Rhythmic Body Movements of Laughter

Radoslaw Niewiadomski
DIBRIS, University of Genoa
Viale Causa 13
Genoa, Italy
radek@infomus.org

Maurizio Mancini
DIBRIS, University of Genoa
Viale Causa 13
Genoa, Italy
maurizio.mancini@unige.it

Yu Ding
CNRS - Telecom ParisTech
37-39, rue Dareau
Paris, France
yu.ding@telecom.paristech.fr

Catherine Pelachaud
CNRS - Telecom ParisTech
37-39, rue Dareau
Paris, France
catherine.pelchaud@
telecom.paristech.fr

Gualtiero Volpe
DIBRIS, University of Genoa
Viale Causa 13
Genoa, Italy
gualtiero.volpe@unige.it

ABSTRACT

In this paper we focus on three aspects of multimodal expressions of laughter. First, we propose a procedural method to synthesize rhythmic body movements of laughter based on spectral analysis of laughter episodes. For this purpose, we analyze laughter body motions from motion capture data and we reconstruct them with appropriate harmonics. Then we reduce the parameter space to two dimensions. These are the inputs of the actual model to generate a continuum of laughs rhythmic body movements.

In the paper, we also propose a method to integrate rhythmic body movements generated by our model with other synthetized expressive cues of laughter such as facial expressions and additional body movements. Finally, we present a real-time human-virtual character interaction scenario where virtual character applies our model to answer to human's laugh in real-time.

Categories and Subject Descriptors

H.1.2 [**Information Interfaces and Presentation**]: User/Machine Systems—*Human factors*; I.3.8 [**Computing Methodologies**]: Computer Graphics—*Applications*; H.5.1 [**Information Interfaces and Presentation**]: Information Interfaces and Presentation—*Artificial, augmented, and virtual realities*

General Terms

Human Factors

Keywords

laughter; nonverbal behaviors; realtime interaction; virtual character

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
ICMI '14 November 12 - 16 2014, Istanbul, Turkey
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-2885-2/14/11...$15.00.
http://dx.doi.org/10.1145/2663204.2663240

1. INTRODUCTION

During laughter, the whole body (head, torso, shoulders, face) is involved [11, 22]. Body movements are important markers allowing distinguishing laughter from smiling (besides facial and vocal expressions). The presence of body movements influences the perceived intensity of laughter [6]. Some body movements, but not all, are due to the respiration phases [22]. Despite the important role of body movements, most of the laughter synthesis algorithms for virtual characters (VCs) are limited to facial animation [5, 17].

In this paper, we propose a model of *rhythmic body movements* (RBM) based on spectral analysis of laughter episodes. We describe a multimodal synchronization schema to integrate all facial and body movements of laughter including RBM. We also present a human-virtual character (H-VC) interaction scenario where laughter is triggered: the VC's rhythmic laughter behaviors are generated in real-time and in link with human's laughter.

The bodily expressive patterns of laughter are complex. According to Ruch and Ekman [22], a laughter episode is usually composed of several repetitive laugh pulses that are caused by forced exhalation. The maximum number of these pulses is between 9 and 12 depending on volume of the lungs. Long laughter episodes include additional inhalations that can result in a higher number of pulses. Several attempts to measure laughter pulses frequency (e.g., [19] in audio modality, [2] in respiration) converge to similar results: laughter pulses have the frequency between 4 and 6 Hz. The body movements observed during laughter can be of two types [22]. Most of laughter body movements are associated with the respiratory patterns of laughter [22]. Examples of such movements are the backward movements of head and trunk that appear to ease the forced exhalations, or the upward movements of torso that are related to the inhalation phase. Laughter pulses generate also visible vibrations of the shoulders and of the trunk. Since these movements are related to respiration phases, one expects they are repetitive and have a fixed rhythm. The movements related to respiration can be accompanied by other body movements such as sideways rocking, which are, however, of different origin [22].

In the first part of the paper, we focus on a significant subset of laughter body movements, the rhythmic body movements. We use Fast Fourier Transform (FFT) to build a

RBM model. Using computer vision algorithms we extract the position of upper-body points from several laugh episodes. We conduct spectral analysis on this data. We associate the prevalent harmonics resulting from this analysis to different types of laughter movements (e.g., shake shoulders). Next we conduct statistical analysis on the harmonics' parameters (i.e., amplitude, frequency). Finally, to allow a human animator an easy control over VC's RBMs, we introduce two high-level normalized variables: *intensity* and *velocity* and we provide a mapping between these two variables and the harmonics' parameters needed to synthesize RBM. Our model allows us to control VC laughter rhythmic body movements via high-level commands like: "shake shoulders with $intensity = X$ and $velocity = Y$".

The rhythmic body movements are just one element of laughter expressive pattern. Other body and face movements have important role in laughter [22] and need to be included. We present how RBMs can be synchronized with facial and other body animations [8]. To generate a multimodal laughter animation, our model takes as input the phonetic transcription of laughter [27]. We end the paper by presenting a scenario of H-VC realtime interaction.

The paper is organized as follows. The next section presents a survey of works on visual laughter synthesis, laughter based interactive systems, and nonverbal synchronization of movements in general. Section 3 is dedicated to the analysis and synthesis of the rhythmic body movements of laughter. Section 4 focuses on multimodal synchronization and Section 5 presents an interactive scenario. We conclude our paper in Section 6.

2. STATE OF THE ART

2.1 Visual Laughter Synthesis

Few models of laughter animations were proposed recently [5, 17] but they are restricted to facial displays. The major exception is by Di Lorenzo et al. [7], who proposed an audio driven model of upper body animation during laughter.

