Text-to-movie authoring of anatomy lessons

Vaishnavi Ameya Murukutla, Elie Cattan, Benjamin Lecouteux, Remi Ronfard, Olivier Palombi

Text-to-Movie Authoring of Anatomy Lessons

Vaishnavi Ameya Murukutla[a], Elie Cattan[b], Benjamin Lecouteux[c], Remi Ronfard[a], Olivier Palombi[a]

[a] *Univ. Grenoble Alpes, LJK, Inria, CNRS, 38000 Grenoble, France*
[b] *Anatoscope, 38330 Montbonnot-Saint-Martin, France*
[c] *Univ. Grenoble Alpes, LIG, 38000 Grenoble, France*

Abstract

There is a need for a simple yet comprehensive tool to produce and edit pedagogical anatomy video courses, given the widespread usage of multimedia and 3D content in anatomy instruction. Anatomy teachers have minimal control over the present anatomical content generation pipeline. In this research, we provide an authoring tool for instructors that takes text written in the Anatomy Storyboard Language (ASL), a novel domain-specific language (DSL) and produces an animated video. ASL is a formal language that allows users to describe video shots as individual sentences while referencing anatomic structures from a large-scale ontology linked to 3D models. We describe an authoring tool that translates anatomy lessons written in ASL to finite state machines, which are then used to automatically generate 3D animation with the Unity 3D game engine. The proposed text-to-movie authoring tool was evaluated by four anatomy professors to create short lessons on the knee. Preliminary results demonstrate the ease of use and effectiveness of the tool for quickly drafting narrated video lessons in realistic medical anatomy teaching scenarios.

Keywords: Anatomy pedagogy, Domain Specific Languages, Finite state machines, Computer animation.

1. Introduction

Anatomy is one of the most essential yet challenging subjects in medical education. It is introduced very early in the curriculum, and the student must retain this knowledge throughout their practice. One of the significant challenges faced in anatomy pedagogy is presenting complex three-dimensional body parts in classes. Multimedia methods such as animated videos and 3D anatomical models are popular to visualise the human body. However, the issue of developing an easy-to-use software that enables teachers to create their own lessons without the need for design experts is ongoing. Anatomy professors create all

Email address: olivier.palombi@univ-grenoble-alpes.fr (Olivier Palombi)

the teaching materials used in lessons, either by themselves or by others following their close instructions. In France, until a few years ago, the teachers would draw the part they were teaching on the chalkboard in front of the class as they led the lesson. In this paper, we present a Text-to-movie authoring system for anatomy professors. Our goal is to have fast production, quick visualisation and easier editing giving teachers more freedom at different stages of content creation. We chose text as the method for input in our system as this fits seamlessly with the current way teachers create their lessons and it is also particularly favourable as the teachers did not have to learn a new and often intimidating user interface for design software, making our system intuitive and easy to learn. We present a domain-specific language called the Anatomy Storyboard Language (ASL) which forms the basis of input for our authoring system. The users have to write scripts in ASL in which they list all the anatomical parts that they want to show in the video and outline the directions for the camera movements and animation of the parts. Our system reads the script, and the resulting animation is played in a Unity player. Then we present the results of our evaluation of the software done with four anatomy professors. The teachers could make their own narrated, animated video lessons on the knee. The study showed that their limited experience in animation or video editing did not hold them back from creating an informative course.

2. State of the art

Anatomy learning has two parts: theory and practical. The lesson starts with classroom lectures with visual aids such as chalkboard drawings and slide presentations. These sessions are for the entire class, and then students are divided into smaller groups for practical dissection sessions on cadavers and prosections. The students have better access to the teachers to ask questions and are informally evaluated on their knowledge by oral questioning during the practical sessions. However due to increasing class sizes, reduced faculty and less time allotted for the anatomy curriculum it has become imperative that traditional methods of teaching be augmented with multimedia approaches. In our survey, presented further in the paper, teachers have stated that they are willing to invest their time and expertise to develop digital content by themselves.

Various multidisciplinary techniques have been introduced to make anatomical learning more engaging and effective [1]. While there cannot be a complete digitalisation of anatomical lectures, it is beneficial to use multimedia tools as a part of the blended learning approach to better prepare the students for the hands-on practical sessions [2, 3, 4]. The use, for instance, of a virtual dissection table, with its straightforward handling, enables the instructor to create multimedia content that is easy to record. These live sessions are often highly valued by students [5]. This approach is not only engaging for the students but also cost-effective [4, 6]. As anatomy involves the study of complex spatial structures, it can benefit from recent results in computer graphics and informatics [7]. The effectiveness of new methods has been confirmed in previous studies [8, 9, 10]. We can particularly mention immersive approaches based on

3D anatomy software like Z-anatomy software, which allow, among other things, the use of 2D sections with didactic value well known to anatomists. Student engagement and immersion can be further enhanced by using 3D glasses like Oculus, which can be simultaneously used by multiple users [11]. All of these tools can be combined with audio recording capabilities for course recording. However, the approaches can be more sophisticated. Pereira et al. use a mix of teacher-created recordings and previously available material, whereas Hoyek et al. use the extensive library of 3D animated videos created by Lyon 1 university in collaboration with a graphics team [1]. Both these studies highlight the need for a teacher authoring system for animated content creation. The videos made by the Anatomie 3D Lyon team have been integral in our process of designing and developing an authoring system. As these videos fit the standard practices of teaching anatomy in France we modelled our final output to be similar to them. Animated videos are the most commonly used teaching aids in anatomy [12]. Even beyond the course suggestions, students reported using video sharing sites such as YouTube to learn new concepts, clear doubts or better visualise anatomy [13].

Authoring high-quality 3D animation is a costly process requiring the skills of designers, animators and directors as seen in the workflow chart in Figure:1. Likewise, creating 3D animation for a new course in anatomy involves time, effort and money. Few authoring systems are easy enough for an anatomy teacher not trained in computer graphics to create 3D animation by herself. Recently, commercial game engines have started providing visual programming tools to facilitate the creation of 3D animation (Unreal Engine's blueprints [2], Unity's Cinemachine [3] and Timeline Editor, Huttong Games' PlayMaker [4] and even spatial programming tools to facilitate the creation of mixed-reality content (Unity's EditorXR [5]). But they are better suited for expert game developers than medical anatomy teachers.

Clearly, there is a need for authoring tools providing better support for authors who are experts in their own field (human anatomy) and not in computer graphics and animation. In the field of web design, this problem has effectively been addressed by introducing a graphic style sheet, which clearly separates the roles of content experts and style experts. The corresponding notion of an animation stylesheet has been advocated by Ken Perlin [14] but remains an open research question.

