Motion Trajectory Grouping for Human Head Gestures Related to Facial Expressions

Maja Kocoń[1] [0000−0001−7567−9773]

West Pomeranian University of Technology Szczecin,
Sikorskiego 37, Szczecin, Poland
maja.kocon@zut.edu.pl

Abstract. The paper focuses on human head motion in connection with facial expressions for virtual-based interaction systems. Nowadays, the virtual representation of a human, with human-like social behaviour and mechanism of movements, can realize the user-machine interaction. The presented method includes the head motion because head gestures transmit additional information about the interaction's situational context. This paper presents head motion analysis based on the rotation of rigid objects technique for virtual-based interaction systems. First, we captured the head gestures of a human subject, expressing three basic facial expressions. The proposed motion model was described using three non-deformable objects, which reflect the neck and head skeleton movement's character. Based on the captured actions, the motion trajectories were analyzed, and their characteristic features were distinguished. The obtained dependencies were used to created new trajectories using piecewise cubic Hermite interpolating polynomial (PCHIP). Furthermore, the trajectories assigned to the rigid model have been grouped according to their similarities for a given emotional state. This way, using a single master trajectory and a set of coefficients, we were able to generate the whole set of trajectories for joint rotations of the head for the target emotional state. The resulting rotation trajectories were used to create movements on the three-dimensional human head.

Keywords: Trajectory grouping · Emotion intensity · Rigid object · Motion similarity.

1 Introduction

Nowadays, intuitive and robust interaction between human and machine is still a challenging problem. One of the methods to improve traditional communication and to humanize machine is the use of some virtual reality elements. First of all, we need to focus on the face and head motion aspects. Facial expression and head gestures provide information about emotions, intentions and mood, therefore it is the main communication channel in interaction [1]. Although the face mainly transmits the person's emotional state, it does not adequately describe the situational context. Therefore, in the analysis and synthesis of human motion, it is essential to consider head movement aspects. The most popular

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77961-0_26

perception studies about the meaning of the head motion in interpersonal communication have been described in [11], and [8], where the significant role of the movement in nonverbal communication was emphasized. For example, the essential head gesture such as tilting, nodding or shaking is important in active listening, where the head nodding can substitute verbal information like "yes" or "no" [23], [2], [13] or can be used instead "this one" / "that one" when you point at something [28]. Additionally, the head movements determine the meaning of the words [27], [9] simplify the verbal expressions and specify the intensity of the emotions [6].

The importance of the head in emotion perception is described in [12], where Hess et al. proved that head position strongly influences reactions to negative emotions like fear and anger. Head movement with anger and fear expression strengthens the recognition of these emotions. In [21] Mignault et al. demonstrated that gesture called extension is correlated with joy expression, and motion called flexion is associated with "inferiority emotions" like guilt. Additionally, researchers have observed that during the conversation, the head movements are not random; these movements are used to influence interaction [10]. Head pose affects communication, complete verbal message and can improve the virtual person's realism, which is essential for many applications. The 3D model of the human head is increasingly used in many applications related to HMI [30], such as serious games [4], user-friendly interfaces [5], personalized agents in telepresence systems [29], driver assistance systems [26], or social robotics [3], [18].

2 Related Works

There are many different methods for generating head movements. Motion can be generated based on captured characteristic points from video sequences, based on spoken words, or generated randomly. Most human head studies can be categorized as rule-based or data-driven frameworks. The first methods define rules for head gestures that build semantic labels, such as shaking, nodding, and tilting. In this case, the set of head movements are limited, which results in repeatability sequences. In comparison, data-driven methods use motion capture sequences and based on given head motion trajectories, new realizations of head movements are generate [24].

One popular idea used for motion synthesis is the head motion prediction based on the audio signals. Marsella et al. [20] proposed an approach for synthesizing a three-dimensional character movement based on the acoustic signal's prosodic analysis. They used the acoustic analysis to select the essential category of behaviour correlated with emotional state and words. Their proposed system can synthesize the head's different facial expressions and movements by transitioning from one gesture to another using co-articulation with other animations. Their method consists of semantic and prosody, and they can generate more appropriate virtual human motion than only the prosody method.

