

ANALYSIS OF USER BEHAVIOR UNDER ERROR CONDITIONS IN SPOKEN DIALOG



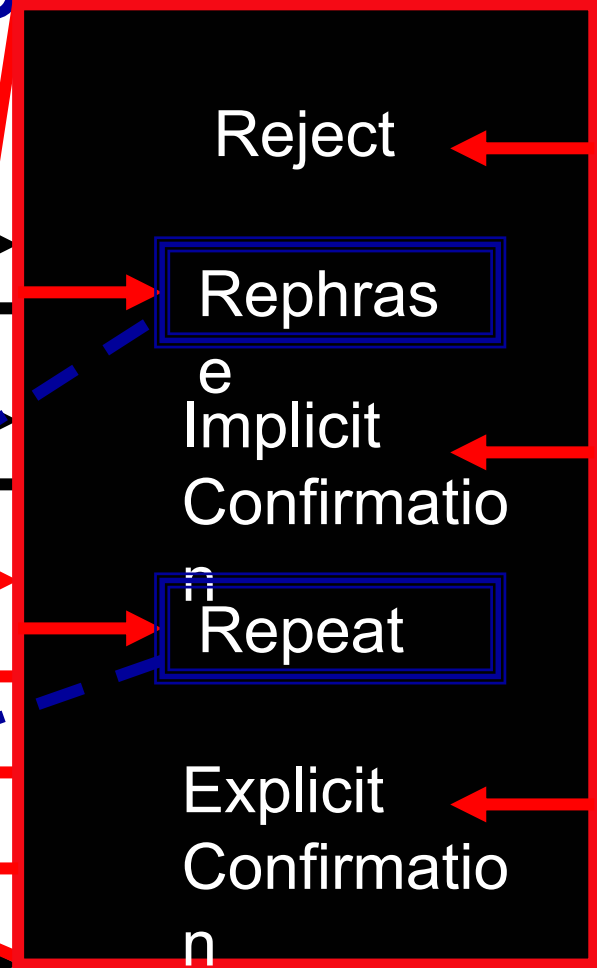
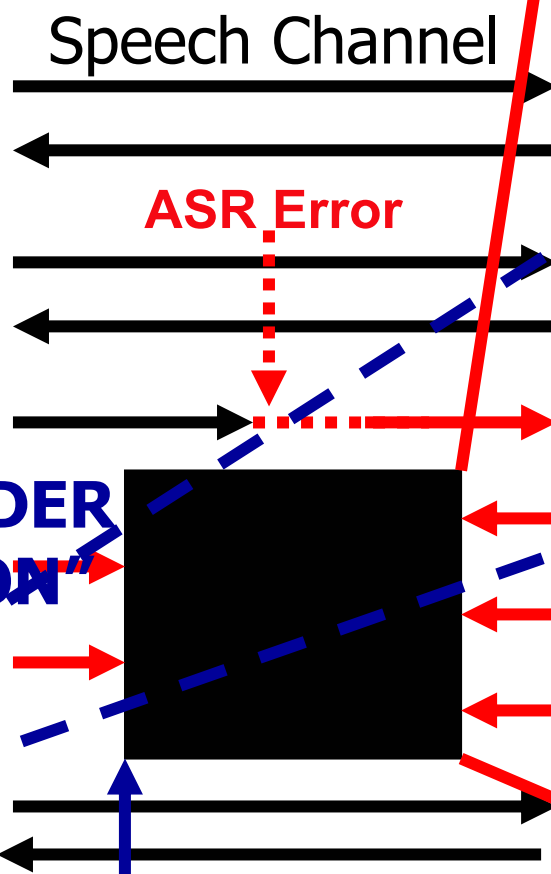
JongHo Shin, Shrikanth Narayanan,
Laurie Gerber, Abe Kazemzadeh, Dani
Byrd,

Speech Analysis and Interpretation
Laboratory,

University of Southern California.