Text-to-movie (or text-to-scene) authoring is a general class of methods that have been proposed for automatically generating 3D graphics and animation from text written by a domain expert. An overview of these methods are given in [15]. Good results have been obtained in limited domains, such as generating 3D scenes from natural language accident reports [16, 17] or generating cartoon

[1] https://www.youtube.com/channel/UCHKlhyLLAxFA69nOONK9wPA
[2] https://docs.unrealengine.com/4.26/en-US/ProgrammingAndScripting/Blueprints/
[3] https://unity.com/unity/features/editor/art-and-design/cinemachine
[4] https://hutonggames.com/
[5] https://github.com/Unity-Technologies/EditorXR

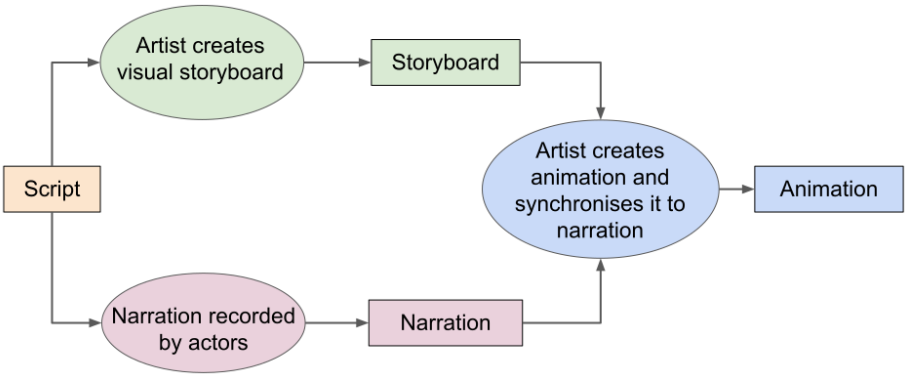


Figure 1: Standard workflow for creation of narrated animation

animation from scripted dialogue scenes [18]. Commercially available text-to-movie systems such as NawmalMAKE [6] and Plotagon Studio [7] are specifically designed to generate dialogue scenes in selected cartoon styles. Xtranormal Technology Inc. first used the term text-to-movie to describe their authoring system that combined the previously used text-to-speech and text-to-scene concepts. Text-to speech enables text to be converted into speech signals that imitate human voice and intonation. This can be used to convert written dialogue in scripts into speech or as voiceovers in instructional videos. Text-to-scene enables the visualisation of natural language descriptions. This needs 3D models that are labelled and positioned in the 3D world depending on the text descriptions. A combination of these two technologies, along with the addition of animations and interactions of virtual actors with the 3D scene based on text descriptions, falls under the domain of text-to-movie authoring. In a system called State developed by Xtranormal, the input is "text" consisting of dialogues and a combination of markers that are similar to emojis. These markers direct the actor movement, facial expressions, posture, interactions and also the camera movements. In essence, the input method is a mix of natural language text and markers that directs both camera and actor actions. The output is synthetically generated and can be visualised immediately as the 3D scene is being built and directed (Figure:2). Generic text-to-movie authoring is also an active area of research. Ye and Baldwin described a system for automatically generating storyboards from natural language movie scripts [19]. More recently, Marti et al. described the CARDINAL system [20] which automatically generates 3D animation of movie scripts for pre-visualisation .

One key requirement for a text-to-movie application is the ability to associate graphic objects and animations with a large vocabulary of concepts expressed in natural language. In the field of medical anatomy teaching, My Corporis

[6] `https://www.nawmal.com/`
[7] `https://plotagon.com/`

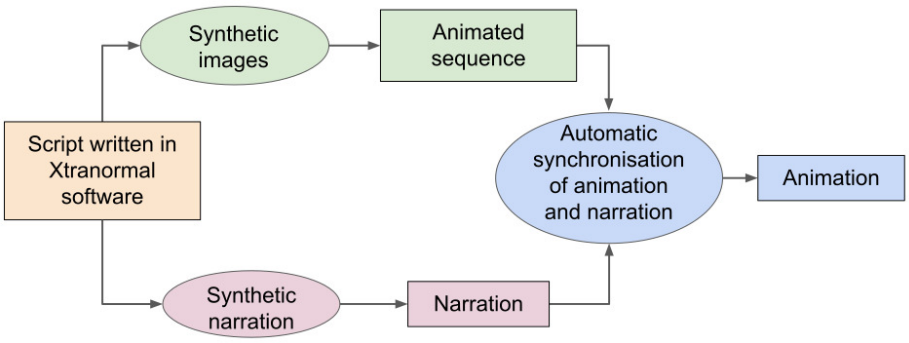


Figure 2: Text-to-Movie workflow for creation of narrated animation - Xtranormal

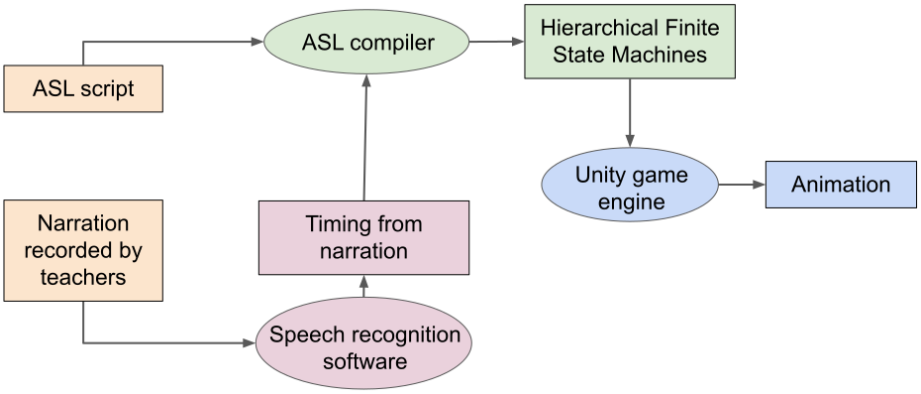


Figure 3: ASL workflow for creation of narrated animation

Fabrica (MyCF), provides an extensive ontology of human anatomy concepts linked to 3D models [21], which have been used to demonstrate quick browsing and visualisation from high-level queries. In this work, we extend the idea by proposing a system that automatically translates anatomical queries into animated movies.

Another key requirement for a text-to-movie application is the ability to automatically place and control the virtual camera so that anatomical entities are shown appropriately and with the right timing. Previous work in intelligent cinematography and editing such as those done by Galvane et al. [22] and He et al.[23] must be carefully adapted for the special case of anatomy. In this work, we introduce a special purpose authoring language with a rich vocabulary of cinematographic terms relevant to medical anatomy as seen in the workflow chart in Figure:3.

