This is a preprint of an article submitted for consideration in the journal *Computer Assisted Language Learning*, published by Taylor & Francis. *CALL* is available online at `http://journalsonline.tandf.co.uk/`.

August 21, 2007

# Learning to Show You're Listening

**Nigel G. Ward, Rafael Escalante, Yaffa Al Bayyari, Thamar Solorio**

**The University of Texas at El Paso**

We thank Lewis Johnson, Jon Amastae, David Novick, Diane Litman, Bill Lucker, Ralph Chatham, Yoshiki Hayashi, Origa Nagai, and Olac Fuentes. This work was supported in part by the Defense Advanced Research Projects Agency and in part by the National Science Foundation under Grant No. 0415150.

## Abstract

Good listeners generally produce back-channel feedback, that is, short utterances such as *uh-huh* which signal active listening. As the rules governing back-channeling vary from language to language, second-language learners may need help acquiring this skill. This paper is an initial exploration of how to provide this. It presents a training sequence which enables learners to acquire a basic Arabic back-channel skill, namely, that of producing feedback immediately after the speaker produces a syllable or two with a sharply falling pitch. This training sequence includes an explanation, audio examples, the use of visual signals to highlight occurrences of this pitch downslope, auditory and visual feedback on learners' attempts to produce the cue themselves, and feedback on the learners' performance as they play the role of an attentive listener in response to one side of a pre-recorded dialog. Experiments show that this enables learners to better approximate proper Arabic back-channeling behavior.

## 1 Introduction

To be a good listener you have to be able to show you're listening. In dialog this includes the active display of attention, interest, understanding and/or willingness to let the other person continue. This is accomplished in part with back-channels, also sometimes known as "response tokens", "reactive tokens", "minimal responses", and "continuers", that is, short utterances produced while the interlocutor has the turn. In English, these are typically short utterances such as *uh-huh*.

One problem for learners is knowing when it is appropriate to produce a back-channel. Many factors are involved, including both speaker-related factors and listener-related factors. A listener is free to produce back-channels based on his/her own understanding and intentions, but these back-channels are especially welcome at certain times in the dialog, and these times are determined in part by paralinguistic factors. Specifically, these times are indicated in part by intonation: in several languages there is a prosodic signal, that is a "cue" involving a specific pitch pattern, that a speaker can use to indicate when he/she welcomes a back-channel from the listener (Ward & Tsukahara, 2000; Fujie *et al.*, 2005; Wesseling & Van Son, 2005; Ward & Al Bayyari, 2007).

In many languages back-channels are very common. For example, in casual English conversation, some 20% of the utterances are back-channels (Shriberg *et al.*, 1998), and they appear about 4

times per minute (Ward & Tsukahara, 2000). For second language learners also, command of this form of interaction is important. A learner who lacks back-channeling skills, even if a master of the vocabulary and grammar, can easily appear uninterested, ill-informed, thoughtless, discourteous, passive, indecisive, untrusting, dull, pushy, or worse. For example, not only do Japanese back-channel twice as often as Americans (Maynard, 1989), but the response to the prosodic cue from the speaker is typically twice as fast (Ward & Tsukahara, 2000). The potential for awkward intercultural interactions here is clear, and specific problems are well attested (LoCastro, 1987; Ikeda, 2004).

This paper addresses the question of how to teach proper back-channeling behavior in Arabic; specifically, how to help students learn to detect the prosodic cue and respond to it swiftly. Section 2 overviews research and practice in the teaching and learning of turn-taking behaviors in general and back-channeling in particular. Section 3 presents our formulation of the learner's task. Section 4 discusses back-channeling in Arabic and the prosodic cue used by speakers to welcome a back-channel. Section 5 presents our systems for learning to produce back-channels in response to this cue. Section 6 presents our experimental method for evaluating the effectiveness of these learning activities, Section 7 presents the results, and Section 8 identifies directions for further work.

## 2 Related Work

Several skeins of research bear on the question of how to teach back-channeling skills.

### 2.1 Back-Channeling and Related Phenomena

Back-channeling is part of a cluster of related activities and skills, variously referred to as "turn-taking", "interactive functions", "how to keep the communicative channel open" (Canale & Swain, 1980, pg. 25), "conversational management" (Hurley, 1992, pg. 266), "discourse functions" (Chun, 2002), and "track 2", namely that involved in creating a "successful communication" as opposed to conveying the "official business" of the exchange (Clark, 1996, pg. 242). This cluster includes the ability to handle expressions of "finality vs. non-finality; marking shared knowledge, presuppositions; turn-taking, interrupting; discourse management," among others (Chun, 2002).

Turn-taking in particular is the process by which dialog participants manage who should talk when. Within the study of turn-taking, back-channeling has often been seen as a central, even prototypical phenomenon (Yngve, 1970; Duncan & Fiske, 1985; Schegloff, 1982). In the 1970s it seemed as if linguistics and psychology were on the verge of discovering the details of how this is done (Yngve, 1970; Sacks *et al.*, 1974), but accurate description of the signals that people use to manage this has proven elusive, and only today are these finally being identified.

### 2.2 Learning How to Back-Channel in Another Language

The importance of back-channeling as something language learners need but lack has often been noted (White, 1989; Rost, 1990; Allwood, 1993; Horiguchi, 1997), and acquisition of the back-channeling conventions of a second language is known to be difficult. Problems with turn-taking more generally are not uncommon, even among those who have spent years in a foreign language environment and are otherwise very advanced. Lacking these skills in the second language, learners frequently fall back on a rigid, pause-based turn-taking style. Although this is adequate in many contexts, it can appear cold and formal in others, and can be limiting to the speaker's social and communicative success.

Although learners often have trouble choosing an exactly appropriate back-channel (which is not surprising given the multiple levels of meaning that can be packed into a single feedback utterance

(Brennan & Hulteen, 1995) and the "varied work" that back-channels can do (Gardner, 2001) ), the more salient problems are those of timing: failing to produce back-channels when they should or producing them when they shouldn't.

Thus there is a need to teach back-channeling timing patterns and rules. Today, unfortunately there are only a handful of known quantitative facts that are useful for learners, for example: that overlaps are longer and more common in Spanish than in general American English (Berry, 1994), that back-channels are very frequent in Japanese, less frequent in English, Spanish, German, and Arabic, and very infrequent in Chinese (Maynard, 1989; Clancy *et al.*, 1996; Heinz, 2003; Acosta, 2004; Ward & Tsukahara, 2000), that back-channels in Japanese come more swiftly than those in English (Ward & Tsukahara, 2000), and that certain prosodic features cue back-channels in American English, Japanese, Mexican Spanish, Korean, and Finnish (Ward & Tsukahara, 2000; Rivera & Ward, 2007; Young & Lee, 2004; Ogden & Routarinne, 2005).

In general, for lack of an understanding of the "causal dynamics" (Allwood, 1993), that is, the detailed rules involved, those wanting to teach these skills have been forced to fall back on exercises in which the learners observe and comment on turn-taking behavior (Demo, 2001; Berry, 2003). Although teaching learners to be careful observers of language and motivating them to pay attention to details of turn-taking behavior are clearly of long-term value, it seems that the specific effectiveness of such exercises has not been evaluated. It may be possible for some learners to figure out for themselves, from such exercises, the rules that govern such behavior — probably not consciously, but perhaps well enough to acquire the skill — but in general such exercises are probably not adequate to give learners information specific enough to apply.