WHAT IS THE "USER BEHAVIOR UNDER AN ERROR CONDITION" ?



"USER BEHAVIOR UNDER AN ERROR CONDITION"



Get Back on the Track

MOTIVATION



- The error condition in spoken dialog system
- Realistic case studies can be good indicators for how user and the system behave under error conditions
- The understanding and modeling user behavior to enable realistic optimization of dialog strategies

DATA USED



- Annotated travel arrangement dialogs
 - DARPA Communicator project recorded in 2000
- 85 experimental subjects interacted with 9 different “travel agent” systems
- Worked with 141 dialogs out of 765 dialogs
- Example
 - System said:** And on what date did you want to fly?
 - Recognizer heard:** NOVEMBER FIRST
 - User said:** NOVEMBER FIRST



ANNOTATION

(http://sail.usc.edu/dialog/model_tags)

- Manual tagging scheme
 - Total 23 tags monitoring 3 dimensions of dialogs
 - User behavior, system behavior and task status
 - Extended version of DAMSL to include additional behaviors

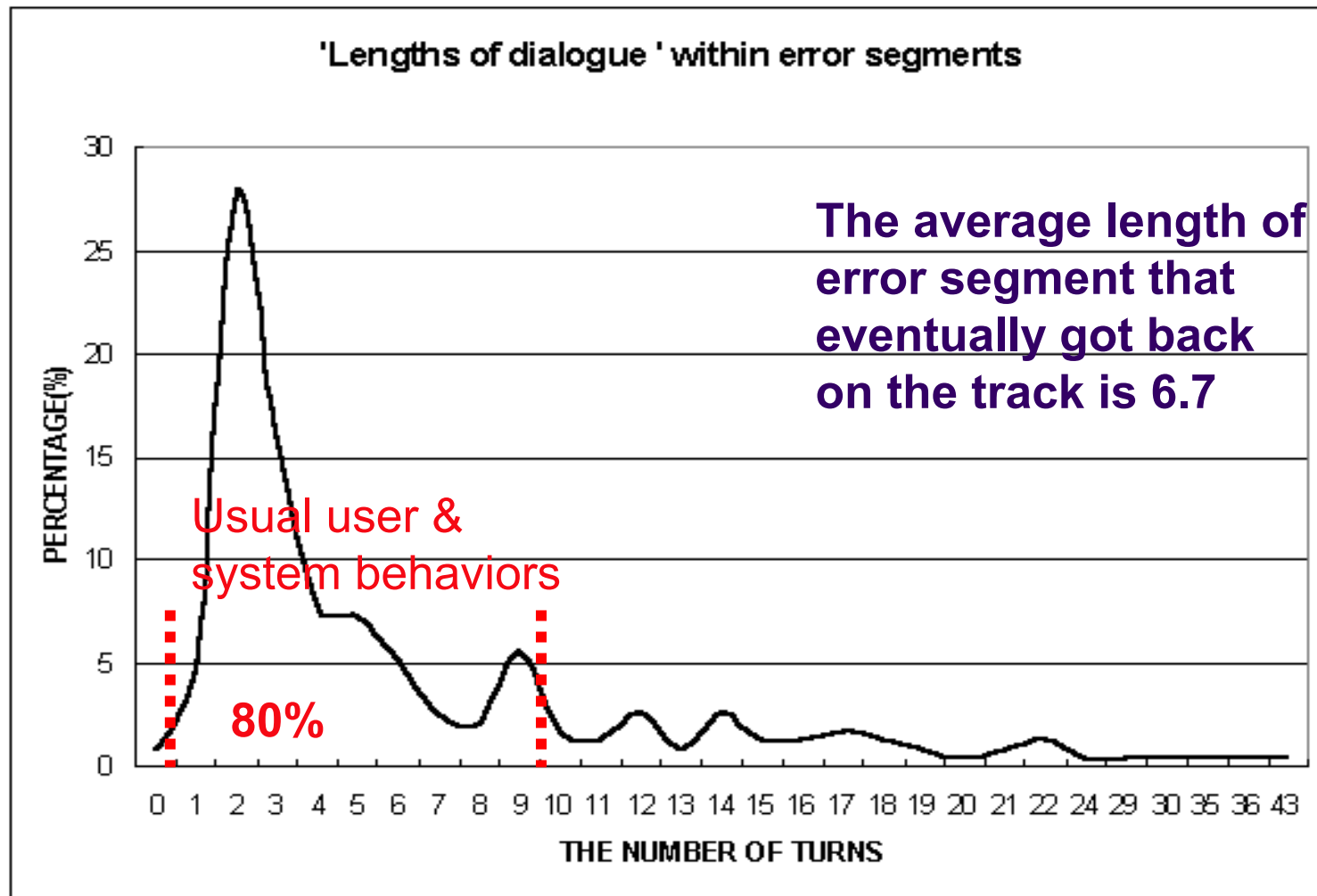
- Definition of Error segment
 - Beginning of segment : Initial ASR error
 - End of the segment : Back-on-track(BOT), End of the dialog, or Hang up

