

# Using multimodal speech production data to evaluate articulatory animation for audiovisual speech synthesis

Ingmar Steiner

Korin Richmond

Slim Ouni



University College Dublin  
& Trinity College Dublin



CSTR  
University of Edinburgh



Loria



UNIVERSITÉ  
DE LORRAINE

LORIA  
Université de Lorraine

Vienna, September 21, 2012

# Motivation

Data-driven animation for speech articulators

# Motivation

Data-driven animation for speech articulators *within* the vocal tract

# The mngu0 Corpus

## Multimodal speech corpus

- one male speaker of British English
- electromagnetic articulography (EMA)
- magnetic resonance imaging (MRI)
- dental cast scans

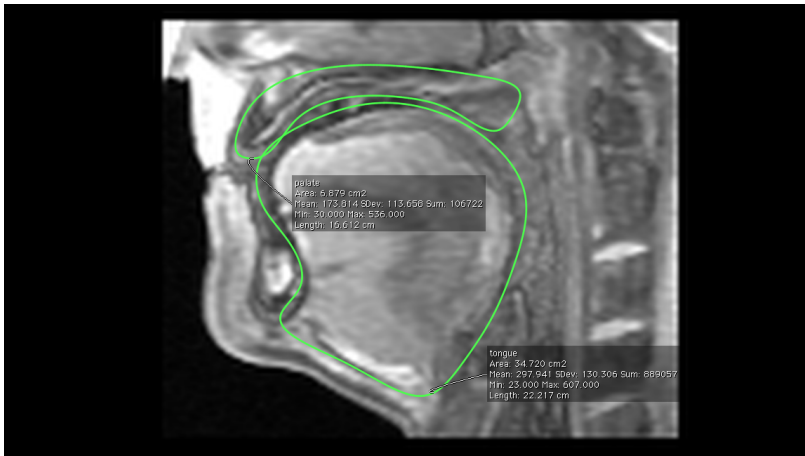


<http://mngu0.org/>

# Electromagnetic articulography

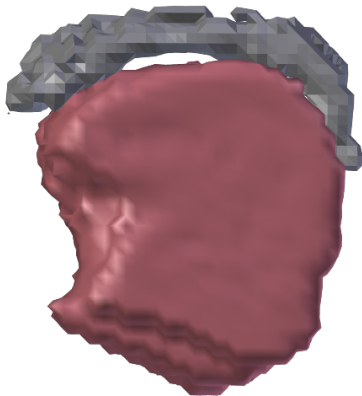


## MRI



manual regions of interest (ROIs)

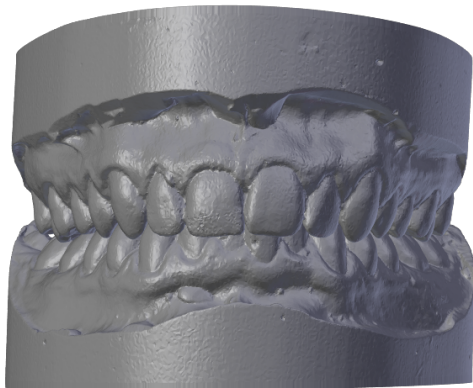
# MRI



isosurfaces within ROIs

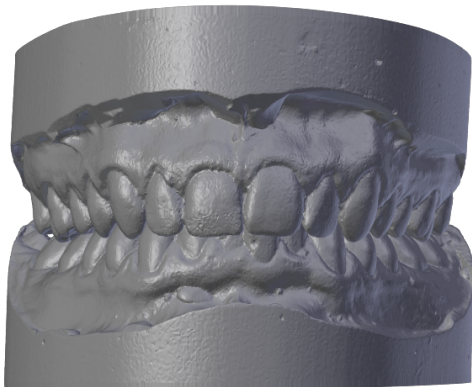


## Dental scans



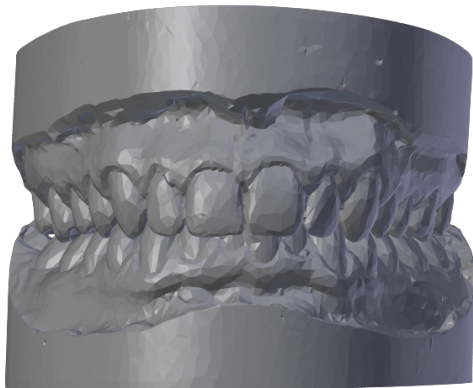
vertex count 927 282 (maxilla), 836 892 (mandible)

