EXPRESSION
Expressiveness in Human Centered Data/Media

Université Bretagne Sud
ubs

Coarticulation Analysis for Sign Language Synthesis

Lucie NAERT, Caroline LARBOULETTE and Sylvie GIBET

July 12th, 2017

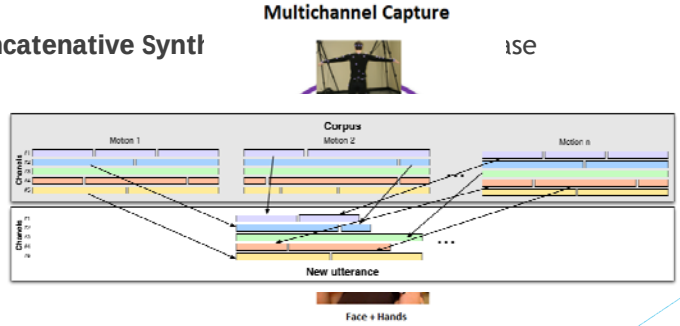
lucie.naert@univ-ubs.fr

UMR IRISA

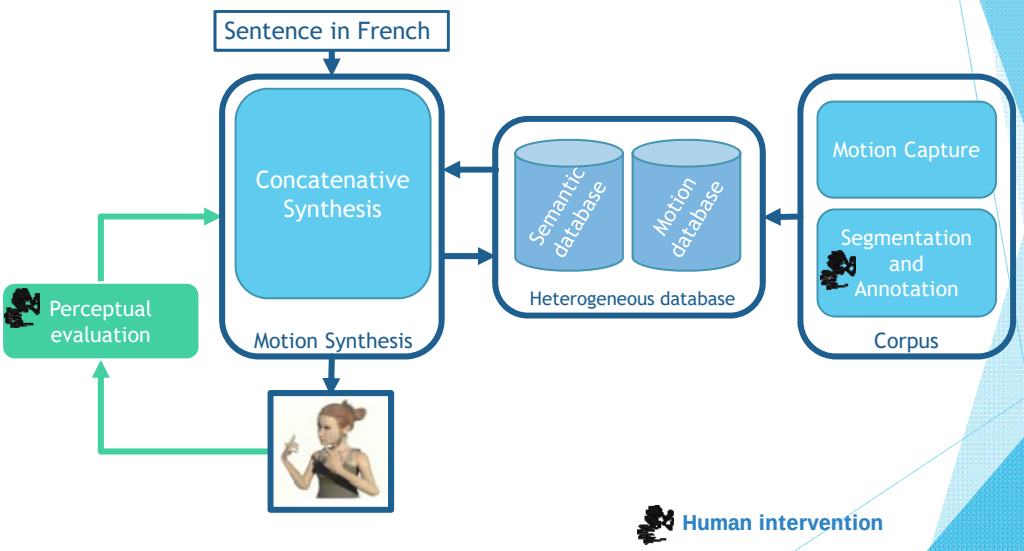
Motivation

Build French Sign Language (LSF) sentences from pre-recorded gestures

- **MoCap data** to improve the naturalness of the animations
- **Concatenative Synth** ıse



