Effects of Culture on the Credibility of Robot Speech: A Comparison between English and Arabic

Sean Andrist,[1] Micheline Ziadee,[2] Halim Boukaram,[2] Bilge Mutlu,[1] Majd Sakr[3]
[1] University of Wisconsin–Madison, WI, USA
[2] American University of Science and Technology, Beirut, Lebanon
[3] Carnegie Mellon University, Pittsburgh, PA, USA
sandrist@cs.wisc.edu; mziadee@aust.edu.lb; hboukaram@aust.edu.lb;
bilge@cs.wisc.edu; msakr@cs.cmu.edu

ABSTRACT

As social robots begin to enter our lives as providers of information, assistance, companionship, and motivation, it becomes increasingly important that these robots are capable of interacting effectively with human users across different cultural settings worldwide. A key capability in establishing acceptance and usability is the way in which robots structure their speech to build credibility and express information in a meaningful and persuasive way. Previous work has established that robots can use speech to improve credibility in two ways: expressing practical knowledge and using rhetorical linguistic cues. In this paper, we present two studies that build on prior work to explore the effects of language and cultural context on the credibility of robot speech. In the first study (n = 96), we compared the relative effectiveness of knowledge and rhetoric on the credibility of robot speech between Arabic-speaking robots in Lebanon and English-speaking robots in the United States, finding the rhetorical linguistic cues to be more important in Arabic than in English. In the second study (n = 32), we compared the effectiveness of credible robot speech between robots speaking either Modern Standard Arabic or the local Arabic dialect, finding the expression of both practical knowledge and rhetorical ability to be most important when using the local dialect. These results reveal nuanced cultural differences in perceptions of robots as credible agents and have important implications for the design of human-robot interactions across Arabic and Western cultures.

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine Systems—*human factors, software psychology*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*evaluation/methodology, user-centered design*

General Terms

Design, Experimentation, Human Factors

HRI'15, March 02 - 05 2015, Portland, OR, USA

ACM 978-1-4503-2883-8/15/03$15.00
http://dx.doi.org/10.1145/2696454.2696464

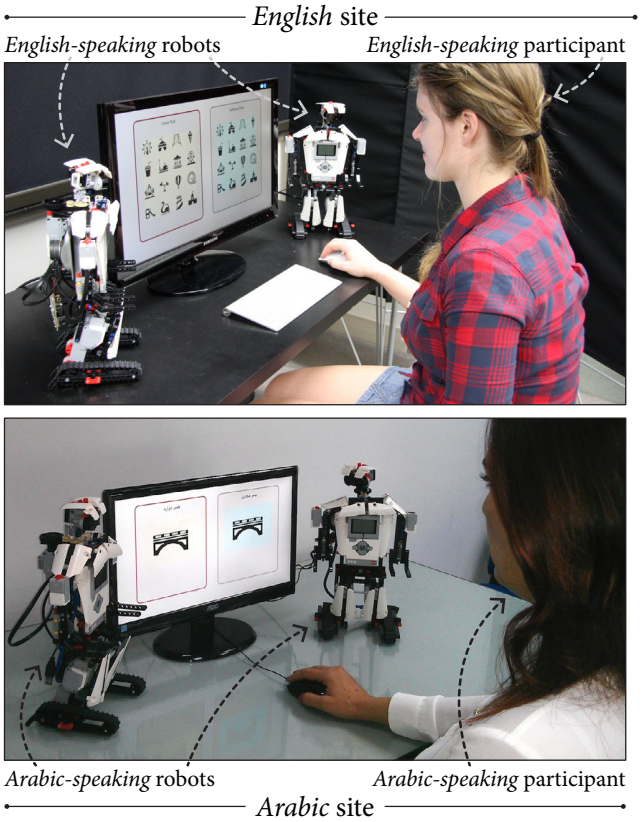

Figure 1: The first study was conducted in both the United States (top) and Lebanon (bottom). Participants interacted with two robots acting as competing tour guides, each speaking with a different degree of practical knowledge and/or rhetorical ability.

1. INTRODUCTION

People seek information for a variety of reasons, including comfort, support, empowerment, learning, and obtaining knowledge to act on. Social robots are increasingly taking on roles in the home and workplace as providers of such information to human users in a number of contexts. Critical to a robot's success in these roles is the perceived credibility of its speech in order to be persuasive and induce a positive change in knowledge, attitude, and behavior [2]. However, credibility is a multi-faceted construct that includes more than just the presentation of accurate information. Human communication research suggests that credibility is composed of

the message content, the receiver's background, and the speaker's characteristics [26]. While little is known about how robots can be perceived as credible sources of information, recent work provides a foundation for structuring the content of robot messages using both practical knowledge and rhetorical linguistic cues [2]. In this paper, we build on this foundation by investigating the effects of the receiver's background, particularly language and cultural background, on how a robot's users perceive and respond to its attempts to produce credible messages. Because language and culture are tightly intertwined and difficult—if not impossible—to study separately [1], we use the term *culture* to encompass both the native language and cultural background of the users of social robotic products. Previous research has used different terms for the same or similar construct, including "language community," [16] "languaculture," [1] and "ethnicity," [17] (ethno-linguistic or ethno-national groups).

A key motivation for studying the effects of culture on people's perceptions of and responses to robots as credible agents is the emerging global market for social robotic products. Cultures vary in terms of their perspectives on the role of language, structuring credible messages, and communicating effectively with audiences, all of which must be considered when designing informational robots that are to be deployed across cultures. One important cultural dimension is *context*—the environment and circumstances surrounding a speech act, including history, knowledge, use of pauses and silences, assumptions of values, and so on [6, 10]. For example, Western culture is described as "low-context," in which more meaning is placed in the language code itself and communication tends to be specific, explicit, and analytical. Alternatively, Arabic culture, which comprises the speakers of the Arabic language, is characterized as "high-context," in which there is extensive assumed shared context and much of the burden of meaning falls on the listener. We hypothesize that this cultural dimension will have a substantial impact on the credibility of robot speech and therefore used these two cultural contexts to conduct a comparison between English-speaking and Arabic-speaking robots.