## Dental scans



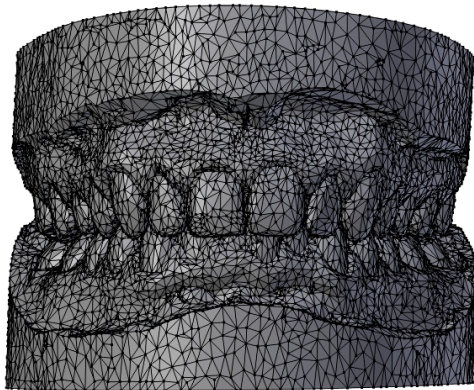
deduplication: vertex count 154 549 (maxilla), 139 484 (mandible)

# Dental scans



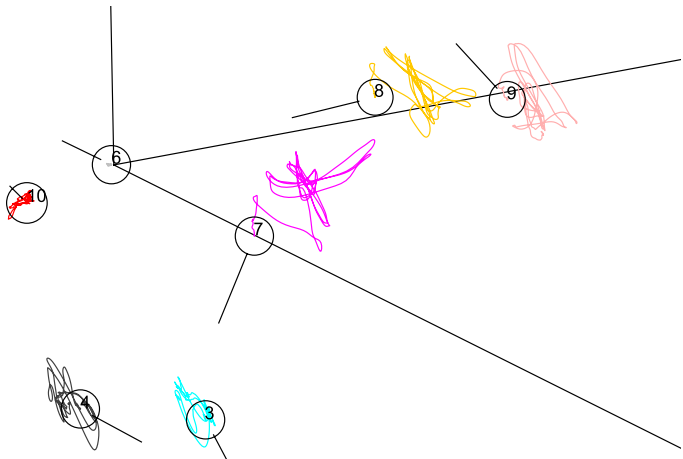
decimate (5%)

# Dental scans

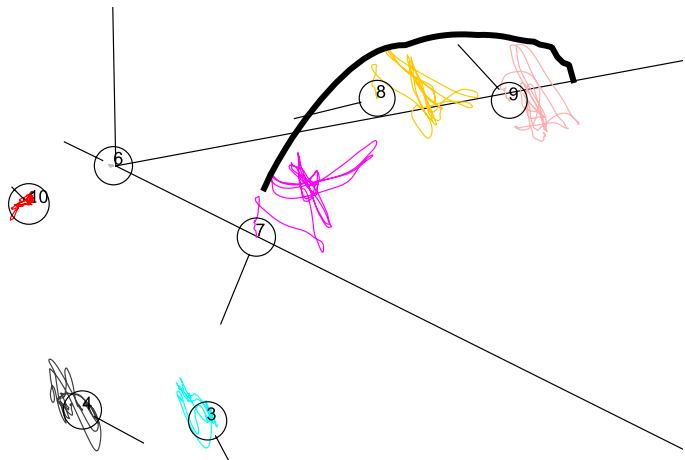


vertex count 7729 (maxilla), 6976 (mandible)

