MODELING OF USER STATE — ESPECIALLY OF EMOTIONS

Elmar Nöth

University of Erlangen–Nuremberg, Chair for Pattern Recognition, Erlangen, F.R.G.

email: noeth@informatik.uni-erlangen.de

Dagstuhl, October 2001

HOW TO FIND TROUBLE IN COMMUNICATION

- Emotion detection (especially of anger, rage, frustration)
  - makes it possible to react adequately to angry users
  - will lead to a higher degree of acceptance
  - is essential for call center applications:

Annoyed users may never call again!

- better: not 'full–blown' anger, but slight/medium irritation indicating a critical phase in dialogue
- look for peculiarities w.r.t. prosody, lexical content, grammatical structure, and dialogue structure
- combine with other modalities, i.e. interpretation of facial expression (not dealt with here)

Batliner et al.: How to Find Trouble in Communication.
To appear in Speech Communication

PROSODY AND EMOTIONS: RESEARCH APPROACHES

- actors (classic, basic research)
- readers (classic, basic research)
- simulation in a WOZ–scenario, closer to *real life*
- human beings in *real life* (not yet)

PROSODIC CLASSIFICATION

- (normally Neural Networks: MLP)
- Linear Discriminant Analysis LDA
- Cart and Regression Trees CRT

- cross–classification (leave–one–out)
- equal distribution
- other default settings

special interest here: comparison of the three scenarios, not optimization of classification

PROSODIC FEATURE SET

features	context size				
	-2	-1	0	1	2
DurTauLoc; EnTauLoc; F0MeanGlob			•		
Dur: Norm,Abs,AbsSyl; En: RegCoeff,MseReg,Norm,Abs,Mean,Max,MaxPos; F0: RegCoeff,MseReg,Mean,Max,MaxPos,Min,MinPos		• • •	• • •	• • •	
Pause-before, PauseFill-before; F0: Off,Offpos		•	•		
Pause-after, PauseFill-after; F0: On,Onpos			•	•	
Dur: Norm,Abs,AbsSyl En: RegCoeff,MseReg,Norm,Abs,Mean F0: RegCoeff,MseReg	• • •			• • •	
F0: RegCoeff,MseReg; En: RegCoeff,MseReg; Dur: Norm		•			
API,APN,AUX,NOUN,PAJ,VERB	•	•	•	•	•

OVERALL PERCENTAGE OF CORRECTLY CLASSIFIED CASES

	cross-classified				$l \neq t$
features	prosodic–acoustic		pros.–acoust. + word info		prosodic acoustic
	LDA	CRT	LDA	CRT	MLP
actor	89	81	97	91	86
read	73	69	82	78	54
WOZ	69	65	71	68	63

prosodic–acoustic: 'pure' prosody, computed globally for whole turn
pros.–acoust. + word info: prosodic features based on output of word recognition, normalized, computed globally for whole turn

WHERE HAVE ALL THE EMOTIONS GONE?

- only 5 subjects (out of 62) reported that they have **not** been angry but amused
- problematic operationalization: emotional = marked prosodically?
- actors: act emotionally because they are supposed to do 'as if'
- 'readers': can only vary prosody
- 'normal' human beings in (semi–) natural situations have much more possibilities to express their state, for instance:

repetitions per phase

phase in dialogue	0	1	2	3	4	5
# of occurrences	0	29	74	66	69	46

TRANSCRIPTION CONVENTIONS FOR PROSODIC PECULIARITIES

<B> = breathing,

<P> = pause,

*1 = Pauses between syntactic/semantic units,

*2 = hyper-clear speech,

*3 = strong emphasis,

*4 = pauses between words,

*5 = very strong emphasis,

*6 = pauses inside words,

*7 = syllable lengthening,

*8 = hyper-articulation (with phoneme changes),

*9 = speech distorted by sighing or laughter.

