Submitted to the *Journal of Cross-Cultural Psychology*, April 23, 2008

American and Arab perceptions of an Arabic turn-taking cue

Nigel G. Ward, Yaffa Al Bayyari
University of Texas at El Paso, 79968 USA
nigelward@acm.org

Abstract: Languages differ in the way that speakers coordinate their interaction moment-by-moment. It has been suggested that misinterpretations of these behaviors can be an important cause of intercultural misunderstandings. We explore this in the domain of listening behavior. One way that listeners show interest and attention is by producing back-channel feedback (short utterances such as *okay* and *hmm*) at appropriate times, and these times are determined, in part, by the interlocutor, who signals when such feedback is welcome with various cues. In Arabic these cues include a prosodic feature in the form of a steep continuous drop in pitch. Here we show that English speakers tend to misinterpret this, perceiving it as an expression of negative affect. We further show that this tendency is substantially alleviated by about 25 minutes of training.

Keywords: cross-cultural interaction, dialog, listener behavior, prosody, back-channel feedback

Acknowledgments: We thank Bill Lucker, David Novick, Maissa Khatib, Rafael Escalante, Lewis Johnson, and Ralph Chatham. This work was supported by the Department of Defense and its Advanced Projects Research Agency.

1 Background

Some social behaviors can carry an emotional punch, whereas others seem mundane. In intercultural interaction, if the participants don't know which are which, then small misinterpretations can lead to deep misunderstandings. At first glance, the mechanics of interaction in dialog seem to fall in the mundane category, and indeed "interaction appears to have detailed universal properties ... for a wide range of features, from turn taking [to] ... greetings ... the languages and cultural systems that have been studied reflect very similar, in some cases eerily similar, subsystems" [1]. On the other hand, it also appears that "the speakers' mutual judgments of abilities and intentions are profoundly influenced, if not determined, by automatic uses of the nuts and bolts of language — pitch and amplitude; intonational patterns; pacing and pausing ..." [2]. A number of studies have identified cases where such behaviors have been or can be misinterpreted [3].

Most evidence on intercultural misinterpretations of such interactional behaviors has been anecdotal or indirect. In this paper we study this issue experimentally, by examining how a prosodic pattern used in one culture as a mechanism for coordinating dialog is interpreted by members of another culture.

1.1 Turn-Taking Behaviors

Turn-taking, the way in which interlocutors coordinate whose turn it is to speak, is an ever-present aspect of almost all interactions [4]. The mechanisms by which this is accomplished have become clearer in recent years,

thanks to the availability of dialog corpora, to the development of methods and tools for analyzing them, and to the need to model such behaviors for dialog systems. Among the methods for managing turn-taking, prosodic cues play a large role: speakers systematically use the pitch, energy, and timing of their utterances to express turn-taking intentions [5].

Back-channeling is one aspect of turn-taking. Back-channels, also sometimes known as "response tokens," "reactive tokens," "minimal responses," and "continuers," are short utterances produced while the interlocutor has the turn [6]. They are one way by which a listener shows that he or she is listening: they typically express attention, interest, understanding and/or willingness to let the other person continue. In English, common back-channel expressions include *uh-huh*. *yeah*, *hm*, *right* and *okay*. Back-channels are pervasive: a typical frequency is 5 per minute in casual English conversation, and on one count 20% of all utterances in English dialog are back-channels [7].

Although not salient — indeed back-channeling seems to happen largely below the level of conscious awareness — back-channels are important both for the successful communication of content and for the development of rapport, as shown by studies in which confederate listeners or computer listeners back-channel or fail to back-channel in various ways [8, 9, 10, 11].

Although the production or non-production of a back-channel is ultimately up to the listener, there are places in dialog where back-channels are especially welcome, and these can be marked by a prosodic cue from the speaker, as shown by statistical studies of dialog corpora and by experiments in which subjects back-channel to a pre-recorded track [12, 13].

People belonging to the same culture generally solicit, produce, and interpret back-channels with no problems and no awareness; thus they seem to be part of the mundane "nuts and bolts of language." However it is also known that cultures differ back-channeling behavior [14], perhaps most saliently in the typical frequencies of back-channels in various languages [15, 16]. There are also cases where such differences have caused cross-cultural misunderstandings [17]. For example, since back-channels are more frequent and occur swifter in Japanese than in English, Americans can seem uninvolved and uninterested to Japanese and Japanese can seem shallow and over-dependent to Americans. In general, second language learners need to be able to back-channel appropriately but often can not [18, 19].

