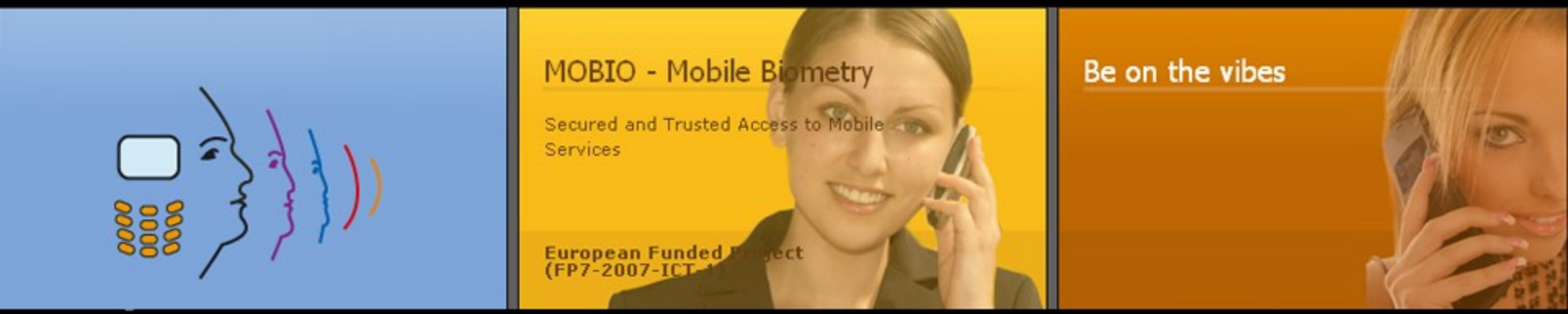
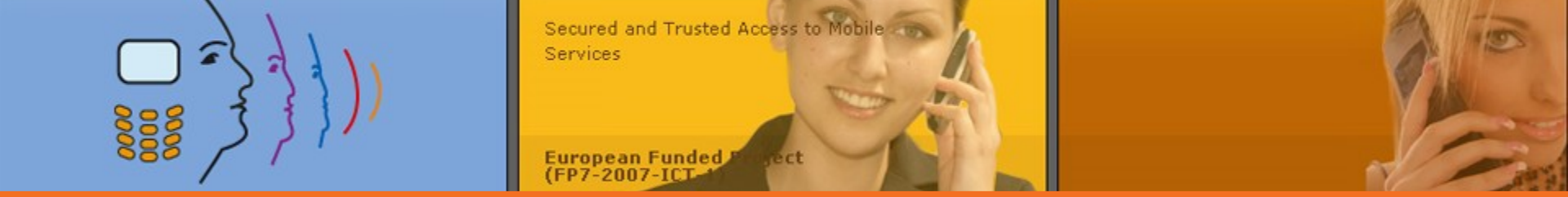
WP5 - Scalability

