

Dynamic modeling of interpersonal stances towards an ECA based on linguistic and acoustic features

A.Batliner², F. De Rosis¹,
*N.Novielli*¹ and S.Steidl²

¹Department of Informatics
University of Bari
<http://www.di.uniba.it/intint/>

²Department of Computer Science
University Erlangen
<http://www.informatik.uni-erlangen.de/>

Users Do Not Necessarily Display A 'Warm' Social Attitude Towards Artificial Agents

What May Influence This Attitude?

Application's characteristics

Users' **personality traits**

(extrovert vs introvert,...)

Users' **background**

(culture, education,...)

Users' **expectations** and **goals** in interaction

(to be informed, to get advice, to enjoy, to challenge the application,...)



Social attitude needs to be recognized;
interaction needs to be adapted accordingly

Which Theory Of Social Attitude?

Various terms have been employed to denote closely related concepts:

- *Interpersonal Stance* (Scherer, Wichmann,...).
- *Empathy* (Paiva, Poggi, Vaknin,...),
- *Social Presence* (Polhemus, Kimble, Rettie, ...)

We will adopt Andersen and Guerrero's concept of 'interpersonal warm relationship as:

"the pleasant, contented, intimate feeling that occurs during positive interactions with friends, family, colleagues and romantic partners."

How To Recognize Social Attitude?

Various 'signs' (facial expression, gesture, posture, ...)

In particular, to Pennebaker (2001)

"Natural language is a marker of social interaction"

Signs of Social Attitude In The Language

from Andersen and Guerrero:

Sence of intimacy (use of a common jargon)

Attempt to establish a common ground

Humour

Benevolent/polemic attitude towards the system failures

Interest to protract or close interaction

from Polhemus et al:

Personal address

Paralanguage

Feeling

Value

Self disclosure

Social sharing

Humor

Acknowledgement

Social motivation

Negative comments

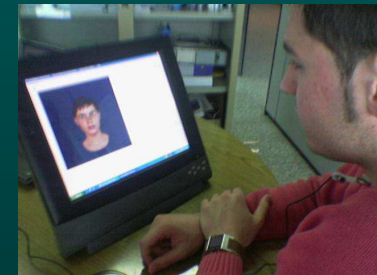
We study 'Almost Naturalistic' Dialogs
collected With Wizard of Oz Studies

This Presentation Is Based On Two Corpora*

text-based:



speech-based:



34 dialogs, 1146 moves overall

Subjects

23 to 30 years

students in Humanities vs Computer Science

equidistributed by gender

Effects Of Interaction Mode On Dialog Dynamics

	keyboard	speech
Av.evaluation (of message and character)	3.0	3.0
Av. n of moves per dialog	23.9 (9-60)	31 (14-55)
Av. move length per dialog (in characters)	48.5 (29-95)	80 (17-276)
% n. of questions per dialog	30.5	23.4
% of social moves	30.5	46.8

In speech-based interaction:

- higher level of involvement
- lower level of initiative
- higher level of social attitude

Previous Work On Recognition Of Affective States from linguistic and/or acoustic analysis of 'natural' corpora

- *personality traits*
'Big Five' traits in weblog texts: Gill and Oberlander. 2003;
- *valence:*
in spoken tutoring dialogs: Litman and Forbes-Ridley, 2003;
in call center dialogs: Lee et al, 2003;
- *individual emotions:*
anger, fear, relief, sadness in call center dialogs: Devillers and Vidrascu, 2006;
annoyance and frustration in telephone dialogs: Ang e al, 2002
- *emotional states / Interpersonal relations:*
call center dialogs and Wizard of Oz children-robot dialogs: Batliner et al, 2003/05

**Analysis of individual dialog turns.
Integration of acoustic and verbal features
with various classification methods**

But:

Social attitude is an emotional state
which evolves gradually all along the dialog
and requires
incremental and dynamic modeling

Steps In The Definition Of A User Model

- Define a markup language
- Label the corpus (3 raters)
- Evaluate the markup language
- Define, build and evaluate a parser
- Define and build a speech analyzer
- Define a user dynamic user modeling method
- Learn the user model from the annotated corpus
- Build a user modeling module

Labelling The Two Corpora:

