2018 IEEE International Conference on Robotics and Automation (ICRA)
May 21-25, 2018, Brisbane, Australia

Using Constrained Optimization for Real-Time Synchronization of Verbal and Nonverbal Robot Behavior

Aravind Elanjimattathil Vijayan[1], Simon Alexanderson[2], Jonas Beskow[3] and Iolanda Leite[4]
{araev[1], simonal[2], beskow[3], iolanda[4]}@kth.se

***Abstract*— Most of the motion re-targeting techniques are grounded on virtual character animation research, which means that they typically assume that the target embodiment has unconstrained joint angular velocities. However, because robots often do have such constraints, traditional re-targeting approaches can originate irregular delays in the robot motion. With the goal of ensuring synchronization between verbal and nonverbal behavior, this paper proposes an optimization framework for processing re-targeted motion sequences that addresses constraints such as joint angle and angular velocities. The proposed framework was evaluated on a humanoid robot using both objective and subjective metrics. While the analysis of the joint motion trajectories provides evidence that our framework successfully performs the desired modifications to ensure verbal and nonverbal behavior synchronization, results from a perceptual study showed that participants found the robot motion generated by our method more natural, elegant and lifelike than a control condition.**

I. INTRODUCTION

Motion re-targeting has been widely used in computer animation for mapping human motion capture data to virtual characters [1]. In the recent years, this technique also became popular in robotics not only for its potential for generating expressive robot behavior [2], [3], [4] but also for teleoperation applications [5], [6].

Most of the existing motion re-targeting methods rely on techniques for mapping human motion into embodiments with a different number of Degrees of Freedom (DoF) [1], [7]. Moreover, because these methods were mainly developed for virtual characters, they typically assume that the target embodiment has unconstrained joint angle limits and angular velocities. However, such an assumption does not always hold true for robots because quite often a compromise between cost and the quality of the actuators (and consequently speed) needs to be made.

Differences between the maximum attainable velocity of the robot's joints and the velocity of the re-targeted sequence may result in motion delays. These delays can be critical especially when speech is conveyed together with the motion. In these cases, simply stretching the re-targeted motion can cause synchronization problems between the robot's verbal and nonverbal behavior. It must be noted that any violation of individual joint constraints would require the whole-body motion sequence (i.e., all the remaining joints) to be stretched, under the penalty of propagating asynchrony within joints and consequently altering the meaning of the expression. These issues can bring serious negative implications, as motion timing [8], [9], [10] and legibility [11] have been shown to play a major role in behavioral perception.

In this work, we propose a solution for ensuring verbal and nonverbal synchronization of robot behavior from re-targeted motion capture data. Our approach is formulated as a constrained optimization problem over sliding windows of joint motion sequences that take into account the robot's dynamic constraints while maintaining behavioral expressivity.

To validate the proposed approach, a dataset of human behavioral expressions including both speech and gestures was captured and re-targeted to an animated character with the same DoF as NAO robot, and then sent to our optimization framework to solve the verbal and nonverbal synchronization issues before being conveyed by the robot. The optimized sequences resultant from our framework were compared to a baseline condition where such constraints were not accounted for. Objective experiments comparing motion trajectories and a subjective user perceptual study showed that our approach was able to keep verbal and nonverbal synchrony while ensuring minimal or no loss in perceived expressivity.

In Section II, previous work on motion re-targeting is presented. Our motion re-targeting pipeline, along with the optimization framework for addressing robot velocity constraints, is discussed in Section III. A case study is provided in Section IV, followed by an experimental evaluation of its performance in Section V.

II. RELATED WORK

Research on motion re-targeting methods have received a lot of attention in computer graphics [1],[12],[13]. Choi et al. [14], for example, introduced an online re-targeting method that used inverse rate control to compute changes in the joint angles corresponding to changes in end-effector positions of virtual characters. Baran et al. [7] proposed another method to perform automatic rigging and modeling surface deformations. In another recent example, Wang and colleagues [15] aimed at creating natural motion for animated human-like characters in varying physically simulated environments. To compensate for latency errors, a predictive component is combined with a Model Predictive Controller (MPC) to perform the re-targeting to animated characters. In our work, we use a similar linearized model for rigid body dynamics, but we also take into consideration limitations in joint velocity (more relevant for physical robots than to

*All authors are with the School of Electrical Engineering and Computer Science at KTH Royal Institute of Technology Stockholm, Sweden. This work was supported by KTH's visionary project SONAO (Robust nonverbal expression in artificial agents: identification and modelling of stylized gesture and sound cues).

virtual characters) which was not accounted for in previous research.

While several authors have proposed motion re-targeting methods from human motion to robots, these works were mainly focused on aspects other than expressivity such as effective goal-based teleoperation [6], imitation [16] or ensuring robot stability [17], [18], [2]. One of such exceptions has employed re-targeting for generating robot motion for artistic performances [3]. In this work, the motion trajectory was re-optimized to mimic human dance performance while ensuring that the Zero Moment Point (ZMP) lies within a supporting area. However, unlike our proposed approach, the motion constraints were handled by saturating motion when limits were violated.

In a similar work, motion data of actors performing a children song was captured and re-targeted to a humanoid robot [19]. A non-uniform local scaling method was proposed to modify motion sequences violating the robot joint limits. In this work, velocity was constrained by saturating it upon reaching its limit. The computed trajectory was then used as input to a trajectory tracking control system to execute the motion on the robot. Unlike this work, our method does not require any feedback control loop in the robot performing the action.

Agarwal et al. [5] employed motion re-targeting for real time teleoperation of a Furhat robotic head using human head motion. The authors estimated head pose using a depth sensor and filtered noisy measurements before mapping the head pose using a neural network model. The robot head motion is then smoothed and constrained using a constrained-state Kalman Filter. Despite positive perceptual results, the authors acknowledge limitations in their approach for mapping high frequency motions such as head nods.