Based on the analysis of the relation between head gestures and conversation actions, Liu et al. [19] proposed an approach to generate head nodding and

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77961-0_26

tilting where motion rules were obtained from human interaction features. In the first step, they labelled the sentence in the conversation database with selected head gestures, and then they created a correlation between sentence and proper head movement. They found an association between head nodding and the last syllable of phrase limits and head tilting when the subject was thinking or embarrassed. They used fixed shapes trajectories for head nods and head tilts, where for example, the for nodding and tilting intensity and the duration was kept the same; therefore, the effects of timing can be estimated. To assess the naturalness of the head motion on human-like robots, they used perceptual evaluation. They find that the naturalness is improved when head nodding and tilting are incorporated into the structure.

Lee et al. [17] describes a framework to generate head motion and other motion like eyelid and eye gaze motion. Proposed approach based on Gaussian Mixture Models (GMM) and gradient descent optimization algorithm to create head motion from speech attributes. Nonlinear Dynamic Canonical Correlation Analysis model determines the eye gaze from head motion synthesis. To obtain the current head gesture, they need two previous frames and prosodic features for the actual frame. Then head postures are calculated by maximization the last joint distribution exploiting gradient descent.

In [25] for head poses parallel generation frame with emotional, synthetic speech aligned with the real speech from motion capture sequences is creating. This frame is used to train the initial models or adapt the previous models obtained with natural speech. The main advantage of this solution is the reduction of mismatch. Besides, head movement with speech-driven methods can ignore the message context, even when speech is synchronized with head movements.

Another idea used for head motion synthesis based on the audio signals is described in [9]. Greenwood et al. have explained the concept of generating head movements from voice. For several conversational scenarios, six hours of natural and expressive speech were captured. Then, head motion and speech have been connected using deep two-way Long Short Term Memory networks, which allows analyzing the language's long structure. Finally, they have obtained and extended the model by conditioning with previous movements using a Conditional Variational Autoencoder.

3 Proposed Approach

Human motion synthesis is an essential topic in man-machine interaction systems. In this case, a machine definition can refer to a human-like robot or avatar – a virtual representation of the human displayed on the screen. HMI researchers' main goal is to design the interaction to be more like interpersonal communication, closer to human-like perception and understanding. For this purpose, non-verbal signals such as facial expressions and head gestures are widely used.

The method presented in this paper is preliminary and can be extended with additional rigid elements for more head gestures correlated with different emotional intensity.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77961-0_26

3.1 Head Gestures

Head gestures were analyzed for three universal emotional states [7]: joy, sadness and fear. The selected emotions were characterized by the most significant facial expression muscles' activity and the most active head movements. Emotions can also be expressed with different intensity. It depends on the situational context and the person's character or mood. Therefore, we took the intensity of selected emotions into account, and we have used three intensities of emotions: weak, medium, strong. Head movement is based on rotation, and possible rotations for the human head are shown in Figure 1. They are based on our previous work [15], where for each subtype of expression, we have estimated ranges of movements.

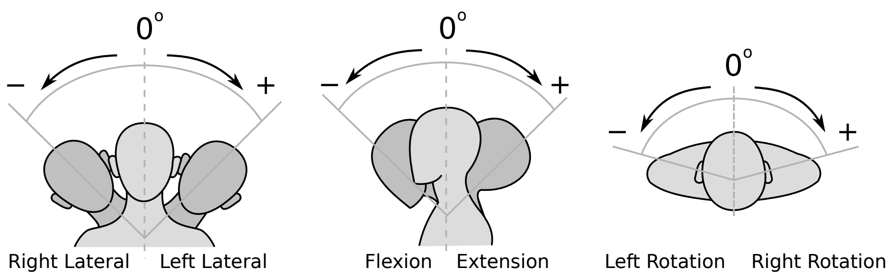

Fig. 1. Selected gestures of the human head.

3.2 Rigid Head Model

In contrast to the face, where we have elastic movements of mimic muscles, the human head can be described as a solid body. Therefore, we have used the set of non-deformable, rigid elements connected by joints for the head's action. Based on the anatomical structure of the head, we have selected three rigid elements that correspond to the head and neck. In this way proposed head consists of three segments that correspond to the neck and skull, which indicates the natural movements of the head [16]: C_1, C_2, C_3, where element C_1 indicate pitch motion, element C_2 roll rotation and element C_3 yaw motion (as depicted in Figure 2). We used a kinematic chain with three degrees of freedom because head gestures generation with three degrees of freedom is more useful than only one or two angles [22]. The full description of the motion for three rigid elements for one of the emotion called joy with high intensity is shown in the Figure 3, where C_1, C_2, C_3 refers to the rigid elements, and x, y, z rotations around the axis, as shown in Figure 2. Analyzing the individual rotational motion of components, we decided to examine the relations between rotations of specific joints.