Signs of social attitude

linguistic

negative comments

joke, humor

talks about self

questions about the agent

familiar style

-

-

-

-

-

positive comments

friendly self introduction

friendly farewell

acoustic

negative comment

joke

-

-

-

laughter

apologize

doubt

I'm thinking

encouragement

positive comments

agreement

friendly intonation

-

-

Evaluating The Markup Language for the two sets of signs

Markup language	Signs of social attitude	Inter.-rater agreement	Avg. agreement
Text corpus	Friendly self-introduction	.98	.80
	Familiar style	.33	
	Talks about self	.73	
	Question about the agent	.70	
	Positive comments	.82	
	Negative comments	.86	
	Friendly farewell	.93	
	Humor and Irony	.84	
Speech corpus	Smiling/Laughters	.98	.93
	Positive comments	.99	
	Agreement	.96	
	Friendly Intonation	.83	
	Encouragement	.97	
	Apologizing	.99	
	Doubt	.93	
	Negative comments	.94	
	Humour and Irony	.98	
	I'm thinking	.91	
	Neutral (nessuna label indicata)	.76	

A Few Examples Of 'Social Moves' in the language

A: Hi, my name is Valentina and I'm here to suggest you how to improve your diet. (agent move)

U: Hi Valentina, my name is Carlo and I'm very happy to interact with you (move to annotate)

signs: *friendly self-introduction*

A: Are you attracted by sweets? (agent move)

U: Enough. I like both preparing and eating them (move to annotate)

signs: *talks about self*

A: What do you think of a good dish of spaghetti? (agent move)

U: A good dish of spaghetti is tempting and would be OK now! (move to annotate)

signs: *familiar style*

Examples of 'social moves' in the speech

Sense of intimacy

friendly intonation

([SPImma](#), [SPMarino](#), [NiceEyesMarino](#))

encouragement,

agreement,

([IKnowM](#), [URightM](#), [WeAgree](#))

apologizing

([SorryM](#))

Humour

smiling

([Imma](#), [MensSana](#))

Benevolent/polemic attitude towards the system failures

favourable/negative intonation

([PosComm](#), [negComm7](#), [NegComm17](#))

'canonical' pronunciation or reducing speech, to favour system understanding of the subject utterance

+ signs of reflection (*I'm thinking, doubt*)

([ThinkI](#), [ThinkM](#))

Recognizing Signs Of Social Attitude In Language: probabilistic parsing

We denote:

- * with S , a set of *signs of social attitude*: $S = \{s_1, s_2, \dots, s_i, \dots, s_n\}$
- * $\forall j$, with C^j a set of *semantic categories of word sequences* which are 'salient' for s_j : $C^j = \{c^j_1, c^j_2, \dots, c^j_j, \dots, c^j_m\}$
(with $(C^1 \cap \dots \cap C^j \cap \dots \cap C^n \neq \text{Nil})$ and $(C^1 \cup \dots \cup C^j \cup \dots \cup C^n = C)$).

Given:

- * a prior probability $P(s_i)$ for the sign s_i and
- * $\forall j$, a probability $P(V(C^j))$ for the combination $V(C^j)$ of values of elements in C^j (each value being = T/F)
- * a probability $P(V(C^j) | s_i)$ for the combination $V(C^j)$ of values of elements in C^j in text strings displaying the sign s_i ,

and given

- * a result of lexical analysis of the user move m_h , as a set of values for all elements in C ,

Compute, $\forall j$, the probability of sign s_i as:
$$P(s_i | V(C^j)) = P(V(C^j) | s_i) * P(s_i) / P(V(C^j))$$

Some Semantic Categories

(largely domain-independent)

Greetings: good morning, nice to meet you, ...

Self Introduction: my name is, I am, ...

Ciao: ciao,cia'

Farewell: good bye, see you soon, ...

Thanking: thanks, thank you,...

Friendly Self-Introduction

Friendly Farewell

Generic Neg Comments: you scare me, you depress me, stop please!,...

Expr of Disagreement: no, but, what's bad?, I don't agree, ...

Neg Evaluation Of Message: it's too much, it's not enough, ...

Neg Evaluation Of Agent's Politeness: you are rude, ...

Negative Comments

Neg Evaluation Of Agent's Competence:

you don't know, you are not able to, you are narrow-minded, ...

Negative Eval Of Agent's Repetitiveness: you repeat the same things,
you already told that,...

Negative Eval Of Agent's Understanding Ability: you did not understand,
you don't understand, ...