Assuming static postures, a unified framework for human pose estimation and motion re-targeting is presented in [20]. Human posture is captured by a depth camera and joint parameters are globally fit into a human robot model which was parametrically modelled. Such approach based on fitting joint parameters globally for re-targeting is effective for static postures but might not be suitable for dynamic sequential behaviors, as previous research has shown that legibility is extremely important in Human-Robot Interaction [11].

Perhaps most similar to our work, a joint optimization framework for optimizing joint trajectories was proposed in [21]. However, because the states included human joint angles in addition to robot joint angles, the global optimization problem was very complex to solve. Instead of posing a global optimization problem, we break it down into a per-joint optimization problem where the goal is to optimize a single joint angle profile over a pre-defined time interval because simpler optimization problems can be solved in real time [22].

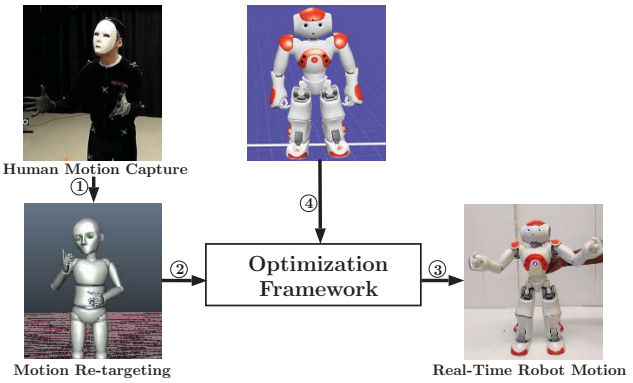

Fig. 1. Proposed pipeline for online optimization of human motion capture data. 1) MoCap sequence; 2) Unconstrained motion sequence; 3) Constrained optimized sequence; 4) Robot dynamic constraints.

III. APPROACH

In this work, we propose a motion re-targeting framework from human motion capture data to robot motion which addresses the following fundamental challenges:

- Ensures verbal and nonverbal behavior synchronization by planning the motion execution considering the joint angle and velocity constraints of the target robot;
- Given the robot constraints, re-targets motion while attempting to minimize loss of behavioral expressivity.

An overview of our pipeline is displayed in Fig. 1. A motion capture (MoCap) system records human motion that is then re-targeted to a virtual model with the same DoF as the target robot. Until here, this is a typical re-targeting workflow for animated characters because attainable joint angles and angular velocities are not a concern in that case.

Our main contribution lies in the optimization framework module that takes into account the robot's velocity constraints and, using a minimal set of weight parameters, computes a modification to the unconstrained motion sequence to be executed by the robot such that the verbal and nonverbal behavior remains synchronized. The remaining of this section describes the steps for this motion optimization in more detail.

A. Constrained Linear Quadratic Optimization

Consider a motion sequence consisting of human joint trajectories recorded by a motion capture system over the time interval, $t \in [t_0, t_f]$. Let the sampled capture be the re-targeted motion to an animated character which consists of $\boldsymbol{k} \in \{1, 2, \ldots, N\}$ samples, each denoted by a state vector $\boldsymbol{\zeta_k}$, where $\boldsymbol{\zeta_k} \in \mathbb{R}^P$. Here, $\boldsymbol{\zeta_k} = [{}^1\boldsymbol{x}_k \ {}^2\boldsymbol{x}_k \ \ldots \ {}^P\boldsymbol{x}_k]^T$ denotes the vector consisting of joint angles of all the $\mathbf{P}$ degrees of freedom of the robot at sampling instant $\boldsymbol{k}$, where ${}^i\boldsymbol{x}_k \in \mathbb{R}$. If ${}^i\boldsymbol{x}_k$ denotes the rotation angle of the i^{th} joint of the robot and ${}^i\boldsymbol{u_k} = {}^i\dot{\boldsymbol{x}}_{\boldsymbol{k}}$ denotes the corresponding angular velocity, the joint angle at any given sampling time k_f, given the angular velocities from $\boldsymbol{k} = \boldsymbol{k_i}, \ldots, \boldsymbol{k}_{f-1}$, can be approximated as

$$ {}^i\boldsymbol{x}_{k_f} = {}^i\boldsymbol{x}_{k_i} + \sum_{\boldsymbol{k}=\boldsymbol{k_i}}^{\boldsymbol{k}_{f-1}} {}^i\dot{\boldsymbol{x}}_{\boldsymbol{k}} h, \tag{1} $$

where h denotes the motion sampling period.

The output of the initial motion re-targeting may not satisfy the various constraints of the robot. Consider a joint motion trajectory of the i^{th} robot joint starting from ${}^{i}\boldsymbol{x_0}$ for a short interval of time consisting of N samples ${}^{i}\boldsymbol{\chi} = [{}^{i}\boldsymbol{x_1}\ {}^{i}\boldsymbol{x_2}\dots\ {}^{i}\boldsymbol{x_N}]^T$, where ${}^{i}\boldsymbol{\chi} \in \mathbb{R}^N$. For the motion sequence, let the angular velocities be defined by the vector ${}^{i}\boldsymbol{\upsilon} = [{}^{i}\boldsymbol{u_1}\ {}^{i}\boldsymbol{u_2}\dots\ {}^{i}\boldsymbol{u_N}]$, ${}^{i}\boldsymbol{\upsilon} \in \mathbb{R}^N$, where ${}^{i}\boldsymbol{u_k}$ follows a first order derivative approximation and is defined as ${}^{i}\boldsymbol{u_k} = \frac{{}^{i}\boldsymbol{x_k} - {}^{i}\boldsymbol{x_{k-1}}}{h}$. Further, we define the state vector for optimization as ${}^{i}\boldsymbol{\xi} = [{}^{i}\boldsymbol{\chi}^T\ {}^{i}\boldsymbol{\upsilon}^T]^T$ and denote the original re-targeted motion trajectory for the joint by ${}^{i}\boldsymbol{\xi}_{rt} = [{}^{i}\boldsymbol{\chi}_{rt}^T\ {}^{i}\boldsymbol{\upsilon}_{rt}^T]^T$. Given a sequence of ${}^{i}\boldsymbol{\xi}$, we define the robot joint rotational dynamics as equality constraints which will be used in the optimal control problem formulation as

