University of Texas at El Paso, Department of Computer Science
Technical Report UTEP-CS-08-34
September 5, 2008

Additional information about 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: This technical report is a supplement to "American and Arab perceptions of an Arabic turn-taking cue", a paper submitted to the *Journal of Cross-Cultural Psychology*. It provides additional details, discussion, figures, tables, and references relating to the main finding, that English speakers tend to misinterpret the prosodic pattern used in Arabic to cue back-channel responses, perceiving it as an expression of negative affect. It also describes an experimental demonstration that being able to detect and respond to this prosodic pattern in dialog can increase native-speaker perceptions of the social effectiveness of learners.

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 Additional Background

The present work builds on the large body of research, in various traditions (1; 2), on the ways that interlocutors coordinate whose turn it is to speak. 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 (3). The detailed mechanics of back-channeling in particular, and the importance of this, had recently become clearer (4; 5; 6; 7; 8; 9; 10; 11).

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 languages and cultures differ back-channeling behavior (12), perhaps most saliently in the typical frequencies of back-channels in various languages (13; 14). There are also cases where such differences have caused cross-cultural misunderstandings (15). 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 (16; 17).

Studies of intercultural interpretations and misinterpretations of prosody have mostly focused on prosody as an indicator of the speech act (18) and on prosody as an overt expression of emotion; this study examines instead an interactional use of prosody.

2 Additional Information about 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.

The stimuli were various pairings of cues and responses, as suggested by Figure 1, including the downslope+back-channel combination and various controls. The first control 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 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 Arab subjects represented a number 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 subjects in Qatar were mostly acquaintances of the second author.

The "exposed American" subjects were those who had earlier had about 25 minutes of training in how to act as a good listener to an Arabic speaker by responding with back-channels to pitch downslope cues. The training sequence 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.

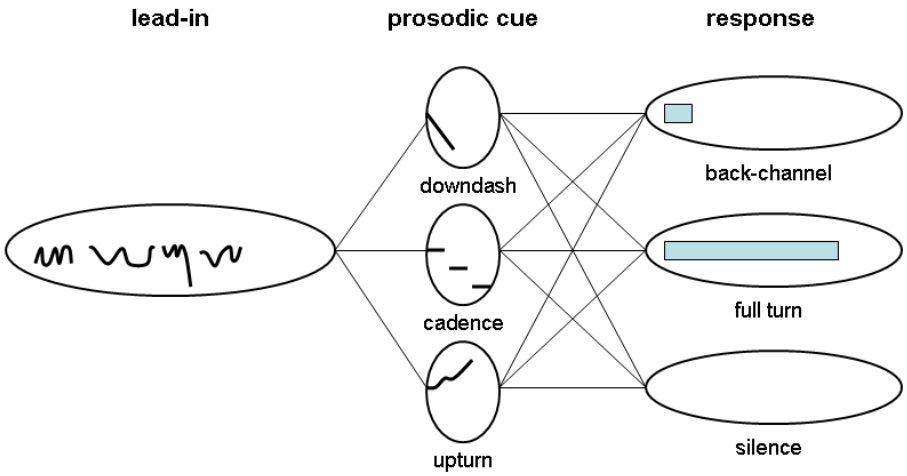

Figure 1: Schematic diagram of the stimuli for Experiment 1. Thick lines represent the pitch of utterances by the person in the talker role. Shaded rectangles represent the duration of utterances by the person in the listener role.

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)

The results are seen in Table 1.

3 Additional Information about 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.

Figure 3 is a schematic of the stimuli. The quantitative results are given in Table 2.

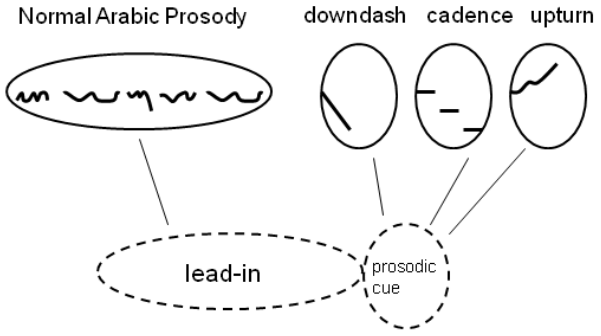

Figure 2: Stimuli for Experiment 2

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

	downslope	control 1 (cadence)	control 2 (upturn)
Arab	**4.0** (1.3)	**3.7** (1.3)	**4.9** (1.4)
naive Am.	**3.2** (1.3)	**3.2** (1.3)	**5.3** (1.2)
exposed Am.	**4.0** (1.0)	**4.2** (0.9)	**6.0** (0.9)

In addition to rating each stimulus, subjects were also asked to "please write two or three adjective describing the feeling (sad, angry, happy, surprised, scared, disgusted, etc.)." 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 (34) to compute the average valences: for the Americans the average rating 2.9, versus 3.5 for the Arabs, on a 9-point scale.