This work investigates two research questions that motivated two studies. First, how does culture affect the efficacy of the expression of both practical knowledge and rhetorical linguistic cues in producing credible robot speech? To answer this question, we conducted a study in two sites—one in the United States and the other in Lebanon—that manipulated the use of knowledge and rhetoric for robots interacting with English-speaking and Arabic-speaking users. We chose the official Arabic language, Modern Standard Arabic (MSA), as the robot's speech in this study to better generalize our findings to Arabic culture as a whole. However, Arabic speakers' use of colloquial dialects when conversing in everyday situations motivated a second research question: how does an Arabic-speaking robot's use of either MSA or the local dialect affect the robot's credibility? To answer this question, we conducted a second study in Lebanon in which we manipulated robots' use of practical knowledge and rhetorical linguistic cues while speaking with MSA or the local dialect. The findings of the two studies have important implications for the design of human-robot interactions across Arabic and Western cultures.

2. BACKGROUND

In this section, we present previous cross-cultural research in human-robot interaction (HRI), background on Arabic culture that is relevant to designing informational robots in that cultural context, and human communication research on the use of rhetorical linguistic cues for producing credible and persuasive speech.

2.1 Cross-Cultural HRI

Culture underlies every aspect of social behavior and influences an individual's communication style, character, personality, knowledge, and motivation [21]. Šabanović [24] offers a perspective on how scientific fields approach the concept of culture in different ways. Social scientists tend to critique and problematize the concept of culture in ways that not only make it difficult to implement in technology but call into question the entire pursuit of culturally adaptive robots. On the other hand, robotics researchers seek to computationally define culture so that it can be boiled down and easily implemented in robotic products. In this paper, we take the perspective of HRI researchers, situated between social science and robotics, and seek to develop a deeper understanding of how cultural factors can be applied in the design of human-robot interactions.

A number of studies have explored the attitudes that people of different cultures hold with regards to their acceptance of social robots. In a questionnaire on attitudes towards interaction with robots across different cultures, American participants were found to be the least negative toward robots while Mexican participants were the most negative [3]. Lee et al. [15] used a generative design approach to determine expectations and desires in Korea and the United States for domestic robots, discovering two themes reflecting the ideals of each culture: relational and interdependent robots among Korean participants and utilitarian and independent robots among American participants. In a study including Chinese participants, a Chinese robot was perceived as more likely to know famous Hong Kong landmarks than an American robot, suggesting that people generalize from information about a particular robot to construct a mental model of the robot's capabilities dependent on the cultural context in which the robot operates [14].

Previous work has also examined cross-cultural differences in the ability of robots to change people's decisions and behaviors. Wang et al. [31] examined the effect of robots communicating with people while respecting or not respecting cultural norms. The study found that Chinese participants were more likely to change previously made decisions to align with a robot's recommendations when that robot was using an implicit communication style while U.S. participants aligned more with a robot using an explicit communication style. Other previous work focused on the effects of communication style and cultural background (Chinese or German) on the acceptance of recommendations from a robot [22]. Researchers found a strong preference among participants for the robot that communicated in the style more familiar to them (explicit for Germans, implicit for Chinese).

A small number of previous studies have investigated human-robot interaction in Arabic culture. Trovato et al. [30] found that Egyptian participants preferred interacting with a robot that used Arabic gestures and ways of speaking and felt uncomfortable when interacting with a robot using Japanese gestures and ways of speaking, while the opposite was true for Japanese participants. Another study examined politeness strategies for robots conversing with English and Arabic speakers, finding that Arabic participants gave higher ratings of politeness and competency to the robot than English speakers [25]. Verbal and nonverbal behaviors can be utilized to evoke associations between a robot's behaviors and its attributed ethnicity among English and Arabic speakers [17], and these groups also differ in how they perceive a robot's personality based on the amount of verbosity, hedging, alignment, and formality in its speech [16]. To our knowledge, there has been no previous work in HRI that explores the comparative effectiveness of robot speech designed to be credible across Western and Arabic culture, and the first study presented in this paper seeks to bridge this knowledge gap.

2.2 Arabic Language

When designing speech for robots in Arabic culture, it is important to consider that there is not a single Arabic language. Modern Standard Arabic (MSA) is the primary language of media, government, education, and public and religious speakers [13]. However, colloquial Arabic dialects have developed within countries and are the languages of everyday interaction. This development is the result of the interaction between different ancient dialects of Classical Arabic and other languages that existed in, neighbored, and/or colonized what is today the Arab World. Arabic dialects are generally restricted in use for informal daily communication and are not taught in schools, although there is a rich popular dialect culture of folktales, songs, movies, and television shows. Arabic dialects include Egyptian Arabic (Egypt and Sudan), Levantine Arabic (Lebanon, Syria, Jordan, Palestine and Israel), Gulf Arabic (Kuwait, United Arab Emirates, Bahrain, and Qatar), North African Arabic (Morocco, Algeria, Tunisia and Mauritania), and more.

There is a complex relationship between MSA and a region's specific dialect that linguists refer to as *diglossia*—the coexistence of two languages in everyday communication [8]. Diglossia typically entails the coexistence of a primary spoken dialect and a different, highly codified language contained in written literature and used in formal settings. Arabic speakers do not think of MSA and their particular dialect as separate languages; instead they coexist, serving different purposes. Although some Arabic speakers might have difficulty expressing themselves in MSA, most speakers understand it, as MSA is commonly used in written and formal communication, including speeches, newspapers, and news broadcasts [7]. MSA was chosen as the Arabic language in the first study in order to generalize to the wider Arabic society, but the second study investigates the diglossia that exists in Arabic culture, particularly the effect of a robot using either MSA or Levantine dialect when attempting to convey credible speech to Lebanese speakers.