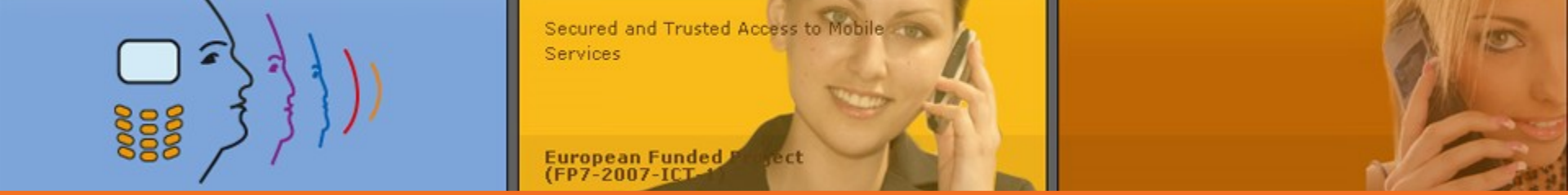
MOBIO Technical Meeting, Dec.10-11 2009

Brno - Czech republic



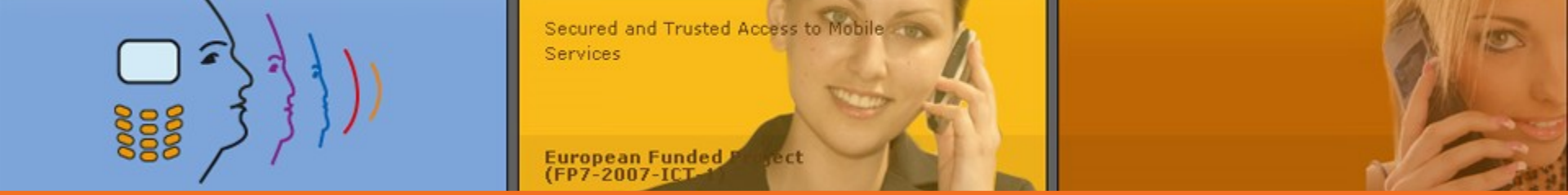


- Status:
  - Speaker recognition: done
  - Face detection: partially
  - Face localization: ok
  - Face verification: done



Speaker recognition - LIA

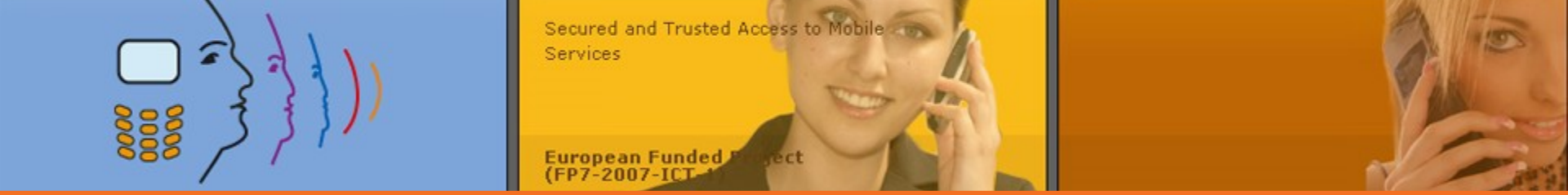
- Feature extraction
  - Limitation of size
    - Several acoustic vectors evaluated (from 20 to 50 coeff.)
- GMM statistics computation and scoring
  - Study of size models
    - different UBM sizes tested (from 512 to 32)
  - Investigation into selection of frames for GMM evaluation
    - Some ratio recorded/estimated frames achieved (from 100% to 25%)



Speaker recognition - LIA

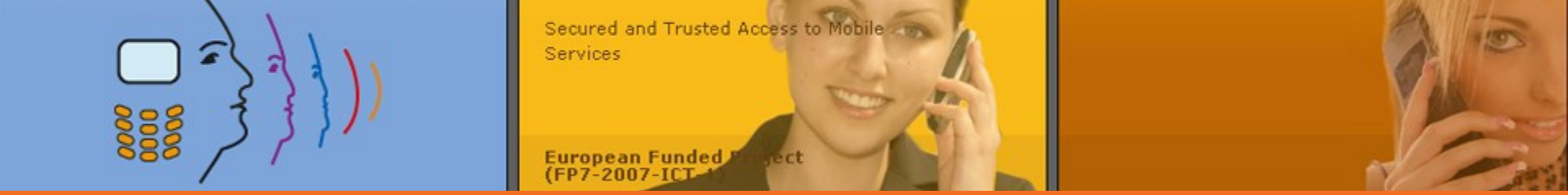
	Baseline	Minimal System	Best comp. System
G1 (EER)	3.48%	7.72%	**4.77%**
G2 (EER)	2.94%	7.34%	**2.94%**
mem. Peak	7.84 MB	1.37 MB	**2.19 MB**
rel. Mem.	100%	17%	**28%**
CPU time	2.06 s.	0.04 s.	**0.24 s.**
rel.time	100%	2%	**12%**
RT	0.0052	0.0001	**0.0006**