1.2 Back-Channel Behavior in Arabic and English

Arabic and English are two languages which are genetically unrelated but which are increasingly in contact. High-level cultural differences between the two speaker populations are well known, and these can explain some differences in discourse patterns and account for many misunderstandings [20, 21], however the possible effects of differences in the low-level mechanics of interaction have not previously been explored.

Semantically and pragmatically, back-channels in Arabic appear to bear the same basic functions as in English [22], and the overall frequency of back-channeling is similar between the two languages. Corpus-based studies indicate that the production of certain prosodic cues by the speaker are strongly associated with the subsequent production of a back-channel response by the listener, in both Arabic and English [12, 23, 24], although the specific cues involved are different. In English (at least in parts of North America), times when the listener is especially welcome to back-channel are indicated by a region of low pitch. In Arabic (at least in parts of Egypt and Iraq), these times are commonly indicated by a different prosodic cue, a prosodic feature complex which includes a steep pitch downslope, a "downdash."

This pattern is not common in English. To the ears of the first author, it seemed to bear a meaning of accusation or feelings like resignation.

Thus we performed three experiments. The first was done to confirm the corpus-based finding, that pitch downdash is perceived as an invitation to back-channel by Arabic speakers but not by naive Americans. The second was done to determine whether the pitch downdash is indeed liable to misinterpretation: we hypothesized that this prosodic pattern would be perceived negatively by Americans but not by Arabs. The third was done to evaluate the importance of back-channel responses in cross-cultural encounters.

2 Experiment 1: Perceptions of Discourse Function

We expected that a back-channel response following an utterance ending in a downdash would be judged as a natural pairing by the Arabic speakers but that naive Americans would not have this perception.

2.1 Methods

2.1.1 Stimuli

The stimuli were various pairings of cues and responses, as suggested by Figure 1, including the downslope+back-channel combination and various controls.

Each stimulus was composed of three parts, a fixed lead-in, one of three prosodic features of interest, and one of three listener responses, as illustrated in Figure 1. Three prosodic features were used: the pitch downdash and two controls. Three types of response were used: a back-channel, a full turn, or silence. The controls were two other pitch patterns with turn-taking functions. The first was a downward staircase of three flattish pitch regions, that is, a "cadence" pattern, identified by Bergsträsser [31]) as an indication of "finality," and sometimes associated in an Egyptian corpus with turn yields. The second was a pitch pattern associated with giving one item in a list while indicating that there are more to come.

The fragments used to create the stimuli were extracted from dialog AR_4023_1_pt1 in a corpus of Egyptian Arabic telephone conversations [32], with the exception of the pitch downdash itself, which was taken from Track 13 of an Iraqi Arabic corpus [33]. The prosodic cue fragments were chosen by the second author, a native speaker of Arabic, to be clear yet typical examples.

To ensure that listener judgments were based on the pitch patterns alone, uninfluenced by lexical information, all fragments were resynthesized using Praat [34] to discard the segmental and volume information. The pitch points used to specify the contours for resynthesis were extracted automatically. The resulting stimuli sounded, for example, like *babaa-ba ba-bababa*, but were recognizably human.

The lead-in was a fragment that included no pragmatically or emotionally salient pitch movements and that sounded natural when spliced with each of the prosodic cues. The pitch-cue fragments were scaled up by multiplying the pitch across each fragment by a constant factor sufficient to make the transitions unnoticeable. The pieces were assembled using Reaper. The quality of the resulting stimuli was evaluated by a second native speaker: although all were somewhat unnatural due to the resynthesis, none was perceived to be particularly bad. The stimuli were also evaluated by a question during the second experiment: we asked the Arab subjects whether the speaker sounded "like a native Arabic speaker?" on a scale from 1 to 5. The stimulus ending in an upturn was rated 4.1, in downturn 3.5, and in cadence 3.3, so there were differences. However these do not correlate with the judgments reported earlier, so it is unlikely that the results were artifacts of synthesis problems.

The pause between the end of the cue and the start of the response was 500ms, a typical value in the corpus both for the time gap between a pause onset and a back-channel response, and for the time-gap between utterances at a turn hand-off.