2.3 Credibility, Persuasion, & Rhetoric

People encounter a great deal of information every day, most of which is filtered out in favor of what is deemed useful. One particularly important criterion used to filter information is its *credibility* [32]. New information that is rejected as not credible will not be learned or have any persuasive impact. Thus, credibility is a necessary but not sufficient component of any message meant to persuade in order to change mental states, attitudes, and behaviors [20]. The credibility of a message is a product of an interaction among source characteristics (e.g., appearance and trustworthiness), message characteristics (e.g., information content and rhetoric), and receiver characteristics (e.g., culture and previously held beliefs) [26]. Previous research found participant opinions on various topics to change more when presented with opposite opinions from "high prestige" (more trustworthy) sources, e.g., an academic journal, rather than less trustworthy sources, e.g., a gossip magazine [12]. Demonstrating credibility with skilled rhetoric can also lead to a listener's compliance, both in thought and behavior [18].

Findings on credibility from human-human interactions have been found to apply to human-computer interactions in which computers act as knowledge sources [9]. For example, recommender systems are sources designed to persuade their users. Recent research on these systems emphasizes the social aspects of their design, particularly the importance of integrating social cues to create more credible and persuasive systems [33].

Previous research in HRI has examined how a robot should establish credibility when interacting with users. Torrey et al. [29] demonstrated that robots should adapt the amount of information they present to the level of expertise of its user and the context of the interaction. Roubroeks et al. [23] studied reactions to robots either opposing or agreeing with their users and concluded that messages should be carefully worded to avoid a poor impression of the robot. Andrist et al. [2] demonstrated that two dimensions of speech—*practical knowledge*, which captures prior knowledge and experience, and *rhetorical ability*, which refers mainly to speaking prowess—contribute to the credibility of a robot's messages.

In this paper, we utilize the model of speech presented by Andrist et al. [2] to produce credible robot speech (illustrated in Figure 2). In this model, high rhetorical ability is created through the use of linguistic cues including goodwill, prior experience, metaphors, organization, and fluency. Expressions of *goodwill* indicate that the speaker wants what is best for the listener [11]. Effective speakers will also convey that the speaker has had *prior experience* giving credible information to previous listeners [27]. *Metaphors* help establish common ground between the speaker and the listener and indicate that the speaker is making an effort to connect with the listener and share his or her experience and expertise [11]. Highly metaphoric language and strong emotion are especially important in Arabic persuasive speech [28]. Rhetorical ability is also shaped by the *organization* of utterances [19]. Poor organization can damage the credibility of the speaker by creating the impression that he or she is not well-versed in the subject. Finally, the timing or *fluency* of speech is a key para-verbal cue for rhetorical ability [4]. There is some evidence that members of Arabic societies tend to speak quickly, possibly pointing to the particular importance of this cue for effective speech in Arabic [7]. In this work, we explore the relative effectiveness of a robot utilizing these rhetorical linguistic cues when speaking to English-speaking and Arabic-speaking users.

3. STUDY 1

The first study aimed to explore the effect of practical knowledge and rhetorical ability on the persuasiveness and perceived credibility of robots across Western and Arabic cultures. The study was conducted in two locations—a university campus in the United States and a university campus in Lebanon—to explore the commonalities and differences between robots using the speech model from prior work [2] to persuade English-speaking and Arabic-speaking participants. Both sites used identical robots and comparable lab spaces. The experimenter in Lebanon was a native Arabic speaker, while the experimenter in the United States was a native English speaker. They followed identical scripts and procedures, differing only in the language used. All survey instruments, written materials, verbal instructions, and robot speech were in Modern Standard Arabic for the Arabic participants and English for the American participants.

3.1 Participants

In total, 96 participants (45 females and 51 males) were recruited for this study, 48 in Lebanon and 48 in the United States. Participants

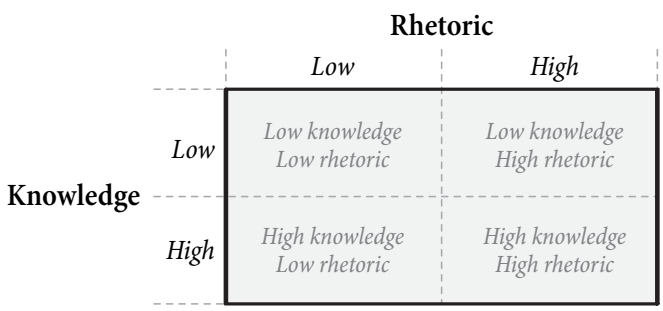

Figure 2: A model of expert speech adapted from Andrist et al. [2], divided into dimensions of knowledge and rhetorical ability.

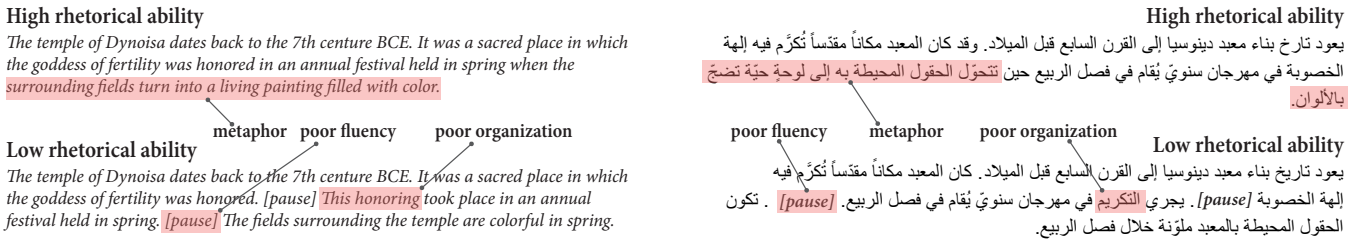

Figure 3: An example script utilizing high and low rhetorical ability in both English and Arabic. The script with high rhetorical ability is utilizing three of the five linguistic cues used in this work: metaphor, organization, and fluency.

in Lebanon were native Arabic speakers and participants in the United States were native English speakers. Ages ranged from 18 to 62 at the site in Lebanon (M = 28.27, SD = 10.24) and 18 to 46 at the site in the United States (M = 21.65, SD = 4.91).

3.2 Study Design

Participants interacted with two robots simultaneously, each embodying a different quadrant of the expertise model shown in Figure 2. We tested each of the six possible pairings of the four quadrants between-participants in both study locations, assigning eight participants at each site to each of the six pairing conditions (8 participants $\times$ 6 conditions $\times$ 2 locations = 96 total participants).