Another fallback expedient is to describe the desired behavior at a higher level. For example some descriptions of Japanese language and culture point out that Japanese speakers tend to be less aggressive, to emphasize harmony, to be courteous, and to empathize (LoCastro, 1987). For Arabic the list includes tendencies to speak using elaboration, repetition, emotional expressions, indirectness and to give importance to values such as hospitality and pride (Almaney & Alwan, 1982; Ellis & Maoz, 2006). It is easy to talk to learners about such tendencies, and suggest ways in which these are reflected in turn-taking behavior (Mizutani, 1981; Matsuda, 1988; Yamada, 1992; Mukai, 1999; Ohama, 2000). Doing so is clearly useful, but leaving it to the learner to make the connections between the high-level cultural features and the specific actions they map to is again probably not adequate in general. Another problem is that attempts to follow such cultural norms are likely to use the learner's existing inventory of language behaviors, but behaviors that count as being polite (for example) in the learner's language may not be perceived as such by speakers of the other language (Young & Lee, 2004).

### 2.3 Learning Second-Language Prosody

Turn-taking relies heavily on prosodic signals. Prosodic ability in general is a major component of second language competence, and recent years have seen major advances in techniques for teaching this (Chun, 2002). Some aspects, such as the various prosodic markings of questions, are relatively well known, and there are established techniques for helping learners both detect and produce such prosody. The prosody of turn-taking itself, however, appears not to have received much attention. The essential difference is that these discourse functions are intrinsically interactive, which affects the pedagogy in two ways. First, although one can identify question types, for example, in isolation, and practice them in isolation, turn-taking behaviors only exist, and can only be presented, in dialog contexts. Second, a response to a question, for example, can be delivered at any time, within reason, but a back-channel response must be produced within a fairly narrow time window for it to count as a back-channel at all. Indeed, a late response can disrupt the flow of the conversation and confuse the

speaker. Thus it is necessary to develop learners' "conversational reflexes" so that they can respond within a fraction of a second.

Learning these behaviors is quite different from most language production skills: here the aim is to help the learner, not with "what should I say" or "how should I say it", but with "when exactly should I say it". This sort of precision timing has apparently not previously been addressed in language teaching.

The relative lack of knowledge regarding pedagogy in the prosody of turn-taking can be ascribed in part to the above factors. Another factor may be the tendency of researchers to focus on speaking and listening as discrete skills, with "coordinated" use receiving less attention (Gardner, 1998).

### 2.4 The Need for Computer-Based Training

A CALL solution for conversational behaviors makes sense because they are hard to acquire from classroom learning alone. Pair work and teacher-fronted activity can help develop some kinds of listener display (Ohta, 2001, pg. 210), but are probably not ideal for highly interactive behaviors. Ideally each learner should be provided the opportunity to learn and practice one-on-one with a native speaker conversation partner, with a teacher observing and providing critiques, however this is seldom possible. Thus there is a need for systems that can simulate conversational partners.

### 2.5 An Early Experiment

At the University of Tokyo, Atsuko Kondo, Yoshiki Hayashi, Origa Nagai, and the first author did a preliminary exploration of issues in teaching back-channeling (Ward *et al.*, 2001). The results were never published and so are summarized here.

The aim was to help Japanese become better listeners in English. Three facts are relevant: that the cue to back-channels in English is a region of low pitch lasting 110 milliseconds or more (essentially the same as in Japanese), that back-channels in English are not given as swiftly after the cue as in Japanese, and that back-channels in English occur about half as often as in Japanese.

The system was designed to train learners by having them take the listener role. The system played one side of a conversation and the learner was to produce back-channels at appropriate points. Reasoning that a reflex behavior could best be developed by providing immediate rewards, learner vocalizations produced at appropriate times for back-channels were flagged with a blue rectangle on the screen, inappropriate ones with a green rectangle, and missed opportunities with a red rectangle. These were displayed along a timeline, which extended to the right as the track progressed.

Before using the system, learners were told the three facts noted above. Three different dialogs sides were included in the system, and one was presented three times so that the learners would have sufficient opportunity to learn to master the desired behavior for at least that one dialog.

Evaluation was done by measuring the percentages of back-channels produced at appropriate places and comparing these scores before and after training. Although we expected improvements, in fact the results were not significant, and indeed only 9 of the 13 learners showed any improvement. User comments were informative and later affected the design of the Arabic training system, as will be seen.

## 3 Learning Scenario

To recap, our aim is to build a system to help learners acquire the ability to back-channel. This is a very general skill, useful in many contexts, however a specific scenario was used to help guide system design. Reflecting the interests of our sponsor, we chose to work on back-channeling in Arabic. This

worked out well in that this involves a non-trivial learning task, since the prosodic cue to back-channels in Arabic is quite unlike that seen in American English, and since Arabic is a language which many Americans want or need to learn, and where communicative competence is often seen as their only need (Kaplan *et al.*, 1998; Johnson *et al.*, 2005).

Back-channel training was planned to be incorporated into a one week intensive course for US military personnel. This course is based on a game-themed CALL system, in which the player takes the role of an officer trying to find the local leader and establish rapport with him (Johnson *et al.*, 2005). Some of the situations involve training in how to make polite greetings and exchange pleasantries or smalltalk before requesting information, thus there was a good fit with our aims, since the need to listen while showing interest and polite attention during smalltalk requires back-channeling skills.

We decided that our back-channel trainer should address only the audio aspects of the skill. This is not because back-channeling is an audio-only phenomenon. In fact a listener head nod can often substitute for a verbal back-channel, however this is as true in English as in Arabic and also technically challenging to detect, so we decided to leave out this aspect. It is also the case that a speaker head nod or hand gesture can sometimes serve as a back-channel cue or cue intensifier, however such visual cues are less important than prosodic ones (Al Bayyari & Ward, 2007), so we decided to leave out that aspect also. Thus our learning scenario resembles a telephone interaction rather than a face-to-face one. Of course training in the visual aspects of back-channeling would be interesting topic for future work.

We also decided that our back-channnel trainer should not assume that the learner knew any Arabic, for several reasons. First, it seems that back-channeling behavior involves two aspects, one involving detection of the prosodic cues marking opportunities to back-channel, and one involving the application of semantic and subtle pragmatic factors to determine which opportunities to take and exactly how to respond. We chose to teach only the first aspect, since the second aspect is largely beyond the ability of our envisaged learner population. There is also some evidence that the prosody is more important than semantic content as a determinant of when to produce back-channels (Fujie *et al.*, 2005). Certainly producing back-channels "automatically", based on prosody alone, without really paying attention, can be an effective strategy for appearing to listen, at least in the short term (Yngve, 1970). A second reason for assuming no knowledge of Arabic was that this let us avoid one complication that surfaced in the Tokyo experiments: some subjects reported that they were so busy trying to understand the English words they were hearing that they couldn't devote any attention to listening for the prosodic cue or responding. Use of very early-stage learners avoided this problem. A third reason was the feeling that teaching turn-taking before vocabulary may resemble the order in which natives acquire language, picking up some patterns of conversational exchange before acquiring lexical competence. A fourth consideration was the difficulty of locally obtaining subjects with partial knowledge of Arabic. Although we do not foresee that back-channel training for advanced learners would need to differ significantly, this would be an interesting topic for future work, as discussed below.

At this point we must acknowledge that it is not necessarily the case that non-natives always do best by complete socio-cultural adaptation to the native interaction style (Kasper & Blum-Kulka, 1993; Byram, 1997), however enabling learners to back-channel in the native interaction style is still a worthwhile option to give them.

# 4 Back-Channeling in Arabic

To back-channel well requires one to detect and respond to the prosodic feature that cues back-channels (Fujie *et al.*, 2005). This section presents the cue used in Arabic.

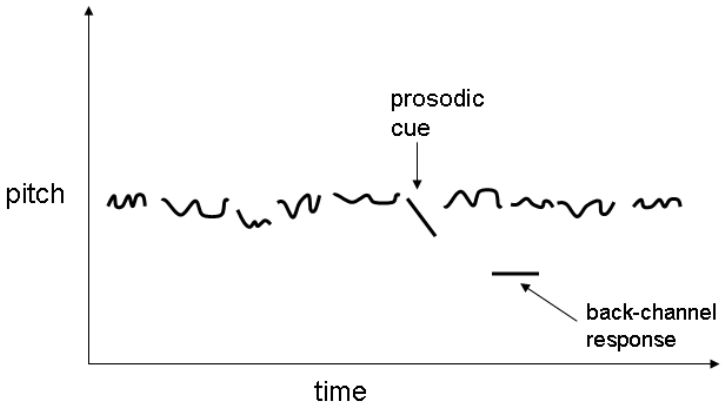


Figure 1: Schematic of a Feature Complex and a Back-Channel Response. The upper wavy line represents the pitch of the speaker's utterance, and the bottom line that of the listener's response.