Stimuli are available at our website, http://www.cs.utep.edu/nigel/abc/.

2.1.2 Participants

There were 3 groups of 18 subjects each: Arabs, English-speaking students who had previously undergone about 25 minutes of training in this dialog pattern ("exposed Americans") , and those without such training ("naive Americans").

The Arab subjects were native Arabic speakers, speaking a variety of dialects. 12 self-identified as native speakers of Palestinian Arabic, 2 as Sudanese, and 1 each as Egyptian, Saudi, Lebanese, and Algerian; however all were familiar with Egyptian Arabic speech patterns. 7 were living in the United States and 11 in Qatar. The El Paso subjects were recruited by word of mouth from the local Arab community; the Qatar subjects were mostly acquaintances of the second author. All were compensated with $20. Subjects gave informed consent and were handled following an IRB-exempted protocol (#2492).

The Arabic-naive and Arabic-exposed subjects were recruited from Introduction to Computer Science

classes, and were compensated either with class credit or $10. Subjects were recruited without regard to linguistic background, however we chose to use data only from those who had being using English since at least age 16 and were judged by the experimenter to have good dialog skills. Five datasets were accordingly discarded and five additional subjects recruited. Even so, most of the subjects were Spanish-English bilinguals; it is therefore worth noting that the back-channel cuing feature-complex in Spanish is different from those in Arabic or English, but the overall frequency, pragmatic contexts, and typical time from cue to response are similar across all three languages [35].

The "exposed American" subjects were those who had earlier been trained in how to act as a good listener to an Arabic speaker by responding with back-channels to pitch downslope cues. This was done as part of an evaluation of the effectiveness of a software suite [36]. The training sequence lasted about 25 minutes and included an explanation, audio examples, the use of visual signals to highlight occurrences of pitch downslopes, auditory and visual feedback on learners' attempts to produce the cue themselves, and feedback on the learners' performance as they played the role of an attentive listener in response to one side of a pre-recorded dialog. These subjects were tested two to three months after this training experience.

2.1.3 Procedure

The subjects were told that "we are interested in patterns of dialog in various languages, especially Arabic and English. This study is about which patterns of interaction are preferred or disliked by various people." The oral instructions were given in Arabic or in English, but the written instructions and surveys were always in English. One experimenter ran all the subjects.

After the demographic survey subjects were told they would be "listening to dialog fragments which have been modified so that you can't identify the speakers or their words." For familiarization with these filtered stimuli, they initially listened to a few English dialog fragments and then resynthesized versions, so they could obtain a sense of how the filtered sounds related to real dialogs.

For the first experiment they were asked for judgments of "a positive or a negative feeling about the second speaker." Subjects ranked each stimulus on a seven-point scale, with 1 being very negative and 7 being very positive. Before the 9 stimuli of interest they were presented with two English-derived example, one with a normal pattern of back-channeling and one with a badly delayed back-channel. The value of the rating for the first sample was used as a benchmark for normalization.

Samples were provided on-screen, on a laptop computer, in stereo, through headphones. Subjects were able to adjust the volume for comfortable listening and were allowed to proceed at their own pace, being able to listen to the stimuli any number of times and in any order. We encouraged them to listen repeatedly to stimuli until they were comfortable making the various judgments. Since pilot studies had shown that repeated listening to the full 16-second clips was tiring, subjects were also provided with abbreviated 6-second versions containing just the cue and the response parts. These short fragments were intended to make it easier for subjects to focus on the transition between the turns when making judgments.

Asked whether making the judgments was difficult, most subjects said that it was, to varying degrees, and most said that this was because the words were masked. Three subjects indicated that the downslope and cadence patterns sounded the same to them. No subject indicated that they had difficulty perceiving the stimuli as filtered versions of two people talking.

The order in which samples appeared was balanced across subjects. These orders were fixed beforehand and each group of participants saw the same orders; subject number n in the Arab group was given the stimuli in the same order as subject number n in the naive American group, and so on. This made matched-pairs analysis possible.

2.1.4 Data Analysis

The raw values were normalized based on the average rating of the benchmark for each group, to compensate for any tendency for one group to be overall more positive than the other. Both the Arab and the Arabic-naive groups rated the benchmark stimulus 5.06 on average and the Arabic-exposed group rated it 5.28 on average. Thus the results in Table 1 are the values after subtracting of .22 from all ratings by the exposed American group.