To determine the properties of human head movements, we have captured trajectories of joints for every element of the rigid model. The data describes three emotional states: joy, sadness and fear with various intensities grades: four

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77961-0_26

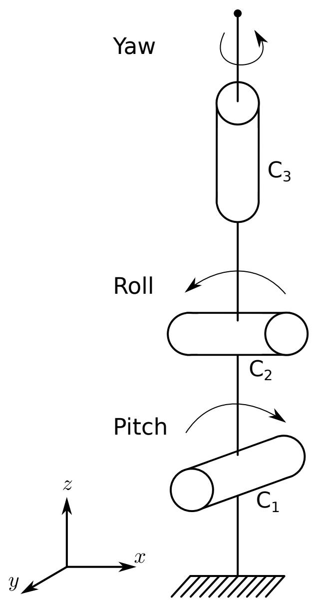

Fig. 2. The chain of rigid elements defined for the head.

for joy, seven for sadness and six for fear. Since three coordinates represent every joint position, the resulting number of trajectories is equal to $3 \cdot 3 \cdot 17 = 153$.

4 Rotation Trajectories Analysis

Analysis of the movements of rigid elements was made for three primary emotions with different intensities. In this case, we have used three rigid items in three-dimensional space for head movements modelling, as shown in Figure 2. Motion trajectories were obtained for all rigid elements in all axes. Motion data were obtained from captured characteristic points of the head, and from the collected motion data, a trajectory was obtained using piecewise cubic Hermite interpolating polynomial [14].

4.1 Ranges of Rotation Angles

In the first stage, an analysis of rotation ranges was performed, and we were able to determine the maximum deflection of a rigid element that can be achieved during rotation. The range of rotation angles in trajectory was calculated using the following formula:

$$R = |max[p(n)] - min[p(n)]| \,, \tag{1}$$

where $p(n)$ - trajectory of angle values $n = 0, \ldots, N-1$, and N is the number of points in the trajectory.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77961-0_26

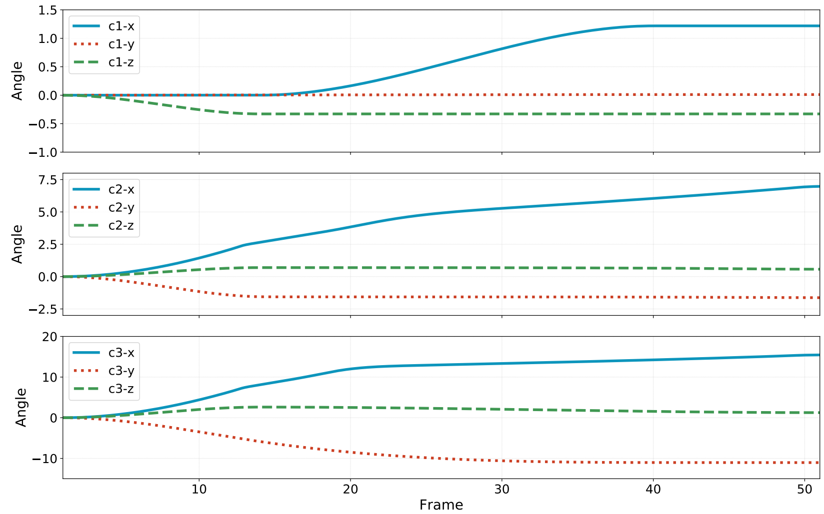

Fig. 3. Motion trajectories for strong intensity joy emotion.

Based on the analysis, we observed that no rotation actions occur in 35% cases with different emotions, rigid elements and axes. On the other hand, the participation of individual rigid elements in the rotation is shown in Table 1.

Table 1. Cases with lack of rotation actions for each rigid element.

