Automatic annotation of head velocity and acceleration in Anvil

Jongejan, Bart

Published in:
Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)

Publication date:
2012

Document version
Peer reviewed version

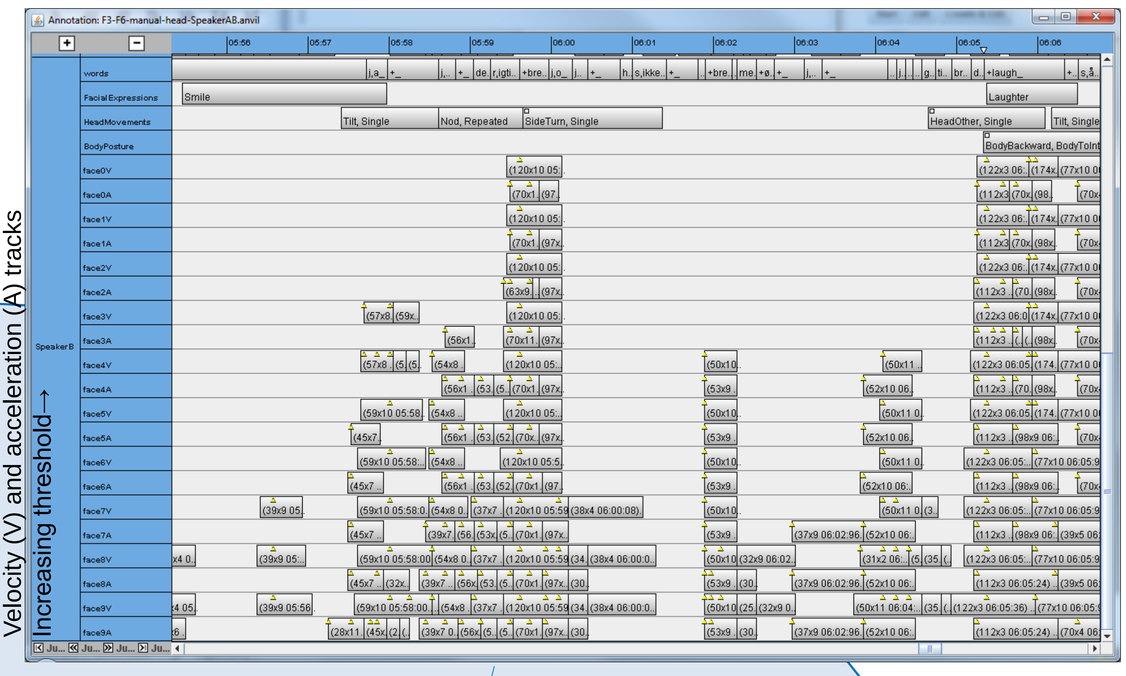
FACULTY OF HUMANITIES

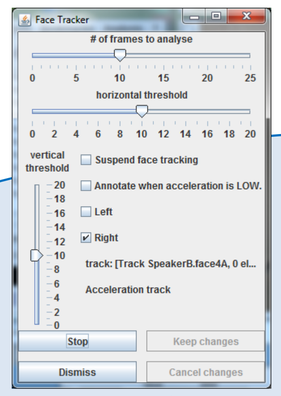
UNIVERSITY OF COPENHAGEN

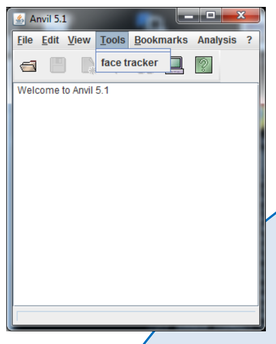
Automatic annotation of face velocity and acceleration in Anvil

Bart Jongejan

We describe an automatic face tracker plugin for the ANVIL annotation tool. The face tracker produces data for velocity and for acceleration in two dimensions. We compare the annotations generated by the face tracking algorithm with independently made manual annotations for head movements. The annotations are a useful supplement to manual annotations and may help human annotators to quickly and reliably determine onset of head movements and to suggest which kind of head movement is taking place.