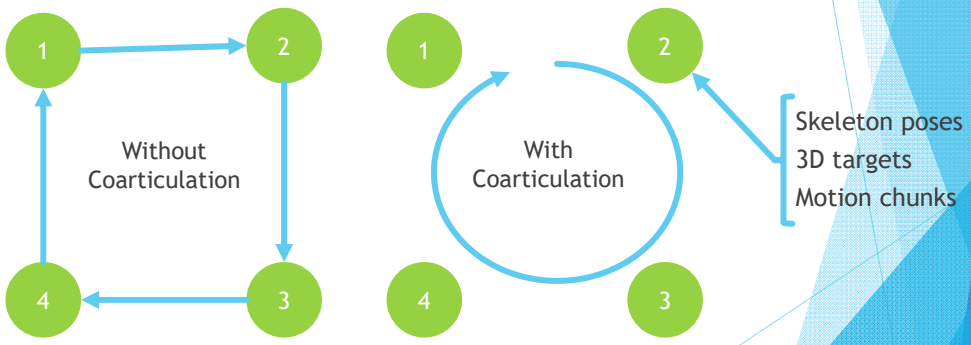
Our Synthesis System



Coarticulation & Transitions

Coarticulation

- Articulatory influence of one "phonetic" item on another *[Grosvald, 2010]*



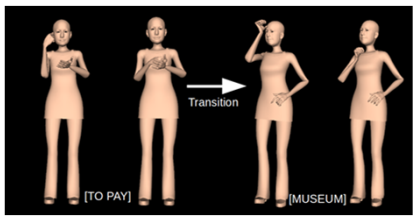
Coarticulation & Transitions

Coarticulation

- Articulatory influence of one "phonetic" item on another *[Grosvald, 2010]*

Transition

- Intermediary movement between two consecutive signs



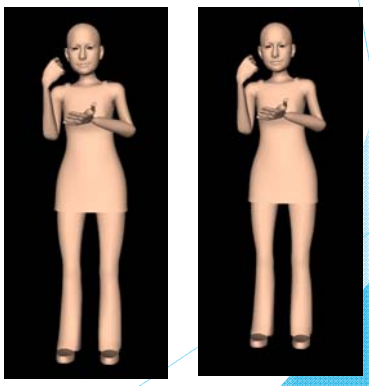
Transitions are manifestations of inter-signs coarticulation

Challenges of Coarticulation

- The animation depends on
  - the quality of the original segmentation
  - the context: some transitions or sign forms do not exist
- 3 recurrent issues
  - Bad segmentation

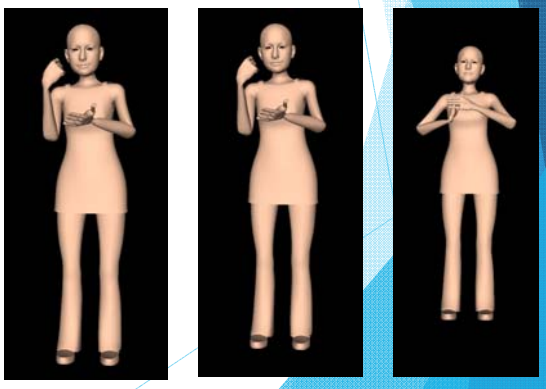
Challenges of Coarticulation

- The animation depends on
  - the quality of the original segmentation
  - the context: some transitions or sign forms do not exist
- 3 recurrent issues
  - Bad segmentation
  - Kinematic problems

