Communication Quarterly

Routledge
Taylor & Francis Group

ISSN: 0146-3373 (Print) 1746-4102 (Online) Journal homepage: http://www.tandfonline.com/loi/rcqu20

# "You Can Do It Baby": Non-Task Talk with an In-Car Speech Enabled System

**Elizabeth Molina-Markham, Brion van Over, Sunny Lie & Donal Carbaugh**

**To cite this article:** Elizabeth Molina-Markham, Brion van Over, Sunny Lie & Donal Carbaugh (2016): "You Can Do It Baby": Non-Task Talk with an In-Car Speech Enabled System, Communication Quarterly, DOI: 10.1080/01463373.2015.1103289

**To link to this article:** http://dx.doi.org/10.1080/01463373.2015.1103289

*Communication Quarterly*
*Vol. 00, No. 00, 2016, pp. 1–24*

Routledge
Taylor & Francis Group

# "You Can Do It Baby": Non-Task Talk with an In-Car Speech Enabled System

Elizabeth Molina-Markham, Brion van Over, Sunny Lie, & Donal Carbaugh



*Natural language speech enabled systems are an attractive option for in-car infotainment. Differences in cultural expectations in communication, however, can pose difficulties for interface developers and cause frustration for users. We adopt the perspective of cultural discourse theory to analyze 26 drivers interacting with an in-car speech interface. We focus here on directive sequences and the phenomenon of participants using non-task talk. The analysis of these sequences reveals a norm that one ought to engage in non-task talk with the system. We suggest that this norm is grounded in a user premise that the system's interactional status involves the ability to speak. We find that this norm lacks crystallization among participants, and we formulate a competing norm that helps to account for this. The second norm reveals an underlying belief that the system's status as a machine is the basis for how it should be treated.*

*Keywords:* *Cultural Discourse Theory; Directives; Face; Politeness; Speech Enabled System*

There is increasing interest in developing natural language speech enabled interfaces for use when interacting with mobile technologies, such as mobile phones. Winter, Shmueli, and Grost (2013) wrote that these interfaces also "offer a compelling choice of user interaction for the automotive market" because they

Elizabeth Molina-Markham (Ph.D., University of Massachusetts Amherst, 2011) is an independent scholar. Brion van Over (Ph.D., University of Massachusetts Amherst, 2012) is an instructor in the Department of Communication at Manchester Community College. Sunny Lie (Ph.D., University of Massachusetts Amherst, 2013) is an assistant professor in the Department of Communication Studies at St. Cloud State University. Donal Carbaugh (Ph.D., University of Washington, 1984) is a professor in the Department of Communication at the University of Massachusetts Amherst. *Correspondence*: Elizabeth Molina-Markham, P.O. Box 1048, Arlington, MA 02474; Email: emolmark@gmail.com

ISSN 0146-3373 print/1746-4102 online 
DOI: 10.1080/01463373.2015.1103289

can be easier to use than touch interfaces and do not require users to memorize key words for commands (p. 1). However, designing these systems is more complicated for the developer because of the need to create a highly flexible interface that takes into account user experiences and expectations in communication (Carbaugh, Molina-Markham, van Over, & Winter, 2012, 2013). Researchers have suggested that variations in communication practices have a cultural dynamic and are important to consider in designing a speech enabled interface because they have a strong influence on users' understandings of and receptivity to interactions (Tsimhoni, Winter, & Grost, 2009). In our analysis, we employ cultural discourse analysis to examine the characteristics of communication events when a driver interacts with an in-car speech enabled interface. Our goal is twofold: (a) to formulate communication norms and cultural premises that guide these events in order to provide a better understanding of these types of interactions, and, in turn, (b) to suggest ways for improving the design of future technology.

In this article, we perform cultural discourse analysis to examine directive sequences in which participants interact with an in-car speech enabled interface. Participants use the system to accomplish the tasks of listening to the radio, listening to their music, or making a phone call. In human-human interaction, the act of giving a directive (for example, telling another person to act in a certain way) has implications for the relationship between individuals and thus for issues concerning considerations of face and politeness (Brown & Levinson, 1978; Goffman, 1967; Grice, 1975; Lakoff, 1973; Searle, 1990; Shahrokhi & Bidabadi, 2013). Giving a directive has the potential to impinge upon the sense of self of the person to whom the directive is given, so how one chooses to formulate a directive reveals a great deal about the speaker's view of the addressee (Blum-Kulka, 1997). Research on interactions between people and computers has suggested that people will often interact with a system similarly to how they would interact with a person, even though they may recognize that their interactional partner is non-human (Nass & Brave, 2005; Turkle, 2011). In our analysis, we find that two competing cultural norms are active in the interaction when drivers engage with a speech-enabled interface in their car—one norm that one ought to engage in non-task talk with the system and a competing norm that non-task talk is unnecessary—and we suggest that the norms we identify activate cultural premises of personhood that constitute the car as an interactional partner in distinctive ways. In other words, how participants go about telling their in-car system what to do speaks about how they view that system and its potential for interaction—that is, whether they view the non-human system as similar to a human interactant or not.

These competing cultural norms in communication have likely developed as a result of new speech-enabled technologies that force participants to consider different types of interactional partners, such as computers. Whereas previously, it was often possible to engage with these types of technology through tactile means, such as pressing a key or touching a screen, interactions through voice are becoming more prevalent in daily life, and individuals must now consider

what type of interactional partner a computer actually is or indeed what they want it to be. These findings have significant implications for the development of new culturally adaptive speech systems and for increased understanding of how people orient to technology more generally, such as in the form of their smartphones or tablets.

## Research on Design, Human-Machine Interaction, and In-Car Communication

Our examination of cultural norms in communication brings together several research areas, including research on design, on human-machine interaction, and on in-car communication between humans. Jokinen and McTear (2009) provided an overview of the design and development of spoken dialogue systems, dividing these systems into two main types: "task-oriented systems," which "involve the use of dialogues to accomplish a task," and "nontask-oriented systems," which "engage in conversational interaction, but without necessarily being involved in a task that needs to be accomplished" (p. 1). According to Jurafsky and Martin (2009), the goal of the field of speech and language processing is "to get computers to perform useful tasks involving human language, tasks like enabling human-machine communication, improving human-human communication, or simply doing useful processing of text or speech" (p. 1). Models and theories for research on speech and language processing draw from research in computer science, mathematics, electrical engineering, linguistics, and psychology, among others, in order to design systems that can use a knowledge of language in order to accomplish speech recognition and speech synthesis. Speech recognition and synthesis require knowledge of phonetics and phonology, morphology, syntax, semantics, pragmatics, and discourse. Models and algorithms are used to resolve ambiguities in language usage so that computers can determine the meaning of utterances by determining such elements as parts of speech, the sense in which words are used, and sentence types.

In recent years, the design and development of systems for interaction with humans has moved away from an engineering-driven approach and toward user-centered design (UCD), or "a multidisciplinary design approach based on the active involvement of users to improve the understanding of user and task requirements, and the iteration of design and evaluation" (Mao, Vredenburg, Smith, & Carey, 2005, p. 105). Designers view UCD as a way to make systems more useful and usable by taking into account how people use technology in their larger environmental context (Holtzblatt & Beyer, 2014). For example, previous research has emphasized the general importance of cultural context in design. Based on case studies in India and South Africa, Mäkäräinen, Tiitola, and Konkka (2001) proposed the need to consider that cultural factors in design should go beyond language issues. Drawing on a case study in Namibia on cultural biases concerning the concept of usability, Winschiers and Fendler (2007) argued that design methods result in locally inappropriate evaluations of usability. They suggest that researchers should not rely on their own assumptions about usability but instead should

actively and explicitly confirm through empirical studies the contextual meaning of these criteria in different places.