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?

These questions relates to the more general question of the value of the non-verbal modality in communication. In the past strong claims have been made (35), despite severe methodological problems facing those who attempt to quantify this (36; 37). Without seeking to establish any general claims, the purpose of our experiment was only to obtain a rough feeling for whether back-channeling matters.

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 a fellow student 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 are 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.

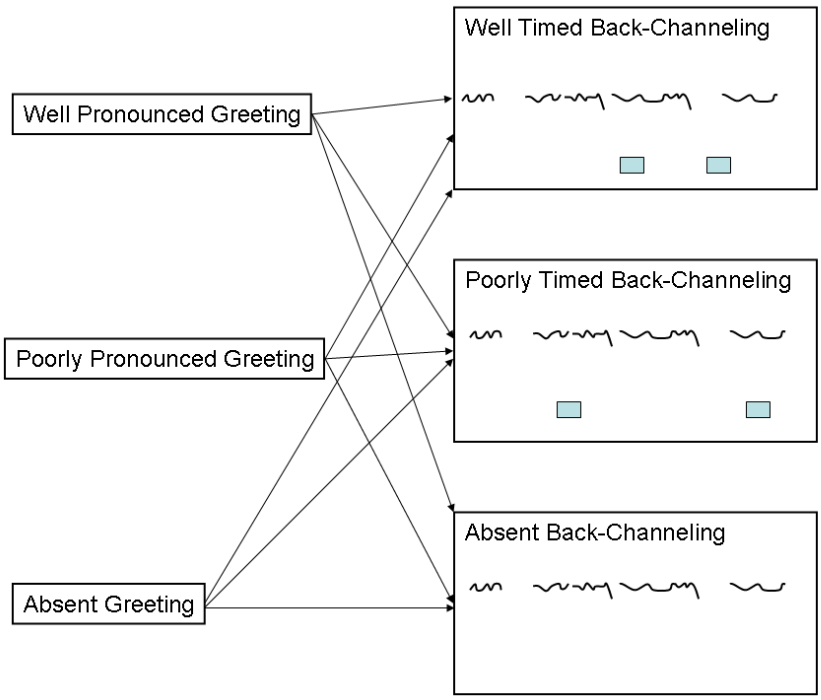

Figure 3: Schematic diagram of the stimuli for Experiment 3

4.1.2 Participants and Procedure

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

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 7-point scale. 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 Additional 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.

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 (22; 23; 24; 25). 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. As there is only a limited prosodic repertoire is used for both functions, and thus it is easy to undrstand why the same patterns can be used in different ways in different languages.

We have identified a new factor that could be leading to systematically mistaken impressions of

Arabic speakers, in addition to known high-level cultural differences between Arabs and Westerners (28; 29), and other known features of the language, notably 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).

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

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

References

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

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

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

[4] 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.

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

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

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

[9] 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*

[10] 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*.

[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] 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.

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

[14] 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.

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

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

[17] 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.

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

[19] El-Hassan S (1988) The intonation of questions in English and Arabic. in *Papers and Studies in Contrastive Linguistics, Volume Twenty-Two* pp 97-108

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

[21] 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.

[22] 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.

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

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

[25] Pell, Marc D, Vera Skorup (2008) Implicit processing of emotional prosody in a foreign versus native language. *Speech Communication*, 50: 519-530.

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

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

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

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

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

[35] Mehrabian A, Ferris S (1967) Influence of attitudes form non-verbal communication in 2 channels. *Journal of Consulting Psychology*, 31, pp 348-352.

[36] Furnham A, Trevethan R, Gaskell G (1981) The relative contribution ov verbal, vocal, and visual channels to person perception: Experiment and critique. *Semiotica*, 37, pp 29-57.

[37] Lapakko D (1997) Three Cheers for Language: A closer examination of a widely cited study of nonverbal communication. *Communication Education*, 46, pp 63-67.

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

[39] Ward N G, Rivera A G (2008) Prosodic Cues that Lead to Back-Channel Feedback in Northern Mexican Spanish. in Proceedings of the Seventh Annual High Desert Linguistics Society Conference, University of New Mexico.

[40] 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.