2.2 Results

As seen in Table 1, the back-channel response to a downslope cue was ranked relatively highly by the Arabic speakers, averaging 4.7 on a scale from 1 to 7, but lower by the naive Americans, averaging 3.7, and this difference was significant ($p < .02$, matched-pairs one-tailed t-test).

The exposed Americans were different: they rated the downslope+back-channel pairing 4.5 on average, close to the rating by the Arabs and significantly different from that of the naive Americans ($p < .01$).

Although Arabic-naive Americans tended to rank the Arabic stimuli lower overall (an average over all nine of 4.06, vs 4.35 for the Arabs) this was not an undifferentiated dislike. In fact, they rated four of the nine pairings more positively than did the Arabs.

3 Experiment 2: Perceptions of Affect

In the second experiment we set out to determine whether the pitch-downslope could be misinterpreted as an expression of affect.

3.1 Method

3.1.1 Stimuli

Subjects were given the initial fragments of the same stimuli as before, without the response parts. Thus there were three stimuli, identical up to the last few hundred milliseconds, where one ended in the pitch downslope pattern and the two controls ended in other patterns.

3.1.2 Participants

The participants were the same as for the first experiment.

3.2 Procedure

Experiment 2 was done immediately after Experiment 1. Subjects were asked to "listen to three audio fragments from one speaker and try to infer their emotional state." For each sample, subjects were also asked to "please write two or three adjective describing the feeling (sad, angry, happy, surprised, scared, disgusted, etc.)."

Before the three Arabic stimuli, subjects were given two emotional American English utterances, one happy and one angry, both in original and masked form. These were included, first, in order to reinforce to subjects the relation between the original and resynthesized utterances, and second, to provide a benchmark. Thus the positive/negative ratings were normalized by subtracting out the inter-group differences in the evaluation of the benchmark. Considering only those who perceived the benchmark the same way, that is, who chose to label it as "happy" and included no other term, the averages for the naive Americans and for the Arabs were both 6.5, and the average for the Arabic-exposed Americans was 6.3. Thus normalization of responses to the question "do you get a negative or positive impression of this speaker?" was done by adding .2 to all the ratings by the Arabic-exposed Americans.

3.3 Results

There was great variety in the adjectives chosen, and a clear pattern did not emerge. For the downslope pattern, the naive American top selections were sad and scared, the exposed American top selections was angry or mad, and the Arab choices were more evenly distributed. Overall there was a weak tendency for Americans to describe the downslope stimulus using more negative adjectives than the Arabs; using the affective norms for English words [37] to compute the average valences, on a scale from 1 to 9, for the Americans this was 2.9 versus 3.5 for the Arabs, on a scale from 1 to 9.

However the rankings (Table 2) clearly supported the hypothesis: the Arabs overall ranked the downdash pattern as neutral or slightly positive (average 4.2) and the naive Americans as slightly negative (3.2), and the difference was significant ($p < .02$, one-tailed matched pairs t-test). This was not just a reflection of a general American dislike for all Arabic stimuli; in fact they rated one of the controls more positively than did the Arabs.

In contrast the exposed Americans did not perceive a negative affect in the downslope cue; they rated it 4.0 on average, neutral and close to the Arab ratings, but significantly different from the ratings of the naive Americans ($p < .04$). However the exposed Americans rated all samples higher than did the naive Americans, so this effect seems to reflect in part a general tendency for exposure to Arabic to reduce negative perceptions.

4 Experiment 3: Perceptions of Personality and Social Effectiveness

The first two experiments show that naive Americans misinterpret both the pragmatic significance and the affective weight of these prosodic cues. This suggests that in intercultural encounters Americans hearing the downslope may incorrectly ascribe negative affect to Arabic speakers with no such intention. This leads to the reciprocal question of how Arab perceptions in intercultural encounters may be affected. Specifically, might Americans be likely to be misjudged if they do not correctly interpret these cues? or, conversely, will Americans trying to interact with Arabs be perceived more favorably if they back-channel according to the rules of Arabic?

4.1 Method

4.1.1 Stimuli

The first experiment showed that Arabic speakers perceive a back-channel following a downslope cue to be an appropriate response. Here we reexamined this perception using longer stimuli.