In Arabic, back-channels are frequently preceded by a certain feature complex produced by the other speaker (Ward & Al Bayyari, 2007, 2006). The most distinctive feature of this feature complex is a region of pitch which is sharply falling: a "downdash", to use Bolinger's term (Bolinger, 1989). This fall is generally steady; almost linear when viewed in log scale. There are other common characteristics but these are less important. A back-channel often occurs in response to this feature complex, typically around 500 milliseconds later. Figure 1 illustrates.

[position Figure 1 about here]

Pauses often, but not invariably, follow the downdash. Being followed by a pause seems to make the cue stronger. Such pauses are not always directly following the downdash; an additional word may come between the cue and the pause. (Similar things are seen in English, where utterances ends do not always co-occur with prosodic cues, for example in cases of "post-completion", as in *"At the mall it was crazy. Just crazy."* where the prosodically marked turn end may be after the first word *crazy*, with the subsequent comment not intended to hold the floor.) Significantly, the pause is often so short that any back-channel response will necessarily overlap the next utterance of the speaker, meaning that learners also need to be able to produce back-channels while the other is speaking. Despite dialect differences, approximately the same feature complex is used in at least Egyptian and Iraqi Arabic. Detailed quantitative descriptions appear elsewhere (Ward & Al Bayyari, 2007, 2006).

The pitch downdash is certainly not the only factor affecting back-channeling behavior; downdashes and back-channels do not always co-occur. In a sense, this cue is an invitation to back-channel, but whether this actually leads to a back-channel response depends on many factors. In addition to meaning and gesture, as noted above, another important factor is individual differences; no two people have exactly the same conversational style.

This Arabic feature-complex is completely unlike the cue to back-channels used in American English, which is a longish region of low pitch (Ward & Tsukahara, 2000). In other respects back-channeling behavior in the two languages is similar: the overall frequency is similar, the delay between occurrence of a cue and the back-channel response is similar, and the semantic and pragmatic contexts

of back-channeling are also similar (Hafez, 1991). Given that most of our subjects were also familiar with Spanish, it is worth noting that the feature-complex seen in Spanish is different again, but again the overall frequency, delay, and pragmatic contexts of back-channels are similar (Rivera & Ward, 2007).

# 5 Teaching Back-Channel Skills

## 5.1 The Core Trainer

The core of our system was designed, like its predecessor, to play back a recorded conversation to a learner and give the user feedback on his or her performance. The main difference was that immediate visual feedback was left out, as users of the Tokyo system had commented that this was not very helpful, indeed, it could be frustrating to see a red flag on the screen when it was too late to do anything about it.

[position Figure 2 about here]

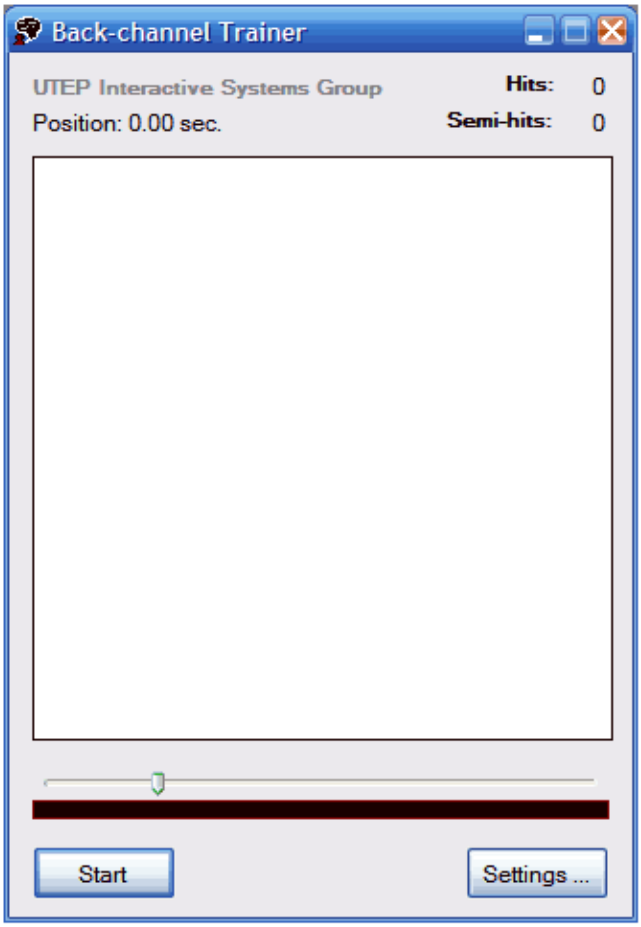


Figure 2: Screenshot of the Core Trainer Main Window

Thus the system is primarily audio-based, however, as seen in Figure 2, some information is presented to the user visually. This includes a black bar at the bottom which turns pink when the system detects a user vocalization; this is so that learners can tell if they are being heard, which is

important especially for those who tend to speak quietly.

In pilot runs several users commented that the system would be more useful if the display could give a visual indication of when to respond. We therefore added a visual cue, not so much to tell them when to respond, but to help them recognize places where the prosodic pattern appears. These places are flagged by having the large central rectangle turn green for 250 milliseconds approximately during the pitch downdash. The settings enable the user to turn this visual hint on until they feel ready to practice without it. The visual hint is of course turned off for the evaluation runs.

Users are told to produce back-channels, to just say anything. The experimenter further informs them that the English back-channels *mm* and *okay* are also common as Arabic back-channels, and reassures them that it is more important to just produce something rather than to produce any specific word.

Adopting an iterative design methodology, we began working with subjects as soon as the first version of this core system was ready. Sixteen subjects participated in this first set of pilot studies. Based on their comments and our own observations we made extensions and improvements, as described below, cycling through design-test-refine phases.

### 5.2 Tutorial

Detecting the prosodic cue, the downdash, seemed difficult for many in the pilot studies, so we decided to replace the brief explanation with a longer tutorial. This was developed by having a native speaker work with various learners in various ways, until she came up with a satisfactory way of explaining things, that is, an explanation that seemed to enable learners to distinguish pitch downdashes from the rest of the speaker's utterances. The examples provided with this explanation also served to familiarize the learners with the general rhythms of Arabic prosody, which were unfamiliar to our subjects.

This explanation was then converted into a multimedia presentation, Flash-based. This includes text on screen, with the text also being read out loud for reinforcement. The entire presentation lasts about 5 minutes, unless the learner chooses to replay some parts. The content of the presentation includes a brief explanation of the role of back-channels in communication and a description of the pitch downdash as a cue in Arabic. There iss also discussion of the role of pauses, pointing out that the cues occur before pauses frequently but not invariably, and that pauses by themselves are not a cue for back-channels. The presentation also includes many examples; in some of these the listener's attention is drawn to the presence of the pitch downdash by a simultaneous on-screen pulsating animation lasting about a half second. In contrast to the role of the visual signal used in the predecessor system, this was timed to overlap the cue.

A key part of the presentation was the examples, and these were ordered to help the learners, in several ways. For the first few examples, presentation of the entire phrase containing the cue was followed by just the cue itself, typically appearing on one word. The early examples were presented in a clear and familiar voice, the voice of the tutorial presenter, and later examples used a variety of speakers from the corpus. The early examples featured cues which occur before pauses, and later examples illustrated how the cues could also occur in the middle of an utterance. The animation was also introduced at this stage, as we wanted learners to focus on the audio at first, but had found that a visual signal was needed to help them notice cues in the middle of utterances. The early examples illustrated strong cues, and later examples included some weaker, more ambiguous ones. The early examples were all single-channel, but the last examples included two channels, so that the learners could hear how native speakers actually responded to the cues.