WOZ–DIALOGUE, FAULTY SYSTEM UTTERANCE, EARLY PHASE IN DIALOGUE

system: *die Urlaubszeit ist vom fünfzehnten Juni bis zwanzigsten Juli.*
vacation time is from 15th of June to 20th of July.

user: *ja, das hat ja auch nicht viel damit zu tun, da wir uns im Januar befinden, ne?*
yes, and this has not much to do with the fact that we are talking about January, has it?

WOZ–DIALOGUE, FAULTY SYSTEM UTTERANCE, LATER PHASE

system: *die Urlaubszeit ist vom fünfzehnten Juni bis zwanzigsten Juli.*
vacation time is from 15th of June to 20th of July.

user: *<B> ja, klasse. <P> Dienstag, zwölfter erster,*
*achtzehn *3 bis zweiundzwanzig *2 Uhr *2.*
<B> yes, great. <P> Tuesday the 12th of January, 6 to 10pm.

<B> = breathing

<P> = pause

*2 = hyperclear speech

*3 = strong emphasis

WOZ–DIALOGUE, FAULTY SYSTEM UTTERANCE, EVEN LATER PHASE

system: *die Urlaubszeit ist vom fünfzehnten Juni bis zwanzigsten Juli.*
vacation time is from 15th of June to 20th of July.

user: *dich sollte man feuern. <B> sechster *4 Januar *4,*
*<P> zwanzig *2 bis zweiundzwanzig Uhr.*
you should be thrown out. <B> 6th of January,
<P> 8 to 10 p.m.