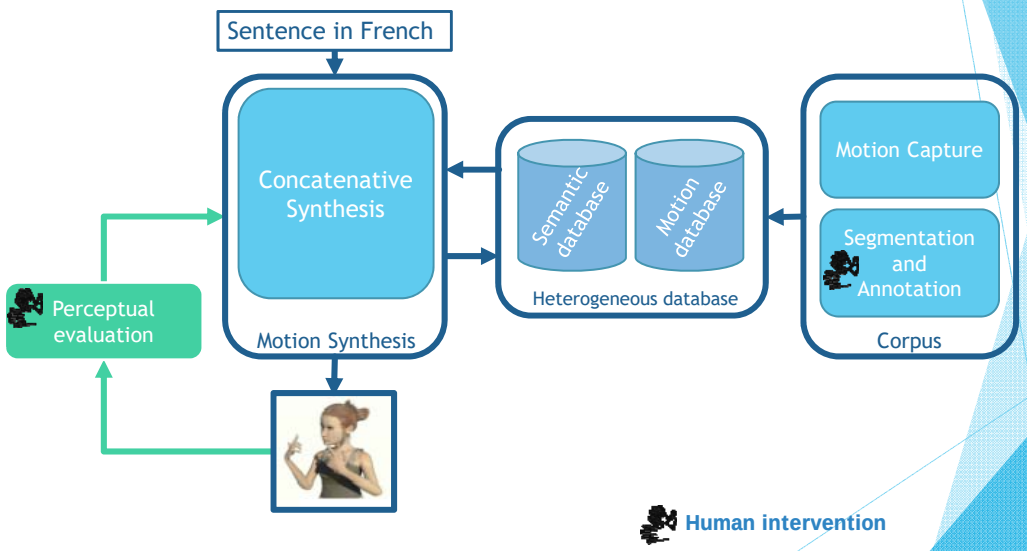


Challenges of Coarticulation

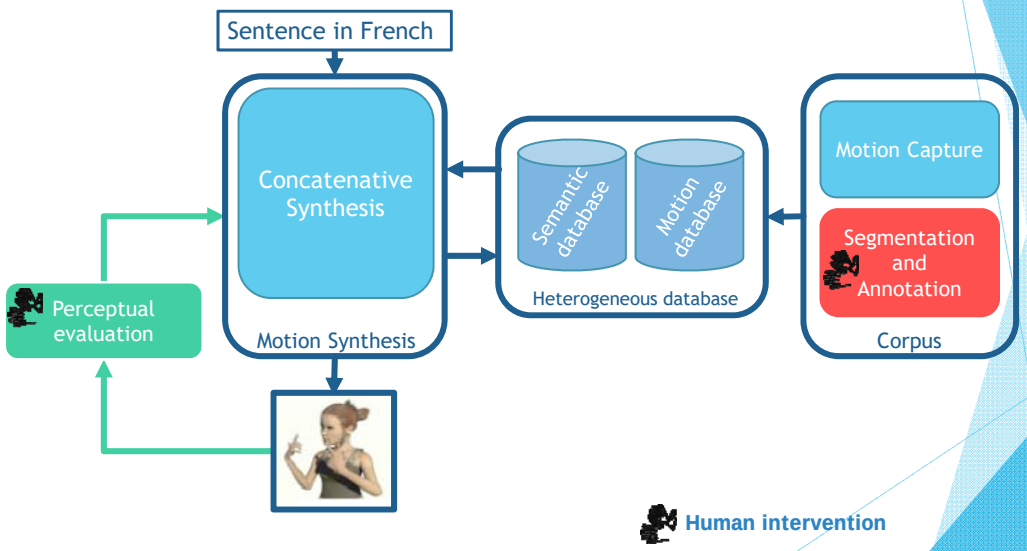
- The animation depends on
  - the quality of the original segmentation
  - the context: some transitions or sign forms do not exist
- 3 recurrent issues
  - Bad segmentation
  - Kinematic problems
  - Artifacts due to an unsuitable context



Overview



Segmentation

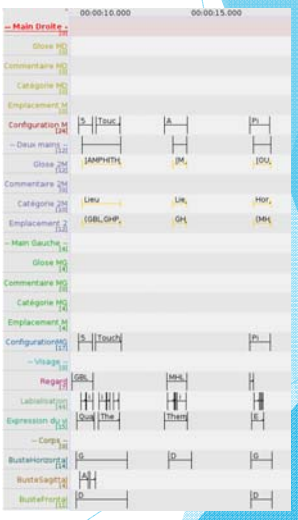


A Need for an Accurate Segmentation

- To avoid artifacts in the reconstruction/synthesis
- To analyze correct transition data

Manual Segmentation

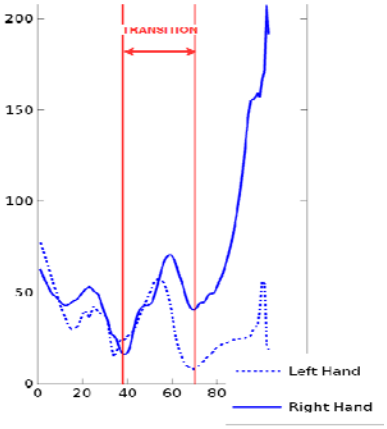
- Tedious and time-consuming
- Sometimes inaccurate
- Subject to variability depending on the annotator and on its criteria *[Hanke, 2012]*



Manual annotation using the ELAN software *[Lefebvre-Albaret, 2013]*

Observations on the Ground Truth

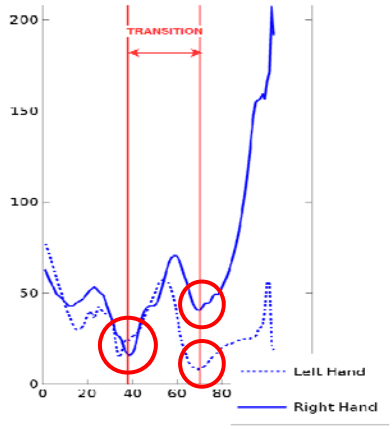
Speed Norm – Motions [TO ENTER] & [TO PAY]



- Extraction of various sequences of two consecutive signs from the manually annotated database
- Computation of kinematics features

Observations on the Ground Truth

Speed Norm – Motions [TO ENTER] & [TO PAY]