9 stimuli were derived from one actual conversation, a rehearsed but natural-sounding 11-second exchange between the second author and someone with no previous knowledge of Arabic. In this the learner asked how to get to a campus building and, as he was given step-by-step directions, showed that he was listening by back-channeling twice.

The greeting was not resynthesized, but the direction-giving and the back-channel responses were. During the direction-giving the back-channels were designed to be appropriate (BC 1), inappropriate (BC 2) or missing. BC 1 was the actual behavior of the listener in the original track; this seemed appropriate in our judgment and also matched our model in that the back-channels came after the downslopes. BC 2 was created by modifying this: one of the two back-channels was deleted and another was inserted. The insertion was in a place where the direction-giver paused, thus it was not interrupting the speaker, but we nevertheless felt that it was an inappropriate place for a back-channel to occur. The third variant had no back-channels, representing the behavior of a silent listener.

We wanted to demonstrate not only that there statistically significant differences in perception, but also that the differences were substantial. The stimuli accordingly crossed pronunciation quality with listening quality, as shown in Figure 2. The greeting, when present, was "assalaamu alaykum," pronounced either well or poorly.

For lack of an obvious way to cross-calibrate a pronunciation quality scale and a back-channeling quality scale, we strove to make the two dimensions of manipulation roughly comparable. The good greeting was designed to be very good, although recognizably non-native, and the bad one was chosen as one which was sloppy but intelligible. Similarly, the appropriate back-channeling was designed to sound fully natural and the bad one to be awkward.

4.1.2 Participants

Judgments were obtained from 18 subjects, the Arab participants from Experiments 1 and 2.

4.1.3 Procedure

In contrast to Experiment 1, this time the stimuli were presented as representing the behavior of learners, and the participants were asked to rate them not only in terms of linguistic naturalness but also in terms of perceptions of the personality and social effectiveness of the learner. Specifically, we asked them to "listen to 9 audio fragments of dialogs between two speakers, a native-Arabic speaker and a learner, and ... judge how the learner sounds." For each stimulus subjects were asked to respond to three questions, each on a scale from 1 to 7. The order of presentation was balanced. At the end we presented two sets of dialog pairs with forced choices between them. One dialog pair contrasted BC1 and a poor greeting with BC2 and a good greeting; the other contrasted BC1 and a missing greeting with no back-channels and a good greeting.

4.2 Results

As seen in Tables 3 – 5, back-channeling generally contributed positively to impressions. Sometimes the difference was significant. For example, comparing the most highly rated behavior, BC 2, with the no back-channel condition across all greeting types, the difference was significant for all three dimensions of rating ($p < .03$ for knowledge of Arabic, $p < .01$ for personality, $p < .02$ for social effectiveness, t-test, one-tailed matched pairs, 54 pairs for each comparison).

However back-channeling generally contributed far less than good greetings. Only in one case was the perceived value of back-channeling comparable to that of good pronunciation: in a forced choice between the good greeting / no back-channel stimulus and the missing greeting / good back-channeling stimulus, the judges were evenly split (9 to 9) on which of the hypothetical learners was "more likely to succeed in making someone want to help him."

An incidental finding was that the track intended to represent exemplary listening, BC 1, was not consistently rated better than the one designed to represent poor listening behavior, BC 2; indeed, it was ranked worse on the social effectiveness question. This could perhaps be due to an expectation by the judges that learners of Arabic will not behave like native speakers.

5 General Discussion

Experiment 1 showed that the pitch downdash is a cue to back-channels in Arabic, but that it is not perceived as such by naive Americans. Thus the interpretation of this cue is indeed culture-dependent rather than universal.

Experiment 2 showed that this cue is instead generally misinterpreted by naive Americans as expressing negative affect, and that a small amount of training suffices to prevent these misperceptions.

Experiment 3 showed that . producing back-channels can increase perceptions of the social effectiveness of a second-language learners.

The mechanics of turn-taking may seem harmless and mundane. Unlike gesture or emotional expressions, people in intercultural encounters may not expect these to be a source of misunderstandings. However these results show that misperceptions of turn-taking cues can occur.