Reviewing research on human-machine interaction reveals that in many situations, people treat machines similar to human interactants (Friedman, 1997; Jurafsky & Martin, 2009; Nass & Yen, 2010; Reeves & Nass, 1996). Jurafsky and Martin (2009) wrote, "It is now clear that regardless of what people believe or know about the inner workings of computers, they talk about them and interact with them as social entities" (p. 8). According to Nass and Brave (2005), numerous experiments show that "the human brain rarely makes distinctions between speaking to a machine—even those machines with very poor speech understanding and low-quality speech production—and speaking to a person" (p. 4). In her research on human relationships with sociable robots, Turkle (2011) described the current state of human interaction with computers as the "robotic moment" (p. 9). With this phrase, she described users' willingness or "readiness" to treat robots as valid partners with whom they can have relationships (p. 9). Turkle (2011) argued that people are not necessarily deceived into thinking technology is alive, but they are willing to "fill in the blanks" and interact with robots as if they could form a relationship with them because robots "perform understanding" (p. 24–25). Turkle's concept of a robotic moment emphasizes that how a robot interacts is more important in shaping how some individuals will respond to the robot than is the fact that the robot is not a person. As a consequence, within the field of speech and language processing, some researchers focus on the design of "conversational agents," or "artificial entities that communicate conversationally" (Jurafsky & Martin, 2009, p. 8).

Prior research on in-car communication has primarily focused on human-human interaction. For example, researchers have examined fine-grained sequencing of human interactions within the car (Haddington, 2010; Laurier, 2005; Laurier, Brown, & Lorimer, 2007; Laurier et al., 2008), the interactional management of distractions in the car (Haddington & Keisanen, 2009; Koppel, Charlton, Kopinathan, & Taranto, 2011), and how speaking within the car is tied to social roles (Laurier, 2011; Laurier, Brown, & Lorimer, 2012). However, research has only just begun to explore interaction between humans and speech interfaces in the car as a site of cultural variation (Carbaugh et al., 2012; Tsimhoni et al., 2009; Winter, Tsimhoni, & Grost, 2011; Winter et al., 2013). This article builds on this past work in order to explore cultural norms that are active in these situations (Carbaugh et al., 2013). One of the cultural norms—that of using non-task talk in interaction with the in-car system—suggests a case in which human-machine interaction parallels human-human interaction. We note, however, that this norm of interaction seems to lack support among some users, perhaps because competing norms exist that are informed by differing cultural premises about the proper relationship between machines and humans.

## Approach

In this article, we adopt the perspective of the ethnography of communication (Hymes, 1962, 1972). The general approach we employ in analyzing communication

practices used by drivers when interacting with a speech enabled system is cultural discourse analysis (Berry, 2009; Carbaugh, 2007, 2012; Scollo, 2011). Focal concerns in cultural discourse analysis (CuDA) include how a message is put together (its discursive structuring), how it flows over time (its sequencing), what it does in social situations (its interactional functions), and participants' meanings about all of this. In CuDA, analysts examine key cultural sequences, terms, or vocabulary, which can be used by analysts to formulate cultural propositions, or statements that capture participants' meanings about communication itself (Carbaugh, 2007, 2012; Scollo, 2011). Propositions can be formulated eventually more abstractly as cultural premises, which are statements representing participants' beliefs and values about communication conduct, as represented both in their conduct and as a basis for that conduct. These premises provide interpretations of the typically unspoken beliefs and values of a speech community, highlighting them for further systematic thought, action, and design.

Cultural norms are statements, which, implicitly or explicitly, are granted legitimacy in a speech situation or community (Carbaugh, 2007). Norms vary in strength and influence (Hall, 1988/1989, 2005; Jackson, 1975). We build on Jackson's (1975) concepts of intensity and crystallization to help distinguish the variety of norms. Intensity refers to how strongly interactants feel about a norm—for example, do participants feel very strongly that a particular norm should or should not be followed or are they relatively indifferent about it? Crystallization, on the other hand, refers to the general agreement among participants about a particular norm. Thus, for example, there may be norms about which participants feel strongly, but, if there is a split between those who strongly support and those who strongly do not support the norm, we would say that the norm lacks crystallization because there is not general agreement about it. We formulate norms in a prototypical four-part formula. This formula addresses: (a) where the norm takes place (the setting or context), (b) who the person is engaging in the norm (or what role they are taking on), (c) what the strength or force is of this norm, and (d) what action is being accomplished (Carbaugh, 2007). Formulating norms in this way makes explicit participants' expectations for how an interaction within this type of situation should unfold and provides a basis for comparison with other communication situations, such as those involving in-car systems in other cultures.

In order to conceptualize the interactive field in which norms are used, we employ the concepts communication situation and communication event from the ethnography of communication. The concept *communication situation* grounds our analyses in actual settings where communication is done, such as interacting with one's car. In this analysis, we treat the car—and interactions within and about it—as a communication situation. Communication situations can be analyzed in more detail as involving specific sequences of acts, or communication events. *Communication events* are "activities, or aspects of activities, that are directly governed by rules or norms for the use of speech" (Hymes, 1972, p. 56). The concept of communication event offers a way to understand smaller cultural sequences within the communication situation of the car, draws attention to ways participants go about doing each sequence, and

emphasizes that each is governed by cultural norms. Communication events are made up of communication acts, such as the giving of a directive, which is a type of communication act designed to get the hearer—in this case, the car—to do some action (Blum-Kulka, 1997; Ervin-Tripp, 1976; Ervin-Tripp, Guo, & Lampert, 1990; Fitch, 1994; Goodwin, 2006; Searle, 1990). We should note that in our analysis, we are using the concept communication act as it is used in the ethnography of communication—see the exchange among John Searle, Michelle Rosaldo, and Dell Hymes (in Carbaugh, 1990).

An analysis of directives as such can reveal users' considerations of face and politeness (Brown & Levinson, 1978; Goffman, 1967; Grice, 1975; Lakoff, 1973; Shahrokhi & Bidabadi, 2013). Goffman (1967) proposed that personal conduct during an interaction is "the combined effect of the rule of self-respect and the rule of considerateness" (p. 11). In other words, a concern for one's own face and for the face of one's interlocutors guides a person's actions. Building on this idea, Brown and Levinson (1978) used the notion of "politeness" to draw attention to the systematic reasons users will deviate from a basic principle of efficiency. Similarly, Grice's (1975) Cooperative Principle includes four maxims (quantity, quality, relevance, and manner) that provide a basis for understanding times when speakers may seem to violate one of these maxims (such as quantity, by giving more information than is necessary) in order to create meaning through inference. According to Lakoff (1973), speakers may in these cases of violation be following a politeness rule. Lakoff (1973) proposed "rules of pragmatic competence," including that one ought to "be clear" and "be polite" (p. 296). Her first rule of clarity drew on Grice's maxims. Lakoff's (1973) second politeness rule includes three sub-rules that one ought to not impose, to give options, and to make the addressee feel good (p. 298). Blum-Kulka (1997) explained the connection between an analysis of directives and issues of face and politeness, noting that, "All types of social control acts impinge on the recipient's freedom of action and constitute a threat to face; therefore, politeness becomes a major consideration in the choice of mode of performance" (p. 142). Thus, the concepts of face and politeness form a basis for our formulation of norms of communication with an in-car system. We link the idea of a communication act (giving a directive) with the larger interactional sequence of which it is a part (a directive sequence) and wider cultural norms users assume for facework and politeness, for proper enactment in and with the car. This theory and methodology, as adapted for the study of in-car communication, is more fully detailed elsewhere (Carbaugh et al., 2012).