Element	X axis	Y axis	Z axis
C_1	10%	95%	95%
C_2	0%	29%	41%
C_3	23%	41%	23%

According to the measurements, the most extensive range of motion was happen for C_2 and C_3 elements for pitch and roll head movements, while the smallest occurs for the yaw head movement. Element C_1 has the lowest range of rotations for all actions. In the next step, we have computed ranges for all trajectories in all emotional states. Table 2 presents a small subset of trajectories with the highest and the lowest values of R (constant trajectories with $R = 0$ were ignored). For all rigid elements and coordinates, the most extensive range was observed for strong joy and strong sadness in pitch movement. The C_2 and C_3 elements are responsible for the upper part of the head rotation and are very active for sadness state. However, the component C_1 is responsible for the lower part of the neck shows the smallest range for the fear state. The roll movements are the most active part when sadness and joy expressions occur. The C_2 and

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77961-0_26

C_3 elements are responsible for the movement of the upper part of the head and demonstrate the significant activity. In the case of yaw movement, the highest activity was observed for C_2 and C_3 elements.

Table 2. Example cases with the highest and lowest R values.

R	Emotion	Element	Rotation
28.6059	strong sadness	C_2	pitch
28.6059	strong sadness	C_2	pitch
18.6531	medium joy	C_2	pitch
16.7949	strong sadness	C_2	roll
15.4474	strong joy	C_3	pitch
15.0925	strong sadness	C_2	pitch
15.0925	medium sadness	C_2	pitch
...	...	...	...
0.2767	medium sadness	C_2	roll
0.2401	weak sadness	C_3	yaw
0.2207	weak fear	C_1	pitch
0.0275	strong sadness	C_3	roll
0.0112	strong joy	C_1	roll
0.0078	weak joy	C_3	roll

4.2 Trajectories Similarity

A preliminary analysis of obtained trajectories for considered emotional states indicates a noticeable similarity between them. Therefore, we have decided to compare them to create groups where all trajectories but one can be obtained by scaling and translating the single selected trajectory. For this purpose, we have compared all combinations of trajectories in pairs and compute the mean squared error (MSE) since the length of every trajectory is the same:

$$M = \frac{1}{N} \sum_{n=0}^{N-1} \left[p_1(n) - p_2(n)\right]^2 , \tag{2}$$

where: $p_1(n)$ represents the first trajectory, $p_2(n)$ denotes the second trajectory and N is the number of points in single trajectory. The case when $M = 0$ shows that $p_1(n) = p_2(n)$ for every n. Thus, the value M can be treated in this study as a value of similarity.

To calculate the similarity between trajectories, we create a set of pairs for all cases in our dataset. The number of combinations without repetitions was equal to $\binom{2}{17 \cdot 9}$ resulting in $\binom{2}{153} = 11628$ pairs to compare. An interesting situation occurs when $M = 0$ and $R > 0$, which means that both trajectories are identical and represents the change over time (for $R = 0$ both trajectories are constant).

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77961-0_26

We have calculated all combinations between trajectories and measured their ranges for the whole set. In the result, we found seven pairs satisfying such condition and Table 3 contains the found pairs.

Table 3. Trajectories that satisfy the similarity criterion.

Set	Emotion	Element	Rotation
1	medium sadness	C_1	pitch
	strong sadness	C_1	pitch
2	medium sadness	C_2	pitch
	strong sadness	C_2	pitch
3	medium sadness	C_2	roll
	strong sadness	C_2	roll
4	medium sadness	C_2	yaw
	strong sadness	C_2	yaw
5	medium sadness	C_3	pitch
	strong sadness	C_3	pitch
6	medium sadness	C_3	roll
	strong sadness	C_3	roll
7	medium sadness	C_3	yaw
	strong sadness	C_3	yaw

The connections between trajectories concern the sadness emotion with various intensities. In the figure 4, the first trajectory is connected with the C_3 element in yaw motion for the medium intensity, which is similar to the high intensity of sadness (for the same element). The same situation can be found for C_1 element in sadness state with medium and strong intensities for pitch motion. In the evaluated comparison with the MSE criterion, we have concluded that compared trajectories have to the same phase when it reaches its target angle of rotation which is marked in that figure with the dashed line.

The duration of a single-phase animation for all emotional states was determined based on the most extended emotional reaction for an event. In our case, the longest period of movement phase is equal to 2.125 seconds (assuming the animation frame rate equal to 24 frames per second). For all trajectories in our dataset, we have calculated the time spans for rotation phases (where the angle is changing in time), and the results are presented in Figure 5. It is visible that the occurrence of short periods of rotations is frequent in the considered dataset.