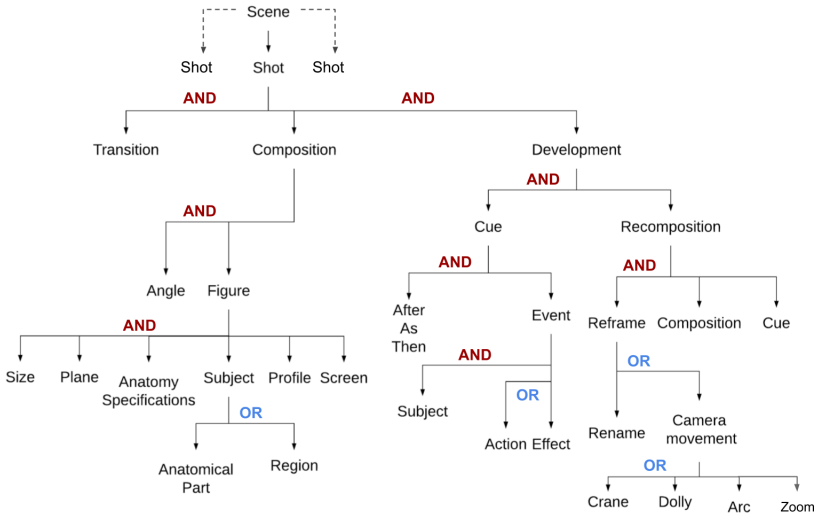


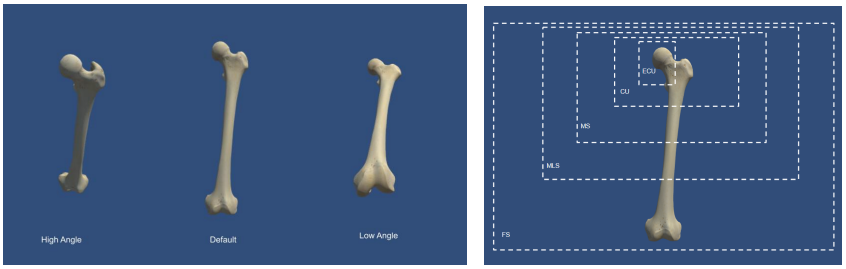
Figure 4: And/Or Graph representation of the Anatomy Storyboard Language grammar. ASL scenes are made of shots containing an initial composition and one or more optional developments.

3. Text-to-movie

3.1. Anatomy Storyboard Language

The input for our system is the text written in a formal language called the Anatomy Storyboard Language (ASL). It is a domain-specific language in which the video to be produced is written as a set of unique sentences. Each sentence describes all the visual elements, camera actions and animations seen from the start of the recording till the camera stops. It is an anatomical extension of the Prose Storyboard Language [24] that was designed for annotating and directing movies.As each sentence generates a complete shot, it must have all the information necessary to transition into the shot, build the composition, direct camera movements, record all the developments from the initial composition and finally describe the last composition before the camera stops. the detailed And/Or graph of the grammar for Anatomy storyboard language is presented in Figure: 4.

ASL is a context-free language with terminals such as anatomical entities and cinematographic terms, and non-terminals, such as initial compositions and subsequent developments. The terminals are either generic terms used for camera action or animation, or specific terms referring to the subjects described in the shot such as anatomical parts and regions. The nomenclature of these specific terminals is derived from My Corporis Fabrica (MyCF) [25, 21] , an extensive ontology that describes structural and functional relations of different parts of human body and their relations in 3D models.



(a) Angles (b) Anatomy sizes

Figure 5: Angles and sizes in Anatomy Storyboard Language

Compositions are descriptions of all the elements that are seen in a particular frame. They need to be comprehensive in detailing the angle, size, the plane of view, anatomical location or specification, profile and relative screen position of the subjects viewed. The subjects in our case are anatomical parts and regions in the complete 3D male Zygote model for the human body [8]. The *Size* term specifies the extent of the subject that is visible within the camera frame. Describing the size for anatomical subjects is more difficult as it has to be relevant for the complete composition. In Figure: 5(b) sizes are illustrated for the femur with the head of the femur as the default centre but in Figure: 8 it is written as a Close Up on Femur, Tibia and Patella thereby making Patella the centre of the composition. *Screen* describes the position of the subject in terms of screen coordinates. This is useful if there is more than one connected body part in the scene. For now we concentrate on teaching one anatomical region at a time and this region will automatically be centred on the screen.

The most important descriptive elements that are essential in building the composition are *Plane*, *Anatomical specification*, *Profile*. *Plane* refers to the hypothetical planes that divide the human body. In ASL they define the view in which we see the anatomical parts and direct camera position accordingly. The three principal planes are *sagittal* (divides the body into left and right halves), *frontal* (divides the body into anterior and posterior parts) and *transverse*(divides the body into upper and lower parts) as seen in Figure: 6. If a plane is not mentioned in a composition then the system will automatically assign a plane in the vertical axis *(sagittal or frontal)* based on the *Profile* information but if the desired composition is in horizontal axis *(transverse)* then it must be mentioned in composition.

It is also important to note, especially in *transverse* plane, whether the anatomical part is viewed from the proximal end (close to the centre of the body) or the distal (further away from the centre). These details are mentioned in *Anatomical specification*. Finally, *Profile* describes the orientation of the part

[8] `https://www.zygote.com`

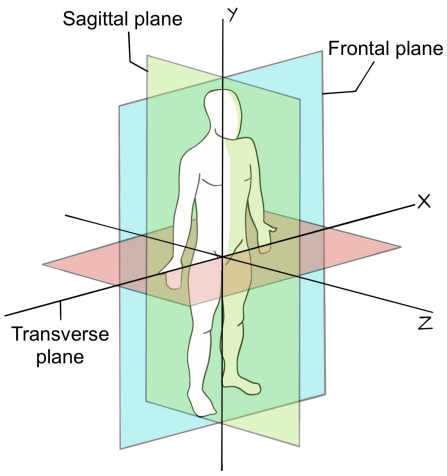


Figure 6: Anatomical planes

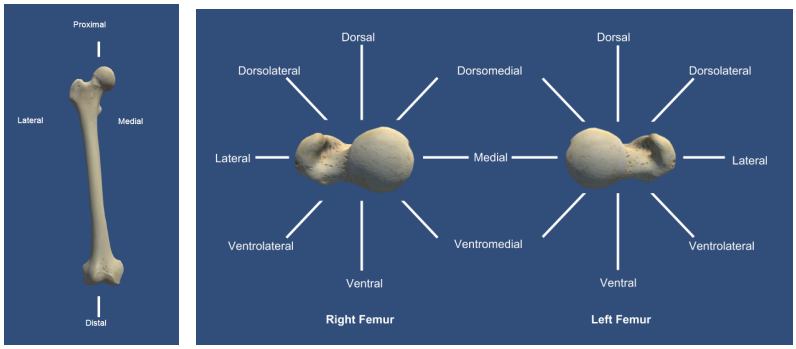
in relation to the camera. It is the side of the subject that is viewed by the camera. The anatomical profile is shown in Figure: 7(b).

3.2. Animation

Sentences written in ASL are parsed via the Parsimonious[9] parser in the python language. The parsed sentences are then translated to a hierarchical finite state machine (HFSM), with one state per composition or development. The HFSM is described in a XML format with specific tags to define each state of the machine. Finally, the HFSM is interpreted and executed at runtime to generate the desired animation. We now describe each of those steps separately.