## Data Collection

### *Participants*

This research was conducted in western Massachusetts in an area that contained primarily rural roads, as well as some urban and suburban driving. The roads in this area do not generally experience a high volume of traffic. There were a total of 26 participants (14 female and 12 male). Nineteen of the participants were originally

from the northeastern United States. Participants ranged in age from 26 to 64 years old and had between 1 and 48 years of driving experience. The average income of participants was around $65,000 (US dollars). Participants spent on average around nine hours driving per week. Nineteen of the participants were smartphone users. Participants were recruited through flyers posted in public areas or through snowball sampling procedures as a result of direct contact with researchers.

### *Technology: Multimodal Capacities*

For this study, we equipped participants' own cars with an infotainment interface application with multimodal capabilities. Previous research has suggested that users tend to prefer computer interfaces that enable multiple modalities, including voice and touch (Oviatt, 2003). These multimodal interfaces are often viewed as more flexible, reliable, and efficient and as offering "a more 'human' way of interacting with computers, by means of speech, gestures or other modalities" (Dumas, Lalanne, & Oviatt, 2009). In our research, the driver interacted with the interface via a touch screen tablet PC, which we attached to the dashboard of the vehicle. Although the system was only temporarily installed in the drivers' cars, they were informed that the future goal of car manufacturers was to install this system directly into new cars prior to sale. There were a total of four researchers participating directly in data collection, who alternated roles in the car—two men and two women. During the drive, one researcher sat in the front seat of the car, and another researcher sat in the rear seat. The researcher in the back seat used a laptop that was connected to the system to respond to the driver's spoken requests. In this way, the researcher in the back seat took over the role of a speech recognizer—the researcher in this role is referred to as the "wizard" because he or she controlled the system behind the scenes in a manner similar to the wizard in the popular novel, *The Wonderful Wizard of Oz* (Baum, 1900). This "wizard" technique is widely used by researchers studying human-machine interaction (Dahlbäck, Jönsson, & Ahrenberg, 1993; Fraser & Gilbert, 1991; Jokinen & McTear, 2009; Passonneau, Epstein, Ligorio, & Gordon, 2011; Rieser & Lemon, 2008). It is employed because current voice recognition technology cannot yet support full natural-language processing. Fraser and Gilbert (1991) provided an overview of the requirements and different variables that play a role in Wizard of Oz experiments. They explained, "The designer is caught in a vicious circle—it is necessary to know the characteristics of dialogues between people and automata in order to be able to build the system, but it is impossible to know what such dialogues would be like until such a system has been built" (Fraser & Gilbert, 1991, p. 81). With the Wizard of Oz set-up, participants perceive that they are interacting with a computerized system with speech recognition capabilities, while in fact voice recognition technology is replaced by the researcher in the role of the wizard (Winter et al., 2013). During the driving sessions, drivers were told that the researcher in the back seat was overseeing the recording of the session, not that he or she was controlling the system. In this way, 25 of the 26 participants were unaware that the system was controlled by a researcher-wizard and did not give any indication of knowing the "machine" was being guided by a person;

the one participant who was aware was a pilot participant, the first in the study. There were also two small audio speakers connected to the wizard's laptop, which were used to broadcast the system's spoken utterances in a female voice, as well as to play the music that the driver requested and to broadcast the recipient's response when a call was placed. The entire session was recorded using three small cameras mounted at different angles in the car.

### *Driving Session*

Upon first meeting the research team, participants were asked to sign a consent form, which introduced them to the project. If a participant had a smartphone, the researchers uploaded his or her contacts and music content to the system so that he or she could access them during the drive. Participants were then introduced to the capabilities of the speech interface through the tablet, with which they were invited to experiment. Following this, they engaged in an off-road (i.e., parking lot) test drive, before engaging in the on-road driving session. Participants were invited to drive a route that they felt comfortable with, and researchers suggested a route if drivers were unsure of where to go. Although participants could ask questions of the researchers during the initial off-road test, they were asked not to speak to the researchers during the on-road session. About halfway through the on-road driving session, the driver was asked to park the car for a brief, semi-structured, mid-session interview. At this point, the researcher sitting in the front seat would ask the driver about his or her impressions of the system thus far, answer any questions that the driver might have, and introduce the driver to any functions of the system that the driver may not yet have noticed. After the complete on-road driving session, the researcher in the front seat would conduct a second longer in-car interview session. During this interview, the researcher would ask the driver about aspects of the driving session, such as what he or she felt did or did not work well, as well as his or her specific needs and desires, for example preferred sequencing of tasks. The overall session with each participant lasted approximately one and a half to two hours. After the participant was dismissed, the research team met for a debriefing session to discuss the driving session, possible analyses, and methodological matters.

## Data Analysis

Our analysis seeks to explore the characteristics of the communication situation when a driver interacts with his or her in-car speech enabled system and to identify communication norms and cultural premises that are active in order (a) to understand such dynamics and (b) to develop suggestions for the design of future technology. We begin with an overview of the directive sequences that we analyzed. We then explore the phenomenon we found of *non-task talk* that occurred during directive sequences. We focus on dynamics of turn-taking and functions of non-task talk, as well as participants' own reflections. These

examinations are the basis for formulating norms of interacting with an in-car speech system—in other words, statements about conduct that resonate with participants' understandings of how this interaction should take place—which will explicate participant expectations, provide a basis for comparison with other cultures, and enable the application of this knowledge to the design of future systems.

### *Directives*

Participants would initiate a speech interaction by touching the microphone button on the tablet screen. Touching the button would cause a chime or "ding" to sound, and the microphone button would become illuminated with a green light to indicate that the system was ready to receive verbal input. Participants would then speak to the system in order to accomplish one of the possible tasks of listening to the radio, listening to their music, or making a phone call. We analyze these participant utterances as directives. As Goodwin (2006) explained, directives are "utterances designed to get someone else to do something" (p. 515). Searle (1990) argued that speech acts varied by three primary dimensions: by the expressed psychological state of the speaker, by the point or purpose of the act, and by the fit of the words to the world. In the case of a directive, the act expresses something that the speaker wants to happen, and the point or purpose of the act is to get the hearer-car to do something. The speaker is attempting to use words to create a new state of affairs.

Researchers have examined directives in terms of directive sequences (Fitch, 1994; Goodwin, 2006), which we found in our data. Below is an example of one such directive sequence. It occurred at the beginning of Participant 23's on-road drive, approximately two and a half minutes after he had pulled onto the highway.

Instance A
Participant 23
21:47

| | | |
|---|---|---|
| 1 | Participant: | ((touches microphone button and system dings)) My music. |
| 2 | System: | What kind of music would you like to hear? |
| 3 | Participant: | Artists. |
| 4 | System: | What artist would you like? |
| 5 | Participant: | LMFAO. |
| 6 | System: | One moment please. ((2)) Playing lmfao or select an album. |
| 7 | | ((music plays)) |

Directive sequences varied in length and complexity. In reviewing our data, we noted that at times participants produced talk during directive sequences that did not seem directly relevant to accomplishing a system task. This led us to question the form and function of this type of non-task talk in interaction with the in-car system.