The robots acted as competing tour guides helping the participant plan a virtual tour through a fictional city. The experiment consisted of ten trials, each with a pair of landmarks for the participant to choose from. Each robot individually uttered a short description of the landmark while a computer monitor displayed the names and images of the landmarks. Following the robot's description, the user selected one of the landmarks by clicking on it (Figure 1).

Practical knowledge of the robot's utterances was manipulated by varying the number of discrete facts included in the descriptions of the landmarks; high knowledge scripts contained four discrete pieces of information, while low knowledge scripts contained two. Rhetorical ability was manipulated by varying the number of linguistic cues of expertise present in the robot's speech; high rhetorical ability scripts contained three of the linguistic cues presented in Section 2.3, while low rhetorical ability scripts contained none. An example script for one of the landmarks used in the study, a temple, is presented in Figure 3, highlighting the presence or absence of linguistic cues for both languages.

All scripts were first created in English and then translated into both Modern Standard Arabic and Lebanese dialect by an accredited translator. This translation was an iterative process. We conducted a short pretest to assess whether the rhetorical cues retained their effectiveness when translated into Arabic. A total of 24 participants, including 13 participants (six females and seven males) for MSA and 11 participants (six females and five males) for dialect, took part in the pretest. Each participant was presented with audio recordings of 20 pairs of sentences describing the landmarks. Each pair of sentences expressed the same level of knowledge but varied in rhetoric (high and low). Participants were asked to pick the sentence showing higher rhetorical ability. The data indicated that our manipulation of rhetorical ability using linguistic cues was successful for both MSA and dialect with the exception of "expressions of goodwill." With this cue, the speaker addresses the listener directly, e.g., "this is a unique experience that *you* will really enjoy." The initial translation used the gender-neutral Arabic plural pronoun for "you" because Arabic inflects for gender. This plural form, however, is a formal type of address in Arabic that automatically creates distance between the speaker and the listener. We decided to replace the use of the plural with the creation of a masculine and a feminine version of every utterance to be used for male and female participants. Finally, these new Arabic scripts were back-translated into English to ensure parallel content for both cultural sites.

3.3 Implementation & Procedure

This study made use of Lego Mindstorm robots in the humanoid ev3rstorm configuration[1] with minor modifications to the neck to allow the head to turn. They were placed on either side of the computer monitor displaying the software used for selecting between pairs of landmarks (Figure 1). The desktop software communicated through sockets to the robots running a custom leJOS operating system[2] that allowed the use of a custom service that accepted regular head-turning commands from the main software in order to emphasize the robots' embodied nature. The robots were given gender-neutral names and voices. Small external speakers placed behind them augmented their audio capabilities. The audio used for the robot voices were recorded using a female human voice and then pitch-shifted (one robotic voice is higher and one robotic voice is lower) and given a subtle robotic echo effect. The voices were randomly assigned to prevent any bias toward either of these voices.

Each participant was first greeted by the experimenter and signed a consent form. Participants were then seated directly facing the monitor and the experimenter explained that the robots would be acting as tour guides that would help the participant plan a virtual tour to a fictional city. The experimenter then left the room. Participants clicked a start button to begin the experiment. Each selection screen consisted of two landmark images representing the same type of landmark, e.g., two different zoos or two different museums. The border around each landmark lit up in turn and the respective (right/left) robot turned its head toward the user, provided some information about the landmark (appropriate to its knowledge/rhetoric condition), and then turned its head back to the monitor. The voice assigned to each robot was randomly generated for each participant. The order in which the pairs of landmarks were displayed as well as the order in which the robots took turns to describe a landmark were also randomized. After finishing all ten pairs of landmarks, the participants were given a tablet computer on which they completed a questionnaire dealing with their perceptions of the two robots. In total, the study took approximately 20 minutes to complete.

3.4 Measures & Analysis

We measured the perceived credibility of the robots' speech both objectively and subjectively. Objective credibility was measured indirectly through the robot's persuasiveness, operationalized as the amount of participant compliance to the robot's suggestions. We also created subjective measures of the participants' perceptions of the robots' persuasiveness, credibility, and sociability. Seven-

[1] http://www.lego.com/en-us/mindstorms/products/starter-robots/ev3rstorm/

[2] http://www.lejos.org/

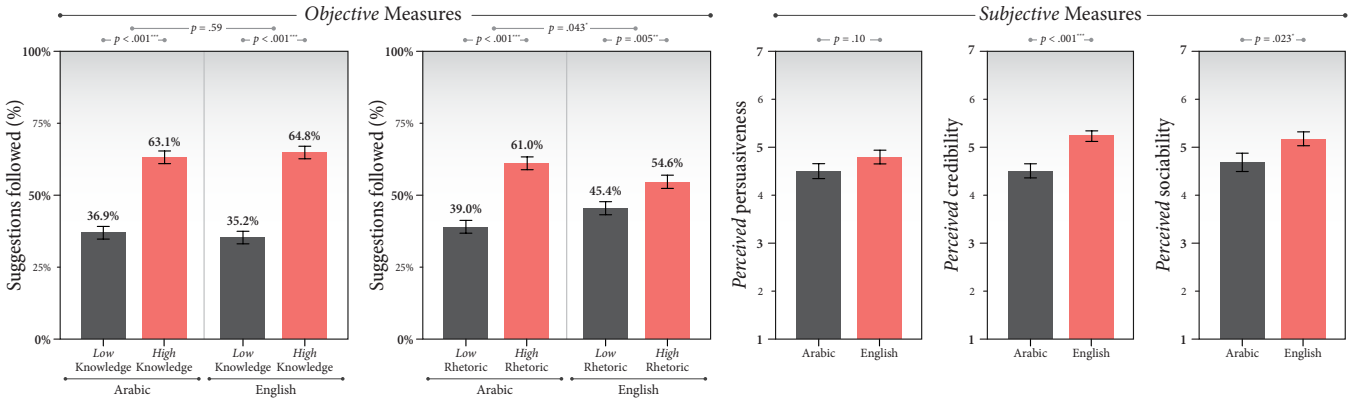

Figure 4: Results from Study 1. High knowledge and high rhetoric were effective in both English and Arabic, with rhetoric having more of an effect in Arabic. English-speaking participants assigned the robots higher ratings of credibility and sociability.

point rating scales were used for all items in these measures. Item reliabilities—measured by Cronbach's α—were sufficiently high for all measures (Table 1).

