Monitoring attentional processes for intelligent channelling of educational tasks

Patrick Heyer Wollenberg, Felipe Orihuela-Espina, Javier Herrera Vega, Luis Enrique Sucar

National Institute of Astrophysics, Optics and Electronics (INAOE), Puebla, Mexico

Background

- Iberoamerican countries consistently rank low in the PISA report.

Prueba de evaluación PISA

Posiciones del año 2012 y del 2009

País	Posición 2012	Posición 2009
España	33º =	33º
Chile	51º ↓	44º
México	53º ↓	48º
Uruguay	55º ↓	47º
Costa Rica	56º	N/D
Brasil	58º ↓	53º
Argentina	59º ↓	58º
Colombia	62º ↓	52º
Perú	65º ↓	63º

Fuente: OECD, *PISA 2013*

Infobae

Figure from: [http://peru21.pe/actualidad/informe-pisa-2012-peru-esta-ultimo-lugar-nivel-latinoamerica-2159914]

Background

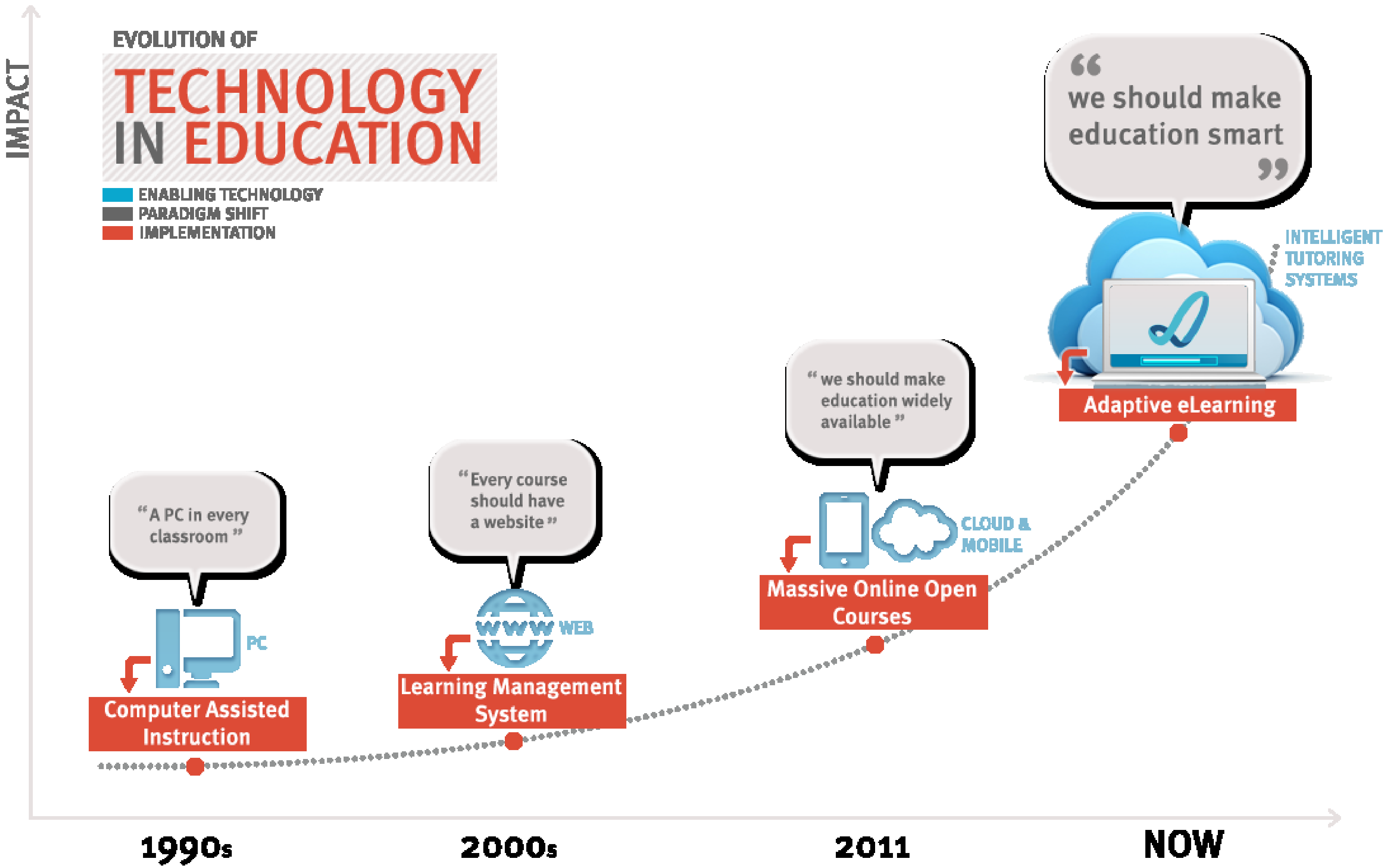


Figure from: [https://www.smartsparrow.com/adaptive-elearning/]

Background

- Intelligent tutor systems (ITS) are educative tools that capitalize on artificial intelligence to support teaching and education [Carbonell 1970].