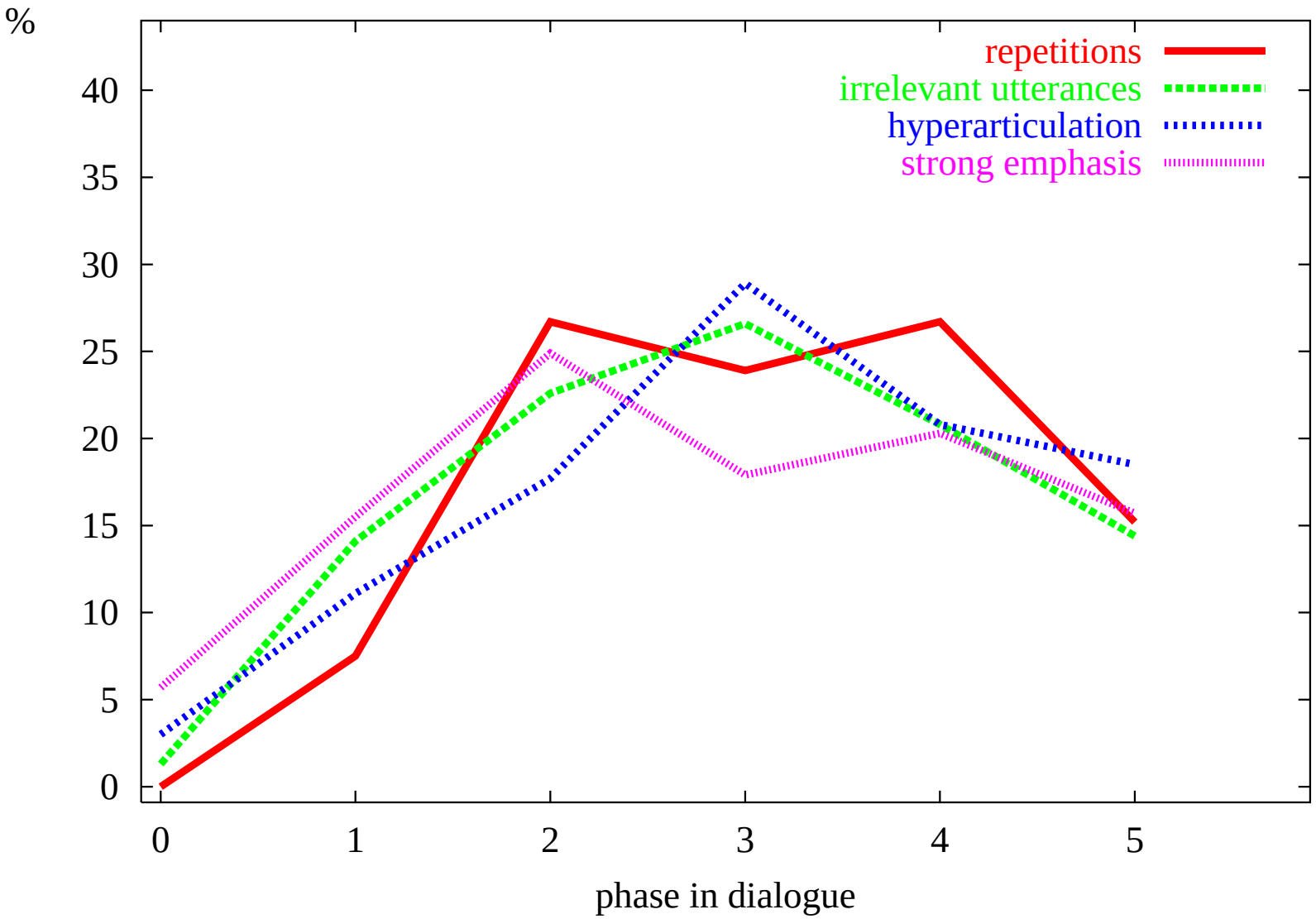
<B> = breathing

<P> = pause

*2 = hyperclear speech

*4 = pauses between words

OCCURRENCE OF CONVERSATIONAL AND PROSODIC PECULIARITIES



CONCEPTS USED FOR WOZ CLASSIFICATION

phenomena	#	**source**
prosodic features *	91	extracted automatically
part–of–speech features POS *	6	annotated in the lexicon by hand
dialogue act features DA	18	LM: trained with VERBMOBIL data, automatic annotation
prosodic peculiarities	10/2	annotated by hand
repetitions	2	annotated automatically (Levenshtein distance)
syntactic–prosodic S boundaries	5	LM: trained with VERBMOBIL data, automatic annotation

CLASSIFICATION OF ANGER VS. NO ANGER

- LDA, leave-one-out, best classification result in percent
- different feature combinations for Actor, Read, and WOZ

	Actor	Read	WOZ
# of cases	10316	13053	28649
features	avRec	avRec	avRec
prosodic	95.4	77.4	73.2
POS	72.2	63.0	66.1
POS, only 0	72.4	57.6	64.1
pros./POS	95.7	79.6	73.7

WOZ	
# of case	28649
features	avRec
DA	56.1
POS/DA	66.8
pros./DA	73.4
pros./POS/DA	**74.2**

WHY NO SATISFYING CLASSIFICATION OF WOZ–DATA?