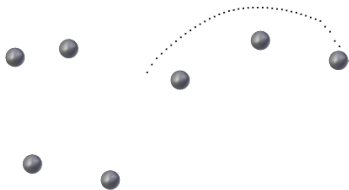
# Palate contour



# Palate contour

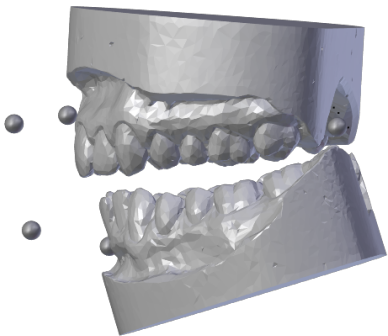


# Model rigging



EMA motion capture data

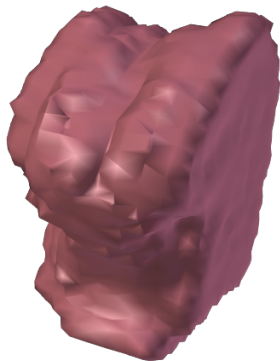
# Model rigging



maxilla/mandible track ref/jaw coils

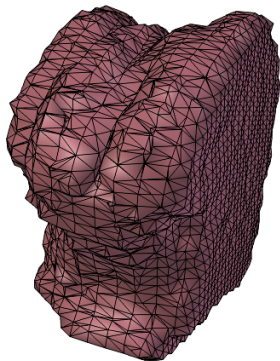


# Tongue mesh retopology



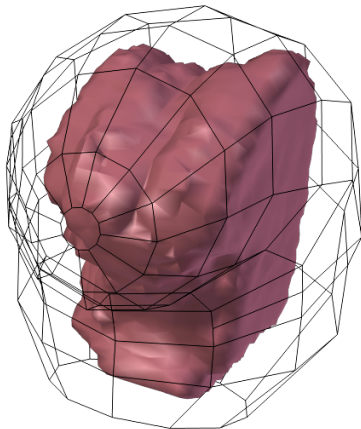
crude isosurface

# Tongue mesh retopology



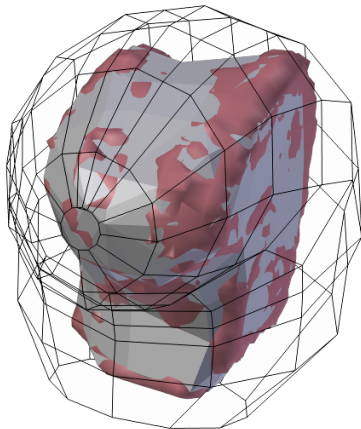
tesselation from MRI voxels

# Tongue mesh retopology



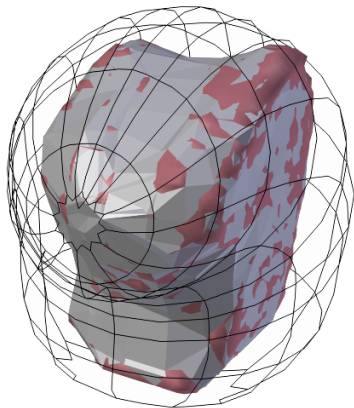
simple cage

## Tongue mesh retopology



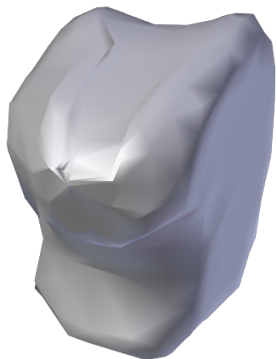
“shrinkwrapped” to isosurface

# Tongue mesh retopology



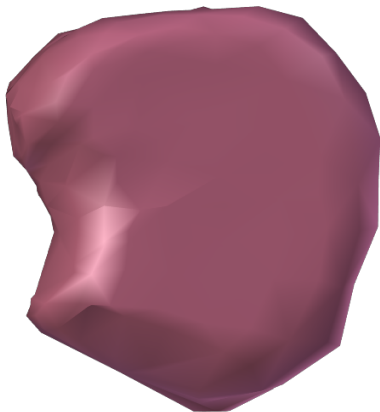
Catmull-Clark subdivision