How does an ITS look like?

From rather naïve…

Educational software for hearing impaired children
Figure from:
http://www.annarenov.fr/ucolmenar/joomla/index.php?option=com_content&task=view&id=7&Itemid=29

Autotutor
Figure from: http://nvate.com

…to extremely sophisticated

ITS for Training US Navy Tactical Action Officers
Figure from:
http://www.stottlerhenke.com/solutions/training/taoits.htm

Background

- To be effective the ITS has to be able to sense the cognitive state of the educand and translate this into task scheduling actions to channel the educational pathway for optimizing the learning process.

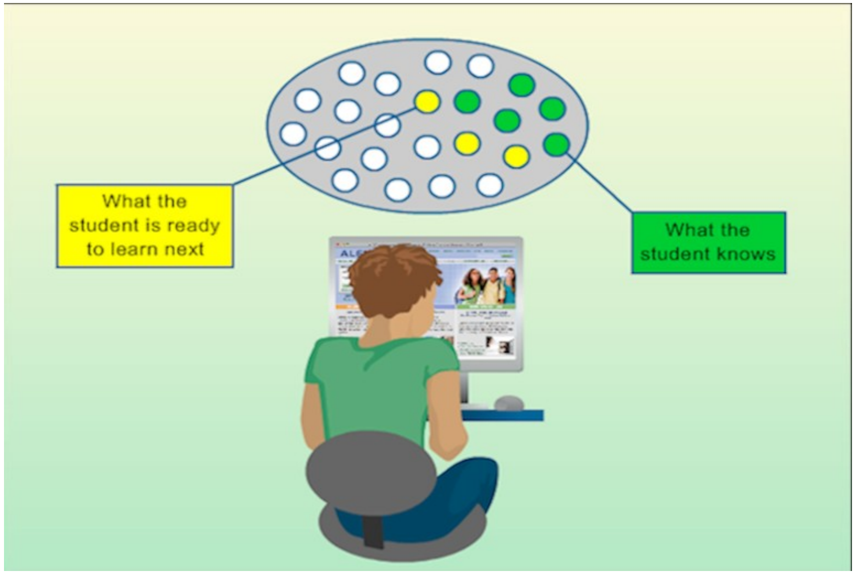


Figure from: [http://inside.at.utep.edu/?p=717]

Hypothesis

- Static body postures can be mined for regulators communicating the attentional and affective state of subjects during normal human communication process [Ekman y Friesen (1969)]

- Posture analysis is a plausible transparent communication channel to enhance human-computer interaction (HCI) [Castellano et al (2007), Kapoor et al (2001), Dmello et al (2008)]

- Hypothesis: *Non-observable attentional state of a system user can be inferred from body posture proxy of observable conduct*.

Aims

- **Aims**:
  - Detection of attention: Map a lexicon of body postures to binarized attentional levels (Experiment I)
  - Attribution of attention: Identify postural features leading to appreciation of attention by third parties (e.g. educators) (Experiment II).

EXPERIMENT I: DETECTION OF ATTENTION

Experimental Aim

- [Assumption] Unacted posture conveys cues about people's attentional disposition.

- [Aim] To **identify robust markers of attention** from posture while people carry out their duties seated in front of their computers
  - Robust here means;
    - High discriminative power
    - Reliable and reproducible across computational models.

- [Impact] This set of features would allow us to infer the attention level of a student interacting with educational software.
  - Human-computer interaction systems can benefit from this knowledge to customize the experience to the user changing attentional state.
  - Adapting the software to the student can in theory improve learning.