4.3 Trajectories Grouping

An essential part of the rotation trajectory is the phase when the angle remains constant up to the end of the animation. Having this phase in mind and considering all trajectories, a set of the group may be created with the same phases. In any of such groups, the trajectories are interrelated, which means that selecting one parent trajectory are interrelated from the group, the remaining trajectories

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77961-0_26

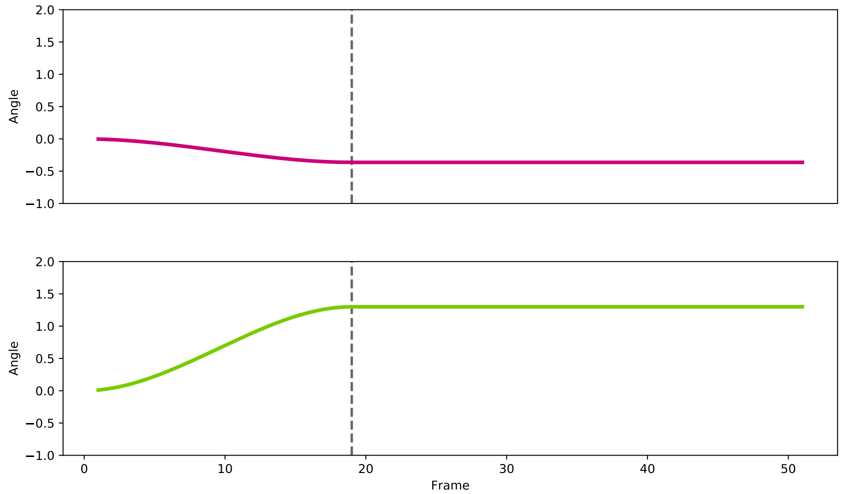

Fig. 4. Comparison of trajectories showing the similarity between elements C_3 in yaw and C_1 in pitch rotations.

can be computed by scaling and translating the parent trajectory. In the result of grouping for our dataset, twelve groups were obtained. Having selected parent trajectory $p_1(n)$, the rest of the trajectories can be described by the following formula:

$$p_j(n) = c \cdot [p_1(n) - d_1] \cdot \frac{r_2}{r_1} + d_2, \tag{3}$$

where $j = 2, \ldots, J-1$ and J denotes the number of trajectories and $p_j(n)$ is the j-th child trajectory. The remaining parameters determine the transformation properties. The selection of $p_1(n)$ from the group is arbitrary.

The first parameter $c \in \{-1, 1\}$ defines if the child trajectory has to be mirrored vertically and can be determined as follows:

$$c = \begin{cases} 1 \text{ if } s_x = s_y \\ -1 \text{ if } s_x \neq s_y \end{cases}, \tag{4}$$

where: $s_x = sgn(p_1(0) - p_1(N-1))$, $s_y = sgn(p_j(0) - p_j(N-1))$ and $sgn(x)$ is signum function. The d_1, d_2 parameters are used to align both trajectories:

$$\begin{aligned} d_1 &= max[p_1(n)], \\ d_2 &= \begin{cases} min[p_j(n)] \text{ if } c = -1 \\ max[p_j(n)] \text{ if } c = 1 \end{cases}. \end{aligned} \tag{5}$$

Finally, the adaptation of ranges for both trajectories is realized by scaling the $p_1(n)$ trajectory by the ratio of ranges r_2/r_1, where r_1 is the range of parent and r_2 child trajectory:

$$\begin{aligned} r_1 &= \left| max[p_1(n)] - min[p_1(n)] \right|, \\ r_2 &= \left| max[p_j(n)] - min[p_j(n)] \right|. \end{aligned} \tag{6}$$

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77961-0_26

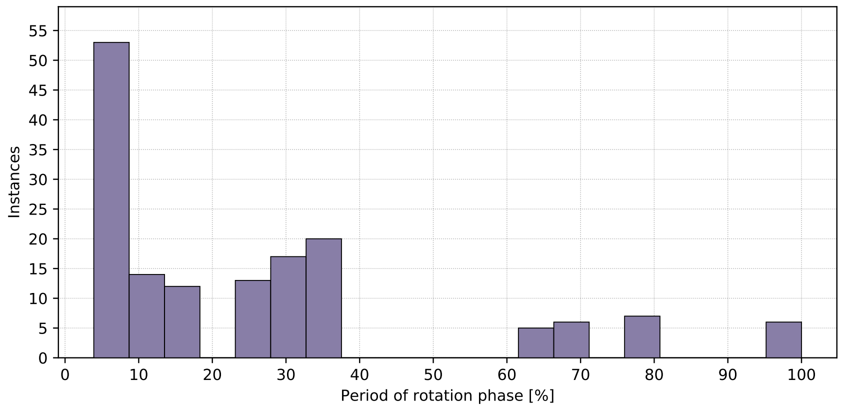

Fig. 5. The length of rotation actions for the length of the single phase of animation.

