Expecting the Unexpected: Warehousing and Analyzing Data from ITS Field Use

W. Lewis Johnson, Naveen Ashish, Stephen Bodnar, Alicia Sagae

Key Ideas and Background

- Learner behavior with a deployed system may not be predictable from pilot studies
- Learners are "not like me," i.e. not like developers (Koedinger, 2001)
- Longer periods of use, more user sessions, user-selected ordering of curriculum material
- Components like ASR perform differently in the field than in lab tests
- To study the differences: ***Hoahu***[1] data warehousing system for extracting learner behavior data from logs
- Explore statistics concerning user behavior
- Obtain overall measures of user/system performance
- Discover unexpected behavior patterns
- Identify opportunities for design improvements

1 From the Hawaiian verb meaning "to collect". Alelo draws inspiration from Hawaii as a cross-cultural crossroads, using Hawaiian names for many of its technologies.

Application Example

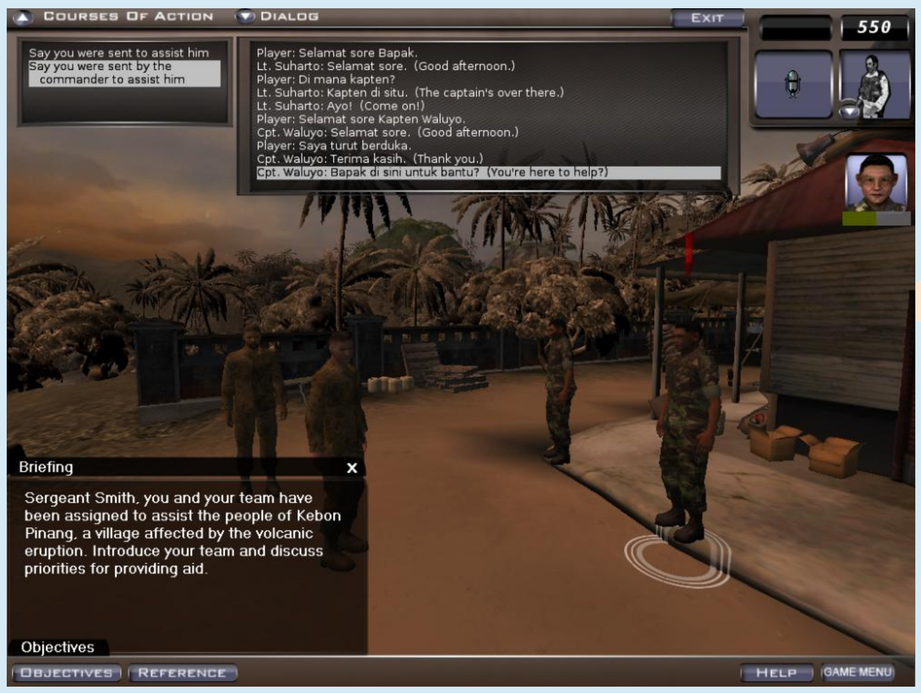

Fig 1. Screenshot of the dialog environment in Operational Indonesian Language and Culture Training System

	NPC:	Votre passeport, s'il vous plaît	(1.1)
On hearing this Utterance Attempt, the system fails to recognize the player's speech	**Player:**	**Bien sûr, voilà.**	(1.2)
	NPC:	Je ne comprends pas.	(1.3)
The system recognizes the player's speech but fails to find an appropriate response	**Player:**	**D'accord.**	(1.4)
	NPC:	[no response]	(1.5)
The system recognizes and responds, marking the start of a new turn	**Player:**	**Voilà**	(1.6)
	NPC:	Quel est votre nom?	(2.1)

Fig 2. Excerpt from a dialog between a learner and non-player character (NPC)

- ITS for languages in a culture- and task-specific setting (Johnson & Valente, 2008)
- Learners interact with computer-controlled Non-player characters in spoken dialog exercises with ASR technology
- ***Dialog breakdown***: the system fails to understand the learner's previous turn, yielding an inappropriate response

Hoahu Technology

- Anonymization: replace personally-identifiable user names
- Extract-Transform-Load (ETL): apply templates to mine behaviors of interest from the logs
- Relational database (DB) representation
- Annotation tools add layers of information to the DB

Results

- Pilot users (Pi): employees, intermediate speakers
- Field users (Fd): Naval personnel, volunteers, self-study
- Severe breakdown rate (3+ consecutive breakdowns):
  - Pi: 23% of dialogs, Fd: 14.4% of dialogs
- Very severe breakdown rate: (4+):
  - Pi: 18.6% of dialogs, Fd: 7.9% of dialogs
- Mean utterances per turn:
  - Pi: 2.3; Fd: 1.6

Future Work

- Follow-up work on annotation: isolate system-triggered breakdowns from learner communication errors (Workshop of UMAP 2010)
- Server-client logging architecture to adapt to multiple platforms: web, desktop, handheld; single- and multi-user settings
- Regression testing to drive ASR improvements

alelo™

Visit us on the web at www.alelo.com