Recognition Power Of The Parser

Sign	(different cut-off values for the probabilities of the various signs)		
	Sensitivity	Specificity	Accuracy
Friendly SelfIntro	1.00	.96	.96
Friendly Farewell	.88	.94	.94
Questions Ab Ag	.92	.86	.87
Talks About Self	.84	.83	.84
Positive Comm	.80	.88	.96
Negative Comm	.73	.93	.91
Familiar Style	.41	.79	.70

Between-Signs Confusion Matrix

	FrSIn	FFar	Talk	Que	NegCo	PosCo	FStyle	Unclassif
threshold	0,10	0,15	0,40	0,30	0,20	0,65	0,25	<0,4
FrSelfInt	1,00	0,86	0,43	0,00	0,00	0,00	0,00	0,00
FrFarew	0,50	0,88	0,00	0,06	0,00	0,00	0,63	0,13
TalksAbS	0,00	0,01	0,84	0,01	0,09	0,03	0,16	0,13
QuesAgt	0,00	0,00	0,08	0,92	0,08	0,00	0,12	0,04
NegCom	0,00	0,05	0,32	0,41	0,73	0,00	0,14	0,18
PosCom	0,00	0,20	0,25	0,25	0,00	0,80	0,35	0,00
FamStyle	0,00	0,00	0,13	0,04	0,00	0,04	0,41	0,48

Also some problem with speech-analysis

affect		Predicted Group Membership (chance level of correct classification 16.7%)					
		Laughter	Friendly	Agreement	Thinking	Neutral	Negative
(a) Original correctly classified cases = 70.9%							
%	Laughter	72.1	.0	7.0	2.3	9.3	9.3
	Friendly	.8	67.8	3.4	2.5	17.8	7.6
	Agreement	2.4	4.8	83.3	.0	7.1	2.4
	Thinking	.0	4.5	.8	78.2	9.0	7.5
	Neutral	2.3	12.3	2.8	5.9	66.8	9.8
	Negative	3.6	4.8	1.2	6.0	8.4	75.9

Confusion matrix - training set = test set

Recognition power of LDA classifier

Accuracy

- test set = training set: accuracy of 70.9%
- Loo-cross validation: accuracy of 37.3%

Why such a big gap?

- We have to further investigate the reasons
- Probable problems due to very unequal distribution of labels among speakers (speaker-specific)

Role of the User Model

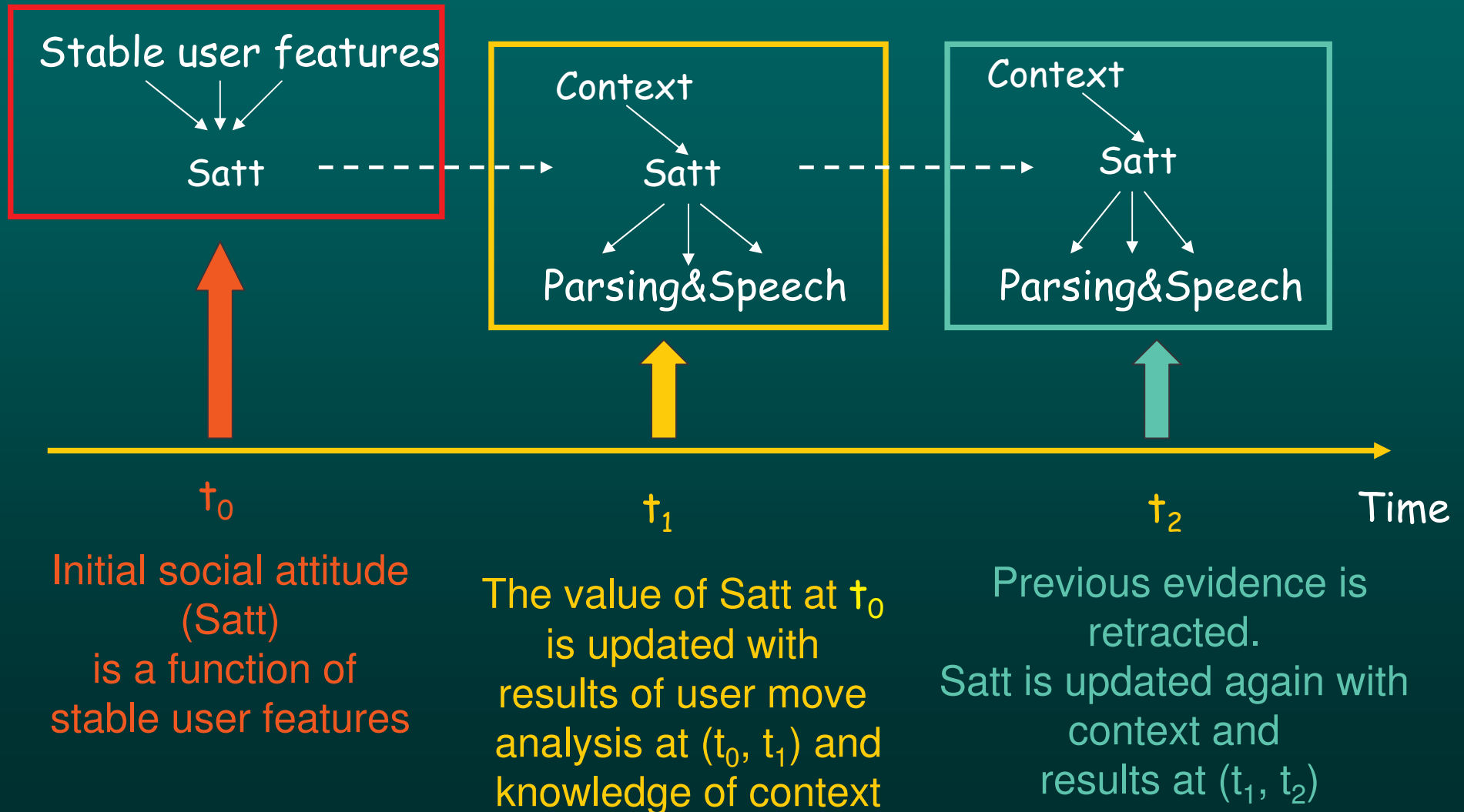
- to *combine results* of linguistic and acoustic analysis ;
- to consider *the context* in which the move was uttered;
- to build a *dynamic, incremental* image of the social attitude (late fusion rather than early fusion)