The model defines the relationship between lungs pressure and laughter phase that can be derived from the amplitude of the audio signal. It applies an anatomically-inspired and physics-based model of human torso that is a combination of rigid and deformable components. The model is able to produce realistic animations but it is computationally expensive.

In other works, algorithms based on FFT were applied to animate VCs. Unuma et al. [25] used Fourier decomposition of motion capture data to model emotional walks, whereas Troje [24] successfully modeled walks with the first two components of Fourier series. More recently, Tilmanne and Dutoit [23] extended this approach by applying Fourier decomposition to angular representation of various style walking sequences. In a perceptive study they showed that various walking styles can be successfully modeled with the first two components.

2.2 Nonverbal Behavior Mirroring

Several virtual character systems were build that synchronize certain nonverbal behaviors. Bailenson and Yee [1] studied the Chameleon Effect in the interaction with virtual characters. In their study, a VC that mimics head nods of the human interaction partner is shown to increase perceived effectiveness. Prepin and Pelachaud [18] presented a dynamic coupling model based on the concept of an oscillator able to synchronize different nonverbal behaviors of virtual characters or of a human and a virtual character. In their model the interactants' movements become synchronized to communicate reciprocal understanding of the verbal content. Riek et al. [21] studied how a robot mimicking the human head nodding influences human perception of interaction. No effect of mimicry was observed.

2.3 Laughter Based Interactive Systems

Urbain et al. [28] developed a system able (i) to detect human laugh and (ii) to answer it by displaying virtual character pre-synthetized laugher responses. The system contains a database of pre-synthetized laugher episodes extracted from real human audio recordings that are directly replayed, and the corresponding facial animations obtained by retargeting the motion capture data. The choice of the VC's response from the database is done algorithmically and it is based on the real-time analysis of acoustic similarities with the input laughter. Fukushima et al. [9] studied laughter contagion in human-robot interaction. They used simple toy robots that play pre-registered laughter sounds when the system detects the initial human laughter. The evaluation study showed that robots' laughter was contagious and increased human laughing. Niewiadomski et al. [14] proposed an interactive TV watching scenario in which a human participant and a virtual agent watch together funny content. The agent is able to detect human laughs and can laugh in response to the detected behavior by using synthetized facial expressions and audio. Such laughter response is generated by taking into the consideration the duration and the intensity of the human laugh.

Figure 1: A frame from MMLI recordings [15]. On the left: a participant is wearing green markers. On the right: another participant is wearing a motion capture suit and green markers.

2.4 Beyond the State of the Art

Existing works may produce even very realistic animations of laughter (see Section 2.1). In this work our main aim, however, is not to improve the realism of laughter animation but to build a simple and computationally cheap method for laughter synthesis, which can be used and controlled in realtime in H-VC interaction. Importantly, we propose a procedural approach in which the synthetized body movements can be easily controlled manually by a human animator (by using a high level language such as BML, see Section 3.3) or by a machine (e.g. see Section 5).

Secondly, unlike the existing interactive realtime systems (see Section 2.3) which do not analyze human body movements when generating VC laughter, we focus on body movements of both a human and a VC. However, we do not use

the copy-synthesis method to reconstruct the whole nonverbal behavior of the human as it might be perceived to be awkward or even as mockery. Inspired by previous works on H-VC interaction (see Section 2.3), in which only some dynamic characteristics of movement are considered, we propose to tune one aspect only: laughter rhythm. Thirdly, although spectral analysis was previously used in procedural synthesis of nonverbal behaviors e.g., walk (see Section 2.1), applying this method to analyze and synthetize body movements of laughter is a novel approach.

3. MODEL OF RHYTHMIC BODY MOVEMENTS OF LAUGHTER

3.1 Data Analysis

Two types of movements are involved in laughter RBM: *torso leaning* and *shoulder vibration*. The first one is rather slow and it may have different directions: linear, e.g., front/back and left/right, or circular. The shoulder movement is much faster and it is always repetitive. It is also symmetric - both shoulders are involved and move synchronously [22, 11].

The freely accessible MMLI corpus[1][15] was used to analyze laughter body movements as it is, to our knowledge, the largest freely accessible corpus dedicated to full-body expressions of laughter. Participants wear mocap sensors, and two green markers (lightweight polystyrene balls) on shoulders. Data consists of synchronized 3D coordinates streams, 3 audio streams, and 6 video streams (640x480, 30fps).

Despite the variety of sensors used in the MMLI corpus, in our analysis only RGB video streams are used. In particular, the inertial motion capture sensors used in MMLI were not placed on participant's shoulders so the resulting data could not capture shoulder vibration.

Thirty-five episodes ($N = 35$) from five participants (2 males) containing both types of RBMs have been chosen for analysis. They contain spontaneous laughter that is either induced (watching videos task) or interactive (social games task). During all episodes participants were standing. Additional selection criteria required that (i) the green markers had to be well visible during the whole segment and (ii) that the person is not walking. An example of a video frame taken from the MMLI corpus is displayed in Figure 1. The mean duration of an episode is $2.87s$ ($\sigma = 2.22$, $min = 1.4s$, $max = 13.43s$).

