HeadBox: A Facial Blendshape Animation Toolkit for the Microsoft Rocketbox Library

Matias Volonte, Eyal Ofek, Ken Jakubzak, Shawn Bruner, and Mar Gonzalez-Franco

Abstract—HeadBox is a series of opensource tools to do facial animation on the Microsoft Rocketbox avatar library. It includes a tool to create blendshapes out of the facial bones inside Maya and transfer the new blendshapes to the other avatars in the library. We have created a total of 15 visemes, 48 FACS, 30 for the Vive facial tracker. These blendshapes have been released with the original library. An additional Unity demo shows the use these tools with Openface and Oculus Lipsync.

Index Terms—Microsoft Rocketbox, Avatars, Virtual Humans, Open-Source, Blendshapes, Animation, Facial, Rigging

1 INTRODUCTION

Avatars are digital representations of humans that can be used in Virtual Reality and digital media to replicate their appearance and behavior. In particular avatars are critical for tasks that demands social interactions, where facial expressions are a vital component. Indeed, facial animation is a key component of avatar animation, in the same way that moving our body will allow us to interact with our environment our facial expressions will express emotions and help us interact with others. For avatars, facial expressions are also important not only to have good Lipsync but also to provide better non verbal cues during virtual interactions. We need them to feel avatars are alive inside VR.

There is evidence that facial animation has a deep impact during simulated social interactions. Previous research has shown that faces of avatars that are animated receive more attention than those that are not [11]. Studies have also compared the effect of facial cues during interactions and found that avatars providing facial expressions were more appealing, and the interaction was rated more positively than when avatars lack facial expressions [14].

The Enfacement Illusion is also greater when avatars are animated, specially if the avatars are presented in first person perspective in front of a mirror [11]. Animation realism can also affect perceived appeal on characters [13]. Enfacement is the illusion wherein the cognitive representation of one’s face allows a person to embody another individual’s face. Emotional facial expressions may affect the enfacement illusion by increasing the spectator’s grounds to comprehend the cognitive condition of the avatar [2]. In fact, facial animation perception can alter our understood personality of avatars even when these are not human-like [8].

Even though facial animation of avatars still poses many challenges, such as overcoming the uncanny valley, overall it has a lot of potential for our future. Current technology advancements offer promising advancement in the area of avatars, such as the Metahumans library [7], the Microsoft’s Rocketbox library [10], or MB-Lab for blender [15]. These open-source options can provide a good starting alternative for the community to explore the effects of facial avatar animation in virtual reality (VR) and augmented reality (AR).

However, for the case of the Microsoft Rocketbox avatar library [10], the animation is currently restricted to the facial bones, which are not the most widely used form nor convenient form of facial animation, as opposed to blendshapes. Here we present a series of tools that will help bridge that gap. We describe the process of developing facial expressions attributes to the existing Microsoft’s Rocketbox Avatars library [10]. The Rocketbox package is a popular open-source library of 115 human characters representing different demographics (gender, age, race, or professions), fully rigged, and animation ready. We extended the attributes of the existing Rocketbox package by providing facial expressions and lip sync-capabilities. In total we modeled 110 facial blendshapes. These includes 15 visemes, 48 facial action units, 30 expressions compatible with the vive facial tracker, and 17 custom gaze blendshapes interesting for future users.

In all this paper contributes with a tool to create and export procedurally blendshapes out of facial bones to all the Microsoft Rocketbox avatars, an upgraded release of the library with series of embedded blendshapes on the avatars. A Unity demo using the visemes for Lipsync and the FACS [6] animations, using Openface [1] for real-time animation of the face using web camera input. We hope more people will now be able to use our Headbox Maya tool, to customize and generate even more blendshapes.

Authorized licensed use limited to: MICROSOFT. Downloaded on May 10,2022 at 03:41:20 UTC from IEEE Xplore. Restrictions apply.
Finally, this open-source package has the goal of promoting a free exchange of ideas within the virtual human community to drive creative, scientific, and technological advancement.

2 RELATED WORK

Our work builds on top of existing standards and is inspired by other opensourse tools.