*Non-Task Talk in Directive Sequences*

In this section, we present the results of our analysis of 79 instances of non-task talk in directive sequences. These instances were produced by 15 participants. We use the term *task* to refer to the system's action of making a phone call, playing a radio station, or playing a song. By *non-task talk*, we mean participant talk within the directive sequences that was not directly connected to telling the system what to do. In other words, this talk was not part of the minimum requisite information for the system to accomplish one of its tasks. There is an important difference here between what another person could understand versus what a speech enabled system can understand. Furthermore, there is a distinction to be drawn in the analysis of human-computer interaction between speech that is formulated with an expectation of response and that which is not. We apply this distinction here to *task talk*, which we suggest is formulated as hearable by, and relevant to, the system's ability to accomplish the task at hand, and *non-task talk*, which, while it may occupy a conversational position that in human-human interaction would necessarily inform the next speaker's turn, does not do so here, as the system is not assumed by the speaker to be "listening" or equipped to understand the utterance.

We want to emphasize that the 79 instances of non-task talk analyzed here all addressed the system, rather than the two researchers in the car. Many of these instances included terms of address that selected the in-car system, rather than the researchers, as recipients. We identified instances when participants addressed their utterances to the system by finding when participants used the pronoun "you" following a system utterance, for example when drivers thanked the system or encouraged the system. We also understood the system to be the intended recipient of the utterance when a participant said please to the system immediately after a directive or when a participant answered a question posed by the system. However, in future research, it would also be interesting to consider instances of non-task talk that appeared less obviously addressed to the system, as well as to consider the role that researchers played as audience members or intended overhearers of non-task talk (Goffman, 1981; Gordon, 2013).

Following the completion of data collection, recordings were consulted for instances of non-task talk. Some participants produced only one instance, while others produced many. Directive sequences that contained non-task talk were transcribed and analyzed in terms of sequencing and functions. The mid- and post-driving session interviews were consulted for instances when the researcher in the front seat asked participants to reflect back on what they had said when interacting with the system. These reflections informed analysis of the transcribed instances, and together, both transcribed instances and interview data were the basis for formulating cultural norms and premises of communication.

Line 5 in Instance B below is an example of non-task talk that occurred during Participant 7's off-road testing session. In this instance, Participant 7 expressed appreciation for the system having returned to the "home" screen as she had requested.

Instance B
Participant 7
25:35

| | | |
|---|---|---|
| 1 | Participant: | Can you go back to the original? ((turns to the researcher)) Can I |
| 2 | | give it a name? 2 ((system changes to Home screen)) |
| 3 | Participant: | Oh! |
| 4 | System: | OK. |
| →5 | Participant: | That was good! Good for you, System! Oh, OK, I'm going to call you Denise. |

Participant 7's statement in Line 5 was addressed to the system, which she referred to as "you," "System," and "Denise." It was an instance of non-task talk in that it was not directly related to giving the system information so that it could accomplish a task (because the system had already accomplished its task). Instead, this particular instance of non-task talk appeared to serve relational functions of congratulations.

Non-task talk also sometimes occurred throughout a directive sequence and not just at the end. The instance below occurred after the mid-session interview, about 30 seconds after the participant had started to drive again. In it, Participant 18 responded to the system's request for him to "Hold on" with "You bet." The system did not require information at that point as it undertook the task of playing a jazz station, until it asked the question, "What music station do you want to hear?," at which point Participant 18 provided the needed information (task talk) of "Real jazz" in Line 5. Similarly to Participant 7, Participant 18 expressed appreciation in Line 7 when the system accomplished the task of playing 67 XM Real Jazz, even though, at that point, the task was accomplished and no more information was required.

Instance C
Participant 18
51:35

| | | |
|---|---|---|
| 1 | Participant: | ((touches microphone button and system dings)) Uh, jazz. |
| 2 | System: | Hold on. |
| →3 | Participant: | You bet. |
| 4 | System: | What music station do you want to hear? |
| 5 | Participant: | Uh, what was it called? Real jazz? |
| 6 | System: | Tuning radio to 67 XM real jazz. |
| →7 | Participant: | Thank you. |

Given that the non-task talk was not required for the system to accomplish its tasks, the question arises of what role this talk serves in the interaction and how this connects to participants' understandings of the communication situation. Here we draw further on cultural discourse analysis to analyze this non-task talk in terms of some of its key characteristics, including turn-taking and functions, and to explore how participants themselves accounted for their talk. We then use this analysis as a basis for formulating two competing norms of communication that appear to be active when participants in this context talk with and about their in-car speech enabled system.

*Turn-taking*

One notable characteristic of the non-task talk demonstrated in the data is that participants seem to use this non-task talk to coordinate turn-taking in interaction with the system. For example, in some cases, participants appeared to employ non-task talk to indicate that the system was not starting its turn at the appropriate time—either it was speaking too quickly or not quickly enough. For example, in Lines 2 and 10 of Instance D, as follows, which took place after the mid-session interview about three minutes after the participant had started to drive again, Participant 9 implied that the system had taken too long to respond.

Instance D
Participant 9
50:13

| | | |
|---|---|---|
| 1 | Participant: | ((touches microphone button and system dings)) Let's listen to |
| →2 | | WNNZ. ((4)) Come on, what's taking you so long. |
| 3 | | ((Researchers indicate that the system needs a restart and refresh the system.)) |
| 4 | Participant: | ((touches button and system dings)) Let's listen to WNNZ. ((spoken slowly)) |
| 5 | System: | Please wait. |
| 6 | Participant: | OK |
| 7 | | ((10)) |
| 8 | System: | Just a second. |
| 9 | | ((8)) |
| →10 | Participant: | That's a second. |
| 11 | System: | Please review your station. It may not be valid. |
| 12 | Participant: | Why not? ((5)) WNNZ ((spoken slowly with emphasis)) 91.7 = |
| 13 | System: | = Pardon? |
| 14 | Participant: | 91.7 |
| 15 | System: | Tuning radio to… |

In Line 2, Participant 9 questioned why the system was taking a long time. This comment did not provide the system with needed information for completing its task, and the system did not respond to Participant 9's statement. However, at this point the reseachers did restart the system, suggesting that they may have been responding to Participant 9's utterance as possibly expressing concern or frustration. In Line 10, Participant 9 told the system, "That's a second" after approximately eight seconds of silence, implying that the system had taken too long. Again, this did not include information necessary to the task at hand that the system could respond to.

Another example of non-task talk being used to coordinate turn-taking occurred in the following instance, which took place after the mid-session interview about a minute after the participant had started to drive again. In this instance in Line 4, Participant 14 told the system, "I'm thinking, shut up," indicating that the system had not allowed enough time between turns.

Instance E
Participant 14
1:06:39

| | | |
|---|---|---|
| 1 | Participant: | ((touches microphone button and system dings)) |
| 2 | | ((8)) |
| 3 | System: | Pardon? |
| →4 | Participant: | I'm thinking, shut up. |
| 5 | System: | Could you repeat that please? |
| 6 | Participant: | Um, play Rage Against the Machine—actually play Tool. |
| 7 | System: | Just a second. ((7)) I am sorry. I am unable to find the music item |
| 8 | | that you are looking for. ((Participant presses the End button.)) |

While Participant 14's utterance in Line 4 could be understood as a directive in that he was telling the system what to do, this utterance did not provide information that the system needed to accomplish its task, and there did not appear to be an expectation of reciprocity on the part of Participant 14 that the system would respond to this utterance. The participant did not initiate a repair sequence when the system asked him "Could you repeat that please?" after he had told it to "shut up," but instead, continued with the interaction and asked the system to play music by a particular artist. This suggests that the user did not have an expectation of response or system comprehension in the formulation of this utterance and was, therefore, not oriented to the accomplishment of the task.