Evaluated on standard PC using only 1 core



Speaker recognition - BUT

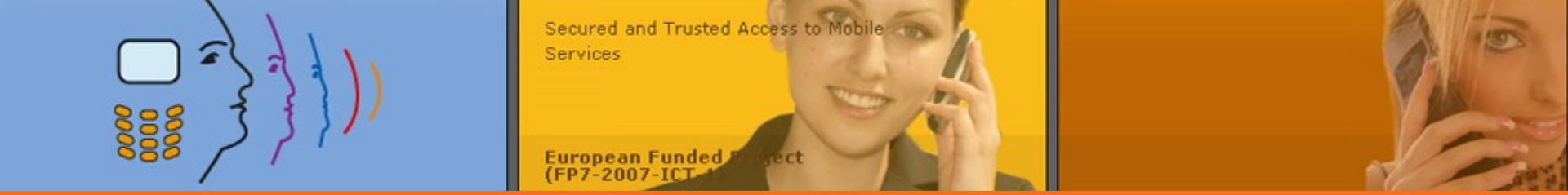
- VAD
  - Different approaches evaluated
    - NN, fast NN, GMM
- Feature extraction
  - Feature dimensionality
    - 39 HLDA, 39, 26
- GMM statistics computation and scoring
  - UBM size
    - from 2048 to 256
  - JFA U matrix rang
    - from 50 to 10



Speaker recognition - BUT

	Baseline	Best comp. System
G1 (EER)	7.16%	8.46%
G2 (EER)	5.27%	5.19%
mem. Peak	48.48	14.04
rel. Mem.	100%	29%
CPU time	20.6 s.	10,9 s.
rel.time	100%	53%
RT	0.0522	0.0277

Evaluated on standard PC using only 1 core

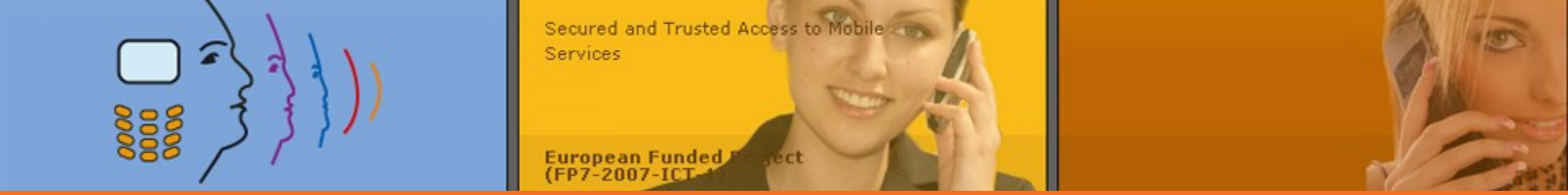


Face detection - UOULO

- fixed point system
- Number of windows processed

- No news related to:
  - System performance
  - memory & computational resource save

- Hoping for now ? 



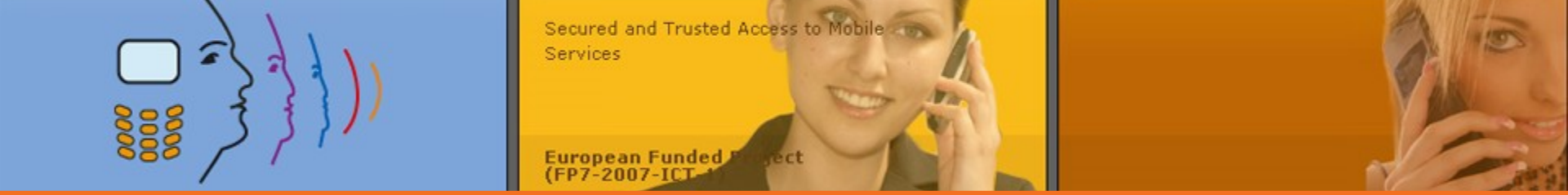
Face detection - IDIAP

- fixed point system:
  - Use of *N* bits to store “floating value” ; *N* varying from 8 to 22

    No modification of computational time & memory consumption ?