$$\begin{aligned} {}^{i}\boldsymbol{x_1} - {}^{i}\boldsymbol{u_1}.h &= {}^{i}\boldsymbol{x_0} \\ {}^{i}\boldsymbol{x_k} - {}^{i}\boldsymbol{x}_{k-1} - {}^{i}\boldsymbol{u_k}.h &= 0 \qquad \forall\ \boldsymbol{k} \in \{2,\dots,\mathrm{N}\}, \end{aligned} \tag{2}$$

which can be expressed as a matrix equality of the form

$$\boldsymbol{A}_{eq}\,{}^{i}\boldsymbol{\xi} = \boldsymbol{b}_{eq}. \tag{3}$$

In this work, we are mainly interested in the joint angle and angular velocity limits of robots, although this formulation could be extended to include stability constraints and other global motion requirements. Since the feasible solution for the joint trajectory is a subspace of all the possible trajectories, the constraint manifold needs to be formulated into a set of inequality expressions representing the various constraints on the state vector for the joint being optimized.

Assuming that the i^{th} joint angle at any given time is required to lie within its joint angle limits ${}^{i}\underline{x}$ and ${}^{i}\overline{x}$ respectively, where ${}^{i}\underline{x} \le {}^{i}\overline{x}$, we have ${}^{i}\underline{x} \le {}^{i}\boldsymbol{x_k} \le {}^{i}\overline{x}\ \forall\ \boldsymbol{k} \in \{1,\dots,\mathrm{N}\}$. Denoting the identity matrix of size N using $\boldsymbol{I_N}$ and a row vector of ones using $\boldsymbol{1_N}$, the inequality can be expressed in terms of ${}^{i}\boldsymbol{\chi}$ as

$$[-{}^{i}\boldsymbol{\chi}^T\ {}^{i}\boldsymbol{\chi}^T]^T \le [-{}^{i}\underline{x}.\boldsymbol{1_N}\ {}^{i}\overline{x}.\boldsymbol{1_N}]^T \tag{4}$$

Similarly, constraints on ${}^{i}\boldsymbol{\upsilon}$ can be formulated as

$$[-{}^{i}\boldsymbol{\upsilon}^T\ {}^{i}\boldsymbol{\upsilon}^T]^T \le [-{}^{i}\underline{u}.\boldsymbol{1_N}\ {}^{i}\overline{u}.\boldsymbol{1_N}]^T \tag{5}$$

where, ${}^{i}\underline{u}$ and ${}^{i}\overline{u}$ are, respectively, the minimum and maximum allowable angular velocities for the i^{th} joint. Further, an inequality expression over the states of the system in the form $\boldsymbol{A}_{neq}\,{}^{i}\boldsymbol{\xi} \le \boldsymbol{b}_{neq}$ can be obtained from (4) and (5).

The cost function for this problem was formulated in a quadratic form such that it can be solved as a Linear Quadratic (LQ) control problem. Representing the error between the re-targeted and the optimized joint trajectory using ${}^{i}\boldsymbol{\chi}_{\Delta} = {}^{i}\boldsymbol{\chi} - {}^{i}\boldsymbol{\chi}_{rt}$ and corresponding error in joint angular velocity using ${}^{i}\boldsymbol{\upsilon}_{\Delta} = {}^{i}\boldsymbol{\upsilon} - {}^{i}\boldsymbol{\upsilon}_{rt}$, the cost function $\boldsymbol{J}({}^{i}\boldsymbol{\xi}_{\Delta}) : \mathbb{R}^{2\mathrm{N}} \to \mathbb{R}$ for the optimization problem is formulated as

$$\boldsymbol{J}({}^{i}\boldsymbol{\xi}_{\Delta}) = {}^{i}\boldsymbol{\chi}_{\Delta}^T Q_1\,{}^{i}\boldsymbol{\chi}_{\Delta} + {}^{i}\boldsymbol{\upsilon}_{\Delta}^T Q_2\,{}^{i}\boldsymbol{\upsilon}_{\Delta} \ :\ {}^{i}\boldsymbol{\xi}_{\Delta} = [{}^{i}\boldsymbol{\chi}_{\Delta}^T\ {}^{i}\boldsymbol{\upsilon}_{\Delta}^T]^T, \tag{6}$$

where Q_1 and Q_2 are positive definite weight matrices. Combining (2)-(6), the optimal control problem can be defined as

$$\min_{{}^{i}\boldsymbol{\xi}} \left\{ {}^{i}\boldsymbol{\chi}_{\Delta}^T Q_1\,{}^{i}\boldsymbol{\chi}_{\Delta} + {}^{i}\boldsymbol{\upsilon}_{\Delta}^T Q_2\,{}^{i}\boldsymbol{\upsilon}_{\Delta} \right\}$$
$$\text{subject to} \begin{cases} \boldsymbol{A}_{eq}\,{}^{i}\boldsymbol{\xi} = \boldsymbol{b}_{eq} \\ \boldsymbol{A}_{neq}\,{}^{i}\boldsymbol{\xi} \le \boldsymbol{b}_{neq} \\ {}^{i}\boldsymbol{x_0} = \text{given} \\ {}^{i}\boldsymbol{\xi} \in \mathbb{R}^{2\mathrm{N}}, \boldsymbol{J}({}^{i}\boldsymbol{\xi}) \in \mathbb{R} \end{cases} \tag{7}$$