A dynamic bayesian network:

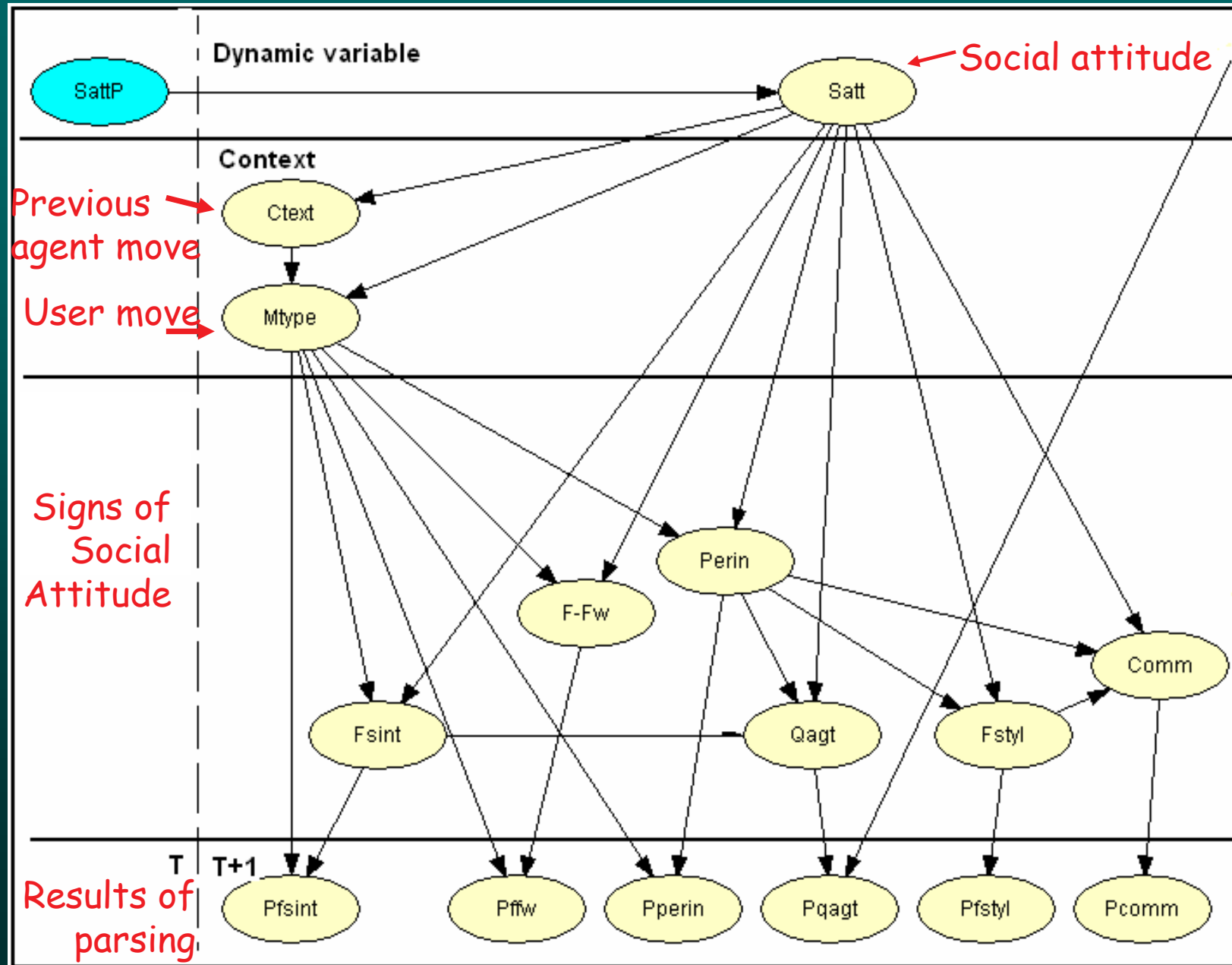
- ‘**Observable**’ variables are the ‘stable’ user characteristics, the results of language and speech analysis and the previous agent move.
- ‘**Hidden**’ variables are the signs of social attitude and the social attitude itself (this is a ‘**dynamic, monitored** variable’)