Thus the aim of the tutorial was to help train the ear to detect the pitch downdash: thus this was entirely about detection, not production. To enable learners to judge whether they were correctly

detecting the cue, for two of the audio examples they were asked to count the number of cues heard, and they were then given the right answer for comparison.

### 5.3 Cue Production Exercise

As another way to help learners understand the nature of the prosodic cue, we added a cue production exercise. Here the student is asked to mimic the pitch downdash. This was initially done with a native speaker listening and giving feedback. Some users reported that this increased their confidence in their ability to detect the cue.

[position Figure 3 about here]

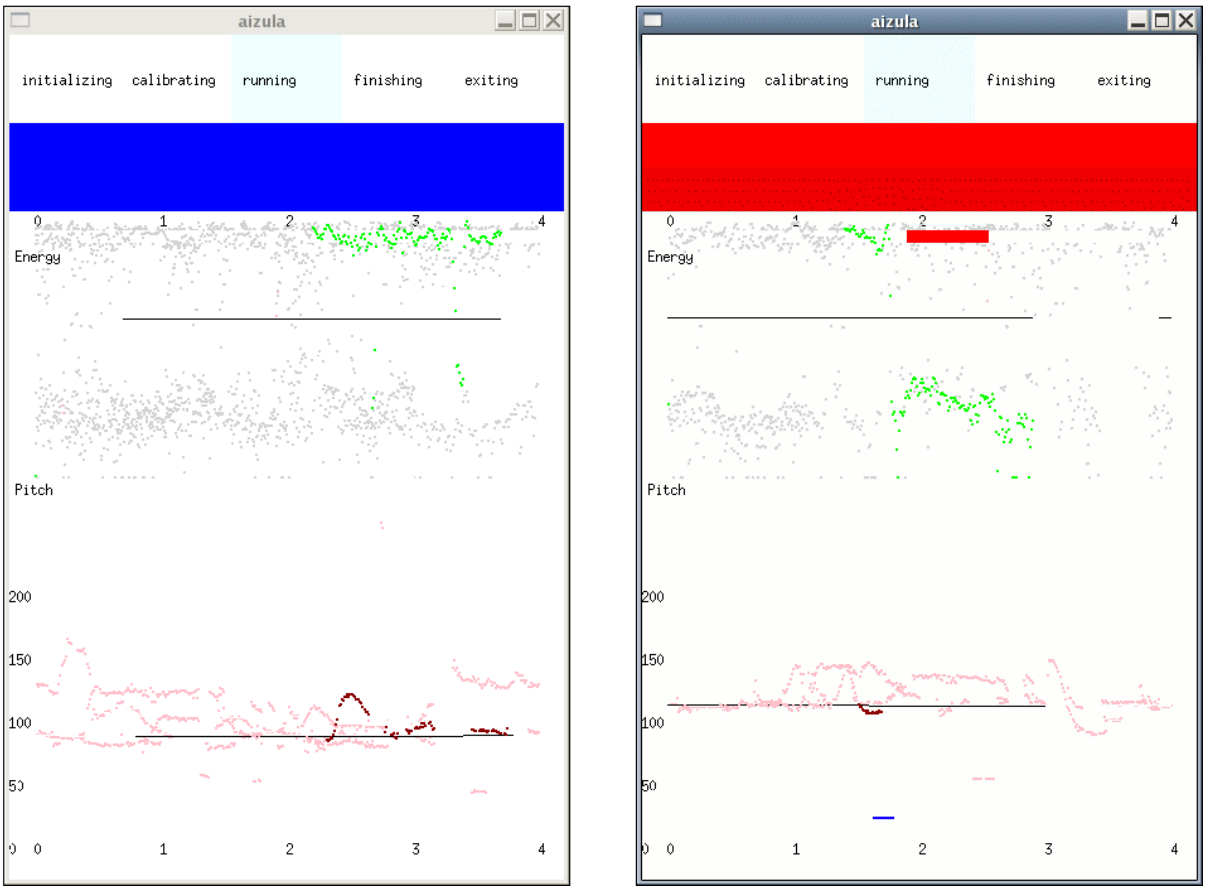


Figure 3: Cue Production Exercise Screenshots. The left shows the situation after the user has just produced a pitch peak followed by a region of flat pitch. The right display shows the visual feedback (a red flash) that appears together with the back-channel sound after the user produces the target prosodic cue.

This feedback was then automated by adapting an existing back-channeling engine (Ward & Tsukahara, 1999). This engine computed the pitch and energy of audio input in real time, displaying these to the screen. This is seen in Figure 3. Here the display grows from left to right, and every four seconds it wraps around and continues from the left again. The middle region of the display shows the energy of the signal from the microphone. Recent energy levels are displayed as dark dots, and the entire history is displayed as faint dots. The horizontal line here is the threshold; energy levels

above this point indicate that the user is speaking. The bottom region of the display shows pitch information in the same way. Subjects found it interesting to watch the system graph their pitch in real-time as they spoke.

The system also monitored for presence of the pitch downslope, and 300 milliseconds after detecting this it producing a back-channel sound and an accompanying red flash. It typically took a minute or so for subjects to learn to produce a pitch downslope long and steep enough to elicit this response; when it did they generally showed pleasure.

### 5.4 Feedback

Giving learners feedback on their performance is complicated by the fact that there is more than one way to be a good listener: there is substantial individual variation in back-channel performance. Given the same stimulus, even native speakers will have different opinions about where back-channels should appear (Wesseling & Van Son, 2005). This means that it is impossible to say that any single learner action, or inaction, is incorrect. Thus feedback must be given at the level of the overall pattern of behavior.

Although the general aim of the feedback is to shape the learner's behavior to be closer to that of a native speaker, some obvious ways to measure closeness (Ward & Tsukahara, 2000) may not be pedagogically useful. For example, one early version of the system assessed a penalty for productions at inappropriate places, and these were flagged to the learner as "misses". However this had the potential to discourage some learners from even trying, and was anyway inappropriate because even mis-timed feedback can be better than silence. These were therefore renamed "semi-hits".

We finally came up with an acceptable way to evaluate performance and give feedback. This decomposes the back-channeling skill into two almost independent aspects, each easy to measure and each easy to give clear feedback on.

The first aspect is timing. Back-channels should appear in places where a native speaker would produce them. These places were defined as times where the respondent in the corpus actually produced a back-channel, or where, in the judgment of a native speaker, a back-channel would be appropriate. If at least half of the learner's back-channels were in these places, this was considered acceptable, otherwise the system suggested paying more attention to the downdash cues.

The second aspect is frequency. The rate of back-channeling is important: if you produce too few, you will appear inattentive; if you produce too many, you may seem over-eager or superficial. Between these extremes there is a wide range of acceptable performance. If the learner's back-channel rate is less than 0.5 or more than 1.5 times that the target, then the system suggests producing more or fewer. For teaching purposes, this aspect may be less critical, since the overall average rate in Arabic is about the same as in English. However for evaluation purposes it is necessary to include a frequency factor, otherwise learners could obtain a high score by producing only one perfectly timed back-channel.

[position Figure 4 about here]
[position Figure 5 about here]

As illustrated in Figure 4, the aim of the system is to move the user towards the 1,1 point. Feedback is given by the core trainer after each track, as illustrated in Figure 5.

### 5.5 Scoring

For purposes of measuring performance — ultimately for evaluating learners' competence but here mostly for measuring whether the training had any effect — we use an overall metric combining both aspects of the skill: the product of the frequency quality and the timing quality.