Style sheets. Some of the descriptive terms of the ASL need to be converted to numerical values in HFSMs. For example, in all the lessons written in this paper, we specified that the ASL term "high angle" will be translated to a 45 degrees bird's eye view. This numerical value of 45 is defined in an animation style sheet [14] along with other parameters such as the camera speed. The teachers can edit this stylesheet depending on their preferences, giving them more nuanced control over the video-making process. Some aspects are not included in the ASL grammar, such as time spent on a composition or default time taken for a camera movement. These are global values that change the

[9] `https://github.com/erikrose/parsimonious`



(a) Specifications (b) Profiles in transverse plane

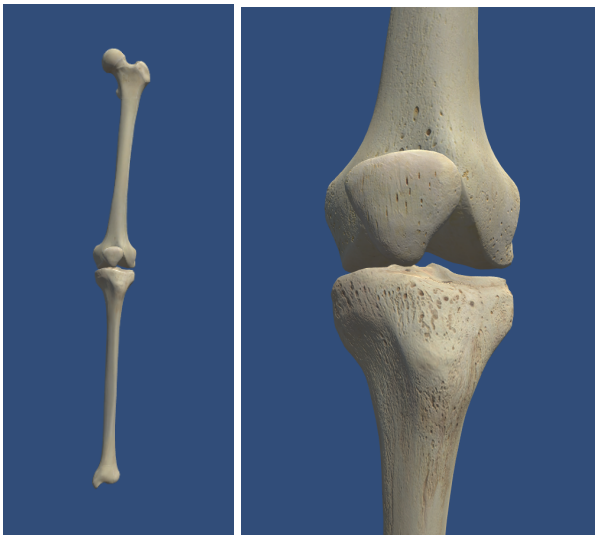
Figure 7: ASL specifications and profiles.

total run-time of the video and can be edited directly in the animation style sheet.

My Corporis Fabrica. All the anatomical terms used in the ASL authoring system are derived from MyCF. It is an ontology connecting the details of anatomical parts with their functions and as MyCF is queryable, we can write queries to obtain the list of parts we need to include in our models.

The list of parts is added to the grammar under the Anatomical Parts non-terminal. They can be updated or changed according to the lesson. When transitioning to a new *state* of an HFSM, the application first analyses this list of components named after the MyCF ontology. The names of parts in MyCF need to match the terms of 3D objects in the Zygote model used in Unity player. If the names do not match, we use a dictionary to convert MyCF names to the 3D objects of the zygote model. This conversion is not trivial since the relation is not bijective. Some MyCF objects are divided into sub-parts represented as one unique entity in the 3D model or vice versa. For example, "head_of_right_femur" exists in MyCF, but the femur is not subdivided in the zygote model, and only the whole bone can be displayed in this case and in the other case, "Right _gastrocnemius" is listed as one part in MyCF and is divided into its component parts of "Right_gastrocnemius_medial_head", "Right_gastrocnemius_lateral_head" and "Right_Achilles_tendon" in Zygote model. In these cases, we need the dictionary to convert the terms used in ASL scripts to match the terms used to build the 3D model in Unity. The dictionary is also built by using the querying feature of MyCF. In our Gastrocnemius example, we query in MyCF to find all the parts that make up the muscle and then find their names in the Zygote model.

Unity player. Our collaborators in the Anatoscope startup developed an application using the Unity 3D game engine to generate the desired animation at runtime from the HFSM obtained from the ASL script. The application is thus an interpreter from a specific XML format to 3D videos of anatomy.



(a) FS Left Femur Tibia Patella ventral (b) CU Left Femur Tibia Patella ventral

Figure 8: Sizes in HFSM

The player first builds a 3D space with the zygote model at the center with appropriate lighting. Then it adds the camera to look at the anatomical parts that were mentioned in the script. These parts are listed under a specific anatomy tag in the XML file for each *state* of the HFSM. The placement of the camera is computed automatically using both internal scene data and data from the ASL script converted through the animation style sheet. We use the bounding box of the 3D models to encompass the whole composition when view angles are setup depending on the ASL information. As the shot develops and there is a camera movement, the application computes a path from its previous position to the next one. It adjusts its orientation according to new parameters. The time taken for this movement is set either in the style sheet if there is no narration or by the time set by the narration if there is. If an *animation* tag is present in a XML state, the 3D models get animated following a previously registered animation stored in the application, such as the flexion of the knee. This animation database has been manually built and can be expanded as per necessity using Anatoscope software.

3.3. Narration

We then extended the system to include narrated voiceovers for the video lessons. The teachers have a choice to add narrations to their lessons. Suppose they want to have a narrated lesson. In that case, they record it and pass that recording through an Automated Speech Recognition(ASR) software developed by our collaborators at Laboratoire d'Informatique de Grenoble. ASR systems

convert audio files and convert them into a sequence of words without a prior transcript of the narration. It works perfectly for us as the teachers can record themselves directly without providing a script for that narration. The input for this system is a .wav audio file. The first step is to perform speech recognition to generate a transcription. Then we use the Kaldi system [26] of forced alignment to align audio with the text. Kaldi is an open-source C++ toolkit for ASR and speech processing. Forced alignment is a process that takes the text from the transcript of audio and finds the time stamp of every word as they occur in the audio segment. At the end of this process, we have the narration transcript used to build the ASL script and time stamps for each spoken word used to retime the videos.

After we get the time for narrated segments and match them to their corresponding ASL counterparts for the video, divide the time between the time spent in that HFSM state and the time taken for the camera movement, if there is any. If there is no narration, the time spent in each state or the delay and the time for camera movement is set by the style sheet's values. It is the usual time taken for the state to progress in the absence of narration. If the time taken for narration is more than twice the sum of the delay and time for camera movements, we calculate the extra time taken by the narration. This extra time is divided and added to the delay and the time taken for camera movements accordingly. If the time taken for narration is more than twice the time taken for the usual state time, we add two-thirds of this extra time to the delay and the remaining one third to the camera movement. We add more to the delay as more narration signifies that the teacher wants to spend more time describing the anatomical parts seen in that state. We don't want to extend the camera movement time as that will interfere visually with the shot as that, unlike the delay, can be seen in the video. If the narration time is less than the usual shot time, the difference between these two values is divided and subtracted from the delay and camera times. If narration is associated with a lesson, but not narration is linked to a state, then that will be assigned the delay and camera time from the style sheet for videos with narration.

4. Experimental evaluation

After we completed the authoring system, we evaluated its ease of use with four anatomy professors from Laboratoire d'Anatomie Des Alpes Françaises (LADAF) surgical school. They were from varied surgical specialities such as orthopaedics and trauma surgery, pediatric and urologic surgery, cardiothoracic surgery and endocrinology and public health medicine. The evaluation was done over two sessions. In the first session we conducted a short interview of their current teaching practices and then introduced the Anatomy Storyboard Language. At the end of the session, they were given the outline for the second session and were asked to record four short audio narrations for lessons on the knee. The four lessons were on the articulations, ligaments, muscles and the movements of the knee joint. Teachers then had to record the audio on any of the devices they were comfortable with and send them to us before the second

session. The audio files were then proceeded using Automatic Speech Recognition (ASR) alignment software to extract the time stamps for each spoken word and get the text transcript of the narration.