B. Extension for Piecewise Behavior Compression

The main limitation of the above formulation is that when the behavior includes high amplitude motions (e.g., when the human waves or swings hands very fast), attempting to track the re-targeted motion trajectory as accurately as possible may deteriorate the naturalness of the behavior. In order to ensure smoothness of behavior even in such cases, high amplitude motion trajectories need to be replaced by lower amplitude ones. To provide a solution for such scenarios, we propose an nonlinear optimization problem in addition to (7) which enables scaling modifications to the original trajectory that will result in more natural motion. This extension to the initial problem formulation will ensure motions with high amplitude to be executed in a smoother manner, without loosing continuity with respect to the overall motion sequence.

To perform the picewise behavior compression, we divide each of the short horizon motion trajectories further and assign a scaling parameter ${}^{i}\boldsymbol{a_r}$ to each of these intervals. Let the trajectory be divided into Q sub-sequences where ${}^{i}\boldsymbol{a_r} \in \mathbb{R}_s, \boldsymbol{r} \in \{1,\dots,\mathrm{Q}\}$ denotes a scaling parameter assigned to each of the sub-sequence. The scaling parameter defines the factor by which the amplitude of the motion needs to be reduced. The vector ${}^{i}\boldsymbol{\alpha} = [{}^{i}\boldsymbol{a_1},\dots,{}^{i}\boldsymbol{a_Q}]^T$ consisting of the scaling parameters is appended to the state vector ${}^{i}\boldsymbol{\xi}$, forming the new state vector ${}^{i}\boldsymbol{\xi}_{new} = [{}^{i}\boldsymbol{\chi}^T\ {}^{i}\boldsymbol{\upsilon}^T\ {}^{i}\boldsymbol{\alpha}^T]^T$. Further, we modify the equality constraints in (2) as follows:

$$\begin{aligned} {}^{i}\boldsymbol{x_1} &= \boldsymbol{x}_{ref} \\ {}^{i}\boldsymbol{x_k} - {}^{i}\boldsymbol{x}_{k-1} - \frac{{}^{i}\boldsymbol{u_k}.h}{{}^{i}\boldsymbol{a_r}} &= 0 \qquad \forall\ \boldsymbol{k} \in \{2,\dots,\mathrm{N}\} \end{aligned} \tag{8}$$

In (8), dividing the control signal ${}^{i}\boldsymbol{u_k}$ by the scaling parameter magnifies its effect, which due to the dynamic constraint forces the control signal to be scaled down. By performing this computation instead of stretching the duration of the sequence, an alternate scaled down version of the motion trajectory is calculated. However, it should be noted that this trajectory modification does not affect the synchrony between verbal and nonverbal actions. To ensure that the scaling action does not damp unnecessary actions, inequality expressions ensuring that the scaling parameter lies within its domain $\mathbb{R}_s$ are added to the problem statement.

Finally, the cost function is modified such that the scaling factor tends to remain close to 1. In order to compute the deviation of the modified trajectory compared to the

Algorithm 1: Motion trajectory synchronization

Input: Non-optimized re-targeted motion sequence $\boldsymbol{\xi}$
Output: Optimized re-targeted motion sequence obeying the robot's velocity constraints $\boldsymbol{\xi}_{opt}$
let $P \leftarrow$ number of joints
define Q_1, Q_2, Q_3 //weight matrices
$\{\boldsymbol{A}_{eq}, \boldsymbol{A}'_{eq}, \boldsymbol{b}_{eq}, \boldsymbol{b}'_{eq}\} \leftarrow$ getRobotDynamics()
$\{\boldsymbol{A}_{neq}, \boldsymbol{A}'_{neq}, \boldsymbol{b}_{neq}, \boldsymbol{b}'_{neq}\} \leftarrow$ getRobotConstraints()
function optimizeSequence($\boldsymbol{\xi}$)
  for $i : i \in \{1, \ldots, P\}$ **do**
    Extract ${}^i\boldsymbol{\xi}$ where ${}^i\boldsymbol{\xi} \subset \boldsymbol{\xi}$ //per joint sequence
    if *jointConstraintViolated*(${}^i\boldsymbol{\xi}$) **then**
      if *notLargeAmplitudeMotion*(${}^i\boldsymbol{\xi}$) **then**
        //Perform constrained optimization
        // in eq. (7)
        ${}^i\boldsymbol{\xi}_{opt} \leftarrow$ constrainedOptimization(${}^i\boldsymbol{\xi}, Q_1, Q_2, \boldsymbol{A}_{eq}, \boldsymbol{b}_{eq}, \boldsymbol{A}_{neq}, \boldsymbol{b}_{neq}$)
      else
        //Scaling factor for sub-sequences
        //generated from ${}^i\boldsymbol{\xi}$
        let ${}^i\boldsymbol{\alpha} = [{}^i\boldsymbol{a_1}, \ldots, {}^i\boldsymbol{a}_{\mathrm{Q}}]$
        define ${}^i\boldsymbol{a_r} \in \mathbb{R}_{\mathrm{S}} \forall r \in \{1, \ldots, \mathrm{Q}\}$
        //Perform piecewise optimization
        // in eq. (10)
        ${}^i\boldsymbol{\xi}_{opt} \leftarrow$piecewiseOptimization(${}^i\boldsymbol{\xi}, Q_1, Q_2, Q_3, \boldsymbol{A}'_{eq}, \boldsymbol{b}'_{eq}, \boldsymbol{A}'_{neq}, \boldsymbol{b}'_{neq}$)
  $\boldsymbol{\xi}_{opt} \leftarrow \{{}^1\boldsymbol{\xi}_{opt}, \ldots, {}^P\boldsymbol{\xi}_{opt}\}$
  return $\boldsymbol{\xi}_{opt}$

reference trajectory, a scaled version of reference velocity sequence is computed using the scaling parameters.