Researchers have found that delays in responding to requests in interactions between people often precede rejections, whereas grantings are not delayed in this way because of a preference principle for avoiding or minimizing disagreements, disconfirmations, and rejections (Pomerantz & Heritage, 2013). Thus, participants who responded to system delays with non-task talk may have been interpreting the system's delay as an inability to accomplish the directed task—in other words, a rejection of the directive. This interpretation was apparent in the instance below when Participant 13 said, "No you don't like that one?" in Lines 1–2, following approximately nine seconds of silence. This instance occurred about 20 minutes after the participant had started the on-road driving session and approximately 10 minutes prior to the mid-session interview. Participant 13's question about liking her directive did not provide information that could help the system accomplish its task. In Line 9, she called the system a "goofball," which was again not something that the system required in order to act, though it could be hearable as an evaluation of the system's performance in task accomplishment. In Lines 12 and 16, Participant 13 encouraged the system, and in Line 18, she expressed appreciation for the system's action, similar to Participants 7 and 18 above. Again this final statement was not related to telling the system what to do since the task had already been accomplished.

Instance F
Participant 13
39:53

| | | |
|---|---|---|
| →1 | Participant: | ((touches microphone button and system dings)) Play XFM. ((9)) |
| 2 | | No you don't like that one? |
| 3 | System: | Please wait. |
| 4 | Participant: | Ah |
| 5 | | ((9)) |
| 6 | System: | Please review your station it may not be valid. |
| 7 | Participant: | ((laughs)) Ok mm. ((5)) Play my music. |
| 8 | System: | Please let me know what you want. |
| →9 | Participant: | Ok. You are a goofball. ((4)) |
| 10 | | I'm going to [call you-] |
| 11 | System: | [Wait a] moment. What kind of music would you like to hear? |
| →12 | Participant: | My music. You can do it. Come on. |
| 13 | System: | What song do you want to hear? |
| 14 | Participant: | Hmm. Play Creep. |
| 15 | System: | Hold on. |
| →16 | Participant: | You can do it baby. |
| 17 | System: | Playing Creep. ((music plays)) |
| →18 | Participant: | There you go. |

In this case, Participant 13 indicated in Lines 1–2 that she believed the system was unable to accomplish the task she had given it of playing the radio by interpreting the silence as the system "not liking" her directive. Her non-task talk could have been an attempt to determine if she should reformulate her directive so that it could be accomplished. However, while a human interactant may have been able to interpret her question about "liking" a directive as an attempt to obtain more information about why there was a delay, the system could not, and her non-task talk did not provide additional information that would have allowed the system to accomplish its task.

*Functions*

In addition to coordinating turn-taking, there is evidence in the data of participants employing non-task talk to accomplish several functions. For example, participants used the non-task talk to *encourage or support* the system. This function is evident in Participant 13's statement, "You can do it baby" in Instance F, which was said after she had asked the system to play music, and it had responded with "Hold on." In addition, the non-task talk in these instances often consisted of an *evaluation* of the system's actions. Frequently this evaluation was positive, as when participants acknowledged that something had been done correctly—for example, "That was good! Good for you, System" said by Participant 7 in Instance B at the end of the sequence after the system had accomplished its task. The function of evaluating the system's actions seems to align with the participants' understanding that they were participating in the driving session in order to test a prototype, as was explained to them before the driving session.

Evidence that participants oriented to the system as something that they had been asked to test can also be found in the explanations that participants gave to the friends and family members whom they called during the session. Several participants asked the people they called about the sound quality of the call. They stated that they were driving with researchers in their car and testing a new system. For example, when Participant 5's mother answered the phone and commented that Participant 5 had appeared on the caller ID as "anonymous," Participant 5 responded, "Yes, I'm testing out an in-car system thing. So, that would be why." During his driving session, Participant 17 left a message on the voicemail of one of the people whom he called, saying, "Hi [name]. I'm calling you from my car on a phone—on a computer. So I'm curious to know what the sound quality is like." In response to this explanation, call recipients would then provide feedback about how the driver sounded on the phone—Participant 8's mother told her that it was difficult to hear her, and said, "So you might give that feedback to your researchers." Thus, drivers viewed their role to be that of testing the in-car system. Their non-task talk can be understood as further *constructing their relationship* with the system as one who has been asked to evaluate it.

In this context, participants generally did not address the system angrily or yell at the system. This may have been largely due to the fact that the researchers were also present in the car, and participants did not feel comfortable speaking negatively in front of the researchers, as this might imply a negative evaluation of the researchers' system. Thus, while participants were instructed not to directly address researchers during the driving session, it is possible that system-addressed non-task talk overheard by the researchers could also have been intended to function to *construct participant identities* and *construct relationships with the researchers*—this would be similar to practices that Tannen (2004) and Gordon (2013) studied of pets and technology used to mediate interactions between people and perform identities, such as "compliant study participants" (Gordon, 2013, p. 306). Just as participants were constructing their relationship with the system as one who has been asked to evaluate it, they were likely, in addition, constructing an identity of evaluator for the researchers' benefit by showing themselves to be cooperative and helpful evaluators. We focus here on participants' understandings of their relationship with the system, but future analyses could explore how participants use speech enabled technology to negotiate relationships with other people, such as researchers or others in their car, and to create identities in interaction.

In two of the instances included above, participants did speak critically to the system. Participant 9 in Instance D asked the system, "what's taking you so long," and Participant 14 in Instance E told the system to "shut up." Interestingly, both of these instances took place after the mid-session interview when participants were encouraged by the researchers to talk to the system as they would do if they had been alone. These instances may indicate that participants would likely speak to a system in their car in both a positive and a negative manner, if they were not being observed by researchers. Although Participant 13's statements to the system were generally positive during the session, she claimed during the mid-session interview that if she were to

have this system in her car on a daily basis, she would speak to it both positively and negatively. When the interviewer observed that during the driving session, Participant 13 had apologized to the system, Participant 13 explained,

> Well I would also yell at it. I mean I would treat it like a person. I'm generally nice and polite, but then, if it was being smart-ass, I might yell back at it. But I wouldn't, you know, I wouldn't get mad. Not really.

As this quote suggests, we must be careful not to overestimate the emotional force of critical remarks directed at the system, since at least for some participants while they might yell at it, they "wouldn't get mad. Not really." This comment is further suggestive of the potentially playful nature of some in-car interactions with the system.



*Participant reflections on interacting with the system*

In the tradition of CuDA, we examined not only communication occurring with the in-car system but also communication about the system—such as Participant 13's comment about speaking negatively above—for evidence of cultural norms and premises being used in interaction, in particular conceptualizations of the system as interactional partner. The mid- and post-session interviews provided useful sources of communication about the system. During these interviews, several participants expressed that they felt the need to produce some non-task talk to the system. Some participants described producing non-task talk to the system as "tempting" or "natural." For example, Participant 9 said during the interview that he thought "it's natural to sort of, converse with it, you know, as I would a person." Participants observed that they talked to the system without really meaning to or thinking about it. When the interviewer commented during the interview at the end of the session that Participant 22 had said "all right" in response to the system's saying "Wait a moment," the participant noted that she often speaks to systems that produce speech, such as her GPS, in this way. She explained,

> I always do that to these. I almost like jokingly, like "Oh OK, thank you." Like whenever it says, "You have arrived." I'm like "All right! Thank you!"…I mean it's so anthropomorphized that like, I don't know. I'm used to it, I guess.

When the interviewer noted during the mid-session interview that Participant 15 had said "thank you" to the system and asked if that is the way he would interact with the system, Participant 15 commented,

> When you're talking to somebody and they do something, you say thank you….I kind of thought to myself after I said it, I was like, "Jesus, dude. It's not like she has a real name."…I guess it just popped out. I was like "Thank you." It was, it just seemed like it was a natural reaction to asking somebody to do something for me.