In the second session, the teachers built the videos based on their recorded narrations. They referred to the transcripts and created the scenes for their lessons one development at a time. Scripts are written in comma-separated values (CSV) files. The first column of the file has the narration and the second column has the ASL sentence fragment that describes the visual elements the teachers want to show for the given narration. Narration for each development or state was added from the transcript. In this way, our authoring system puts 'narration first' as the visual components are built off of the audio parts, and the time spent on the visual states is also based on the duration of the narration for each state. They first referred to and divided the transcript into parts, and for each part, they wrote the ASL fragments. They could visualise their videos and animations immediately and could edit them in real-time as they built them. They could also expand the list of anatomical parts in the grammar to include the parts they were interested in showing. The first session was an hour long and the second session lasted for two hours.

After the second session, we recorded the videos lessons and added the narrations and sent the completed animated, narrated anatomical lessons on the knee joint back to the professors. Then we organised a short, 15-minute video call to perform the NASA task load test and analysed the results.

Pre-test interviews.
The object of the pre-testing interviews was to understand the current system of teaching anatomy in LADAF and assess the teacher's familiarity with video or graphic authoring systems.

1. What is the typical structure of the anatomy classes?
   The general anatomy curriculum in French medical education is divided over two years. The first-year courses have a strength of about 1800 students and are all held online. The first-year courses are similar for students from a wide range of medical and related fields of dentistry, physiotherapy, nursing and others. As the students are from diverse fields, the courses are pretty generalised. They also have in-person teaching sessions in the amphitheatre for smaller groups of about 100 students each. After the students finish the online courses, if they have any questions, they send them to the professors 3 to 4 days before the live teaching sessions. The professors incorporate these questions into their lessons and discuss problem areas in detail with the students. At the end of the first year, examinations enable the students to choose their field. Around 1/10 of the students from the first year pass the exams to join the medical track. These students then start their second year of anatomy courses. The second-year courses are more clinically oriented and focus on problem-based learning by introducing clinical case studies.
2. What are your prefered tools/ multimedia methods for teaching?
   During the first-year courses, both online and in the amphitheatre sessions,

the preferred teaching method is via PowerPoint slides. For the online classes, these slides have a voice-over from the professors. For the in-person sessions, they reuse the slides without the narration and can show either surgical data or animated videos and talk over them.

3. How were these videos/other media made?
   The teachers created their slides based on the learning objectives they wanted them to achieve for that lesson.A learning objective is a clear and precise goal set before the lesson by the instructors or institutions for the students. At the end of a lesson, a student should successfully reach these objectives. All the teachers have experience in drawing their illustrations. They either draw on paper and scan them or draw on graphic tablets. In addition to their illustrations, they also use images from surgical procedures, dissections and 3D models. After they create the slides, they record their narrations using the sound stage in LADAF. These narrations are then processed and synchronised to the slides by an audio engineer. The use of videos and 3D models in teaching is based on personal preferences. As stated earlier, they are more commonly used during the amphitheatre sessions rather than the main online classes. One professor had also commissioned a series of videos for his lessons on the pericardium (the double-layered protective sac around the heart). The 3D model was built from computerised tomography (CT) scans of the heart, and then the labels and camera movements were added to this model. The work was done by an engineer under the direction of the professor and took six months to complete.
4. How familiar are you with the animation authoring pipeline?
   All the professors that we interviewed stated that they did not have any experience with animation software or were familiar with creating an animation.
5. Have you used any authoring systems before?
   None of the professors used an authoring system for creating animated videos before. They were interested in testing a system in which the input for the video was a text-based script. This method aligned with the pipeline they had already in place for creating their slides and narration.

After the interview, the teachers were introduced to ASL grammar using the And/Or tree. Then the professors were asked to record four short audio clips on four concepts of the knee joint, the articulations, ligaments, musculature and movements. They could record this when convenient, and on any system they liked. They were asked to send the audio clips to us before we conducted the second session. We analysed the audio files sent by the teachers using the ASR alignment software. The input for this software is the .wav file of the audio narration, and the output is a text file with the time stamp for each spoken word. We then create a text file of this audio transcription which serves as a guide and a script for the teachers to develop their ASL visual directions. The metrics for the lessons are presented in table 1. The storyboards for the lessons made by one of the professors on Articulation are provided in Figure: 10 and

Table 1: Metrics for the 4 lessons made by the 4 teachers given in minutes and seconds

Lessons	**Teacher 1**	**Teacher 2**	**Teacher 3**	**Teacher 4**
Articulation	1:42	1:05	0:33	1:22
Ligaments	2:17	2:05	1:03	1:38
Muscles	1:37	0:31	1:23	0:55
Movements	2:00	0:32	1:09	0:17
Total	7:36	4:13	4:08	4:12

11.

Performance assessment using the NASA Taskload test.
The mental workload for creating video lessons using our system was measured using the NASA Task Load Index or NASA-TLX. It is a commonly used multidimensional assessment tool for measuring subjective mental workload developed in the 1980s by Sandra Hart in the Human Performance Group at NASA's Ames Research centre [27]. It measures the mental workload of a task using six subscales or dimensions, mental demand, physical demand, temporal demand, performance, effort and frustration. The subscales are nearly independent of each other and are general. It is also important to note that all tasks are not influenced equally by six subscales. Some of the variables may be irrelevant in specific tasks. Still, this issue is addressed in the weighting procedure, where the user is presented with 15 pairwise comparisons of the six subscales. The user has to chose which of the pair of variables affected the task more. This way, the subscale that was not as important while performing the task will be assigned a lower weight, and their influence on the final score will reflect their contribution to the mental workload. We used software created by Dr Keith Vertanen to perform the test [10].

Results.
We analysed the six subscales' raw scores from the rating stage and the final workload after the weighing process. We present the average for each subscale in Figure: 9. This is before the weights have been applied and represent the quantitative subjective rating for each subscale from 1 to 20 (raw scores are given in table 2). We find that our system scores well for physical demand, frustration and effort. It also scores favourably for performance. This means that the users found our authoring system to have minimal physical demand and that it did not have a high contribution to the overall mental workload. The physical components of our system were typing an ASL script and recording narrations. The teachers were very familiar with both of these actions. They compared this with other 3D software they were familiar with, in which they had to manipulate the 3D model using an onscreen navigation bar that required a lot of clicking. They found our system physically more intuitive. This was reflected

[10] https://www.keithv.com/software/nasatlx/nasatlx.html

Table 2: Ratings for each subscales of NASA Taskload test before weighing

Sub scale	**Teacher 1**	**Teacher 2**	**Teacher 3**	**Teacher 4**
Mental Demand	14	10	16	13
Physical Demand	2	1	2	2
Temporal Demand	10	13	14	8
Performance	6	5	10	11
Effort	8	3	8	7
Frustration	8	1	2	4

in the low scores for frustration which meant the teachers were not discouraged or annoyed by the system. They did not have to learn any new actions, and for the effort they put in, they achieved the goal they were expecting. The average performance score was 8 out of 20.