Representing the error in trajectory as ${}^i\boldsymbol{\chi}_\Delta$, the error in angular velocity vector as ${}^i\boldsymbol{\upsilon}_\Delta$ and the deviation of scaling parameter as ${}^i\boldsymbol{\alpha}_\Delta$, the modified cost function is given by

$$\begin{aligned} J'({}^i\boldsymbol{\xi}_{new}) &= {}^i\boldsymbol{\chi}_\Delta^T Q_1 {}^i\boldsymbol{\chi}_\Delta + {}^i\boldsymbol{\upsilon}_\Delta^T Q_2 {}^i\boldsymbol{\upsilon}_\Delta \\ &\quad + {}^i\boldsymbol{\alpha}_\Delta^T Q_3 {}^i\boldsymbol{\alpha}_\Delta \\ &\quad : {}^i\boldsymbol{\xi} = [({}^i\boldsymbol{\chi}_{new}^T \; {}^i\boldsymbol{\upsilon}^T \; {}^i\boldsymbol{\alpha}^T]^T \end{aligned} \tag{9}$$

Combining (8)-(9), an extension of (7) is derived which performs piecewise compression of joint trajectories. The optimization problem is defined as

$$\min_{{}^i\boldsymbol{\xi}_{new}} \; {}^i\boldsymbol{\chi}_\Delta^T Q_1 {}^i\boldsymbol{\chi}_\Delta + {}^i\boldsymbol{\upsilon}_\Delta^T Q_2 {}^i\boldsymbol{\upsilon}_\Delta + {}^i\boldsymbol{\alpha}_\Delta^T Q_3 {}^i\boldsymbol{\alpha}_\Delta$$

$$\text{subject to} \begin{cases} \boldsymbol{A}'_{eq} {}^i\boldsymbol{\xi}_{new} = \boldsymbol{b}'_{eq} \\ \boldsymbol{A}'_{neq} {}^i\boldsymbol{\xi}_{new} \leq \boldsymbol{b}'_{neq} \\ {}^i\boldsymbol{x_0} = \text{given} \\ {}^i\boldsymbol{\xi}_{new} \in \mathbb{R}^{2\mathrm{N}+\mathrm{Q}}, J'({}^i\boldsymbol{\xi}_{new}) \in \mathbb{R} \end{cases} \tag{10}$$

The proposed method for optimization of re-targeted motion is summarized in Algorithm 1.

IV. CASE STUDY

This section describes a case study in which our proposed optimization framework was used to re-target human motion capture data to a NAO robot.

A. Dataset

We used a corpus of highly expressive motion performed by a mime artist [23]. Mime artists are trained experts in adapting their body movements and gestures to make them clearly readable to an audience [24], and use the similar principles of simplification and exaggeration as in traditional animation [25]. In robotics, the use of animation principles has shown to have positive effects in interactions with robots with limited expressive capabilities [26], [9]. The data base consists of short scripted dialogues, were the mime displays various intensity levels of five inner states: level of certainty, attentiveness, engagement and positive and negative emotion. The complete data base contains custom performances for three different embodiments: one with face and body, one with head-and-face (no body) and one with only body (no face). In this case study, we use the subset of the data performed for embodiments with only body. For more details about the corpus design, recording and processing we refer to [23].

B. Motion Re-targeting for Virtual Model

In a prior step to using our method, we re-targeted the motion capture data to a skeleton model with the same DoFs as the NAO robot at a sampling frequency of 25Hz. This was done by adding joint constraints on the skeleton model using Autodesk Maya. We then transferred the data from the original recordings by solving an inverse kinematics (IK) problem with weighted positional and rotational targets applied on the hands, feet, torso and head. For this we used the IKinema Action plugin to Maya with manually tuned weight parameters. The process resulted in motion constrained to the skeleton DoFs, but still following the global characteristics of the original data.

C. Executing Motion Sequences in the NAO Robot

In order to test and validate our approach, we performed the optimized and non-optimized motion sequences on the NAO robot. For performing the optimization as stated in (7) and (10), the joint angle and angular velocity constraints of the NAO robot were passed to the framework[1]. The optimization was carried out using diagonal weight matrices Q_1, Q_2 and Q_3 with corresponding diagonal elements being equal to 5, 0.1 and 10 respectively. The domain $\mathbb{R}_s$ for the scaling parameter was set to $[0.5, 1]$. Since the mime artist performs very minimal lower body movements in the dataset, some motors in the lower body were not used during the animation. This was done to make sure that the robot does not loose contact with the floor.

The optimization problem is run over short horizons (approx. 5 seconds) to keep the state vector small in dimension.

[1] http://doc.aldebaran.com/2-1/family/nao_h25/joints_h25.html

The optimization vectors were initialized with the re-targeted sequence obtained from the re-targeting software to ensure fast convergence and proximity to this reference sequence. This could be done because we expect the modified trajectory to not deviate a lot from the source trajectory, except for the required modifications to satisfy various robot joint constraints. As detailed in Algorithm 1, the optimization was solved independently for each joint violating velocity constraints and combined to form the optimized robot trajectory for a given sub-sequence. This process was performed for all the sub-sequences violating constraints. The combined sub-sequences were then executed in the robot along with the audio file from the original motion data.