Participant 15's mentioning of the characteristic of having a "real name" reinforces Participant 22's comment that the system has some human-like qualities. It is also significant that Participant 15 refers to the system with the pronoun "she," rather than "it," as this pronoun would also appear to indicate an orientation to the system as a

female person rather than as a computer. In this way, some participants found it natural to interact with the system using non-task talk, and they appeared to view the system to a certain extent as if "she" were a female person.

The key terms "nice" or "polite" were drawn on by participants as a way of describing and understanding their behavior when asked about it during interviews. During the mid-session interview, when the interviewer commented to Participant 20 that he had said "please" to the system and asked if it would be his preference to talk in this way with the system when using it, Participant 20 reflected, "Yeah, I mean, why not? It's helping you out. You might as well be nice to it." Similarly, when the interviewer noted that Participant 13 had apologized to the system, Participant 13 claimed, "Well, it's technology. That doesn't mean you don't have to be polite." At the end of the session, during the final interview, the interviewer observed to Participant 18 that he had inserted some phrases into the interaction with the system, such as "oh, great, thanks" or "oh, sure." The interviewer asked if Participant 18 inserted these types of comments for fun, and Participant 18 observed,

> No, I do it subconsciously. I'm just trying to be polite to the system….You got to understand from a social science standpoint, I think we are a very impolite culture, and I'm trying to change that so, it sort of creeps in.

Thus, one way that participants had of describing their use of non-task talk was as being "nice" or "polite." Furthermore, some participants even admitted that they would feel uncomfortable not talking to the system in this way. At the beginning of the mid-session interview, the interviewer asked Participant 7 what her experience had been like so far interacting with the system. Participant 7 said that she had never interacted with this type of a system before in which she had to "give orders," and she did not really like the feeling that she was giving orders, or, as she also called it, "barking things at it." In other words, engaging in non-task talk with the system was, for Participant 7, a way to avoid feeling that she was giving the system orders. These interactional dynamics exhibit a participant stance of being a good, polite conversational partner—one both self-reflective and considerate of the other/system.

There were participants, however, who did not share Participant 7's discomfort. A key idea that emerged in the data when those who did not produce non-task talk were asked about their interactions was that the system was just a "machine." For example, when asked during the interview at the end of the session about how she relates to her car and if she would prefer her car to speak to her as another person would, Participant 5 said that she thinks of her car as a "tool," and she does not see it as something that she would relate to as she would another person. She said,

> I think I see it more as a tool. I do not name my car, and you know, that type of thing….I think of it still, still a functional tool, that, you know, I'm using for directions, or to call people, or radio…it—it's a machine, technology.

Similarly, during the mid-session interview, when the interviewer asked Participant 16 about a complaint that he had because he felt the system was interrupting him, the participant observed that he felt that these interruptions emphasized for him the

feeling that he was interacting with a computer and not another person. He went on to note, however, that this was "fine" because "I don't need her to be my friend." He explained,

> I don't need this to be human. I just need it to recognize what I tell it to do and do it. I don't need any of the, "Ok, see you later." You know, I don't take any like comfort from it having a couple of pre-programmed personable phrases….I'd rather it just was straight and to the point.

Thus, although for some participants, it seemed difficult to not engage in what they perceived to be a "nice" manner with the system by using non-task talk, others recognized it as a "machine" and were not concerned with engaging in this type of talk. This distinction appears to be linked to a view of the system as either having some person-like qualities or as being non-human.



## Findings: Competing Cultural Norms Active During Driving Sessions

Based on the above analysis, we suggest that a key finding of this research is the presence of at least two competing cultural norms in participants' communication with and about the in-car voice activated system. We also propose that these two norms are based on cultural premises about personhood—or how participants understand being a person in this context. One cultural norm that we found to be active in these types of interactional sequences could be formulated as: *when interacting with another, or a system participant (i.e., the in-car system), if this is done properly, one should engage in some non-task talk with the in-car system*. In other words, the user can be understood to act so as to maintain the "face" of the in-car system by offering encouragement, such as "You can do it baby," acknowledging when something has been done correctly, or calling the system by a personal name, such as "Denise." This encouragement, acknowledgement, and naming could be understood as instances of what Brown and Levinson (1978) called positive politeness strategies, in that they indicate some approval of the system or its actions and emphasize intimacy with the system. In talking in this way, participants flout Grice's (1975) maxims by including more in their utterances than is required at that point in the interaction.

Participant 20 noted during the mid-session interview that the system "talks to you a little bit, and, you know, takes on some humanish, you know, characteristics when she's talking." For Participant 20, being spoken to in a "humanish" way seems to merit a human response regardless of the material embodiment of that speech. Thus, we suggest that the above norm is based on a cultural premise of personhood that we formulated as: *social interaction is based on an agent's (i.e., the system's) way of communicating rather than its status as a human*. This premise underlies drivers' non-task talk to the system, and also their reference to the system with the female pronoun, "she." Nass and Brave (2005) observed that users will tend to assign gender to a voice, even if the voice is clearly mechanically produced, responding to synthetic voices "as if they reflected the biological and sociocultural realities of sex and gender" (p. 15). In this way, for these participants, treating the system as a female person is connected to its speaking in a female human way, regardless of whether it is a person or not.

Although some participants treated the system similar to a human interactant, in spite of its status of non-human, our data indicate that even these participants did still recognize that "she" is not human. For example, some participants spoke negatively toward the system, although Participant 13 observed that even if she were to yell at the system, she would not "really" be mad. This would seem to indicate that even though some participants are interacting with the system "as if it were human" and appear to demonstrate concern for the system's face through routine uses of politeness, the absence of expectation of response from the system to those forms of interaction free the user from serious concern over potentially face-threatening actions. This enables participants to "yell at the system," a risk that human interactants would be unlikely to take given the possible social consequences.

Participants' reflections on their interactions with the system during the sessions indicate a lack of crystallization, or agreement among participants, of a norm of engaging in some non-task talk with the in-car system. While participants, such as Participant 7, expressed feeling this norm strongly, to the point of discomfort if they did not engage the system with some non-task talk, many participants used relatively little to no non-task talk with the system, and when asked about their view of talking with the system, noted that it was just a "machine" or "tool." Given the novelty of this type of system and the fact that participants had generally not interacted with many other systems like it, it seems likely that while a norm regarding being "nice" or "polite" when interacting with other humans by engaging in non-task talk is highly crystallized, the norm of being "nice" when talking with a non-human other is less crystallized. The lack of non-task talk among some participants suggests a competing cultural norm that we formulated as: *when interacting with another that is non-human (i.e., the in-car system), if this is done properly, the interaction should be efficient and one does not need to engage in non-task talk*. This lack of norm crystallization appears to be present even within the talk of individual participants, such as Participant 16, who uses both "her" and "it" to refer to the system in the statements cited above. He observes, "I don't need *her* to be my friend," and then later claims, "I just need *it* to recognize what I tell *it* to do and do it." In the case of the second norm, an underlying cultural premise is again involved: *social interaction is based on an agent's (i.e., the system's) status as a human rather than its way of communicating*. While for some participants the "anthropomorphized" system warrants a certain way of interacting, for others the system is just a tool, regardless of how it speaks, and it does not need to be talked to as if it were a human.

## Implications for Speech Enabled Interface Design and Suggestions for Future Research

In this section, we address the second goal of our analysis—to draw on findings regarding cultural norms and premises formulated above to suggest improvements for future speech enabled technology. First, we propose that designers draw on our

findings to make better use of the fact that some users make supportive and evaluative comments. Systems could be designed in the future to monitor for non-task talk and attempt to create a "log of complaints" or "feedback log" that designers, or the system itself, could use to modify the system more largely or tailor the system more specifically to the users' preferences. Designers could thus benefit from the ongoing feedback that some users already provide to the system verbally.