The main contributors to the workload of the authoring task were the mental and temporal demands. The teachers had to learn the partonomy and taxonomy of a formalised cinematographic language, ASL. They are not used to thinking of anatomy in terms of the shot sizes or angles. The placement of a camera and its related views and movements were wholly new and took time to be internalised. Another factor was that this was also in English, which is not the native language for the professors. So, they had to think of the visual description of the video with two inbuilt levels of difficulty, a new cinematographic vocabulary and a non-native language. They also specified that they do not have any prior experience in authoring to compare the mental demand of this system. The weighing system helps us in this regard as the teachers can choose which subscale contributed more to the workload, but the comparison is only within the subscales and not with another system. It would be ideal if the teachers used other approaches of creating videos to compare to ours. The temporal demand was high as a consequence of high mental demand. The task required new concepts that took a lot of trial and error to master. Again the teachers were comparing the time taken to create the videos using our system with the time taken to create slides. They found it challenging to unhook this comparison and to treat this task as a stand-alone entity. Despite these issues, the teachers were very enthusiastic about using the system in the future. They recognised the potential for creative freedom in designing their own digital content and were willing to invest the time and expertise to become more familiar with the authoring tool. They were very enthusiastic about creating their library of videos that can be edited and fine-tuned on a lesson to lesson basis. As the professors involved in our study were at different stages of their teaching careers ranging from veteran teachers with over twenty years of experience to relative newcomers, we had interesting and inclusive feedback. The teachers just starting their teaching tenure were significantly invested in learning and using our system and agreed to further studies to improve the software. They also decided to use the lessons made using our system in their actual classes to get feedback from the students.

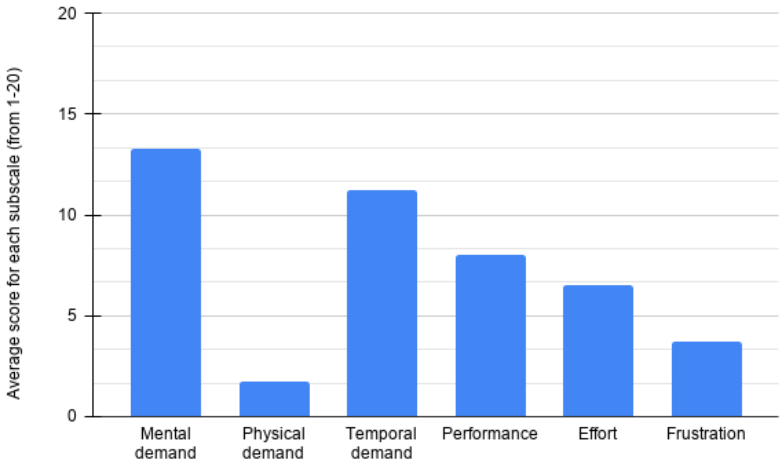


Figure 9: Average score for each subscale before weighing

Qualitative feedback and further comments.

After the complete sessions of authoring and evaluating the new authoring systems, the teachers expressed interest in further testing the system to get used to this content creation method. The main areas where they faced issues were learning the cinematographic vernacular and estimating the speed of narration. They wanted to use the system more to get used to these new elements. Most of them do not use videos in their regular classes as they did not have the means to create the videos themselves. The teachers stated that they would like to include videos in their lessons if making them were easy. They were very enthusiastic about our authoring system and planned on using it further to create longer lessons.

Concerning the videos themselves, the teachers gave us a list of features they would like to incorporate in the future. These include controlling the opacity of specific parts to see through them, having animated labels of parts that moved as the camera moved and incorporating their current slides and figures into the video lessons. We are already working on these suggestions and chose not to include some of the features into the test version to concentrate more on the ASL authoring aspect. We will incorporate them in the later versions of our system. We plan to conduct the same experiment with the new features and the teachers with prior experience using our authoring system. We believe that will decrease the mental load for the task.

The authoring tool we propose is generic enough to be able to manipulate other 3D content libraries. The ASL language does not depend on the models used; it uses identifiers that are linked to anatomical entities. It is sufficient to provide a correspondence map between anatomical entities and the corresponding models. This flexibility also allows, for example, to associate a collection

of reference anatomical cross-sections (2D) that could be called via ASL when needed.

5. Limitation of the project

Attempting to intervene in the production pipeline of animated videos for anatomy education is essential to facilitate the inclusion of 3D digital content in teachers' courses. Our solution introduces cinematic concepts that are unfamiliar to educators. To master the authoring tool, one must learn these concepts and become proficient in the ASL language. This can be a barrier to its use. This difficulty seems more easily overcome by younger anatomists who are more comfortable with computer languages and more inclined to seek assistance in designing their courses. This limitation is even more true as it appears much simpler to learn how to manipulate a virtual table or 3D anatomy software and improvise animations to illustrate discourse than to learn a formal language (ASL) that is constraining. However, the perspectives offered by ASL go far beyond the production of a single video. Beyond simplified content updates, transparent changes for authors of the source 3D library, the real added value lies in the concept of a formalised language. Indeed, ASL opens the doors wide to generative AI. The creation of a language model that includes ASL allows for the envisioning of high-level assistance in the use of 3D in anatomy courses. The power of generative AI in the service of anatomy education will inevitably require a formalisation of the visual dimension of training materials

6. Conclusion

We introduced and evaluated a new system for authoring anatomy video lessons with four professors of anatomy from the Grenoble Centre Hospitalier Universitaire (teaching hospital). The teacher had no prior experience with video creating tools. After a short introduction and hands-on tutorial session, the teacher made and manipulated the 3D scenes from the zygote model of the human body using our text-based input system. They learnt the Anatomy Storyboard Language, which combines cinematographic and anatomical terms that describes all the visual elements seen in the video. The teachers learnt to incorporate their audio narrations into the video lessons and found the authoring system intuitive and interesting. The learning curve is a bit steep as there were new terms and concepts from cinematography that they had to learn and use, but the overall mental workload for the complete authoring task was not high. This workload will only reduce with further use of the system as their familiarity with these new concepts improves. Based on this feedback and the enthusiasm shown by the teachers to learn and use our tool, we will continue to make improvements they suggested, such as including animated labelling and the ability to change the opacity of parts. We focus on the authoring tool as the current thesis aims to teach anatomy pedagogy. The logical next step is to use the lessons made using our system in the anatomy curriculum and get

feedback from the students. These validation experiments are being planned for the future and in collaboration with the Anatosocpe start-up team. Given the positive feedback for the authoring tool, we are confident of improving its features and presenting a robust, easy-to-use system for anatomy teachers to use.