Experimental Set-Up

- Body postures were randomly captured from (n=)6 subjects while at work
  - Self evaluated attentional state
  - Bespoken data acquisition software
  - Position from Kinect; 40 postural features
  - Samples acquired randomly every 40 60 mins

- Computational solution
  - 4 well-known classifiers for reproducibility.
  - Supervised model learning

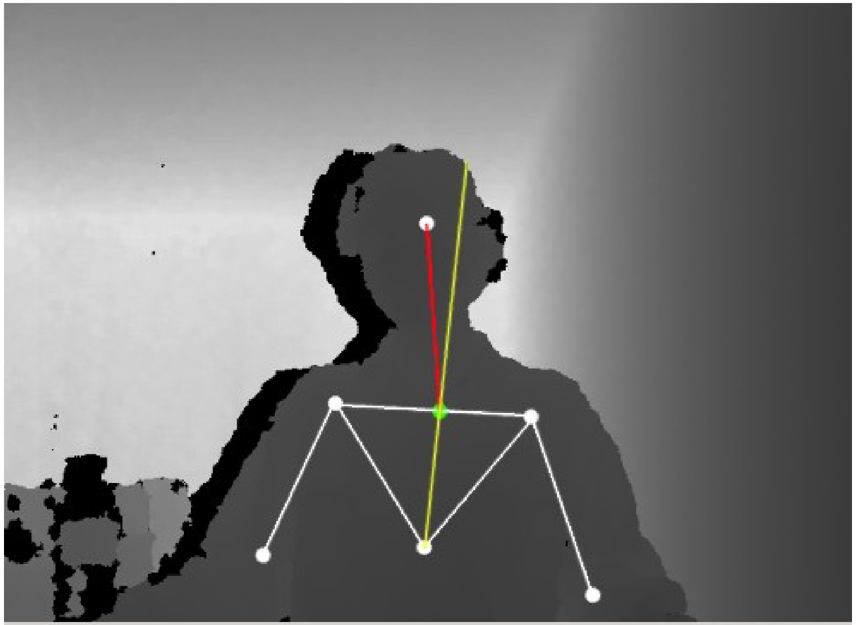
- Feature discriminative power evaluation:
  - 3 class separability criteria.

Posture representation

- Using the Kinect sensor skeletal landmarks for the head, torso, shoulders and elbow joints are tracked and recorded.

- Other features are calculated based on this initial markers.

An exemplary postural sample with the features shown as lines and points



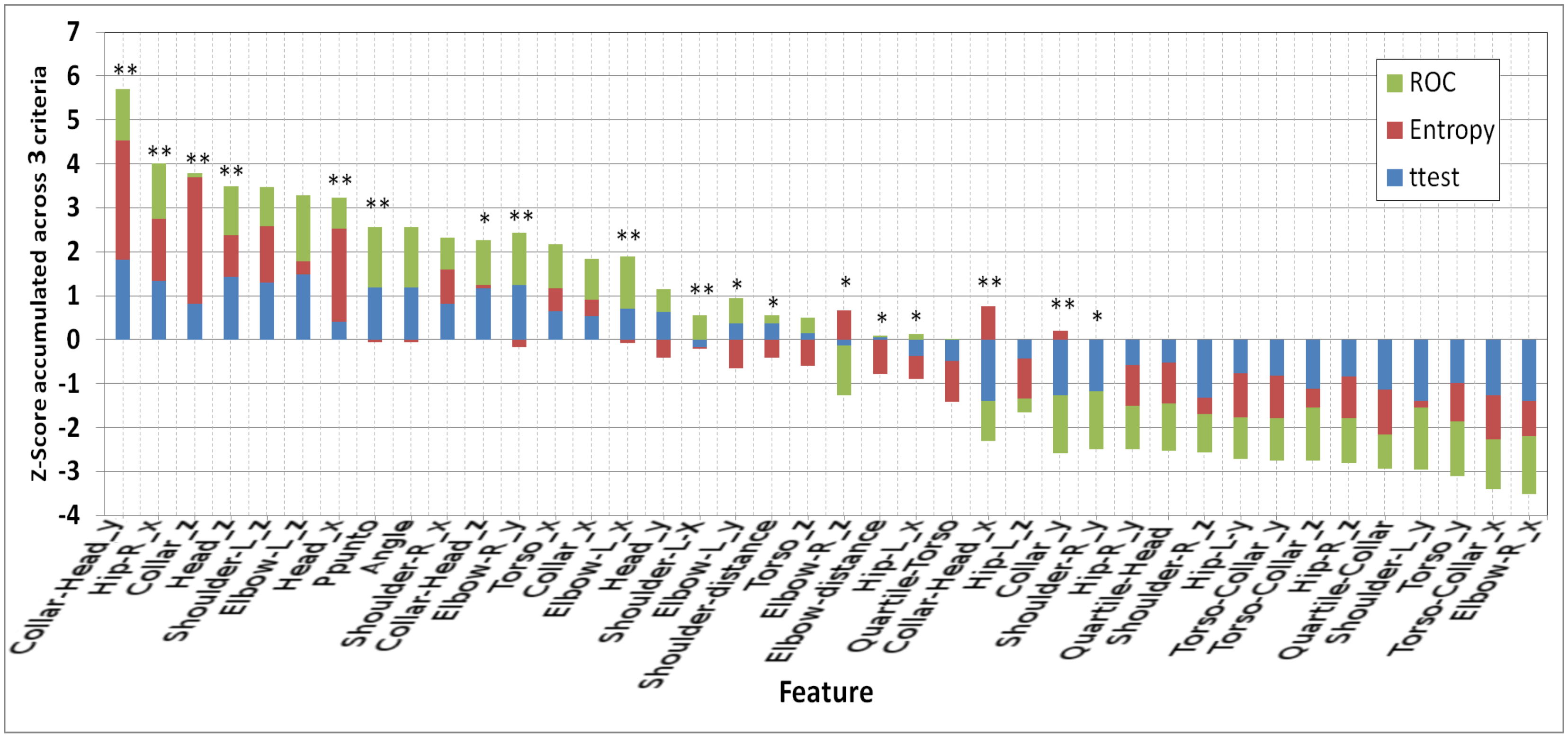
Feature	Description
Elbow distance	Euclidean distance between elbows
Shoulder distance	Euclidean distance between shoulders
Collar	Midpoint R-Shoulder/L-Shoulder
Torso-Collar	Vector formed by Torso-Collar
Collar-Head	Vector formed by Collar-Head
PPunto	Cross product of Torso-Collar Collar-Head vectors
Angle	Angle on the YZ axis between Torso-Collar and Collar-Head
Quartile head	Quartile occupied by the head position of the observation with respect to the head position throughout the subject's session
Quartile torso	Quartile occupied by the torso position of the observation with respect to the torso position throughout the subject's session
Quartile collar	Quartile occupied by the collar position of the observation with respect to the collar position throughout the subject's session