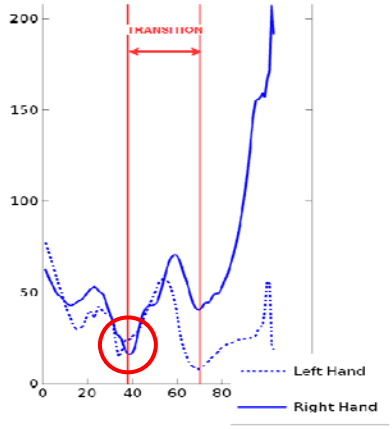
Observation #1

Local minima seem to delimit signs

Observations on the Ground Truth

Introduction
Segmentation
Transition Duration
Analysis
Synthesis
Conclusion

Speed Norm – Motions [TO ENTER] & [TO PAY]



Observation #1

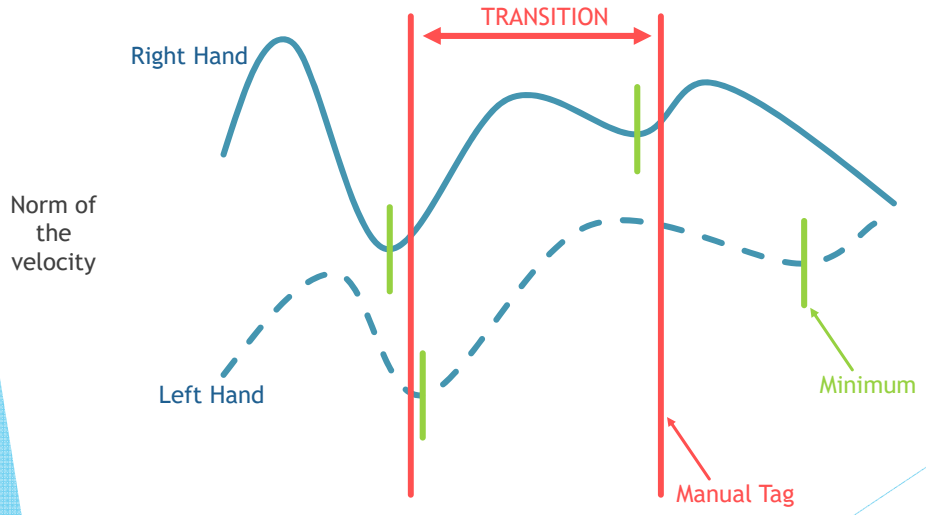
Local minima seem to delimit signs

Observation #2

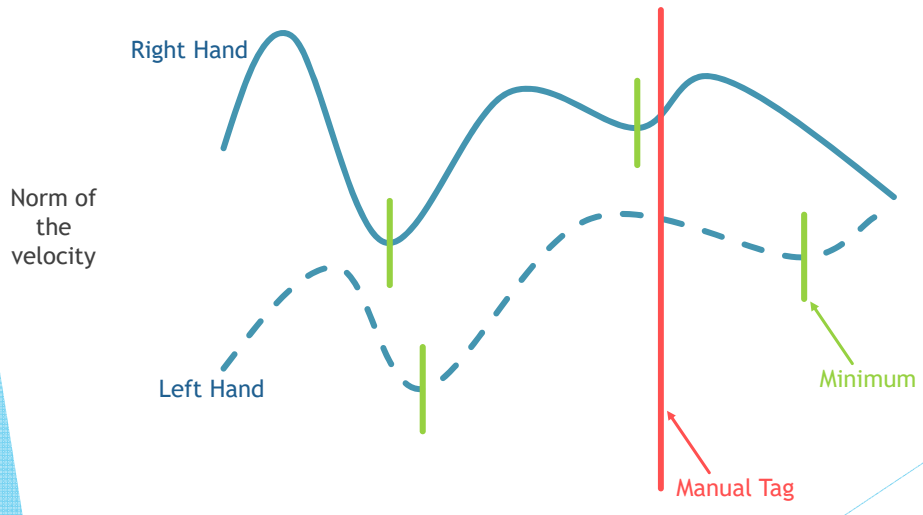
The two hands are not perfectly synchronized

Toward a segmentation per hand at a glose level?