- Example
 - System said: What is your destination ^^sysrepeat
 - Recognizer heard: PHOENIX ARIZONA
 - User said: Phoenix Arizona ^^repeat

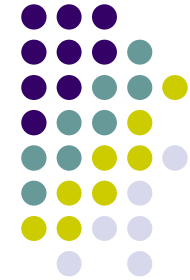


REALISTIC CASE STUDIES

HOW MANY TURNS FOR THE ERROR RECOVERY ?



ERROR AFFECTS THE TIME AND SUCCESS RATE FOR ERROR RECOVERY



User's error perception	# of error segments	Average error length for BOT	Average error length for not BOT	BOT %
Reject	35	6	7.8	83%
Implicit	25	9.6	14.6	68%
Repeat	21	5.8	13	90%
Explicit	10	5.5	8.75	60%
Non-sequitur	9	6	7.5	77%

High frequency

More time

Less success rate

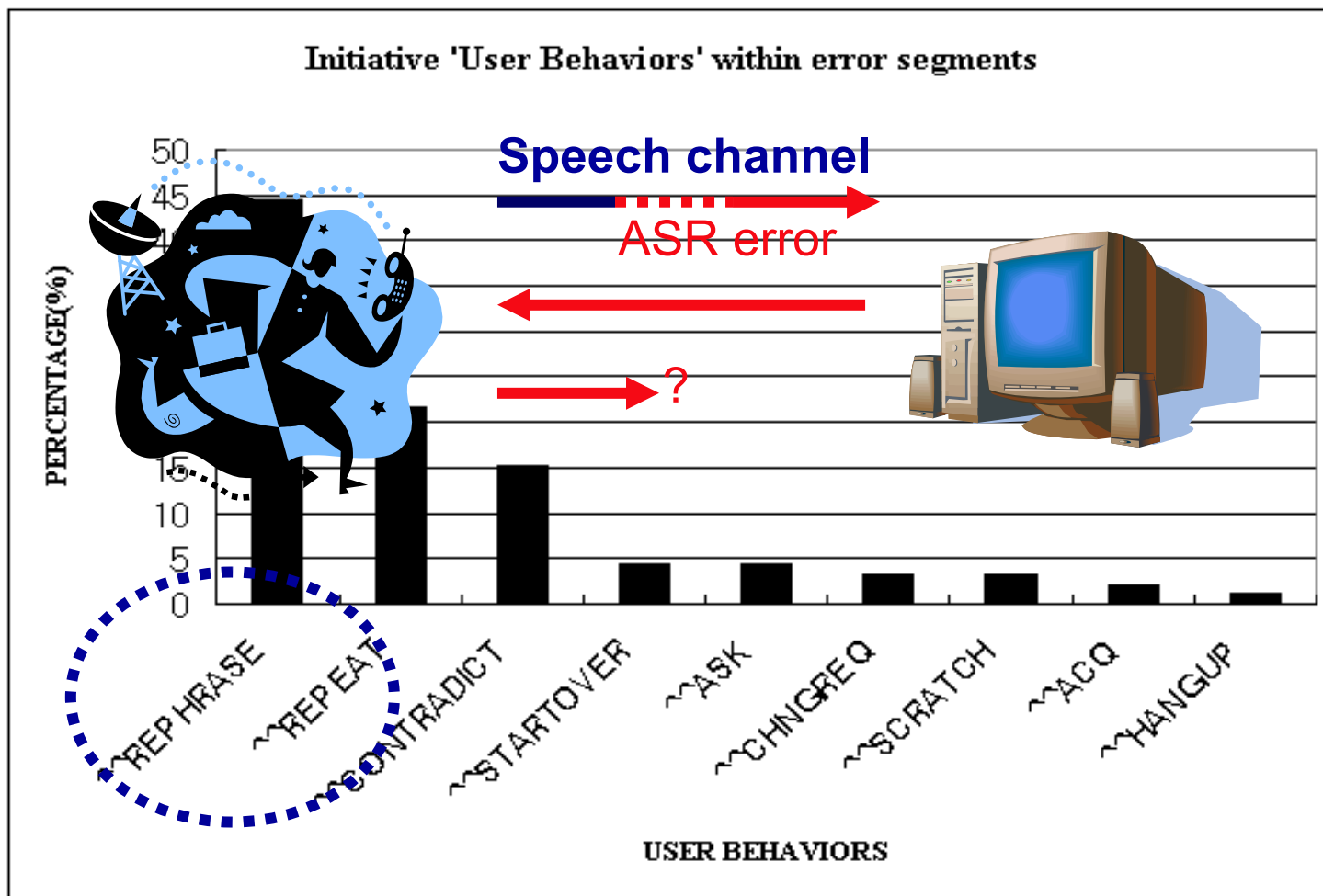
DIFFERENT USER BEHAVIORS IN SUCCESSFUL AND UNSUCCESSFUL ERROR RECOVERY



frequency normalized for length of errors	User strategy in Errors that got back-on-track
0.130	Repeat
0.117	Rephrase
0.077	Contradict
0.055	Start over
0.045	Ask
0.022	Change request
0.015	Scratch
0.005	Acquiesce to error

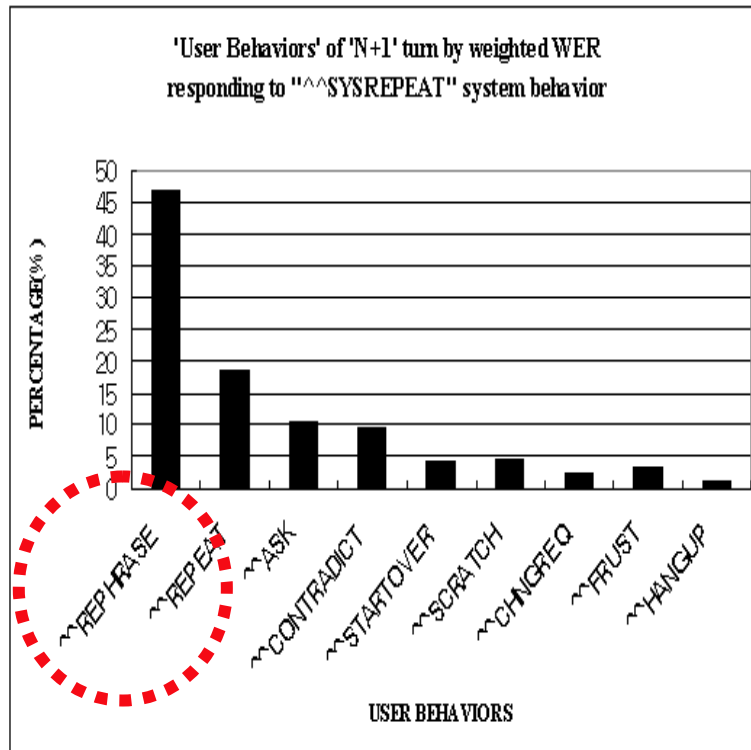
frequency normalized for length of errors	User strategy in NON-back-on-track
0.114	Repeat
0.102	Contradict
0.071	Rephrase
0.055	Hang up
0.031	Start over
0.024	Ask
0.012	Scratch
0.012	Acquiesce to error