V. EVALUATION

In this section, the objective and subjective experiments conducted to evaluate our re-targeting approach to synchronize robot gestures with voice sequence are described.

A. Materials

The materials for the evaluation included 6 motion sequences of approximately 10 seconds from the dataset described in Section IV. As a selection criteria, we chose sequences where the re-targeted motion violates joint constraints of the robot. Our final selection included 4 motion sequences where the actor displayed negative emotional expressions and 2 sequences of positive expressions.

Three conditions were used to evaluate our method:

- *Optimized* – refers to the motion sequence resultant from our proposed optimization framework, including constrained optimization and, when necessary, piecewise compression.
- *Reference* – the trajectory resultant from the commercial motion re-targeting tool applied to the virtual model of the robot. Note that this condition may include trajectories that violate the robot's joint angle and angular velocity constraints.
- *Control* – due to the constraint violations, certain subsequences of the reference trajectory cannot be executed by the robot. This condition stretches such sequences so that we can obtain a control condition to compare our optimized approach in the perceptual study.

B. Objective Evaluation

To understand how our *optimization* framework compares with regard to the *reference* trajectory, we plotted the sequences generated by these two conditions for all the robot joints in the aforementioned evaluation sequences[2]. We discuss the obtained results in the remaining of this section, providing illustrative examples of these plots to highlight the effects of both constrained optimization and piecewise compression on the resultant robot motion.

1) Constrained optimization results: Fig. 2 contains a snippet of the left shoulder pitch of the robot while performing a positive expression (the corresponding angular velocity is plotted in the bottom chart of the same figure). It can be observed from the circled portion that the reference motion violates the velocity constraints of the joint, which requires the angular velocity to lie within -8.2 rad/sec and 8.2 rad/sec. In the optimized condition, the robot starts corrective measures ahead of time by slowly increasing the angular velocity of the joint and saturating at the peak allowed velocity. This alternate velocity profile ensures that the robot motion trajectory closely follows the reference trajectory while satisfying constraints (top chart in Fig. 2).

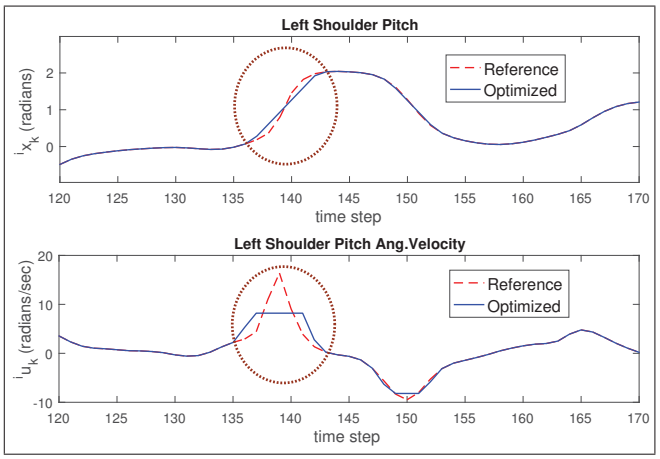

Fig. 2. Comparison of left shoulder pitch trajectory optimized with constrained optimization.

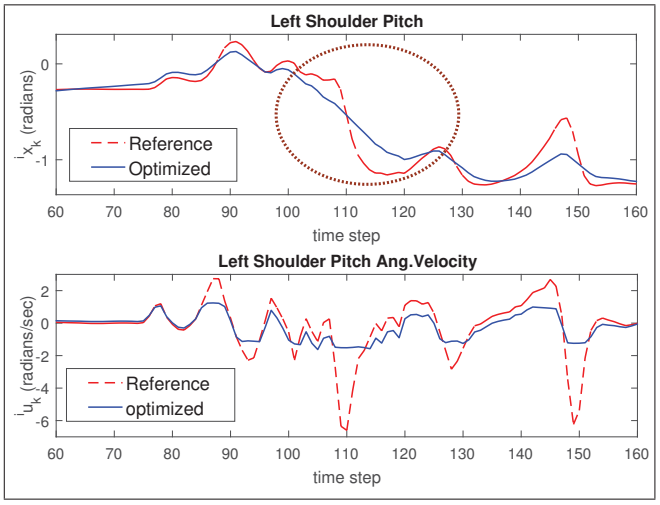

Fig. 3. Comparison of left shoulder pitch trajectory with trajectory optimized using piecewise compression.

2) Piecewise compression results: In Fig. 3, a portion of a left shoulder pitch trajectory of the robot while performing a high frequency motion sequence is plotted. It can be observed that while the human artist performs a fast gesture starting with the quick motion encircled in the figure, the optimized sequence uses a scaled down velocity profile and later saturating it at a lower speed to create a smother trajectory for the shoulder joint. It can also be noticed that the shoulder starts moving earlier performing a slower motion to avoid a fast but short motion. However, such modification does not cause delays with regard to speech, but rather alters

[2]We deliberately left the control condition out of the objective evaluation for being an extension of the reference condition.

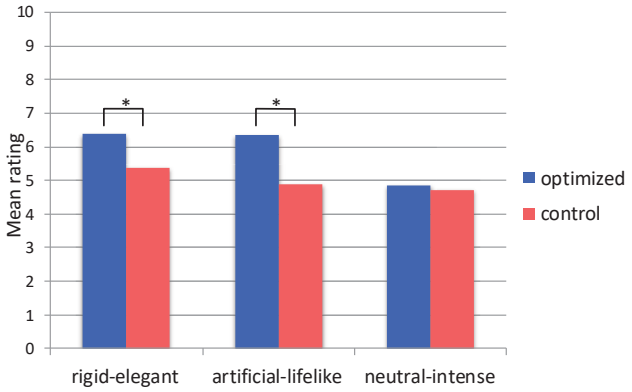

Fig. 4. Mean ratings of the semantic scales in the optimized and control conditions. (*) denotes $p < .0005$.

the way that the joint motion is executed.