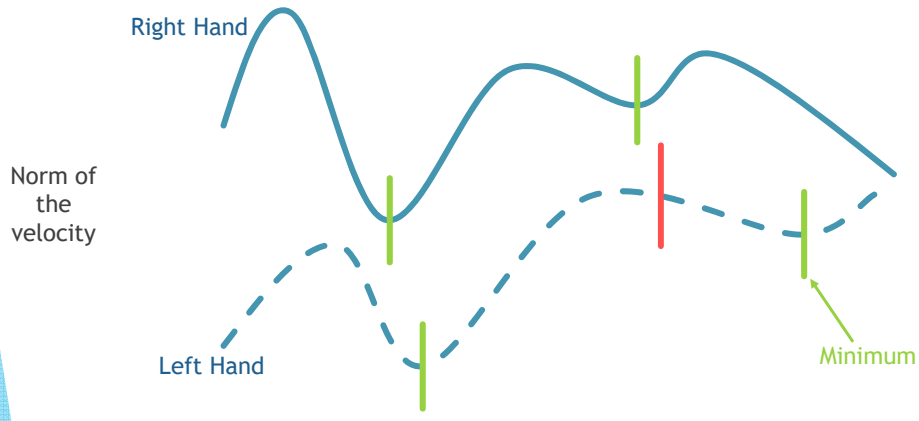
Refinement of the Segmentation



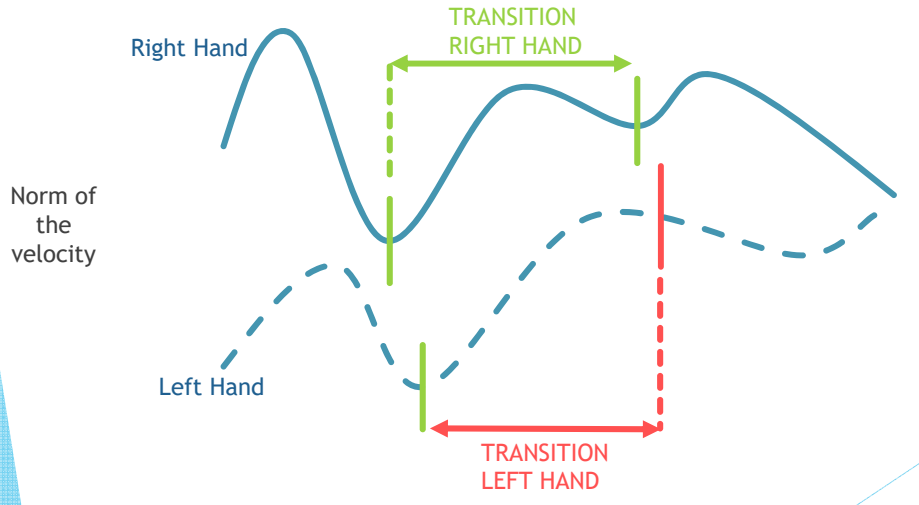
Refinement of the Segmentation



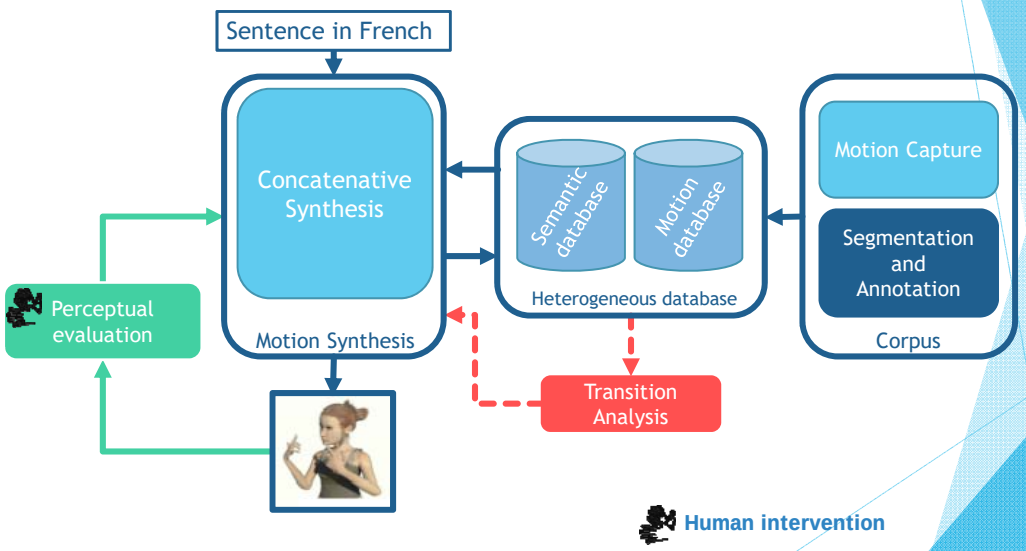
Refinement of the Segmentation



Refinement of the Segmentation

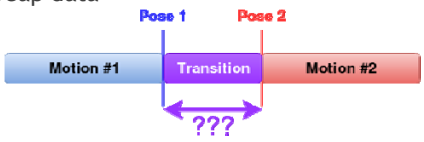


Transition Analysis

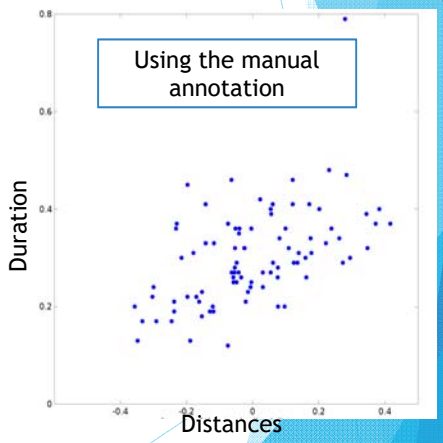


Analysis of Transition Duration

- **89** transitions extracted from **2** sequences of MoCap data



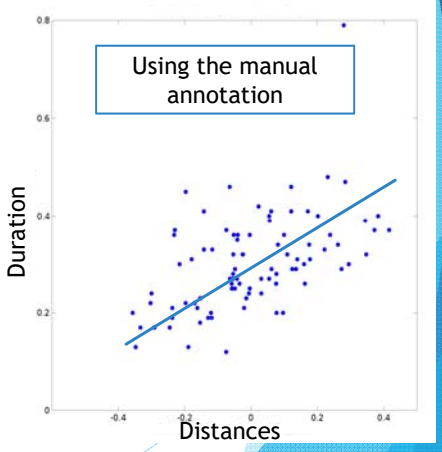
- Duration as a function of the distance between Pose 1 and Pose 2
