

Table 1. The controlled modalities for generating five motion types

Motion	Controlled modalities
A	Face (eyelids + lip corners) + eye blink + head
B	Face (eyelids + lip corners) + head
C	Face (eyelids + lip corners) + eye blink
D	Face (eyelids + lip corners) + eye blink + head + idle smiling face
E	Face (eyelids + lip corners) + eye blink + head + idle smiling face + upper body

Figure 7 illustrates examples of the actuator command values for upper body motion control, based on equations (2) and (3).

Based on the analysis results of upper body motion during laughter events in Section III.D, $upbody_{target}$ was adjusted to the mean body pitch angle range of −10°, and the time interval T_{max} to reach the maximum angle was adjusted to 1.5 s (a bit longer than the human average time, to avoid jerky motion in the android).

The lip motion is controlled based on the formant-based lip motion control method proposed in [1]. In this way, appropriate lip shapes can be generated in laughter segments with different vowel qualities (such as in "hahaha" and "huhuhu"), since the method is based on the vowel formants. The jaw actuator (actuator 13) is controlled for the lip height.

B) Evaluation of the proposed motion generation method

We extracted two conversation passages of about 30 s including multiple laughter events, and generated motion in our android, based on the speech signal and the laughing speech interval information. One of the conversation passages includes social/embarrassed laughter events ("voice 1"), while the other conversation passage includes emotional/funny laughter events ("voice 2").

In order to evaluate the effects of different modalities, each motion was generated in the android according to the five conditions described in Table 1.

"Lip corners" indicates the motion control to raise the lip corners, which is also accompanied by a cheek raising motion in our android. "Eyelids" indicates the motion control to narrow the eyes. These are default control for facial expression (corresponding to Duchenne smile faces) during laughter, and are present in all conditions.

Table 2. Motion pairs for comparison of the effects of different modalities

Motion pair	Differences in the controlled modalities
A vs B	Presence/absence of "eye blink" control ("eyelids", "lip corners", and "head" are in common)
A vs C	Presence/absence of "head" control ("eyelids", "lip corners", and "eye blink" are in common)
A vs D	Absence/presence of "idle smiling face" control ("eyelids", "lip corners", "eye blink" and "head" are in common)
D vs E	Absence/presence of "upper body" control ("eyelids", "lip corners", "eye blink", "head" and "slightly smiling face" are in common)

"Eye blink" indicates the eye blinking control when turning the face expression from the smiling face to the neutral face. "Head" indicates the motion control of the head pitch from the voice pitch. "Idle smiling face" indicates the generation of a slightly smiling face during non-laughter intervals. "Upper body" indicates the motion control of torso pitch, according to the laughter duration.

Video clips were recorded for each condition and used as stimuli for subjective experiments. Pairwise comparisons were conducted in order to investigate the effects of the different motion controls. The evaluated motion pairs are described in Table 2.

In the evaluation experiments, video stimuli pairs were presented for the participants. The order of the videos for each pair was randomized. The videos were allowed to be re-played at most two times each.

After watching each pair of videos, participants are asked to grade the preference scores for pairwise comparison, and the overall naturalness scores for the individual motions, in seven-point scales, as shown below. The numbers within parenthesis are used to quantify the perceptual scores.

The perceptual preference scores for comparisons between two videos are:

- Video A is clearly more natural (−3)
- Video A is more natural (−2)
- Video A is slightly more natural (−1)
- Difficult to decide (0)
- Video B is slightly more natural (1)
- Video B is more natural (2)
- Video B is clearly more natural (3)

The perceptual naturalness scores for individual videos are:

- Very unnatural (−3)
- Unnatural (−2)
- Slightly unnatural (−1)
- Difficult to decide (0)
- Slightly natural (1)
- Natural (2)
- Very natural (3)

For conditions A and D, which appear multiple times, individual scores are graded only once, at the first time the videos are seen. Besides the perceptual scores, participants are also asked to write the reason of their judgments, if a motion is judged as unnatural.

The sequence of motion pairs above was evaluated for each of the conversation passages (voices 1 and 2).

Twelve subjects (remunerated) participated in the evaluation experiments.

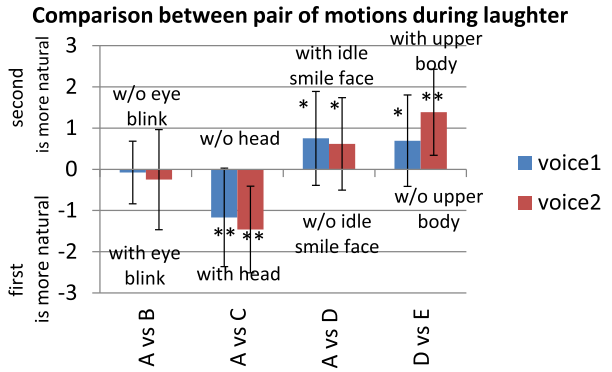

Fig. 8. Subjective preference scores between condition pairs (average scores and standard deviations). (Negative average scores indicate the first condition was preferred, while positive average scores indicate that the second condition was preferred.)