It is interesting to compare this finding to other work on cross-cultural interpretations of prosodic features. Studies on the ability to identify the emotion expressed from the prosody of a speech sample have shown that the ability to correctly interpret emotions in such stimuli is weaker for speakers of other languages and members of other cultures, despite universal tendencies [25, 26, 27]. The novelty of this study lies in the focus on a turn-taking use of prosody, and in showing there are affective misinterpretations here too. It would be convenient if there were a clear distinction between the emotional and linguistic uses of prosody; that is, between the interactionally significant and the mundane. Within a single language such a distinction may exist, maybe even down to the neural level [29]. Cross-culturally, however, our results indicate that there is no clear line between emotional prosody and mundane linguistic prosody.

Since the current study used controlled stimuli, the findings clearly need further examination in more realistic contexts. However they do suggest that the study of intercultural interaction must address not only gesture, facial expression, politeness routines, cultural values etc., but also the prosodic mechanisms of

turn-taking and probably other phenomena of dialog mechanics and "close timing" in social interaction [1].

As the prosodic patterns relating to turn-taking are probably frequently encountered in intercultural exchanges (and perhaps more so than the prosodic patterns conveying emotion), the potential significance of this finding is large. For example, Americans needing to interact with Arabs probably need to learn, among many other things, the meaning of this prosodic pattern, both to give a good impression and to avoid misunderstanding. This is probably equally true whether the words used to communicate are English or Arabic, as speakers of foreign languages generally do not manage to overcome their native turn-taking patterns [28].

This finding may also be significant for those who, although not involved in intercultural interactions themselves, overhear exchanges from other cultures. For example, many radio interviews with speakers of foreign languages do not begin the translation voiceover until a few seconds after the speaker has started. The audience is thus led to gauge the speaker's personality and emotional state from those few seconds of speech. For interviews with Arabic speakers, it seems that a common place to start the first voiceover is after the first pause, which is in fact commonly a place where a back-channel would be welcome, and is often marked with this downslope. Along with other properties of Arabic — such as the pharyngeal phonemes, which are often associated with disgust in other languages, and the lack of de-accenting of stressed syllables, leading to wide pitch and volume range, which is associated with anger in some languages [30] — the pitch downslope could be leading to systematically mistaken impressions of Arabic speakers.

References

[1] Levinson S C (2006) On the Human 'Interaction Engine'. in *Roots of Human Sociality: Culture, Cognition and Human Interaction*, eds Enfield N J, Levinson S C (Berg, Oxford), pp 39-69.

[2] Tannen D (2005) Interactional Sociolinguistics as a resource for Intercultural Pragmatics. *Intercultural Pragmatics* 2:205-208.

[3] Gumperz J J (1982) *Discourse Strategies* (Cambridge University Press, Cambridge).

[4] Sacks H, Schegloff E A, Jefferson G (1974) A Simplest Systematics for the Organization of Turn-taking for Conversation. *Language* 50:696–735

[5] Shriberg E E (2005) Spontaneous Speech: How People Really Talk, and Why Engineers Should Care. in *Proc. Interspeech 2005.*

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

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

[8] Kraut R K, Steven H. Lewis S H, Swezey L W (1982) Listener Responsiveness and the Coordination of Conversation. *Journal of Personality and Social Psychology* 43:718-731.

[9] Bavelas J, Coates L, Johnson T (2000) Listeners as Co-Narrators. *Journal of Personality and Social Psychology* 79:941-952.

[10] Fujie S, Fukushima K, Kobayashi T (2005) Back-channel feedback generation using linguistic and non-linguistic information and its application to spoken dialogue system. in *Proc. Interspeech 2005* 889-892.

[11] Gratch J, Wang N, Gerten J, Fast E, Duffy R (2007) Creating Rapport with Virtual Agents. in *IVA 2007, LNAI 4722* (Springer, Berlin) pp 125–138.

[12] Ward N, Tsukahara W. (2000) Prosodic Features which Cue Back-Channel Feedback in English and Japanese *Journal of Pragmatics* 32:1177–1207.

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

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

[15] Senko K. Maynard S K (1989) *Japanese Conversation* (Ablex, Norwood NJ).

[16] Clancy P M, Thompson S A, Kuzuki R, Tao H (1996) The conversational use of reactive tokens in English, Japanese and Mandarin. *Journal of Pragmatics* 26:355–387.

[17] Yamada H (1992) *American and Japanese Business Discourse: A comparison of interactional styles*, (Ablex, Norwood NJ).

[18] Rost M (1990) *Listening in Language Learning* (Longman, London).

[19] Allwood J (1993) Feedback in Second Language Acquisition. in *Adult Language Acquisition: Cross Linguistic Perspectives, II: The Results*, ed Perdue C (Cambridge University Press, Cambridge) pp 196-235.