C. Perceptual Study

While the results reported in the previous subsection indicate that our optimization framework can re-target robot motion in a much smoother manner than a reference trajectory, it is also important to evaluate whether such differences are perceivable to non-experts observing robot behavior. For that, we conducted a perceptual user study to determine how robot motion sequences generated using our *optimized* approach are perceived when compared to the same sequence generated in the control *condition* (stretched reference motion that can be executed by the robot).

1) Method: We recruited 38 adult participants (27 male and 11 female) through Amazon Mechanical Turk to complete a video-based online survey. The stimuli for the survey included videos of the NAO robot expressing behaviors re-targeted according to the optimized or the control condition. A total of 12 videos were included in the survey (6 different behaviors * 2 conditions). The videos show the robot displaying both verbal and nonverbal behavior. The audio file containing the actor's voice in the original motion capture sequence was overlayed to the video of NAO performing that sequence, meaning that each pair generated from the same source motion capture data has the same verbal behavior. The audio file included utterances such as "No problem, I'll fix it" or "I am not done yet!" that complemented the behavioral gesture.

The survey consisted of two main parts: pair-wise comparisons and individual ratings. In the pair-wise comparisons, participants watched two videos side by side, each one with the same behavioral sequence being re-targeted either with the optimization or control conditions, and were asked to submit answers to the following questions: "In which video the movement of the robot seems more natural?"; 2) "In which video the voice and movement of the robot seem more in sync?". In both questions, participants could chose a third option if they considered that the videos seemed equally natural/in sync. In the second part of the survey, the 12 video stimuli were presented again, but one at a time. After watching each video, participants were asked to rate the robot motion on three different 10-point semantic differential scales based on prior related research [8], [27], [28]: rigid-elegant, artificial-lifelike and neutral-intense.

Participants took, on average, 40 minutes to complete the survey and were compensated for their time. Different versions were created to counterbalance order effects by randomizing the order that each video was presented and by switching the order of the video pairs (left/right).

2) Subjective results: To analyze the pair-wise comparisons between the same motion sequence re-targeted with and without optimization, we coded participants' responses as follows: +1 for responses considering the video showing the optimized condition to be more natural/in sync, -1 for responses favoring the control condition, and 0 when participants considered both videos to be equally natural/in sync. We then averaged all the 6 responses of each participant to obtain two continuous variables ranging between -1 and 1 that represented a participant's preferential method in our two measures of interest, *naturalness* and *synchrony*. One-sample t-tests were run to determine whether one of the conditions (optimized or control) was significantly different to the neutral option of both videos appearing equally natural/in sync, defined as a score of 0. The mean naturalness score ($M = .49$, $SD = .36$) was significantly higher than the neutral score, $t(37) = 8.38$, $p < .0005$, such that participants found the motion resultant from the proposed optimization approach to be more natural than the control condition. A similar trend was found in the synchrony measure, where the mean synchrony score ($M = .49$, $SD = .37$) was significantly different than the neutral score, $t(37) = 8.29$, $p < .0005$, in the direction of the optimized condition.

For the semantic scales (*rigid-elegant*, *artificial-lifelike* and *neutral-intense*), we conducted repeated-measures ANOVA with 2 within-subject factors, study condition (optimized vs. control) and the 6 different source behaviors. In the *rigid-elegant* measure, we found a significant main effect for condition, $F(1,37) = 36.10$, $p < .0005$, $\eta^2 = .49$, such that participants considered the robot movement generated from the optimized condition ($M = 6.37$, $SE = .22$) more elegant than the movement generated from the non-optimized condition ($M = 5.36$, $SE = .20$). Similarly, a significant main effect for condition was found in the *artificial-lifelike* scale, $F(1,37) = 63.10$, $p < .0005$, $\eta^2 = .63$, such that participants found the movement in the optimized condition more lifelike ($M = 6.36$, $SE = .23$) than the movement in the non-optimized condition ($M = 4.88$, $SE = .22$). However, we found no significant main effects on the *neutral-intensity* scale, $F(1,37) = 1.24$, $p = .27$, suggesting that the movement generated from both conditions was portrayed with similar intensity. Fig. 4 illustrates these results.

VI. CONCLUSIONS

This paper presents a motion re-targeting optimization framework which addresses the joint angle and angular velocity constraints of a robot to ensure that verbal and nonverbal actions remain synchronized during real-time behavior realization. Because optimizing high amplitude mo-

tion sequences in real-time can lead to unnatural behavior, the method proposes a non-linear optimization formulation extension to compress and smooth high amplitude sequences.

The presented method was applied to a NAO robot displaying behaviors requiring verbal and nonverbal synchronization. An analysis of the joint motion trajectories showed that our framework successfully performs the desired modifications to re-targeted robot motion sequences. Additionally, a perceptual user study showed that non-experts found the robot displaying behavior generated from our optimization method to be more natural and to results in a better synchronization between voice and gestures than a control condition. While no significant differences were found with regard to perceived intensity, the behaviors resultant from the optimization framework were rated as significantly more elegant and lifelike than the behaviors generated by the control condition.

The proposed optimization framework can be extended to account for other limitations such as the stability of the robot. Future work will also explore the use of this approach to generate similar behaviors with slightly different trajectories, for example, to express the same behavior with varying levels of intensity.

ACKNOWLEDGMENTS

The authors thank Diogo Almeida, Roberto Bresin and Emma Frid for their valuable feedback.

REFERENCES

[1] M. Gleicher, "Retargetting motion to new characters," in *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '98. New York, NY, USA: ACM, 1998, pp. 33–42.