- Number of windows processed:
  - Stop as soon the first face was detected

- Tests using only the 20 specified tests for memory & CPU consumption



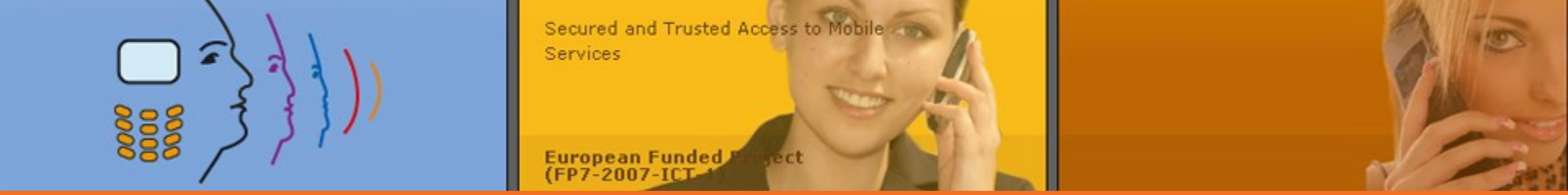
Face detection - IDIAP

	Baseline	**Best comp. System**
Accuracy (%)	99.37%	**98.95%**
mem. Peak	???	**???**
rel. Mem.	100%	**100%**
CPU time	41 m. 20 s.	**8 m. 40 s.**
rel.time	100%	**21%**
RT	???	**???**

What does it mean ?

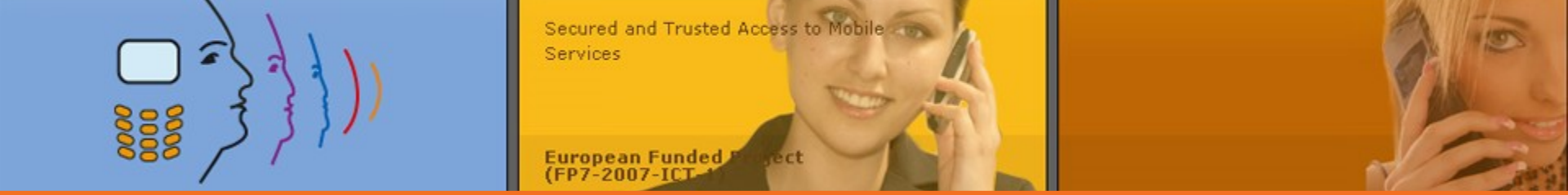
To evaluate for the D5.1

Evaluated on what type of machine ?



Face localization - UMAN

- A lot of scaled parameters:
  - # iterations for non-linear minimization
  - # of modes of the appearance subspace
  - # of facial features localized
  - Size of the image templates
  - Texture representation
  - Points prediction from the provided bounding box of the face
  - Points optimization
- Results on banca database available ?



Face localization - UMAN

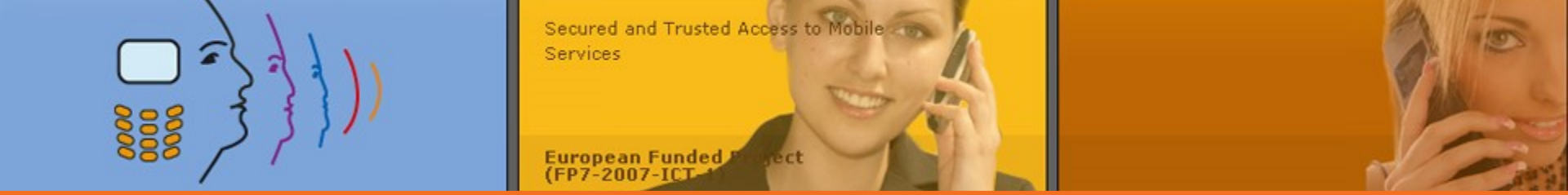
	Eye points d_{max}		All points d_{max}		All points d_{90}		All points d_{mean}	
	Med.	90%	Med.	90%	Med.	90%	Med.	90%
baseline	0.049	0.110	0.189	0.331	0.123	0.231	0.066	0.107
compromise	0.047	0.123	0.089	0.204	-	-	0.048	0.088

Table 19: Accuracy of compromise system.