[20] Almaney A J, Alwan A J (1982) *Communicating with the Arabs* (Waveland Press, Prospect Heights, IL).

[21] Zaharna R S (1995) Bridging Cultural Differences: American Public Relations Practices and Arab Communication Patterns. *Public Relations Review* 21:241-255.

[22] Hafez O M (1991) Turn-taking in Egyptian Arabic: Spontaneous speech vs drama dialogue. *Journal of Pragmatics* 15:59-81.

[23] Ward N G, Al Bayyari Y (2006) A Case Study in the Identification of Prosodic Cues to Turn-Taking: Back-Channeling in Arabic. in *Proc. Interspeech 2006*.

[24] Ward N G, Al Bayyari Y (2007) A Prosodic Feature that Invites Back-Channels in Egyptian Arabic, in *Perspectives in Arabic Linguistics XX*, ed Mughazy M (John Benjamins, Amsterdam), pp 186-206.

[25] Scherer K R, Banse R, Wallbott H G (2001) Emotion Inferences from Vocal Expression Correlate Across Languages and Cultures. *Journal of Cross-Cultural Psychology* 32:76-91.

[26] Elfenbein H A, Ambady N (2002) On the Universality and Cultural Specificity of Emotion Recognition. *Psychological Bulletin* 128:203–235.

[27] Chen A, Gussenhoven C, Rietveld T (2004) Language-specificity in perception of paralinguistic intonational meaning. *Language and Speech* 47:311-350.

[28] Chun D M (2002) *Discourse Intonation in L2: From theory and research to practice* (John Benjamins, Amsterdam).

[29] Grandjean D, Banziger T, Scherer K R (2006) Intonation as an interface between language and affect. *Progress in Brain Research* 156:235-268.

[30] Hellmuth S. (2005) No de-accenting in (or of) phrases: Evidence from Arabic for cross-linguistic and cross-dialectal prosodic variation. in *Prosodies: With special reference to Iberian languages* eds Frota S, Vigario M, Freitas M J (Mouton de Gruyter, Berlin) pp 99-121.

[31] Bergsträsser G (1968) Zum arabischen Dialekt von Damaskus (On Damascene Arabic), (Georg Olms Verlagbuchhandlung, Hildesheim). originally published in 1924 by Orient-Buchhandlung Heinz Lafaire, Hannover, in the series Beitraäge zur semitischen Philologie und Linguistik

[32] Canavan A, Zipperlen G, Graff D (1997) CALLHOME Egyptian Arabic Speech (Linguistic Data Consortium, Philadelphia).

[33] Ward N G, Novick D G, Salamah S I (2006) The UTEP Corpus of Iraqi Arabic. Technical Report UTEP-CS-06-02 (University of Texas at El Paso Department of Computer Science, El Paso TX).

[34] Boersma P, Weenink D (2005) Praat: doing phonetics by computer [Computer program]. retrieved from http://www.praat.org/

Table 1: Average normalized appropriateness ratings (and standard deviations)

	downslope+BC	eight other pairings
Arab	**4.7** (1.3)	**4.3** (1.7)
naive Am.	**3.7** (0.9)	**4.1** (1.5)
exposed Am.	**4.5** (1.0)	**4.2** (1.5)

Table 2: "Does the speaker sound more positive or more negative?" average ratings (and standard deviations)

	downslope	control 1 (cadence)	control 2 (upturn)
Arab	**4.2** (1.3)	**3.9** (1.3)	**5.1** (1.4)
naive Am.	**3.2** (1.3)	**3.2** (1.3)	**5.3** (1.2)
exposed Am.	**4.0** (1.0)	**4.1** (0.9)	**6.0** (0.9)

[35] Ward N G, Rivera A G (2008, to appear) Prosodic Cues that Lead to Back-Channel Feedback in Northern Mexican Spanish. in *HDLS-7 Proceedings* (Cambridge Scholars Press, Cambridge).

[36] Ward N G, Escalante R, Al Bayyari Y, Solorio T (2007) Learning to Show You're Listening. in *Computer Assisted Language Learning* 20:385-407.

[37] Bradley M M, Lang P J (1999) Affective Nomrs for English Words (ANEW): Instruction Manual and Affective Ratings. Technical Report C-1 (The Center for Research in Psychophysiology, University of Florida, Gainesville FL).