Table 1: Description of derived postural feature subset. Raw skeletal features can be appreciated in Figure 1.

INSTITUTO NACIONAL DE ASTROFÍSICA, ÓPTICA Y ELECTRÓNICA

Results

- Ranking of features by class separability criteria



The higher the bar, the more individual discriminative power by the feature. Combined discriminative power of feature sets is the result of the feature selection strategies. The **asterisks** above each bar indicates when the feature has been filtered by the feature selection strategies; one asterisk indicate that it has been selected only by one strategy, and the double asterisk indicate that the feature has been selected by both feature selection strategies.

Conclusions

- [Main conclusion] Unobtrusively monitoring posture of users while working in front of a computer can reliably be used to infer attentional disposition from the user.
- [Performance] Average classification of attention from posture reached 76.47%±4.58% (F-measure).
  - This is competitive with state-of-the-art ad-hoc solutions; despite no effort to specialised our model!
  - Best classification achieved 88.55%
  - The model and the feature selection strategy (satistically significant) affects the performance.
- [Finding] A total of 40 postural features were tested and those proxy of head tilt were found to be the most stable markers of attention in seated conditions based upon 3 class separability criteria over self-reported attentional state.
  - [Future work] Maybe a bit of leap thinking here… requieres more thorough testing
- [Limitations]
  - Assumption that attention must be directed to the computer system.
  - Small cohort size
  - A leave-one-out validation will explore generalization to subjects out-side the cohort.

EXPERIMENT II: ATTRIBUTION OF ATTENTION

Experimental Aim

- [Assumption] Unacted posture conveys cues about people's attentional disposition. (Same as for experiment I)
- [Aim] Identify **postural features leading to appreciation of attention** by third parties (e.g. educators)
- [Impact] This set of features would allow us to infer the attention level of a student interacting with educational software.
  - Human-computer interaction systems can benefit from this knowledge to customize the experience to the user changing attentional state.
  - Adapting the software to the student can in theory improve learning.

Experimental Set-Up

- Synthetic postural repository construction
  - Pre-discretized posture space
    - 432 images
  - 4 human models mimic a manek
  - Aged 11-15
  - Posture represented by Kinect

- Attribution of attention labelling
  - Crowd-sourcing
  - Raters blind to the task