	Time (ms)		Mem. (Mb)
	Med.	Mean	Peak
baseline	99	100	13.4809
compromise	31	30	7.61973

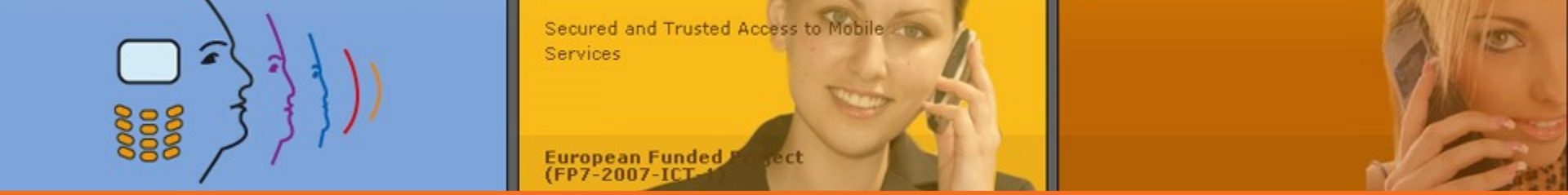
Table 20: Efficiency of compromise system.

Evaluated on what type of machine ?



Face verification - UNIS

- Multi-scale Local Binary Pattern Histogram Linear Discriminant Analysis (MLBPHLDA) approach:
  - Evolution of the number of LPB operators
  - Different sizes for non-overlapping regions
- Results on Feret & Banca



Face verification - UNIS

Memory consumption in MB

Minimal system 15% of “baseline”

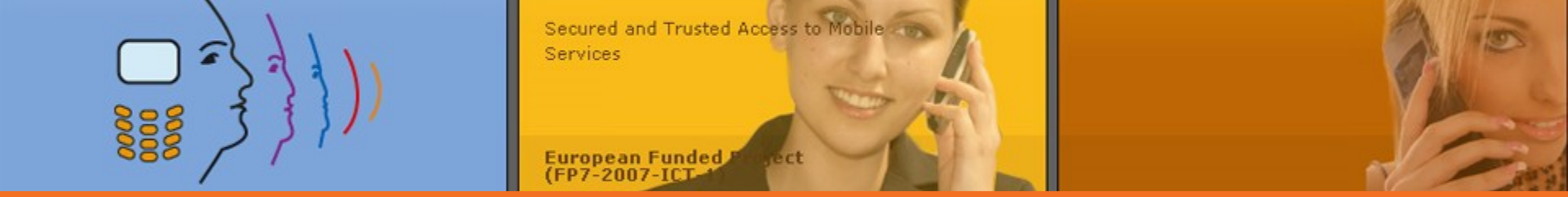
	$k = 5$	$k = 10$
9 LBP operators	133.356	412.805
5 LBP operators	88.169	229.691
3 LBP operators	64.020	135.929
9 LBP operators_T-norm	134.394	416.947
5 LBP operators_T-norm	89.140	233.474
3 LBP operators_T-norm	64.839	139.135

	$k = 5$	$k = 10$
9 LBP operators	74.39	122.99
5 LBP operators	51.72	79.48
3 LBP operators	40.70	55.03
9 LBP operators_T-norm	74.28	133.06
5 LBP operators_T-norm	52.51	81.33
3 LBP operators_T-norm	40.86	56.24

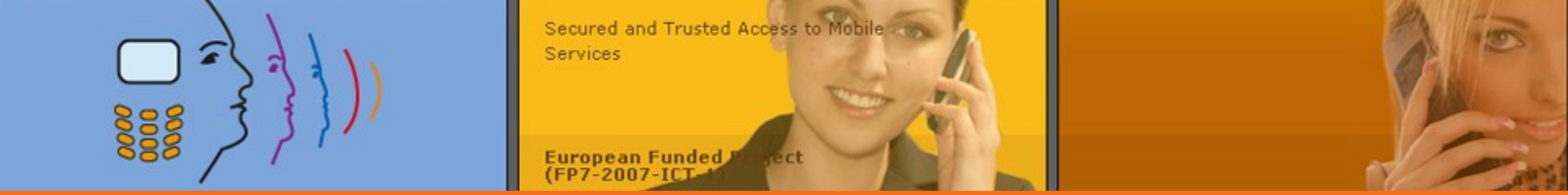
CPU consumption in MB

Minimal system 30% of “baseline”

Evaluated on what type of machine?