C) Evaluation results

Results for pairwise comparisons are shown in Fig. 8. Statistical analyses were conducted by t-tests (* for $p < 0.05$ and ** for $p < 0.01$ confidences). For the preference scores in the pairwise comparison, significance tests are conducted in comparison to 0 scores.

The differences between conditions A and B (with and without eye blinking control) were subtle, so that most of the participants could not perceive differences. However, subjective scores showed that the inclusion of eye blinking control was judged to look more natural for both conversation passages ("voice 1" and "voice 2").

Regarding the comparison between conditions A and C (with and without head motion control), the differences in the motion videos were clear, so that the participants' judgments were remarkable. Subjective scores indicated that the inclusion of head motion control clearly improved naturalness ($p < 0.01$) for both "voice 1" and "voice 2".

Regarding the comparison between conditions A and D (without or with idle smiling face), it was shown that keeping a slightly smiling face in the intervals other than laughing speech was judged to be look more natural ($p < 0.01$).

For the comparison between conditions D and E, the inclusion of upper body motion in condition E was judged as more natural ($p < 0.05$ for "voice 1", $p < 0.01$ for "voice 2"). The reason why differences were more evident in "voice 2" than in "voice 1" is that the conversation passage "voice 2" contained longer laughter intervals, so that the upper body motions were more evident.

Figure 9 shows the results for perceived naturalness graded for each condition.

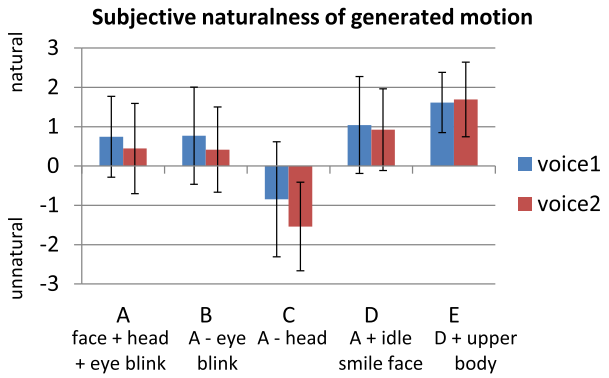

Fig. 9. Subjective naturalness scores for each condition (average scores and standard deviations).

Condition C received negative average scores, meaning that if the head does not move, the laughter motions look unnatural. Conditions A and B received slightly positive scores. Condition E received the highest scores. Overall, slightly natural to natural motions could be achieved by the proposed method including all motion control types.

V. DISCUSSION

Regarding the five evaluated motion types, firstly the difference between the conditions A and C (presence/absence of head motion) was clear in the video, and the subjective results also indicated that the proposed method for head motion generation based on *F*o was effective to improve naturalness of the android's motion. However, some of the participants pointed out that the motions would look more natural, if other axes of the head also moved.

Regarding the insertion of eye blinking at the instant the facial expression turns back to the neutral face (condition B), our preliminary evaluation had shown that it was effective to relieve unnaturalness of a sudden change in facial expression. However differences between the conditions A and B (presence/absence of eye blinking control) were not statistically significant in the evaluation results of the present work. The reason was that most participants could not notice a difference between these two conditions, since the visual difference is subtle. Nonetheless, the participants who perceived the difference judged the presence of eye blinking to be more natural. The eye blinking control is thought to work as a cushion to alleviate the unnaturalness caused by sudden changes in facial expression. The insertion of such a small motion as the eye blinking could possibly be used as a general method for other facial expressions.

The control of idle slightly smiling face in non-laughter intervals (condition D) was shown to be effective to improve the naturalness, since the conversation context was in joyful situations. However, for a more appropriate control of slightly smiling face, detection of the situation might be important.

The reason why condition E (with upper body motion) was clearly judged as more natural than condition D (without upper body motion) for "voice 2" is that it looks unnatural if the upper body does not move during long and strong emotional laughter. The proposed upper body motion control was effective to relieve such unnaturalness. Regarding intensity of the laughter, although it was implicitly accounted in the present work, by assuming high correlation between pitch and duration with intensity, it could also be explicitly modeled on the generated motions.