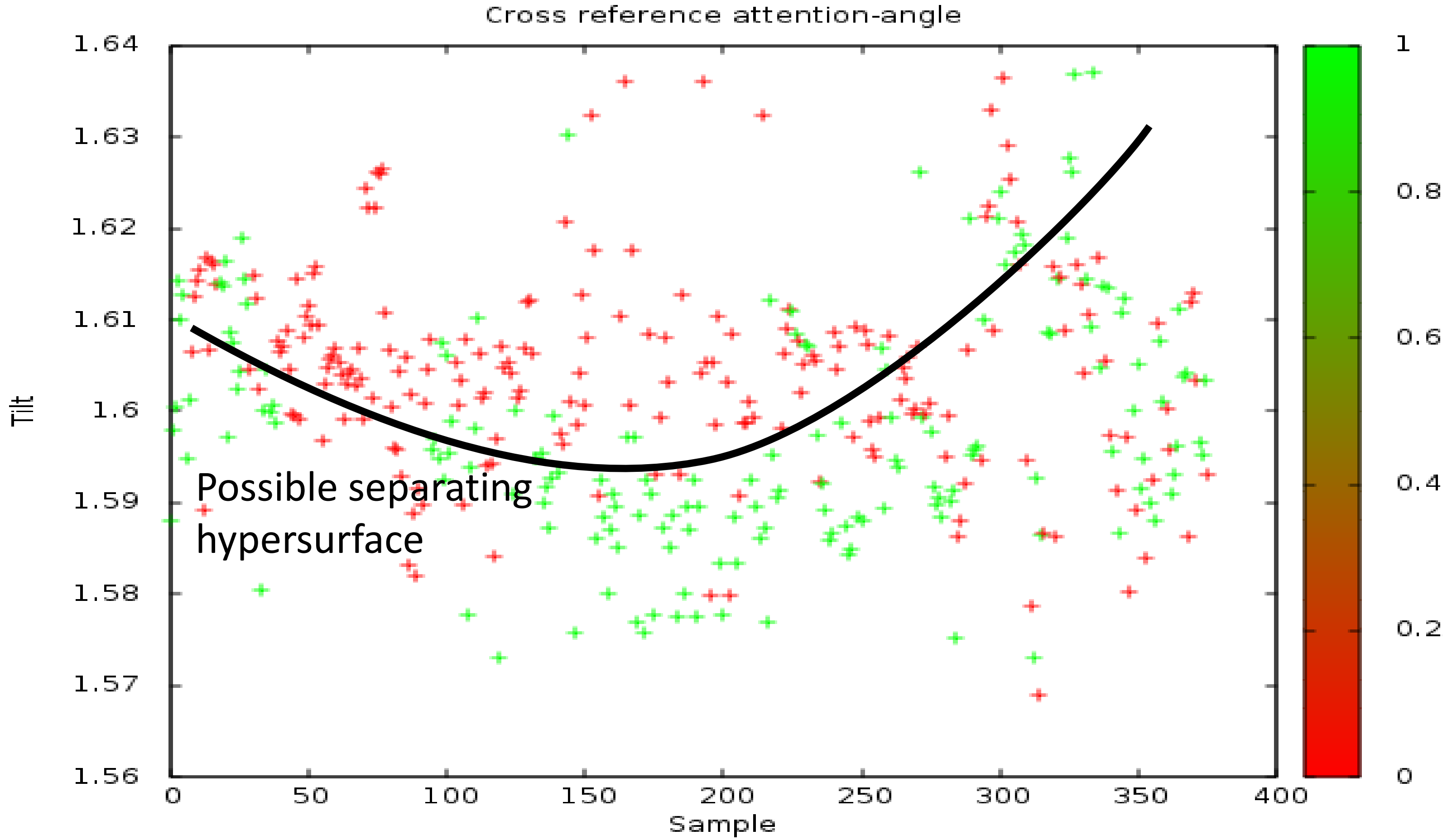
- Computational modelling:
  - Supervised learning.
  - In progress; still in prelimnary visual exploration

Crowd sourcing

- +1800 evaluations

Preliminary results

- Cross reference attributed attention / tilt of the head



REAL SET-UP: ECOLOGICAL VALIDITY

Intelligent Tutoring Systems (ITS)