Since the ratio is always positive, its value determine if the dynamic range of parent trajectory is extended ($r_2/r_1 > 1$) or compressed ($r_2/r_1 < 1$).

4.4 Final Rotation of the Head

As an example of using the proposed mechanism, we have selected a group of trajectories with rotation phase in 19 frames. We have chosen a first trajectory (intense sadness emotional state) in the group depicted in Figure 6, where $d_1 = 0$ and $r_1 = 1.3564$. Obtained rotations based on the trajectory are presented in Figure 7.

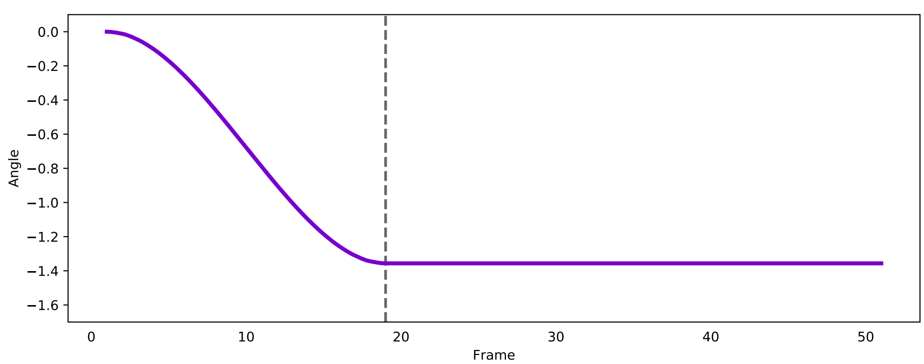

Fig. 6. The parent trajectory of child trajectories in the group for strong sadness state rotations.

Next, using it as the parent trajectory, we have calculated the parameters for the rest trajectories in the group. Table 4 summarizes the coefficients obtained

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77961-0_26

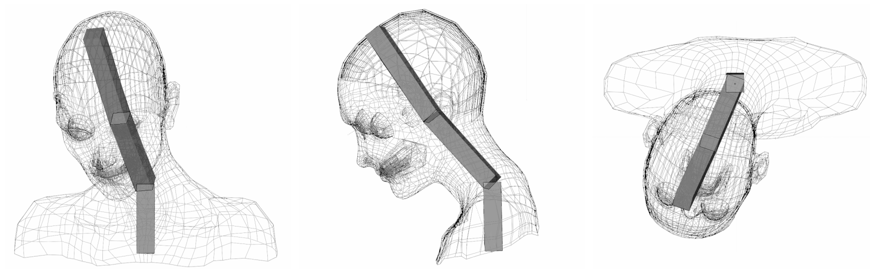

Fig. 7. Example configuration of rigid elements in strong sadness state for the front, side and top view.

for one group where $p_1(n)$ represents weak joy state, element C_1 with pitch movement. Selecting all trajectories for intense sadness emotional state, we have rendered a few frames of animation for this case presented in Figure 8.

Fig. 8. Example phases of head animation for strong sadness emotional state.

The final data for animation can be expressed as a single reference trajectory with set parameters to generate the rest of the trajectories for a specific emotional state.

5 Conclusions

The principal aim of our work was to create relations between the facial expressions and head movements for the better virtual human-machine interaction. For this purpose, we analyzed motion trajectories for three basic emotions for three different intensities of these emotions. The obtained results were used to create

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77961-0_26

Table 4. The coefficients of child trajectories obtained for one group with rotation phase performed in 19 frames.