  - Isochrony? $Velocity = k.Distance$
  - Fitts's law? $Duration = a + b.\log(\frac{Distance\ from\ Target}{Radius\ of\ Target})$



Analysis of Transition Duration

Observations

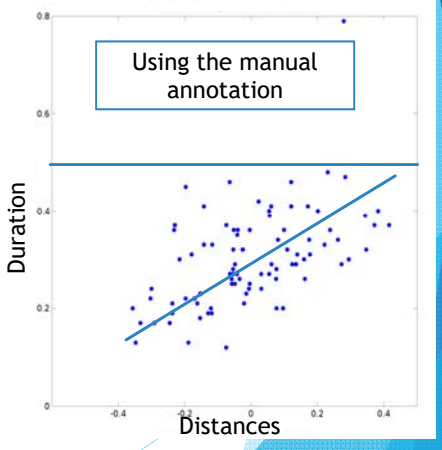
- General tendency to increase with the distance



Analysis of Transition Duration

Observations

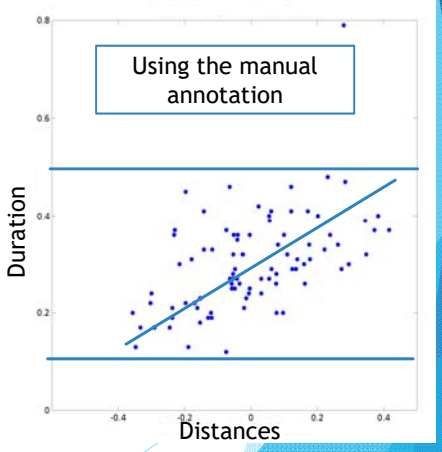
- General tendency to increase with the distance
- Duration ≤ 0.5s



Analysis of Transition Duration

Observations

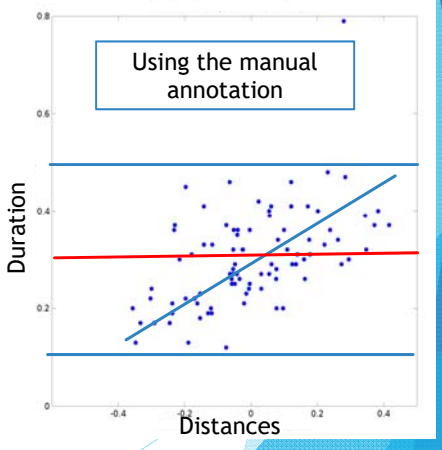
- General tendency to increase with the distance
- Duration ≤ 0.5s
- Duration ≥ 0.1s



