Speech animation using electromagnetic articulography as motion capture data

Ingmar Steiner
ingmar.steiner@dfki.de

Korin Richmond
korin@cstr.ed.ac.uk

Slim Ouni
slim.ouni@loria.fr

Overview

Electromagnetic articulography (EMA) captures the position and orientation of a number of markers, attached to the articulators, during speech. As such, it performs the same function for speech that conventional, marker-based motion capture does for full-body movements acquired with optical modalities.

We present an approach to processing EMA data from a motion-capture perspective and applying it to the visualization of an existing multimodal corpus of articulatory data,[a] creating a kinematic 3D model of the articulators by adapting a conventional motion capture based animation paradigm.
Such an animated model can then be easily integrated into multimedia applications as an animation asset, allowing the visualization of speech production in an intuitive and accessible manner.

[a] http://mngu0.org/

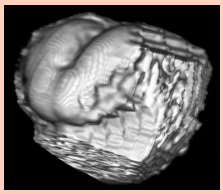
EMA data as motion capture

```
HIERARCHY
ROOT T3
{
  OFFSET   −0.08384774 −0.68610632 0.72265321
  CHANNELS 6 XPosition YPosition ZPosition XRotation YRotation ZRotation
  End Site
  {
    OFFSET   0.00000000   −1.00000000 0.00000000
  }
}
ROOT upperlip
{
  OFFSET   −0.94566417 0.[illegible]  −0.13029577
  CHANNELS 6 XPosition Y[illegible]on ZPosition XRotation YRotation [illegible]tation
  End Site
  {
    OFFSET   [illegible]0000   −1.00000000 0.00000
  }
}
MOTION
Frames: 724
Frame Time: 0.005
6.78959465  12.2727[illegible] 1.20037365  −8.13653286 56.68709013 −1.78241577 ...
        −6.08323193 [illegible]89048 0[illegible]787  19.70638666 −11.45516554 −52.56656678 ...
        0.17483102  [illegible]558  −[illegible]645 3.93864096  24.40054857 −51.69048915 ...
        0.05262496  −[illegible]40909 −2[illegible]731 56.49315374 9.14393027  −2.77820826 ...
        0.00545616  −0.00614901 5.71[illegible]687  56.73992164 1.99134613  7.70857967 ...
        −0.00336362 −0.02571771 0.00576410  −57.27868696  −1.34396448 0.38998995 ...
        0.25838542  1.55094087  −0.98620248 7.96369151  −30.90922537 −47.58157029 ...
        0.10555566  3.21496391  0.25413591  −8.54249967 −56.29318629 −6.39601923 ...
        0.11248340  5.05562544  −0.04919426 −4.80412156 −39.31099667 41.40497899 ...
        0.09581324  −1.06100082 −0.19431658 −54.18256563  17.06821414 −7.46539762
```

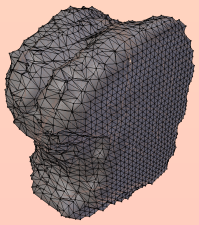
Fragment of one EMA sweep from `mngu0` database in Biovision Hierarchy format. EMA coils are rendered over a 250 ms window (frame step = 3); lips, incisors at left, tongue coils 1 to 3 at right.

Articulatory model

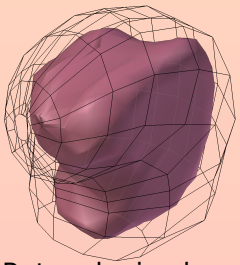
The tongue is segmented from a volumetric magnetic resonance imaging (MRI) scan and retopologized into a mesh. The tongue mesh can be deformed using spline inverse kinematics (IK); the spline's control points are modified by the EMA coils. Dental scans are registered into the same space and added to the rig.



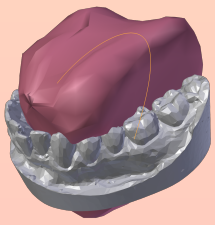
Volume rendering of raw MRI data



Voxel-tesselated iso-surface from MRI

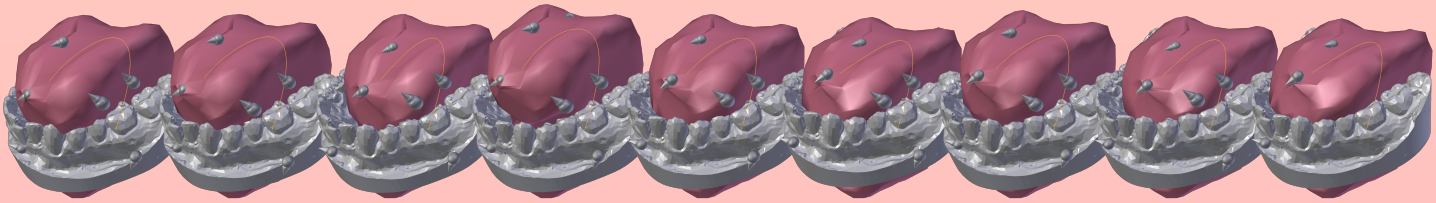


Retopologized tongue mesh



Tongue model rigged for spline IK, with mandible

Animation



For speech animation, the EMA data drives the animation rig. The `ref` and `jaw` coils control the maxilla and mandible, respectively; the tongue coils move the IK control points, which in turn deform the tongue mesh.