We learned a first version of the model
from the text-based corpus of dialogs
(with the K2 algorithm, by Cooper and Herskovitz)

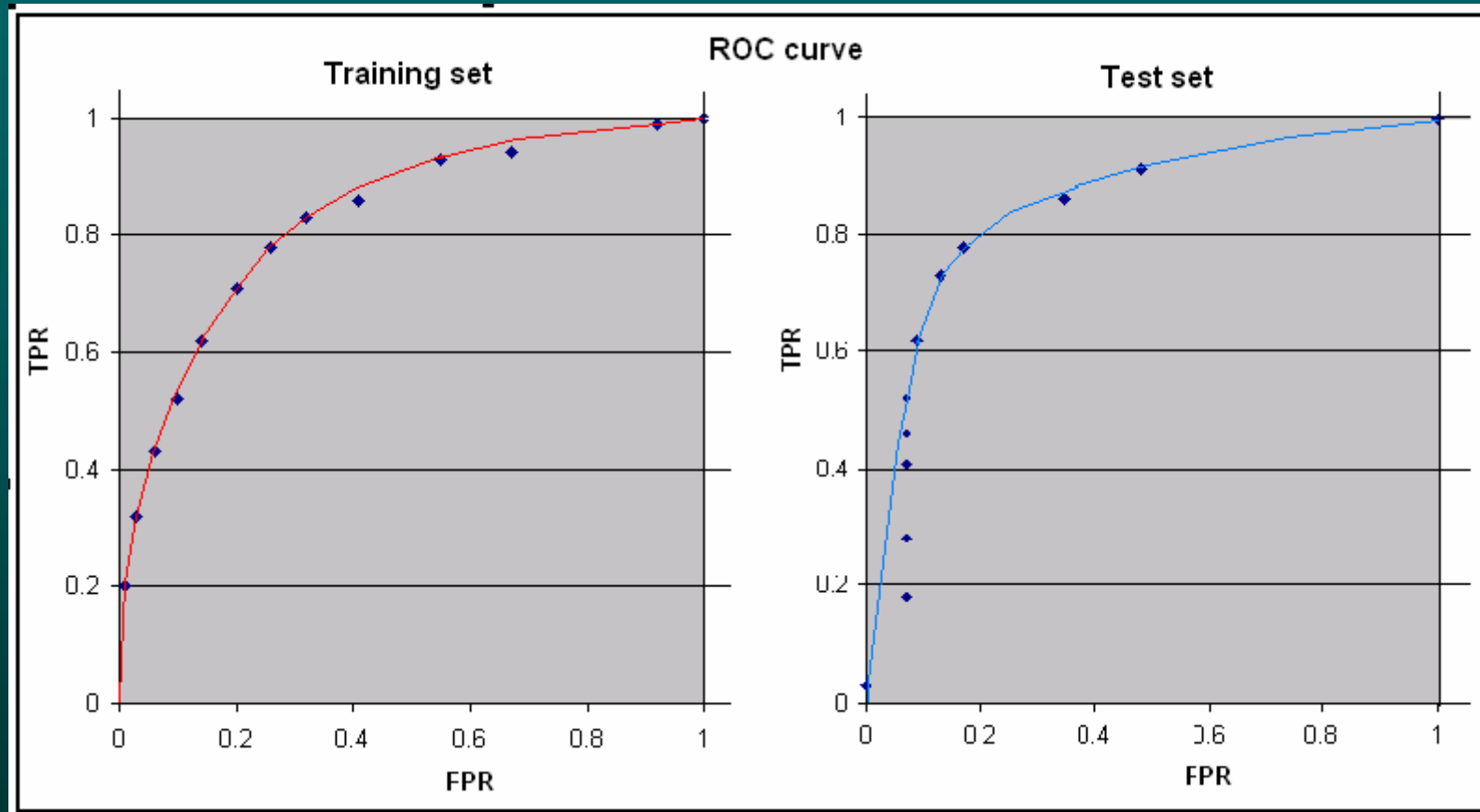
A Dynamically And Incrementally Updated User Model