- color tracking is performed by the EyesWeb XMI platform[2]: we isolate the green component of the input RGB 30 fps video stream and we apply a threshold on it; we determine the barycenter of the 2 largest pixel groups corresponding to the green markers; since shoulder movements are usually symmetric, for each episode we consider only one marker (left or right shoulder) and only its vertical coordinate $y_i(t)$;
- since the values of $y_i(t)$ are not comparable across different videos (i.e., different cameras configurations were used to record the participants), we normalize $y_i(t)$ by the participant's head height: e.g., normalized $||y_i(t)||$ is equal to 2 if the marker vertical position is equal to two times the head height, where coordinate $||y_i(t)|| = 0$ corresponds to the top image border;

[1]http://www.infomus.org/people/mmli/
[2]http://www.eyesweb.infomus.org

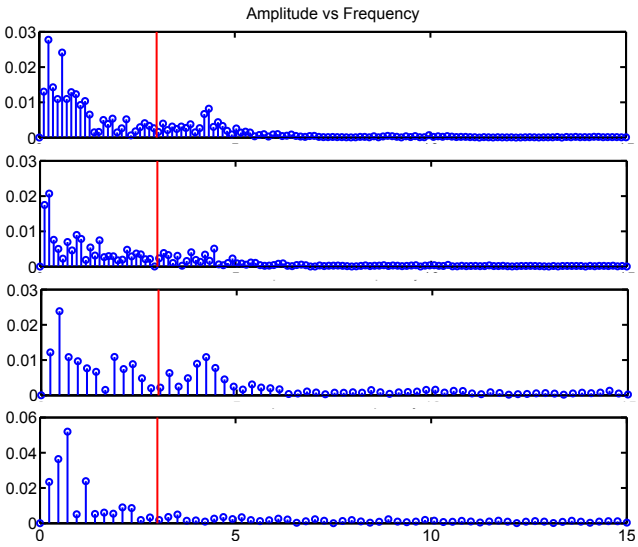

Figure 2: Examples of spectrum analysis for four different episodes

- FFT is applied to $||y_i(t)||$, that is, it is applied to the green marker's vertical coordinate for the entire episode i.

We observe that two frequency intervals can be identified in the resulting frequency spectra for each episode (see Figure 2). The first one - around $0.5 - 1.5Hz$ - contains the fundamental frequency. However, relatively strong peaks can be identified also around $3.5 - 6Hz$. Following the literature (see Section 1) we assume that: slower components, that is, frequencies around $0.5 - 1.5Hz$, correspond to torso forward/backward leaning whereas faster components, that is, frequencies around $3.5 - 6Hz$, correspond to shoulder vibration.

3.2 Signal Reconstruction with Estimation of Error

We process the input signal $s_i = ||y_i(t)||$ in 2 steps: (i) filtering and (ii) error computation. By performing low and high pass filtering we extract torso and shoulder movements. Then, we reconstruct the input signal to estimate the reconstruction error and validate the process.

3.2.1 *Filtering*

We apply two filters to $s_i = ||y_i(t)||$: the first one (low pass) allows us to extract only the torso leaning component; the second one (high pass) allows us to extract only the shoulder vibration component. Starting from the literature on laughter (see Section 1) we define a threshold at $Th = 3Hz$: the first filter allows one to isolate only frequencies in the interval $I_1 = [0, 3)Hz$; the second one allows one to isolate only frequencies in the interval $I_2 = [3, 15]Hz$.

Next, in each frequency interval I_1, I_2, we find the harmonics exhibiting the two highest local maxima of amplitude and we sort them in frequency ascending order. More precisely, for I_1, we define a matrix L of size $N \times 6$ (where $N = 35$ is the number of laughter episodes). Let $L_{i,\_}$ be the i-th row of matrix L, $L_{i,\_}$ contains the 6 torso leaning parameters of episode i, that is, $L_{i,\_} = [a_{1,i}, f_{1,i}, \phi_{1,i}, a_{2,i}, f_{2,i}, \phi_{2,i}]$,

components	mean	σ	min	max
all	0.16	0.18	< 0.01	0.72
2	0.21	0.20	0.05	0.77
4	**0.18**	**0.15**	**0.05**	**0.54**
6	0.18	0.16	0.04	0.57

Table 1: Signal reconstruction error computed on 26 signals

where $a_{j,i}$ means amplitude, $f_{j,i}$ - frequency, $\phi_{j,i}$ - phase of $\{j,i\}$-harmonics, and $f_{1,i} < f_{2,i}$, $f_{1,i}, f_{2,i} \in [0,3)Hz$. Similarly, for I_2, we define a matrix V of size $N \times 6$ such that the i-th row $V_{i,\_}$ of V contains the 6 shoulder vibration parameters of episode i, that is, $V_{i,\_} = [A_{1,i}, F_{1,i}, \Phi_{1,i}, A_{2,i}, F_{2,i}, \Phi_{2,i}]$, where $F_{1,i} < F_{2,i}$ and $F_{1,i}, F_{2,i} \in [3,15]Hz$. Also, we indicate with $L_{\_,j}$ and $V_{\_,j}$, the j-th columns of matrices L and V, respectively. For example, $L_{\_,2}$ (resp. $V_{\_,2}$) of L (resp. V) is a column vector that contains the lower frequencies, among the above selected ones, in $[0,3)Hz$ (resp. $[3,15]Hz$); while $L_{\_,5}$ (resp. $V_{\_,5}$) is a column vector that contains the higher frequencies, among the above selected ones, in $[0,3)Hz$ (resp. $[3,15]Hz$).