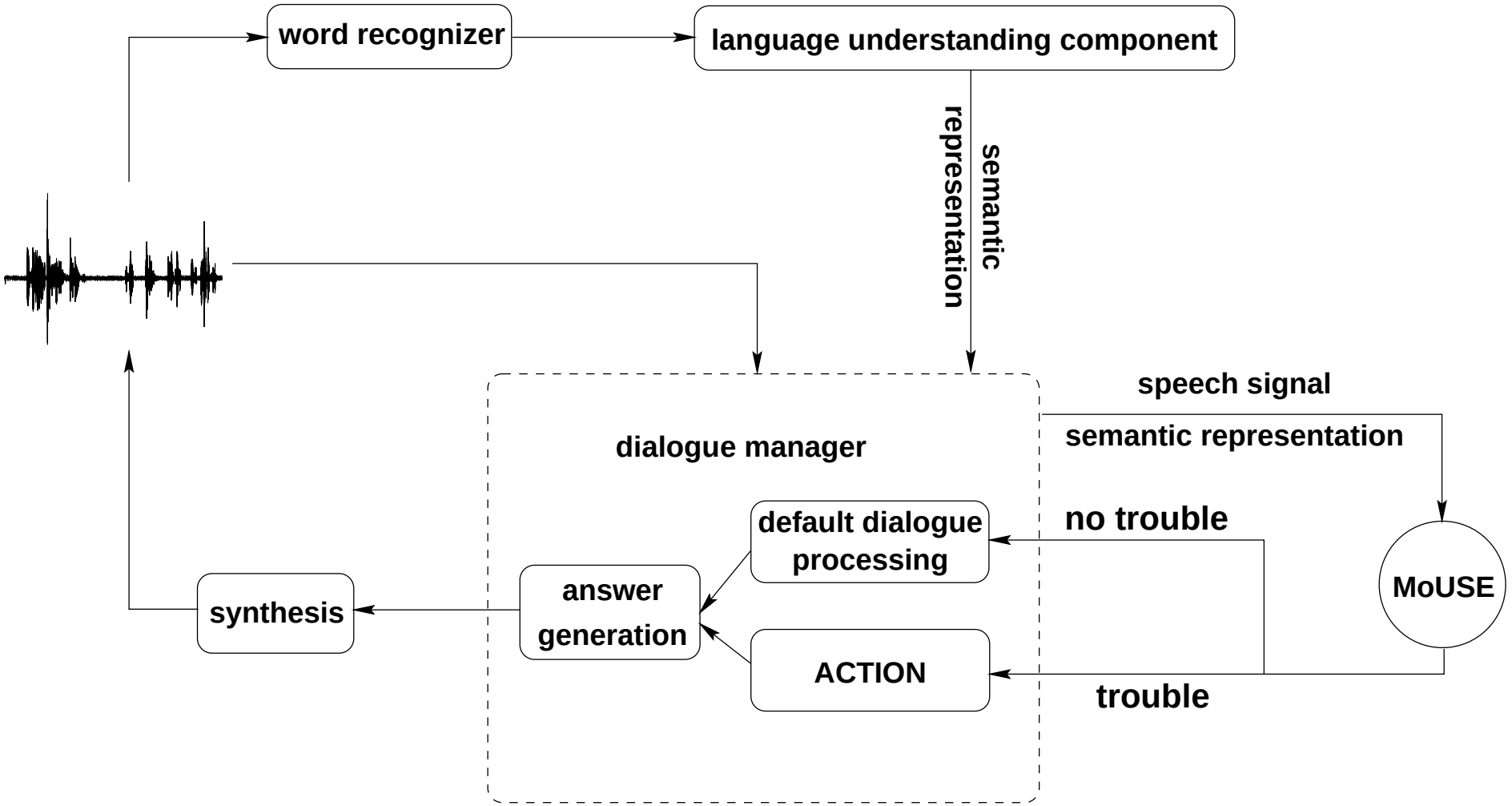
- suboptimal classifiers?
- suboptimal features?
- 'prosodic' WOZ–labels: rather 'segmental' labels?
- not enough training material
- 'noise': multi-functionality of prosody (boundaries, accents, etc.)
- most important however:
  - camouflage of emotions esp. in transactional situations
  - marking of emotion not only by prosodic means