Zooming on a generic time slice



Model validation: ROC analysis



Sensitivity: from .78 to .83
Specificity: from .68 to .74
Best %CCC: 75 %

Best %CCC: 79 %

Application of the user modeling agent to dialog simulation

- At the beginning of interaction, the model is initialized by introducing evidence about stable *user characteristics*;
- after every user move, *results of parsing* and of acoustic analysis are introduced and propagated in the network, with evidence about the *previous agent move* (context);
- the new values of the *signs of 'social attitude'* are read and contribute to formulate the short-term agent behavior (next agent move);
- the new value of *'social attitude'* (*dynamic variable*) is read and contributes to the planning of the long-term agent behavior.

An Example Dialog

W1: Hi, my name is Valentina. I'm here to suggest you how to improve your diet. Do you want to talk about your eating habits?

S1: Yes

W2: Do you like eating?

S2: Reasonably

W3: What would you think of a good dish of spaghetti?

S3: A good dish of spaghetti is tempting and would be OK now! (fam. style)

W4: I agree! Do you think that your diet is correct or would you like to change it?

S4: I think I would need a more regular regimen. (talks about self)

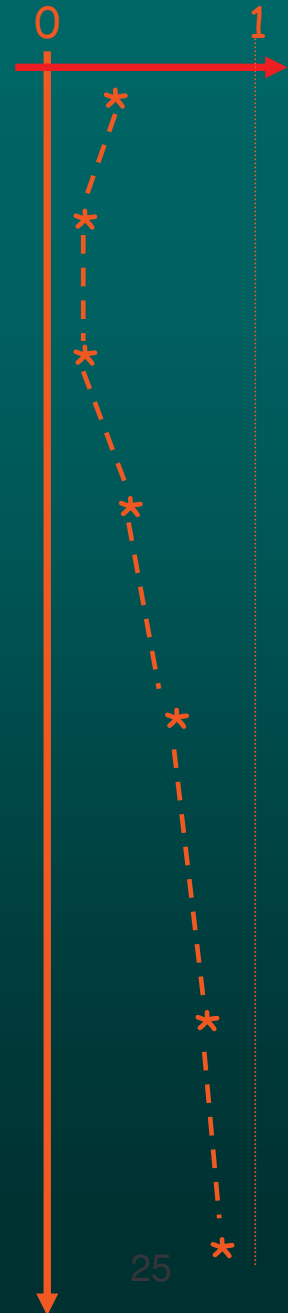
W5: Good, let's try to do something then. Did you ever think to change your diet?

S5: Yes, sometimes I did it and got very good results. Now, either because of my laziness or my stressing rythms I can't lay myself into more regular habits. (talks about self)

W6: You seem to be in a more positive attitude now. Do you believe your weight is correct or would you like to change it in any way?

S6: Unfortunately it is not ideal. I would like to change it and come back to my weight of a few years ago. (talks about self)

P(SocialAttitude)



Future work in the scope of HUMAINE Interchange(s)

- To repeat analysis on a larger corpus, in order to reduce the problem of sparse data (with Uni-Erlangen);
- To test the level of complementarity between linguistic and acoustic analysis on signs that may be analyzed with the two methods (with UNI-Erlangen);
- To revise the user model (DBN) so as to enable introducing and propagating 'soft' rather than 'hard' evidence (results of probabilistic parsing and classification of acoustic features);
- To investigate the possibility of connecting our probabilistic parser to WordNet, so as to increase corpus-independence of linguistic analysis (with ITC-IRST, in the scope of WP8).