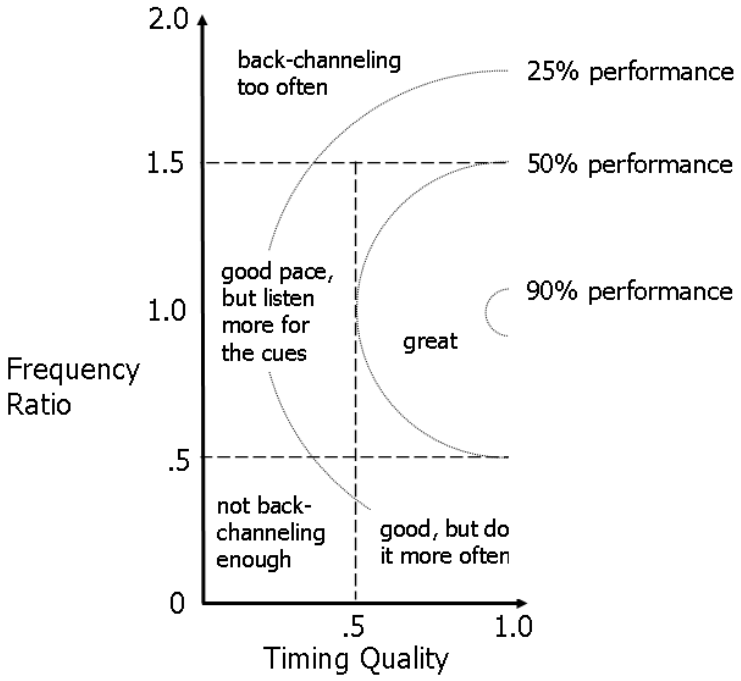


Figure 4: The Performance Space. The text in each dotted region summarizes the feedback given for that type of performance. The curved lines indicate overall performance levels.

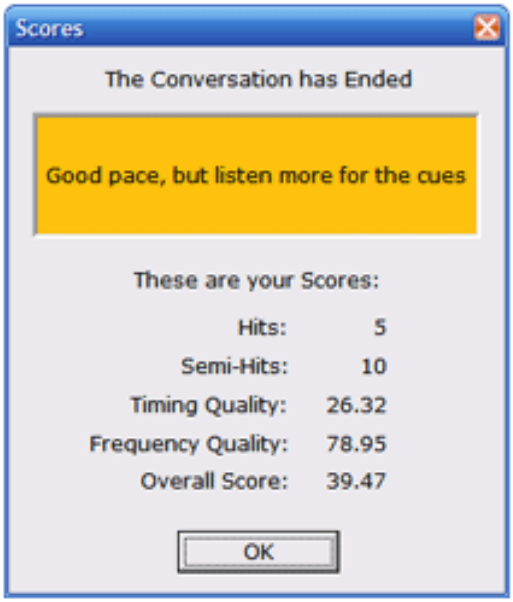


Figure 5: Screenshot of the Core Trainer's Feedback Window

First, the learner should produce approximately as many back-channels as were produced by the live listener responding to that track in the corpus. (This number is typically somewhat less than the number of opportunities.) We therefore divide the number of back-channels by the target number to

get a frequency ratio. If this ratio is 1 then the frequency quality is also 1 (100%), if this ratio is 0 or exceeds 2 then the quality is 0, and if this ratio is between these extremes, the quality is computed by linear interpolation.

The timing quality, computed as the fraction of back-channel productions that were at an appropriate place, also ranges from 0 to 1 (100%). However there is an effective lower bound of about 30%, that being the probability that a back-channel produced at random will fall in one of the windows of opportunity.

Figure 4 illustrates that the overall score, the product of these two, increases as the learner approaches the 1,1 point. Learners with an overall performance of 50% or higher are considered to have mastered the skill. At that point they are within the range of normal native variation, and are doing about as well as is possible based on prosody alone.

### 5.6 Summary of the Training

1. The learner uses the tutorial.
2. The learner learns to produce the pitch downdash, using the back-channeling engine.
3. The learner tries to produce back-channels in response to downdashes in an audio track, using the core trainer, as many times as desired, with the visual indicator on or off as desired.

As the software was not highly polished, and the three pieces were not integrated, an experimenter was always present to start the learner on each system.

### 5.7 Dialog Fragment Selection

A key question is that of what dialog fragments to use in the core trainer. These were chosen using three criteria. First, they were chosen to be less than a minute long, since learners' attention tends to fade fairly quickly. Second, they were chosen to be rich in back-channel opportunities, so that learners receive numerous examples of the cue. Third, since learners find it annoying to listen to a track containing multi-second silences, tracks were chosen from times when one person is talking almost continuously. Since long continuous talk was rare, one track was spliced together from smaller clips.

Each track is from a real conversation between two native Arabic speakers, taken from a corpus of telephone conversations (Canavan *et al.*, 1997). Chosen according to the criteria above, these tracks are atypically rich in back-channel opportunities; indeed most of the pause-delimited utterances ended with a back-channel cue. For each track, the listener's side, that is, the side containing the back-channel productions to be emulated by the learner, was excised and the remaining track is used as the stimulus.

In a second pilot study, with 21 subjects, we noticed that most subjects seemed to improve on the frequency quality but not much on the timing quality. This was a positive result: it meant that the training was enabling learners to produce back-channels in more places (to be more active listeners) while producing these extras in mostly the right places. However we still wanted subjects to attain better timing quality. We therefore decided to add a second training sample, one where the back-channel cues appeared much less frequently, reasoning that this would force subjects to pay more attention to the back-channel cues rather than relying heavily on frequency-matching. Doing this also addressed the most common suggestion by the pilot subjects, that we provide more training data.

We also discovered a complication involving scoring: subjects tended to do better on one of the tracks. This was the one with more back-channel opportunities, meaning that the chance of even a

random back-channel being scored as appropriate was greater For purposes of measuring learning we therefore normalized the scores, by decreasing the raw scores on the easier track by the difference in average scores between this track and the harder track.

### 5.8 Technical Details

Identification of the places where a back-channel response would be appropriate was done by a native Arabic speaker, the third author. She identified actual back-channels as they occurred in the corpus and also places where a back-channel could have occurred. In the tracks chosen about three quarters of these places were preceded by the pitch downslope cue, with the others generally preceded by another of the prosodic features identified as cues for back-channels (Ward & Al Bayyari, 2007). Each of these times was considered to be an appropriate place for the learner to back-channel. Incidentally, the visual hint calling attention to the cue, when used, lasts from 250 milliseconds before the window startpoint until the startpoint; thus it almost always occurs together with the cue.

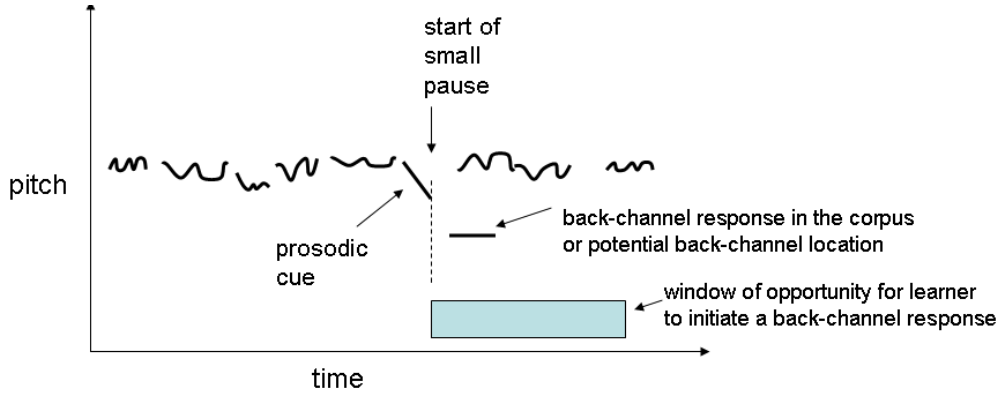


Figure 6: Determining Whether a Learner's Back-Channel is Appropriately Timed.