2.1 Facial Animation

We aim for our blendshapes to be as compatible as possible with FACS. The Facial Action Coding Systems [6] is a manual that describes multiple facial muscle configurations. The FACS describes the subtle differences in the configuration that outcomes from different muscle movements.

In fact FACS has been shown a viable choice for avatar animation and there are even opensourse libraries like FACSatars that have built on top of it. FACSatars implements openFace [1] for facial landmark detection, head pose estimation, facial action unit recognition, and eye-gaze estimation to output an animated character in Unity or Blender.

Additionally we aim to implement visemes for lip sync performance and additional precombined blendshapes with facial expressions and emotions.

2.2 Avatar Libraries

There are multiple avatar libraries available. However not all of them are opensourse. Here we review a couple of the most used ones and they to see what facial animation characteristics they have.

USC Institute for Creative Technologies: The University of Southern California (USC) Institute for Creative Technologies (ICT) developed the Virtual Human Toolkit. This package includes a set of modules, tools, and libraries conceived for researchers and developers with the creation of virtual human conversational characters [12]. Users can create dialogue responses to input to one or multiple characters, manipulate the agent’s non-verbal expressions, animate the agent’s body, gaze, lip-sync in real time using the Behavior Markup Language (BML), and track the user’s facial and head behaviors through webcam to trigger the avatar’s reactions.

Mixamo: This open-source system allows users to access free characters and an extensive animation library [3]. Users can select motion capture animations in 3D applications like Autodesk Maya (see Figure 3) and apply them to digital characters. Additionally, users can upload their custom characters and auto rig them using Mixamo’s proprietary tool. This process requires the user to place makers on the agent (wrists, elbows, knees, and groin) and the tool creates the body rig automatically. Several of the example avatars that Mixamo provides have facial blendshapes that are compatible with HeadBox, however not all of them include blendshapes nor are following the FACS standard per se, and also sometimes they don’t have visemes.

Epic Games MetaHumans: MetaHuman is a tool for creating high-quality digital character for the Unreal Engine. Users can create photorealistic digital humans, fully rigged and complete facial blendshapes with hair and clothing. This package provides an advanced facial rig system (named “Puppet”) to animate the agent’s facial expressions. The MetaHuman virtual agents support motion capture animation and they can be animated in 3D applications like Autodesk Maya.

MakeHuman: This package is a professional open-source software for creating virtual human characters [4]. MakeHuman allows users to edit granularly the attributes of the virtual agents. Users can modify the agent’s gender, age, muscle, weight, height, physical proportions, skin, and race. Also, the characters are rigged and support motion capture animation. Users can select different outputs (Game, Mocap, Detailed hands, Facial Bones, Detailed Hands, etc.), that will determine the rig structure.

Microsoft Rocketbox: This library contains 115 characters representing different races, ages, and genders. The library is designed for research mainly, and even though the license is MIT free it is mostly in use by researchers in areas such as sociology, psychology, VR, crowds, and HCI. One particular aspect of the library is that its avatars represent everyday characters and professions, which makes it very attractive for research. The avatars are fully rigged and their topology is optimized for performance. These agents are mid-to-high fidelity looking and support motion capture animation. A Rocketbox agent possess a total of 81 joints.

3 SYSTEM DESCRIPTION

Our goal is to develop a system to create and distribute accurate and consistent facial expressions to 115 different characters of the Microsoft Rocketbox Avatar Library. This process involved a first step for selecting the visemes, action units [6], and custom facial expressions to be developed. We want to provide new blendshapes that followed the FACS standard, (Facial Action Code System), based on muscular movements [6] and blendshapes that would allow people to do good lip-sync with visemes. We also account for compatibility with facial animation tools like Vibe Facial Tracker, oculus lip-sync and openface which means we needed to create more blendshapes. While Microsoft Rocketbox Avatars were not designed to support blendshapes, but rather based on facial bones, we developed a new tool, supplied with the opensource Headbox toolkit, that can generate equivalent results.

3.1 HeadBox Blendshape Creation Tool

We developed a Maya based tool, named Headbox blendshape creation tool, which is a custom pipeline for developing, storing, and transferring the blendshapes to all the Microsoft RocketBox characters. In a single step, a designer of a particular blendshape can export a particular blendshape to all the avatars in the library. This system was developed in Python in Autodesk Maya (see Figure 3).

