Beyond Words: Identification of Back-Channel Communication Rules in Arabic and Development of Training Methods

Funded by DARPA’s Defense Sciences Office, via the Department of the Interior under Grant NBCH1050025 to the Information Sciences Institute of the University of Southern California under Subaward PO#105814.

Quarterly Progress Report 5

July 27, 2006 – October 26, 2006

The University of Texas at El Paso
Nigel G. Ward, Principal Investigator

Computer Science, UTEP, 79968-0518
915 747-6827, fax 915 747-5030
nigel@cs.utep.edu

1 Tasks from the Project Charter

1.1 Rule Discovery

Our first task was to “discover the basic rules governing back-channel behavior (production of uh-huh, etc. as a display of serious listening) in Eastern Arabic”. Discover of the three prosodic cues involved was largely completed in the fourth quarter.

In the fifth quarter we began to explore the role of gestures as additional cues for back-channels. Using Anvil we labeled the gestures seen just before each back-channel in the corpus. Several gestures are involved, but head nods from the speaker appear to be the best (but still weak cue) for back-channels. We also began to explore the role of gestures as substitutes for verbal back-channels. Here also head nods seem to be the most frequent gesture for displaying understanding and attention.

In the sixth quarter we plan to quantify these tendencies, and to submit the results for publication.

In the fifth quarter we also donated a copy of the UTEP Iraqi corpus to AFRL/HECP at the request of Tim Anderson.

1.2 Toolset Production

As stated in the project charter, we will “produce a toolset for the automatic discovery of new rules in new languages and cultures”.
In the fifth quarter we completed a plug-in to enable Wavesurfer to work well with stereo dialog data (using mostly NSF funds). This is described in Technical Report UTEP-CS-06-04.

We also organized and held a very successful Special Session, “The Prosody of Turn-Taking and Dialog Acts”, as part of the Interspeech 2006 conference in September; this brought together about 50 researchers in this area, including some of the major speech analyses tool builders. My talk was on Methods for Discovery, as noted in the last report.

As a test of the discovery methodology, an undergraduate discovered and quantified the prosodic rules governing back-channelling in Spanish. This took over 1000 person-hours, which underlines the need to produce a discovery toolkit to fully support such analysis. The rules she discovered for Spanish do, however, constitute a new finding, and will be presented orally in the sixth quarter at the High Desert Linguistics Symposium in November in Albuquerque, as “Prosodic Cues that Lead to Back-Channel Feedback in Northern Mexican Spanish”.

Hedging our bets, we also continued to explore a 100% automatic approaches to the discovery of new rules. Earlier we found that ramp and random filters applied to both the energy and the fundamental frequency yielded good predictions of back-channel opportunities. We are now developing a way to explore the space of features using evolution strategies. A preliminary study suggests that this may give better performance than hand-crafted rules, and may also lead to more concise descriptions. In the sixth quarter we will determine if this result is real and if so publish it and exploit it.

In the sixth quarter, if time permits, we plan to continue to lay groundwork for the production of the toolset, including requirements analysis, and prototyping. Other activities, including a decision as to which tool to use as a base for further development and the formation of partnerships with other developers of such tools and users of such tools, will come later.

1.3 Training Methods

The charter also calls for us to develop methods for training American speakers to understand and emulate the rules in Arabic, and to apply these training methods in ISI’s Tactical Language Trainer (TLT).

In the fifth quarter we added some features:

1. An improved scoring mechanism.

2. A visual indicator to give immediate feedback on whether a back-channel produced by the learner was timed correctly or not.

3. A preliminary training phase in which the learner is initially given with the relatively simple task of listening to short audio clips and indicating whether or not he heard the back-channel cue in each one.

We also surveyed related work in intercultural communication, interlanguage pragmatics, and turn-taking. This background, the system design, and the training sequence were written
up in draft form. An experimental protocol was written up, submitted to the IRB, and approved.

We presented this work so far, under the title “Learning to Listen: A Back-Channel Trainer for Arabic”, at two venues: the NLP Group at the University of Pittsburgh in September and the Virtual Humans Group at USC’s Institute for Creative Technologies in October.

We have also submitted an abstract to the 2007 Calico (Computer Assisted Language Instruction Consortium) Conference.

**In the sixth quarter** we will do the experiments. As noted in the last report, our hypotheses are that using the system for about 10 minutes increases the learner’s:

- ability to notice (some of) the dialog cues active in Arabic
- awareness of the importance of demonstrating attention while listening
- ability to be an effective listener in Arabic
- confidence that they could politely engage an Arabic speaker
- awareness that Arabic dialog cues differ from those of English

The results will be slotted into the draft article and submitted for publication, probably to the journal *Computer-Assisted Language Learning*.

This is our top priority for the sixth quarter.

In the sixth quarter we will explore ways to integrate non-verbal training into the classroom, at the request of the chair of Linguistics who has asked for our input on this as part of the development of an Arabic curriculum at UTEP, to start in the Spring semester.

Also as soon as ISI is ready to work on this, we will assist them with the integration of the tool into TLT, including not only the technical aspects, but also in tuning the presentation to be better for soldier-learners.

### 1.4 Emulating Back-Channel Behavior

As part of the training sequence, we include a phase in which learners try to produce the prosodic cue, to make sure that they understood what it is.

**In the fifth quarter**, we added a parameter to our system to control the utterance size this responds to; this made it more appropriate for learners, who typically like to test their production skills with just a single-word utterance. In future, we may port this from Linux to Windows and integrate it with the ABC Trainer, although this is not currently a priority.

### 1.5 Evaluating the Importance of Proper Back-Channel Behavior

In addition to the experiments noted above, we plan to do another set of three experiments, to evaluate three hypotheses:
1. The pitch down-dash is a cue to back-channels in Arabic, but is not perceived as such by Americans.

2. Responding to pitch down-dashes with a back-channel is an important part of being a good listener in Arabic.

3. Good back-channeling matters more than good pronunciation, for giving a good first impression.

In the fifth quarter, we simplified the experimental design, determined the statistical tests to use and ran a small pilot study. We also reviewed work in related fields — linguistic pragmatics, interactional sociolinguistics, intercultural communication studies, interpersonal psychology, nonverbal communication, international relations, dialog systems engineering, artificial intelligence, and language teaching — and determined that this work will be (if it succeeds) an advance with respect to all of these areas.

In the sixth quarter, we will complete this work: preparing the audio stimuli, writing the precise protocol, recruiting subjects, doing the experiments, and writing up the results. This is our second priority.

2 Tasks Noted in Previous Quarterly Reports

2.1 Endpointing

As noted in the first quarterly report, an understanding of how turn-ends are signaled in Iraqi Arabic would make the Trainer more compelling and also allow the teaching of more turn-taking skills to users. In the fifth quarter we used machine learning techniques to create an initial system for predicting English turn ends before they actually happen and measured its performance. In the sixth quarter we will continue improving this until it meets or exceeds state-of-the-art performance, and then apply this to the Iraqi data. This is not on any critical path.

2.2 Inter-Cultural Misinterpretations

As noted in the second quarterly report, the fact that Americans can find Arabic grating may be due in part to the prosodic features we have uncovered. That is, we hypothesize that these prosodic patterns are perceived negatively by Americans, but not by Arabs.

In the fifth quarter we slightly refined our experiment design.

In the sixth quarter we will do the experiment, probably together with the experiment for evaluating the importance. This is our third priority.

3 References

Project documents and software are available at the private URL, http://www.cs.utep.edu/nigel/dso/.