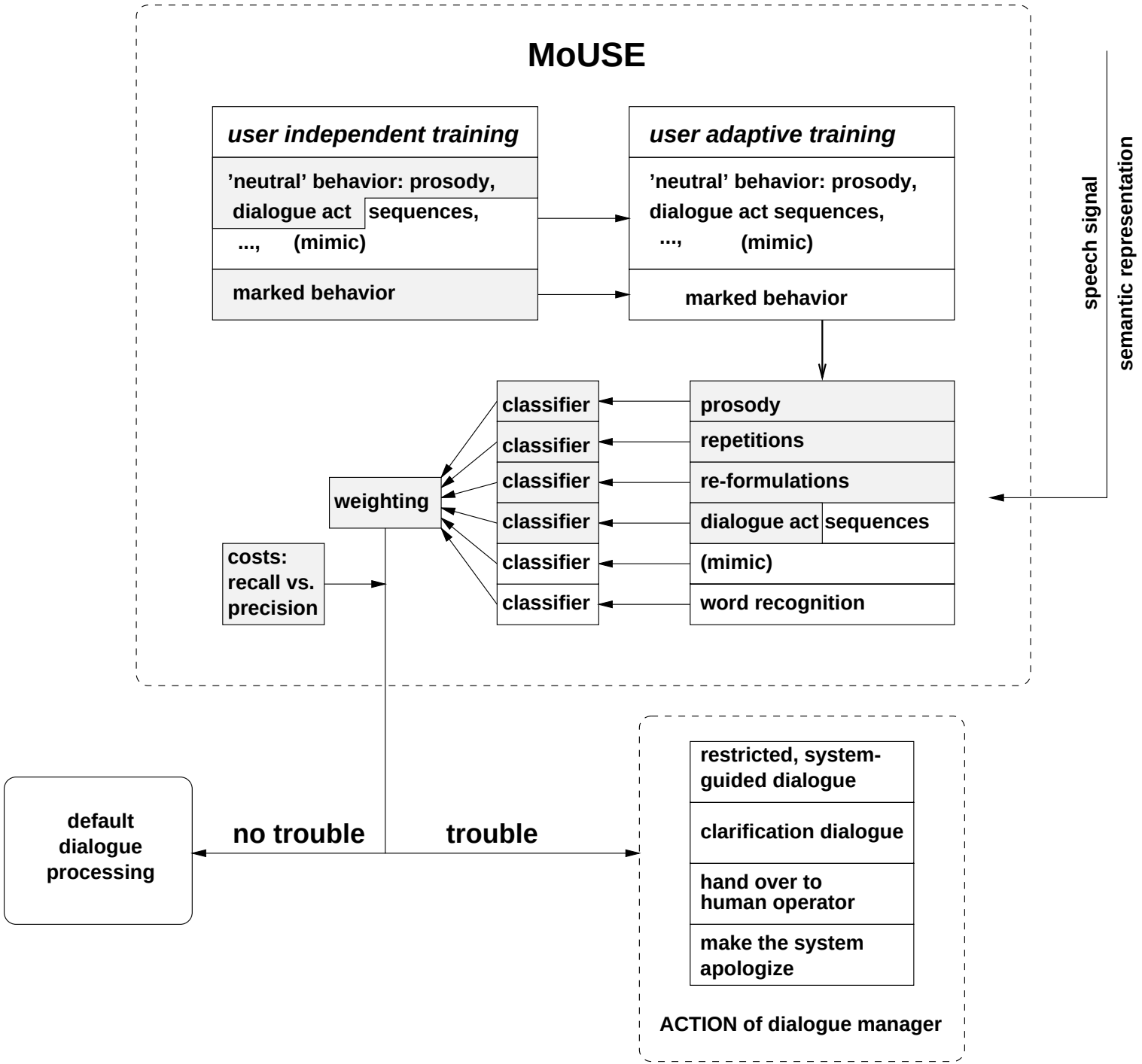
  $\Rightarrow$
- multi–knowledge approach:
  M*onitoring* **o***f* **U***ser* **S***tate [especially of]* **E***motion* MoUSE

WHAT CAN BE DONE?

- **Monitoring of User State [esp. of] Emotion (**MoUSE**)**
  - context-independent strategies
    - prosody
    - lexical material (cursing, etc.)
  - context-sensitive strategies
    - repetitions
    - re-formulations
    - metalanguage, self-talk
- detection of Trouble in Communication

MOUSE: POSSIBLE INTEGRATION IN AN AUTOMATIC DIALOGUE SYSTEM




MoUSE
user independent training
'neutral' behavior: prosody,
dialogue act sequences,
..., (mimic)
marked behavior
user adaptive training
'neutral' behavior: prosody,
dialogue act sequences,
..., (mimic)
marked behavior
speech signal
semantic representation
classifier
classifier
classifier
classifier
classifier
classifier
weighting
costs:
recall vs.
precision
prosody
repetitions
re-formulations
dialogue act sequences
(mimic)
word recognition
default
dialogue
processing
no trouble
trouble
restricted, system-
guided dialogue
clarification dialogue
hand over to
human operator
make the system
apologize
ACTION of dialogue manager