Table 3: Ratings of "How well does the listener seem to know Arabic?" as a function of back-channeling and greeting quality

	Good Gr.	Poor Gr.	No Gr.	avg.
BC 1	**5.5** (1.5)	**3.8** (1.8)	**4.4** (1.5)	4.6
BC 2	**5.3** (1.4)	**4.0** (1.9)	**4.3** (1.0)	4.5
no BC	**5.2** (1.3)	**3.4** (2.0)	**3.8** (1.4)	4.1
avg.	5.3	3.8	4.2	4.4

Table 4: Ratings of "Does this person sound like a nice person?"

	Good Gr.	Poor Gr.	No Gr.	avg.
BC 1	**5.4** (1.2)	**4.6** (1.5)	**3.6** (1.8)	4.5
BC 2	**5.8** (1.7)	**4.9** (1.2)	**3.7** (1.5)	4.8
no BC	**5.0** (1.7)	**4.3** (1.7)	**3.3** (1.6)	4.4
avg.	5.4	4.6	3.5	4.5

Table 5: Ratings of "Is this person likely to succeed in making someone want to help him?'

	Good Gr.	Poor Gr.	No Gr.	avg.
BC 1	**5.4** (1.0)	**4.7** (1.2)	**4.2** (1.6)	4.8
BC 2	**5.7** (1.6)	**5.3** (1.4)	**4.1** (1.5)	5.0
no BC	**5.3** (1.3)	**4.7** (1.4)	**3.7** (1.7)	4.5
avg.	5.5	4.9	4.0	4.8

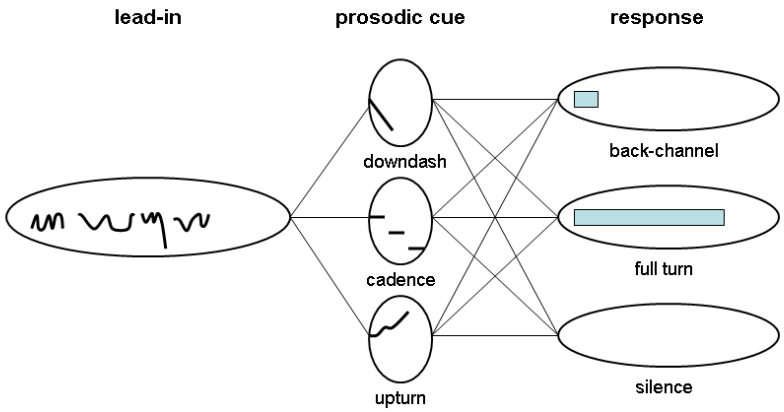

Figure 1: Schematic diagram of the first set of stimuli. Thick lines represent the pitch of utterances by the person in the talker role. Shaded rectangles represent the duration of utterances by the person in the listener role.

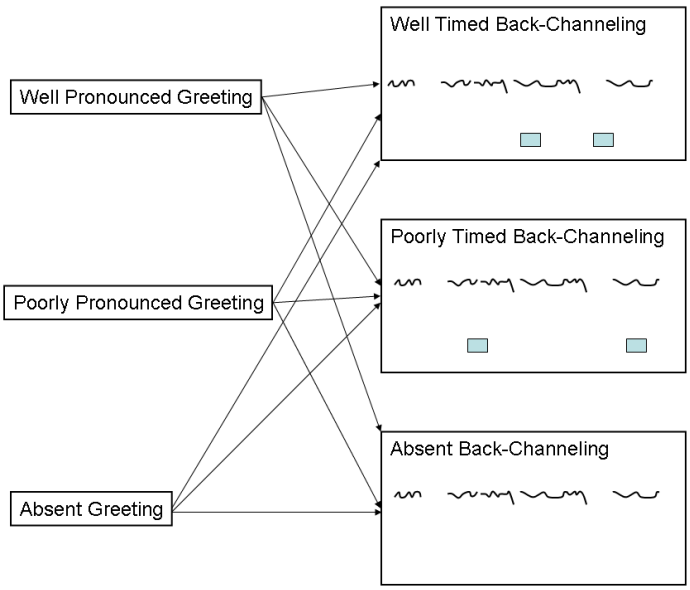

Figure 2: Schematic diagram of the third set of stimuli