[position Figure 6 about here]

The learner was of course not required to back-channel precisely at a specified time, as indeed native speakers also vary in their precise timing (Wesseling & Van Son, 2005). Rather, anything within a certain range was considered to fall within the window of opportunity. Determining the appropriate range accurately is difficult, so we used a simple method. The earliest acceptable back-channel point was taken to be the end of the phrase containing the previous back-channel cue, as suggested by Figure 6. In the tracks used the cue almost always came exactly at the end, and this was almost always followed by a substantial pause, so this point was easy to identify. The end of the acceptable window was considered to be 1400 ms after this timepoint, based on some simple solicitation of native speaker impressions of back-channels at various timings.

A minor issue here is that of how to handle cases where the learner produces two back-channels in quick succession. Sometimes only one is appropriate, sometimes both are independently appropriate, and sometimes the two function together as one reduplicated back-channel. For lack of a principled way to distinguish among these, the system simply ignores any learner vocalization starting within 900ms of an earlier start. Apart from this, any vocalization falling within a window of opportunity is counted as a back-channel hit.

# 6 Evaluation Method

We performed an experiment to determine whether this training sequence did in fact help learners acquire the Arabic back-channeling skill.

The subjects were 24 students, 13 female and 11 male, recruited from the subject pool of those taking the first computer science course at a large public university in the Western United States on the Mexican border. No subjects reported any knowledge of Arabic. 19 self-reported as English-Spanish bilinguals.

The protocol, approved by the University's Institutional Review Board, was as follows:

1. The experimenter gave subjects the consent form and collected demographic data.
2. The experimenter briefly explained back-channels and their function within dialogs.
3. The subject had a familiarization run with the trainer tool, given a dialog in either English or Spanish, according to preference.
4. The subject used the introductory part of the tutorial, including the explanation and a few examples. This lasted about 70 seconds total.
5. The learner used the trainer tool and performance was measured, this was taken as the baseline.
6. The learner followed the training sequence detailed above (Section 5.6). The experimenter was available to provide technical help, but declined to answer questions about Arabic, back-channeling, or the specific cue. This phase lasted until subjects had done two runs with each training track or when they achieved 50% overall performance, whichever came first. Subjects took from 2 to 7 minutes, averaging about 5 minutes, for this.
7. The learner again used the training tool, with a different dialog.
8. The learner filled out a questionnaire to assess his or her views of the training process.
9. Looking at the questionnaire, the experimenter had an open-ended discussion with the subject, taking notes.

Different tracks were used in steps 5 and 7. One was 57 seconds taken from a dialog between two women and including 19 back-channel opportunities, and the other was 51 seconds taken from a dialog between two men and including 13 opportunities. Odd-numbered subjects were given the longer track first; even numbers the shorter first.

# 7 Results and Discussion

## 7.1 Learning Outcomes

We hypothesized that learners would perform better after the training than before. Table 1 shows that average frequency quality, timing quality, and overall quality all increased after training. The values seen here are the normalized ones; the specific correction factors were 2% for the frequency scores and 20% for the timing scores (explainable as due to the higher "effective lower bound" for the easier track, as discussed in Section 5.5).

Regarding overall performance, although there was great variation, 16 of the 24 subjects did show improvement. Applying matched-pairs t-tests to the individual scores, the tendency to improve was

| | normalized timing quality | | normalized frequency quality | | normalized overall quality | |
|---|---|---|---|---|---|---|
| before training | **57%** | (11%) | **71%** | (20%) | **40%** | (14%) |
| after training | **66%** | (18%) | **72%** | (20%) | **48%** | (20%) |
| gain | **9%** | (19%) | **1%** | (27%) | **8%** | (21%) |

Table 1: Normalized Average Learner Performance before and after training, with standard deviations in parentheses.

significant ($p < .05$). Regarding the timing quality, 18 of the subjects improved, and this was also significant ($p < .02$). There was, however, no significant improvement on the frequency quality.

Determining how much learning this actually represents is not easy. At first glance an 8% improvement in overall quality may not seem impressive, however for two reasons this understates the learning that took place. First the timing quality has a sort of lower bound of around 30%, as noted above, and a sort of upper bound: even for a native speaker perhaps 90% would be a likely maximum. Thus, the actual possible range of improvement was not from 0% to 100% but rather less. Second, the baseline performance was already fairly high, meaning that, for many of our subjects, just 70 seconds of exposure to a simple explanation and a few examples was enough to become able to recognize the cues and respond to them rather well, limiting the potential for further gains from the full training sequence.

Subjects' own view of their learning was seen in responses to the questionnaire item asking them to describe what they learned in their own words. Most expressed what they had learned at a high level, commenting that they had learned what back-channels are, how they are important in dialog, and how they can be produced as responses to prosodic cues.

Subjects generally thought that the training could be worthwhile, as seen by their responses to the questions, "Suppose you somehow found yourself at a cafe table where everyone was talking in Arabic. Do you think that having done this exercise would be helpful?" and "Why?" There was wide variation, but the average rating was 4.6 on a scale from 1 to 7, where 4 was labeled "somewhat helpful".

### 7.2 Determinants of Learning

We also asked subjects to rank the usefulness of each part of the training sequence. Almost all felt that all parts were valuable, with the core trainer generally ranked highest, the tutorial second, and the cue production exercise third. However this was not consistent, suggesting that different people may have different learning styles and benefit more from different methods.

There were no clear correlations between learning on the various metrics and other factors, such as time on task or order of exposure to the training samples. Comparing the results of the second pilot study with the main experiment, choice of training samples does seem to allow control over whether learners improve more on timing quality or more on frequency quality.

### 7.3 User Attitudes

Since back-channeling is not what most people think of as a key part of language learning, we were initially concerned that subjects might not take the training seriously, however this was not a problem. During the initial explanation, all subjects easily accepted the idea that back-channeling

skills are worth learning, and that a prosodic cue is involved. It may be that back-channeling is one of the universal fundamental "human interactional abilities" (Levinson, 2006). Subjects were also comfortable with the idea of learning this skill in the abstract; no one suggested adding more realism to the core trainer exercise. On the other hand, some subjects were skeptical about one aspect of the learning scenario: 9 out of the 24 subjects answered "yes" or "somewhat" to a question asking whether they "thought it strange to practice back-channeling with no knowledge of Arabic".

No users appeared to be "gaming the system" to increase their score. This was an advance over what we saw in early pilot studies, where some subjects back-channeled far too often, perhaps for lack of any clear idea how they were supposed to behave, due to a weak tutorial at that point.

We were also concerned that some users might resist this kind of training. Turn-taking patterns in general, and back-channel frequency in particular, often seem to reflect personality; for example, extroverts may back-channel more readily than introverts. In the early pilot experiments, it seemed that some subjects consistently back-channeled relatively infrequently, both before and after training, however no such correlation was seen here, perhaps due to the more thorough explanations and training. We explored this with the question "Some people find it hard to produce back-channels in certain ways, because doing so is 'not true to their personality'. Do you feel this way?" 8 of the 24 subjects answered "yes" or "somewhat". According to Crookall and Oxford (cited by Williams and Burden (1997)), "learning a second language is ultimately learning to be another social person,"' and there can be limits to how much a learner is willing to adopt the patterns of the target culture (Cutrone, 2005).

Another finding was that most subjects considered this topic interesting. People do not generally pay conscious attention to the prosody of turn-taking information when they are engaged in dialog, and learning to notice such low-level, automatic behaviors seemed to be novel and engaging. To the question "compared to other topics in foreign languages, is back-channeling interesting", the average response was 5.4 on a scale from 1 to 7, that is, halfway between "somewhat" and "exceedingly", suggesting that students may be intrinsically motivated to learn these skills.

# 8 Future Work

As this was just an initial exploration, many tasks and questions remain. One obvious task is to improve the implementation, making the training system fully automatic, fully integrated, fully portable, and integrated with other training tools. The remainder of this section points up three other issues.