Our analysis of the objective results on participant compliance with robot suggestions was conducted using Pearson's Chi-Squared test on the categorical data of landmark choices. The statistical model included three independent variables: practical knowledge (high or low), rhetorical ability (high or low), and culture (American or Lebanese). Participant gender, robot voice, and robot speaking order were found to be nonsignificant predictors on all measures and excluded from further analysis. The impact of the continuous variable of participant age was analyzed with a nominal logistic regression in order to address the difficulty of achieving identical age distributions at both study sites. Finally, an Analysis of Variance (ANOVA) was utilized to determine the effect of the independent variables on the outcomes of the subjective rating scales.

3.5 Results

We first confirmed that the use of both dimensions of expert speech—practical knowledge and rhetorical ability—resulted in more persuasive utterances for the robots across both cultural groups. Participants chose landmarks presented by a robot with high knowledge (64%) more than those presented by a robot with low knowledge (36%), $\chi^2(1, N = 1920) = 149.63, p < .001$. Similarly for the rhetoric dimension of expert speech, participants chose landmarks presented by a robot with high rhetorical ability (57.8%) more than those presented by a robot with low rhetorical ability (42.2%), $\chi^2(1, N = 1920) = 46.88, p < .001$.

Table 1: The subjective measures used to capture the participants' impressions of the robots.

Measure *(Reliability)*	**Items**
Persuasiveness *(Cronbach's α = 0.94)*	*The robot was persuasive.* *The robot made travel recommendations like an expert.* *The robot was knowledgeable on the landmarks.* *The robot had a lot of expertise on the travel destinations.* *The robot seemed to have deep knowledge of the city and its landmarks.* *The robot convinced me to choose its landmarks.* *The robot was like a good salesperson.* *I would use the robot's suggestions in the future.* *The robot was a collaborator that helped me make good travel decisions.*
Credibility *(Cronbach's α = 0.84)*	*The robot was believable.* *The robot was trustworthy.* *The robot was informative.* *The robot was helpful.* *The robot was clear.* *I trust the information that the robot gave me.*
Sociability *(Cronbach's α = 0.83)*	*The robot was sociable.* *The robot was friendly.*

We next looked at the results within and across the two cultural groups. We found that high knowledge robots were complied with more than low knowledge robots in both English, $\chi^2(1, N = 960) = 84.02, p < .001$, and Arabic, $\chi^2(1, N = 960) = 66.15, p < .001$. We found no significant difference between the Arabic and American participants in terms of the number of high knowledge robot suggestions chosen, $\chi^2(1, N = 960) = 0.29, p = .59$. Participants also followed the suggestions of a robot with high rhetorical ability more than a robot with low rhetorical ability in both English, $\chi^2(1, N = 960) = 8.07, p = .005$, and Arabic, $\chi^2(1, N = 960) = 46.82, p < .001$. We discovered that Arabic participants were significantly more likely than the American participants to comply with suggestions from a robot with high rhetorical ability, $\chi^2(1, N = 960) = 4.01, p = .043$.

We also examined the difference in compliance between cultural groups within each of the four quadrants in Figure 2. We found no significant difference between Arabic and American participants in their choice of suggestions produced by a robot with low knowledge and low rhetorical ability, $\chi^2(1, N = 480) = 0.71, p = .40$, or by a robot with low knowledge and high rhetorical ability, $\chi^2(1, N = 480) = 2.15, p = .14$. We found that the American participants were significantly more likely than the Arabic participants to comply with a robot with high knowledge and low rhetorical ability, $\chi^2(1, N = 480) = 4.54, p = .033$, and found no significant difference in their compliance with robots with both high knowledge and high rhetorical ability, $\chi^2(1, N = 480) = 2.21, p = .14$.

Next we conducted a nominal logistic regression to determine the possible effect of participant age on the likelihood of complying with robot suggestions. We found a significant negative relationship between the age of the Arabic participants and their choice of landmarks described by a robot with high knowledge, $\chi^2(1, N = 480) = 5.81, p = .016$. No other significant relationships involving age were observed for either of the two cultural groups.

Finally, we analyzed the three subjective measures of persuasiveness, credibility, and sociability using an Analysis of Variance (ANOVA) with practical knowledge, rhetorical ability, and culture modeled as fixed effects. Our analysis found high knowledge robots to be rated as more persuasive than low knowledge robots, $F(1, 186) = 63.51, p < .001$, and similarly for robots with high rhetorical ability, $F(1, 186) = 12.59, p < .001$. We found no significant effect of culture on ratings of robot persuasiveness, $F(1, 186) = 2.66, p = .10$. Our analysis found high knowledge robots to also be rated as more credible than low knowledge robots, $F(1, 186) = 29.08, p < .001$, but found no significant difference on the dimension of rhetorical ability, $F(1, 186) = 0.89, p = .35$. We also found that the American participants gave significantly higher ratings of credibility than the Arabic participants, $F(1, 186) =$

$17.94, p < .001$. Finally, for the measure of sociability, our analysis uncovered higher ratings for high knowledge robots than low knowledge robots, $F(1, 186) = 26.34, p < .001$, and similarly for robots with high rhetorical ability, $F(1, 186) = 24.37, p < .001$. American participants also assigned higher ratings of sociability to the robots than the Arabic participants, $F(1, 186) = 5.28, p = .023$. The objective and subjective results are summarized in Figure 4.

3.6 Discussion

In this study, we first replicated the results found in previous work [2]. Consistent with the findings of that work, we found that a robot that described the landmarks with high knowledge—more discrete facts about the landmark—was more persuasive (measured through compliance) than robots with low knowledge and less facts. Similarly, we found that robots with high rhetorical ability—using more linguistic cues of expertise—were also complied with more than robots with low rhetorical ability.