Regarding the comparison between "voice 1" (which includes social/embarrassed laughter events) and "voice 2" (which includes emotional/funny laughter events), the results in Figs 7 and 8 indicated that the differences on presence/absence of head and body motion were more evident in "voice 2". As stated in Section IV.C, one reason is that the conversation passage in "voice 2" contained longer laughter intervals, resulting in more evident body motions. As the proposed method constrains the amount of body movement depending on the laughter length, small body motion was generated in "voice 1". The effects of forcing different modalities in different laughter types could be an interesting topic for future investigations.

In the present study, it was shown that with a limited number of DOFs (lip corners, eyelids, head pitch, torso pitch), natural laughter motion could be generated. Although the android robot ERICA was used as a testbed for evaluation, the proposed motion generation approach can be generalized for any robot having equivalent DOFs.

The proposed generation method is useful not only for autonomous robots but also for tele-operated robot systems. In conventional tele-operated android systems, it is possible to create a function (e.g. a button) for generating smiling faces. However, in involuntary laughter, where the speaker unconsciously starts laughing in reaction to a funny situation, the operator has no time to press a smiling face button. Obviously, the facial expressions could be reproduced through image processing, by detecting the face parts (including eyes and mouth). However, face parameter extraction is not robust to light conditions, besides the need of the face parts being in the camera's field of view, with enough resolution, and with compensation of head rotation. Therefore, the method to automatically generate the laughter motion from the speech signal is important for a tele-operated robot system.

VI. CONCLUSION

Based on analysis results of human behaviors during laughing speech, we proposed a laughter motion generation method by controlling different parts of the face, head, and upper body (eyelid narrowing, lip corner/cheek raising, eye blinking, head pitch motion, and upper body pitch motion). The proposed method was evaluated through subjective experiments, by comparing motions generated with different modalities in an android robot.

Results indicated "unnatural" scores when only the facial expression (lip corner raising and eyelid narrowing) is controlled, and the most "natural" scores when head pitch, eye

blinking (at the instant the facial expression turn back to neutral face), idle smiling face (during non-laughter intervals), and upper body motion are controlled.

Topics for a further work include the control strategy of head tilt and shake axes, the investigation of eye blinking insertion for alleviating unnaturalness caused by sudden changes in other facial expressions, the detection of situation for slightly smiling face control, and the explicit modeling the laughter intensity on the generated motions. We are also currently developing a system to automatically detect laughing speech intervals from acoustic features, so that we will be able to automate the laughter motion generation from the speech signal.

ACKNOWLEDGEMENTS

We thank Mika Morita and Megumi Taniguchi, for contributing in the experiment setup and data analysis.

FINANCIAL SUPPORT

This work was supported by JST, ERATO, Grant Number JPMJER1401.

STATEMENT OF INTEREST

None.

ETHICAL STANDARDS

None.

REFERENCES

[1] Devillers, L.; Vidrascu, L.: Positive and negative emotional states behind the laughs in spontaneous spoken dialogs, in *Proc. of Interdisciplinary Workshop on The Phonetics of Laughter*, 2007, 37–40.

[2] Campbell, N.: Whom we laugh with affects how we laugh, in *Proc. of Interdisciplinary Workshop on The Phonetics of Laughter*, 2007, 61–65.

[3] Bennett, M.P.; Lengacher, C.: Humor and laughter may influence health: III. Laughter and health outcomes. *Evidence-Based Complementary Altern. Med.*, **5** (1) (2008), 37–40.

[4] Esposito, A.; Jain, L.C.: Modeling emotions in robotic socially believable behaving systems, in Esposito A. and Jain L.C. (eds.), *Toward Robotic Socially Believable Behaving Systems - Volume I*, Ch. 2, Springer, Cham, 9–14, 2016.

[5] Mehu, M.: Smiling and laughter in naturally occurring dyadic interactions: relationship to conversation, body contacts, and displacement activities. *Hum. Ethol. Bull.*, **26** (1) (2011), 10–28.

[6] Dupont, S. *et al*. Laughter research: a review of the ILHAIRE project, in Esposito A. and Jain L.C. (eds.), *Toward Robotic Socially Believable Behaving Systems – Volume I*. Ch. 9, Springer, Cham, 147–181, 2016.

[7] Ishi, C.; Liu, C.; Ishiguro, H.; Hagita, N.: Evaluation of formant-based lip motion generation in tele-operated humanoid robots, in *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems* (*IROS* 2012), October, 2012, 2377–2382.

[8] Ishi, C.T.; Liu, C.; Ishiguro, H.; Hagita, N.: Head motion during dialogue speech and nod timing control in humanoid robots, in *Proc. of 5th ACM/IEEE Int. Conf. on Human-Robot Interaction* (*HRI* 2010), 2010, pp. 293–300.

[9] Liu, C.; Ishi, C.; Ishiguro, H.; Hagita, N.: Generation of nodding, head tilting and gazing for human-robot speech interaction. *Int. J. Humanoid Robotics* (*IJHR*), **10** (1), (2013).