[2] J. Koenemann, F. Burget, and M. Bennewitz, "Real-time imitation of human whole-body motions by humanoids," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, May 2014, pp. 2806–2812.

[3] S. Nakaoka, A. Nakazawa, K. Yokoi, H. Hirukawa, and K. Ikeuchi, "Generating whole body motions for a biped humanoid robot from captured human dances," in *IEEE International Conference on Robotics and Automation*, vol. 3, Sept 2003, pp. 3905–3910 vol.3.

[4] A. Nakazawa, S. Nakaoka, K. Ikeuchi, and K. Yokoi, "Imitating human dance motions through motion structure analysis," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 3, 2002, pp. 2539–2544 vol.3.

[5] P. Agarwal, S. A. Moubayed, A. Alspach, J. Kim, E. J. Carter, J. F. Lehman, and K. Yamane, "Imitating human movement with teleoperated robotic head," in *2016 25th IEEE International Symposium on Robot and Human Interactive Communication*, 2016, pp. 630–637.

[6] D. Rakita, B. Mutlu, and M. Gleicher, "A motion retargeting method for effective mimicry-based teleoperation of robot arms," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '17. New York, NY, USA: ACM, 2017, pp. 361–370.

[7] I. Baran and J. Popović, "Automatic rigging and animation of 3d characters," in *ACM SIGGRAPH 2007 Papers*, ser. SIGGRAPH '07. New York, NY, USA: ACM, 2007.

[8] E. J. Carter, L. Sharan, L. Trutoiu, I. Matthews, and J. K. Hodgins, "Perceptually motivated guidelines for voice synchronization in film," *ACM Trans. Appl. Percept.*, vol. 7, no. 4, pp. 23:1–23:12, July 2010.

[9] T. Ribeiro and A. Paiva, "The illusion of robotic life: principles and practices of animation for robots," in *Human-Robot Interaction (HRI), 2012 7th ACM/IEEE International Conference on*. IEEE, 2012, pp. 383–390.

[10] A. Zhou, D. Hadfield-Menell, A. Nagabandi, and A. D. Dragan, "Expressive robot motion timing," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '17. New York, NY, USA: ACM, 2017, pp. 22–31.

[11] A. D. Dragan, K. C. T. Lee, and S. S. Srinivasa, "Legibility and predictability of robot motion," in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, March 2013, pp. 301–308.

[12] Y. Nakamura and K. Yamane, "Dynamics computation of structure-varying kinematic chains and its application to human figures," *IEEE Transactions on Robotics and Automation*, vol. 16, no. 2, pp. 124–134, Apr 2000.

[13] E. J. Haug, *Computer Aided Kinematics and Dynamics of Mechanical Systems. Vol. 1: Basic Methods*. Needham Heights, MA, USA: Allyn & Bacon, Inc., 1989.

[14] K.-J. Choi and H.-S. Ko, "On-line motion retargetting," in *Proceedings. Seventh Pacific Conference on Computer Graphics and Applications (Cat. No.PR00293)*, 1999, pp. 32–42.

[15] M. Da Silva, Y. Abe, and J. Popovi, "Simulation of human motion data using short-horizon model-predictive control," *Computer Graphics Forum*, vol. 27, no. 2, pp. 371–380, 2008.

[16] M. Do, P. Azad, T. Asfour, and R. Dillmann, "Imitation of human motion on a humanoid robot using non-linear optimization," in *Humanoids 2008 - 8th IEEE-RAS International Conference on Humanoid Robots*, Dec 2008, pp. 545–552.

[17] K. Yamane, S. O. Anderson, and J. K. Hodgins, "Controlling humanoid robots with human motion data: Experimental validation," in *2010 10th IEEE-RAS International Conference on Humanoid Robots*, Dec 2010, pp. 504–510.

[18] K. Yamane and J. Hodgins, "Simultaneous tracking and balancing of humanoid robots for imitating human motion capture data," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct 2009, pp. 2510–2517.

[19] N. S. Pollard, J. K. Hodgins, M. J. Riley, and C. G. Atkeson, "Adapting human motion for the control of a humanoid robot," in *Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No.02CH37292)*, vol. 2, 2002, pp. 1390–1397 vol.2.

[20] S. Wang, X. Zuo, R. Wang, F. Cheng, and R. Yang, "A generative human-robot motion retargeting approach using a single depth sensor," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, May 2017, pp. 5369–5376.

[21] K. Ayusawa, M. Morisawa, and E. Yoshida, "Motion retargeting for humanoid robots based on identification to preserve and reproduce human motion features," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sept 2015, pp. 2774–2779.

[22] A. Zanelli, A. Domahidi, J. Jerez, and M. Morari, "FORCES NLP: an efficient implementation of interior-point methods for multistage nonlinear nonconvex programs," *Int. Journal of Control*, 2017.

[23] S. Alexanderson, C. O'Sullivan, M. Neff, and J. Beskow, "Mimebot: Investigating the expressibility of non-verbal communication across agent embodiments," *ACM Transactions on Applied Perception*, 2017.

[24] J. Lecoq, *The moving body (Le corps poetique): Teaching creative theatre*. A&C Black, 2009.

[25] F. Thomas, O. Johnston, and F. Thomas, *The illusion of life: Disney animation*. Hyperion New York, 1995.

[26] M. J. Gielniak and A. L. Thomaz, "Enhancing interaction through exaggerated motion synthesis," in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*. ACM, 2012, pp. 375–382.

[27] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International journal of social robotics*, vol. 1, no. 1, pp. 71–81, 2009.

[28] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner, "The robotic social attributes scale (rosas): Development and validation," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '17. New York, NY, USA: ACM, 2017, pp. 254–262.