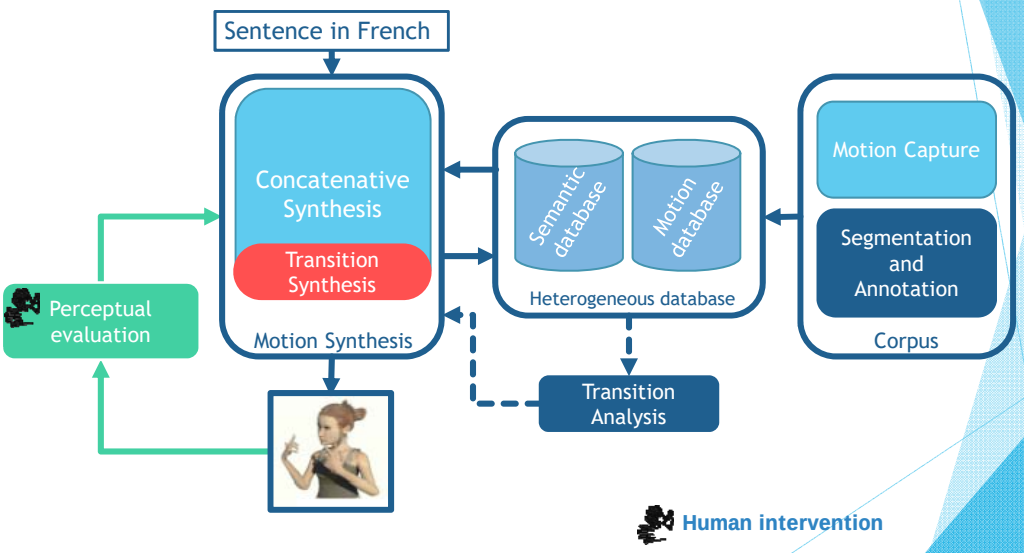
Analysis of Transition Duration

Observations

- General tendency to increase with the distance
- Duration ≤ 0.5s
- Duration ≥ 0.1s
- Mean = 0,303s & SD = 0,098s

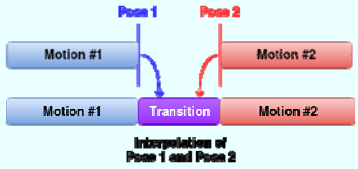


Transition Synthesis

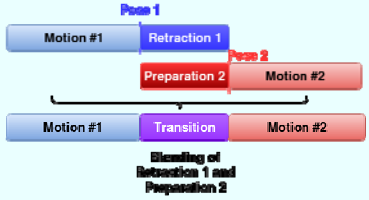


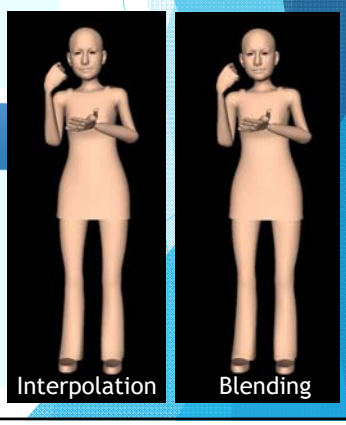
Transition Synthesis Techniques

Pose Interpolation



Motion Blending





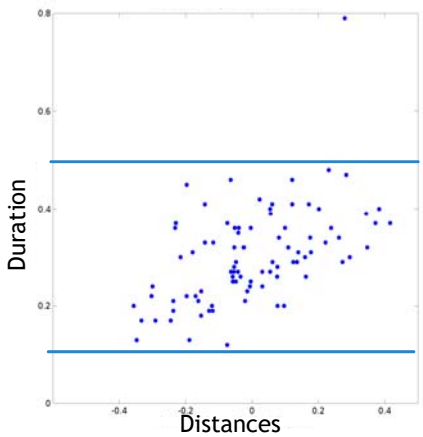
Computation of Transition Duration

- SIMPLE

$$Duration = \alpha * \frac{2*EuclideanDistance(Pose1,Pose2)}{MeanVel_{M1}+MeanVel_{M2}} + (1-\alpha) * \frac{2*GeodesicDistance(Pose1,Pose2)}{MeanAngVel_{M1}+MeanAngVel_{M2}}$$