Acknowledgement

This work was funded by the French National Research Agency through An@tomy2020 project (ANR-16-CE38-0011). A CC-BY public copyright license has been applied by the authors to the present document and will be applied to all subsequent versions up to the Author Accepted Manuscript arising from this submission, in accordance with the grant's open access conditions.

References

[1] M. Estai, S. Bunt, Best teaching practices in anatomy education: A critical review, Annals of Anatomy - Anatomischer Anzeiger 208 (2016) 151–157.

[2] J. Majerník, L. Szerdiová, Preparation of medical students for cadaveric anatomy using multimedia education tools, in: 2017 International Conference on Information and Digital Technologies (IDT), 2017, pp. 252–255. doi:10.1109/DT.2017.8024305.

[3] A. Stirling, J. Birt, An enriched multimedia ebook application to facilitate learning of anatomy, Anatomical Sciences Education 7 (1) (2014) 19–27. arXiv:https://anatomypubs.onlinelibrary.wiley.com/doi/pdf/10.1002/ase.1373, doi:https://doi.org/10.1002/ase.1373.
URL https://anatomypubs.onlinelibrary.wiley.com/doi/abs/10.1002/ase.1373

[4] M. K. Khalil, E. M. Abdel Meguid, I. A. Elkhider, Teaching of anatomical sciences: A blended learning approach, Clinical Anatomy 31 (3) (2018) 323–329. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/ca.23052, doi:https://doi.org/10.1002/ca.23052.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/ca.23052

[5] W. Alasmari, Medical students' feedback of applying the virtual dissection table (anato-mage) in learning anatomy: A cross-sectional descriptive study, Adv Med Educ Pract 12 (2021) 1303–1307. doi:10.2147/AMEP.S324520.

[6] W. Clifton, A. Damon, E. Nottmeier, M. Pichelmann, The importance of teaching clinical anatomy in surgical skills education: Spare the patient, use a sim!, Clinical Anatomy 33 (1) (2020) 124–127. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/ca.23485, doi:https://doi.org/10.1002/ca.23485.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/ca.23485

[7] R. B. Trelease, Anatomical informatics: Millennial perspectives on a newer frontier, The Anatomical Record 269 (5) (2002) 224–235.

[8] N. Hoyek, C. Collet, F. D. Rienzo, M. D. Almeida, A. Guillot, Effectiveness of three-dimensional digital animation in teaching human anatomy in an authentic classroom context, Anatomical Sciences Education 7 (6) (2014) 430–437.

[9] J. Pereira, A. Meri, C. Masdeu, M. Molina-Tomas, A. Martinez-Carrio, Using videoclips to improve theoretical anatomy teaching, Eur. J. Anat. 8 (2004) 143–146.

[10] D. P. Tim Vernon, The benefits of 3d modelling and animation in medical teaching, Journal of Audiovisual Media in Medicine 25 (4) (2002) 142–148.

[11] U. Uruthiralingam, P. Rea, Augmented and virtual reality in anatomical education - a systematic review, Adv Exp Med Biol 1235 (2020) 89–101. doi:10.1007/978-3-030-37639-0_5.

[12] A. Hulme, G. Strkalj, Videos in anatomy education: History, present usage and future prospects, International Journal of Morphology 35 (4) (2017) 1540–1546, copyright the Author(s) 2017. Version archived for private and non-commercial use with the permission of the author/s and according to publisher conditions. For further rights please contact the publisher. doi:10.4067/S0717-95022017000401540.

[13] A. A. Jaffar, Youtube: An emerging tool in anatomy education, Anatomical Sciences Education 5 (3) (2012) 158–164. arXiv:https://anatomypubs.onlinelibrary.wiley.com/doi/pdf/10.1002/ase.1268, doi:https://doi.org/10.1002/ase.1268.
URL https://anatomypubs.onlinelibrary.wiley.com/doi/abs/10.1002/ase.1268

[14] K. Perlin, Toward interactive narrative, in: Proceedings of the Third International Conference on Virtual Storytelling: Using Virtual Reality Technologies for Storytelling, ICVS'05, 2005, pp. 135–147.

[15] R. Ronfard, Film Directing for Computer Games and Animation, Computer Graphics Forum 40 (2) (2021) 713–730, eurographics State of the Art Report (STAR). doi:10.1111/cgf.142663.
URL https://hal.inria.fr/hal-03225328

[16] O. Akerberg, H. Svensson, B. Schulz, P. Nugues, Carsim: An automatic 3d text-to-scene conversion system applied to road accident reports, in: Proceedings of the Tenth Conference on European Chapter of the Association for Computational Linguistics - Volume 2, EACL '03, 2003, pp. 191–194.

[17] M. O'Kane, J. Carthy, M. Bertolotto, Text-to-scene conversion for accident visualization, in: ACM SIGGRAPH 2004 Posters, SIGGRAPH '04, 2004.

[18] L. M. Seversky, L. Yin, Real-time automatic 3d scene generation from natural language voice and text descriptions, in: Proceedings of the 14th ACM International Conference on Multimedia, MM '06, 2006, pp. 61–64.

[19] P. Ye, T. Baldwin, Towards automatic animated storyboarding, in: Proceedings of the 23rd National Conference on Artificial Intelligence - Volume 1, AAAI'08, 2008, pp. 578–583.

[20] M. Marti, J. Vieli, W. Witoń, R. Sanghrajka, D. Inversini, D. Wotruba, I. Simo, S. Schriber, M. Kapadia, M. Gross, Cardinal: Computer assisted authoring of movie scripts, in: 23rd International Conference on Intelligent User Interfaces, IUI '18, 2018, pp. 509–519.

[21] O. Palombi, F. Ulliana, V. Favier, J.-C. Léon, M.-C. Rousset, My corporis fabrica: an ontology-based tool for reasoning and querying on complex anatomical models, Journal of Biomedical Semantics 5 (1) (2014) 20.

[22] Q. Galvane, R. Ronfard, M. Christie, N. Szilas, Narrative-Driven Camera Control for Cinematic Replay of Computer Games, in: MIG'14 - 7th International Conference on Motion in Games , ACM, Los Angeles, United States, 2014, pp. 109–117. doi:10.1145/2668064.2668104.
URL https://hal.inria.fr/hal-01067016

[23] L.-w. He, M. F. Cohen, D. H. Salesin, The virtual cinematographer: A paradigm for automatic real-time camera control and directing, in: Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '96, 1996, pp. 217–224.

[24] R. Ronfard, V. Gandhi, L. Boiron, V. A. Murukutla, The prose storyboard language: A tool for annotating and directing movies (2020). arXiv:1508.07593.