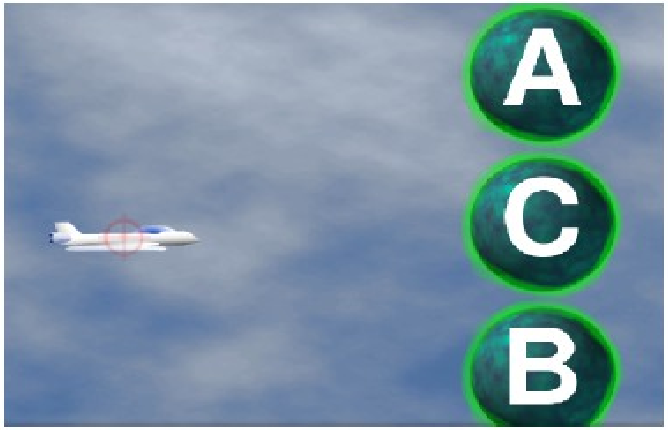
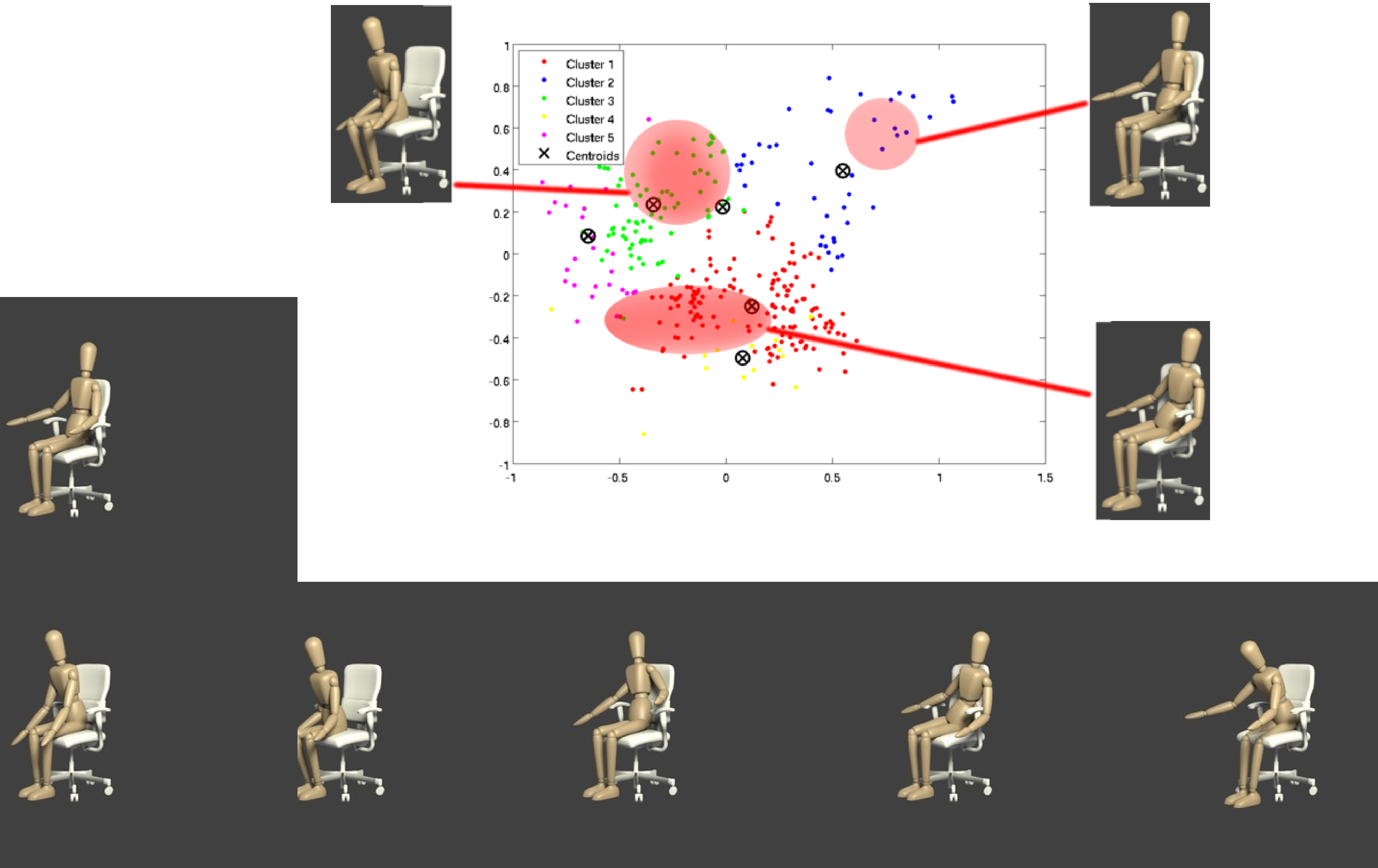


Figure 1. Interface of an Intelligence Tutoring System implemented to teach phonological awareness (Hurtado and Soto, 2011). The child is expected to drive the plane to the object representing the sound previously produced by the ITS whilst avoid colliding with erroneous solutions.

Experimental set-up

- 8 subjects
- Over 7000 data samples from Kinect
- Additional data (EEG, mouse interaction, autotutor)
- Expert (psychologist) analisis; labelling of attention

Preliminary results



Final conclusions

- Results suggests:
  - It is possible to monitor posture of a user unobtrusively
  - Unacted posture conveys attentional disposition from the user
- Conclusion:
  - With mature appropriate computational models, it should be possible to establish a relation between seated posture and attention
  - …and that relation can be exploited to dynamically adapt the ITS behaviour for enhancing educational tasks
- Future Work
  - The presented is still ongoing research; evidence must be piled up, computational models have to be polished, ITS must be enhanced with AI for adaptation to the student cognitive status and most importantly proof of educational benefits must be given.