An important finding of this study is that the rhetorical dimension of expert speech appears to be more effective in Arabic than in English. This finding aligns with the common "form over function" description of Arabic culture's perspective on language in comparison with Western culture [34]. Previous research has shown that Arabic speakers tend to view someone who uses language cleverly as well-educated and therefore qualified to make judgements, and that person's advice will then be followed [28]. On the other hand, information presented without linguistic eloquence may appear cold, impersonal, and ultimately unpersuasive, as supported by our finding that the Arabic participants complied significantly less than American participants with a robot that expressed high knowledge but low rhetorical ability. In Western culture, *ideas* (the knowledge content) are more central to persuasion while one's status or use of language (rhetorical ability) are less powerful [7]. We also found high knowledge to be more effective for the younger Arabic participants compared with older Arabic participants, similar to its greater effectiveness for the American participants. We speculate that this result might be attributed to a greater Western cultural influence on younger members of Arabic societies, but this claim would need further substantiation in future work.

In general, American participants had more positive perceptions of the robots' credibility and sociability compared with the Arabic participants. Previous work has similarly found that high-context cultural beliefs are associated with lower perceptions of a robot's credibility [31]. The authors posited that users in high-context cultures are less able to accept a robot with narrow communication skills and minimal nonverbal behaviors.

4. STUDY 2

Our first study compared robots' use of persuasive speech in English and Modern Standard Arabic. MSA was chosen in order to generalize to the rest of Arabic culture, but it is important to consider the diglossia present in Arabic culture. People use their own local dialect when speaking to each other in everyday situations. We designed a study to address this situation by directly comparing the persuasiveness of two robots using either the more formal Modern Standard Arabic or the less formal local dialect. The goal of the study was to investigate how the persuasive differences from using one or the other language might be affected by different levels of knowledge and rhetorical ability being expressed by each robot.

The second study was conducted in a single location—the same university in Lebanon—to explore the effect of speaking with either Modern Standard Arabic or the local Levantine dialect on a robot's persuasiveness and perceived credibility. The implementation, task, procedure, and measures are identical to Study 1.

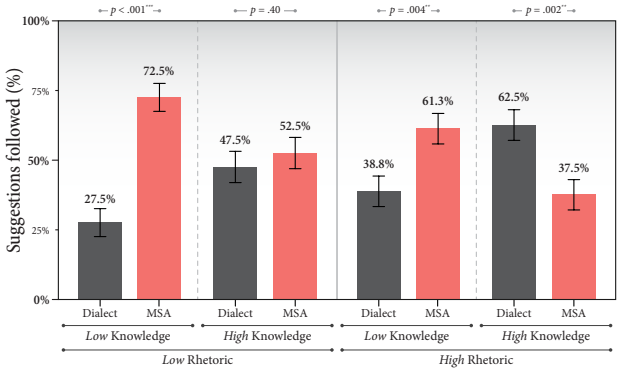

Figure 5: Results from Study 2. MSA was more effective when knowledge was low with both low and high rhetoric. Dialect was more effective when both knowledge and rhetoric were high.

4.1 Participants

Thirty-two participants (16 females and 16 males) who had not previously participated in Study 1 were recruited for this study in Lebanon. All participants were again native Arabic speakers and ages ranged from 18 to 57 ($M = 29.13$, $SD = 9.23$).

4.2 Study Design

This study utilized a mixed design with language (MSA or local dialect) varying within participants and knowledge (high or low) and rhetorical ability (high or low) varying between participants. Participants interacted with two robots simultaneously, each acting as a competing tour guide helping the participant plan a virtual tour and providing information on various fictional landmarks. The primary difference from Study 1 is that each robot embodied the same rather than different quadrants of the expertise model in Figure 2 and instead used a different Arabic dialect. One robot spoke using MSA and the other robot spoke using the local dialect. The assignment of language to each robot was random for each participant. We tested this language comparison in each of the four quadrants of expertise, with eight participants randomly assigned to each quadrant (8 participants $\times$ 4 quadrants = 32 total participants). For example, both robots might convey low knowledge and high rhetoric, but one spoke using MSA and the other with local dialect. All other aspects of this study were identical to Study 1.

4.3 Results

Similar to Study 1, our analysis of objective participant compliance was conducted using Pearson's Chi-Squared test with practical knowledge, rhetorical ability, and language (MSA or local dialect) as independent variables. Participant gender was found to be a significant predictor of compliance and is thus also considered in this analysis. The objective results are presented in Figure 5.

We found that robots speaking MSA were complied with more than robots speaking local dialect, $\chi^2(1, N = 640) = 9.03, p = .003$. Drilling down into each quadrant of knowledge and rhetorical ability in the model of expert speech (Figure 2), we found that with low knowledge and low rhetorical ability, robots speaking MSA were more persuasive than robots speaking local dialect, $\chi^2(1, N = 160) = 32.40, p < .001$. When speaking with low knowledge and high rhetorical ability, robots using MSA were also complied with significantly more than robots using local dialect, $\chi^2(1, N = 160) = 8.10, p = .004$. We found no significant language effect for robots speaking with high knowledge and low rhetorical ability, $\chi^2(1, N = 160) = 0.40, p = .53$. However, our analysis revealed

that when speaking with both high knowledge and high rhetorical ability, a robot using local dialect was complied with more than a robot using MSA, $\chi^2(1, N = 160) = 10.00, p = .002$.

Among female participants, we found no difference in compliance with robots using MSA or local dialect, $\chi^2(1, N = 320) = 0.00, p = 1.00$. However, we found a significant preference among male participants to comply with a robot using MSA than a robot using local dialect, $\chi^2(1, N = 320) = 18.05, p < .001$.

Our analysis revealed no significant effect of language on any of the subjective ratings. These results are excluded for brevity.

4.4 Discussion

The presence of a diglossia—the coexistence of two languages in everyday communication—is a nuance of Arabic culture that must be considered in cross-cultural language comparisons between Arabic and Western cultures. Our second study sought to address this complication in the context of credible speech by directly comparing the persuasiveness of two robots using either the more formal Modern Standard Arabic or the less formal local dialect. This study revealed a tradeoff in effectiveness between the formal use of MSA and the overall level of expertise conveyed in the robot's speech. When both robots had a relatively small amount of factual content to present about each landmark in the virtual tour, regardless of rhetorical ability, participants perceived the formality of MSA as a cue indicating higher credibility for that robot, and thus were persuaded more to comply with it. On the other hand, when both robots expressed high knowledge and high rhetorical ability for their landmark descriptions, participants complied more with the robot speaking their own local dialect, possibly indicating a familiarity preference for interacting with a robot that speaks more like them when it has more to say.