In addition, the prompts and dialog flow of a system could be adapted to respond to a driver's non-task talk by producing non-task utterances, which could create for the driver a sense of having a relationship with his or her system. Developers might even define certain characteristics or personalities for systems, using a variety of voice qualities, and perhaps even give systems personal backgrounds and likes or dislikes, which could further support a driver's treatment of the system as a human interactant. At the same time, developers should also be aware of the competing preference of those drivers who would rather that their system did not respond to or produce non-task utterances; systems should also be developed to accommodate this manner of interacting. The system could monitor user non-task talk and be programmed to produce more non-task talk if the user produced this type of talk, or be programmed not to produce this type of talk if the user did not produce it. The system might also pay attention to those occasions when participants respond (either positively or negatively) to non-task talk that it produces in order to see to what degree and when in the interaction these utterances are appreciated or not.

The competing norms that we have identified here also suggest an area of potential future change, for which designers could prepare. Current drivers, such as our participants, who have not frequently been exposed to this type of technology may be more likely to view "speaking" as indicative of a certain status of agency or personhood. However, in the future, drivers who have more experience with speech enabled systems may tend to view these systems as simply another form of technology, rather than a conversational interactant with whom one might develop a relationship. Developers could work to create flexible systems that might accommodate this type of a change, even in the same user over time. Thus, the system's monitoring of non-task talk would continue over time, and the system might gradually decrease the amount that it produced, if it found that a driver was decreasing the amount of non-task talk that he or she was producing. Alternatively, the system could also gradually produce more if the driver were producing more. This monitoring could also help systems to be more adaptable to different situational or cultural contexts in which different amounts of non-task talk may be deemed appropriate (for example, the situation would likely be different if one were driving his or her spouse in the car versus his or her boss). As mentioned previously, a potential area of future study could focus on how individuals employ interaction with speech enabled technology in order to construct a particular identity for themselves and different types of

relationships with other riders in their cars. Findings from this type of research could inform when and to what extent systems are programmed to produce non-task talk.

The ideas we present here may also be relevant for designers working on other forms of technology that make use of a dialogue system, such as smartphones or tablets, and, perhaps in the future, smart homes and smart businesses. Jurafsky and Martin (2009) emphasized the importance of such work on language processing and human-machine interaction when they noted:

> The critical connection between language and thought has placed speech and language processing technology at the center of debate over intelligent machines. Furthermore, research on how people interact with complex media indicates that speech and language processing technology will be critical in the development of future technologies. (p. 15)

As speech enabled technology becomes more prevalent, individuals will increasingly be forced to consider what type of interactional partner a computer represents. Our findings support past research indicating that people will at times respond to a machine that talks as if it were another person, regardless of being aware that the system is a machine. The interactions and perhaps even relationships that those in the future have with their speaking machines may have repercussions, in turn, for the way in which they understand interactions and relationships with other people, as the quality of being able to speak becomes less exclusively connected with being human and the readiness and willingness of individuals to interact with their computers as people changes.

## Conclusion

In this analysis, we have explored the phenomenon of participants producing non-task talk while engaging in directive sequences. We have proposed two competing cultural norms that we suggest are active in this practice. While some participants appear to engage in conduct that could be seen to protect the face of the in-car speech system, others treat the system as a mechanical "tool" with few face concerns, that is merely a device to accomplish a task. We have connected these communicative norms to cultural premises of personhood and differences in how being able to speak is viewed as aligned with an interactant's status. Our future research will explore other communication events, sequences, norms, and premises in our data from the north-eastern United States, as well as a corpus of data we collected in Shanghai, China. We expect task and non-task talk to have some similarities and differences relative to our findings here.

## Acknowledgments

The authors would like to thank Ute Winter, Research Specialist of the Human-Machine Interface at General Motors Research and Development, Herzeliya, Israel for

her collaboration on this project. The authors also gratefully acknowledge their colleagues who are part of the field research team in China: Libin Hang, Associate Professor of Communication at Donghua University, Shanghai, China and Pei Wang who is a Research Specialist of the Human-Machine Interface at General Motors in Shanghai, China.

## Funding

This work was supported by General Motors, Donal Carbaugh, Principal Investigator in collaboration with Ute Winter, Research Specialist of the Human-Machine Interface at General Motors Research and Development, Herzeliya, Israel, Elizabeth Molina-Markham, Post-Doctoral Research Fellow, Brion van Over and Sunny Lie, Research Associates.



## References

Baum, L. F. (1900). *The wonderful wizard of oz*. New York, NY: George M. Hill Company.

Berry, M. (2009). The social and cultural realization of diversity: An interview with Donal Carbaugh. *Language and Intercultural Communication*, *9*, 230–241. doi:10.1080/14708470903203058

Blum-Kulka, S. (1997). *Dinner talk: Cultural patterns of sociability and socialization in family discourse*. Mahwah, NJ: Lawrence Erlbaum Associates.

Brown, P., & Levinson, S. (1978). Universals in language usage: Politeness phenomena. In E. N. Goody (Ed.), *Questions and politeness: Strategies in social interaction* (pp. 56–289). Cambridge, United Kingdom: Cambridge University Press.

Carbaugh, D. (1990). *Cultural communication and intercultural contact*. Mahwah, NJ: Lawrence Erlbaum Associates.

Carbaugh, D. (2007). Cultural discourse analysis: Communication practices and intercultural encounters. *Journal of Intercultural Communication Research*, *36*(3), 167–182. doi:10.1080/17475750701737090

Carbaugh, D. (2012). A communication theory of culture. In A. Kurylo (Ed.), *Inter/Cultural communication: Representation and construction of culture* (pp. 69–87). Thousand Oaks, CA: SAGE.

Carbaugh, D., Molina-Markham, E., van Over, B., & Winter, U. (2012). Using communication research for cultural variability in human factor design. In N. Stanton (Ed.), *Advances in human aspects of road and rail transportation* (pp. 176–185). Boca Raton, FL: CRC Press.

Carbaugh, D., Winter, U., van Over, B., Molina-Markham, E., & Lie, S. (2013). Cultural analyses of in-car communication. *Journal of Applied Communication Research*, *41*(2), 195–201. doi:10.1080/00909882.2013.782422

Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993, January). *Wizard of Oz studies: Why and how*. Proceedings of the 1st international conference on intelligent user interfaces (pp. 193–200), Orlando, FL, ACM.

Dumas, B., Lalanne, D., & Oviatt, S. (2009). Multimodal interface: A survey of principles, models and frameworks. In D. Lalanne & J. Kohlas (Eds.), *Human machine interaction: Lecture notes in computer science* (pp. 3–26). Berlin/Heidelberg: Germany: LNCS-Springer-Verlag.

Ervin-Tripp, S. (1976). Is Sybil there? The structure of some American English directives. *Language in Society*, *5*(01), 25–66. doi:10.1017/S0047404500006849

Ervin-Tripp, S., Guo, J., & Lampert, M. (1990). Politeness and persuasion in children's control acts. *Journal of Pragmatics*, *14*, 307–331. doi:10.1016/0378-2166(90)90085-R

Fitch, K. L. (1994). A cross-cultural study of directive sequences and some implications for compliance-gaining research. *Communication Monographs*, *61*(3), 185–209. doi:10.1080/03637759409376333

Fraser, N. M., & Gilbert, G. N. (1991). Simulating speech systems. *Computer Speech & Language*, *5*, 81–99. doi:10.1016/0885-2308(91)90019-M

Friedman, B. (Ed.). (1997). *Human values and the design of computer technology*. New York, NY: Cambridge University Press.