Acknowledgements

- Other collaborators:
  - Universidad Católica de Chile
    - Alvaro Soto
    - Diego Cosmelli
    - Miguel Nussbaum

THANKS, ¿QUESTIONS?

BACK UP SLIDES

Posture representation

Feature	Description
Elbow distance	Euclidean distance between elbows
Shoulder distance	Euclidean distance between shoulders
Collar	Midpoint R-Shoulder/L-Shoulder
Torso-Collar	Vector formed by Torso-Collar
Collar-Head	Vector formed by Collar-Head
PPunto	Cross product of Torso-Collar Collar-Head vectors
Angle	Angle on the YZ axis between Torso-Collar and Collar-Head
Quartile head	Quartile occupied by the head position of the observation with respect to the head position throughout the subject's session
Quartile torso	Quartile occupied by the torso position of the observation with respect to the torso position throughout the subject's session
Quartile collar	Quartile occupied by the collar position of the observation with respect to the collar position throughout the subject's session

Table 1: Description of derived postural feature subset. Raw skeletal features can be appreciated in Figure 1.

- Parameterizations for the 4 classifiers

Classifier	Parameterization
Decision Tree (DT)	•BS=False; CF=0.25; mNO=2; nF=3; REP=False; STR=T; UP=False; UL=False •BS=True; CF=0.25; mNO=2; nF=3; REP=False; STR=T; UP=False; UL=False •BS=False; CF=0.25; mNO=2; nF=3; REP=True; STR=T; UP=False; UL=False •BS=True; CF=0.25; mNO=2; nF=3; REP=True; STR=T; UP=False; UL=False •BS=True; CF=0.25; mNO=2; nF=3; REP=False; STR=T; UP=True; UL=False •BS=False; CF=0.25; mNO=2; nF=3; REP=False; STR=T; UP=True; UL=False •BS=False; CF=0.25; mNO=2; nF=3; REP=False; STR=T; UP=True; UL=True •BS=False; CF=0.25; mNO=2; nF=3; REP=False; STR=False; UP=False; UL=True •BS=False; CF=0.25; mNO=2; nF=3; REP=False; STR=False; UP=False; UL=False
Naive Bayes (NB)	•UKE =False; USD=False •UKE =True; USD=False •UKE =False; USD=True
Support Vector Machine (SVM)	•Polynomial Kernel (p=1) •Polynomial Kernel (p=2) •Polynomial Kernel (p=3) •Polynomial Kernel (p=4) •Polynomial Kernel (p=5) •RBF Kernel (G=0.01) •RBF Kernel (G=0.05)
Tree Augmented Naive Bayes	•Estimator = simple; α=0.25 •Estimator = simple; α=0.5 •Estimator = simple; α=0.75