5. GENERAL DISCUSSION

These studies were designed to explore the impact of language and cultural context—native English speakers in America vs native Arabic speakers in Lebanon—on the ability of a robot to convey credible messages. The first study revealed the particular importance of rhetorical linguistic cues for robots to be perceived as credible to Arabic users. It also showed Arabic users to be generally more critical of the perceived credibility and sociability of robots, as compared with American users. The second study explored the use of either Modern Standard Arabic or the local dialect for robots speaking to Arabic users, revealing the use of dialect to be most effective when the robot is knowledgeable and expresses high rhetorical ability.

These studies suggest a number of key implications for the design of credible robot speech. First, it is important that designers consider the cultural context in which their robots will be embedded, especially when the users' culture is different from that of the designer. Different cultures take different perspectives on the use of language to inform and persuade, as exemplified by the differences found in high-context and low-context cultures. Specific cultural nuances must also be accounted for in design decisions, such as the presence of diglossia within Arabic culture.

A specific implication of this work relates to the finding that rhetorical linguistic cues were more effective for the Arabic-speaking robots than for the English-speaking robots. To date, HRI designers have mostly focused on building robots that are knowledgeable and accurate with conveying information, but we have shown that rhetoric is critical in an Arabic cultural context for conveying credibility. Robots that are to be deployed in this culture will need to make use of these linguistic cues in order to be utilized and recognized as a valued source of information.

This research also illuminates a potential risk in designing effective robot speech with high rhetoric and low knowledge, especially in Arabic. Speaking persuasively without expressing enough factual content could be dangerous and designers should exercise some caution in this situation. For example, if a robot is presenting prescription drugs to patients, sounding like an expert without actually possessing the requisite knowledge about the drug or the user's specific situation might have unintended harmful consequences. In general, if a robot has low knowledge on a specific subject, it should use low rhetorical ability and remain harmless. If it has high knowledge, it should attempt to generate the rhetorical ability to be perceived as a true expert on the subject.

The final implication of this work has to do with the choice a designer will have to make when deciding between MSA or a local dialect for the robot's speech. In settings where the robot has a relatively small amount of information to express to its users, the study results indicate that MSA affords the perception of credibility. Using MSA will also make the robot easily adaptable to all countries in the Arab World. However, if the robot is to reach its full potential for producing credible speech with both practical knowledge and rhetorical linguistic cues, it should use the local dialect of its users.

While it is possible that not all Arabic users are fully literate in MSA and thus might always prefer a robot to speak the local dialect, most Arabs at least understand MSA when it is spoken to them, as it is commonly used in many daily interactions, including all written and formal communication, speeches, newspapers, and news broadcasts [7]. Furthermore, we did not observe in either Study 1 or 2 any indication of participants having difficulty understanding MSA, from the reading of consent forms written in MSA to the instructions of the experimenter spoken in MSA. Education level may be an important predictor of MSA literacy. However, as we did not measure education level and recruited our Lebanese participants from a university population with at least some college education, further research is essential to understand whether the broader population of Arabic speakers are equally literate in MSA.

5.1 Limitations & Future Work

Credibility is composed of the source characteristics, message characteristics, and receiver characteristics [26]. In this work, we manipulated aspects of the message—knowledge and rhetoric—and compared the effect across two different receiver groups—American and Lebanese users. The robot (source) was held constant throughout, but future research should also examine the effect of physical characteristics of the robot itself—including size, shape, and humanlikeness—on its credibility and persuasiveness.

This work utilizes and extends a previous model of expert speech for robots that was developed in a Western culture, so it is potentially missing some rhetorical linguistic cues that are specific to Arabic culture. For example, cues of persuasive rhetoric in Arabic also include repetition, exaggeration, and assertion [7, 28], which should be further explored. Nonverbal communication patterns such as gesture, eye contact, touch, and interpersonal distance are also very important for facilitating the perceived credibility of robots [5], particularly in high-context cultures [7], and should be further explored in future work. Furthermore, we only investigated the effects of the presence and absence of linguistic cues in the credibility of robot speech. Future work should also explore the relative importance and effectiveness of each verbal or nonverbal cue.

Study 2 was designed from a practical HRI perspective; designers must choose between MSA and local dialect for robots to speak when deployed in any Arabic context. We explored this choice in Lebanon within the framework of an existing model of expert speech, revealing interesting differences and potential tradeoffs. However, future work should further explore the nuances of the differences

that arise from other cultural factors, such as attitudes toward each dialect and different regional factors, as well as demographic factors, such as gender and religion.

In order to be useful in real settings outside of the laboratory, robots will need to express high rhetorical ability by generating linguistic cues of expertise automatically. Although there is currently a lack of computational methods for generating linguistic cues such as metaphors and expressions of goodwill, automatic methods for organizing fluent utterances should be possible. In addition, culturally adaptive agents should be designed to automatically detect different forms of communicative behaviors across various cultures and react to them appropriately.

6. CONCLUSION

In this paper, we presented work to explore two research questions. First, how does language and cultural context impact the effectiveness of robot speech that has been designed to be credible? And second, how does the use of either Modern Standard Arabic or the local colloquial Arabic dialect impact the robot's credibility? Our first study found that rhetorical linguistic cues are particularly important for robots to speak effectively in Arabic cultural contexts, while our second study revealed a tradeoff in effectiveness between the credibility of the robot's speech and its use of either MSA or the local dialect. This work addresses particular challenges for HRI in Arabic culture, a large and diverse set of people worldwide that have been largely unrepresented in previous work. In general, it is imperative that social robots conform to the cultural norms of their intended users, in the same way that people must be cognizant of the cultural norms of people with whom they work and places they visit. Social interaction, culture, and language are complex constructs and designing for them is difficult, but by making progress toward understanding the nuances of these concepts, we can create effective and rewarding robot behaviors that positively affect people's lives throughout the world.