3.2.2 Error Computation

For each laughter episode i we reconstruct the original signal using the harmonics described by matrices L and V. We use 4 harmonics: $\{(a_{k,i}, f_{k,i}, \phi_{k,i}) | k \in \{1,2\}\} \in L$ and $\{(A_{k,i}, F_{k,i}, \Phi_{k,i}) | k \in \{1,2\}\} \in V$. Thus the reconstructed signal r_i is:

$$r_i = \sum_{k=1..2} (a_{k,i} cos(2\pi f_{k,i} t + \phi_{k,i}) + A_{k,i} cos(2\pi F_{k,i} t + \Phi_{k,i})) \quad (1)$$

Then we compute the reconstruction error for episode i as the normalized maximum difference between the original signal s_i and r_i:

$$E_i = \frac{\max(|s_i - r_i|)}{\max(s_i) - \min(s_i)} \quad (2)$$

The mean error E, computed according to the above equation on 26 episodes, is $E = 0.18$ (see Table 1).[3] We finally check how the reconstruction mean error depends on the number of harmonics. We compare 4 harmonics reconstruction error with the following additional cases: (i) reconstruction performed on the entire FFT spectrum; (ii) reconstruction performed on 2 harmonics, the prevalent one in $[0,3)Hz$ and the prevalent one in $[3,15]Hz$; (iii) reconstruction performed on 6 harmonics, 3 in $[0,3)Hz$ and 3 in $[3,15]Hz$ selected in a similar way as for L and V. The corresponding computed errors are reported in Table 1. It can be seen that the signal reconstructed with the use of four harmonics provides satisfactory results (see Figure 3). Adding more harmonics does not reduce significantly the mean error. Thus, it is a good compromise between the accuracy and the number of parameters needed to describe the signal.

[3]Remaining 9 episodes had only one local maximum of amplitudes in one of two considered ranges and, thus, they have been discarded.

3.3 Model of Rhythmic Movements in Laughter

To generate RBMs with our model one needs to specify the type of the movement (i.e., torso leaning, shoulders vibration), the desired duration of the movement t, and two additional variables that describe the movement: intensity $INT \in [0,1]$ and velocity $VEL \in [0,1]$ (see Section 3.3.2 for details). For a frame to be displayed at time t, the model outputs two values: torso leaning $l(t)$ and shoulder vibration $v(t)$. Accordingly to results presented in Section 3.1 and 3.2 each of these body movements is modeled using two harmonics:

$$\begin{aligned} l(t) &= a_1 cos(2\pi f_1 t + \phi_1) + a_2 cos(2\pi f_2 t + \phi_2) \\ v(t) &= A_1 cos(2\pi F_1 t + \Phi_1) + A_2 cos(2\pi F_2 t + \Phi_2) \end{aligned} \quad (3)$$

Thus, body movements in our model can be fully controlled within 12 parameters. However, controlling the animation generation with such a high number of parameters would be not intuitive for a human. Thus in the next step we reduce the dimensionality of the body animation control space and we map the parameters of Equation 3 onto high-level input variables. For this purpose, we use the values of matrices V and L and a mapping for input parameters: high values of the intensity variable INT correspond to strong movements (i.e., movement with high amplitude) whereas high values of the velocity variable VEL correspond to high frequencies. Next, for each column of V and L we compute its mean and standard deviation. Shoulder animation $v(t)$ is finally computed using Equation 3 where:

$$\begin{aligned} A_1 &= \overline{V}_{\_,1} + \sigma_{V_{\_,1}}(2\ INT - 1) \\ A_2 &= \overline{V}_{\_,4} + \sigma_{V_{\_,4}}(2\ INT - 1) \\ F_1 &= \overline{V}_{\_,2} + \sigma_{V_{\_,2}}(2\ VEL - 1) \\ F_2 &= \overline{V}_{\_,5} + \sigma_{V_{\_,5}}(2\ VEL - 1) \\ \Phi_2 &= \Phi_1 + C \\ C &= \frac{\sum_{i=1..N}(|V_{i,3} - V_{i,6}|))}{N} \end{aligned} \quad (4)$$

where Φ_1 is a random value in $[0,\pi]$, $\overline{V}_{\_,i}$ is the mean value of the elements of the i-th column of matrix V and $\sigma_{V_{\_,i}}$ is the standard deviation of the elements of the i-th column. For example, $\overline{V}_{\_,4}$ is the mean of the amplitudes of the higher frequencies among the selected ones. Thus, the values A_i, F_i of synthetized RBM are in the range of the values $a_{k,i}$, $f_{k,i}$ observed in the real data. If the amplitudes and frequencies in matrices V do not vary a lot then amplitudes and frequencies of the synthesized movements do not vary a lot either, independently of the values of INT and VEL. If the amplitudes and frequencies in matrices V vary a lot then the values of the variables INT and VEL provided by a human animator influence strongly the final animations (i.e., generated RBMs vary a lot depending on INT and VEL).

Torso leaning $l(t)$ is computed using a similar set of equations defined on L.

3.3.1 Virtual Character Animation

The duration $t = [t_1, t_2]$ is the third input to our model. As shoulders, during vibration, firstly move up (and then down), we find a time t_{min} that corresponds to the minimum value of $v(t)$ on the time interval $[t_1, t_1 + Period]$, where $Period$ is the actual period of $v(t)$. Next the animation $v(t)$ is computed on the interval $VI = [t_{min}, t_{min} + t_2 - t_1]$. This ensures that the shoulders movement will always start from