Emotion	**Element**	**Rotation**	**c**	d_2	r_2	$\mathbf{r_2/r_1}$
weak joy	C_2	pitch	-1	0	4.8811	3.5986
weak joy	C_3	pitch	-1	0	5.1591	3.8035
weak joy	C_3	roll	1	0	0.0078	0.0058
weak joy	C_3	yaw	-1	0	4.5821	3.3781
medium sadness	C_1	pitch	-1	0.0104	1.2895	0.9507
medium sadness	C_2	pitch	1	-0.1220	15.0925	11.1269
medium sadness	C_2	roll	-1	0.0777	9.6130	7.0871
medium sadness	C_2	yaw	1	-0.0144	1.7840	1.3152
medium sadness	C_3	pitch	1	-0.0168	2.0816	1.5347
medium sadness	C_3	roll	1	-0.0002	0.0275	0.0203
medium sadness	C_3	yaw	1	-0.0029	0.3605	0.2658
strong sadness	C_1	pitch	-1	0.0104	1.2895	0.9507
strong sadness	C_2	pitch	1	-0.1220	15.0925	11.1269
strong sadness	C_2	roll	-1	0.0777	9.6130	7.0871
strong sadness	C_2	yaw	1	-0.0144	1.7840	1.3152
strong sadness	C_3	pitch	1	-0.0168	2.0816	1.5347
strong sadness	C_3	roll	1	-0.0002	0.0275	0.0203
strong sadness	C_3	yaw	1	-0.0029	0.3605	0.2658

movement on the three-dimensional head of a human. The results indicate that based on the similarity of the trajectories, the rotations can be performed using a limited set of the original trajectories in the database. We have presented a technique on how to group a set of trajectories using similarity measure between them and how to restore them inside the group using a single trajectory and additional coefficients. We have used the mean square error measure to determine the similarity level between pairs of trajectories. The mean absolute error, root mean squared error or other measurements can also be used for this task. However, selecting a proper similarity measure taking the problem's specificity into mind is not a trivial task and needs further investigation. The presented work can be extended by adding new rigid elements and increasing the number of head gestures correlated with different emotional intensities. In future work,

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77961-0_26

a scheme for selecting a base trajectory in groups using optimization techniques is planned.

References

1. Boker, S.M., Cohn, J.F., Theobald, B.J., Matthews, I., Brick, T.R., Spies, J.R.: Effects of damping head movement and facial expression in dyadic conversation using real–time facial expression tracking and synthesized avatars. Philosophical Transactions of the Royal Society of London B: Biological Sciences **364**(1535), 3485–3495 (2009)
2. Boker, S.M., Cohn, J.F., Theobald, B.J., Matthews, I., Mangini, M., Spies, J.R., Ambadar, Z., Brick, T.R.: Something in the way we move: Motion dynamics, not perceived sex, influence head movements in conversation. Journal of Experimental Psychology: Human Perception and Performance **3**(37), 874–891 (2011)
3. Breazeal, C., Thomaz, A.L.: Learning from human teachers with socially guided exploration. In: ICRA. pp. 3539–3544. IEEE (2008)
4. Cai, Y., van Joolingen, W., Walker, Z.: VR, Simulations and Serious Games for Education. Springer Publishing Company, Incorporated, 1st edn. (2018)
5. Clavel, C., Plessier, J., Martin, J.C., Ach, L., Morel, B.: Combining facial and postural expressions of emotions in a virtual character. In: Ruttkay, Z., Kipp, M., Nijholt, A., Vilhjalmsson, H.H. (eds.) Intelligent Virtual Agents, Lecture Notes in Computer Science, vol. 5773, pp. 287–300. Springer Berlin Heidelberg (2009)
6. Cohn, J.F., Reed, L.I., Moriyama, T., Xiao, J., Schmidt, K.L., Ambadar, Z.: Multimodal coordination of facial action, head rotation, and eye motion during spontaneous smiles. Sixth IEEE International Conference on Automatic Face and Gesture Recognition, Proceedings. pp. 129–135 (2004)
7. Ekman, P.: Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life. Owl Books (2007)
8. Graf, H.P., Cosatto, E., Strom, V., Huang, F.J.: Visual prosody: Facial movements accompanying speech. 5th IEEE International Conference on Automatic Face and Gesture Recognition pp. 396–401 (2002)
9. Greenwood, D., Laycock, S., Matthews, I.: Predicting head pose from speech with a conditional variational autoencoder. In: Proc. Interspeech. pp. 3991–3995 (2017)
10. Gunes, H., Pantic, M.: Dimensional emotion prediction from spontaneous head gestures for interaction with sensitive artificial listeners. In: Proceedings of the 10th International Conference on Intelligent Virtual Agents. pp. 371–377. IVA 2010, Springer-Verlag, Berlin, Heidelberg (2010)
11. Harrigan, J., Rosenthal, R., Scherer, K.R.: The New Handbook of Methods in Nonverbal Behavior Research. Series in Affective Science, Oxford University Press (2005)
12. Hess, .U., Jr, R.B.A., Kleck, R.E.: The influence of head orientation on the signal value of emotional facial expressions, motivation and emotion. Motivation and Emotion **31**(2), 137–144 (2007)
13. Heylen, D.: Challenges ahead: Head movements and other social acts during conversations. In: Halle, L., Wallis, P., Woods, S., Marsella, S., Pelachaud, C., Heylen, D.K. (eds.) Proceedings of the Joint Symposium on Virtual Social Agents. pp. 45–52. The Society for the Study of AI and the Simulation of Behav. (2005)
14. Kahaner, D., Moler, C., Nash, S.: Numerical Methods and Software. Prentice-Hall, Inc., USA (1989)