- ANVIL -
Video
Annotation
Research Tool

OpenCV
Open Source
Computer
Vision

Anvil
Main window

JavaCV

Video window

Face tracker

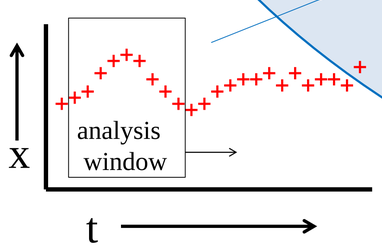
J A V A

C

math

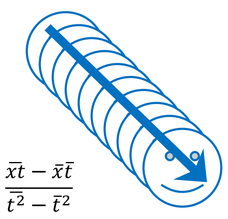
Annotation window

linear regression analysis

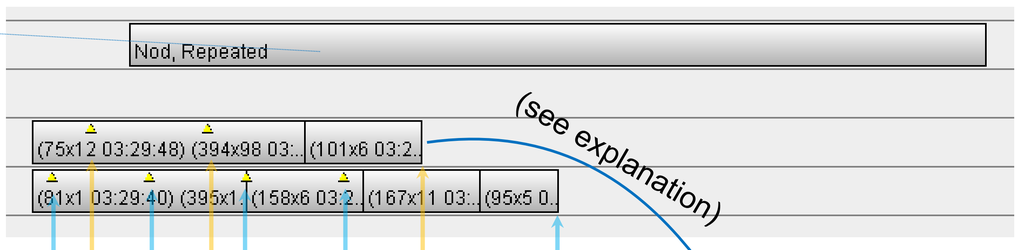


“nod”

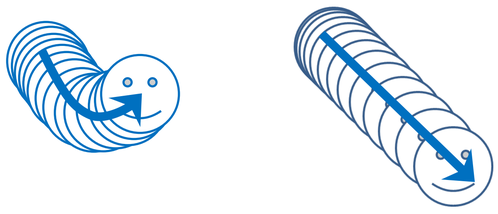
Velocity



$$v_x = \frac{\bar{xt} - \bar{x}\bar{t}}{\bar{t^2} - \bar{t}^2}$$



Acceleration



$$a_x = \frac{\overline{xt^2}\left(\bar{t^2} - \bar{t}^2\right) - \bar{xt}\left(\bar{t^3} - \bar{t}\bar{t^2}\right) + \bar{x}\left(\bar{t}\bar{t^3} - \bar{t^2}^2\right)}{\bar{t^4}\left(\bar{t^2} - \bar{t}^2\right) - \bar{t^3}\left(\bar{t^3} - 2\bar{t}\bar{t^2}\right) - \bar{t^2}^3}$$

time when strength is at maximum
(typically near beginning of annotation)

(75x12 03:29:48)(394x98 03:.. (101x6 03:2..

This marker belongs to this annotation!

(clock) direction when strength is at maximum

maximum strength during annotation

Experiment

5-minute video of a conversation between two people
Manually annotated on beforehand
0.4 s analysis window (10 frames)
Threshold values 5-14 (for both velocity and acceleration)

Analysis

Cohen’s kappa as measure for agreement between human and machine.
Framewise, no coarse graining of time domain. (≈8000 frames)

Results

Good indication of begin and, to lesser degree, end time of head gesture.
All kinds of communicative gestures with the head are detected: *HeadForward*, *(down-)Nod*, *Shake*, *SideTurn*, *Tilt*, *Waggle*, and *HeadOther*

Some types of head movements can be categorized automatically, but with low reliability.
up/down = *Nod* or *HeadForward*
left/right = *Shake* or *SideTurn*

The statistical analysis of a single 5-minute video of a conversation between two people has teached us that there are no threshold values that are optimal for detecting all kinds of head movements.

Future

Combine sequences of automatically detected head movement phases into phrases, aiming at 1:1 correspondence with manual annotations.

Combine velocity and acceleration to detect movements along curved paths.