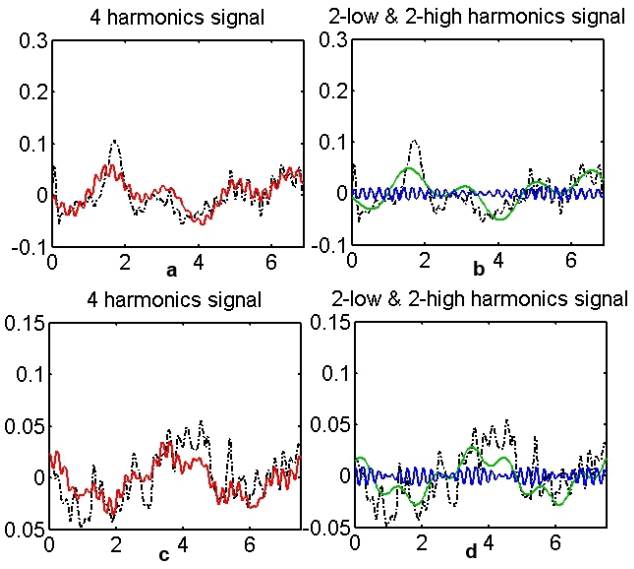

Figure 3: Reconstruction of two first signals from Figure 2. The original signal is dotted; the red curve shows signal reconstructed using 4 components (2 above and 2 below 3Hz), the green curve shows the sum of the two prevalent harmonics below 3 Hz, while the blue curve shows the sum of the two prevalent harmonics above 3 Hz.

the neutral (lowest) position.
Torso leaning can be performed in both directions (forward, backward). To ensure that torso animation starts from the neutral position, it is computed from time $t_0 > t_{min}$, where $l(t_0) < 0.001$. Next the animation $l(t)$ is computed on the interval $LI = [t_0, t_{min} + t_2 - t_1]$ while $l(t) = 0$ for $t \in [t_{min}, t_0]$. This ensures that torso movement is still synchronized with shoulders movement, but it does not need to start from the first frame of the animation.
At the end of the animation, additional frames may be added to allow shoulders and torso to come back to the neutral position. In the last step, the values of $v(t)$ and $l(t)$ computed on the intervals VI and LI are applied to joints of the VC skeleton but starting from the original time t_1.

3.3.2 *Controlling Behaviors with BML Language*

BML[4] is an XML-based language used to control behaviors of VCs. Shoulder movements are not encompassed in version 1.0 of the BML standard. In order to control such movements, we propose to extend it. In particular, to control RBMs we add the tag *shoulder* with the lexeme "shake" as well as the lexeme "leaning" to the repository of *pose* tag. Both tags have all obligatory BML's attributes such as *id*, *start*, and *end* as well as two additional attributes, *intensity* and *velocity*, taking values in $[0, 1]$ (see Table 2). For example, to generate a two seconds shoulder vibration one may use the following command:

<shoulder id="s1" start="1.0" end="3.0" lexeme="shake" side="both" intensity="1" velocity="1">.

The full description of the new tags is out of the scope of this paper.

[4]http://www.mindmakers.org/projects/bml-1-0

Attribute	Type	Required	Default
lexeme	UP|FRONT|BACK|SHAKE	Yes	-
intensity	float $\in$ [0..1]	Not	1
velocity	float $\in$ [0..1]	Not	0.5
Side	LEFT|RIGHT|BOTH	Not	BOTH

Table 2: Attributes of the shoulder tag

4. INTEGRATION: MULTIMODAL LAUGHTER SYNTHESIS

In this section we briefly introduce our laughter expressions model for torso movement and facial expression. Next, we describe how to generate the complete multimodal expression of laughter by combining rhythmic body movements, facial expressions, additional body movements and by synchronizing them with laugh sound.

4.1 Contextual Gaussian Model and PD Controller Laughter Synthesis

To model laughter animation, we rely on our previous work [8]: lip and jaw animations are modeled with a Contextual Gaussian Model (CGM) while head, eyebrow, eyelid and cheek animations are generated by concatenating segmented motions. Finally torso animations are synthesized with a proportional-derivation (PD) controller.

Animations were developed on the data of the freely available AVLC dataset [28] that contains facial motion capture data synchronized with audio. Fourteen pseudo-phonemes have been defined (see [27] for their detailed description). Phonetic transcription of audio stream with these 14 pseudo-phonemes was extracted [27]. Then we divided motion capture data of facial motions for the laugh episodes in AVLC into segments where each segment has a duration of one pseudo-phoneme. All segments were labelled by their corresponding pseudo-phoneme and clustered into 14 sub-datasets corresponding to the 14 pseudo-phonemes.

The input I to the laughter animation model consists of the phonetic transcription of the laughter to be generated. It contains the following information:

- pseudo-phonemes sequence as well as their respective duration, which can be extracted automatically from a laughter audio file (e.g., [27]),
- prosodic features of laughter. We consider two features, pitch and energy, at each frame. These features are extracted using PRAAT [3].

The animations of different parts of the body are generated with different methods: lip and jaw animations are modeled with a Contextual Gaussian Model (CGM). The output animation is described using 23 animation parameters: 22 parameters for the lip movements and 1 for the jaw motion.

A CGM is a Gaussian distribution whose mean vector depends on a set of contextual variable(s). In our case, contextual variables are the values of pitch and energy at each frame. When such CGM with a parameterized mean vector is used to model lip and jaw motions, the mean of the CGM obeys the following equation:

$$\hat{\mu}(\theta) = W^{\mu}\theta + \bar{\mu} \quad (5)$$

where W^{μ} is a 23×2 matrix and $\bar{\mu}$ is an offset vector. The mean μ is a 23-dimension vector containing values of all 23

animation parameters per frame. θ stands for a 2-dimension vector containing values of the pitch and energy (contextual variables) at a given frame. Each of the 14 pseudo-phonemes has its own CGM model. Each sub-dataset is used to build the CGM model, which is learned through Maximum Likelihood Estimation (MLE). So, each laughter pseudo-phoneme of the input I is used to select one from 14 CGMs; the values of pitch and energy are applied to the selected CGM to determine lip shape and jaw motion.