The tool supports the following properties:

- Save and load joints transform data.
- Transfer joints information to all characters.

To transfer the facial motions among different avatars, we take advantage of a similar facial bone rigging across all of the Microsoft RocketBox avatars. However, different avatars have different meshes, as well as bone length. Applying similar rotations of corresponding bones may generate different facial expressions. To compensate for this difference, we applied a correction which is unique per avatar.

Within Maya, we can apply a desired blendshape to the mesh of the avatar’s face, and compare it to the result of applying rotations on the avatar’s facial bones. We estimate a additional transformation of the joint position, that will minimize the difference between the two transformations. Such offset takes into account the variation of the bones lengths.

\[ \delta = \text{JointTargetPosition} - \text{JointOriginPosition} \]
Applying these offset translations to bone origins, which means we make use of the neutral character mesh, generates meshes that resembles the expression over different avatars, while also maintain the facial bone dynamics.

A part of the development of the tool, we used image references from the existing literature to manually model the facial expressions by manipulating the facial joints present of a specific avatar of the library. After generating similar expressions to the reference ones, all model joints’ transform data was saved to a Json file. This process was repeated for all the facial expressions. Using this data, we successfully transferred modeled expressions all other avatars in the library.

In total we modeled 111 facial blendshapes. These includes 15 visemes, 48 facial action units, 30 expressions compatible with the vive facial tracker, and 17 custom facial that show combinations of emotional, tongue, and gaze blendshapes interesting for future users.

### 3.2 Headbox Realtime Demonstration

To test the new blendshapes we developed a real-time tool in Unity that includes openface compatibility [1] and Oculus Lipsync (see attached video).

The lip-sync animation was tested on Oculus LipSync [17] and on Salsa LipSync [16]. Salsa provides the possibility of using 17 blendshapes by default with the option of increasing this number compared to the 15 that are provided by Oculus LipSync. The results with both packages were satisfactory.

![Image of a person demoing the Unity tool with Openface.](image)

One thing to note is that lip-sync animation was significantly better when done from audio than from the computer vision approach provided by Openface. This could be due to limitations on Openface where only 17 action units are represented for the whole face, while visemes alone on oculus lip sync, based only at the mouth already account for 15 variations. This could be overcome if more mouth positions would be detected from a different computer vision solution in the future. Therefore, we recommend using openface for the rest of the face and audio driven animation for speech to blendshape.

The toolbox also includes blendshapes to be used with the Vive facial tracker, however despite them being implemented we didn’t test them for performance.

All the blendshapes can be modulated at runtime inside Unity to reduce their dynamic range using our demonstration starter kit on github.

Although not included in this toolbox, previous work has also shown that full facial animation can be achieved only by audio, like the one proposed by JALI [5].

### 4 Discussion

We decided to use a procedural approach for modeling the facial blendshapes by manipulating generating approximating facial bone transforms, supported by Microsoft Rocketbox avatars (position, rotation, and scale). The current facial rig included a total of 29 joints. This small number of joints can limit the accuracy and quality of the blendshapes. Traditionally artist animated blendshapes also take into account individual vertices, which in our case wasn’t an option if we wanted to be compatible and procedural transfer between different avatars. However, using the Maya tool we developed to transfer blendshapes between avatars, users can now improve blendshapes or create new ones.

The facial mesh resolution, used by the avatar library has limited resolution, optimized for performance. This may have affected the modeling of the facial expressions, in the same way that the restriction on facial bones can also have an effect.

This new toolbox is compatible with other Microsoft Rocketbox tool-sets such Movebox recording studio [9], for recording and playing body motions along with the facial animations.

The Headbox with the Maya tool for generating new blendshapes will be available in a new GitHub https://github.com/openVRlab/Headbox. Also, The Microsoft Rocketbox library with updated facial expressions will be freely accessible inside the original Microsoft Rocketbox avatar GitHub https://github.com/microsoft/Microsoft-Rocketbox.

### References

15. MB-Lab-Community. Lab.