Goffman, E. (1967). *Interaction ritual: Essays in face to face behavior*. Chicago, IL: Aldine.

Goffman, E. (1981). *Forms of talk*. Philadelphia, PA: University of Pennsylvania Press.

Goodwin, M. H. (2006). Participation, affect, and trajectory in family directive/response sequences. *Text & Talk—An Interdisciplinary Journal of Language, Discourse Communication Studies*, *26* (4–5), 515–543. doi:10.1515/TEXT.2006.021

Gordon, C. (2013). Beyond the observer's paradox: The audio-recorder as a resource for the display of identity. *Qualitative Research*, *13*(3), 299–317. doi:10.1177/1468794112442771

Grice, H. P. (1975). Logic in conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics: Speech acts 3* (pp. 41–58). New York, NY: Academic Press.

Haddington, P. (2010). Turn-taking for turntaking: Mobility, time, and action in the sequential organization of junction negotiations in cars. *Research on Language & Social Interaction*, *43* (4), 372–400. doi:10.1080/08351813.2010.518068

Haddington, P., & Keisanen, T. (2009). Location, mobility and the body as resources in selecting a route. *Journal of Pragmatics*, *41*, 1938–1961. doi:10.1016/j.pragma.2008.09.018

Hall, B. (1988/1989). Norms, action, and alignment: A discursive perspective. *Research on Language & Social Interaction*, *22*, 23–44. doi:10.1080/08351818809389296

Hall, B. (2005). *Among cultures: The challenge of communication*. New York, NY: Wadsworth.

Holtzblatt, K., & Beyer, H. R. (2014). Contextual design. In M. Soegaard & R. F. Dam (Eds.), *The encyclopedia of human-computer interaction* (2nd ed.). Aarhus, Denmark: The Interaction Design Foundation. Retrieved from https://www.interaction-design.org/encyclopedia/contextual_design.html

Hymes, D. (1962). The ethnography of speaking. In T. Gladwin & W. Sturtevant (Eds.), *Anthropology and human behavior* (pp. 13–52). Washington, DC: The Anthropological Society of Washington.

Hymes, D. (1972). Models for the interaction of language and social life. In J. J. Gumperz & D. Hymes (Eds.), *Directions in sociolinguistics: The ethnography of communication* (pp. 35–71). New York, NY: Basil Blackwell.

Jackson, J. M. (1975). Normative power and conflict potential. *Sociological Methods & Research*, *4*, 237–263. doi:10.1177/004912417500400205

Jokinen, K., & McTear, M. (2009). *Spoken dialogue systems*. San Rafael, CA: Morgan & Claypool.

Jurafsky, D., & Martin, J. H. (2009). *Speech and language processing: An introduction to natural language processing, speech recognition, and computational linguistics* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall.

Koppel, S., Charlton, J., Kopinathan, C., & Taranto, D. (2011). Are child occupants a significant source of driving distraction? *Accident Analysis & Prevention*, *43*(3), 1236–1244. doi:10.1016/j.aap.2011.01.005

Lakoff, R. (1973). The logic of politeness; or minding your p's and q's. *Papers from the ninth regional meeting* (pp. 292–305). Chicago, IL, Chicago Linguistic Society.

Laurier, E. (2005). Searching for a parking space. *Intellectica*, *41–42*(2–3), 101–115.

Laurier, E. (2011). Driving: Pre-cognition and driving. In T. Cresswell & P. Merriman (Eds.), *Geographies of mobilities: Practices, spaces, subjects* (pp. 69–82). Farnham and Burlington, VT: Ashgate.

Laurier, E., Brown, B., & Lorimer, H. (2007). *Habitable cars: The organisation of collective private transport: Full research report ESRC end of award report, RES-000-23-0758*. Swindon, UK: ESRC.

Laurier, E., Brown, B., & Lorimer, H. (2012). What it means to change lanes: Actions, emotions and wayfinding in the family car. *Semiotica*, *191*, 117–135.

Laurier, E., Lorimer, H., Brown, B., Jones, O., Juhlin, O., Noble, A.,…Weilenmann, A. (2008). Driving and "passengering": Notes on the ordinary organization of car travel. *Mobilities*, *3* (1), 1–23. doi:10.1080/17450100701797273

Mäkäräinen, M., Tiitola, J., & Konkka, K. (2001). How cultural needs affect user interface design? In M. Reed Little & L. Nigay (Eds.), *Engineering for human-computer interaction* (pp. 357–358). New York, NY: Springer.

Mao, J., Vredenburg, K., Smith, P., & Carey, T. (2005). The state of user-centered design practice. *Communications of the ACM*, *48*(3), 105–109. doi:10.1145/1047671

Nass, C. I., & Brave, S. (2005). *Wired for speech: How voice activates and advances the human-computer relationship*. Cambridge, MA: MIT press.

Nass, C. I., & Yen, C. (2010). *The man who lied to his laptop: What machines teach us about human relationships*. New York, NY: Current.

Oviatt, S. (2003). Advances in robust multimodal interface design. *IEEE Computer Graphics and Applications*, *23*(5), 62–68. doi:10.1109/MCG.2003.1231179

Passonneau, R. J., Epstein, S. L., Ligorio, T., & Gordon, J. (2011, June). *Embedded wizardry*. Proceedings of SIGDIAL 2011 conference (pp. 248–258). Portland, OR.

Pomerantz, A., & Heritage, J. (2013). Preference. In J. Sidnell & T. Stivers (Eds.), *Handbook of conversation analysis* (pp. 210–228). Cambridge, UK: Cambridge University Press.

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. New York, NY: Cambridge University Press.

Rieser, V., & Lemon, O. (2008, June). *Learning effective multimodal dialogue strategies from wizard-of-oz data: Bootstrapping and evaluation*. Proceedings of ACL (638–646). Columbus, OH.

Scollo, M. (2011). Cultural approaches to discourse analysis: A theoretical and methodological conversation with special focus on Donal Carbaugh's cultural discourse theory. *Journal of Multicultural Discourses*, *6*(1), 1–32. doi:10.1080/17447143.2010.536550

Searle, J. (1990). A classification of illocutionary acts. In D. Carbaugh (Ed.), *Cultural communication and intercultural contact* (pp. 349–372). Hillsdale, NJ: Lawrence Erlbaum.

Shahrokhi, M., & Bidabadi, F. S. (2013). An overview of politeness theories: Current status, future orientations. *American Journal of Linguistics*, *2*(2), 17–27.

Tannen, D. (2004). Talking the dog: Framing pets as interactional resources in family discourse. *Research on Language & Social Interaction*, *37*(4), 399–420. doi:10.1207/s15327973rlsi3704_1

Tsimhoni, O., Winter, U., & Grost, T. (2009, August). *Cultural considerations for the design of automotive speech applications*. Paper presented at the 17th World Congress on Ergonomics IEA, Beijing, China.

Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other*. New York, NY: Basic Books.

Winschiers, H., & Fendler, J. (2007). Assumptions considered harmful: The need to redefine usability. In N. Aykin (Ed.), *Usability and internationalization, Part I* (pp. 452–461). Mahwah, NJ: Lawrence Erlbaum Associations.

Winter, U., Shmueli, Y., & Grost, T. (2013, July). *Interaction styles in use of automotive interfaces*. In Proceedings of the Afeka AVIOS 2013 Speech Processing Conference, Tel Aviv, Israel.

Winter, U., Tsimhoni, O., & Grost, T. (2011, June). *Identifying cultural aspects in use of in-vehicle speech applications*. In Proceedings of the Afeka AVIOS 2011 Speech Processing Conference, Tel-Aviv, Israel.