Head, eyebrow, eyelid, and cheek animations are determined by selecting and concatenating existing (captured) motion segments, which are scaled to the pseudo-phonemes duration specified in the input I. The synthetized animation is obtained by concatenating the segments that correspond to the input pseudo-phonemes sequence. Each sub-dataset for a given pseudo-phoneme contains many examples of motion segment. The choice of the motion segment from a sub-dataset depends on the sum of two criteria: *duration cost* and *continuity cost*. Duration cost is computed as the difference between expected (target) duration and the candidate segment duration; continuity cost is the distance between the end position of the previously selected segment and the beginning position of the candidate segment. The candidate with the least sum of the two costs is selected as output. Finally, the outputted motion is obtained by concatenating and interpolating the selected samples between two successive segments.

To overcome the unnatural effect of a moving head attached to a rigid body, the third part of our model generates additional torso movements. Thus we compute the torso movement from the head movement. The torso animations are generated by PD controllers. The input to the PD controllers are the three rotation angles of the head. Three PD controllers are independently built to control the three torso rotations (pitch, yaw, and roll). The output of the PD controllers depends on the input at the current frame and the output at the previous frame. Consequently, the torso rotations follow the head rotation, and the head rotations depends on the currently considered pseudo-phoneme. For example, when the head moves down (pitch rotation) then the torso leans forward (pitch rotation). In the PD controllers, the parameters, which define the relation between the head and the torso rotations, are manually defined.

4.2 Synchronization between Modalities

The model presented above is controlled by the phonetic transcription I (list of phonemes and their durations). The same input I is used to synchronize all the modalities: 1) to generate appropriate facial expressions with CGMs and torso movements by concatenation of motion segments, 2) to control (indirectly) the additional body movements (computed from head movement done at item 1), 3) to compute the rhythm of laughter and to control the model of rhythmic body movements, 4) to synchronize with real or synthesized sounds of laughter.

The synchronization procedure is presented in Figure 4. The animation can be synchronized with real or synthesized sounds of laughter. When using natural (real) laughter audio, we need to extract automatically the sequences of pseudo-phonemes. We use Urbain et al.'s laughter segmentation tool [27]. The topic of generating laughter audio synthesis is beyond the scope of this paper. We only mention that there exist several techniques for that, e.g., [26]. The

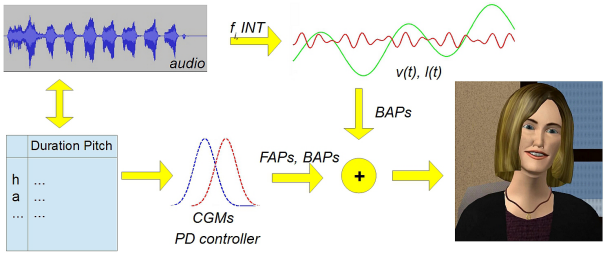

Figure 4: Synchronization procedure

animations generated with the model presented in Section 4.1 take as input the pseudo-phoneme lists and compute the animation based on the duration of each pseudo-phoneme.

Rhythmic body movements can also be synchronized with the whole animation. In the case of multiple modalities to be synchronized the values INT and VEL (see Section 3.3) are not controlled anymore by a human but they are computed from audio, and then applied to Equation 4. Both audio bursts and shoulder movements are caused by repetitive forced exhalations and display similar ranges of frequencies (see Section 1). Thus, we assume that to generate synchronized audio-visual laughter it is sufficient to align the main frequency of the shoulder movements F_1 with the mean frequency of the laughter bursts in the audio sample; this latter can be obtained from the phonetic transcription: $F_1 = \overline{|\, pho \,|}$. Using it and reversed Equation 4 one can compute the value of VEL and thus F_2. The laughter intensity of the audio sample can be estimated in real-time, we use a method proposed in [14]. Once we have the value of INT, we compute the remaining parameters of Equation 4, namely A_1, and A_2. Finally we know the value of $v(t)$.

Figure 5 shows an example of the synchronized multimodal animation. We used a freely available MPEG-4 compatible virtual character called Greta[5][13].

5. INTERACTION SCENARIO: REAL-TIME ADAPTATION OF LAUGHTER RHYTHM

The approach presented in Section 3 has one important advantage: it can be applied during real-time H-VC interaction. In this section we propose how, using our models, a VC tunes the rhythm of its laughter to the nonverbal behavior of a human interacting with it. The ability to perform such an adaptation could contribute to, for example, increase entitativity or affiliation between the human and the VC [10]. We describe here the concept of such real-time interaction, leaving the system implementation and evaluation for future work.

Several interactive H-VC systems considering laughter for the VC side have been proposed, as discussed in Section 2.3. While in these works the VC is able to respond to the user by producing laughs, none of them takes into account human's laughter body movements. Our concept of interaction described below could be integrated within any of these systems.