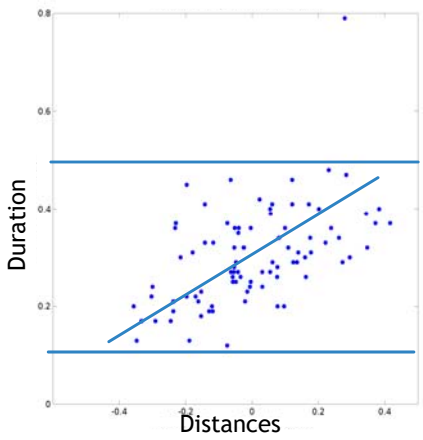
Computation of Transition Duration

- SIMPLE
- SIMPLE and BOUNDED



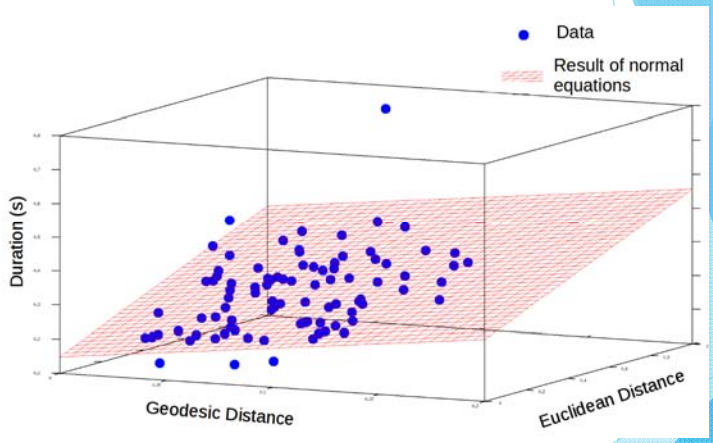
Computation of Transition Duration

- SIMPLE
- SIMPLE and BOUNDED
- LINEAR



Computation of Transition Duration

- SIMPLE
- SIMPLE and BOUNDED
- LINEAR
- SURFACE

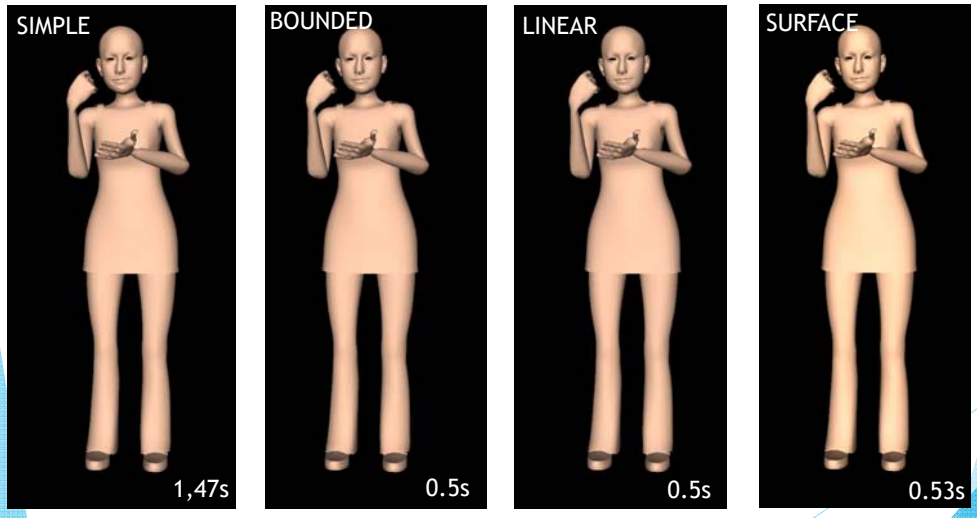


- The 3 methods have to be **perceptually evaluated**

$$\theta = (X^T * X)^{-1} * X^T * y$$

Computation of Transition Duration

Visual Comparison



Results

- Automatic refinement of the manual segmentation
  - By using local minima in the norm of the velocity of the two wrists
  - By processing the two hands separately
- Improvement of the rendering of the animation by modifying the computation of the transition duration
  - Adding lower and higher limits based on our analysis
  - Using basic statistics on the data

For our dataset, the nature of the surrounding signs (symmetry, number of hands) did not impact significantly the duration of the transition.

Future Work

- Analysis on a higher number of transitions
- Elaboration of a new corpus to study more specifically transitions, coarticulation and synchronization between channels (hand configurations)
- Extension of the approach: segmentation/annotation of other channels
- Study of the synchronization of the sign language channels

Thank you !

Questions?