## 8.1 Improving the Training

Although we know that this training sequence works, we do not know whether all components are in fact required; there may be a more streamlined way to give learners this skill. On the other hand, there are many possible extensions and enhancements that should be considered.

One is to explore ways to give learners feedback on each action, that is, their production or non-production at specific times. Although this is bound to be error prone, as noted in Section 5.4, it may still be of value. It may be useful, for example, to let learners replay specific parts of the track and hear how their behavior compares to that of the native speaker in the corpus.

Another improvement would be the addition of a demonstration that back-channeling is indeed important and can indeed be done without understanding. In the early informal studies many of the subjects had Spanish as a native language, and the human tutor, who knew no Spanish, used this to make a point. She would have a subject tell her a story, in Spanish, and demonstrate that she

could be a more or less attentive listener by attempting to back-channel cooperatively or not. Such an exercise could perhaps be automated.

There are many other possibilities to examine in future, including the use of pitch contour graphs (Chun, 2002; Hirata, 2004), making the trainer more game-like (Garcia-Carbonnell *et al.*, 2001; Johnson *et al.*, 2005), giving real-time feedback on frequency quality or timing quality, or giving diagnostic feedback regarding common problems, such as responding to the cue too slowly or relying too much on pauses.

### 8.2 Measuring the Value of Training

While back-channeling is clearly valuable in general, we have not yet demonstrated specifically that the back-channeling skill acquired in these training sessions is actually worthwhile in real situations. We have begun a new sequence of experiments in which native speakers of Arabic judge learners exhibiting either poor or good back-channel behavior, and exhibiting either poor or good pronunciation. Our expectation is that if a learner mispronounces a phoneme he or she will probably be seen as merely sloppy or childish, but if a learner mis-handles back-channeling, he or she can easily be seen as rude; conversely, we expect that good back-channeling can give a very positive impression.

Beyond this, ultimately we need to do more detailed studies to examine how well our overall metric of performance (and possible rivals such as d-prime and the F measure) correlates with actual judgments, by native Arabic speakers, of back-channeling skills. It would also be interesting to explore whether the timing quality could be measured better by assigning numeric scores to each back-channel based on its exact timing, rather than simply using a binary appropriate/inappropriate judgment. Ultimately we need to determine whether and when back-channeling skills actually facilitate successful intercultural communication (Li, 2006, pg. 111).

It is also necessary to examine the contributions of the various parts of our training sequence to the retention of back-channeling skills over time.

### 8.3 Training for Other Populations

The training sequence has been tested so far only with subjects with no knowledge of Arabic. Acquiring only this one skill may not be entirely useless: a person dealing with an Arabic speaker through an interpreter may still want to act as a polite listener.

However we also think this skill should be taught to learners who already have some knowledge of Arabic, which raises the question of whether our training methods will still be appropriate. One consideration is cognitive load, which may be higher if a learner already knows some words of Arabic. Users of the core trainer tend to be fully engaged in the task of detecting and responding to the prosodic cues; if they were simultaneously expected to recognize and process lexical information, that may overload them, as we saw in some of our Tokyo users. Thus it is possible that more advanced learners should be exposed to low-pass filtered stimuli, allowing them to concentrate on the prosody only.

Other applications may also be possible. It is not invariably the case that native speakers are socially adept, and one aspect of this is often non-verbal communication problems. Computer tutors can be useful for children with autism (Bosseler & Massaro, 2003; Tartaro & Cassell, 2006), and the techniques used here may also be helpful.

## 9 Summary

The main contribution of this paper is the presentation of a method for training learners to emulate back-channel behavior in a second language. A secondary contribution is the presentation of a way to measure back-channel performance, a skill which has not previously been quantified.

The application of these findings may be slow initially, due to the paucity of quantitative knowledge about back-channeling and turn-taking norms in various languages. However basic research on these topics is proceeding quickly, thanks to the development of better tools for working with corpora of conversation data, so the lessons learned should eventually find wide general use, leading, we hope, to increased cultural proficiency and more satisfactory intercultural communications.

## References


Acosta, Luis Hector (2004). Prosodic Features that Cue Back-Channel Responses in Northern Mexican Spanish. University of Texas at El Paso, Computer Science Department Masters Thesis.

Al Bayyari, Yaffa & Nigel Ward (2007). The Role of Gesture in Inviting Back-Channels in Arabic. In *presented at the 10th Meeting of the International Pragmatics Association*.

Allwood, Jens (1993). Feedback in Second Language Acquisition. In Clive Perdue, editor, *Adult Language Acquisition: Cross Linguistic Perspectives, II: The Results*, pp. 196–235. Cambridge University Press, Cambridge.

Almaney, A. J. & A. J. Alwan (1982). *Communicating with the Arabs*. Waveland Press, Prospect Heights, Illinois.

Berry, Anne (1994). Spanish and American Turn-Taking Styles: A Comparative Study. In L. F. Boulton, editor, *Pragmatics and Language Learning Monograph Series, Volume 5, 1994*, pp. 180–190. University of Illinois, Urbana-Champaign: Division of English as an International Language, Urbana, Illinois.

Berry, Anne (2003). Are You Listening? (Backchannel Behaviors). In K. Bardovi-Harlig & R. Mahan-Taylor, editors, *Teaching Pragmatics*. US Department of State, Office of English Language Programs, Washington, D. C.

Bolinger, Dwight (1989). *Intonation and Its Uses*. Stanford University Press, Stanford, California.

Bosseler, Alexis & Dominic W. Massaro (2003). Development and Evaluation of a Computer-Animated Tutor for Vocabulary and Language Learning in Children with Autism. *Journal of Autism and Developmental Disorders*, 33:653–672.

Brennan, S. E. & E. A. Hulteen (1995). Interaction and Feedback in a Spoken Language System: A theoretical framework. *Knowledge-Based Systems*, 8(2–3):143–151.

Byram, Michael (1997). *Teaching and Assessing Intercultural Communicative Competence*. Multilingual Matters, Clevedon.

Canale, Michael & Merrill Swain (1980). Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing. *Applied Linguistics*, 1:1–47.

Canavan, Alexandra, George Zipperlen, & David Graff (1997). *CALLHOME Egyptian Arabic Speech*. Linguistic Data Consortium, Philadelphia, Pennsylvania. LDC Catalog No. LDC97S45, ISBN: 1-58563-114-0.

Chun, Dorothy M. (2002). *Discourse Intonation in L2: From theory and research to practice*. John Benjamins.

Clancy, Patricia M., Sandra A. Thompson, Ryoko Suzuki, & Hongyin Tao (1996). The conversational use of reactive tokens in English, Japanese and Mandarin. *Journal of Pragmatics*, 26:355–387.

Clark, Herbert H. (1996). *Using Language*. Cambridge University Press, Cambridge.

Cutrone, Pino (2005). A case study examining backchannels in conversations between Japanese-British dyads. *Multilingua*, 24:237–274.

Demo, Douglas A. (2001). Discourse Analysis for Language Teachers. Technical Report EDO-FL-01-07, Center for Applied Linguistics, Washington, D. C.

Duncan, Jr., Starkey & Donald W. Fiske (1985). The Turn System. In Starkey Duncan, Jr. & Donald W. Fiske, editors, *Interaction Structure and Strategy*, pp. 43–64. Cambridge University Press, Cambridge.

Ellis, Donald G. & Ifat Maoz (2006). Dialogue and Cultural Communication Codes between Israeli Jews and Palestinians. In Larry A. Samovar, Richard E. Porter, & Edwin R. McDaniel, editors, *Intercultural Communication: A Reader, 10th edition*, pp. 231–237. Thompson-Wadsworth, Belmont, California.

Fujie, Shinya, Kenta Fukushima, & Tetsunori Kobayashi (2005). Back-channel feedback generation using linguistic and nonlinguistic information and its application to spoken dialogue system. In *Proc. 9th European Conf. on Speech Communication and Technology, Interspeech 2005*, pp. 889–892, Lisbon, Portugal.