At least two solutions can be proposed to map human nonverbal behavior onto the animation of a VC: (i) a low-level mapping in which the RBM parameters of the human

[5]http://perso.telecom-paristech.fr/~pelachau/Greta/

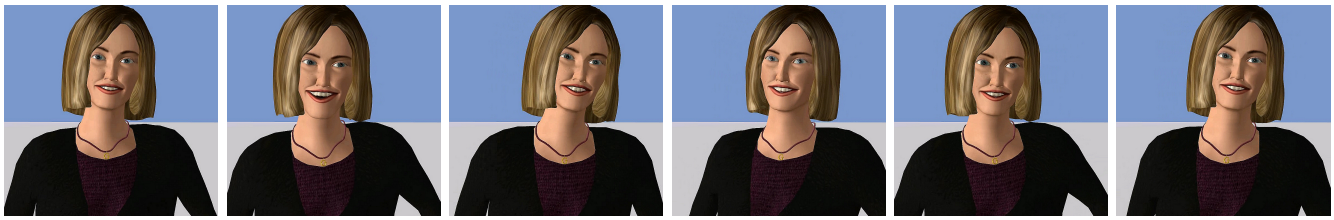

Figure 5: Example of animations created with the model

behavior are directly copied to VC's RBM, (ii) a high-level mapping in which more generic characteristics of human expressivity during laughter are captured and mapped onto the VC's RBM parameters. In this interaction scenario we focus on the second solution. Instead of directly copying the harmonics of rhythmic laughter movements, we measure higher-level expressive qualities of human nonverbal behavior during laughter and we propose a mapping of these values onto the laughter RBM harmonics from Equation 4. The idea of mapping human's expressive features to VC's expressive features without copying the movement itself is not novel [4, 20], however, to our knowledge, no previous work focused on body movements of laughter.

The EyesWeb XMI platform is used to analyze the expressive qualities of nonverbal behaviors in laughter. It is equipped with several algorithms to analyze expressive qualities of movement, such as quantity of motion or smoothness. Human data is captured and processed in real-time using EyesWeb XMI and the human's silhouette is automatically extracted from the depth map captured by a Microsoft Kinect sensor. From the silhouette, three expressive features are computed - Quantity of Motion, Contraction Index, and Smoothness Index - which have been widely discussed in literature, see for example [16]. Such features are then mapped onto the VC's RBM parameters $\{(a_k, f_k, \phi_k)|k \in \{1, 2\}\}$ and $\{(A_k, F_k, \Phi_k)|k \in \{1, 2\}\}$.

In particular, we propose the following mapping:

- Quantity of Motion, that is, the detected amount of human's body movement, is mapped onto the VEL variable (which determines the frequency of shoulder and torso pulses, see Equation 4);
- Contraction Index, that is, the detected amount of horizontal and vertical human's body contraction, is mapped onto the INT variable (which determines the amplitude of shoulder and torso movement, see Eq. 4);
- Smoothness Index, that is, the amount of continuity in human's body movement, is mapped onto the $C = \Phi_2$ - Φ_1 (that determines the distance between the phases of the 2 harmonics of the generated movement, see Equation 4).

Consequently: (i) the more the human produces body movements during laughter, the more frequent are the VC's RBMs movements, (ii) the larger are the human's movements during laughter, the stronger are the RBM pulses; (iii) the smoother are the movements, the smoother are the trajectories of the VC's RBM movements.

The problem of the detection of human laughter is out of the scope of this paper. There exist several algorithms working in real-time for that, e.g., [12, 14].

6. CONCLUSION AND FUTURE WORK

In this paper we studied and modeled the rhythmic body movements of laughter. The spectral analysis of expressive body movements confirms the findings in literature about the rhythmic movements of laughter. We also showed that only 4 harmonics can effectively reconstruct RBMs. Our RBM model combines the ease of controlling an animation obtained by procedural approach with the accuracy and naturalness of animation based on experimental data. Our model is based on real data analysis but it allows a human to control the animation via a small set of parameters. Importantly, our approach is computationally light and it can be applied to any virtual character. It enables to map human movements to virtual characters in real-time. We believe that it could be easily extended to other rhythmic movements performed while, e.g., coughing or crying. In this paper we also showed how rhythmic animations can be synchronized with the other synthetized expressive cues of laughter such as facial expressions and other body movements. Finally we have also discussed an interactive application integrating our RBM model.

Several limitations and future works should be considered. Our current RBM model is based on a relatively small dataset. Anatomical differences as well as stances (e.g., sitting) may strongly influence movements in laughter. We plan to analyze more data of different gender and stances. Regarding the synthesis part, we are going to extend our model by considering other movement types, such as arm throwing, and knees bending. To validate our model we will perform a perceptive study. The study will provide feedback on the pertinence of our animation model of multimodal laughter. Participants will be asked to evaluate different stimuli of the virtual agent laughing. Three conditions will be compared; they will all use the same audio episode synchronized with facial animation, only the RBM movements will vary between conditions as follow: 1) animations with RBM movements generated with our model, which are synchronized with audio the episode, 2) RBM movements of frequency and amplitude beyond the intervals specified in Equation 4 of our model, 3) animations without RBM movements (the facial animation is still synchronized with audio). The believability, realism and naturalness of the animations will be compared.

7. ACKNOWLEDGEMENTS

The research leading to these results has received funding from the EU 7th Framework Programme under grant agreement n 270780 ILHAIRE. The authors would like to thank Jérôme Urbain (University of Mons) for providing the phonetic transcription of the audio samples used for synthesis.

8. REFERENCES

[1] J. N. Bailenson and N. Yee. Digital chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological Science*, 16(10):814–819, 2005.

[2] S. Block, M. Lemeignan, and N. Aguilera. Specific respiratory patterns distinguish among human basic emotions. *International Journal of Psychophysiology*, 11(2):141–154, 1991.

[3] P. Boersma and D. Weeninck. PRAAT, a system for doing phonetics by computer. *Glot International*, 5(9/10):341–345, 2001.

[4] G. Castellano, M. Mancini, C. Peters, and P. McOwan. Expressive copying behavior for social agents: A perceptual analysis. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 42(3):776–783, May 2012.

[5] D. Cosker and J. Edge. Laughing, crying, sneezing and yawning: Automatic voice driven animation of non-speech articulations. In *Proc. of Computer Animation and Social Agents*, pages 21–24, 2009.