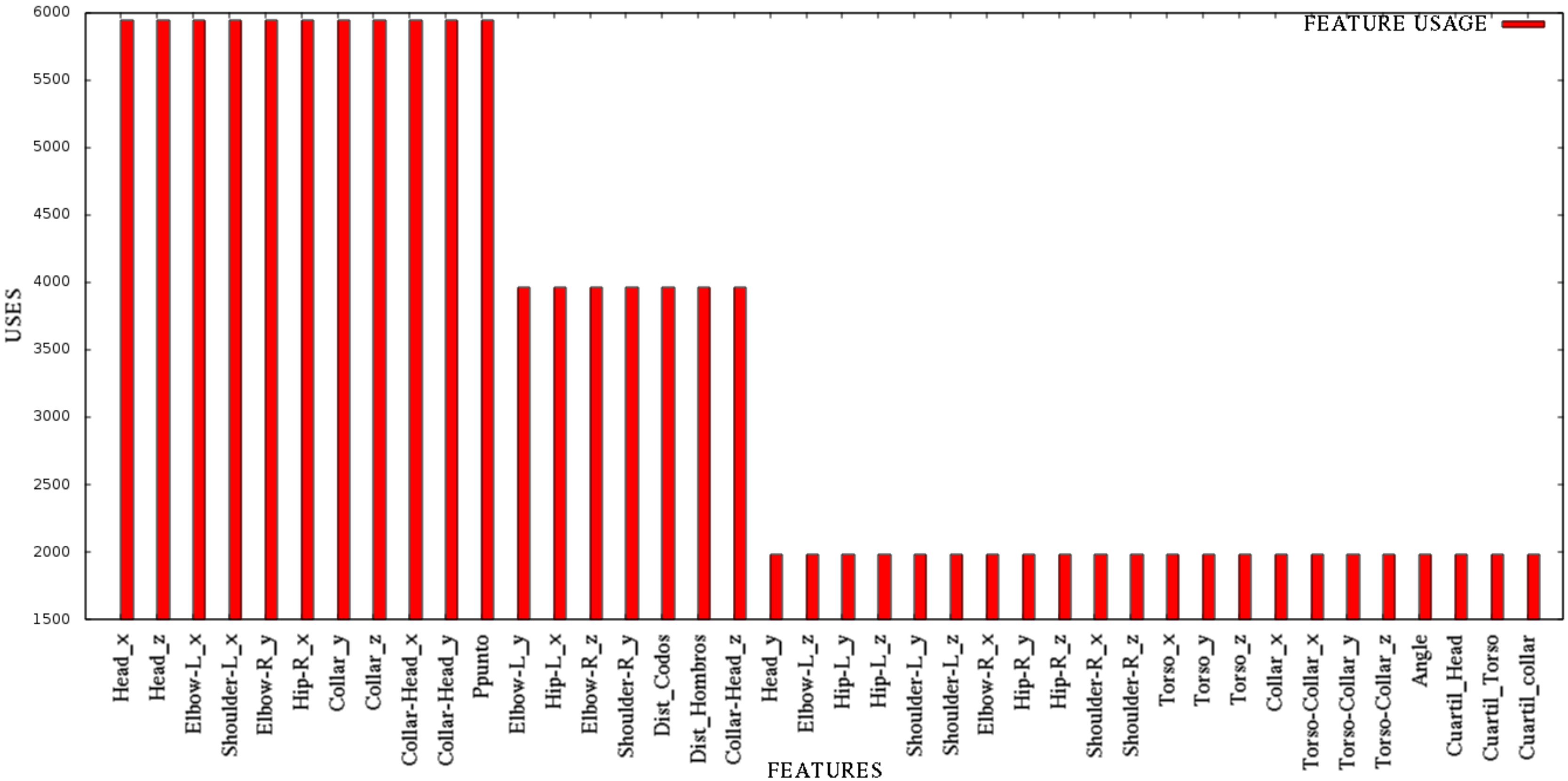
INSTITUTO NACIONAL DE ASTROFÍSICA, ÓPTICA Y ELECTRÓNICA

Computational modelling

- 60 frames long videos of each posture capture event were split into 6 equal-length chunks of 10 frames.
- In total, 377 labelled samples were obtained.
- A total of 5940 (= 3 feature selection strategies x 22 parameterization x 3 dataset partitions x 30 fold) classification exercises were carried out in Weka

Results

- Histogram of feature usage across feature selection strategies



Results

- Summary of classification results by classifier and feature selection strategy

Precision			
Classifier \Selection	NoAttSelection	CfsSubsetEval	ConsistencySubsetEval
Decision Tree	77.58±6.37	77.93±7.52	78.04±7.15
Naive Bayes	78.35±6.62	78.58±5.69	78.90±6.50
Tree augmented NB	75.55±6.05	76.43±6.57	76.43±6.52
SVM	70.46±5.94	69.01±4.94	69.42±5.27
Recall			
Classifier \Selection	NoAttSelection	CfsSubsetEval	ConsistencySubsetEval
Decision Tree	75.90±7.23	77.77±0.50	78.51±9.24
Naive Bayes	69.09±6.77	72.40±6.43	72.14±6.59
Tree augmented NB	75.93±6.66	77.35±6.49	78.40±7.02
SVM	84.32±5.93	86.36±5.30	86.91±5.22
F-Measure			
Classifier \Selection	NoAttSelection	CfsSubsetEval	ConsistencySubsetEval
Decision Tree	76.39±4.60	77.02±5.32	77.63±4.72
Naive Bayes	73.09±4.68	75.10±4.29	75.04±4.37
Tree augmented NB	75.43±4.11	76.63±4.79	77.07±4.59
SVM	76.49±3.84	76.55±3.69	77.02±4.01

Results

- Statistically significant difference between classifiers (one-way ANOVA ; $F_{(3,5936)} = 69.91$, $p<.000$).
  - Tukey post-hoc test revealed that the F-Measure was statistically significantly different for Naive Bayes.

- Statistically significant difference between feature selection strategies (one-way ANOVA; $F_{(2,5937)} = 3.17$, $p<.000$).
  - Tukey post-hoc test revealed that there was significant differences between not carrying out feature selection and other strategies, as well as between feature selection strategies.

INSTITUTO NACIONAL DE ASTROFÍSICA, ÓPTICA Y ELECTRÓNICA

Experimental set-up

CONACYT 40 años