Acknowledgments

We would like to thank Brandi Hefty and Jieni Peng for their help in conducting the experiments.

7. REFERENCES

[1] M. Agar. *Language shock: Understanding the culture of conversation*. William Morrow and Company, 1994.

[2] S. Andrist, E. Spannan, and B. Mutlu. Rhetorical robots: making robots more effective speakers using linguistic cues of expertise. In *Proc. HRI '13*, pages 341–348, 2013.

[3] C. Bartneck, T. Nomura, T. Kanda, T. Suzuki, and K. Kato. Cultural differences in attitudes towards robots. In *Proc. AISB '05: Symposium on Robot Companions*, pages 1–4, 2005.

[4] R. Carlson, K. Gustafson, and E. Strangert. Cues for hesitation in speech synthesis. In *Proc. Interspeech '06*, volume 6, 2006.

[5] V. Chidambaram, Y.-H. Chiang, and B. Mutlu. Designing persuasive robots: how robots might persuade people using vocal and nonverbal cues. In *Proc. HRI '12*, pages 293–300, 2012.

[6] J. B. Dozier, B. W. Husted, and J. T. Mcmahon. Need for approval in low-context and high-context cultures: A communications approach to cross-cultural ethics. *Teaching Business Ethics*, 2(2):111–125, 1998.

[7] E. Feghali. Arab cultural communication patterns. *Int J Intercult Rel*, 21(3):345–378, 1997.

[8] C. A. Ferguson. Diglossia revisited. *Southwest J Linguistics*, 10(1):214–234, 1991.

[9] B. Fogg and H. Tseng. The elements of computer credibility. In *Proc. CHI '99*, pages 80–87, 1999.

[10] E. T. Hall. *Beyond culture*. Anchor Books, 1976.

[11] E. J. Hartelius. *The rhetoric of expertise*. PhD thesis, The University of Texas at Austin, 2008.

[12] C. I. Hovland and W. Weiss. The influence of source credibility on communication effectiveness. *Public Opin Quart*, 15(4):635–650, 1951.

[13] A. S. Kaye. Diglossia: the state of the art. *Int J Sociol Lang*, pages 117–130, 2001.

[14] S. Kiesler. Fostering common ground in human-robot interaction. In *Proc. RO-MAN '05*, pages 729–734, 2005.

[15] H. R. Lee, J. Sung, S. Šabanović, and J. Han. Cultural design of domestic robots: a study of user expectations in Korea and the United States. In *Proc. RO-MAN '12*, pages 803–808, 2012.

[16] M. Makatchev and R. Simmons. Perception of personality and naturalness through dialogues by native speakers of american english and arabic. In *Proc. SIGDIAL '11*, pages 286–293, 2011.

[17] M. Makatchev, R. Simmons, M. Sakr, and M. Ziadee. Expressing ethnicity through behaviors of a robot character. In *Proc. HRI '13*, pages 357–364, 2013.

[18] G. Marwell and D. Schmitt. Dimensions of compliance-gaining behavior: An empirical analysis. *Sociometry*, pages 350–364, 1967.

[19] J. Noel. On the varieties of phronesis. *Educ Philos Theory*, 31(3):273–289, 1999.

[20] D. J. O'Keefe. Trends and prospects in persuasion theory and research. *Perspectives on persuasion, social influence and compliance gaining*, 2004.

[21] P. O'Neill-Brown. Setting the stage for the culturally adaptive agent. In *Proc. AAAI Fall Symposium on Socially Intelligent Agents*, pages 93–97, 1997.

[22] P. Rau, Y. Li, and D. Li. Effects of communication style and culture on ability to accept recommendations from robots. *Comput Hum Behav*, 25(2):587–595, 2009.

[23] M. Roubroeks, J. Ham, and C. Midden. The dominant robot: Threatening robots cause psychological reactance, especially when they have incongruent goals. In T. Ploug, P. Hasle, and H. Oinas-Kukkonen, editors, *Persuasive Technology*, pages 174–184. Springer, 2010.

[24] S. Šabanović, C. C. Bennett, and H. R. Lee. Towards culturally robust robots: A critical social perspective on robotics and culture. In *Proc. HRI '14: Workshop on Culture-Aware Robotics*, 2014.

[25] M. Salem, M. Ziadee, and M. Sakr. Marhaba, how may I help you?: effects of politeness and culture on robot acceptance and anthropomorphization. In *Proc. HRI '14*, pages 74–81, 2014.

[26] C. S. Self. Credibility. *An integrated approach to communication theory and research*, 1:421–441, 1996.

[27] J. Sniezek and L. Van Swol. Trust, confidence, and expertise in a judge-advisor system. *Organ Behav Hum Decis Process*, 84(2):288–307, 2001.

[28] J. Suchan. Toward an understanding of Arabic persuasion: a Western perspective. *Int J Bus Commun*, 51(3):279–303, 2014.

[29] C. Torrey, A. Powers, M. Marge, S. Fussell, and S. Kiesler. Effects of adaptive robot dialogue on information exchange and social relations. In *Proc. HRI '06*, pages 126–133, 2006.

[30] G. Trovato, M. Zecca, S. Sessa, L. Jamone, J. Ham, K. Hashimoto, and A. Takanishi. Cross-cultural study on human-robot greeting interaction: acceptance and discomfort by Egyptians and Japanese. *Paladyn*, 4(2):83–93, 2013.

[31] L. Wang, P.-L. P. Rau, V. Evers, B. K. Robinson, and P. Hinds. When in Rome: the role of culture & context in adherence to robot recommendations. In *Proc. HRI '10*, pages 359–366, 2010.

[32] C. N. Wathen and J. Burkell. Believe it or not: Factors influencing credibility on the web. *J Am Soc Inf Sci Tec*, 53(2):134–144, 2002.

[33] K.-H. Yoo, U. Gretzel, and M. Zanker. *Persuasive Recommender Systems*. Springer, 2012.

[34] R. S. Zaharna. Understanding cultural preferences of Arab communication patterns. *Public Relat Rev*, 21(3):241–255, 1995.