[6] C. Darwin. *The expression of emotion in man and animal*. John Murray, London, 1872.

[7] P. C. DiLorenzo, V. B. Zordan, and B. L. Sanders. Laughing out loud: control for modeling anatomically inspired laughter using audio. *ACM Transactions on Graphics (TOG)*, 27(5):125, 2008.

[8] Y. Ding, K. Prepin, J. Huang, C. Pelachaud, and T. Artières. Laughter animation synthesis. In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems*, AAMAS '14, pages 773–780, 2014.

[9] S. Fukushima, Y. Hashimoto, T. Nozawa, and H. Kajimoto. Laugh enhancer using laugh track synchronized with the user's laugh motion. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '10, pages 3613–3618, New York, NY, USA, 2010. ACM.

[10] D. Lakens and M. Stel. If they move in sync, they must feel in sync: Movement synchrony leads to attributions of rapport and entitativity. *Social Cognition*, 29(1):1–14, 2011.

[11] M. Mancini, J. Hofmann, T. Platt, G. Volpe, G. Varni, D. Glowinski, W. Ruch, and A. Camurri. Towards automated full body detection of laughter driven by human expert annotation. In *Proceedings of Affective Interaction in Natural Environments (AFFINE) Workshop*, pages 757–762, Geneva, Switzerland, 2013.

[12] M. Mancini, G. Varni, D. Glowinski, and G. Volpe. Computing and evaluating the body laughter index. *Human Behavior Understanding*, pages 90–98, 2012.

[13] R. Niewiadomski, E. Bevacqua, Q. A. Le, M. Obaid, J. Looser, and C. Pelachaud. Cross-media agent platform. In *Web3D ACM Conference*, pages 11–19, Paris, France, 2011.

[14] R. Niewiadomski, J. Hofmann, J. Urbain, T. Platt, J. Wagner, B. Piot, H. Cakmak, S. Pammi, T. Baur, S. Dupont, M. Geist, F. Lingenfelser, G. McKeown, O. Pietquin, and W. Ruch. Laugh-aware virtual agent and its impact on user amusement. AAMAS '13, pages 619–626, 2013.

[15] R. Niewiadomski, M. Mancini, T. Baur, G. Varni, H. Griffin, and M. Aung. MMLI: Multimodal multiperson corpus of laughter in interaction. In A. Salah, H. Hung, O. Aran, and H. Gunes, editors, *Human Behavior Understanding*, volume 8212 of *LNCS*, pages 184–195. Springer, 2013.

[16] R. Niewiadomski, M. Mancini, and S. Piana. Human and virtual agent expressive gesture quality analysis and synthesis. In M. Rojc and N. Campbell, editors, *Coverbal Synchrony in Human-Machine Interaction*, pages 269–292. CRC Press, 2013.

[17] R. Niewiadomski and C. Pelachaud. Towards multimodal expression of laughter. In *Proceedings of IVA'12*, pages 231–244. Springer-Verlag Berlin, Heidelberg, 2012.

[18] K. Prepin and C. Pelachaud. Effect of time delays on agents' interaction dynamics. In *The 10th International Conference on Autonomous Agents and Multiagent Systems - Volume 3*, AAMAS '11, pages 1055–1062, 2011.

[19] R. R. Provine and Y. L. Yong. Laughter: A stereotyped human vocalization. *Ethology*, 89(2):115–124, 1991.

[20] R. Pugliese and K. Lehtonen. A framework for motion based bodily enaction with virtual characters. In *Intelligent Virtual Agents*, volume 6895 of *LNCS*, pages 162–168. Springer Berlin Heidelberg, 2011.

[21] L. D. Riek, P. C. Paul, and P. Robinson. When my robot smiles at me: Enabling human-robot rapport via real-time head gesture mimicry. *Journal on Multimodal User Interfaces*, 3(1-2):99–108, 2010.

[22] W. Ruch and P. Ekman. The expressive pattern of laughter. In A. Kaszniak, editor, *Emotion, qualia and consciousness*, pages 426–443. World Scientific Publishers, Tokyo, 2001.

[23] J. Tilmanne and T. Dutoit. Stylistic walk synthesis based on Fourier decomposition. In *Proceedings of the INTETAIN*, Mons, Belgium, 2013.

[24] N. Troje. The little difference: Fourier based synthesis of gender-specific biological motion. In R. Würtz and M. Lappe, editors, *Dynamic perception*, pages 115–120. Berlin: AKA Verlag, 2002.

[25] M. Unuma, K. Anjyo, and R. Takeuchi. Fourier principles for emotion-based human figure animation. In *Proceedings of SIGGRAPH 1995*, pages 91–96, 1995.

[26] J. Urbain, H. Cakmak, A. Charlier, M. Denti, T. Dutoit, and S. Dupont. Arousal-driven synthesis of laughter. *Selected Topics in Signal Processing, IEEE Journal of*, 8(2):273–284, April 2014.

[27] J. Urbain, H. Cakmak, and T. Dutoit. Automatic phonetic transcription of laughter and its application to laughter synthesis. In *Proceedings of ACII '13*, pages 153–158, 2013.

[28] J. Urbain, R. Niewiadomski, E. Bevacqua, T. Dutoit, A. Moinet, C. Pelachaud, B. Picart, J. Tilmanne, and J. Wagner. AVLaughterCycle: Enabling a virtual agent to join in laughing with a conversational partner using a similarity-driven audiovisual laughter animation. *Journal on Multimodal User Interfaces*, 4(1):47–58, 2010. Special Issue: eNTERFACE'09.