Garcia-Carbonnell, Amparo, Beverly Rising, Begona Montero, & Frances Watts (2001). Simulation/gaming and the acquisition of communicative competence in another language. *Simulation and Gaming*, 32:481–491.

Gardner, Rod (1998). Between Speaking and Listening: the Vocalisation of Understandings. *Applied Linguistics*, 19:204–224.

Gardner, Rod (2001). *When Listeners Talk: Response tokens and listener stance*. John Benjamins, Amsterdam.

Hafez, Ola Mohamed (1991). Turn-taking in Egyptian Arabic: Spontaneous speech vs drama dialogue. *Journal of Pragmatics*, 15:59–81.

Heinz, Bettina (2003). Backchannel responses as strategic responses in bilingual speakers' conversations. *Journal of Pragmatics*, 35:1113–1142.

Hirata, Yukari (2004). Computer Assisted Pronunciation Training for Native English Speakers Learning Japanese Pitch and Durational Contrasts. *Computer Assisted Language Learning*, 17:357–376.

Horiguchi, Sumiko (1997). *Nihongo Kyoiku to Kaiwa Bunseki (Japanese Conversation by Learners and Native Speakers)*. Kuroshio, Tokyo.

Hurley, Daniel Sean (1992). Issues in Teaching Pragmatics, Prosody, and Non-Verbal Communication. *Applied Linguistics*, 13:259–281.

Ikeda, Keiko (2004). "Listenership" in Japanese: An Examination of Overlapping Listener Responses. NFLRC NetWork #32, National Foreign Language Resource Center, University of Hawai'i at Manoa.

Johnson, W. Lewis, Carole Beal, Anna Fowles-Winler, Ursula Lauper, Stacy Marsella, Shrikanth Narayanan, Dimitra Papachristou, Andre Valente, & Hannes Vilhjalmsson (2005). Tactical Language Training System: An Interim Report. USC ISI, adapted from a conference paper presented at the Intelligent Tutoring Systems Conference, September 2004.

Kaplan, Jonathan D., Mark A. Sabol, Rober A. Wisher, & Robert J. Seidel (1998). The Military Language Tutor (MILT) Program: An Advanced Authoring System. *Computer Assisted Language Learning*, 11:265–287.

Kasper, Gabriele & Shoshana Blum-Kulka (1993). Interlanguage Pragmatics: An Introduction. In Gabriele Kasper & Shoshana Blum-Kulka, editors, *Interlanguage Pragmatics*, pp. 3–17. Oxford University Press, Oxford.

Levinson, Stephen C. (2006). On the Human 'Interaction Engine'. In Nicholas Enfield & Stephen C. Levinson, editors, *Roots of Human Sociality: Culture, Cognition and Human Interaction*. Berg, Oxford.

Li, Han Z. (2006). Backchannel Responses as Misleading Feedback in Intercultural Discourse. *Journal of Intercultural Communication Research*, 35:99–116.

LoCastro, Virginia (1987). Aizuchi: A Japanese Conversational Routine. In L. E. Smith, editor, *Discourse Across Cultures*, pp. 101–113. Prentice-Hall, Upper Saddle River, New Jersey.

Matsuda, Yoko (1988). Taiwa no Nihongo Kyouikugaku: Aizuchi ni Kanren shite (Teaching Japanese Interaction: on back-channel feedback). *Nihongogaku*, 7:59– 66.

Maynard, Senko K. (1989). *Japanese Conversation*. Ablex, Norwood, New Jersey.

Mizutani, Osamu (1981). *Japanese: The Spoken Language in Japanese Life*. The Japan Times, Tokyo.

Mukai, Chiharu (1999). The Use of Back-channels by Advanced Learners of Japanese: Its Qualitative and Quantitative Aspects. *Sekai no Nihongo Kyouiku*, 9:197–217.

Ogden, Richard & Sara Routarinne (2005). The Communicative Functions of Final Rises in Finnish Intonation. *Phonetica*, 62:160–175.

Ohama, Rui (2000). Nihongo Kyoiku niokeru Aizuchi Shido no tame ni (Groundwork for Teaching Back-channeling in Japanese Language Education). In Ichiro Marui, editor, *Ibunka Tekio to Gengo Kyoiku (Symposium on Intercultural Adaptation and Language Teaching)*, pp. 37–43. Kochi University Department of International Society and Communication.

Ohta, Amy Snyder (2001). *Second Language Acquisition Processes in the Classroom: Learning Japanese*. Lawrence Erlbaum Associates, Mahwah, New Jersey.

Rivera, Anais G. & Nigel G. Ward (2007). Three Prosodic Features that Cue Back-Channel Feedback in Northern Mexican Spanish. Technical Report UTEP-CS-07-12, University of Texas at El Paso, Department of Computer Science.

Rost, Michael (1990). *Listening in Language Learning*. Longman, London.

Sacks, Harvey, Emanuel A. Schegloff, & Gail Jefferson (1974). A Simplest Systematics for the Organization of Turn-taking for Conversation. *Language*, 50:696–735.

Schegloff, Emanuel A. (1982). Discourse as an Interactional Achievement: Some Uses of "Uh huh" and Other Things that Come Between Sentences. In D. Tannen, editor, *Analyzing Discourse: Text and Talk*, pp. 71–93. Georgetown University Press, Washington, D. C.

Shriberg, Elizabeth, R. Bates, A. Stolcke, P. Taylor, D. Jurafsky, K. Ries, N. Coccaro, R. Martin, M. Meteer, & C. Van Ess-Dykema (1998). Can Prosody Aid the Automatic Classification of Dialog Acts in Conversational Speech? *Language and Speech*, 41:439–487.

Tartaro, Andrea & Justine Cassell (2006). Authorable Virtual Peers for Autism Spectrum Disorders. In *Proceedings of the Workshop on Language-Enabled Educational Technology at the 17th European Conference on Artificial Intelligence (ECAI06)*.

Ward, Nigel & Yaffa Al Bayyari (2006). A Case Study in the Identification of Prosodic Cues to Turn-Taking: Back-Channeling in Arabic. In *Interspeech 2006 Proceedings*.

Ward, Nigel & Yaffa Al Bayyari (2007). A Prosodic Feature that Invites Back-Channels in Egyptian Arabic. In Mustafa Mughazy, editor, *Perspectives in Arabic Linguistics XX*. John Benjamins, Amsterdam. to appear.

Ward, Nigel, Atsuko Kondo, & Origa Nagai (2001). A Conversational-Reflex Training System for Second Language Learners (in Japanese). Final grant report, University of Tokyo.

Ward, Nigel & Wataru Tsukahara (1999). A Responsive Dialog System. In Yorick Wilks, editor, *Machine Conversations*, pp. 169–174. Kluwer.

Ward, Nigel & Wataru Tsukahara (2000). Prosodic Features which Cue Back-Channel Feedback in English and Japanese. *Journal of Pragmatics*, 32:1177–1207.

Wesseling, W. & R.J.J.H. Van Son (2005). Timing of Experimentally Elicited Minimal Responses as Quantitative Evidence for the Use of Intonation in Projecting TRPs. In *Interspeech 2005*.

White, Sheida (1989). Backchannels across cultures: A study of Americans and Japanese. *Language in Society*, 18:59–76.

Williams, Marion & Robert L. Burden (1997). *Psychology for Language Teachers: a Social Constructivist Approach*. Cambridge University Press, Cambridge.

Yamada, Haru (1992). *American and Japanese Business Discourse: A comparison of interactional styles*. Ablex, Norwood, New Jersey.

Yngve, Victor (1970). On Getting a Word in Edgewise. In *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society*, pp. 567–577, Chicago.

Young, Richard F. & Jina Lee (2004). Identifying Units in Interaction: Reactive Tokens in Korean and English Conversations. *Journal of Sociolinguistics*, 8:380–407.