[10] Kurima, S.; Ishi, C.; Minato, T.; Ishiguro, H.: Online speech-driven head motion generating system and evaluation on a tele-operated robot, in *Proc. IEEE Int. Symp. on Robot and Human Interactive Communication* (*ROMAN* 2015), 2015, 529–534.

[11] Ishi, C.; Hatano, H.; Ishiguro, H.: Audiovisual analysis of relations between laughter types and laughter motions, in *Proc. Speech Prosody* **2016**, 2016, 806–810.

[12] Ishi, C.; Funayama, T.; Minato, T.; Ishiguro, H.: Motion generation in android robots during laughing speech, *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems* (*IROS* 2016), Oct., 2016, 3327–3332.

[13] Niewiadomski, R.; Pelachaud, C.: Towards multimodal expression of laughter, in *Proc. of Int. Conf. on Intelligent Virtual Agents* (*IVA* 2012), 2012, 231–244.

[14] Niewiadomski, R. *et al*. Laugh-aware virtual agent and its impact on user amusement, in *Proc. of Int. Conf. on Autonomous Agents and Multiagent Systems* (*AAMAS*2013), 2013, 619–626.

[15] Ding, Y.; Prepin, K.; Huang, J.; Pelachaud, C.; Artieres, T.: Laughter animation synthesis, in *Proc. of Autonomous Agents and Multiagent Systems* (*AAMAS*2014), 2014, 773–780.

[16] Niewiadomski, R.; Mancini, M.; Ding, Y.; Pelachaud, C.; Volpe, G.: Rhythmic Body Movements of Laughter, in *Proc. of 16th Int. Conf. on Multimodal Interaction*, 2014, 299–306.

[17] Breazeal, C.: Emotion and sociable humanoid robots. *Int. J. Hum. Comput. Stud.*, 59, 119–155 (2003).

[18] Zecca, M.; Endo, N.; Momoki, S.; Itoh, K.; Takanishi, A.: Design of the humanoid robot KOBIAN - preliminary analysis of facial and whole body emotion expression capabilities-, *Proc. of the 8th IEEE-RAS Int. Conf. on Humanoid Robots* (*Humanoids* 2008), 2008, 487–492.

[19] Dobs, K.; Bulthoff, I.; Schultz, J.: Use and usefulness of dynamic face stimuli for face perception studies – a review of behavioral findings and methodology. *Front. Psychol.*, **9** (1355) (2018), 1–7.

[20] Ochs, M.; Niewiadomski, R.; Brunet, P.; Pelachaud, C.: Smiling virtual agent in social context. *Cogn. Process*, **13** (2) (2012), 519–532.

[21] Ekman, P.; Davidson, R.J.; Friesen, W.V.: The Duchenne smile: emotional expression and brain physiology II. *J. Pers. Soc. Psychol.*, **58** (2) (1990), 342–353.

[22] Yehia, HC; Kuratate, T; Vatikiotis-Bateson E. Linking facial animation, head motion and speech acoustics. *J. Phonetics*. **30**, (2002) 555–568.

Carlos T. Ishi received the B.E. and M.S. degrees in electronic engineering from the Instituto Tecnológico de Aeronáutica (Brazil) in 1996 and 1998, respectively. He received the PhD degree in engineering from The University of Tokyo (Japan) in 2001. He worked at the JST/CREST Expressive Speech Processing Project from 2002 to 2004 at ATR Human Information Science Laboratories. He joined ATR Intelligent Robotics and Communication Labs, since 2005, and is currently the group leader of the Dept. of Sound Environment Intelligence at ATR Hiroshi Ishiguro Labs. His research topics include speech processing applied for human-robot interaction.

Takashi Minato obtained a Ph.D. degree in Engineering from Osaka University in 2004. In December 2001, he was a researcher at CREST, JST and worked as an assistant professor of the Department of Adaptive Machine Systems, Osaka University from September 2002 until 2006. In June 2006, he began to work as a researcher for ERATO, JST. Since January 2011, he has been working as a researcher in the ATR Hiroshi Ishiguro Laboratory.

Hiroshi Ishiguro received a D. Eng. in systems engineering from the Osaka University, Japan in 1991. He is currently a professor in the Department of Systems Innovation in the Graduate School of Engineering Science at Osaka University (2009-) and a distinguished professor of Osaka University (2017-). He is also a visiting director (2014-) (group leader: 2002–2013) of the ATR Hiroshi Ishiguro Laboratories and an ATR fellow. His research interests include sensor networks, interactive robotics, and android science. He received the Osaka Cultural Award in 2011. In 2015, he received the Prize for Science and Technology (Research Category) from the Minister of Education, Culture, Sports, Science and Technology (MEXT).