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77961-0_26

15. Kocoń, M.: Influence of facial expressions on the human head movements. In: 41st International Conference on Telecommunications and Signal Processing, TSP 2018, Athens, Greece, July 4-6, 2018. pp. 1–5. IEEE (2018)
16. Kocoń, M.: Head movements in the idle loop animation. IADIS International Journal On Computer Science and Information Systems **15**(2), 137–147 (2020)
17. Le, B., Xiaohan, m., Deng, Z.: Live speech driven head-and-eye motion generators. IEEE transactions on visualization and computer graphics **18** (02 2012)
18. Ligthart, M., Hindriks, K., Neerincx, M.A.: Reducing stress by bonding with a social robot: Towards autonomous long-term child-robot interaction. In: Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction. pp. 305–306. HRI 2018, ACM (2018)
19. Liu, C., Ishi, C.T., Ishiguro, H., Hagita, N.: Generation of nodding, head tilting and eye gazing for human-robot dialogue interaction. In: 2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI). pp. 285–292 (2012)
20. Marsella, S., Xu, Y., Lhommet, M., Feng, A., Scherer, S., Shapiro, A.: Virtual character performance from speech. In: Proceedings of the 12th ACM SIGGRAPH/Eurographics Symposium on Computer Animation. pp. 25–35. SCA 2013, Association for Computing Machinery (2013)
21. Mignault, A., Chaudhuri, A.: The many faces of a neutral face: Head tilt and perception of dominance and emotion. J. Nonverbal Behav. **27**, 111–132 (06 2003). https://doi.org/10.1023/A:1023914509763
22. Mukherjee, S., Robertson, N.: Deep head pose: Gaze-direction estimation in multimodal video. IEEE Transactions on Multimedia **17**, 1–1 (11 2015)
23. Munhall, K.G., Jones, J.A., Callan, D.E., Kuratate3, T., Vatikiotis-Bateson, E.: Visual prosody and speech intelligibility: Head movement improves auditory speech perception. Psychological Science **15**(2) (2004)
24. Sadoughi, N., Busso, C.: Head Motion Generation, pp. 2177–2200. Springer International Publishing (2018)
25. Sadoughi, N., Liu, Y.P., Busso, C.: Meaningful head movements driven by emotional synthetic speech. Speech Commun. **95**, 87–99 (2017)
26. Schwarz, A., Haurilet, M., Martinez, M., Stiefelhagen, R.: Driveahead - a large-scale driver head pose dataset. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 1165–1174 (2017)
27. Sun, X., Truong, K.P., Pantic, M., Nijholt, A.: Towards visual and vocal mimicry recognition in human-human interactions. In: IEEE International Conference on Systems, Man, and Cybernetics. pp. 367–373 (2011)
28. Tojo, T., Matsusaka, Y., Ishii, T., Kobayashi, T.: A conversational robot utilizing facial and body expressions. In: Systems, Man, and Cybernetics, 2000 IEEE International Conference on. vol. 2, pp. 858–863 vol.2 (2000). https://doi.org/10.1109/ICSMC.2000.885957
29. Vidrascu, L., Devillers, L.: Real-life emotion representation and detection in call centers data. In: Tao, J., Tan, T., Picard, R.W. (eds.) Affective Computing and Intelligent Interaction, Lecture Notes in Computer Science, vol. 3784, pp. 739–746. Springer Berlin Heidelberg (2005)
30. Wang, K., Zhao, R., Ji, Q.: Human computer interaction with head pose, eye gaze and body gestures. In: 2018 13th IEEE International Conference on Automatic Face Gesture Recognition. pp. 789–789 (05 2018). https://doi.org/10.1109/FG.2018.00126

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77961-0_26