WHAT IS THE FIRST USER BEHAVIOR UNDER ERROR CONDITION ?

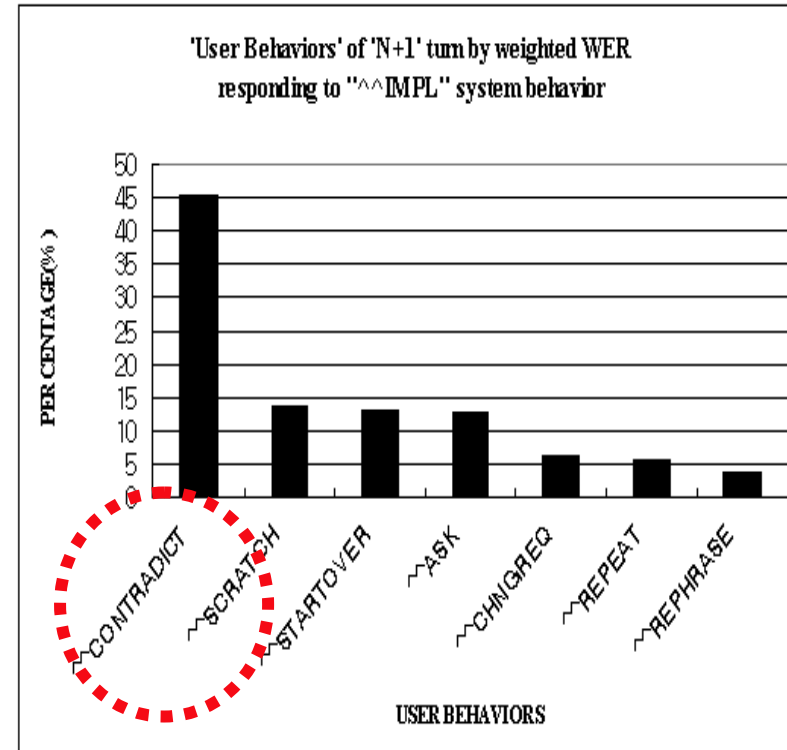


USER BEHAVIORS UNDER TWO SITUATIONS

- SYSTEM KNOWS ERROR AND DOESN'T KNOW ERROR



System repeat



Implicit confirmation

WHEN DOES USER USE INITIATIVE MORE ?



User Initiative tag	Frequency in error segments	Frequency in NON-error segments
ASK	0.0319	0.0060
Contradict	0.0707	0.0121
General Initiative	0.1647	0.0424

User initiative behavior is significantly more in “Error segments” than “NON-Error segments”

SUMMARY



- Empirical analysis and modeling using real cases helps to illuminate user behavior patterns
- General user behaviors can be observed within 10 dialog turns between human and machine
- Under error conditions, user tends to use initiative more and most frequent user behaviors are “Rephrase” and “Repeat”
- “Implicit confirmation” makes users less likely and take a longer time to get back on the track
- For successful error recovery, user uses more “Rephrasing”, less “Contradicting” and “Termination” of the error episodes rather than “Repairing” a chain of errors



ONGOING WORK

- Incorporate user behavior priors in the user modeling

User ID	# of dials	Errs/dial	BOT%	Avg length of error segment
1	9	1.4	0.69	8.9
2	9	1.4	0.76	8.9
3	8	2.9	0.87	7.8

- User modeling and optimization of dialog strategies under error condition
- Two agents game for spoken dialogs