## Tongue mesh retopology



smooth, tongue-shaped mesh with simple topology

# Tongue rigging

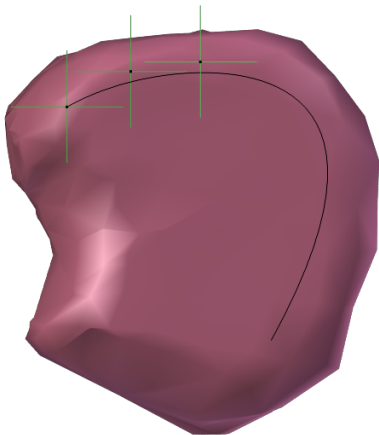


static mesh



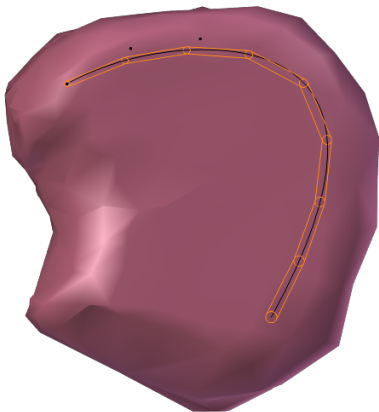


# Tongue rigging



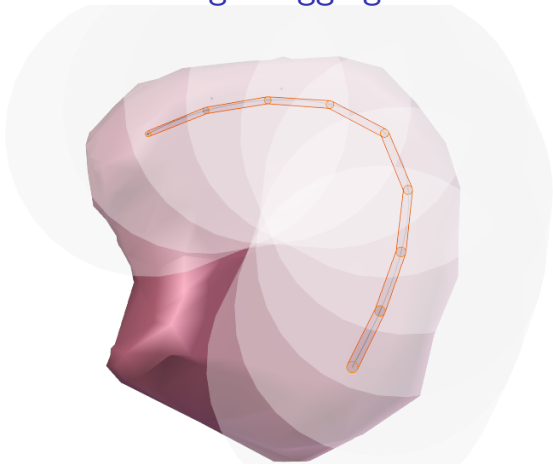
modified by hooks tracking tongue coils

## Tongue rigging



armature follows spline through inverse kinematics (IK)

# Tongue rigging



tongue mesh deformed by armature

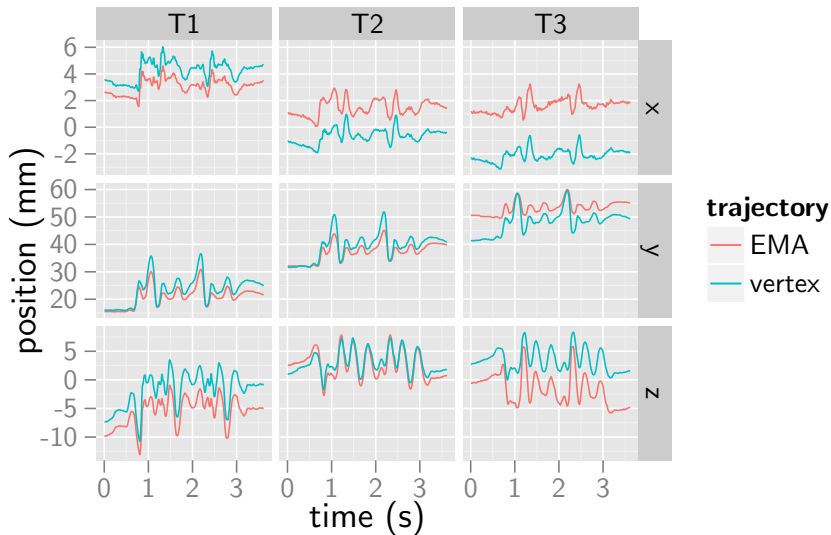
# Animation

# Animation

# Animation

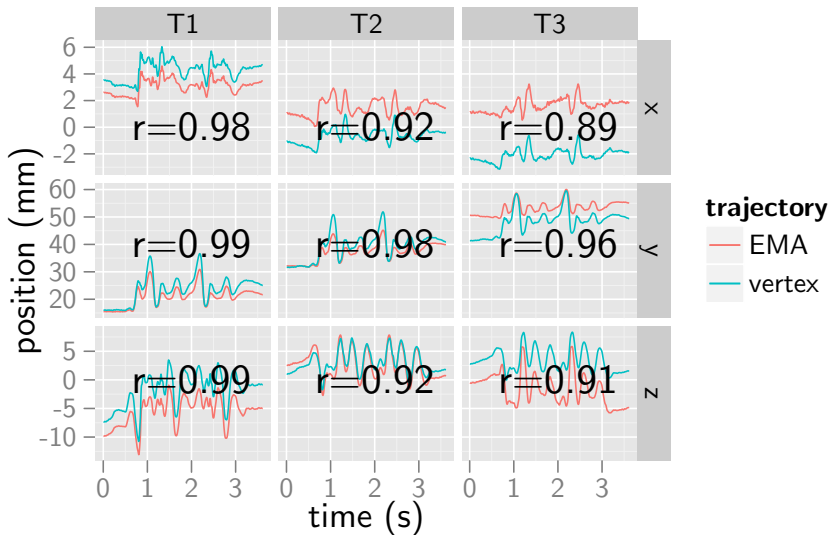
# Animation

# Vertex tracking





## Vertex tracking



# Conclusion

Skeletal animation of articulatory movements from speech production data seems promising, but depends on

- model topology
- data quality
- registration (incl. posture effect)