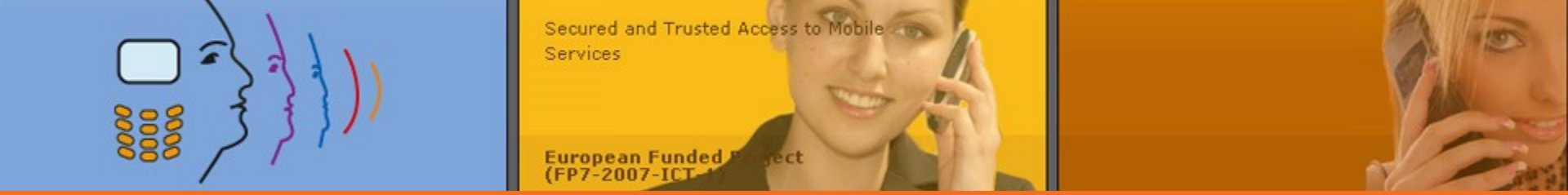


- Previous planning:
  - D5.1 ready for end of november:
  - Status at 8th of december: partially done
  - 15/10/2009:
    - Each partner upload on the SVN their scalable systems
    - UOULU system is missing
  - Preliminary Report on the uni-modal scalable systems
    - Good advance



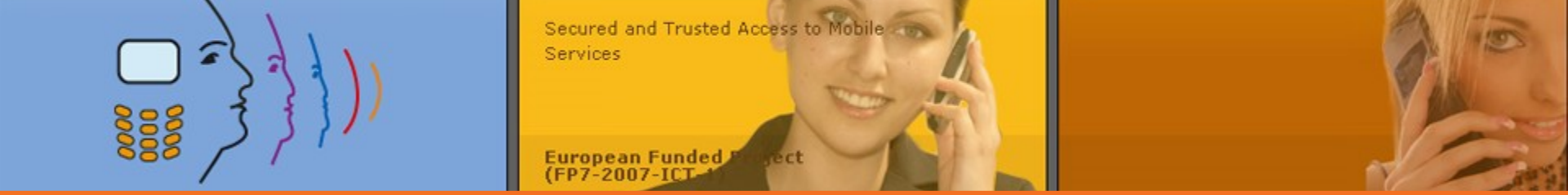
Reminder

- Report content: (*cf.* Previous meeting slides & email from Christophe 11/5)
  - Description of the systems (each scalable parameters)
  - Evaluation of scalable parameters
    - Performance / CPU / memory by using
      - “time” cmd and “valgrind”
  - Performance should be given both absolute and relative to the baseline system.
  - Evaluation of the « best » system (considering all scalable parameters together)



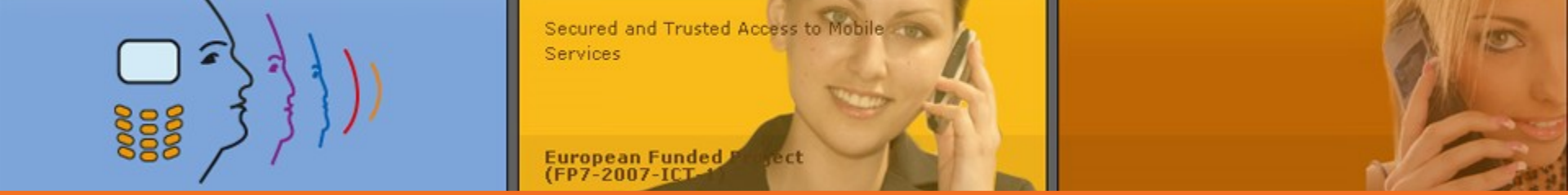
TODO LIST (1/2)

- Speaker recognition:
  - OK
- Face detection:
  - UOULU : all
  - IDIAP : banca tests for mem & cpu consumption ?
- Face localization:
  - banca tests for mem & cpu consumption ?



TODO LIST (2/2)

- Face recognition:
  - OK
- Face “toolchain”:
  - A full scalable system to use previous step at each level
- ALL:
  - improve the uni-modal report even if the final report (uni & multi) is due for m33 (→ WP6)



Next steps

- D5.2: (due to m33 ↔ 9/2010)
  - Part 1: bi-modal scalable system
    - What is a bi-modal system ?
    - Not really defined
  - Part 2: report
    - Urgently : standardization of the results table
      - Face “toolchain” able to provide EER ?
    - How to evaluate the bi-modal system ?
- Link with WP6