[25] O. Palombi, G. Bousquet, D. Jospin, S. Hassan, L. Reveret, F. Faure, My Corporis Fabrica: a Unified Ontological, Geometrical and Mechanical View of Human Anatomy, in: 3DPH2009 - 2nd Workshop on 3D Physiological Humal, Vol. 5903 of Lecture Notes in Computer Science, Springer, 2009, pp. 209–219.

[26] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, J. Silovsky, G. Stemmer, K. Vesely, The kaldi speech recognition toolkit, in: IEEE 2011 Workshop on Automatic Speech Recognition and Understanding, IEEE Signal Processing Society, 2011, iEEE Catalog No.: CFP11SRW-USB.

[27] S. G. Hart, L. E. Staveland, Development of nasa-tlx (task load index): Results of empirical and theoretical research, in: P. A. Hancock, N. Meshkati (Eds.), Human Mental Workload, Vol. 52 of Advances in Psychology, North-Holland, 1988, pp. 139–183. doi:https://doi.org/10.1016/S0166-4115(08)62386-9.

URL https://www.sciencedirect.com/science/article/pii/S0166411508623869

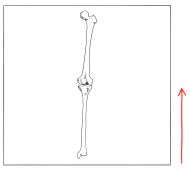
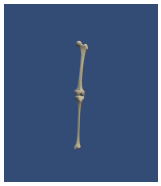
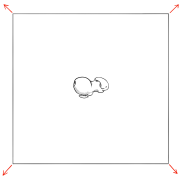
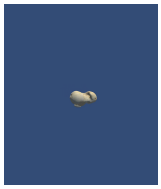
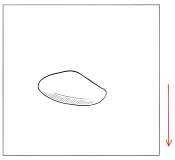
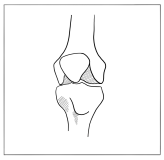
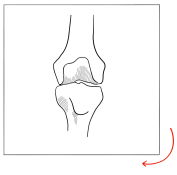
Narration	ASL script	Storyboards	Video screenshots
L'articulation du genou met en rapport trois os: l'épiphyse distale du fémur, l'épiphyse proximale du tibia et enfin la patella	cut to FS Left Femur Tibia Patella ventral		
En réalité le genou comporte trois articulations distinctes	then arc up to FS transverse Left Femur Tibia Patella ventral		
qui sont: l'articulation fémoro-patellaire entre	then dolly in to CU transverse Left Femur Tibia Patella ventral		
la trochlée fémorale et la surface cartilagineuse de la patella	then arc down to CU Left Femur Tibia Patella ventral		
et les deux articulation	then as Patella disappears continue to CU Left Femur Tibia ventral		

Figure 10: Storyboard for lesson on articulation - Part 1

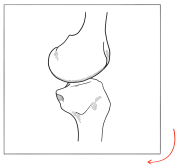
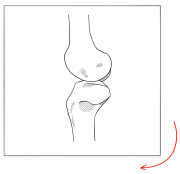
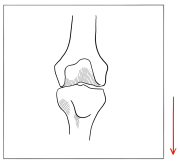
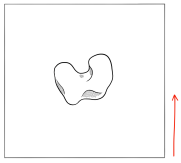
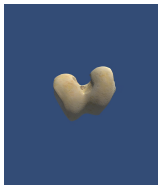
Narration	ASL script	Storyboards	Video screenshots
fémoro-tibiales médiale	then arc clockwise to CU Left Femur Tibia medial		
et latérale	then arc clockwise to CU Left Femur Tibia lateral		
Sur cette vue ventrale d'un genou droit nous mettons en place les deux condyles fémoraux qui sont recouverts de cartilage et qui s'articulent avec les surfaces articulaires de l'épiphyse tibiale proximale pour former les articulations fémoro-tibiales médiale et latérale	then arc clockwise to CU Left Femur Tibia ventral		
Au centre, entre les deux condyles fémoraux il existe une dépression, non recouverte de cartilage appelée fosse ou échancrure inter-condylienne. Au sein de cette fosse se trouvent les deux ligaments croisés du genou.	then as Tibia disappears arc down to CU transverse Left distal Femur ventral		
Ventralement il existe une autre zone recouverte de cartilage: il s'agit de la trochlée fémorale qui va s'articuler avec la surface articulaire de la patella.	then arc up to CU Left Femur ventral		

Figure 11: Storyboard for lesson on articulation - Part 2

```
Shot              = Transition? _ Composition? _ (Development)*
Development       = Cue? _ Recomposition?
Recomposition     = Reframe? _ Composition? _ Cue?
Reframe           = Rename/CameraMov
Cue               = After / As / Then
After             = "then"? _ "after" _ Time? _ Event?
As                = "then"? _ "as" _ Event _ ("and"? _ Event)*
Then              = "then" _ Event? _ ("and"? _ Event)*
Composition       = Angle? _ Figure _ (Angle? _ Figure)*
Figure            = Size? _ Plane? _ AnatSpecs _ Subject _ Profile _ Screen?
Subject           = ((AnatomicalPart _ ("with" _ AnatomicalPart)?) / Region)*
Rename            = ("keep" / "continue") _ "to"
CameraMov         = Camera _ ("to" / "with")?
Camera            = Speed? _ ( Dolly / Crane / Zoom / Arc)
Dolly             = "dolly" _ ("in" / "out")?
Crane             = "crane"  _ ("up" / "down") _ ("left"/"right")?
Arc               = "arc"  _ ( "up" / "down")? _ ("clockwise" /"anticlockwise")?
Zoom              = "zoom" _ ("in" / "out")
Speed             = "slow" / "quick" / ("following" Subject)
Transition        = ("cut to" / "dissolve to" / "fade in to")
Cross             = "crosses" _ ("over" / "under") _ Subject _ (Subject)?
Flex              = "flexes"_("partially"/"completely")
Rotate            = "rotates"_("internally"/"externally")
Angle             = "low angle" / "high angle"
Size              = "ECU" / "CU" / "MCU" / "MS" / "MLS" / "FS" / "LS" / "ELS"
Profile           = "ventral" / "dorsal" / "medial" / "lateral" / "ventromedial"
                    / "dorsomedial" / "ventrolateral" / "dorsolateral"
AnatSpecs         = ("Left"/"Right") _ ("proximal" /"distal")?
Screen            = "screen" _ ("top" / "bottom")? _ ("left" / "center"/ "right")?
Plane              = "frontal"/ "transverse" / "sagittal"
Time              = ~r"[0-9]*"i _ ("seconds" / "s" )
String            = ~r"[A-Z 0-9]*"i
Action            = Cross / Flex /  Rotate
Effect            = "appears" / "disappears"
Event             = Subject _ (Action / Effect)
AnatomicalPart    = "Patellar ligament" / "Femur" / "Patella" / "Medial meniscus"
                     / "Fibula" / "Hip bone" /"Tibia"
                     /"Lateral meniscus" /"Articular capsule of left knee joint"
                     / "Posterior cruciate ligament"
Region            = "Knee"
space             = ~r"\s"
_                 = (space)*
```

Figure 12: ASL Grammar.